MEDIDAS DE LOCALIZACIÓN Y ASIMETRÍA BASADAS EN EL ÍNDICE DE GINI

ROMERO GARCÍA, JOSÉ ENRIQUE
Departamento de Economía Aplicada I
Universidad de Sevilla
correo-e: romerogje@us.es

GAMERO ROJAS, JAVIER
Departamento de Economía Aplicada I
Universidad de Sevilla
correo-e: jgam@jet.es

BASULTO SANTOS, JESÚS
Departamento de Economía Aplicada I
Universidad de Sevilla
correo-e: basulto@us.es

RESUMEN

Se estudia la relación entre la “*G-media*”, Berrebi y Silber (1987), y otras “medias” habituales, atendiendo a la estructura de su ponderación.

Como consecuencia, se relaciona la medida de asimetría construida a partir de la *G-media*, también introducida por Berrebi y Silber, con los coeficientes de asimetría de Pearson. Se generaliza a otros coeficientes de asimetría posibles.

Palabras clave: Localización, dispersión, asimetría, *G-media*, índice de Gini.

1. Introducción

Comenzamos exponiendo notaciones y resultados básicos y establecemos una expresión del índice geométrico de Gini como una media ponderada de desviaciones absolutas respecto a la mediana. A partir de la G-media introducida por Berrebi y Silber (1987), definimos la *G-media complementaria* y analizamos las propiedades de robustez de ambas.

Haciendo un paralelismo entre los cuartiles y la definición de la G-media, introducimos la noción de *G-cuartiles* y demostramos las propiedades de optimalidad que verifican. Construimos medidas de dispersión absoluta y relativa a partir de los G-cuartiles y demostramos las relaciones entre esas medidas y la diferencia media de la distribución.

También construimos un G-coeficiente de desigualdad y estudiamos su relación con el índice geométrico de Gini, según la asimetría de la distribución medida a través de lo que denominamos G-coeficiente de asimetría.

2. Expresiones preliminares

Definiremos algunos términos de importancia para el desarrollo del trabajo.

De una variable estadística observada, anotaremos sus valores como x_i, i = 1,…,N, valores ordenados de menor a mayor, y todos ellos no negativos.

Sea T= $\sum_{j=1}^{k} x_j n_j$ la riqueza total que se está repartiendo

Sea t_i= $x_i n_i$ la riqueza que tienen entre todos los que están en la categoría i (ganan lo mismo)

Sea q_i=t_i/T la proporción de riqueza que tienen los individuos de la categoría i

Sea $T_i = \sum_{j=1}^{i} t_j = \sum_{j=1}^{i} x_j n_j$ la riqueza acumulada entre todos los individuos de la categoría i y las inferiores a ella.

Por tanto, $T_i = \sum_{j=1}^{i} t_j = \sum_{j=1}^{i} x_j n_j = T_{i-1} + x_i n_i = T_{i-1} + t_i$; es decir, $T_i = T_{i-1} + t_i$

Sea $Q_i = T_i/T = \sum_{j=1}^{i} q_j$ la proporción de riqueza acumulada hasta la categoría i.

Igualmente, $Q_i = Q_{i-1} + \frac{x_i n_i}{T} = Q_{i-1} + q_i$

El símbolo P_i representa la frecuencia acumulada relativa hasta la categoría *i*, es decir la proporción de observaciones iguales o menores a x_i.

Seguidamente estableceremos una serie de proposiciones que serán instrumentales para la definición y estudio de la G-media y sus derivaciones.

Proposición 1.

$$P_i - Q_i = \frac{\sum_{k \le i} \left(\overline{x} - x_k\right) f_k}{\overline{x}}$$

Demostración:

Como,

$$Q_i = \frac{\sum_{j=1}^{i} x_j n_j}{\sum_{j=1}^{k} x_j n_j} = \frac{\sum_{j=1}^{i} x_j n_j \Big/ N}{\sum_{j=1}^{k} x_j n_j \Big/ N} = \frac{\overline{x_i}}{\overline{x}}$$

se verifica que

$$P_i - Q_i = \sum_{j \le i} f_j - \frac{\sum_{j \le i} x_j f_j}{\overline{x}} = \frac{\overline{x} \sum_{j \le i} f_j - \sum_{j \le i} x_j f_j}{\overline{x}} = \frac{\sum_{k \le i} \left(\overline{x} - x_k\right) f_k}{\overline{x}}$$

♦

Proposición 2.

$$\sum_{i=1}^{N-1}(P_i - Q_i) = \frac{1}{2\bar{x}N}\left[\sum_{i=1}^{N} x_i\left(2i-N-1\right)\right]$$

Demostración:

Vimos en la proposición 1 que $P_i - Q_i = \dfrac{\sum_{k\le i}\left(\bar{x}-x_k\right)f_k}{\bar{x}}$, de donde se obtiene:

$$\sum_{i=1}^{N-1}(P_i-Q_i) = \sum_{i=1}^{N-1}\frac{\sum_{k\le i}\left(\bar{x}-x_k\right){}^{1}\!/_{N}}{\bar{x}} = \frac{1}{N\bar{x}}\left[\sum_{i=1}^{N-1}\sum_{k\le i}\bar{x} - \sum_{i=1}^{N-1}\sum_{k\le i}x_k\right] =$$

$$= \frac{1}{N\bar{x}}\left[\sum_{i=1}^{N-1}\bar{x}i - \sum_{k=1}^{N-1}\sum_{i,i\ge k}^{N-1}x_k\right] = \frac{1}{\bar{x}}\left[\frac{\bar{x}\left(N-1\right)N}{2}\sum_{k=1}^{N-1}x_k\left(N-1-k+1\right)\right] =$$

$$= \frac{1}{2\bar{x}N}\left[\sum_{i=1}^{N}x_i\left(2i-N-1\right)\right]$$

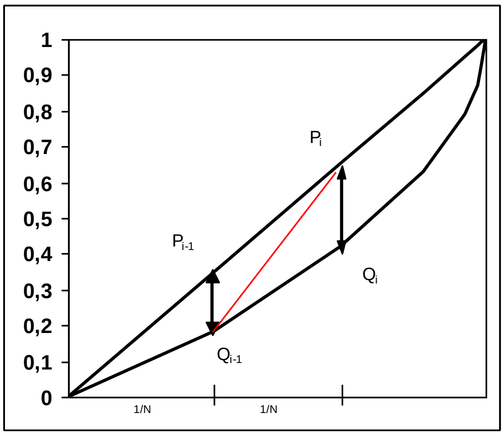
♦

Proposición 3.

$$A_L = \frac{\sum_{i=1}^{N-1}(P_i-Q_i)}{N}$$

Demostración:

En efecto,



obsérvese que cada sector *i* del área de Lorenz se decompone en dos triángulos de áreas A_i y A_i', salvo el primer y el último sector, que tienen cada uno un solo triangulo de área A_1 y A_N' ($A_1' = 0 = A_N$), respectivamente, de manera que

$A_i = \frac{1}{2}(P_i - Q_i)\frac{1}{N}$ (área del triangulo inferior del sector i)

$A_i' = \frac{1}{2}(P_{i-1} - Q_{i-1})\frac{1}{N}$ (área del triangulo superior del sector i)

Por tanto, $A_L = \sum_{i=1}^{N}\left(A_i + A_i'\right) = \sum_{i=1}^{N-1}\left(A_i + A_{i+1}'\right) =$

$$= \sum_{i=1}^{N-1}\frac{P_i - Q_i}{N} = \frac{1}{N}\sum_{i=1}^{N-1}(P_i - Q_i);$$

es decir,

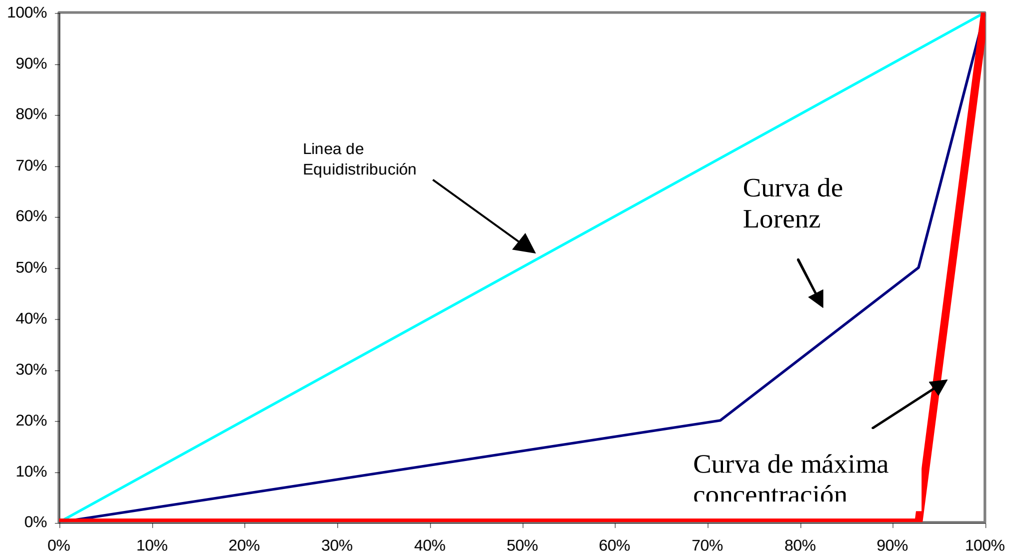
$$A_L = \frac{\sum_{i=1}^{N-1}\left(P_i - Q_i\right)}{N}$$

♦

El índice de Gini, es definido como el cociente entre el área de Lorenz y el área de máxima desigualdad posible con *N* observaciones. Cuando el tamaño *N* es grande, el área de máxima desigualdad tiende a convertirse en el área de todo el triángulo bajo la línea de equidistribución; en tal caso tiene sentido aproximar el índice de Gini mediante el que denominaremos *Índice de Gini Geométrico* (IGG) , definido por:

$$\text{IGG} = \frac{A_L}{\frac{1}{2}} = 2A_L$$

Se sabe que el IGG puede expresarse en función de la diferencia media de la siguiente forma:



$$IGG = \frac{\Delta / 2}{\bar{x}}, \text{ siendo } \Delta = \frac{\sum_i \sum_j | x_i - x_j |}{N^2}$$

Proposición 4.

$$\text{IGG}=\boxed{\frac{\sum_{i=1}^{N}(2i-N-1)x_i}{\bar{x}N^2} = \frac{\sum_{i=1}^{N}\left(\frac{2i-N-1}{N^2}\right)x_i}{\bar{x}} = \frac{\sum_{i=1}^{N} d_i x_i}{\bar{x}}, d_i = \frac{2i-N-1}{N^2}}$$

Demostración:

$$\text{IGG}=2\ A_L$$

Además, ya vimos, en la proposición 3, que

$$A_L = \frac{\sum_{i=1}^{N-1}(P_i - Q_i)}{N}$$

y, en la proposición 2, que

$$\sum_{i=1}^{N-1}(P_i - Q_i) = \frac{1}{2\bar{x}N}\left[\sum_{i=1}^{N} x_i\,(2i-N-1)\right]$$

luego:

$$\text{IGG}=\frac{\sum_{i=1}^{N}(2i-N-1)x_i}{\bar{x}N^2}=\frac{\sum_{i=1}^{N}\left(\frac{2i-N-1}{N^2}\right)x_i}{\bar{x}}=\frac{\sum_{i=1}^{N}d_i x_i}{\bar{x}},\, d_i=\frac{2i-N-1}{N^2}$$

♦

Proposición 5. Sea *N* un número par. Entonces:

$$\text{IGG}=\frac{\sum_{i=1}^{N}c_i\left|x_i-Me\right|}{2\bar{x}} \quad,\ c_i=\frac{\left|2i-N-1\right|}{N^2/2},\ \sum_{i=1}^{N}c_i=1$$

(Una expresión similar ocurriría para el caso impar)

Demostración:

En efecto,

a partir de la expresión IGG= $\frac{\sum_{i=1}^{N}(2i-N-1)x_i}{\bar{x}N^2}$, se prueba que, para el caso N par:

$$\text{IGG}=\frac{\dfrac{\sum_{i=1}^{N}w_i\left|x_i-Me\right|}{\sum_{i=1}^{N}w_i}}{2\bar{x}} \quad,w_i=\left|\frac{2i-N-1}{N^2}\right|,\ \sum_{i=1}^{N}w_i=\frac{1}{2}$$

Y, por tanto, si llamamos

$$c_i=\frac{\left|2i-N-1\right|}{N^2/2},\ \sum_{i=1}^{N}c_i=1$$

se verifica que:

$$\text{IGG}=\frac{\sum_{i=1}^{N}c_i\left|x_i-Me\right|}{2\bar{x}}$$

♦

Definición 1. Definimos el Índice Geométrico Absoluto de Gini como IGGA= $\bar{x}\cdot$IGG,

con lo que, por la propiedad 5, se verifica que

$$\text{IGGA}=\frac{1}{2}\sum_{i=1}^{N}c_i\left|x_i-Me\right|,\ c_i=\frac{\left|2i-N-1\right|}{N^2/2},\ \sum_{i=1}^{N}c_i=1$$

3. La G-media como media ponderada. La $\overline{G}$ -media.

Berrebi y Silber(1987) definen la G-media de la siguiente forma:

Definición 2. Se define G-media de las observaciones $\{x_i\}$ como

$$x_G = \sum c_i x_i \ , \ c_i = \frac{|2i - N - 1|}{N^2 / 2} \ \text{y} \ \sum c_i = 1 .$$

Puede apreciarse que la G-media es una media ponderada de los valores observados x_i, mediante las ponderaciones c_i. Estas ponderaciones tiene la propiedad de ser mayores en las posiciones más alejadas de la mediana y menores en los valores cercanos a esta (siendo cero precisamente en el valor mediano).

Dichas ponderaciones nos indican que x_G será una media que sobrepondera los valores extremos e infrapondera los centrales. Este comportamiento es el opuesto a las medias ponderadas consideradas “robustas”, en las cuales se infraponderan los valores extremos para evitar un exceso de sensibilidad ante valores atípicos o aberrantes.

Berrebi y Silber sugieren que la G-media es una especie de combinación de media aritmética y mediana. En realidad la G-media es, como media ponderada, no un intermedio entre mediana y media aritmética sino que estaría fuera del intervalo, por así decirlo, determinado por ellas. Es decir, tomando la media aritmética como referencia equiponderada, la mediana estaría en una “dirección” (infraponderar los extremos) y la G-media en otra “dirección” opuesta (sobreponderar los extremos). Una media derivada de la G-media y que tendría la propiedad de infraponderar los extremos es lo que denominaremos $\overline{G}$ -media, consistente en una media ponderada con las ponderaciones complementarias a las de la G-media respecto a las de la media. Es decir, si b_i son las ponderaciones de la $\overline{G}$ -media y 1/N son las de la media aritmética, entonces haciendo que se verifique $(b_i + c_i) / 2 = 1/N$, se deduce que $b_i = 2/N - c_i$.

Por consiguiente, podemos establecer la siguiente definición:

Definición 3. Se define la *G-media complementaria* o $\overline{G}$ *-media* de las observaciones $\{x_i\}$ como

$$x_{\overline{G}} = \sum b_i x_i \, , \; b_i = \frac{2}{N} - c_i \; , \; (\; \sum b_i = 1 \;)$$

Puede comprobarse fácilmente que

$$\frac{x_G + x_{\overline{G}}}{2} = \overline{x}$$

El papel de x_G y $x_{\overline{G}}$ como medidas robustas o "antirobustas" es el siguiente: en general una media robusta (es decir, que infrapondere los extremos) es mejor estimador del valor central de una distribución simétrica que la media aritmética si la distribución tiene curtosis elevada (hablando en términos generales) y una media "antirobusta" (que sobrepondera los valores extremos) es mejor que la media aritmética en distribuciones con escasa curtosis (por ejemplo, distribuciones en forma de "U").

4. Otras medidas de localización y dispersión a partir de la G-media

La expresión de la G-media podemos reescribirla como

$$x_G = \sum_{i=1}^{N} c_i x_i = \sum_{i=1}^{N/2} c_i x_i + \sum_{i=N/2+1}^{N} c_i x_i = \frac{1}{2}\left(\sum_{i=1}^{N/2} 2c_i x_i + \sum_{i=N/2+1}^{N} 2c_i x_i \right)$$

Interpretando estos últimos sumatorios, se observa que son, el primero de ellos, una media ponderada de los valores inferiores a la mediana y, el segundo de ellos, una media ponderada de los valores superiores a la mediana. Por tanto hacen una función similar a los cuartiles 1º y 3º (que son la mediana de los valores inferiores a la mediana y la mediana de los valores superiores a la mediana, respectivamente). Por ello los denominaremos *G-cuartiles*.

Definición 4. Definimos el primer G-cuartil, G_1, y el tercer G-cuartil, G_3, como

$$G_1 = \sum_{i=1}^{N/2} e_i x_i \,,\; e_i = 2c_i \,,\; \left(\sum_{i=1}^{N/2} e_i = 1 \right)$$

$$G_3 = \sum_{i=N/2+1}^{N} e_i x_i \,,\; e_i = 2c_i \,,\; \left(\sum_{i=N/2+1}^{N} e_i = 1 \right)$$

Lema 1. Sea $x_w = \sum_{i=1}^{N} w_i x_i$ una media ponderada con $\sum_{i=1}^{N} w_i = 1,\; w_i \geq 0, \forall i$. Entonces x_w es el valor a que minimiza la expresión $\sum_{i=1}^{N} w_i (x_i - a)^2$.

La demostración es similar a la conocida para la media aritmética.

Proposición 6. La G-media y la $\overline{G}$-media minimizan las expresiones cuadráticas $\sum_{i=1}^{N} c_i (x_i - a)^2$ y $\sum_{i=1}^{N} b_i (x_i - a)^2$, respectivamente.

Demostración: Inmediata a partir del lema anterior.

Proposición 7. Los G-cuartiles, G_1 y G_3, minimizan las expresiones cuadráticas $\sum_{i=1}^{N/2} e_i (x_i - a)^2$ y $\sum_{i=N/2+1}^{N} e_i (x_i - a)^2$, respectivamente.

Demostración: Inmediata a partir del lema anterior.

Proposición 8. La G-media es la media de los G-cuartiles

$$x_G = \frac{G_1 + G_3}{2}$$

Demostración: Se deduce de la construcción de los G-cuartiles.

Lema 2.

$$\Delta = \sum_{i=N/2+1}^{N} c_i x_i - \sum_{i=1}^{N/2} c_i x_i$$

Demostración:

Dado que $IGG = \frac{\Delta/2}{\bar{x}}$ y, como por la proposición 5, IGG= $\frac{\sum_{i=1}^{N} c_i \left| x_i - Me \right|}{2\bar{x}}$, se verifica que:

$$\begin{aligned}
\Delta &= \sum_{i=1}^{N} c_i \left| x_i - Me \right| = \\
&= \sum_{i=1}^{N/2} c_i \left(Me - x_i \right) + \sum_{i=N/2+1}^{N} c_i \left(x_i - Me \right) = \\
&= Me \sum_{i=1}^{N/2} c_i - \sum_{i=1}^{N/2} c_i x_i + \sum_{i=N/2+1}^{N} c_i x_i - Me \sum_{i=N/2+1}^{N} c_i = \\
&= Me \left(\sum_{i=1}^{N/2} c_i - \sum_{i=N/2+1}^{N} c_i \right) - \sum_{i=1}^{N/2} c_i x_i + \sum_{i=N/2+1}^{N} c_i x_i = \\
&= -\sum_{i=1}^{N/2} c_i x_i + \sum_{i=N/2+1}^{N} c_i x_i
\end{aligned}$$

♦

Teorema 1. El semirecorrido inter-G-cuartílico coincide con la diferencia media:

$$\frac{G_3 - G_1}{2} = \Delta$$

Demostración:

$$\frac{G_3 - G_1}{2} = \frac{\sum_{i=N/2+1}^{N} e_i x_i - \sum_{i=1}^{N/2} e_i x_i}{2} = \sum_{i=N/2+1}^{N} c_i x_i - \sum_{i=1}^{N/2} c_i x_i$$

Expresión que coincide con Δ según el lema 2.

♦

Con los G-cuartiles podemos construir una medida de dispersión relativa análoga al recorrido intercuartílico relativo.

Teorema 2. El recorrido inter-G-cuartílico relativo es igual a la diferencia media relativa respecto a la G-media.

$$\frac{G_3 - G_1}{G_3 + G_1} = \frac{\Delta}{x_G}$$

Demostración:

Se deduce del Teorema 1 y de la proposición 8.

♦

Definición 5. Sean unos pesos de ponderación w_i que verifican $\sum_{i=1}^{N} w_i = 1,\ w_i \geq 0, \forall i$, definimos la *w-mediana* de las observaciones $\{x_i\}$ como el valor a que minimiza $\sum_{i=1}^{N} w_i |x_i - a|$. Como caso particular, definimos la *G-mediana* cuando $w_i = c_i$.

Un resultado algo menos general al del lema 1 se verifica con desviaciones absolutas.

Lema 3. Sea unos pesos de ponderación w_i que verifican $\sum_{i=1}^{N} w_i = 1$, $w_i \geq 0, \forall i$, siendo simétricos respecto a la posición mediana, $\sum_{i=1}^{N/2} w_i = \sum_{i=N/2+1}^{N} w_i$. Entonces la mediana M_e es el valor *a* que minimiza la expresión $\sum_{i=1}^{N} w_i \left| x_i - a \right|$.

Demostración: Análoga a la demostración de que la mediana minimiza la suma de las desviaciones absolutas.

Proposición 9. La G-mediana coincide con la mediana.

Demostración: A partir de la definición 5 y el lema 3.

De la proposición 6 surgen de forma natural las siguientes definiciones de varianza y coeficiente de variación.

Definición 6. Definimos la *G-varianza, S_G^2*, como

$$S_G^2 = \sum_{i=1}^{N} c_i \left(x_i - x_G \right)^2$$

Definición 7. Definimos el *G-coeficiente de variación,* $CV_G = \dfrac{S_G}{x_G}$

Otra medida de dispersión relativa o de desigualdad surgiría de sustituir la media por la G-media en el índice geométrico de Gini, IGG.

Definición 8. Definimos el *G-coeficiente de desigualdad* como

$$CD_G = \frac{\Delta/2}{x_G}$$

4. Asimetría y desigualdad

Vamos a exponer la relación entre el índice geométrico de Gini (IGG) y el G-coeficiente de desigualdad en función de la asimetría en la distribución de los valores observados.

Definición 9. Se define el *G-coeficiente de asimetría*, A_G , como

$$A_G = \frac{\mu_3 - M_3}{\mu_1 - M_1}$$

siendo

$$M_1 = \frac{\sum_{i=1}^{N/2} x_i}{N/2},\ \mu_1 = \frac{\sum_{i=1}^{N/2} i x_i}{N(N+2)/8}$$

$$M_3 = \frac{\sum_{i=N/2+1}^{N} x_i}{N/2},\ \mu_1 = \frac{\sum_{i=N/2+1}^{N} (i - N/2)\, x_i}{N(N+2)/2}$$

Diremos que la distribución es:

- G-simétrica si A_G=1
- G-asimétrica a la derecha si $A_G > 1$
- G-asimétrica a la izquierda si $A_G < 1$

Teorema 3. Se verifican las siguientes afirmaciones:

- Si la distribución es G-simétrica, CD_G = IGG
- Si la distribución es G-asimétrica a la derecha, CD_G < IGG
- Si la distribución es G-asimétrica a la izquierda, CD_G > IGG

Demostración: Puede comprobarse que se dan las siguientes relaciones:

- $A_G = 1 \Rightarrow x_G = \bar{x}$
- $A_G > 1 \Rightarrow x_G > \bar{x}$
- $A_G < 1 \Rightarrow x_G < \bar{x}$

Y de estas relaciones, junto con las definiciones de IGG y CD_G, se deduce inmediatamente el teorema.

♦

Bibliografía

1. Berrebi, Z. M. Y Silber, J. (1987): “Dispersión, asymmetry and the Gini index of inequality”, *International Economic Review*, **28**, 2, pp.331-3382.

2. Gini, C. (1912), “Variabilità e Mutabilità”, *Studi Economico-Giuridici dell’Univ. Di Cagliari*, **3**, part 2, pp.1-158.

3. Gini, C. (1914), “Sulla misura della concentrazione e della variabilità dei caractteri”, *Atti del R. Istituto Veneto di Scienze, Lettere ed Arti*, Tomo **LXXIII**, pp. 1203-1248.

4. Gini, C. (1935), *Curso de Estadística.* Editorial Labor. Barcelona