

# Convergencia de Estimadores a Mínimo de Entropía Robustos: Aplicaciones en Instrumentación y al PDS

## *Convergence of Minimum-Entropy Robust Estimators: Applications in DSP and Instrumentation*

**José Ismael de la Rosa Vargas**

Universidad Autónoma de Zacatecas - Unidad Académica de Ingeniería Eléctrica,  
Lab. de Procesamiento Digital de Señales, Av. López Velarde No. 821, Col. Centro,  
C.P. 98068, Zacatecas, Zacatecas.  
ismaelrv@ieee.org

*Artículo recibido en junio 21, 2004; aceptado en noviembre 8, 2006*

### **Resumen**

En este trabajo de investigación nos proponemos continuar con la misma línea de investigaciones iniciadas por Pronzato y Thierry (Pronzato *et al*, 2000a), (Pronzato *et al*, 2000b), (Pronzato *et al*, 2001) las cuales se abordaron ya en los trabajos de De la Rosa y Fleury (De la Rosa *et al*, 2002), (De la Rosa *et al*, 2003) en un marco de trabajo perteneciente a la instrumentación, y en donde se establece un modelo estocástico para representar ciertas señales y para el cual se formulan ciertas hipótesis limitadas sobre la naturaleza del ruido o perturbaciones que afectan los sistemas bajo estudio. La utilización de estimadores robustos es importante, ya que los sistemas reales están expuestos a perturbaciones continuas que son de naturaleza desconocida, esto se ha experimentado en aplicaciones propias de la instrumentación médica, en procesos industriales, y en telecomunicaciones entre otros. Presentamos algunos resultados complementarios a los presentados por Pronzato y Thierry sobre la estimación robusta, tanto para modelos lineales como para modelos no lineales.

**Palabras clave:** Entropía, Simulación de tipo Monte-Carlo, Estimación No Paramétrica, Regresión, Estimación Robusta.

### **Abstract**

In this paper we propose to continue in the same research line initiated by Pronzato and Thierry (Pronzato *et al*, 2000a), (Pronzato *et al*, 2000b), (Pronzato *et al*, 2001), recent works inspired in the minimum-entropy estimation have been published by De la Rosa and Fleury (De la Rosa *et al*, 2002), (De la Rosa *et al*, 2003) in the instrumentation framework. An statistical model has been established to represent some instrumental signals, similarly, some limited hypothesis over such a model have been made. In fact, we assume limited knowledge of the noise or external perturbations distribution that interact into the system. The use of robust estimators in such situations is very helpful, since the real systems are always exposed to continuous perturbations of unknown nature. Some applications where the last is true are: medical instrumentation, industrial processes, in telecommunications among others. Some results of new minimum-entropy estimators for linear and nonlinear models are presented, such results complement those presented by Pronzato and Thierry.

**Keywords:** Entropy, Monte Carlo Simulation, Nonparametric Estimation, Regression, Robust Estimation.

## **1 Introducción y estado del arte**

Tanto en procesos industriales como en instrumentación médica, frecuentemente se realiza la adquisición de señales para monitoreo y análisis de los procesos de producción o en el caso de medicina para el monitoreo de pacientes. Cualquiera que sea el fin aplicativo en la industria, la medicina, etc., un problema ampliamente tratado por décadas es el hecho de considerar que las señales adquiridas están contaminadas por perturbaciones correlacionadas con el sistema bajo estudio (errores de redondeo, fallas en el equipo de adquisición) y otras perturbaciones externas.

El estimador clásico conocido mejor como estimador de Máximo de Verosimilitud (MV), minimiza la entropía de una versión muestra de la densidad de probabilidad de los errores ( $p(\mathbf{e})$ ) del sistema, cuando dicha distribución es desconocida el estimador de MV podría ser sub-óptimo, puesto que la distribución considerada podría ser la equivocada. En este caso, en la sección 3 se propone un estimador basado en la minimización de la entropía de una versión estimada de manera no paramétrica de la densidad de probabilidad de los errores ( $\hat{p}_{n,h}(\mathbf{e})$ ). Dicha versión está basada en la

distribución empírica de los errores. En efecto, se considera la utilización de estimadores por núcleos (kernel estimators).

Este trabajo propone también en las secciones 4 y 5, la comparación de estimadores a Mínimo de Entropía (EME), los cuales muestran una mejora en los resultados cuando se intenta conocer cual es la influencia del parámetro denominado ancho de banda ( $h$ ) perteneciente a los estimadores por núcleos clásicos. Así pues, se propone una función más general del estimador no paramétrico de los errores, que nos conducirá a estimadores más eficientes (hablando en términos de tiempo de cálculo) (Eğecioglu *et al*, 2000), (Terrell, 1990) y (Terrell *et al*, 1985). Finalmente en la sección 6 brindamos algunas conclusiones sobre los estimadores EME propuestos.

## 2 El problema general de regresión

### 2.1 Regresión lineal

Muchas aplicaciones del procesamiento digital de señales y de instrumentación se basan en un análisis de modelos estadísticos. Uno de los modelos más utilizados es el modelo regresivo lineal (simple o multi-variable):

$$y_i = x_i^T \boldsymbol{\theta} + e_i, \quad i = 1, 2, \dots, K, n, \quad (1)$$

donde  $y_i$  representa la  $i$ -ésima respuesta (datos observados o censados, o bien, variables a explicar),  $\{x_i\}_{i=1}^n$  es el vector de variables explicativas o protocolo experimental (por ejemplo, los instantes de observación de un instrumento, la posición de sensores (Fleury, 1999), frecuencias utilizadas en corrientes de Foucault (De la Rosa *et al*, 2003), etc.) de dimensión  $p$  (o matriz  $n \times p$ ) que está asociado a  $y_i$  ( $y_i, x_i^T$ ),  $\boldsymbol{\theta} = \{\theta_j\}_{j=1}^p$  es el vector de parámetros funcionales de dimensión  $p$  que tendrán que ser estimados por algún proceso de identificación. Los errores que modelan el sistema como un proceso aleatorio están dados por  $e_i$ . El conjunto  $\Theta_L \subset \mathbb{R}^p$  representa el subconjunto admisible de parámetros  $\boldsymbol{\theta}$  y  $X_L \subset \mathbb{R}^n$  representa el dominio admisible para las variables explicativas.

### 2.2 Regresión no lineal

Una extensión del modelo regresivo lineal, es el modelo regresivo no lineal. Al igual que el modelo anterior, éste caracteriza el modelado de los datos observados  $\mathbf{y} = \{y_i\}_{i=1}^n$ , pero ahora se lleva a cabo por medio de una función  $f(\cdot)$  parametrada (ver ecuación (2)). Dicha función es no lineal con respecto a los parámetros, y también se considera su uso porque en muchas de las aplicaciones del modelado en instrumentación y control se ha demostrado que los sistemas pueden ser modelados de manera más óptima haciendo uso de este tipo de funciones, además se reduce el número de grados de libertad (de parámetros a estimar) obteniendo en ocasiones tiempos de cálculo reducidos. Las perturbaciones que interactúan en el sistema analizado son también modeladas como procesos estocásticos. Entonces, las componentes determinista  $f(\cdot)$  y aleatoria  $e_i$  son modeladas por:

$$y_i = f(x_i, \boldsymbol{\theta}) + e_i, \quad i = 1, 2, \dots, K, n, \quad (2)$$

Algunas técnicas para la estimación de  $\boldsymbol{\theta}$  son por ejemplo, mínimos cuadrados (MC), máximo de verosimilitud (MV), M-estimadores, etc. En este caso, proponemos un estimador robusto basado en la minimización de la entropía de una versión estimada de manera no paramétrica de la densidad de probabilidad de los errores ( $\hat{p}_{n,h}(\mathbf{e})$ ). Dicha versión está basada en la distribución empírica de los errores.

### 3 Estimador a mínimo de entropía (EME)

Los estimadores  $\hat{\boldsymbol{\theta}}$  se obtienen de manera clásica minimizando una función de costo  $J(\boldsymbol{\theta})$  que varía en función  $\psi(\cdot)$  de los residuos o errores  $\mathbf{e}(\boldsymbol{\theta})$ :

$$e_i(\boldsymbol{\theta}) = y_i - x_i^T \boldsymbol{\theta}, \text{ o bien } e_i = y_i - f(x_i, \boldsymbol{\theta}) \quad (3)$$

$$J(\boldsymbol{\theta}) = \sum_{i=1}^n \psi(e_i(\boldsymbol{\theta})) \quad (4)$$

Tal es el caso del estimador de máximo de verosimilitud:

$$\hat{\boldsymbol{\theta}}_{MV} = \arg \min_{\boldsymbol{\theta} \in \Theta} \left[ - \sum_{i=1}^n \log(p(e_i(\boldsymbol{\theta}))) \right] \quad (5)$$

Para su funcionamiento óptimo este estimador necesita del conocimiento previo de  $p(\mathbf{e})$ . Cuando la distribución  $p(\mathbf{e})$  no se conoce de manera precisa, el estimador  $\hat{\boldsymbol{\theta}}_{MV}$  es sub-óptimo. El estimador a mínimo de entropía evita el problema de sub-optimalidad y contiene una robustez intrínseca. Este estimador está basado en la minimización de la entropía de una versión estimada  $\hat{p}_{n,h}(\mathbf{e})$  de la distribución  $p(\mathbf{e})$ . Una primera proposición de Pronzato y Thierry es obtener una aproximación construida a partir de núcleos  $\hat{p}_{n,h}(\mathbf{e})$  los cuales utilizan la distribución empírica del vector aleatorio  $e_1(\boldsymbol{\theta}), e_2(\boldsymbol{\theta}), \dots, e_n(\boldsymbol{\theta})$ , como se expone en la siguiente expresión:

$$\hat{p}_{n,h}(\mathbf{e}) = \hat{p}_{n,h}(\mathbf{e} | e_1(\boldsymbol{\theta}), \dots, e_n(\boldsymbol{\theta})) = \frac{1}{n} \sum_{i=1}^n K_h(\mathbf{e} - e_i) \quad (6)$$

Se formula la hipótesis de que  $p(\mathbf{e})$  es simétrica y que es doblemente diferenciable, por lo tanto se asume que  $K(\cdot)$  es una función núcleo ponderada tal que ésta satisface las condiciones impuestas según los trabajos de Masry (Masry, 1983) y retomadas en los trabajos de Devroye (Devroye, 1992) – (Devroye, 1997), Berlinet (Berlinet *et al*, 1994), y Loader (Loader, 1999) (ver también los trabajos de Parzen – Rosenblatt que datan de 1968). El ancho de banda  $h = h_n$  está dado en función del tamaño de la muestra  $n$ , por lo que  $h_n$  es una secuencia de números positivos que debe satisfacer:  $h_n \rightarrow 0$  y  $nh_n \rightarrow \infty$  cuando  $n \rightarrow \infty$ . La consistencia fuerte de  $\hat{p}_{n,h}(\mathbf{e})$  y su convergencia hacia  $p(\mathbf{e})$ , dependen de un procedimiento adecuado de selección del ancho de banda. Berlinet y Devroye llevan a cabo un estudio muy completo, en donde se comparan diferentes procedimientos (Berlinet *et al*, 1994), (Devroye, 1989). Un procedimiento sencillo y rápido que nosotros decidimos utilizar es el propuesto por Terrell. Se ha demostrado que bajo condiciones de simetría y para distribuciones unimodales el procedimiento de Terrell es consistente, las condiciones necesarias para asegurar la consistencia, eficiencia y convergencia están dadas en (Terrell, 1990), (Terrell *et al*, 1985).

Si se asume que existe la convergencia de  $\hat{p}_{n,h}(\mathbf{e})$  hacia  $p(\mathbf{e})$ , entonces se puede construir el criterio de entropía sobre  $\hat{p}_{n,h}(\mathbf{e})$ . Para llevar a cabo de manera práctica la minimización de la entropía, se debe considerar un vector de

residuos extendido  $\hat{p}_{n,h}(\mathbf{e}_E)$  el cual está dado por:  $\mathbf{e}_E = \{\mathbf{e}, -\mathbf{e}\}$ . El criterio se construye entonces tomando en cuenta  $h$ :

$$J_e(\boldsymbol{\theta}) = H_A(\hat{p}_{n,h}(\mathbf{e}_E)) \quad (7)$$

en donde  $H_A(f) = - \int_{-A_n}^{A_n} f(x) \log(f(x)) dx$ . El estimador a mínimo de entropía (EME) se denota entonces por  $\hat{\boldsymbol{\theta}}_{EME}$

$$\hat{\boldsymbol{\theta}}_{EME} = \arg \min_{\boldsymbol{\theta} \in \Theta} J_e(\boldsymbol{\theta}) \quad (8)$$

Pronzato y Thierry (Pronzato *et al*, 2000a), (Pronzato *et al*, 2000b) analizan algunas propiedades de asintoticidad y de consistencia del estimador a mínimo de entropía, los cuales hemos retomado y estamos generalizando para el caso de núcleos distintos a aquellos que dependen del ancho de banda  $h$ .

Por otro lado, para obtener una caracterización de la distribución de los parámetros, utilizamos un método de simulación sucesiva ( $N$  veces) del proceso de estimación paramétrica (Integración vía Monte-Carlo), a partir de dicha simulación estocástica obtendremos las poblaciones conjuntas y marginales (si se considera que a cada etapa de simulación se tiene un vector diferente de realizaciones de  $\mathbf{e} \sim p(\mathbf{e})$ ). La simulación estocástica depende del conocimiento de  $p(\mathbf{e})$ , en el uso de ejemplos académicos podemos utilizar el método de Monte-Carlo Primitivo (MCP) o bien en aplicaciones prácticas podemos recurrir a la utilización de los métodos bootstrap (De la Rosa *et al*, 2006) en donde no es necesario el conocimiento de la distribución exacta de  $\mathbf{e}$ . Estas técnicas de simulación utilizan el concepto básico de Monte-Carlo, y brindan entonces una aproximación empírica  $\hat{p}(\boldsymbol{\theta})$  de la distribución  $p(\boldsymbol{\theta})$ , siempre y cuando el número de simulaciones sea significativamente grande  $\lim_{N \rightarrow \infty} |\hat{p}(\boldsymbol{\theta}) - p(\boldsymbol{\theta})| = 0$  (respetando la ley de los grandes números y aplicando algunas técnicas de reducción de variancia).

#### 4 Estructuras de los núcleos

Un núcleo de la forma  $K(z)$  es considerado como un núcleo fijo  $K_h(z) = 1/(h^d)K(z/h)$ , en donde  $h > 0$  y es conocido como parámetro de ancho de banda (e.g. factor de suavizado). El problema fundamental en la estimación de densidades de probabilidad haciendo uso de núcleos radica en que se tiene que llevar a cabo una selección adecuada de  $h$  y una selección apropiada del núcleo al mismo tiempo. La selección de un núcleo  $K(z)$  podría depender de qué tan suave (o regular) es la distribución objetivo  $p(\cdot)$ . Adicionalmente al núcleo propuesto por Pronzato y Thierry, nosotros proponemos el uso de otros dos núcleos que no dependen de  $h$  y los cuales introducimos de manera breve en esta sección. El núcleo Gaussiano por su parte, brinda un buen desempeño cuando  $h$  es seleccionado utilizando el principio de “sobre - suavizado” (e.g. oversmoothed principle) introducido por Terrell (Terrell, 1990), (Terrell *et al*, 1985). Sin embargo, no debemos olvidar que dicha selección de  $h$  es asintóticamente óptima, por lo que un trabajo que nos ocupa actualmente es indagar como se comporta el estimador por núcleos cuando se usa un método no asintótico como el de doble núcleo (e.g. double kernel) que es más indicado cuando  $n$  es de tamaño limitado (Berlinet *et al*, 1994), (Devroye, 1989), (Devroye, 1997), y puesto que el vector de errores es generalmente de tamaño limitado. El segundo núcleo que proponemos utilizar está representado por una serie de funciones ponderadas del coseno, en donde  $h$  es visto de manera diferente. Y finalmente, el tercer núcleo propuesto se obtiene a partir de una clase recientemente propuesta denominada núcleos de Hilbert (Devroye *et al*, 1999). Esta nueva clase de núcleos evita la dependencia de la selección adecuada de

$h$  y a cambio hay que seleccionar dos parámetros cuya proposición es mucho más sencilla (los parámetros  $d$  y  $k$  serán definidos en la sección 4.3).

#### 4.1 Núcleo normal o Gaussiano

Entre la clase de núcleos estándar (Berliner *et al*, 1994), elegimos utilizar el núcleo Gaussiano que es uno de los más utilizados y entonces tenemos que el estimador se expresa mediante la suma de exponenciales:

$$\hat{p}_{n,h}(z) = \frac{1}{nh\sqrt{2\pi}} \sum_{i=1}^n \exp\left(-\frac{(z-z_i)^2}{2h^2}\right) \tag{9}$$

Para este caso, y sabiendo que se cuenta con una estructura de núcleo fija Terrell (Terrell, 1990) sugiere un ancho de banda sobre – suavizado  $h_n$  tal que:  $h_n = 3\left(\frac{1}{2\sqrt{\pi}(35)}\right)^{\frac{1}{5}} \sigma^{\frac{1}{5}}$ , este valor para el ancho de banda garantiza que la minimización del MISE (Mean Integrated Squared Error) será obtenida,  $\sigma$  es la desviación estándar de los datos  $z$ , y  $\int K(z)^2 dz = \frac{1}{2\sqrt{\pi}}$ . Suponiendo que existen las condiciones de suficiencia y que cuando se utiliza el principio de sobre suavizado existe consistencia, entonces la DDP de los errores estimada utilizando el método por núcleo Gaussiano es consistente.

#### 4.2 Funciones coseno ponderadas

Una secuencia especial de funciones coseno ponderadas  $c_q(z) = 1/(A_q)\cos^{2q}(z/2)$  requiere tan sólo  $O(q^2n)$  operaciones aritméticas,  $q$  es una cantidad que crece en función de  $n$  sin tener una expresión formal que exprese un límite sobre ésta (i.e.  $q(n) < n$  ó  $q(n) > n$ ). Este estimador es similar al estimador general de núcleos y evalúa una expansión en series de una manera más eficiente. En este caso, el rol que tiene  $h$  se sustituye por  $q$  ( $c_q(z) \equiv K_h(z)$ ), lo cual es ventajoso puesto que la elección de  $q$  es muchísimo más sencilla que la elección de  $h$ . Y entonces la ecuación (6) será equivalente a

$$\hat{p}_{n,q}(z) = \frac{1}{nA_q} \sum_{i=1}^n \left(\frac{1 + \cos(z-z_i)}{2}\right)^q \tag{10}$$

en donde el valor de  $A_q$  puede ser aproximado por:  $A_q \approx \frac{2\sqrt{\pi}}{\sqrt{q}}$ . Las condiciones suficientes para que exista la consistencia y la convergencia se abordan en (Eğecioğlu *et al*, 2000) (revisar los teoremas 3 y 4), en donde se demuestra que el MISE se minimiza de forma asintótica, ofreciendo una estimación eficaz de la DDP.

#### 4.3 Núcleos de Hilbert

Una clase reciente de estimadores por núcleos se basa en los llamados núcleos de Hilbert. El núcleo  $K_h(z) = 1/(h^d)K(z/h)$  se considera equivalente a  $K(u) = 1/\|u\|^d$ , obteniendo así una cancelación del ancho de banda  $h$ :

$$\hat{p}_n(z) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\|z - z_i\|^d} \quad (11)$$

La consistencia de esta clase de estimadores ha sido comprobada en (Devroye *et al.*, 1999) (ver el teorema 2). Basándonos en una sub-clase de estos estimadores, la cual evita la producción de picos infinitos durante la estimación, proponemos utilizar el estimador de Hilbert de orden  $k$ , en donde  $k > 0$  y en este caso  $k = 2$ :

$$\hat{p}_n(z) = \sqrt{\frac{4}{V_d^2 \pi(n-1) \log(n)} \sum_{1 \leq i < j \leq n} \frac{1}{Den_{i,j}}} \quad (12)$$

en donde  $Den_{i,j} = \|z - z_i\|^{2d} + \|z - z_j\|^{2d}$  y  $V_d$  es el volumen de una esfera que se encuentra en  $R^d$  ( $V_d = 0.7071$ ). Esta última expresión también es conocida como el estimador de la distribución de Cauchy. Por otro lado,  $\|\cdot\|$  denota la distancia métrica  $L_2$  sobre  $R^d$  y se considera que  $\hat{p}_n(z) \rightarrow p(z)$  al menos converge en probabilidad para todo valor de  $\mathbf{z}$ .

En la primera introducción del estimador EME, éste demuestra su resistencia a datos aberrantes y de igual manera la generalización que proponemos en esta sección conserva dicha robustez:  $J_{eg}(\boldsymbol{\theta}) = H_A(\hat{p}_n(\mathbf{e}_E))$  y entonces,

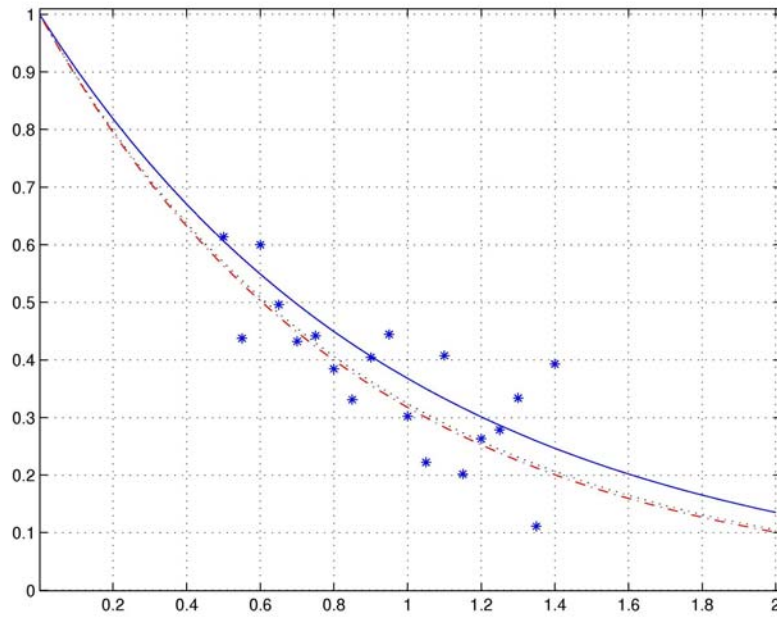
$$\hat{\boldsymbol{\theta}}_{EME} = \arg \min_{\boldsymbol{\theta} \in \Theta} J_{eg}(\boldsymbol{\theta}) \quad (13)$$

Puesto que el estimador  $\hat{\boldsymbol{\theta}}_{EME}$  no es asintóticamente eficaz, podemos considerar que para una selección adecuada de  $A_n$  y en su caso de  $h_n$ , de  $q$ , de  $d$  y de  $k$ , éste puede ser considerado como “asintóticamente eficaz en un sentido limitado” (actualmente estamos trabajando en la parte matemática que justifique dicha suposición).

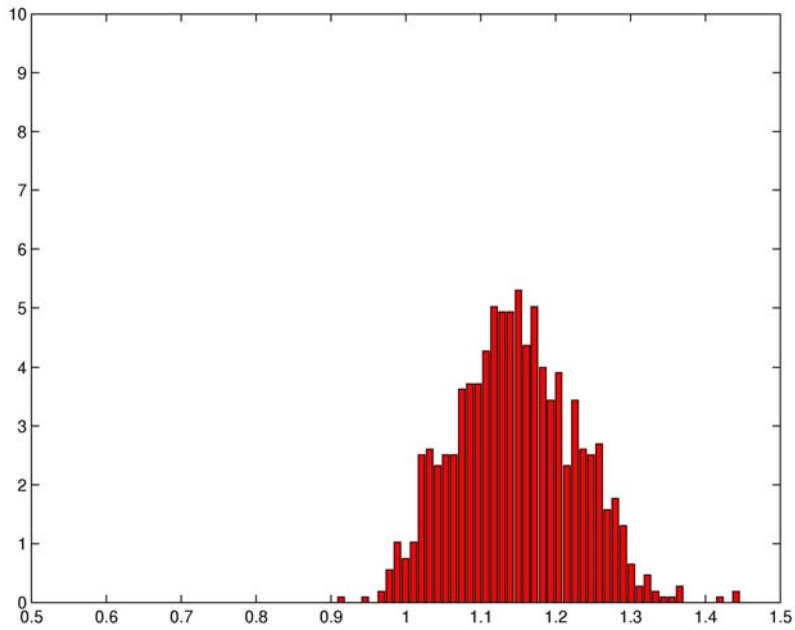
## 5 Resultados con algunos ejemplos

### 5.1 Ejemplo 1

El primer ejemplo que abordamos, es el mismo ejemplo inicial tratado por Pronzato y Thierry (Pronzato *et al.*, 2000a), (Pronzato *et al.*, 2000b), en donde  $f(x_i, \boldsymbol{\theta}) = \exp(-\theta x_i)$  y  $\bar{\theta} = 1$  (parámetro “real”) en la ecuación (2), se toman 20 observaciones  $\{x_i\}_{i=1}^n = [0.5, 0.55, K, 1.45]^T$  y se consideran los tres núcleos propuestos. Se estudiaron dos casos, en el primero se suponen errores con distribución Gaussiana y además un dato aberrante  $y(20) = 5$  que contamina de alguna manera los datos observados,  $e_i \sim N(0, \sigma^2)$  con  $\sigma = 0.1$ . En el segundo caso se considera que los errores siguen una distribución uniforme,  $e_i \sim U(-\alpha/2, \alpha/2)$  con  $\alpha = 0.8$ . En las Figuras 1 y 3 se muestran algunos resultados cuando se obtiene la reproducción de los datos utilizando la primera estimación EME<sup>3</sup>, que representa el estimador que utiliza los núcleos de Hilbert (se pueden comparar con la Figura 5 de (Pronzato *et al.*, 2000b)). Por otro lado, en las Figuras 2 y 4 se muestran las distribuciones empíricas del parámetro  $\theta$  de la función no lineal, obtenidas utilizando Monte-Carlo primitivo (MCP) para 10,000 simulaciones del proceso de estimación cuando se toman en cuenta los dos casos diferentes de perturbaciones (ruido).



**Fig. 1.** Representación de datos según el modelo  $f(x_i, \theta)$  (---), se considera ruido Gaussiano con un dato aberrante en  $y(20) = 5$  (\*) y el estimador  $EME^3$  (- - -) para primera estimación, y (- · - · -) para valor medio.



**Fig. 2.** Distribución empírica del parámetro de  $f(x_i, \theta)$  cuando se usa el estimador  $EME^3$ , hay ruido Gaussiano con un dato aberrante en  $y(20) = 5$ .

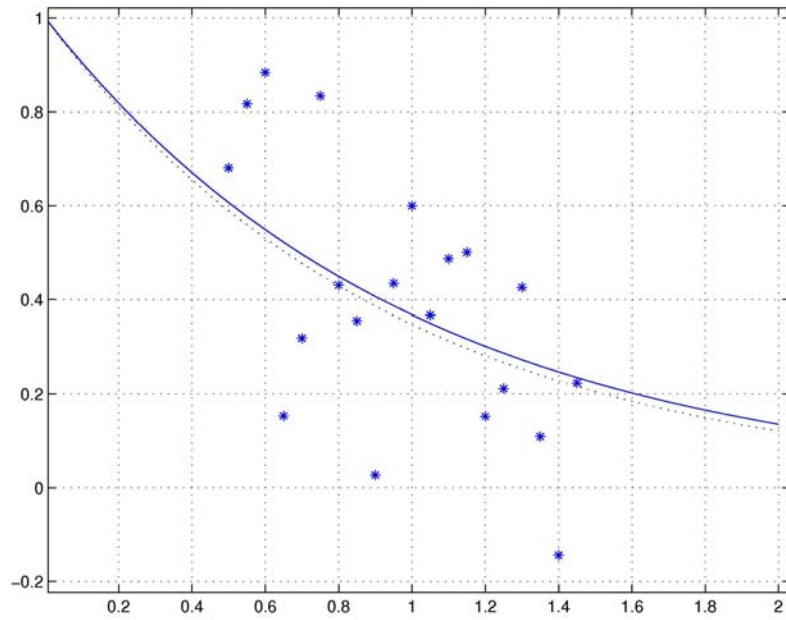


Fig. 3. Representación de datos según el modelo  $f(x_i, \theta)$  (---), se considera ruido uniforme (\*) y el estimador  $EME^3$  (- - -).

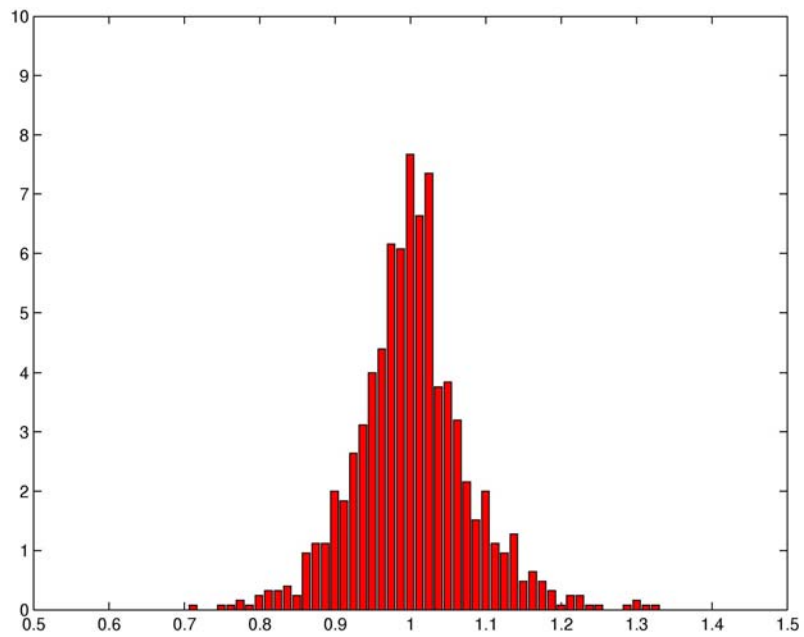


Fig. 4. Distribución empírica del parámetro de  $f(x_i, \theta)$  cuando se usa el estimador  $EME^3$ , hay ruido uniforme.



**5.2 Ejemplo 2**

Otro problema en el cual se considera la estimación paramétrica, es la estimación de los coeficientes de un filtro de respuesta al impulso finita (Finite Impulse Response - FIR), en un marco donde existen interferencias desconocidas. El filtro es  $A(z) = \sum_{j=0}^4 a_j z^{-j}$ , en donde  $z^{-1}$  es el operador de retardo. Los datos  $y_i$  están contaminados por una secuencia de  $e_i \sim N(0, \sigma^2)$  con  $\sigma = 0.1$  y otra secuencia  $\mu_i$  dada por otro filtro FIR,  $B(z) = \sum_{j=0}^2 b_j z^{-j}$  y una señal binaria  $v_i \in \{-1, 1\}$  (en donde  $\mu_i = B(z)v_i$ ). El vector de parámetros a estimar es:  $\theta = [a_0, a_1, \dots, a_4]^T$ , y  $y_i = A(z)u_i + \varepsilon_i$ ,  $i = 1, \dots, 100$  (arreglo como un modelo regresivo lineal), en donde  $\varepsilon_i = e_i + \mu_i$ , las entradas  $u_i \in \{-1, 1\}$ . Los valores reales utilizados para  $\bar{\theta} = [1, -0.5, 0.2, -0.3, 0.1]^T$  y  $\bar{b} = [1, -0.5, 0.2]^T$ . Se obtuvieron 500 simulaciones según el método MCP, en la Tabla 1 se presenta una comparación de las primeras estimaciones obtenidas con el estimador de mínimos cuadrados (MC) y los tres estimadores a mínimo de entropía (EME) propuestos (EME<sup>1</sup> para núcleo normal, EME<sup>2</sup> para núcleo cosenoidal y EME<sup>3</sup> para núcleos de Hilbert). Las distribuciones empíricas obtenidas para los parámetros del primer filtro FIR ( $A(z)$ ) se presentan en las Figuras 6, 7 y 8 las cuales se pueden contrastar con la Figura 5 que muestra las distribuciones empíricas obtenidas con el estimador MC. Se verifica que el valor medio converge asintóticamente a el valor real  $\bar{\theta}$ , y en el caso de los tres estimadores EME se nota que la dispersión (variancia) es mucho menor (aquí hay que cuidar que el límite inferior de Cramer-Rao no tienda rápidamente a cero - asintoticidad eficaz en un sentido limitado).

**Tabla 1** Comparación de las primeras estimaciones de los parámetros utilizando MC, EME y modelo real.

	MC	EME <sup>1</sup>	EME <sup>2</sup>	EME <sup>3</sup>	Ref.
$\hat{a}_0$	0.92111	1.01000	1.00871	1.01072	1.0
$\hat{a}_1$	-0.39534	-0.50103	-0.50312	-0.50150	-0.5
$\hat{a}_2$	0.06132	0.22052	0.22071	0.21911	0.2
$\hat{a}_3$	-0.16180	-0.29027	-0.29293	-0.28988	-0.3
$\hat{a}_4$	0.14790	0.10996	0.10995	0.10994	0.1

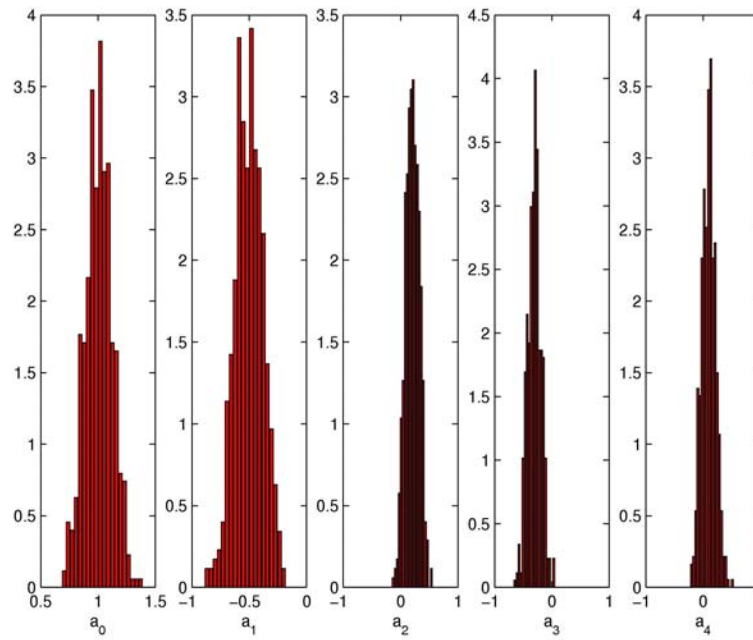


Fig. 5. Distribución empírica de los parámetros del filtro FIR  $A(z)$  cuando se usa el estimador MC.

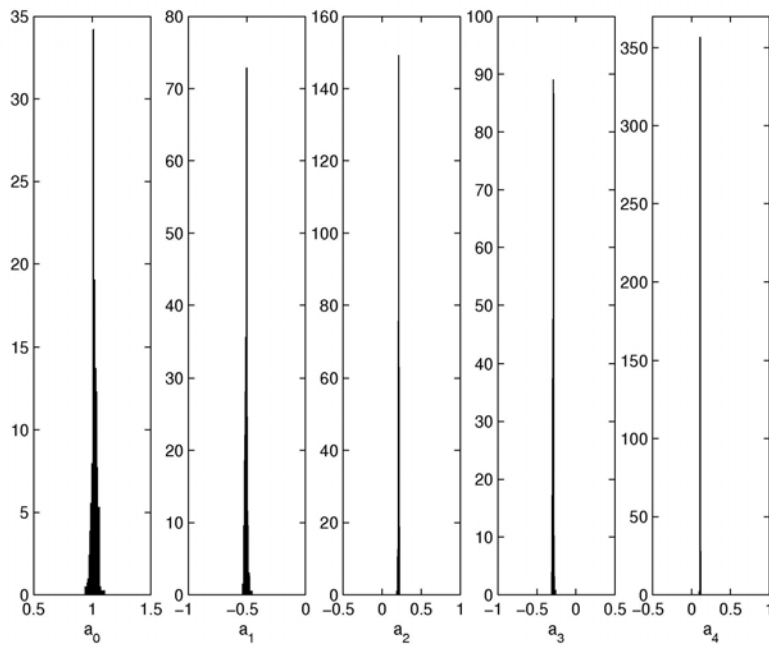


Fig. 6. Distribución empírica de los parámetros del filtro FIR  $A(z)$  cuando se usa el estimador EME<sup>1</sup>.

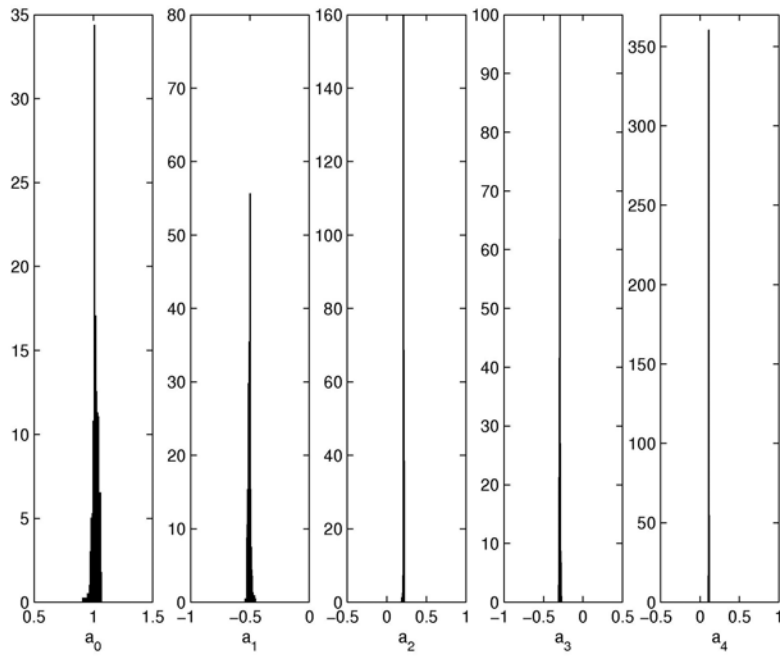


Fig. 7. Distribución empírica de los parámetros del filtro FIR  $A(z)$  cuando se usa el estimador  $EME^2$ .

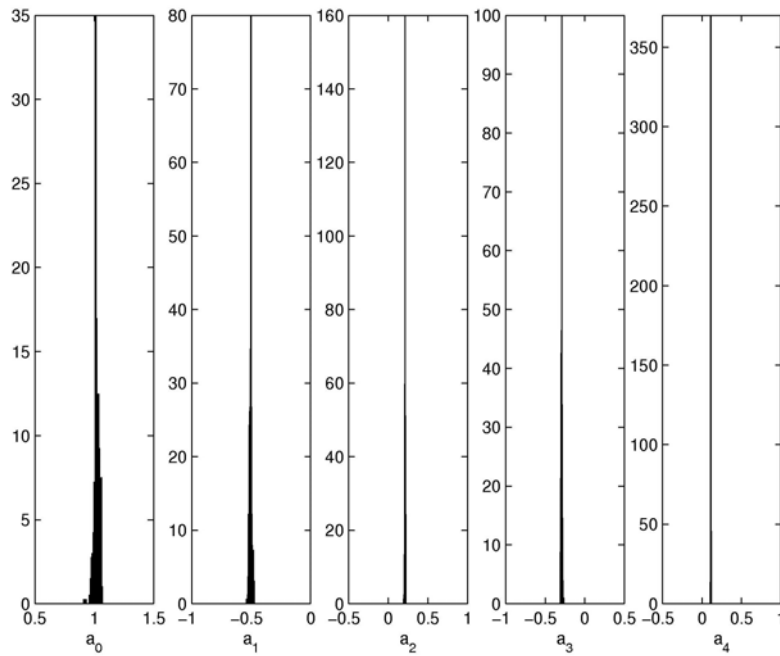


Fig. 8. Distribución empírica de los parámetros del filtro FIR  $A(z)$  cuando se usa el estimador  $EME^3$ .

## 6 Conclusiones

Una selección entre los diferentes núcleos permite mejorar la precisión y el tiempo de cálculo de los estimadores EME. Los resultados obtenidos con los tres núcleos son en general favorables en el sentido de la robustez, sin embargo el estimador EME<sup>2</sup> y EME<sup>3</sup> son los que convergen más rápido (para el estimador EME<sup>2</sup> se utiliza  $q = 40$ , y es interesante observar cómo mejora la precisión cuando variamos  $q$  en el rango  $30 \leq q \leq 90$ ). Más aún, los resultados de simulación en aplicaciones de instrumentación (De la Rosa *et al*, 2002), (De la Rosa *et al*, 2003), otros recientes obtenidos en aplicaciones de telecomunicaciones (Pronzato *et al*, 2001), y los que hemos obtenido en el presente trabajo, nos permiten proponer el uso de los estimadores EME en un marco más general, aún en condiciones donde la no linealidad de los sistemas está presente (Fleury, 1999). Con lo anterior, y como trabajo futuro, pretendemos concretizar de manera más real la programación de algoritmos de estimación en tarjetas DSP, para su utilización en problemas de instrumentación o de tratamiento de señales reales.

## Agradecimientos

Quiero agradecer de la manera más atenta, la disponibilidad, amabilidad y ayuda del Prof. Luc Pronzato del Laboratorio I3S situado en la Universidad de Nice-Shopia Antipolis en Francia. Por otro lado, también agradezco el apoyo brindado por PROMEP para la realización de este trabajo que esta auspiciado por el convenio 103.5/03/1127.

## Referencias

1. (Berlinet *et al*, 1994) **Berlinet A.**, and **Devroye L.**, “A comparison of kernel density estimates,” *Publications de l'Institut de Statistique de l'Université de Paris*, Vol. 38 (3), 1994, pp. 3 – 59.
2. (De la Rosa *et al*, 2002) **De la Rosa J. I.**, and **Fleury G.**, “On the Kernel selection for Minimum-Entropy estimation,” *Proc. of the IEEE Instrumentation and Measurement Technology Conference*, Anchorage, AK (USA), 21 – 23 May 2002, Vol. 2, pp. 1205 – 1210.
3. (De la Rosa *et al*, 2003) **De la Rosa J. I.**, **Fleury G.**, and **Davoust M.-E.**, “Minimum-Entropy, pdf approximation and Kernel selection for measurement estimation,” *IEEE Transactions on Instrumentation and Measurement*, Vol. 52, No. 4, August 2003, pp. 1009 – 1020.
4. (De la Rosa *et al*, 2006) **De la Rosa J. I.**, and **Fleury G.**, “Bootstrap methods for a measurement estimation problem,” *IEEE Transactions on Instrumentation and Measurement*, Vol. 55, No. 3, June 2006, pp. 820 – 827.
5. (Devroye, 1992) **Devroye L.**, “A note on the usefulness of superkernels in density estimation,” *The Annals of Statistics*, Vol. 20, 1992, pp. 2037 – 2056.
6. URL: //jeff.cs.mcgill.ca/~luc/
7. (Devroye, 1989) **Devroye L.**, “The double kernel method in density estimation,” *Annales de l'Institut Henri Poincaré*, Vol. 25, 1989, pp. 533 – 580.
8. (Devroye *et al*, 1999) **Devroye L.**, and **Krzyzak A.**, “On the Hilbert kernel density estimate,” *Statistics and Probability Letters*, Vol. 44, 1999, pp. 299 – 308.
9. (Devroye, 1997) **Devroye L.**, “Universal smoothing factor selection in density estimation: theory and practice,” *Test*, Vol. 6, 1997, pp. 223 – 320.
10. (Ö. Egecioglu *et al*, 2000) **Egecioglu Ö.**, and **Srinivasan A.**, “Efficient non-parametric density estimation on the sphere with applications in fluid mechanics,” *SIAM Journal on Scientific Computing*, Vol. 22, 2000, No. 1, pp. 152 – 176.
11. (Fleury, 1999) **Fleury G.**, “Optimal nonlinear modeling and reparametrization,” *IEEE International workshop on Intelligent Signal Proc.*, sept. 1999, pp. 72 – 76.
12. (Loader, 1999) **Loader C. M.**, “Bandwidth selection: classical or plug-in?,” *The Annals of Statistics*, Vol. 27, No. 3, 1999, pp. 415 – 438.
13. (Masry, 1983) **Masry E.**, “Probability density estimation from sampled data,” *IEEE Trans. on Information Theory*, Vol. IT-29, No.5, Sept. 1983, pp. 697 – 709.

14. (Pronzato *et al.*, 2000a) **Pronzato L.**, and **Thierry E.**, “A minimum-entropy estimator for regression problems with unknown distribution of observation errors,” *MaxEnt 2000*, Edited by A. Mohammad-Djafari, *American Institute of Physics*, July 8 - 13, 2000, pp. 169 – 180.
15. (Pronzato *et al.*, 2000b) **Pronzato L.**, and **Thierry E.**, “A minimum-entropy estimator for regression problems with unknown distribution of observation errors,” Tech. Rep. 00-08, Laboratoire I3S, CNRS-Université de Nice-Shophia Antipolis, France, 2000.
16. (Pronzato *et al.*, 2001) **Pronzato L.**, and **Thierry E.**, “Entropy minimization of parameter estimator with unknown distribution of observation errors,” *Proc. of the IEEE International Conference in Acoustics, Speech and Signal Processing*, Vol. 6, March 2001, pp. 3993 – 3996.
17. (Terrell, 1990) **Terrell G. P.**, “The maximal smoothing principle in density estimation,” *Journal of the American Statistical Association*, Vol. 85, 1990, pp. 470 – 477.
18. (Terrell *et al.*, 1985) **Terrell G. P.**, and **Scott D. W.**, “Oversmoothed nonparametric density estimation,” *Journal of the American Statistical Association*, Vol. 80, 1985, pp. 209 – 214.



**José Ismael de la Rosa Vargas** Ingeniero en Electrónica y Comunicaciones egresado de la Universidad Autónoma de Zacatecas (UAZ) en 1995. Obtuvo el grado de Maestría en Sistemas Digitales en 1998 en CITEDI-IPN y Doctorado con especialidad en Procesamiento Digital de Señales (PDS) y Control en 2002 en la Universidad de Paris XI y SUPELEC, Paris, Francia. Se incorporó a la UAZ en 2003, como profesor-investigador en el laboratorio de PDS de la Unidad Académica de Ingeniería Eléctrica. Actualmente pertenece al SNI (nivel 1) y es profesor con perfil PROMEP.