V. 8, n.º 3 (setembre 1984) pp. 127-132

Concentración y Distribuciones Truncadas

JOAN BARÓ LLINÁS I VICENS MORATAL OLIVER

ESCOLA UNIVERSITÀRIA D'ESTUDIS EMPRESARIALS
BARCELONA

Algunos coeficientes de desigualdad pueden expresarse en función de las medias de una distribución inicial y otra truncada en varios puntos críticos, en especial los cuartiles y la Mediala, aún cuando no es normal encontrarlos así reseñados en la literatura estadístico-económica.

Este trabajo reseña algunas de aquellas medidas: acumulaciones, índices de paridad y disparidad y distancias notables en la curva de concentración, todo ello dentro del marco de aplicación de la distribución personal de la renta española en los años de la planificación económica.

Keywords: TRUNCATED DISTRIBUTION; MEASURE OF INEQUALITY AND CONCENTRATION CURVE.

1. INTRODUCCION.

La relativamente abundante pero dispersa literatura sobre las medidas de desigualdad, - en especial de concentración, ofrece gran variedad de indicadores formalmente válidos pero que no siempre se inspiran en las mismas premisas y que incluso pueden llegar a proporcionarnos interpretaciones dispares. Se impone pues una selección en base a las peculiaridades propias de cada medida, y es por lo que nos centraremos en una gama de indicadores cuya interpretación,tanto estadística como económica, viene favorecida por el empleo que hacen de unos mismos instrumentos.

A continuación reseñaremos índices o coeficientes de desigualdad que pueden expresarse como operaciones entre la media de la variable en su distribución original y la media - resultante en esta misma distribución censurada, persiguiendo el doble objetivo de amalgamar un conjunto de medidas con idénticos ingredientes y que usualmente quedarían incluidas en diferentes epígrafes, a la vez que reseñar las estimaciones empíricas de que han sido objeto para un caso concreto: la distribución de los ingresos en la España de los años 60 y 70.

2. LAS DISTRIBUCIONES INICIAL Y TRUNCADA.

Vamos a referirnos a la variable aleatoria ξ: "nivel de renta", que consideramos continua, al margen de que los datos muestrales aconsejen un tratamiento discreto; asi mismo consideramos un dominio no negativo para la variable ya que razonablemente las cifras de ingreso son positivas o nulas.

Con este recorrido cualquier función teórica de probabilidad será nula cuando la variable tome valores negativos

$$f_\xi(x) = F_\xi(x) = q_\xi(x) = 0 \qquad \text{si } x < 0$$

y no necesariamente nula en otro caso

$$\exists x \geq 0 \quad f_\xi(x) \neq 0, \quad F_\xi(x) \neq 0 \text{ y } q_\xi(x) \neq 0$$

Con funciones de distribución y masa acumulada de variable $F_\xi(x)$ y $q_\xi(x)$ respectivamente, donde ésta última admite idéntico tratamiento y propiedades que aquella, ya que como es sabido responde a la función de distribución de una variable η de densidad.

$$f_\eta(x) = \frac{x}{m} f_\xi(x) \qquad \text{con } m = E(\xi)$$

y logicamente

- Joan Baró Llinas i Vicenç Moratal Oliver - Escola Universitaria d'Estudis Empresarials - Av. Diagonal, 696 Barcelona 08034.

$$q_\xi(x) = F_\eta(x)$$

cuestión que se confirma al ser

$$f_\eta(x) \geq 0, \quad y$$

$$\int_0^\infty f_\eta(x)\, dx = 1$$

ya que con 1/m factor común, la integral que resultase es la propia esperanza matemática de ξ.

Nótese asimismo como en el punto m ambas funciones presentan la misma pendiente y que para

$x < m$ pendiente de q < pendiente de F

$x > m$ pendiente de q > pendiente de F

Conviene también destacar que siempre la función de distribución será superior o igual a la función acumuladora de la variable

$$\forall x \quad F_\xi(x) \geq q_\xi(x)$$

Efectivamente

$$m\,[F_\xi(x) - q_\xi(x)] = \int_0^x F_\xi(x)\,dx - (x-m)F_\xi(x)$$

Si $x < m$ la justificación es inmediata puesto que aquella diferencia por la esperanza matemática es la suma de dos factores positivos; de razonamiento trivial sería $x = m$, y en el caso de que $x > m$ entonces queda asimismo confirmada aquella relación puesto que siempre

$$\int_0^{x>m} F_\xi(x)\, dx \geq (x - m)F_\xi(x)$$

al tener en cuenta la propiedad en que interviene $F_\xi(x)$ para justificar que la media es el centro de gravedad.

La diferencia $F_\xi(x) - q_\xi(x)$ presenta su máximo valor en el punto de abcisa $x = m$, siendo esta desviación tanto más pequeña cuanto más se aleja la variable de su media, así

$$\frac{d(F - q)}{dx} \gtrless 0 \qquad \text{se tiene } x \lesseqgtr m$$

con incrementos de x, para $x < m$ la diferencia aumenta y para $x > m$ la diferencia disminuye hasta conseguir en los casos extremos igualdad entre $F_\xi(x)$ y $q_\xi(x)$.

En relación a la variable económica que nos ocupa, ξ = "nivel de renta", tendremos:

$F_\xi(x)$ = Acumulación porcentual de elementos (individuos, familias,...)

$q_\xi(x)$ = Acumulación porcentual de montante de renta.

Las relaciones existentes entre $q_\xi(x)$ y $F_\xi(x)$ o sus complementos no sólo determinan la función de concentración como es sabido, sino que además nos permiten efectuar interpretaciones concretas de la desigualdad existente.

Para ello plateémonos una distribución truncada de la variable, con niveles de renta acotados entre dos valores cualesquiera x_1 y x_2, recorrido mucho más restrictivo que el que venimos empleando. Es decir, de una variable inicial $\xi \in R^+$, supongamos que prefijamos dos nuevos extremos para los que $x_1 < \xi_T < x_2$; esto es, la consideración de aquella variable truncada en x_1 y x_2.

$$\xi_T = \frac{\xi}{x_1 < \xi \leq x_2}$$

A la que, desde las leyes iniciales, le correspondería una función de distribución

$$F_{\xi_T}(x) = P(\xi_T \leq x) = P\left(\frac{\xi \leq x}{x_1 < \xi \leq x_2}\right) = \frac{F_\xi(x) - F_\xi(x_1)}{F_\xi(x_2) - F_\xi(x_1)}$$

Si $x_1 < x \leq x_2$

siendo $F_{\xi_T}(x) = 0$ para $x \leq x_1$ y $F_{\xi_T}(x) = 1$ con $x > x_2$

y por tanto, densidad

$$f_{\xi_T}(x) \begin{cases} \dfrac{f_\xi(x)}{F_\xi(x_2) - F_\xi(x_1)} & \text{si } x_1 < x \leq x_2 \\ 0 & \text{en otro caso} \end{cases}$$

y acumulación de masa relativa

$$q_{\xi_T}(x) = \int_{x_1}^x \frac{x}{m_T} f_{\xi_T}(x)\,dx = \frac{q_\xi(x) - q_\xi(x_1)}{q_\xi(x_2) - q_\xi(x_1)}$$

si $x_1 < x \leq x_2$

siendo $q_{\xi_T}(x) = 0$, para $x \leq x_1$ y

$q_{\xi_T}(x) = 1$ con $x > x_2$

con media para la distribución truncada

$$m_T = E\left[\frac{\xi}{x_1 < \xi \leq x_2}\right] =$$

$$= \int_{x_1}^{x_2} x f_{\xi_T}(x)\,dx = m.\frac{q_\xi(x_2) - q_\xi(x_1)}{F_\xi(x_2) - F_\xi(x_1)}$$

siendo m la esperanza matemática de la variable en su distribución inicial.

Cabe como interpretación de la relación m_T/m la misma que se establece entre las acumulaciones de masas de renta y población.

Sin acudir al cociente, tanto $q_\xi(x_2) - q_\xi(x_1)$ como $F_\xi(x_2) - F_\xi(x_1)$ constituyen por sí mismos indicadores de fácil manejo y cómoda lectura; en este sentido y para analizar la estructura personal de los ingresos en el decenio 1.964-1.974 Julio y Angel Alcaide /1/ han empleado esta última diferencia, aportando información del porcentaje de hogares con renta imponible inferior al 50% de la media: $F_\xi(m/2)$, entre el 50% y el 100%: $F_\xi(m) - F_\xi(m/2)$ y con más de 200%: $1 - F_\xi(2m)$, por su parte Ramón Tamames /2/ en su manual de economía española, el único comentario estadístico dedicado a la distribución personal de los ingresos es para calcular la proporción de la población que percibe ingresos por debajo de la renta per cápita: $F_\xi(m)$.

Otras diferencias podrían establecerse en puntos notables de la distribución, tales como la Mediana o Mediala, a las que más adelante nos referiremos, al considerar la curva de concentración.

3. INTERPRETACION DE LAS RELACIONES ENTRE LAS MEDIAS PARA ALGUNAS ACOTACIONES.

A partir de esta generalización podemos considerar casos particulares como las acotaciones por la derecha o por la izquierda. Para ingresos no superiores a uno dado se tendrá

$$E\left[\frac{\xi}{\xi \leq x}\right] = m.\frac{q_\xi(x_2)}{F_\xi(x_2)}$$

Pudiendo interpretar que para un cierto valor de la variable x_2 las veces que la renta per cápita supera a la media de los que no lo alcanzan $E\left[\frac{\xi}{\xi \leq x_2}\right]$ es el exceso multiplicativo de proporción de individuos sobre la acumulación de ingresos en aquel nivel, de forma que cuando

$$F_\xi(x_2) = K\, q_\xi(x_2)$$

es por que

$$m = k\, E\left[\frac{\xi}{\xi \leq x_2}\right]$$

De un modo análogo pudo haberse razonado

$$E\left[\frac{\xi}{\xi \geq x_1}\right] = \frac{1 - q_\xi(x_1)}{1 - F_\xi(x_1)}$$

estableciéndose una misma proporción entre las medias como la que hay entre los complementos de las funciones acumuladas de renta y probabilidad en determinado nivel x_1.

En puntos notables como son los centros ordenados: Mediana M_e y Mediala M_l, puede fácilmente llegarse a

x_1	x_2	m/m_T
0	M_l	$2\,F_\xi(M_l)$
0	M_e	$1/\,2q_\xi(M_e)$
M_l	∞	$2\lvert 1 - F_\xi(M_l)\rvert$
M_e	∞	$1/\lvert 1 - q_\xi(M_e)\rvert$

esto es la proporcionalidad existente entre la media de la variable y la conseguida al truncarla en un solo punto (M_l o M_e), por la derecha o por la izquierda.

También para una doble acotación del cominio de la variable caben expresiones concretas, algunas de las cuales coinciden con casos singulares de medidas frecuentemente utilizadas. Tal es el caso del índice de paridad que se emplea en estratificaciones de grupos de renta (provincias, profesiones, niveles culturales, etc.).

$$\frac{\text{\% de hogares de un estrato}}{\text{\% de renta del estrato}}$$

expresión que es trasladable a cualquier intervalo numérico, por ejemplo por división cuartílica.

$$P = \frac{F_\xi(Q_3) - F_\xi(Q_1)}{q_\xi(Q_3) - q_\xi(Q_1)} = \frac{1}{2[q_\xi(Q_3) - q_\xi(Q_1)]}$$

que cabe considerar como la relación que hay entre frecuencias y masa de la variable en el 50% central de la distribución. Al respecto de las rentas españolas sólo existen algunos cálculos efectuados por Sawyer /3/ para una estratificación por categorías sociales en 1.974.

Conviene notar que tal medida responde al cociente $P = m/m_T$, con $m_T = E(\xi/Q_1<\xi\leq Q_3)$, interpretando pues el índice de paridad como la relación que hay entre las medias, de toda la distribución y de aquella otra que ha prescindido de sus valores extremos o "no representativos", el 25% de los más ricos y el 25% de los más pobres.

Con el nombre de índice de Disparidad designamos a la expresión que emplea para cada posible escalón de renta o en general para cada posible estrato, la relación

$$\frac{\%\ \text{población} - \%\ \text{renta}}{\%\ \text{población} + \%\ \text{renta}}$$

de modo que con los cuartiles primero y tercero nos resultaría

$$D = \frac{(F_\xi(Q_3) - F_\xi(Q_1)) - (q_\xi(Q_3) - q_\xi(Q_1))}{(F_\xi(Q_3) - F_\xi(Q_1)) + (q_\xi(Q_3) - q_\xi(Q_1))} =$$

$$= \frac{1 - 2(q_\xi(Q_3) - q_\xi(Q_1))}{1 + 2(q_\xi(Q_3) - q_\xi(Q_1))}$$

lo que podría trasladarse a cualquier otra partición cuantílica o de fractiles. La Confederación Española de Cajas de Ahorros /4/ presenta aquellos cálculos con un detalle - tanto regional como nacional a partir de las estimaciones llevadas a cabo en 1.966 por DATA en la base de los cálculos realizados por la Comisaría del Plan de Desarrollo y Joan Baró /5/ ha aplicado el coeficiente de los años 1.964-67-70-74 a las rentas españolas.

El indicador presentado para el 50% central de la distribución, puede expresarse en función de m y m_T, donde m_T tiene la misma consideración que le hemos otorgado en el índice de paridad, siendo pues:

$$D = \frac{1 - \frac{m_T}{m}}{1 + \frac{m_T}{m}} = \frac{m - m_T}{m + m_T}$$

pudiendo establecer las relaciones

$$D = \frac{P-1}{P+1} \quad \text{y} \quad P = \frac{1+D}{1-D}$$

que admiten interpretaciones extremos en

$P = 1$ y $D = 0$ equidad

$P = \infty$ y $D = 1$ máxima concentración

4. LA CURVA DE CONCENTRACION.

Por su sencillez y cómoda lectura la curva de Lorentz se ha convertido en un instrumento básico para el estudio de la concentración.

Las propiedades matemáticas de la función

$$q_\xi(x) = \Phi[F_\xi(x)]$$

no pugnan con las características de la distribución de las rentas; es decir:

a) La proporción de la renta y de los individuos que la disfrutan no puede ser negativa ni exceder del 100% (dominio entre 0 y 1).

b) A cada proporción de individuos le corresponde un porcentaje de renta, como continuación a la agregación existente un infinitésimo antes. (continuidad).

c) En una ordenación creciente de ingresos, a mayor porcentaje de población le corresponderá un no inferior porcentaje de individuos. (monotonia creciente).

d) Sucesivos incrementos porcentuales de población conllevan aumentos relativos de - grado inferior en la acumulación de la renta, toda vez que la variable se ordenó de un modo creciente. (convexidad).

En este marco general resulta de interés establecer la relación que hay entre $F_{\xi_T}(x)$ y $q_{\xi_T}(x)$ a la vista de la función original ϕ.

Así sustituyendo las variantes, según los cambios encontrados tendremos:

$$[q_\xi(x_2) - q_\xi(x_1) \; | q_{\xi_T}(x) + q_\xi(x_1) =$$

$$= \Phi[(F_\xi(x_2) - F_\xi(x_1))F_{\xi_T}(x) + F_\xi(x_1)]$$

y despejando q_{ξ_T}, será

$$q_{\xi_T}(x) =$$

$$= \frac{\Phi\left[(F_\xi(x_2) - F_\xi(x_1)) \cdot F_{\xi_T}(x) + F_\xi(x_1)\right] - q_\xi(x_1)}{q_\xi(x_2) - q_\xi(x_1)}$$

lo que en definitiva ha de derivar en

$$q_{\xi_T}(x) = \Phi^*[F_{\xi_T}(x)]$$

dependiendo Φ^* de la naturaleza de Φ.

Con independencia de la curva de concentración correspondiente a la distribución truncada, algunas medidas de desigualdad de la distribución original, expresadas como distancias en el diagrama de Lorentz, pueden presentarse como operaciones entre m y m_T.

Tal es el caso de la distancia paralela a la abcisa, que separa la diagonal de la curva, desde el punto central:

$$F_\xi(M_1) - q_\xi(M_1) =$$

$$= \frac{1}{2}\left[\frac{m}{m_T} - 1\right] \; ; \text{ con } m_T = E\left[\frac{\xi}{\xi \leq M_1}\right]$$

mientras que la distancia, desde el mismo centro, perpendicular a la abcisa será

$$F_\xi(M_e) - q_\xi(M_e) =$$

$$= \frac{1}{2}\left[1 - \frac{m_T}{m}\right] \; ; \text{ con } m_T = E\left[\frac{\xi}{\xi \leq M_e}\right]$$

El doble de esta expresión, $1 - 2\,q_\xi(M_e)$, ha sido utilizado por Joan Baró / / como coeficiente de concentración de las rentas en el decenio 1.964-74, en un estudio sobre la evolución de la desigualdad durante los años del "desarrollo" español.

Otros indicadores podrían construirse con distancias en otros puntos notables de la curva, algunos muy relacionados con otras expresiones ya encontradas, en todo caso, cabe idéntica conclusión que en el resto de coeficientes: cuanta menor desproporción haya entre m_T y m tanta más equidad presentará el reparto.

5. REFERENCIAS.

/1/ ALCAIDE, J., y ALCAIDE, A.:"Distribución Personal de la Renta en España y en los paises de la OCDE", Rev. Hacienda Pública Española, nº 47, 1.977.

/2/ TAMAMES, R.: "Estructura Económica de España", Guadiana Publicaciones, 1.977.

/3/ SAWYER, M. : "La repartition des revenues dans les pays de l'OCDE". Perspectives Economiques de l'OCDE. Etudes speciales, 1.976.

/4/ CONFEDERACION ESPAÑOLA DE CAJAS DE AHORROS: "Comportamiento y actitudes de las economías domésticas hacia el ahorro y el consumo". C.E.C.A. 1968.

/5/ BARO, J.: "El uso de los cuartiles en la medición de la desigualdad de la renta". Cuadernos de Economía, vol. II, nº 30. 1.983.