

Estimadores robustos de estadísticos de posición

Mariano Carbonero

Facultad de Ciencias
Económicas y
Empresariales. ETEA.
Univ. de Córdoba
14004 Córdoba
mcarbonero@etea.com

Jesús Ramírez

Facultad de Ciencias
Económicas y
Empresariales. ETEA.
Univ. de Córdoba
14004 Córdoba
jramirez@etea.com

César Hervás

Departamento de
Informática y A. N.
Universidad de Córdoba
14071 Córdoba
chervas@uco.es

Domingo Ortiz

Departamento de
Informática y A. N.
Universidad de Córdoba
14071 Córdoba
ma1orbod@uco.es

Este trabajo ha sido financiado por el MCYT, proyecto TIC2002-04036-C05-02 y con fondos FEDER

Resumen

El trabajo aborda la utilización de estimadores robustos en el caso de que las distribuciones asociadas a las variables implicadas en el problema no tengan distribuciones normales. De entre estos estimadores destacamos la importancia de la mediana como estimador de parámetros de localización y en general de los cuantiles como estimadores de parámetros de dispersión. Proponemos para la mediana una distribución alternativa a la distribución binomial que habitualmente se utiliza y que tiene la dificultad de ser discreta, por lo que es difícil obtener intervalos de confianza con extremos reales. Como aplicación consideramos la construcción de intervalos de confianza bilaterales que sirvan de base a operadores de cruce multipadres para algoritmos genéticos con codificación real

1. Introducción

Existen muchos estimadores sencillos para el centro de una distribución simétrica perteneciente a una clase llamada L-estimadores, esto es estimadores combinación lineal de estadísticos de orden. Hasta ahora, dada la ausencia de un conocimiento preciso acerca de la distribución verdadera subyacente a un conjunto de datos, es habitual considerar la conducta de estos estimadores para distribuciones alternativas que van desde una Gaussiana a una distribución de Cauchy [1]. Este tipo de estimadores están teniendo una gran difusión dentro de la comunidad científica tanto en estadística como en aprendizaje, dado que en minería de datos el efecto de la dimensionalidad del espacio de variables independiente inmersas en la resolución de un problema, así como la suposición de

normalidad de dichas variables, o de los errores cometidos en problemas de clasificación o regresión, hace que las técnicas de comparación de algoritmos o de análisis de sus características sean abordadas mediante métodos robustos de estadística no paramétrica [2]. Es por ello, que con este trabajo intentamos aproximarnos al desarrollo y aplicación de una teoría de los contrastes de hipótesis en el contexto de modelos de regresión de cuantiles lineales (L-estimadores)[3] que se utilizan como una alternativa a los desarrollados en la literatura sobre grandes muestras en múltiples aspectos de la inferencia estadística asociada a la regresión con cuantiles. Ejemplos de ello incluyen, la inferencia basada en rangos [4], así como la inferencia basada en varios procedimientos de bootstrap [5-7]. De esta forma, en una primera aproximación aplicaremos los resultados de distribuciones alternativas asociadas a estimadores robustos, la mediana en este trabajo, que explícitamente incorporen detalles de la forma en la que se implementa la varianza asintótica estimada en los test estadísticos de los parámetros de posición de una distribución.

Proponemos como aplicación de estas distribuciones la definición y propuesta de un método de cruce para algoritmos genéticos con codificación real, AGCR, donde se trata de extraer las características estadísticas más relevantes de localización y dispersión de los individuos más aptos obtenidos en cada generación. Es por ello, un algoritmo donde introducimos un aprendizaje estadístico al considerar, en primer lugar, que para estimar la localización y dispersión de los genes de los mejores individuos sus distribuciones van a ser continuas a lo largo de la evolución [8-9].

Bajo esta consideración, al no conocer la distribución de los genes de los mejores individuos, utilizamos un estimador de localización basado en la norma L_1 , esto es, la

mediana muestral de los genes de los n mejores individuos de cada generación. Como la distribución de la mediana no depende de la distribución de esos genes como proponemos en este trabajo, entonces tendremos intervalos que no dependerán de la distribución de los genes de los mejores individuos lo que en computación evolutiva es muy importante dados los posibles cambios que pueden tener dichas distribuciones a lo largo de la evolución

A partir de esos estimadores de los parámetros de localización y de sus distribuciones asociadas construimos intervalos de confianza bilaterales para producir, en cada generación, tres padres, uno formado por los extremos inferiores de los intervalos de confianza, otro formado por los extremos superiores y otro formado por el estadístico de localización utilizado la mediana dado que utilizamos la norma $L1$.

2. Definición del problema

Aunque presentado bajo distintos aspectos y con diferentes denominaciones, el problema que presentamos es, probablemente, uno de los más importantes y discutidos en la actualidad:

Consideremos una variable aleatoria, Y , y una variable k -dimensional \mathbf{X} sobre las que disponemos de un conjunto finito de n observaciones $\{(\mathbf{x}_i, y_i), i = 1, \dots, n\}$.

Consideremos igualmente una familia de funciones

$$F = \{f_{\theta}(\mathbf{x}), \theta \in \Theta \subset \square^p\} \quad (1)$$

definidas de \square^k en \square , de forma funcional conocida salvo por el valor del parámetro p -dimensional θ perteneciente a un cierto espacio paramétrico Θ .

A partir de la anterior información, el objetivo consiste en determinar, a partir de la información incluida en el conjunto de observaciones, la mejor elección posible del parámetro vectorial θ de manera que resulte mínima la cantidad

$$\Delta(\theta) = \Delta(\mathbf{y}, \mathbf{f}_{\theta}(\mathbf{x})) \quad (2)$$

, siendo \mathbf{y} y $\mathbf{f}_{\theta}(\mathbf{x})$ los vectores n -dimensionales contruidos a partir del conjunto disponible de observaciones para cada valor paramétrico y Δ una distancia o, de forma más genérica, cualquier medida de disimilaridad.

Tan sólo por evidenciar la frecuencia de este tipo de problemas, basta considerar que, si la variable dependiente Y es de tipo continuo, este problema no es más que un modelo de regresión, mientras que si Y es una variable discreta y finita, se tratará de un problema de clasificación supervisada.

Este problema no admite una solución universal, ni en general, ni tan siquiera para alguna de sus variantes, y las diferencias entre las distintas soluciones, así como su nivel de éxito en términos del valor de la medida Δ vienen condicionados por la forma en que tanto F como Δ se elijan, elecciones éstas que a su vez, deberán apoyarse tanto en la información que pueda extraerse del conjunto de datos observados como en las hipótesis formuladas y contrastadas que sobre la población de la que provienen se efectúen, especialmente en lo que a la distribución de probabilidad de la variable aleatoria Y se refiere.

De nuevo, por aproximar estas consideraciones generales a problemas más concretos, baste constatar que el modelo clásico de regresión lineal, utilizando para Δ la norma L_2 , no es más que la siguiente elección

$$F = \left\{ f_{\theta}(\mathbf{x}) = \theta_0 + \sum_{j=1}^k \theta_j x_j, \theta \in \square^{k+1} \right\} \quad (3)$$

$$\Delta(\theta) = \sum_{i=1}^n (y_i - f_{\theta}(\mathbf{x}_i))^2 \quad (4)$$

cuya utilidad y conocidas propiedades estadísticas se derivan del doble supuesto de normalidad de la variable dependiente y linealidad en el conjunto de datos observados.

Enfocado el problema, en este trabajo trataremos de poner de relieve la importancia que tanto la elección de Δ como sus consecuencias tienen, así como los inconvenientes que pueden presentarse si esta elección no se hace cuidadosamente. Supondremos por tanto, en lo que sigue, que el conjunto de funciones F se encuentra perfectamente especificado salvo por el valor del parámetro θ , cuyo valor deberá ser, en consecuencia, estimado a partir de los datos disponibles.

Con este objetivo y bajo el citado supuesto, comenzaremos por describir un esquema de resolución y validación que, si bien no tiene carácter universal, sí que responde a la estrategia seguida por muchos de los métodos actualmente

propuestos y empleados para abordar el problema. En esta descripción se destacarán, sobre todo, las conexiones existentes entre los distintos elementos presentados así como las consecuencias que pueden llegar a tener determinadas decisiones si las hipótesis sobre los que se asientan no están bien contrastadas.

En la segunda parte, basada en los elementos críticos señalados en la primera, se presenta un modelo de trabajo alternativo que, aunque siguiendo el mismo esquema ya presentado, podría resultar preferible en aquellos casos en que los supuestos a que nos referimos en el párrafo anterior no resulten aceptables.

3. Aspectos generales de una solución

Dada la diversidad de soluciones para el problema presentado en el apartado anterior, cualquier nueva propuesta de solución que se haga se desarrollará normalmente en dos fases aparentemente independientes:

Fase 1: De tipo metodológico, consiste en la presentación del procedimiento empleado para la resolución.

Fase 2: De carácter competitivo, tratará de justificar la utilidad de la propuesta, comparando su capacidad de resolución con la de otros procedimientos generalmente aceptados. Para ello suelen compararse los resultados de todos ellos frente a problemas concretos que, como los métodos de referencia, suelen elegirse de entre un conjunto más o menos estandarizado. Analicemos con algo más de detalle cada una de estas fases:

3.1. Fase 1

Los dos elementos que integran esta fase son la elección y cálculo de la medida de discrepancia Δ y la estimación del parámetro θ , cuestiones que, aunque aparentemente puedan parecer independientes no lo son, ni mucho menos, en todos los casos.

Para justificar esta última afirmación, consideremos, de nuevo, que en general es $\Delta(\theta) = \Delta(y, f_{\theta}(x))$, por lo que su forma funcional y el objetivo de minimizarla pueden llegar incluso a definir el procedimiento de cálculo de las estimaciones, como ocurre, por ejemplo, en el método de mínimos cuadrados y, con carácter mucho más general, con

procedimientos algorítmicos como el de Levenberg-Marquardt o cualquier procedimiento de tipo gradiente o newtoniano.

La influencia en sentido contrario resulta implícitamente del razonamiento anterior, ya que si la estimación se hace sin tener en consideración la forma funcional de Δ , la discrepancia entre observaciones y predicciones no resultará necesariamente mínima y la elección de parámetros efectuada será, por tanto, ineficiente.

Trasladada a otro terreno estadístico, el de las medidas de posición y dispersión, esta concordancia entre estimadores y medidas de discrepancia se traduce en el hecho de que cada medida de dispersión se encuentra asociada a una medida de posición que la minimiza. Por citar los dos ejemplos más importantes y que serán los referidos en este documento, si es X una variable aleatoria, su valor medio $\mu = E(X)$, supuesta la existencia, minimiza la dispersión cuadrática media mientras que su mediana $Me = Me(X)$ minimiza el valor esperado de la desviación absoluta [2]

$$\begin{aligned} \min_T E(X - T)^2 &= E(X - \mu)^2 \\ \min_T E|X - T| &= E|X - Me| \end{aligned}$$

situaciones ambas que se mantienen conceptualmente si se consideran conjuntos de datos y no la variable aleatoria en sí.

Por tanto, es de esperar que la utilización de medidas de discrepancia basadas en el criterio mínimo cuadrático determine que en los procesos de estimación subsiguientes la media muestral desempeñe un papel importante, mientras que si se opta por valorar las desviaciones entre lo observado y lo calculado en términos absolutos y no cuadráticos ese papel le corresponda a la mediana. La diferencia entre ambas situaciones es apreciable. Son conocidas las propiedades e inconvenientes de una y otra medida como estadísticos de posición que, naturalmente, serán trasladados a los correspondientes estimadores obtenidos mediante su empleo.

La media muestral resultara una buena base para el cálculo de estimadores cuando la distribución presente evidentes indicios de simetría, sus colas no resulten ser excesivamente pesadas y, por tanto, la presencia de observaciones anómalas sea altamente improbable. El paradigma de esta situación lo constituye la distribución

normal, y de hecho, cuando la hipótesis de normalidad puede ser asumida, los estimadores obtenidos mediante el procedimiento mínimo cuadrático gozan de propiedades óptimas que los hacen prácticamente inmejorables.

Si bien en la práctica esta situación se presenta o puede ser supuesta con bastante frecuencia, también es cierto que no se trata de una hipótesis que deba ser automáticamente asumida y que de hecho es inaceptable en bastantes situaciones, especialmente cuando se opera con conjuntos de datos de pequeño tamaño, procedentes de distribuciones manifiestamente asimétricas o en los que se hallen presentes observaciones atípicas.

En presencia de estas u otras “irregularidades” la media deja de ser la medida óptima de posición, los estimadores basados en su cálculo manifestarán tendencia a comportarse de manera poco precisa y, por las razones anteriormente expuestas, el criterio de medida del error no debiera ser mínimo cuadrático.

3.2. Fase 2

Con aparente independencia del procedimiento o procedimientos empleados en la que hemos llamado fase 1, surge la necesidad de comprobar la eficacia del método construido lo que, como ya hemos señalado, se lleva a cabo haciéndole competir con otros algoritmos generalmente aceptados en la resolución de problemas más o menos estandarizados.

Para llevar a cabo esta tarea de prueba, los métodos actualmente vigentes se basan en la comparación de los resultados ofrecidos por uno y otros mediante el empleo de técnicas estadísticas, normalmente muy estandarizadas: el mero cálculo de estadísticos descriptivos (normalmente medias y desviaciones típicas), el uso de estadísticos *t* de Student y procedimientos de análisis de varianza podrían citarse entre los medios más empleados, si bien es éste un campo en el que el nivel de sofisticación técnica sigue en aumento.

En todo caso e independientemente de la complejidad de los procedimientos que se empleen, en la mayor parte de los casos sigue presentándose una circunstancia que vincula esta fase con la anterior: el implícito supuesto de normalidad. Tanto si se recurre sólo al cálculo de la media y desviación del conjunto de resultados y mucho más aún si se hace uso de la distribución *t* de Student o de los métodos clásicos de análisis de

la varianza el supuesto de normalidad subyace: la distribución *t* de Student surge asociada a la valoración de la media de una distribución normal cuya varianza es desconocida y debe ser estimada, mientras que los procedimientos ANOVA se basan en comparar las desviaciones cuadráticas con respecto a la media que pueden o no ser explicadas por la presencia de un determinada fuente de variación, siendo su principal supuesto la normalidad de la distribución analizada.

De nuevo, por tanto, como en la fase anterior, y quizá de forma mucho más evidente, la asunción de que es la media la medida idónea de posición y por tanto la variabilidad cuadrática la correspondiente medida de dispersión, está implícita en estos procedimientos que, por otra parte, no parecen ser pruebas excesivamente robustas frente a alteraciones del supuesto de normalidad.

4. Estimadores de posición robustos

Nuestra propuesta consiste en considerar, especialmente para aquellos casos en que el supuesto de normalidad deba ser descartado, la estimación de parámetros basada no en la media muestral, sino en la mediana, estadístico mucho más robusto por su mayor insensibilidad a la presencia de “irregularidades”.

Esta decisión debe venir asociada, por las razones expuestas en el punto anterior, a un cambio en la elección de la medida de discrepancia Δ basada consecuentemente en la desviación absoluta y no en la cuadrática, coherente, como ya se ha señalado, con la elección de la media como medida de dispersión.

Como consecuencia de lo anterior, las medidas de comparación propias de la que hemos llamado fase 2 deberán también replantearse en términos de dicho estadístico.

Un esquema de cuál sería el modo de construir uno de estos procedimientos podría ser el siguiente:

4.1. Fase 1

Sean dados el conjunto de datos $\{(x_i, y_i), i = 1, \dots, n\}$ y el espacio funcional $F = \{f_{\theta}(x), \theta \in \Theta \subset \mathbb{R}^p\}$ completamente definido salvo por el valor del parámetro θ .

Si consideremos la medida de discrepancia asociada a la norma L_1

$$\Delta(\theta) = \sum_{i=1}^n |y_i - f_{\theta}(x_i)| \quad (5)$$

la estimación del vector de parámetros θ vendrá determinada por la resolución del problema

$$\min_{\theta \in \Theta} \Delta(\theta)$$

lo que normalmente requerirá de la aplicación de algún algoritmo de optimización de tipo gradiente o newtoniano.

4.2. Fase 2

Si se pretende poder establecer comparaciones objetivas a partir de los resultados obtenidos mediante el procedimiento definido en la fase anterior resulta imprescindible hacer uso de alguna distribución asociada a la mediana, que es la medida de posición asociada a la función de dispersión elegida.

Consideraremos a este respecto el siguiente resultado, adaptación a nuestro problema del presentado por Chuan Goh [5]:

Sea X_1, \dots, X_N una muestra aleatoria simple procedente de una población con función de densidad f continua y positiva en un entorno de $Me(X)$. Sea $X_{(1)} \leq \dots \leq X_{(N)}$ la correspondiente

muestra ordenada, si definimos $k = \lfloor \frac{N}{2} \rfloor + 1$, siendo $\lfloor \cdot \rfloor$ la función parte entera, y $m < k$ es un entero independiente de N .

Si consideramos el estimador de Siddiqui-Bloch-Gastwirth de una función de la mediana [10]

$$\hat{f}(Me(X)) = S_{mN} = \frac{N}{2m} (X_{(k+m)} - X_{(k-m)}) \quad (6)$$

el estadístico

$$t_{mN} = \frac{2\sqrt{N}}{S_{mN}} (X_{(k)} - Me(X)) \quad (7)$$

tiene distribución límite, cuando $N \rightarrow \infty$, S con m grados de libertad, cuya función de densidad viene dada por

$$g_s(x, m) = \frac{1}{2m\Gamma(2m)\sqrt{2\pi}} \int_0^\infty t^{2m} e^{-\frac{t}{2}(\frac{x^2}{4m^2} + 2)} dt \quad (8)$$

$x \in \mathbb{R}$

Se trata de una distribución simétrica con respecto al origen, que se aproxima a la normal cuando $m \rightarrow \infty$ pero de colas significativamente más pesadas que las de ésta, lo que resulta coherente con el contexto de trabajo en que ha sido planteada su construcción y si bien no podrá ser obtenida exactamente la distribución de probabilidad correspondiente a este tipo de variables, la misma puede ser tabulada en función del parámetro m del mismo modo que ocurre con el parámetro asociado a los grados de libertad de la distribución t de Student cuando utilizamos la media como estimador de la localización de una variable distribuida normalmente, y a la que, en cierto modo, viene a sustituir en el contexto de trabajo establecido.

Por tanto, aunque con una distribución de referencia distinta, el tipo de operación presentado en la que hemos llamado fase 2 podrá mantenerse, al menos parcialmente, considerando como estadísticos de comparación la mediana muestral y la desviación media absoluta respecto de la misma (en lugar de la media y la desviación típica) y podremos utilizar el estadístico de contraste S (en lugar del estadístico de contraste t presentado en el apartado anterior) que permitirá contrastar por ejemplo hipótesis de igualdad de eficacia en mediana de un algoritmo frente a otro cuando las distribuciones de los errores que miden la eficacia no sean normales y/o no sean simétricas.

5. Operadores de cruce robustos en algoritmos genéticos con codificación real

Una de las posibles aplicaciones de la distribución propuesta para la mediana estará asociada a la definición de operadores de cruce en algoritmos genéticos con codificación real, AGCR, en la forma:

Sea β_i el gen i -ésimo del cromosoma de un individuo, para $i = 1, \dots, q$ y suponemos que estas son variables aleatorias independientes con distribuciones continuas $H(\beta_i)$, con un

parámetro de localización definido como μ_{b_i} . Además, consideramos el modelo $\beta_i = \mu_{b_i} + e_i$, para cada $i=1, \dots, q$, siendo e_i una variable aleatoria.

Si suponemos que los n mejores individuos forman una muestra aleatoria simple $(\beta_{i1}, \beta_{i2}, \dots, \beta_{iN})$ de la distribución de los mejores individuos de la población utilizada por el AGCR, β_i^b , el modelo se puede escribir como:

$$\beta_{ij}^b = \mu_{b_i} + e_{ij}, \quad j=1, \dots, N \quad (9)$$

Entonces, utilizando este modelo, podemos buscar un estimador para el parámetro de localización μ_{b_i} del gen i -ésimo basándonos en el método propuesto en las secciones anteriores, de forma tal que ahora si utilizamos un algoritmo de gradiente

descendente, $S_1(\mu_{b_i}) = -\frac{d\Delta(\mu_{b_i})}{d\mu_{b_i}}$, donde la

función de dispersión inducida por Δ , según la norma L_1 es $\Delta(\mu_{b_i}) = \sum_{j=1}^N |\beta_{ij} - \mu_{b_i}|$.

El estimador, es la mediana muestral $\hat{\mu}_{b_i} = M_{b_i}$ de la distribución de los β_{b_i} [2]. La mediana muestral es mejor estimador del parámetro de localización que la media muestral cuando se desconoce la forma de la distribución H .

Si consideramos la distribución del estadístico S , dada una muestra de los n genes i -ésimos asociados a otros tantos mejores individuos, el intervalo de confianza para un nivel de significación α , se construye utilizando el método de studentización. Como los resultados de las simulaciones indican la habilidad que los valores críticos de S tienen para controlar su tamaño, para valores relativamente pequeños de m , tomaremos $m=2$, y entonces

$$M_{b_i} = S_{2N} = \frac{N}{4} \left(\beta_{i \left(\left\lfloor \frac{N}{2} \right\rfloor + 3 \right)} - \beta_{i \left(\left\lfloor \frac{N}{2} \right\rfloor + 1 \right)} \right) \quad (10)$$

y utilizando el estadístico t_{2N} , tenemos

$$\frac{8}{\sqrt{N}} \left(\frac{\beta_{i \left(\left\lfloor \frac{N}{2} \right\rfloor + 1 \right)} - Me(b_i)}{\beta_{i \left(\left\lfloor \frac{N}{2} \right\rfloor + 3 \right)} - \beta_{i \left(\left\lfloor \frac{N}{2} \right\rfloor + 1 \right)}} \right) \xrightarrow{N \rightarrow \infty} S_2 \quad (11)$$

y si consideramos un coeficiente de confianza del $(1-\alpha)\%$, el intervalo de confianza bilateral será

$$I_{1-\alpha}(Me(b_i)) = \left[\beta_{i \left(\left\lfloor \frac{N}{2} \right\rfloor + 1 \right)} \pm S_{2, \alpha/2} \frac{\sqrt{N}}{8} (\beta_{i \left(\left\lfloor \frac{N}{2} \right\rfloor + 3 \right)} - \beta_{i \left(\left\lfloor \frac{N}{2} \right\rfloor + 1 \right)}) \right]$$

Si observamos la función de distribución de S_2 , tabulada en [3] y tomamos $1-\alpha=0.95$, tenemos que $S_{2, 0.025}$ toma un valor de 3.1392, con lo que podemos calcular el intervalo para un valor de N dado.

El intervalo propuesto no depende de la distribución H de los genes, lo que es de una importancia manifiesta en nuestro problema de evolución, ya que las distribuciones de los genes de los mejores padres presumiblemente cambiarán a lo largo de la evolución. Otra ventaja que tiene este intervalo frente a otros donde utilizamos la distribución binomial asociada a la mediana [2], es que no es necesario interpolar valores puesto que la distribución asintótica de S es continua. A partir de este intervalo de confianza contaremos en cada generación con tres padres para poder hacer el cruce: la mediana, el extremo inferior del intervalo y el extremo superior, estos padres se han obtenido teniendo en cuenta características estadísticas de los mejores individuos de la población a través de los estadísticos robustos de centralización y de dispersión. Este tipo de operadores de cruce están siendo empleados en optimización de funciones multimodales y epistáticas con resultados prometedores [11-12].

6. Conclusiones

Hemos presentado en este artículo un método alternativo para la construcción de modelos adecuados para la resolución de problemas de regresión o clasificación que, si bien sigue la misma estrategia que los habitualmente empleados, difiere de éstos en la función de valoración del error, el procedimiento para la

estimación de parámetros y las medidas de evaluación de sus resultados.

Con independencia de que dentro de este mismo esquema caben sin duda otras muchas elecciones, queremos resaltar dos hechos implícitos en nuestro planteamiento de carácter general:

En primer lugar, que existe una conexión inevitable entre las fases que integran la construcción de un modelo, la medida interna de su efectividad y la validación de su eficacia frente a otros, de manera que las decisiones que sobre estos aspectos se tomen no deben de ser, a nuestro entender, adoptadas independientemente las unas de las otras.

En segundo lugar que, a su vez, estas decisiones deben apoyarse en la naturaleza del conjunto de datos observados. Esta afirmación es evidente en lo que a la elección de la forma funcional del conjunto F se refiere, pero pensamos que no tanto en cuanto a resolución del problema a partir de ese momento. La información que sobre la estructura probabilística de las variables observadas contiene el conjunto de datos debe servir de apoyo para elegir convenientemente no sólo el tipo de función que se desea emplear, sino también cuál debe ser la forma adecuada de medir las diferencias entre los valores observados para la variables dependiente y su estimación, con todo lo que ello implica si consideramos la anterior conclusión.

Referencias

- [1] Hoaglin D., Mosteller, F., and Tukey, J. Understanding robust and exploratory data analysis. Wiley Classics Library. 2000
- [2] Hettmansperger T, and McKean J. Robust nonparametric statistical methods. Kendall's Library of statistics 5. Arnold. London. 1998
- [3] Chuan G. Smoothing choice and distributional approximations for econometric inference. PhD dissertation, University of California, Berkeley. 2004
- [4] Gutenbrunner, C., Jurecková, J., Koenker R. And Portnoy S. Test of linear hypotheses based on regression rank scores. Journal of Nonparametric Statistics 2, 307-333. 1993
- [5] Horowitz, J. L. Bootstrap methods for median regression models. Econometrica 66, 1327-1351. 1998
- [6] He, X and Hu, F. Markov chain marginal bootstrap. Journal of the American Statistical Association 97, 783-795. 2002
- [7] Koenker, R., and Machado J. A. F. Goodness of fit and related inference processes for quantile regression. Journal of the American Statistical Association 94, 1296-1310. 1999
- [8] Ortiz D., Hervás C. and Muñoz J. Genetic algorithm with crossover based on confidence intervals as an alternative to traditional nonlinear regression methods. European Symposium in Artificial Neural Networks, Brujas. 2001.
- [9] Ortiz D., Hervás C. and Muñoz J. Genetic algorithm with crossover based on confidence intervals as an alternative to least squares estimation for nonlinear models. Metaheuristic International Congress, Oporto. 2001.
- [10] Siddiqui, M. M. Distribution of quantile in samples from a bivariate population.. Journal of Research of the National Bureau of Standards -B. Mathematics and Mathematical Physics 64B, 145-150.
- [11] Ortiz D., Hervás C., García N. CIXL2: A crossover operator for evolutionary algorithms based on population features". Journal of Artificial Intelligence Research. 2004 (Aceptado)
- [12] Hervás C., Ortiz D. Analyzing the statistical features of CIXL2 crossover offspring. Soft Computing. Vol 9, nº 4 pp 270-279. 2005