LA DISTRIBUCION DE CAUCHY, SU UTILIZACION EN EL ANALISIS DE LA REGRESION CON LA NORMA L_∞

Carlos Narciso Bouza Herrera
Universidad de La Habana, Cuba
Luis Carlos Martinez
Universidad de A Coruña, España

RESUMEN:

En el análisis de la regresión, utilizando la norma L_∞ o de Tchebychev para el modelo lineal Y = AX+ ε , esta última componente que es aleatoria, en el caso óptimo se distribuye según una distribución de Cauchy. Su utilización en el modelo es útil si queremos, por ejemplo, estudiar el incremento del valor de las acciones motivado por determinados fenómenos naturales y sociales que afectan a la economía, considerados como “choques”, en los que si usamos la distribución normal, la probabilidad de ocurrencia es prácticamente nula, debido a la delgadez de sus colas. Una vez efectuado el ajuste, trataremos de ver si los errores se pueden considerar que se distribuyen según una distribución de Cauchy. Dicho problema nos lleva a considerar las diversas estimaciones de los parámetros de la distribución de Cauchy. La estimación de dichos parámetros nos conduce a resolver un sistema de 2n ecuaciones, por lo tanto conviene estudiar otros métodos alternativos. Su aplicación en un ejemplo económico es la base de éste trabajo.

Palabras clave: Regresion de Tchebychev
Norma L- infinito
Norma de Tchebychev
Distribuciones Alfa estables
Máxima verosimilitud

Esta función es también llamada Distribución de Lorentz o Distribución de Breit-Wigner. Una interpretación económica que podemos dar se basa en el estudio del incremento del valor de las acciones motivado por determinados fenómenos naturales y sociales que afectan a la economía considerados como "choques", en la que si usamos la distribución normal, su probabilidad de ocurrencia prácticamente sería nula, debido a la delgadez de sus colas, por lo tanto trataremos de que en un ajuste del tipo Y = AX + ε , esta ultima componente ε que es aleatoria se distribuya según una distribución de Cauchy, en cuyo caso el ajuste que debo emplear para ser óptimo es mediante la norma L_∞ o norma de Tchebychev.

Su función de densidad es de la forma $f(x, \theta, \lambda) = (\pi\lambda)^{-1}[1+\{(x-\theta)/\lambda\}^2]^{-1}$ para $\lambda>0$. Su función de distribución es $F(x, \theta, \lambda) = ½ + \pi^{-1} \tan^{-1}[(x-\theta)/\lambda$ Los momentos divergen para $n \geq 1$ por lo que no podemos estimar los parámetros λ, θ por el método de los momentos.

La distribución de Cauchy es un caso particular de una familia de distribuciones conocidas como α-estables, que son una generalización de una distribución de Gauss. El concepto de leyes estables fue introducido en 1925 por el matemático francés Paul Lévy, (1886-1971) como resultado de obtener la suma de variables aleatorias idénticamente distribuidas, que de acuerdo con el Teorema Generalizado del Límite, la suma de un gran número de v.a. idénticamente distribuidas debe ser una distribución estable, Zolatorev (1986).

Definición . Una v.a. X continua se llama "*estable*" si y solamente si $\forall$ k y toda familia X_1, …,X_k idénticamente distribuidos con la misma distribución que X, si $\exists$ $a_k > 0$ y b_k reales, tales que $X_1 + \ldots + X_k \xrightarrow{d} a_k X + b_k$

Cuando $b_k = 0$ se llama "*estrictamente estable*".

Consecuencias. *Se puede demostrar que existe una constante α, $0 < \alpha \leq 2$, tal que a_k = $k^{\frac{1}{\alpha}}$ para todo $k \in N^+$. Puede verse la demostración detallada en Feller (1971, pag 170-171).*

Teorema. Una v.a. continua X es el límite en distribución de la v.a. continua $\dfrac{(X_1 + \Lambda + X_n - b_n)}{a_n} \xrightarrow{d} X$, $a_n > 0$, si y solo si X es estable. La demostración puede verse en Shiryayev (1984, pags 338-339).

La función característica de estas distribuciones viene expresada a partir del siguiente corolario.

Corolario. (Lévy-Khinchin) Si X es una distribución estable, entonces su función característica se escribe:

$$\varphi_X(t) = e^{\left[i\mu t - \gamma|t|\left[1 + i\beta sig(t)\frac{2}{\pi}\ln|t|\right]\right]}, \quad si\ \alpha = 1$$

$$\varphi_X(t) = e^{\left[i\mu t - \gamma|t|^\alpha\left[1 - i\beta sig(t) tg\frac{\alpha\pi}{2}\right]\right]}, \quad si\alpha \neq 1.$$

La demostración puede verse en Gnedenko y Kolmogorov (1968).

Una distribución α- estable, llamada también de Lévy, depende de cuatro parámetros

- α : Es el parámetro principal o índice de estabilidad o índice de cola o exponente de cola o exponente característico $0<\alpha \leq 2$, que determina la razón mediante la cual las colas de la distribución disminuyen. Cuanto más disminuye α mas pesadas o gordas son las colas. Esto es por lo que se llaman α-estables.
- μ : Es el parámetro de localización, la media de la distribución (si $\alpha > 1$)
- γ : Es el parámetro de dispersión. Por ejemplo, si $\alpha = 2$ es la mitad de la varianza.
- β : Es el parámetro de simetría, no confundir con simétrica respecto al origen. Si $\beta=0$, la distribución es simétrica respecto a μ. $-1 \leq \beta \leq 1$. Si β es positivo, la distribución es asimétrica a la derecha, esto es la cola de la derecha es más densa. Si β es negativo es asimétrica a la izquierda. Si lo aplicamos a la distribución Normal y a la distribución de Cauchy, tenemos:

En el caso de la distribución Normal, $\alpha = 2$ y Obteniendo $\varphi_X(t) = e^{it\mu - \frac{\sigma^2}{2}t^2}$ y en el caso de la distribución de Cauchy, $\alpha = 1, \quad \beta = 0$

al poner $\gamma = \lambda \quad y \quad \mu = \theta \qquad \varphi_X(t) = e^{it\mu - \lambda|t|}$. Sabemos que el parámetro de dispersión λ no es la varianza, ya que no la tiene. Dada la función de densidad de Cauchy $f(x, \theta, \lambda) = (\pi\lambda)^{-1}[1+\{(x-\theta)/\lambda\}^2]^{-1}$, en el caso estándar en que $\theta = 0, \lambda = 1$, su función de densidad queda:

$f(x) = \pi^{-1}(1+x^2)^{-1}$. En el caso mas sencillo $f(x)=\pi^{-1}(1+x^2)^{-1}$ es una t de Student con 1 grado de libertad, si la comparamos con la N(0,1) notamos sus diferencias. Comparándola con la distribución Normal ésta tiene las colas mas gruesas. Así al comparar el valor de F(x) en 4,2 que en la distribución normal prácticamente es 1, en la distribución de Cauchy obtenemos el valor de x es 3200 para obtener la probabilidad de 0,999901.Este mayor peso o grueso de las colas se ve muy bien en la siguiente representación gráfica de las funciones de densidad Normal (0,1) y Cauchy (0,1) superpuestas, como se ve en la figura 1.

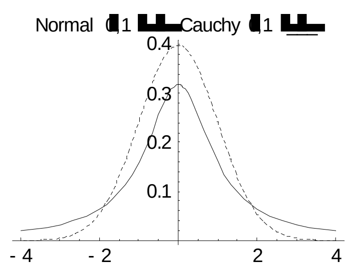


Fig. 1

Vamos a hallar sencillas estimaciones de λ. Para lo cual nos referimos a la figura 2 en donde vemos una distribución de Cauchy junto a una normal (0; 0.5). Si queremos estimar λ, procedemos del modo siguiente: Hallamos la altura máxima y la dividimos por 2, que resulta en la figura igual a 0,3183. Hallamos la proyección sobre el eje x de este valor en la función. La abscisa de este punto es la estimación deseada, en nuestro caso $\lambda = 0.5$.

La justificación de esto es la siguiente:

Sea $Y_{máx} = \dfrac{1}{\pi\lambda}$ que ocurre en x = 0, $\dfrac{y_{máx}}{2} = \dfrac{1}{2\pi\lambda} = f(x) = \dfrac{\lambda}{\pi(x^2+\lambda^2)}$,

De donde despejando x de la ecuación $x^2 - 2\lambda + \lambda^2 = 0$ da como solución doble $x = \lambda$, que es dicha estimación.

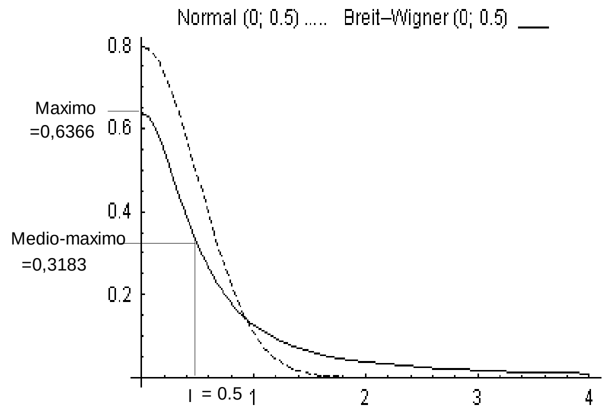
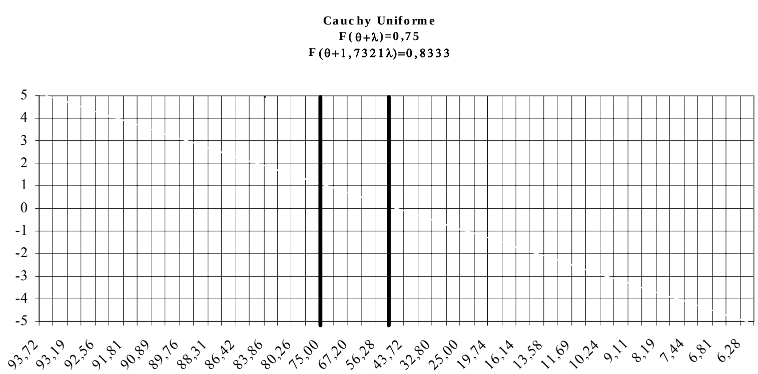


Fig.2

Esto permite hacer la estimación utilizando un histograma con un número suficientemente grande de intervalos. Otra forma de estimarlo es a partir de la función de distribución: F(x) = $\dfrac{1}{2} + arc\ tg(\dfrac{x-\theta}{\lambda})$, en nuestro caso $\theta = 0$ y $x = \lambda$, por lo tanto obtenemos:

F(x) = $\dfrac{1}{2} + arc\ tg\ 1 = \dfrac{1}{2} + \dfrac{1}{\pi}\dfrac{\pi}{4} = 0.75$,que es el percentil 75 el que corresponde a $x = \lambda$. Si hubiese un origen θ, el percentil 75 correspondería a $\theta + \lambda$.

También podemos hacer esto mediante el papel Cauchy-Uniforme.



Otras estimaciones de los parámetros de θ y λ. de la distribución de Cauchy.

La estimación de parámetros por máxima verosimilitud nos conduce a resolver las ecuaciones del sistema:

$$\begin{cases} \displaystyle\sum_{j=1}^{n}\left[1+\{(x_j-\hat{\theta})/\hat{\lambda}\}^2\right]^{-1} = \frac{1}{2}n \\ \displaystyle\sum_{j=1}^{n} x_j\left[1+\{(x_j-\theta)/\lambda\}^2\right]^{-1} = \frac{1}{2}n\hat{\theta} \end{cases}$$

el cual tiene 2n soluciones. Si el valor de λ es conocido, entonces, $\hat{\theta}$ satisface la ecuación:

$$\sum_{j=1}^{n}(x_j-\hat{\theta})\left[\ 1+\{(x_j-\hat{\theta})/\lambda\}^2\ \right]^{-1}=0$$

Otro estimador insesgado de θ es la mediana.Vamos a utilizar la norma $L\infty$, que se sabe que es óptima si los errores se distribuyen según una distribución de Cauchy

La norma L_∞ o de Tchebyshev, se define como:

$\|x\|_\infty$: $(x_1, x_2,\ldots, x_n) \in \mathbf{R}^n \rightarrow \|(x_1, x_2,\ldots, x_n)\|_\infty = \text{máx}\{|x_1|,|x_2|,\ldots,|x_n|\} = \underset{i=1,\ldots,n}{\text{máx}}|x_i|, \forall i$

Minimizando la Norma-∞ es equivalente a minimizar el máximo elemento, u optimización min-max. o mini$\|x\|_\infty$ minimización del máximo de las desviaciones absolutas o (MINMAXAD).Aplicándolo al problema de la regresión lineal nos resultaría un problema de programación lineal. Podemos considerar el problema de la siguiente forma:

Se dispone del modelo:

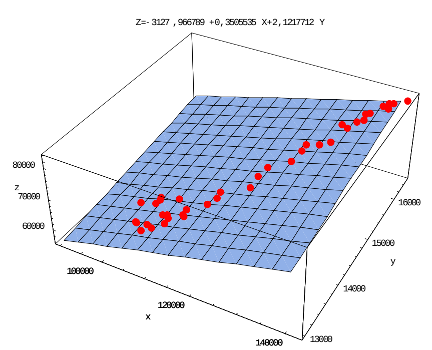
$Y = X\beta + \varepsilon$ $\quad y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \varepsilon_i$, $\quad \varepsilon_i = y_i - \beta_0 - \beta_1 x_{1i} - \ldots - \beta_j x_{ji} - \ldots - \beta_k x_{ki} =$ $y_i - \sum_{j=0}^{k}\beta_j x_{ji}$ Dicho mínimo de $d = \max_{1\le i\le n}|y_i - \sum_{j=0}^{k}\beta_j x_{ji}|$, lo calculamos consideramos la siguiente transformación $d = \frac{1}{b}$ y es equivalente a Maximizar b sujeto al sistema:

$$by_i + \sum_{j=0}^{k} b_j x_{ji} + w_i = 1, \quad 1 \le i \le n, \quad 0 \le w_i \le 2, \quad 0 \le j \le 2 = k$$

los parámetros del plano ajustado vienen dados por $\beta_0 = -db_0 \quad \beta_1 = -db_1 \quad \beta_2 = -db_2$

Como aplicación de las distintas normas se han tomado del INE (Instituto Nacional de Estadística de España www.ine.es), los datos del Gasto en Consumo Final Nacional de los Hogares Españoles a precios constantes(expresada en millones de €, que se ha tomado como variable dependiente Y que es función de X1= Producto Interior Bruto a precios constantes en millones de €. y de X2 = Empleo por rama de actividad (Ocupados) en miles de personas, durante el período 1990-2003 (primer trimestre).Se ajustó el modelo lineal obteniendo

$$\hat{Y} = -3127{,}966789 + 0{,}3505535 X_1 + 2{,}1217712 X_2$$



¿Cómo funcionó la regresión $L\infty$?

Primeramente obtuvimos los errores, y estimamos los parámetros λ y θ de la distribución de Cauchy, ya que sabemos que los errores en el ajuste L∞ son óptimos si la distribución es de Cauchy. Por último efectuaremos una prueba de la bondad del ajuste.

Empezaremos por la estimación de θ y de λ mediante el sistema de ecuaciones:

$$\sum_{j=1}^{n}\left[1+\{(x_j-\hat{\theta})/\hat{\lambda}\}^2\right]^{-1}=\frac{1}{2}n$$

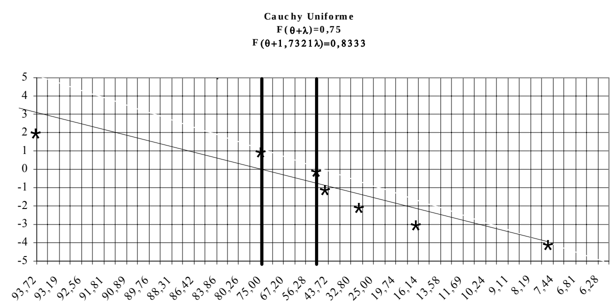
$$\sum_{j=1}^{n}x_j\left[1+\{(x_j-\theta)/\lambda\}^2\right]^{-1}=\frac{1}{2}n\hat{\theta}$$

sabemos que tiene 2n ecuaciones. Se han hecho estimaciones, bien utilizando la hoja de cálculo Excel, mediante tanteo, variado los parámetros hasta obtener valores próximos a cero en la ecuación anterior, y también utilizando las ecuaciones

$$\begin{cases}\tilde{\lambda}=½(\hat{X}_p-\hat{X}_{1-p})tan(\pi(1-p)) \\ \tilde{\theta}=½(\hat{X}_p+\hat{X}_{1-p})\end{cases}$$ con p = 0,8 > ½ y 1- p = 0,2 < ½ obtenemos para

n = 53, como estimaciones $\hat{\theta}=714{,}9067 \quad y \quad \hat{\lambda}=729{,}23$

Vemos que esta solución no nos satisface, pues como otro estimador de θ es la mediana esta será aproximada a 947,5636 . Hacemos otra estimación y obtenemos:



En este último grafico se han agrupado los N= 53 errores en 7 intervalos de clase, de acuerdo con la expresión debida a Sturges, que da el número de intervalos n:

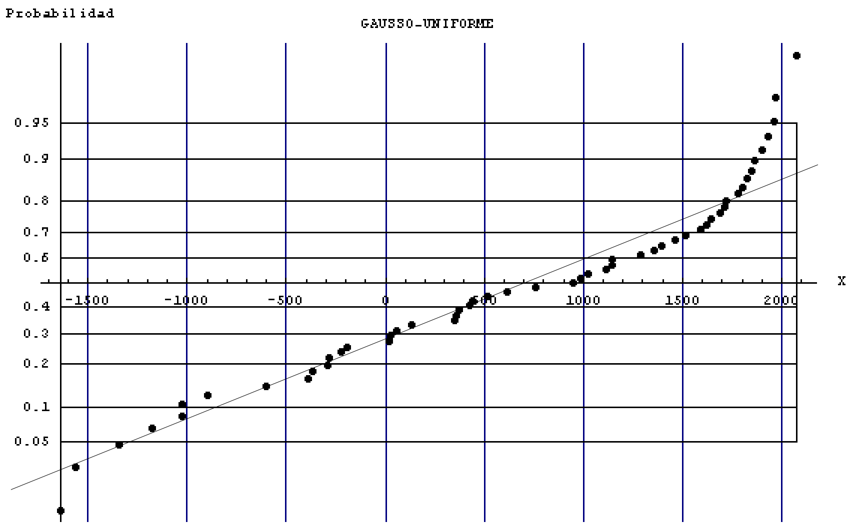
$$n=E\left(\frac{3}{2}+\frac{\ln N}{\ln 2}\right)=E\left(1{,}5+\frac{3{,}97029}{0{,}69315}\right)=E(7{,}2279)=7$$, donde E es la parte entera.

De donde obtenemos $\chi^2_2=23{,}59$ que rechazamos a cualquier nivel usual de significación el que los errores se distribuyan según una distribución de Cauchy.

Se observa un mayor peso en el último intervalo, que es bueno que ocurra, ya que la distribución es simétrica. Haciendo las modificaciones en la escala superior, obtenemos:

Que nos da lugar al gráfico anterior. De dicho gráfico obtenemos las estimaciones:

$\tilde{\theta}=645 \quad y \quad \tilde{\lambda}+\tilde{\theta}=1050 \quad de \quad donde \quad \tilde{\lambda}=405$ Por otra parte utilizando el papel Cauchy-aritmético, pero con las observaciones individuales, obtenemos el siguiente gráfico, de donde observamos también el mayor peso en el extremo derecho, no dando lugar a una recta. Si representamos las frecuencias acumuladas relativas en un gráfico Gausoaritmético, donde en el eje vertical está la función de distribución de la Normal (0,1), tenemos una representación mas lineal (recta de Henry), de donde concluimos que dichos errores no se muy bien a la distribución de Cauchy, en cambio si se ajustan mejor a una distribución normal.



Conclusiones: Los métodos gráficos con el papel Cauchy-uniforme parecen confirmar su rapidez en el cálculo de las estimaciones.

Referencias:

Norman L. Johnson & Samuel Kotz: Continuous univariate distributions-1 y 2. John Wiley & Sons. Ney York

T. S. Arthanari & Yadolah Dodge: Mathematical Programming in Statistics John Wiley & Sons. New York