

La distribución Cuantil-Beta

por

JULIO MIRÁS

Instituto Galego de Estatística

RESUMEN

Se estudia la familia paramétrica de distribuciones univariantes $F(x)$, definida por el modelo de función cuantil $Q(u) = F^{-1}(u) = K u^{a-1} (1-u)^{b-1}$; $K > 0$; $a > 1$; $0 < b < 1$, donde u es una variable con distribución uniforme en $0 < u < 1$. Es una generalización de la distribución Logarítmico-logística que corresponde al caso particular: $a+b=2$. La función cuantil de su logaritmo: $Z(u) = L(K) + (a-1)L(u) + (b-1)L(1-u)$, es una generalización de la distribución Logística.

Palabras clave: Función cuantil, Distribuciones continuas, Distribución Log-logística, Curva de Lorenz, Índice de Gini.

Clasificación AMS: 60E05, 62E10.

1. INTRODUCCIÓN

En este artículo se estudia una familia paramétrica de distribuciones univariantes $F(x)$, para variables no negativas, definida por un modelo de función cuantil $Q(u) = F^{-1}(u)$, donde u es una variable con distribución uniforme en $0 < u < 1$. El análisis cuantil que utiliza $Q(u)$ y $q(u) = dQ(u)/du$, no se enfrenta sino que complementa al análisis frecuencial, que utiliza $F(x)$ y $f(x) = dF(x)/dx$. Parzen (1979) desarrolla los puntos de vista de Tukey (1962, 1977) y con el uso de la función cuantil intenta un punto de vista simultáneamente paramétrico y no paramétrico. Wilk y

Gnanadesikan (1968) fueron pioneros en la promoción de la función cuantil muestral para el análisis estadístico de datos y mencionan una larga lista de ventajas.

Una de las más antiguas técnicas de análisis de distribuciones desde el punto de vista de su desigualdad, que tiene su origen en las ideas de Pareto (1897), Lorenz (1905) y Gini (1922), es esencialmente una técnica de análisis cuantil. En este caso se utiliza frecuentemente la Curva de Lorenz que existe para toda variable positiva con media finita y que determina a $F(x)$ salvo un parámetro de escala.

La Función de tasa de fallo: $h(x) = f(x)/(1-F(x))$, y la Elasticidad de la distribución: $\varepsilon(x) = dL(F(x))/dL(x)$, también son funciones asociadas a una distribución, utilizadas en el análisis estadístico de variables tamaño. La primera determina totalmente a $F(x)$ y la segunda la determina salvo un parámetro de escala.

Estas funciones pueden ser estudiadas haciendo la transformación integral $u=F(x)$, $0 < u < 1$, lo que permite tratar los datos muestrales X_i , $i=1, \dots, n$, como el resultado de n variables aleatorias $X_i = Q(u_i)$, donde u_i se distribuyen con arreglo a alguna distribución en el intervalo unidad $0 < u < 1$; en particular, si el muestreo es aleatorio simple las u_i son uniformes e independientes.

Este trabajo tiene un precedente en Mirás (1995) que propone dos modelos paramétricos de función cuantil para la distribución personal de la renta. Está organizado de la siguiente forma: en §2 se establecen los principales elementos del análisis estadístico cuantil de una distribución univariante. En §3 se define el modelo de función cuantil objeto del artículo y se estudian sus principales características estadísticas. En §4 se estudia brevemente la distribución de variable logaritmo. En §5 se presentan varios criterios de ajuste y un ejemplo. En §6 se presentan las distribuciones muestrales de algunos estadísticos relevantes y un ejemplo. En §7 se estudian dos estadísticos del Índice de Gini y por último, en §8 las principales transformaciones y otro ejemplo.

2. ELEMENTOS PRELIMINARES

2.1. Suponemos que X es una variable aleatoria absolutamente continua, con función de distribución $F(x)$ y de densidad $f(x) > 0$ en el interior de su soporte que es un intervalo de \mathbb{R} denotado por S_X . Denotaremos por \mathbf{F} a la clase de las distribuciones $F(x)$ que cumplen estas condiciones. En particular nos referiremos al caso de variables no negativas con $S_X = \{x > 0\}$.

Denotaremos por $Q(u) = F^{-1}(u)$ la Función cuantil de X que es la función inversa de F . También denotaremos por $X(u)$ el cuantil de orden u de la distribución $F(x)$ y es inmediato ver que si $0 < u < 1$ es una variable uniforme, $X(u) = Q(u)$ se distribuye según F . Seleccionando números aleatorios en el intervalo $0 < u < 1$ podemos generar

muestras aleatorias de X si conocemos la función cuantil. Siguiendo la terminología de Tukey (1965), llamamos *Sparsity* a la función:

$$q(u) = Q'(u) = dQ(u)/du \quad [2.1]$$

y *Densidad cuantil*, a la función:

$$f(Q(u)) = 1/Q'(u) = 1/q(u) \quad [2.2]$$

siendo $f(t) = dF(t)/dt$. La relación entre la función densidad cuantil y la función sparsity es inmediata puesto que $F(t)$ y $Q(u)$ son funciones una inversa de la otra, de modo que sus derivadas son recíprocas. Parzen (1979) denomina *quantile-density* a [2.1] y *density-quantile* a [2.2]. La función [2.2] es la función de densidad de X calculada en el punto $x = Q(u) = F^{-1}(u)$.

2.2. Para el cálculo de momentos y esperanzas matemáticas, tenemos:

$$\alpha_r = E\{X^r\} = \int_0^1 \{Q(u)\}^r du; \quad E\{\varphi(X)\} = \int_0^1 \varphi(Q(u)) du \quad [2.3]$$

2.3. En el contexto de análisis estadístico cuantil, se utiliza la *Función Score*, que se define como: $J(u) = \frac{-f'(Q(u))}{f(Q(u))}$, que también puede escribirse como:

$$J(u) = \frac{-d(1/q(u))}{du} = \frac{q'(u)}{q^2(u)}, \text{ o también: } J(u) = \frac{1}{q(u)} \frac{dL(q(u))}{du} \quad [2.4]$$

2.4. La versión cuantil de la Función de tasa de fallo $h(x) = \frac{f(x)}{1-F(x)}$, es:

$$h(Q(u)) = \frac{1}{(1-u)q(u)} \quad [2.5]$$

2.5. La elasticidad cuantil de la distribución, que se define como: $\varepsilon(x) = \frac{dL(F(x))}{dL(x)}$, se expresa inmediatamente en su versión cuantil, como:

$$\varepsilon(u) = \frac{Q(u)}{uq(u)} \quad [2.6]$$

2.6. Para el estudio de la longitud de la cola de la distribución, denotado por γ_t , adopto el criterio de Parzen, utilizado por Alzaid y Al-Osh (1989) y por Schuster (1984). Se define la longitud de la cola como:

$$\gamma_t = \lim_{u \rightarrow 1} \frac{(1-u)J(u)}{f(Q(u))} \quad [2.7]$$

que por [2.4] puede escribirse como:

$$\gamma_t = \lim_{u \rightarrow 1} \frac{(1-u)dL(q(u))}{du} \quad [2.8]$$

Análogamente defino el peso de la cabeza, g_h , como:

$$\gamma_h = \lim_{0 \leftarrow u} \frac{uJ(u)}{f(Q(u))} = \lim_{0 \leftarrow u} \frac{udL(q(u))}{du} \quad [2.9]$$

Si $\gamma_t < 1$, la distribución se denomina de longitud de cola corta, si es $\gamma_t = 1$, de cola media y si es $\gamma_t > 1$ de cola larga. Esta clasificación toma como referencia el valor $\gamma_t = 1$ que corresponde a la distribución exponencial y a la distribución de Weibull. La clasificación según el peso de la cabeza es un asunto menos tratado en la literatura, aunque tiene importancia en el estudio de la desigualdad de la distribución de magnitudes tales como ingreso o gasto de las familias, salarios, etc. Una distribución tiene cabeza más pesada cuanto mayor sea γ_h ; como referencia tenemos: $\gamma_h = 0$, para la distribución exponencial; para la distribución de Weibull, en todo caso es $\gamma_h > -1$, pudiendo tomar valores positivos.

En todo caso la longitud de la cola γ_t , y el peso de la cabeza γ_h , son invariantes en los cambios de origen y escala. Si $Q^*(u) = x_0 + cQ(u)$, se tiene $q^*(u) = cq(u)$, de donde $dL(q^*(u)) = dL(q(u))$, resultando que [2.8] y [2.9] no se modifican.

2.7. Distribución de los cuantiles muestrales.

a) La distribución asintótica del cuantil de orden $0 < u < 1$, $\hat{Q}(u)$, en muestras de tamaño n , para valores no extremos de u , cuando existe la densidad y es positiva, es Normal con media y varianza:

$$E(\hat{Q}(u)) \approx Q(u) ; V(\hat{Q}(u)) \approx \frac{u(1-u)}{nf(Q(u))^2} = \frac{u(1-u)q^2(u)}{n} \quad [2.10]$$

También es conocido, como resultado más general, véase por ejemplo Galambos (1991), que cualquier subconjunto de k cuantiles empíricos, ninguno próximo a los extremos, tiene como distribución asintótica la Normal k -variante con medias y varianzas [2.10] y covarianzas:

$$\text{Cov}(\hat{Q}(u), \hat{Q}(v)) \approx \frac{u(1-v)}{nf(Q(u))f(Q(v))} = \frac{u(1-v)q(u)q(v)}{n} \quad [2.11]$$

En particular para la Mediana muestral, tenemos:

$$E(\hat{Q}(1/2)) \approx Q(1/2) = M; \quad \sigma(\hat{Q}(1/2)) \approx \frac{q(1/2)}{2\sqrt{n}} \quad [2.12]$$

donde M denota la mediana del modelo.

b) La función de distribución exacta del estadístico de orden $1 \leq j \leq n$, denotado por $X_{(j)}$, en muestras de tamaño n , es:

$$P\{X_{(j)} < x\} = \sum_{r=j}^n \binom{n}{r} u_{(x)}^r (1-u_{(x)})^{n-r} \quad [2.13]$$

para cuyo cálculo en un punto x , es necesario resolver la ecuación inversa $u_{(x)} = Q^{-1}(x)$. También:

$$P\{X_{(j)} < x\} = j \binom{n}{j} \int_0^{u_{(x)}} t^{j-1} (1-t)^{n-j} dt = B(u_{(x)}; j, n-j+1) \quad [2.14]$$

donde $B(t; p, q)$ denota el valor de la distribución beta de parámetros p, q en el punto t .

c) La función de densidad cuantil exacta del estadístico de $X_{(j)}$, se obtiene in-

mediatamente. De: $f_j(x) = j \binom{n}{j} F(x)^{j-1} (1-F(x))^{n-j} f(x)$, tenemos

$$f_j(x) = j \binom{n}{j} u_{(x)}^{j-1} (1-u_{(x)})^{n-j} \frac{1}{q(u_{(x)})} = \beta(u_{(x)}; j, n-j+1) \frac{1}{q(u_{(x)})};$$

$$0 < u_{(x)} = Q^{-1}(x) < 1 \quad [2.15]$$

donde $\beta(t; p, q)$ denota la función de densidad beta.

2.8. Función de intensidad, Curva de Lorenz e Índice de Gini

2.8.1. Definimos la *Función Intensidad* de una distribución F con soporte $S_X = \mathbb{R}^+$ y media finita, como la función cuantil dividida por su media:

$$C(u) = \frac{1}{\mu} Q(u) \quad [2.16]$$

Teniendo en cuenta la definición habitual de la Curva de Lorenz para distribuciones de clase F : $L(u) = \frac{1}{\mu} \int_0^u F^{-1}(t) dt = \frac{1}{\mu} \int_0^u Q(t) dt$, resulta que la derivada de la

Curva de Lorenz es justamente la función intensidad de la distribución que denotamos por: $C(u) = dL(u)/du$. Observemos que $C(u)$ presenta las características formales de una función de densidad en el intervalo unidad y es justamente la derivada de Radon-Nicodým de la curva de Lorenz respecto de $F(x)$ en el cuantil de orden $0 < u < 1$: $dL(F(x))/dF(x) = dL(u)/du = C(u)$. Una imagen práctica es la siguiente: si tenemos que repartir una cantidad total, entre los individuos de una población (familias, empleados, empresas, etc.) $C(u)du$ representa la proporción del total que corresponde a los individuos ubicados en el entorno du , si se ordenasen equidistantes de cero a uno, de menor a mayor participación en el reparto.

La Curva de Lorenz es invariante por un cambio de escala, pero se modifica si se produce una traslación de la función cuantil; si $X^*(u) = x_0 + Q(u)$, su nueva expresión en función del modelo $Q(u)$, tras un sencillo cálculo, es:

$$L^*(u) = \frac{x_0}{\mu} u + \frac{\mu - x_0}{\mu} L(u), \text{ o también: } L^*(u) = L(u) + \frac{x_0}{\mu} (u - L(u))$$

de donde resulta que si el modelo inicial experimenta una traslación positiva $x_0 > 0$, la nueva Curva de Lorenz domina a la anterior.

2.8.2. El índice de Gini, se define como $G = 2S - 1$, donde S es el área comprendida en el cuadrado unidad sobre la curva $L(u)$. Más conveniente que esta definición geométrica es su definición analítica, según la cual G es dos veces la covarianza entre la intensidad y el orden del cuantil:

$$G = 2\text{Cov}\{u, C(u)\} = 2E\{uC(u)\} - 1 \quad [2.17]$$

o también:

$$G = \frac{2}{\mu} \text{Cov.}\{u, Q(u)\} = \frac{2}{\mu} E\{uQ(u)\} - 1 \quad [2.18]$$

La equivalencia de ambas definiciones se comprueba inmediatamente ya que: $S = \int_0^1 \frac{1}{\mu} uQ(u) du = E\{uC(u)\}$, teniendo en cuenta que $E\{u\} = 1/2$ y $E\{C(u)\} = 1$.

El índice de Gini es invariante en los cambios de escala pero no en las traslaciones. Se obtiene inmediatamente el siguiente resultado: Si $X^*(u) = x_0 + cX(u)$, siendo μ la media de $X(u)$, el índice de Gini de X^* , es :

$$G^* = \frac{\mu}{X_0 + \mu} G \quad [2.19]$$

3. DISTRIBUCIÓN CUANTIL-BETA

Denominamos *Distribución Cuantil-beta* a la familia paramétrica de distribuciones univariantes con soporte R^+ , caracterizada por el siguiente modelo de función cuantil:

$$Q(u) = Ku^{a-1}(1-u)^{b-1}; 0 < u < 1 \quad [3.1]$$

donde K es un parámetro de escala o centralización y a, b son parámetros de estructura. En notación abreviada escribiremos: $QB(K; a, b)$. Para establecer el espacio paramétrico, observemos que:

3.a.1. Para $a=1, b=1$, la distribución concentra su masa en un único punto $Q(u)=K$, para todo $0 < u < 1$.

3.a.2. Para $b=1$, tenemos la distribución potencial, $F(x) = A x^B$, con $B=1/(a-1)$, cuyo soporte es un intervalo finito: $0 < x < K = A^{-1/B}$.

3.a.3. Para $a=1$, tenemos la distribución de Pareto $F(x) = 1 - (k/x)^B$, con $B=1/(1-b)$, cuyo soporte es el intervalo real $x > K$.

Estudiaremos el modelo formado por el conjunto de distribuciones definidas por [3.1] cuyo soporte es R^+ en el espacio paramétrico:

$$\Theta : K > 0; a > 1; 0 < b < 1 \quad [3.2]$$

En estas condiciones $Q(u)$ es una función positiva y monótona creciente en $0 < u < 1$ que sirve de modelo para la función cuantil de una variable aleatoria con media finita que toma valores reales positivos. Para valores $a < 1$, o $b > 1$ [3.1] no es una función cuantil y para $b < 0$ la distribución existe pero no tiene media. Excepto en el caso particular $a+b=2$, no es posible resolver explícitamente la ecuación inversa para obtener F , de modo que efectuaremos el estudio de esta familia de distribuciones utilizando su función cuantil, su función sparsity y su densidad cuantil.

b). El caso particular $a+b=2$, corresponde a una variable con función cuantil:

$$X(u) = Q(u) = k \left(\frac{u}{1-u} \right)^{a-1} \quad [3.3]$$

cuya función de distribución es la conocida Distribución Logarítmico-logística:

$$F(x) = \frac{\alpha x^\beta}{1 + \alpha x^\beta}; \beta = 1/(a-1); \alpha = \left(\frac{1}{k} \right)^{1/(a-1)}, \text{ para } x > 0 \quad [3.4]$$

de forma que podemos interpretar el modelo (3.1) como una generalización de la familia Logarítmico-logística. Tomando logaritmos en (3.3), tenemos:

$$Z = L(X(u)) = L(k) + (a-1)L\left(\frac{u}{1-u}\right) \quad [3.5]$$

de modo que el Logaritmo de la variable es una función lineal del Logit de la distribución. Esta propiedad permite un análisis gráfico preliminar: si disponemos de K cuantiles empíricos \hat{X}_j que acumulan una frecuencia \hat{F}_j , la nube de puntos que se

forma con $Z_j = L(\hat{X}_j)$ en ordenadas y $\phi_j = L\left(\frac{\hat{F}_j}{1-\hat{F}_j}\right)$ en abscisas, que denominamos

gráfico Log-logístico, se concentrará en torno a una línea recta. En el caso general el modelo [3.1] permite representar distribuciones cuyo gráfico Log-logístico se aparta de la linealidad, presentando un cierto grado de concavidad o convexidad.

3.1. Linealidad del Logaritmo del cuantil respecto del Logit

Denotando por $\phi(u) = L\left(\frac{u}{1-u}\right)$ el Logit de la distribución y por $Z = L(X(u))$, el logaritmo de la variable, calculamos la primera y segunda derivadas de Z respecto de ϕ

para estudiar el comportamiento del gráfico Log-logit. Tenemos: $Z'_\phi = \frac{dL(Q(u))}{d\phi(u)}$,

$Z''_\phi = \frac{dZ'_\phi}{d\phi(u)}$, teniendo en cuenta que $dL\phi(u) = \frac{du}{u(1-u)}$, el cálculo es inmediato:

$$Z'_\phi = (a-1) - (a+b-2)u ; Z''_\phi = -(a+b-2)u(1-u)$$

y en el espacio paramétrico, $a>1 ; 0<b<1$, tenemos :

- En todo caso $Z'_\phi > 0$ para todo $0<u<1$ que refleja el carácter creciente del Logaritmo del cuantil respecto de su Logit.
- $Z''_\phi < 0$ para $(a+b) > 2$. Z_ϕ es convexo.
- $Z''_\phi > 0$ para $(a+b) < 2$. Z_ϕ es cóncavo.
- Si $(a+b)=2$ el gráfico logit es una recta y el modelo corresponde a la distribución Log-logística.

Definimos entonces el Coeficiente de linealidad:

$$\delta = (a+b)/2 \quad [3.6]$$

como un parámetro descriptivo de la estructura de la distribución. El valor $\delta = 1$ corresponde a la distribución Log-logística. Para $\delta > 1$ el gráfico $L(X(u))$ vs. $L(u/(1-u))$ es convexo, y para $\delta < 1$ es cóncavo. Véase ejemplo 5.7 y figura 3.

3.2. Momentos de la distribución Cuantil-beta y del producto cruzado

$$\eta = XF(X)$$

El momento de orden r respecto del origen existe si $b > \frac{r-1}{r}$, para $r=1,2..$ y viene dado por:

$$\alpha(r) = k^r B(1+r(a-1), 1+r(b-1)) \quad [3.7]$$

donde $B(p,q)$ denota la función beta completa. En particular la distribución tiene media finita si $b>0$, tiene varianza finita si $b>1/2$. Tenemos:

$$E(X) = \mu = KB(a,b); \quad \sigma_X^2 = K^2(B(2a-1,2b-1) - B^2(a,b)) \quad [3.8]$$

Su cálculo es inmediato: $\alpha(r) = E(X^r) = k \int_0^1 u^{r(a-1)} (1-u)^{r(b-1)} du$, y la condición de existencia se deriva de las propiedades de la función beta.

También será útil en el análisis la variable producto cruzado $\eta = XF(X)$. En particular los momentos:

$$E(\eta) = E(uX(u)) = \int_0^1 k u^a (1-u)^{b-1} du = kB(1+a, b) = \frac{\mu}{a+b} \quad [3.9]$$

$$\sigma_\eta^2 = V(\eta) = E(\eta^2) - (E(\eta))^2 = \mu^2 \left(\frac{B(1+2a, 2b-1)}{B^2(a, b)} - \left(\frac{a}{a+b}\right)^2 \right) \quad [3.10]$$

$$\sigma_{\eta, X} = \text{Cov}(\eta, X(u)) = \mu^2 \left(\frac{B(2a, 2b-1)}{B^2(a, b)} - \frac{a}{a+b} \right) \quad [3.11]$$

3.4. Función de densidad cuantil

3.4.1. De acuerdo con §2.1, la función sparsity y la densidad cuantil vienen dadas por:

$$q(u) = \frac{k\{(a-1) + (2-(a+b))u\}}{u^{2-a}(1-u)^{2-b}} \quad [3.12]$$

$$f_X(u) = \frac{u^{2-a}(1-u)^{2-b}}{k\{a-1 + (2-(a+b))u\}}; \quad 0 < u < 1 \quad [3.13]$$

Dentro del espacio paramétrico definido, la densidad en la cola de la distribución ($u \rightarrow 1$), es decreciente. La cabeza de la distribución depende de los valores paramétricos, así resulta: para valores $1 < a < 2$, las densidades tienen forma de U invertida. Para $a=2$, la densidad es monótona decreciente con valor finito (igual a $1/k$) en $u=0$ y para $a > 2$ también es monótona decreciente pero se hace infinita en $u=0$. En el caso particular de la distribución Log-logística, cuando tiene media finita, por ser $(a+b)=2$; $a > 1$ y $b > 0$, es siempre $1 < a < 2$ y por tanto la densidad tiende a cero en la cabeza y en la cola. El cuadro 1 presenta las principales características descriptivas para distintos valores de los parámetros a, b de distribuciones QB con media igual a 1. Se observa que dentro del modelo existen distribuciones con

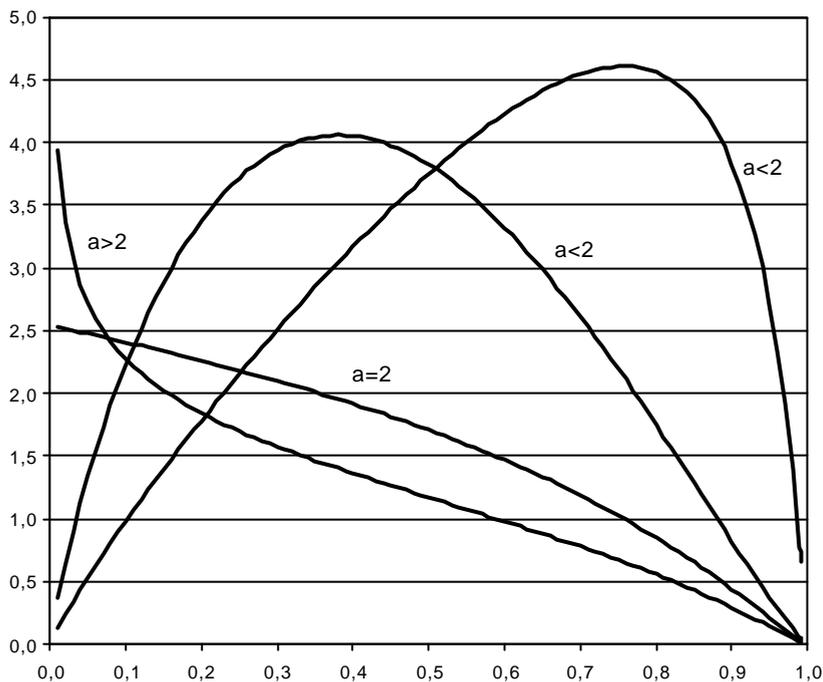
asimetría positiva o negativa ($m_3 / \sigma^{3/2}$ mayor o menor que cero) y con curtosis (m_4 / σ^2) mayor o menor que 3. Véase figura 1.

Cuadro 1

DISTRIBUCIÓN CUANTIL-BETA CON MEDIA=1
CARACTERÍSTICAS DESCRIPTIVAS PARA ALGUNOS VALORES DE LOS
PARÁMETROS ESTRUCTURALES

<i>a</i>	<i>b</i>	<i>Media</i>		<i>Momentos</i>				<i>Índice</i>	<i>Coef. de</i>	<i>Coef. de</i>	<i>Coef. de</i>
		<i>a1</i>	<i>a2</i>	<i>a3</i>	<i>a4</i>	<i>Gini</i>	<i>Variación</i>	<i>Asimetría</i>	<i>Curtosis</i>		
1,10	0,60	1	1,9542	-	-	0,294	0,977	-	-		
1,20	0,60	1	2,1152	-	-	0,333	1,056	-	-		
1,30	0,60	1	2,2807	-	-	0,368	1,132	-	-		
1,40	0,60	1	2,4491	-	-	0,400	1,204	-	-		
1,50	0,60	1	2,6198	-	-	0,429	1,273	-	-		
2,00	0,60	1	3,4909	-	-	0,538	1,578	-	-		
2,00	0,60	1	3,4909	-	-	0,538	1,578	-	-		
2,20	0,60	1	3,8440	-	-	0,571	1,686	-	-		
1,10	0,70	1	1,2966	4,1285	-	0,222	0,545	13,860	-		
1,20	0,70	1	1,3760	4,9160	-	0,263	0,613	12,091	-		
1,30	0,70	1	1,4604	5,7872	-	0,300	0,679	10,904	-		
1,40	0,70	1	1,5481	6,7397	-	0,333	0,740	10,094	-		
1,50	0,70	1	1,6381	7,7723	-	0,364	0,799	9,530	-		
2,00	0,70	1	2,1073	14,119	-	0,481	1,052	8,409	-		
2,20	0,70	1	2,2997	17,207	-	0,517	1,140	8,306	-		
1,10	0,80	1	1,1069	1,4345	2,5918	0,158	0,327	3,250	43,333		
1,20	0,80	1	1,1563	1,6226	3,2747	0,200	0,395	2,485	29,556		
1,30	0,80	1	1,2114	1,8371	4,0974	0,238	0,460	2,086	22,769		
1,40	0,80	1	1,2703	2,0745	5,0671	0,273	0,520	1,874	19,035		
1,50	0,80	1	1,3320	2,3332	6,1932	0,304	0,576	1,764	16,809		
2,00	0,80	1	1,6615	3,9207	14,5421	0,429	0,813	1,740	13,319		
2,20	0,80	1	1,7987	4,6875	19,3573	0,467	0,894	1,809	13,166		
1,10	0,90	1	1,0343	1,1087	1,2369	0,100	0,185	0,937	6,510		
1,20	0,90	1	1,0665	1,2061	1,4449	0,143	0,258	0,387	4,391		
1,30	0,90	1	1,1053	1,3241	1,7048	0,182	0,324	0,241	3,631		
1,40	0,90	1	1,1485	1,4582	2,0136	0,217	0,385	0,225	3,242		
1,50	0,90	1	1,1947	1,6063	2,3715	0,250	0,441	0,259	3,018		
2,00	0,90	1	1,4504	2,5236	4,9647	0,379	0,671	0,570	2,823		
2,20	0,90	1	1,5588	2,9661	6,4255	0,419	0,748	0,693	2,927		
1,10	0,95	1	1,0174	1,0518	1,1039	0,073	0,132	-0,176	3,817		
1,20	0,95	1	1,0432	1,1257	1,2498	0,116	0,208	-0,431	3,256		
1,30	0,95	1	1,0760	1,2197	1,4393	0,156	0,276	-0,396	2,846		
1,40	0,95	1	1,1135	1,3288	1,6672	0,191	0,337	-0,304	2,550		
1,50	0,95	1	1,1542	1,4504	1,9323	0,224	0,393	-0,201	2,347		
2,00	0,95	1	1,3841	2,2107	3,8433	0,356	0,620	0,246	2,066		
2,20	0,95	1	1,4825	2,5779	4,9104	0,397	0,695	0,389	2,121		

Figura 1
DENSIDAD-CUANTIL. MODELO BETA



3.4.2. Función de densidad cuantil del cuantil de orden j .

Sustituyendo (3.12) en (2.15)

$$f_{X_{(j)}}(u) = \binom{n}{j} \frac{u^{j-a+1}(1-u)^{n-j-b+2}}{K\{a-1+(2-a-b)u\}}; \quad 0 < u < 1 \quad [3.14]$$

3.4.3. Longitud de la cola y peso de la cabeza

Efectuando el cálculo según lo dicho en §2.6, teniendo en cuenta que $a > 1$; $0 < b < 1$, tenemos:

$$\gamma_t = \lim_{u \rightarrow 1} \left\{ \frac{(a-2)(1-u)}{u} + (2-b) - \frac{(a+b-2)(1-u)}{(a-1)-(a+b-2)u} \right\} = (2-b)$$

$$\gamma_h = \lim_{u \rightarrow 0} \left\{ (a-2) + \frac{(2-b)u}{1-u} - \frac{(a+b-2)(1-u)}{(a-1)-(a+b-2)u} \right\} = (a-2)$$

Resulta que las distribuciones del modelo cuantil-beta, son todas de cola larga, pero la cabeza puede ser más o menos pesada, siendo los coeficientes:

$$\gamma_t = 2 - b > 1 \quad ; \quad \gamma_h = a - 2 > -1 \quad [3.15]$$

Se observa que la longitud de la cola, siendo siempre de tipo largo, es tanto más corta cuanto mayor sea b , y en límite ($b=1$) correspondería a una distribución (3.a.2), de cola media. La cabeza es muy pesada para $a > 2$, pesada para $a=2$, y moderadamente pesada para $a < 2$. La distribución Log-logística, cuando tiene media finita, es de cabeza moderada por ser $a < 2$.

De acuerdo con estos resultados definimos el *Coefficiente de pesode* las colas:

$$\gamma = \gamma_t + \gamma_h = a - b \quad [3.16]$$

como indicador conjunto del peso de la cabeza y la cola. Cuanto menor es b mayor es la longitud de la cola, y cuanto mayor es a mayor es el peso de la cabeza.

3.5. Curva de Lorenz

De acuerdo con §2.8.1, la curva de Lorenz del modelo cuantil-beta viene dada por los valores de la Función de Distribución beta de parámetros (a,b) en el punto

$$0 < u < 1. \text{ Tenemos: } L(u) = \frac{1}{\mu} \int_0^u K t^{a-1} (1-t)^{b-1} dt = \frac{1}{B(a,b)} \int_0^u t^{a-1} (1-t)^{b-1} dt = B(u; a, b).$$

3.6. Parámetros de la distribución

El modelo de distribución cuantil-beta consta de tres parámetros, uno de centralización o escala y los otros dos de estructura. Si tomamos la media como parámetro de centralización, teniendo en cuenta [3.8] tenemos:

$$Q(u) = \frac{\mu}{B(a,b)} u^{a-1} (1-u)^{b-1}; \quad 0 < u < 1 \quad [3.17]$$

y veamos que es posible tomar como parámetros de estructura del modelo el Índice de Gini G y el Coeficiente de linealidad-logit δ .

3.6.1. Índice de Gini.

El índice de Gini de acuerdo §2.8.2, se obtiene inmediatamente de [2.18] y [3.9]:

$$G = \frac{2a}{a+b} - 1 = \frac{a-b}{a+b} \quad [3.18]$$

También se puede expresar como el doble de la razón entre el coeficiente de peso de las colas y el coeficiente de linealidad-logit: $G = 2\frac{\gamma}{\delta}$. Para el caso particular de la Log-logística tenemos: $G=a-1$.

De [3.6] y [3.18] resulta: $a = \delta(1+G)$; $b = \delta(1-G)$ y sustituyendo en [3.17] el modelo de distribución cuantil-beta puede expresarse como una familia caracterizada por tres parámetros: La media μ , el índice de Gini G y el coeficiente de linealidad δ :

$$Q(u) = \frac{\mu}{B(\delta(1+G), \delta(1-G))} u^{\delta(1+G)-1} (1-u)^{\delta(1-G)-1} \quad [3.19]$$

Como caso particular, la distribución Log-logística queda determinada por su media, cuando existe, y su Índice de Gini ya que en este caso es $\delta = 1$. Tiene como Función Cuantil:

$$Q(u) = \frac{\mu}{B(1+G, 1-G)} u^G (1-u)^{-G} = \frac{\mu}{B(1+G, 1-G)} \left(\frac{u}{1-u}\right)^G \quad [3.20]$$

3.8. La función cuantil de tasa de fallo

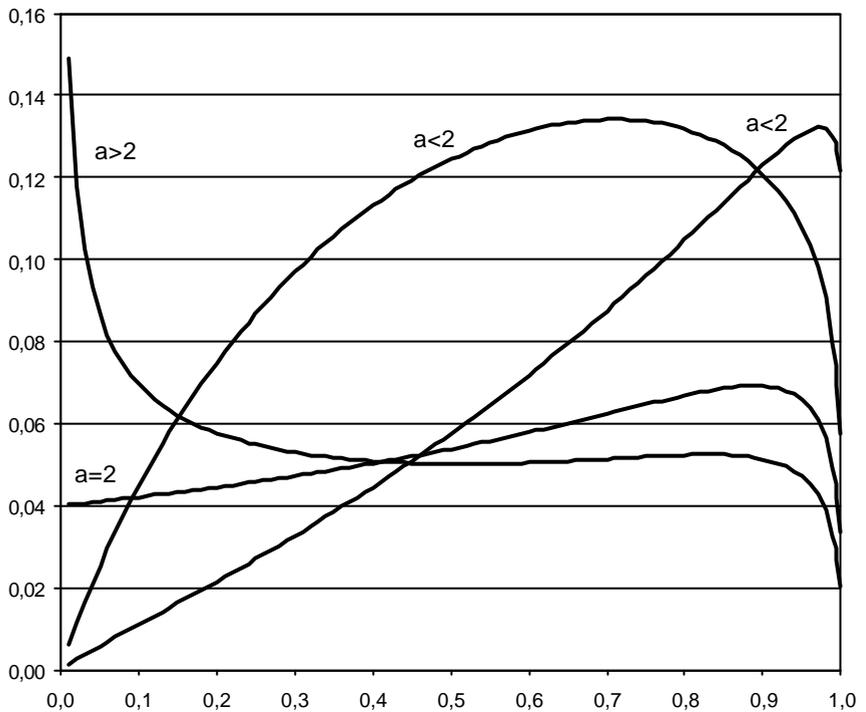
De acuerdo con [2.5], teniendo en cuenta [3.12], la función cuantil de tasa de fallo, viene dada por:

$$h(Q(u)) = \frac{u^{2-a} (1-u)^{1-b}}{k\{a-1+(2-a-b)u\}}; 0 < u < 1 \quad [3.21]$$

que en todo caso tiende a cero en la cola de la distribución ($u \rightarrow 1$). Para $a < 2$ también tiende a cero en la cabeza de la distribución ($u \rightarrow 0$), siendo creciente en un primer tramo y decreciente en el siguiente, si bien para valores de b próximos a 1 es creciente en casi todo su recorrido decreciendo a partir de un valor alto, próxi-

mo a $u=1$. Si $a=2$ toma un valor finito positivo en $u=0$, y si $a>2$ la tasa de fallo se hace infinita en $u=0$. Véase figura 2.

Figura 2
TASA DE FALLO CUANTIL. MODELO BETA



3.9. Elasticidad cuantil de la distribución

De acuerdo con [2.6] y [3.12], se obtiene:

$$\varepsilon(u) = \frac{1-u}{a-1-(a+b-2)u} \quad [3.22]$$

que en el espacio paramétrico donde está definido el modelo, es siempre una función decreciente, desde $\varepsilon(u) = 1/(a-1)$ para $u=0$, hasta $\varepsilon(u) = 0$ para $u=1$. La

distribución Log-logística ($a+b=2$) se caracteriza por ser $\varepsilon(u) = (1-u)/(a-1)$ linealmente decreciente.

4. DISTRIBUCIÓN EXPONENCIAL CUANTIL-BETA

Toda variable X cuya función de distribución sea de clase F con $S_X = \mathbb{R}^+$ tiene asociada la variable $Z=L(X)$ que también es de clase F y cuyo soporte es la recta real $S_Z = \mathbb{R}$. Su función cuantil es $Z(u)=L(X(u))$. Si tomamos el logaritmo de la Función Cuantil-beta tenemos:

$$Z(u) = C + (a-1)L(u) + (b-1)L(1-u) \quad [4.1]$$

que en el espacio paramétrico: $-\infty < C = L(K) < \infty; a > 1; 0 < b < 1$, es la Función Cuantil de una variable Z cuya exponencial tiene distribución cuantil-beta $QB(K=\exp(C); a, b)$. Esta distribución es una generalización de la Distribución Logística que resulta como caso particular de (4.1) para $a+b=2$.

La función Densidad-cuantil, es:

$$f(Z(u)) = \frac{u(1-u)}{(a-1) + (2-(a+b))u} ; 0 < u < 1 \quad [4.2]$$

y la Función Cuantil de Tasa de fallo:

$$h(Z(u)) = \frac{u}{(a-1) + (2-(a+b))u} ; 0 < u < 1 \quad [4.3]$$

La Función Generatriz de momentos, es:

$$M_Z(t) = E(e^{Zt}) = K^t \frac{\Gamma(1+(a-1)t)\Gamma(1+(b-1)t)}{\Gamma(2+(a+b-2)t)} \quad [4.4]$$

y las derivadas de orden $r=1,2$ en $t=0$ del Logaritmo de $M_Z(t)$ que se obtienen de las funciones digamma y trigamma, consultadas en Abramovitz and Stegun (1972), permiten calcular la media y varianza. La media, teniendo en cuenta que $\Gamma'(2) - \Gamma'(1) = 1$, es:

$$\begin{aligned} \mu_Z = E(Z) &= L\left(\frac{\mu}{B(a,b)}\right) + (a+b-2)(\Gamma'(1) - \Gamma'(2)) = \\ &= L\left(\frac{\mu}{B(a,b)}\right) + 2 - (a+b) = L\left(\frac{\mu}{B(a,b)}\right) + 2(1-\delta) \end{aligned} \quad [4.5]$$

La varianza, teniendo en cuenta que $\psi'(1) - \psi'(2) = 1$, y $\psi'(1) = \zeta(2) = \pi^2/6$, donde $\zeta(t)$ es la función zeta de Riemann:

$$\sigma_Z^2 = \frac{\pi^2}{3}(a-1)(1-b) + (a+b-2)^2 = \frac{\pi^2}{3}(a-1)(1-b) + 4(1-\delta)^2 \tag{4.6}$$

Para $(a+b)=2$ tenemos la distribución Logística de media $\mu_Z = C = L(K)$, y varianza $V(Z) = \frac{\pi^2}{3}(a-1)^2$, y para $a=2, b=0$ tenemos la distribución Logística, $F(z) = \exp(z)/(1+\exp(z))$, de media cero y desviación típica $\pi/\sqrt{3}$.

Nota: La función cuantil $Z(u) = \gamma + (a-1)L(u) + (b-1)L(1-u)$ corresponde a una variable cuyo soporte es la recta real, con media y varianza finitas para todo $-\infty < \gamma < \infty; a > 1; b < 1$, pero solo en el caso $0 < b < 1$ la función cuantil $Z = \exp(Z(u))$ corresponde a una variable no negativa con media finita de la familia cuantil-beta.

5. AJUSTE DE LA DISTRIBUCIÓN QB(K;A,B)

Es posible establecer diferentes criterios de ajuste del modelo a una distribución empírica. Veamos algunos.

5.1. Criterio de ajuste 1.- Consiste en realizar el ajuste por regresión lineal mínimo-cuadrática, entre los logaritmos de los cuantiles empíricos y los del modelo. Realmente es un ajuste de la distribución de la variable logaritmo. Utilizando como espaciado del intervalo unidad $u_{(j)} = \hat{F}_j = \hat{F}(X_{(j)}) ; j=1, \dots, K$, en el caso de frecuencias agrupadas, o bien $u_{(j)} = \hat{F}_j = j/(1+n); j=1, \dots, n$, en el caso de frecuencias unitarias, tomando logaritmos en el modelo de función cuantil resulta:

$L(\hat{X}_j) = L(k) + (a-1)L(\hat{F}_j) + (b-1)L(1-\hat{F}_j)$. Con las notaciones simplificadas: $Z_j = L(\hat{X}_j); C = L(k); V_j = L(\hat{F}_j); W_j = L(1-\hat{F}_j)$, tenemos el sistema de ecuaciones normales:

$$\begin{cases} \sigma_V^2 a + \sigma_{V,W} b = \sigma_{Z,V} \\ \sigma_{V,W} a + \sigma_W^2 b = \sigma_{Z,W} \end{cases}$$

donde σ_{-}^2 y $\sigma_{-,-}$ denotan las varianzas y covarianzas de los datos. La solución buscada es:

$$k = \exp(\bar{z} - (a-1)\bar{v} - (b-1)\bar{w})$$

$$a = 1 + \frac{\sigma_{Z,V} \sigma_W^2 - \sigma_{V,W} \sigma_{Z,W}}{\Delta}; \quad b = 1 + \frac{\sigma_{Z,W} \sigma_V^2 - \sigma_{V,W} \sigma_{Z,V}}{\Delta}$$

denotando por $\bar{z}, \bar{v}, \bar{w}$ valores medios y $\Delta = \sigma_V^2 \sigma_W^2 - \sigma_{V,W}^2$. De otra forma:

$$a = 1 + \frac{\sigma_Z}{\sigma_V} \frac{r_{Z,V} - r_{Z,W} r_{V,W}}{1 - r_{V,W}^2}; \quad b = 1 + \frac{\sigma_Z}{\sigma_W} \frac{r_{Z,W} - r_{Z,V} r_{V,W}}{1 - r_{V,W}^2}$$

que existen por ser el coeficiente de correlación lineal: $r_{V,W}^2 \neq 1$, ya que entre V_j y W_j existe una relación funcional que no es lineal. En el caso particular de la Distribución Logarítmico-logística, de acuerdo con (3.20) se ajusta una recta:

$L(\hat{X}_j) = L(k) + (a-1)L(\hat{F}_j / (1 - \hat{F}_j))$ cuya pendiente es una estimación del Índice de Gini que debe estar comprendida entre cero y uno ($1 < a < 2$).

5.2. Criterio de ajuste 2- Consiste en determinar los parámetros por regresión mínimo-cuadrática entre los cuantiles empíricos y el modelo. En este caso se trata de una regresión no lineal que puede intentarse utilizando algún algoritmo de búsqueda de un mínimo local partiendo de una solución inicial.

5.3. Criterio de ajuste 3- Se igualan la media, la varianza y el Índice de Gini de la distribución empírica a sus correspondientes de la distribución modelo. Para el Índice de Gini se utiliza el estadístico:

$$\hat{G} = \frac{2}{\bar{X}} \text{Cov}\{X_{(j)}, u_{(j)}\} \quad [5.1]$$

siendo $X_{(1)} \leq \dots \leq X_{(j)} \leq \dots \leq X_{(n)}$, los cuantiles muestrales y $u_{(j)} = j/(1+n)$, que es un estimador consistente de G . De los resultados obtenidos en párrafos anteriores tenemos el sistema de ecuaciones:

$$\begin{cases} \bar{x} = K B(a, b) \\ \hat{G} = \frac{a-b}{a+b} \\ \hat{\sigma}^2 = K^2 (B(2a-1, 2b-1) - B^2(a, b)) \end{cases} \quad [5.2]$$

y llevando $K = \bar{x}/B(a, b)$ de la primera y $b = \frac{1-\hat{G}}{1+\hat{G}} a$ de la segunda, a la tercera

ecuación, resulta la siguiente ecuación en a que puede resolverse por iteración numérica disponiendo de una herramienta con la función beta o gamma:

$$\hat{\sigma}^2 = \frac{\bar{x} (B(2a - 1, 2a \frac{1 - \hat{G}}{1 + \hat{G}} - 1) - B^2(a, \frac{1 - \hat{G}}{1 + \hat{G}} a))}{B(a, \frac{1 - \hat{G}}{1 + \hat{G}} a)} \quad [5.3]$$

Como caso particular, la distribución Log-logística queda determinada por la media y el Índice de Gini. Efectuando el cálculo de sus valores empíricos:

$$a = 1 + \hat{G}; K = \frac{\bar{x}}{B(1 + \hat{G}, 1 - \hat{G})} \quad [5.4]$$

5.4. Criterio de ajuste 4. Se igualan la media, la mediana y el índice de Gini de la distribución empírica a sus correspondientes de la distribución modelo. De los resultados obtenidos en párrafos anteriores tenemos el sistema de ecuaciones:

$$\begin{cases} \frac{\bar{x}}{\hat{M}} = 2^{a+b-2} B(a, b) \\ \hat{G} = \frac{a-b}{a+b} \end{cases} \quad [5.5]$$

denotando por \hat{M} la mediana empírica. Llevando como antes, b en función de a y \hat{G} de la segunda ecuación de [5.5] a la primera tenemos la ecuación en a :

$$4^{a/(1+\hat{G})} B(a, \frac{1-\hat{G}}{1+\hat{G}} a) = 4 \frac{\bar{x}}{\hat{M}} \quad [5.6]$$

cuya solución puede encontrarse por iteración numérica. El parámetro b se obtiene en el mismo proceso de resolución de [5.6] y K se obtiene de la relación: $K = \bar{x}/B(a, b)$.

5.5. Criterio de ajuste 5. Se igualan la media, el índice de Gini y la media de los logaritmos de la distribución empírica a sus correspondientes de la distribución modelo. En este caso, denotando por \bar{z} la media de los logaritmos de los datos muestrales, teniendo en cuenta [4.5] y relaciones ya utilizadas tenemos:

$$\begin{cases} \frac{\bar{x}}{B(a,b)} = K \\ \hat{G} = \frac{a-b}{a+b} \\ \frac{\bar{x}}{e^{\bar{z}}} e^{2-(a+b)} = B(a,b) \end{cases} \quad [5.7]$$

y llevando b en función de \hat{G} y a , de la segunda ecuación a la tercera de [5.7] resulta la siguiente ecuación que resolvemos en a como antes:

$$e^{-2(1+\hat{G}-a/(1+\hat{G}))} B(a, \frac{1-\hat{G}}{1+\hat{G}} a) = \frac{\bar{x}}{e^{\bar{z}}} \quad [5.8]$$

por último, b se obtiene en el mismo proceso de resolución numérica de [5.8] y K de la primera condición de [5.7].

Estos tres métodos deben de ir precedidos de algún análisis preliminar de los datos que permita admitir la adecuación del modelo y de una comprobación posterior de la precisión de la resolución numérica de las ecuaciones

5.6. Criterio de ajuste 6. Se igualan la media, el índice de Gini y la varianza de los logaritmos de la distribución empírica a sus correspondientes de la distribución modelo. En este caso, el sistema de ecuaciones:

$$\begin{cases} \hat{\sigma}_z^2 = \frac{\pi^2}{3} (a-1)(1-b) + (a+b-2)^2 \\ b = \frac{1-\hat{G}}{1+\hat{G}} a \end{cases} \quad [5.9]$$

se reduce a una ecuación de segundo grado en a , permite obtener los parámetros de estructura a y b , y el parámetro de escala se obtiene de la relación $K = \bar{x}/B(a,b)$.

5.7. Ejemplo.- Distribución de las $n=167$ secciones censales del municipio A Coaña según su población en el censo de 1991. Supongamos que esta distribución empírica procede de un modelo de población infinita $QB(K;a,b)$. Las ecuaciones [5.3], [5.6] y [5.8] se han resuelto con un error de $4,35 \cdot 10^{-8}$, $1,01 \cdot 10^{-10}$ y $7,55 \cdot 10^{-7}$. Los resultados de los seis criterios de ajuste junto con los datos muestrales entre paréntesis, aparecen en el cuadro 2.

Cuadro 2

Parámetros	Datos	Ajustes					
		C(1)	C(2)	C(3)	C(4)	C(5)	C(6)
K		1633	1538	1542	1830	1555	1548
a		1,2550	1,2283	1,2224	1,3267	1,2275	1,2245
b		0,8995	0,8649	0,8724	0,9469	0,8761	0,8740
Media	(1472)	1469	1476	1472	1472	1472	1472
Mediana	(1514)	1467	1442	1444	1514	1448	1445
Indice G	(0,1671)	0,1650	0,1736	0,1671	0,1671	0,1671	0,1671
Var(X)	(202333)	188969	221653	202333	189018	200958	201748
Media Z	(7,2458)	7,2435	7,2450	7,2461	7,2386	7,2458	7,246
Var(Z)	(0,1028)	0,1082	0,1101	0,1023	0,1319	0,1035	0,1028

El criterio $C(1)$ y el criterio $C(4)$ presentan resultados más alejados de los otros cuatro. En el primer caso se explica porque realmente se ha ajustado el modelo de la variable logaritmo. En el caso de $C(4)$ se explica porque la distribución empírica del ejemplo tiene asimetría positiva pero presenta una mediana superior a la media; esta contingencia de los datos se recoge en el criterio $C(4)$ pero no en los otros. Las figuras 3 y 4 presentan el gráfico Logit y la función cuantil del ajuste $C(2)$.

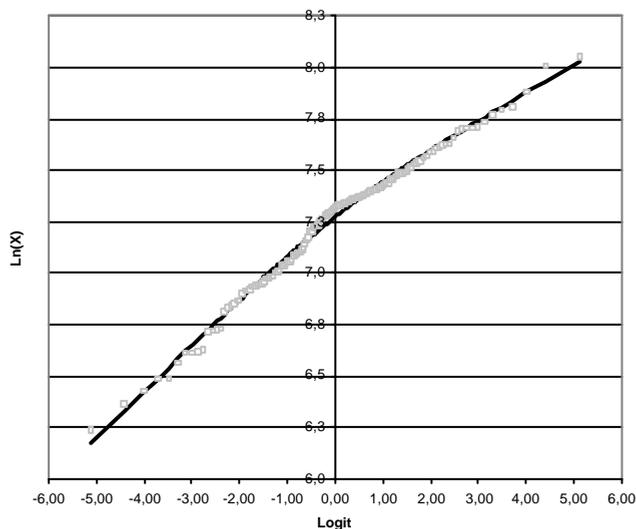
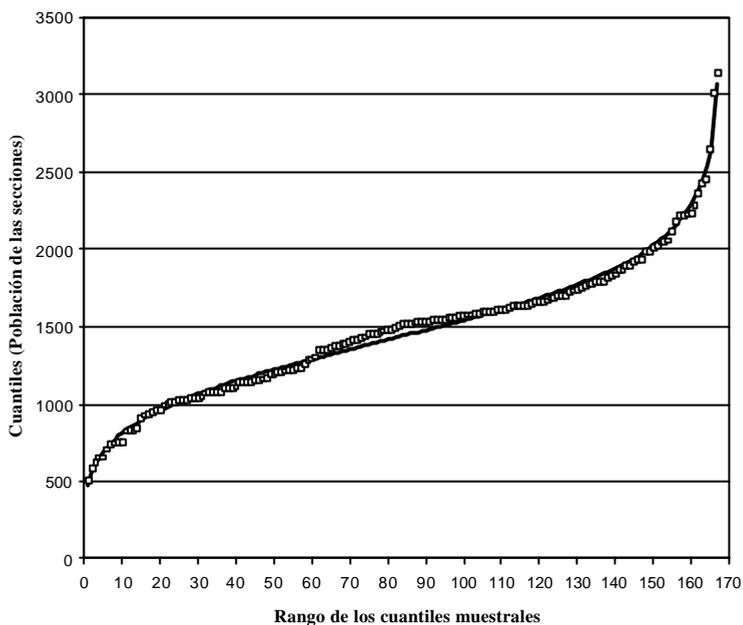
Figura 3**GRÁFICO LOG-LOGIT. AJUSTE C(2)**

Figura 4
FUNCIÓN CUANTIL.AJUSTE C(2)



6. DISTRIBUCIÓN ASINTÓTICA DE ALGUNOS ESTADÍSTICOS $QB(K;A,B)$

Para construir tests de hipótesis sobre las distribuciones de la familia paramétrica Cuantil-beta, es útil considerar los siguientes estadísticos. Suponemos que x_i ; $i=1, \dots, n$, es una muestra aleatoria simple de una variable con distribución $QB(K;a,b)$ completamente especificada.

6.1. Media

La media muestral \bar{x} es un estimador insesgado de la media μ y si $b > 1/2$ (en otro caso no tiene varianza finita) su distribución asintótica es Normal:

$$\bar{x} \approx N(\mu, \sigma_x / \sqrt{n}) \quad [6.1]$$

siendo μ, σ_x^2 la media y varianza de la distribución $QB(K;a,b)$ dadas en (3.8).

6.2. Mediana y cuantiles

La distribución asintótica de estadísticos de orden se obtiene de [2.10] sustituyendo la función $q(u)$ del modelo. En particular de acuerdo con [2.12], la mediana muestral \hat{M} es un estimador consistente de la mediana de la distribución $M=Q_x(1/2)$, y su distribución asintótica es Normal:

$$\hat{M} \approx N(M, q_x(1/2)/2\sqrt{n}) \quad [6.2]$$

siendo $q_x(u)$ dada en [3.11] la función sparsity de la distribución $QB(K; a, b)$

La distribución exacta se obtiene de [2.14] y en particular las probabilidades:

$$P\{X_{(j)} \leq x\} = j \binom{n}{j} B(j, n-j+1) B(u_{(x)}; j, n-j+1) = B(u_{(x)}; j, n-j+1) \quad [6.4]$$

se obtienen resolviendo previamente la ecuación: $u_{(x)} = Q^{-1}(x)$, siendo $Q(u)$ la función cuantil del modelo.

6.3. Coeficiente de concavidad-logit

El estadístico:

$$\hat{\delta} = 1 - \frac{1}{2}(\bar{z} - L(k)) \quad [6.5]$$

donde $\bar{z} = \frac{1}{n} \sum_1^n L(x_i)$ es la media de los logaritmos de los datos muestrales, es un

estimador insesgado del coeficiente de concavidad-logit $\delta = (a+b)/2$, y su distribución asintótica es Normal:

$$\hat{\delta} \approx N(\delta, \sigma_z^2 / 2\sqrt{n}) \quad [6.6]$$

donde $\sigma_z^2 = V(Z)$ ha sido calculada en [4.6], y es finita en cualquier punto del espacio paramétrico $a > 1; 0 < b < 1$.

6.4. Ejemplo

Los seis ajustes del modelo Cuantil-beta a la distribución empírica del ejemplo 5.7, son razonablemente buenos. Si se utiliza la media o el coeficiente de concavidad-logit como estadísticos de un test todos serían aceptados; sin embargo todos

son rechazables si se utiliza como estadístico significativo la mediana o un cuantil próximo al extremo superior. Suponiendo que la distribución empírica es una muestra aleatoria simple de una distribución teórica de la familia paramétrica QB , resulta que bajo las correspondientes a los ajustes $C(1), C(2), C(3), C(5)$ y $C(6)$, la desviación normal de la mediana $\frac{\hat{M}-M}{\sigma(M)}$, utilizando [6.2], es en cada caso: 4,105; 4,791; 4,915; 4,810 y 4,826, obviamente excesiva. El ajuste $C(4)$ iguala la mediana muestral a la del modelo por lo que este test no lo rechazaría. Sin embargo los diez cuantiles muestrales de orden $j=158$ a $j=167$ presentan valores superiores a los esperados según el modelo ajustado. En este caso por tratarse de cuantiles próximos al extremo superior, utilizamos la distribución exacta en el muestreo; para $j=158$, por ejemplo, tenemos $\hat{X}_{(158)} = 2220$ y bajo la hipótesis del ajuste $C(4)$, utilizando [6.4] se tiene: $u_{(x=2219)} = 0,97701$, y $YP\{\hat{X}_{(158)} > 2219\} = 1 - 0,9944 = 0,0056$ que conduce a rechazar la hipótesis significativamente.

7. Índice de Gini

7.1. En el caso paramétrico de una variable X con distribución $QB(K; a, b)$ completamente especificada denotemos por $R = \frac{E(\eta)}{E(X)} = \frac{E(\eta)}{\mu}$ la razón entre el valor esperado del momento cruzado y de la variable objetivo. Teniendo en cuenta que $G = 2R - 1$, consideremos el estadístico:

$$\hat{G}_1 = 2\hat{R}_1 - 1 \quad [7.1]$$

donde $\hat{R}_1 = \frac{\frac{1}{n} \sum_{i=1}^n x_i F(x_i)}{\mu} = \frac{\bar{\eta}}{\mu}$. De [3.9] resulta que [6.8] es un estimador insesgado de G y utilizando [3.10] obtenemos su varianza:

$$E(\hat{G}_1) = 2 \frac{1}{\mu} \frac{\mu a}{a+b} - 1 = \frac{a-b}{a+b} = G \quad [7.2]$$

$$V(\hat{G}_1) = \frac{4}{\mu^2} V(\bar{\eta}) = \frac{4}{n} \left(\frac{B(1+2a, 2b-1)}{B^2(a, b)} - \left(\frac{a}{a+b} \right)^2 \right) \quad [7.3]$$

7.2. Consideremos ahora el estadístico:

$$\hat{G}_2 = 2\hat{R}_2 - 1 \quad [7.4]$$

donde $\hat{R}_2 = \frac{\frac{1}{n} \sum_{i=1}^n x_i F(x_i)}{\bar{x}} = \frac{\bar{\eta}}{\bar{x}}$ es el estimador de la razón R . Se cumple:

Sesgo: \hat{G}_2 es un estimador consistente de G ; su sesgo es:

$$\text{Sesgo}(\hat{G}_2) = 2 \text{Sesgo de } (\hat{R}_2) = -\frac{2}{\mu} \text{Cov}(\hat{R}_2, \bar{x}) \quad [7.5]$$

y en valor absoluto:

$$|\text{Sesgo}(\hat{G}_2)| = \frac{2}{\mu} \left| \rho_{\hat{R}_2, \bar{x}} \sigma_{\hat{R}_2} \sigma_{\bar{x}} \right| \leq \frac{2 \sigma_{\hat{R}_2} \sigma_{\bar{x}}}{\mu} \quad [7.6]$$

y en en relación a su desviación típica:

$$\frac{|\text{Sesgo}(\hat{G}_2)|}{\sigma_{\hat{G}_2}} \leq \frac{\sigma_{\bar{x}}}{\mu} = \frac{\sigma_X}{\mu \sqrt{n}} = \frac{C_X}{\sqrt{n}} \quad [7.7]$$

donde C_X es el coeficiente de variación de la distribución objetivo.

Error cuadrático medio: El E.C.M. aproximado es:

$$\text{ECM}(\hat{G}_2) \approx \frac{4}{n\mu^2} (\sigma_{\eta}^2 + R^2 \sigma_X^2 - 2R \text{Cov}(\eta, X)) \quad [7.8]$$

y para el caso de la distribución cuantil-beta:

$R = \frac{a}{a+b} = \frac{1+G}{2}$, σ_{η}^2 , $\text{Cov}(\eta, X)$, μ , σ_X^2 se han calculado en [3.9]-[3.11] y en [3.8].

Utilizando la variable residuo: $\varepsilon = \eta - RX$, [7.8] puede escribirse como:

$$\text{ECM}(\hat{G}_2) \approx \frac{4}{n\mu^2} V(\varepsilon) \quad [7.9]$$

y una estimación de éste:

$$\hat{\text{ECM}}(\hat{G}_2) = \frac{4}{n\mu^2(n-1)} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 = \frac{4}{n\mu^2(n-1)} \sum_{i=1}^n \varepsilon_i^2 \quad [7.10]$$

con $\varepsilon_i = x_i F(x_i) - \hat{R}_2 x_i$

Los resultados [7.5]-[7.8] se obtienen utilizando la técnica de linearización habitual en el análisis del estimador por razón. Para el sesgo tenemos:

$$\text{Cov}(\hat{R}_2, \bar{x}) = E(\hat{R}_2 \bar{x}) - \mu E(\hat{R}_2) = E(\bar{\eta}) - \mu E(\hat{R}_2) = \mu(R - E(\hat{R}_2))$$

y el sesgo de \hat{R}_2 como estimador de R , es $\text{Sesgo}(\hat{R}_2) = \frac{-\text{Cov}(\hat{R}_2, \bar{x})}{\mu}$ de donde se obtiene (6.12). Para el E.C.M. tenemos:

$$\text{ECM}(\hat{R}_2) = E(\hat{R}_2 - R)^2 = E\left(\frac{\bar{\eta}}{\bar{x}} - R\right)^2 = E\left(\frac{\bar{\eta} - R\bar{x}}{\bar{x}}\right)^2 \approx \frac{1}{\mu^2} E(\bar{\eta} - R\bar{x})^2$$

y siendo $\hat{G}_2 = 2\hat{R}_2 - 1$, se obtiene [7.5].

7.3. Notas: 1/ El uso de los estadísticos \hat{G}_1 y \hat{G}_2 necesita la resolución de la ecuación inversa $F(x_i) = Q^{-1}(u_i)$ para cada uno de los n datos muestrales x_i , pero no necesita el uso de la muestra ordenada. 2/ El E.C.M. de \hat{G}_2 en la práctica puede utilizarse como varianza cuando [7.7] es pequeño. 3/ En muestreo aleatorio simple los estimadores del Índice de Gini presentan una varianza muy grande y solo en el caso de muestras de gran tamaño tendremos una precisión aceptable. En el caso de diseños más sofisticados que incluyen una buena estratificación la precisión aumentará notablemente.

7.4. Ejemplo. Para los datos del ejemplo 5.7, admitiendo la hipótesis de que la distribución tiene como parámetros los obtenidos en el ajuste $C(2)$, tenemos: $\sigma(\hat{G}_1) = 0,0728$ y $S(\hat{G}_2) = 0,0448$, con $\text{Sesgd}(\hat{G}_2) < 0,0247\sigma(\hat{G}_2)$.

8. TRANSFORMACIONES

Si $X(u)$ es una magnitud con distribución Q-beta, $X^*(u) = cX^\theta(u)$, con $c > 0$, $0 < \theta < \frac{1}{1-b}$, también es Q-beta con parámetros:

$$K^* = cK, \quad a^* = 1 + \theta(a-1), \quad b^* = 1 + \theta(b-1) \quad \text{y media finita.}$$

Para $\theta > \frac{1}{1-b}$ es $b^* < 0$ y $X^*(u)$ no tiene media.

8.1. Cambios de escala y origen

El cambio de escala está implícito en el modelo ya que si $\theta = 1$ y $c > 0$, $X^*(u)$ es siempre Q-beta, y no modifica las propiedades estructurales de la distribución.

Sin embargo las traslaciones: $X^*(u) = x_0 + X(u)$ modifican algunas propiedades relevantes. En particular:

El Gráfico Logit: $L(X(u))$ vs. $Logit(u)$, si $a+b > 2$, el es convexo, pero por ser el límite de $L(x_0 + X(u))$ finito para $u \rightarrow 0$, el Gráfico $L(X^*(u))$ vs. $Logit(u)$, tiene un primer tramo cóncavo seguido de otro convexo.

La Curva de Lorenz y el Índice de Gini se modifican según hemos visto en §2.8.2.

8.2. Ejemplo

En algunos casos se obtiene un mejor ajuste si se adopta un modelo con un valor inicial mayor que cero. En el cuadro 3 se presentan los datos de Renta disponible por unidad de consumo en 1997 de los hogares de Francia (en francos de 1996) procedentes de fuentes fiscales, tomados de Hourriez, J-M y Roux, V. (2001). En la tercera columna aparecen los valores ajustados por mínimos cuadrados del modelo Q-beta con una traslación que ubica el origen en $x_0 = 15700$. La tercera columna presenta, en porcentaje, la desviación absoluta relativa de los cuantiles ajustados respecto de los datos. Las dos últimas columnas presentan los cuantiles ajustados del modelo Q-beta y las desviaciones porcentuales. Los parámetros ajustados al pie del cuadro. La figura 5 muestra el Gráfico Logit y la figura 6 la función de densidad del modelo $x_0 + Q(u)$ ajustado, con trazo más grueso en el intervalo comprendido por los datos, desde $u = 0,1$ hasta $u = 0,95$.

Cuadro 3

MODELO X_0+Q -BETA

Orden cuantil	Cuantil de Renta	Modelo $X_0+Q(u)$ Beta		Modelo $Q(u)$ Beta	
		Renta	100*ABS	Renta	100*ABS
u	Dato $X(u)$	Renta	100*ABS	Renta	100*ABS
0,10	43602	43773	0,39	47781	9,59
0,20	55049	54739	0,56	57127	3,77
0,30	64183	63898	0,44	64690	0,79
0,40	72588	72599	0,01	71958	0,87
0,50	81221	81547	0,40	79673	1,91
0,60	91043	91431	0,43	88582	2,70
0,70	103234	103301	0,07	99878	3,25
0,80	119538	119464	0,06	116270	2,73
0,90	148495	147884	0,41	147374	0,75
0,95	179232	179550	0,18	184370	2,87

Parámetros del ajuste $X_0+Q(u)$ Beta

X_0	K	a	b
15700	73184	1,4288	0,7236

Parámetros del ajuste $Q(u)$ Beta

X_0	K	a	b
0	74257	1,2055	0,6929

Figura 5

GRÁFICO LOGIT DEL EJEMPLO 8.2

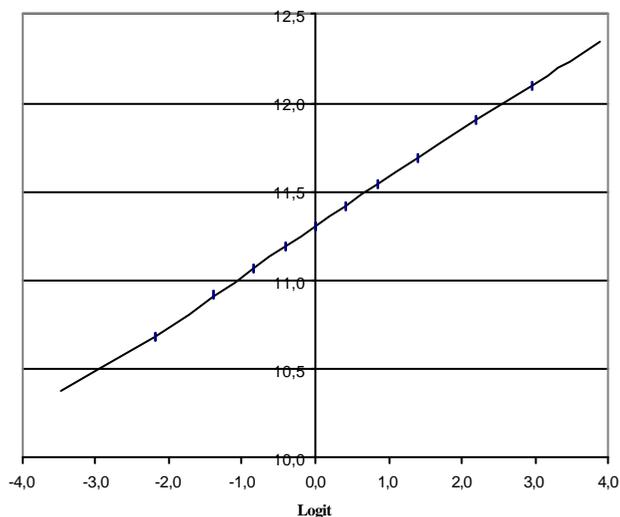
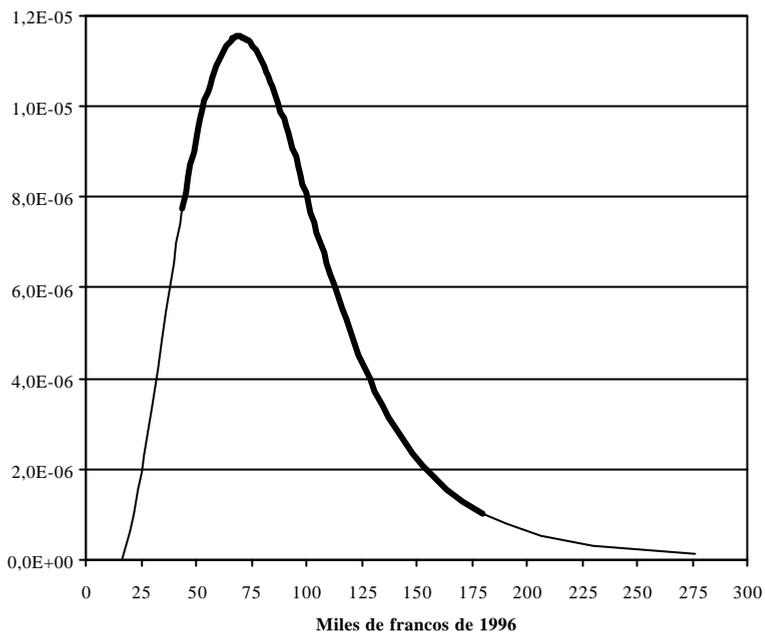


Figura 6
FUNCIÓN DE DENSIDAD DEL EJEMPLO 8.2
RENTA DISPONIBLE DE LOS HOGARES POR U.C.
FRANCIA 1997



REFERENCIAS

- ABRAMOWITZ, M. AND STEGUN, I. A. «Handbook of Mathematical Functions». Dover, New York.
- ALZAID, A. A. AND AL-OSH, M. (1989). «Ordering probability distributions by tail behavior». *Statistics & Probability Letters*, 8, 203-216.
- GALAMBOS, J. (1984). «Order Statistics». P. R. Krishnaiah and P. K. Sen, eds., *Handbook of Statistics*, vol 4, 359-382.
- GINI, C. (1922). «Indici di Concentrazione e di Dipendenza». *Biblioteca dell'Economista*, 5ª serie, vol XX.

- GNANADESIKAN, R. (1977). «Methods for Statistical Data Analysis of Multivariate Observations». New York: John Wiley & Sons.
- HOURRIEZ, J-M Y ROUX, V. (2001). «Vue d'ensemble des inégalités économiques». *Document de Travail F0103*. INSEE.
- LORENZ, M. C. (1905). «Methods of Measuring the Concentration of Wealth». *Publications of the American Statistical Association*, vol. 9, New Series.
- MIRÁS, J. (1995). «La Función Cuantil de la Renta Personal: Un marco general y dos modelos paramétricos». *Estadística Española*, vol. 37, nº. 139. 261-286.
- PARETO, W. (1897). «Cours d'économie politique». Lausanne.
- PARZEN, E. (1979). «Nonparametric Statistical Data Modeling». *JASA*, vol. 74, 365, 105-131.
- SCHUSTER, F. F. (1984). «Classification of probability laws by tail behavior». *JASA*, 74, 936-939.
- TUKEY, JOHN W. (1965). «Which Part of the Sample Contains the Information?», *Proceedings of the National Academy of Sciences*, 53, 127-134.
- TUKEY, JOHN W. (1962). «The Future of Data Analysis». *Annals of Mathematical Statistics*, 33, 1-67.
- TUKEY, JOHN W. (1977). «Exploratory Data Analysis». Reading Mass.: Addison-Wesley Publishing Co.
- WILK, M.B. AND GNANADESIKAN, R. (1968). «Probability Plotting Methods for the Analysis of Data», *Biometrika* 55, 1-17.

THE QUANTILE-BETA DISTRIBUTION.

SUMMARY

In this paper it's considered the parametric family of univariate distributions $F(x)$, definite for the model of quantile function $Q(u)=F^{-1}(u)=Ku^{a-1}(1-u)^{b-1}; K>0; a>1; 0<b<1$, where u is a variable with uniform distribution in $0<u<1$. This family is a generalized Log-logistic distribution which correspond to particular case $a+b=2$. The quantile function of logarithm: $Z(u)=L(K)+(a-1)L(u)+(b-1)L(1-u)$, is a generalized Logistic distribution.

Key words: Continuous distributions, Quantile function, Log-logistic distribution, Lorenz curve, Gini index.

AMS classification: 60E05, 62E10.