# UNA FAMILIA UNIPARAMÉTRICA DE MEDIDAS DE LOCALIZACIÓN CENTRAL A PARTIR DEL ÍNDICE DE GINI

**José Enrique Romero García**
Departamento de Economía Aplicada I
Universidad de Sevilla
e-mail: romerogje@us.es

**Javier Gamero Rojas**
Departamento de Economía Aplicada I
Universidad de Sevilla
e-mail: jgam@us.es

**Jesús Basalto Santos**
Departamento de Economía Aplicada I
Universidad de Sevilla
e-mail: basulto@us.es

**Resumen**

A partir de la representación del índice de concentración de Gini como combinación ponderada de los datos, pueden elaborarse varias medidas de localización que presentan diversos niveles de robustez al estimar el valor central de una distribución.

Estas medidas pueden organizarse en una familia uniparamétrica de las mismas.

Se presentan varios ejemplos de aplicación de esta familia en función de ciertas características distribucionales.

*Palabras clave*: Gini, localización, robustez, G-media.

*Area temática*: Métodos cuanitativos.

## 1. Introducción

Comenzamos estableciendo una expresión del índice geométrico de Gini como una media ponderada de desviaciones absolutas respecto a la mediana.

Consideraremos unos valores $x_1,\ldots,x_N$, ordenados de menor a mayor, posiblemente con elementos repetidos (y por tanto, considerado cada $x_i$ con frecuencia unitaria). Llamaremos $F_i$ a la frecuencia acumulada relativa de $x_i$ para i=1,…,N. Denominaremos $Q_i$ a la cantidad acumulada relativa i-ésima:

$$Qi = \frac{\sum_{j=1}^{i} x_j}{\sum_{j=1}^{N} x_j}$$

Representaremos de la forma habitual las medias y las medianas de las variables involucradas.

En el contexto habitual de los estudios de desigualdad/concentración, los valores $x_i$ son ingresos o rentas o quizás riqueza de cada miembro de una población humana. El análisis que planteamos es, sin embargo, generalizable a cualquier variable no negativa.

**Definición.** Se define el índice geométrico de Gini (IGG) como el doble de área de Lorenz ($A_L$)

$$IGG = 2 \cdot A_L$$

**Proposición.**

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

*Demostración:*

$$IGG=2\ A_L$$

Se sabe que

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

e, igualmente, 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.** Sea *N* un número par. Entonces:

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

*Demostración*

En efecto,

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

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

Y, por tanto, si llamamos

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

se verifica que:

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

♦

Obsérvese, que cada peso $w_i = \left|\frac{2i - N - 1}{N}\right|$ nos mide la posición relativa del individuo i-ésimo, pues es el valor absoluto de la proporción de individuos en peor situación menos la proporción de individuos en mejor situación.

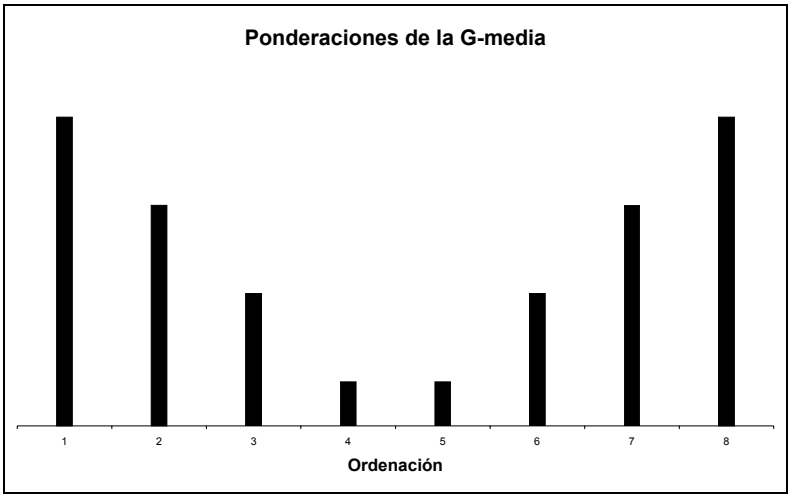
**2. La G-media y la $\overline{G}$ -media.**

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

**Definición.** 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 tienen 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).



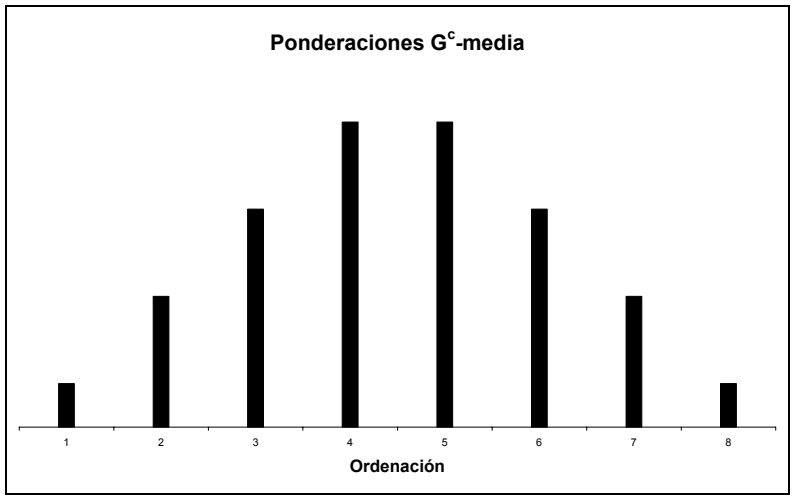

Dichas ponderaciones nos indican que $x_G$ será una media que sobrepondera los valores extremos e infrapondera los centrales. Este comportamiento es el opuesto a las medias ponderadas consideradas "robustas", en las cuales se infraponderan los valores extremos para evitar un exceso de sensibilidad ante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

Hemos introducido una media con pesos complementarios a los de la G-media, de forma que se sobreponderen los valores centrales y se infraponderen los valores extremos. Tal media tendrá características de robustez ante posibles valores anómalos y la denominaremos *G-media complementaria*.

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

$$x_{\bar{G}} = \sum b_i x_i \,,\; b_i = \frac{2}{N} - c_i = \frac{2}{N} - \frac{\left|2i - N - 1\right|}{N^2/2} \text{ y } \sum b_i = 1 \,.$$




### 3. Familia uniparamétrica λ-media.

Definamos una familia paramétrica de estadisticos de localización de tal forma que la G-media, la media, la mediana, el punto medio y la $G^c$-media sean casos particulares de ella, y el valor del parámetro indique en qué grado se sobrepondera/infrapondera las zonas centrales y extremas de la distribución de valores observados.

**Definición.** Sea v1 el vector de ponderaciones de la G-media, v2 el vector de ponderaciones de la $G^c$-media, y sea $\lambda$ un valor real. Se define la $\lambda$-media como la media ponderada cuyo vector de ponderaciones es:

$\max(0 , v1(2 + \lambda - |\lambda|) + v2(2 + \lambda + |\lambda|)) = \max(0 , (2 + \lambda)(v1 + v2) + |\lambda|(v2 - v1))$, $\lambda \in \mathbf{R}$

En particular, se obtienen las siguientes equivalencias:

$\lambda = 0 \rightarrow$ media aritmética

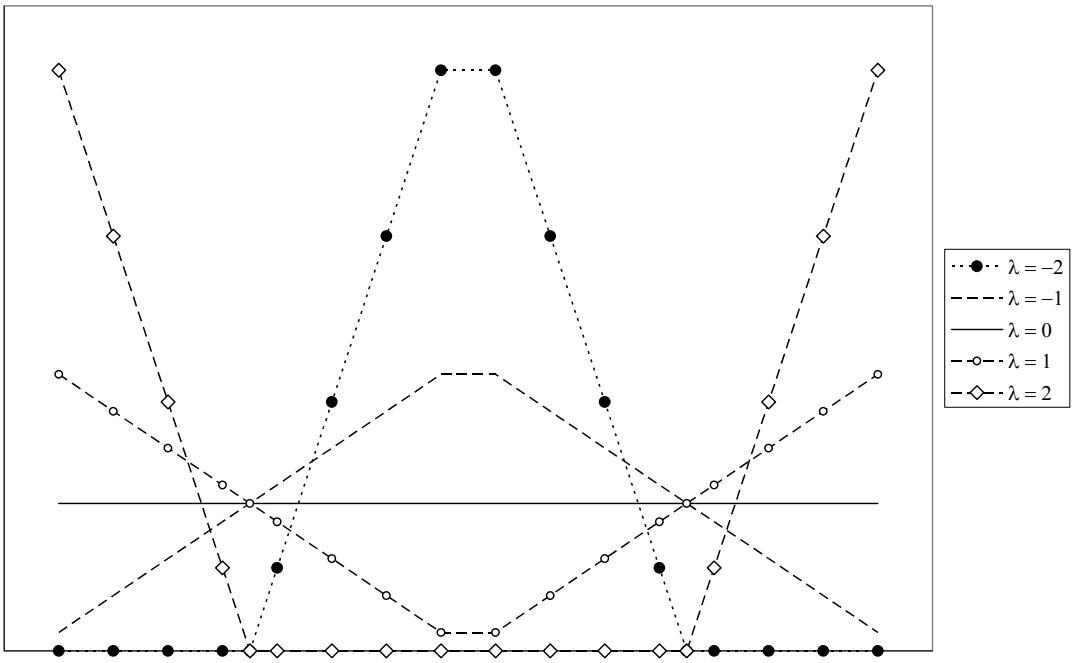
$\lambda = 1 \rightarrow$ G-media

$\lambda = -1 \rightarrow$ Gc-media

$\lambda = \infty \rightarrow$ punto medio

$\lambda = -\infty \rightarrow$ mediana

Un gráfico representando los esquemas de ponderación según varios valores de $\lambda$ es el siguiente:




## 4. Simulaciones con una familia de modelos generadores

### 4.1 Modelos generadores

Para medir la bondad de los diferentes estimadores $\lambda$-media en diferentes escenarios según modelos generadores de diferentes curtosis.

**Definición.** Sea $Z = N(0,1)$, definimos el modelo generador Hr a la variable

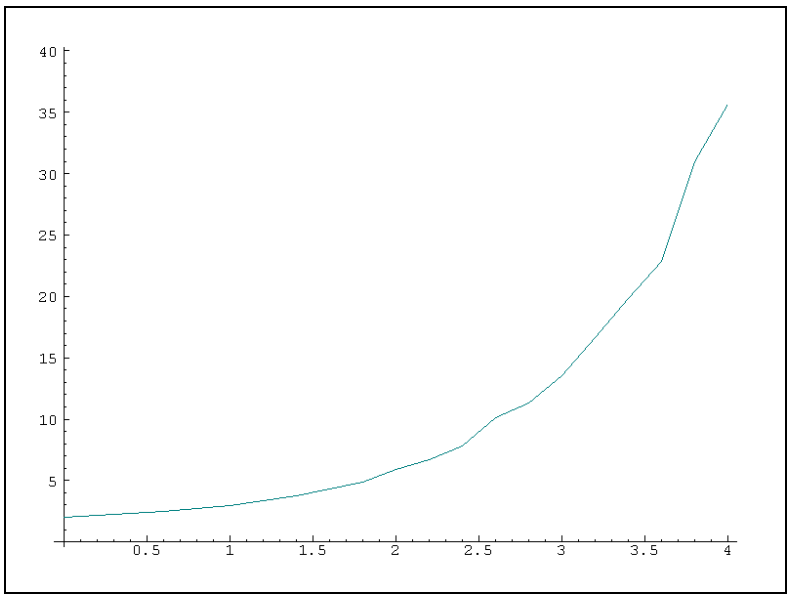
$$Y = \text{sign}(Z)\cdot[(1 + |Z|)^{r} - 1] / r$$

En donde $r \in (0 , \infty)$.

Las características más notables de los modelos de esta familia son:

- Son variables simétricas, con punto de simetría situado en 0.
- A mayor valor del parámetro "r", mayor curtosis. Cuando r = 1, Y = Z y la curtosis es, obviamente normal.
- El espacio total es toda la recta real.

Para cada valor del parámetro r (r = 0, 0.2, 0.4,….4), hemos simulado una muestra de 100.000 elementos, calculando su curtosis, tal como aparece en el gráfico que se muestra a continuación:




Hemos comprobado que existe una fácil relación analítica aproximada entre el parámetro r y la curtosis del modelo generador Hr:

$$\mathrm{Ln}(g_2 - 1) \cong -0.160 + 0.901\ r,\ R^2 = 0{,}9948,$$

donde $g_2$ es el coeficiente de curtosis de Fisher, y $R^2$ el coeficiente de bondad del ajuste.

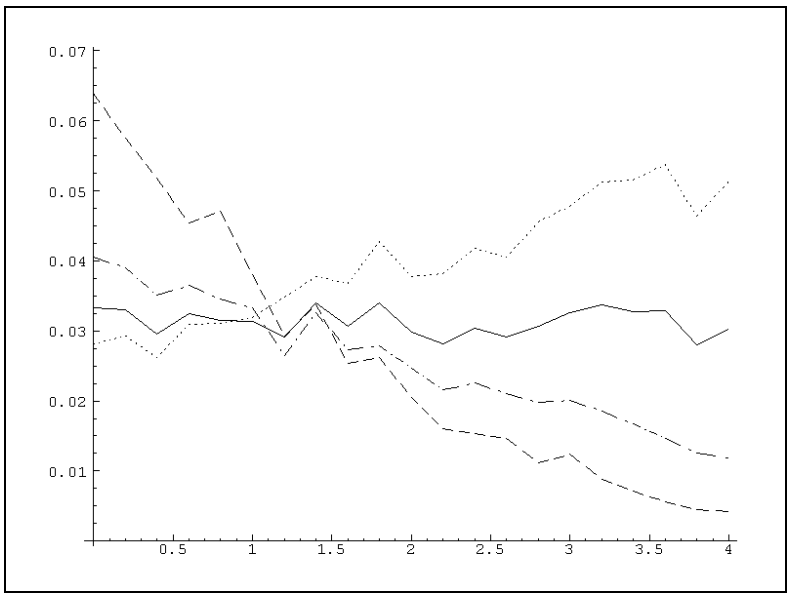
Con lo que obtenemos la siguiente expresión:

$$g_2 \cong 1+ 0.852 \cdot 2.463^{r}$$

**4.2 Simulación**

Para cada valor de r (r = 0, 0.2, 0.4,….4) se han generado mil muestras de tamaño 100, a cada una se le ha calculado el estimador del valor central, en el modelo generador es cero, para diferentes valores del parámetro $\lambda$ ($-\infty$, $-1$,0,1) y, luego, se ha calculado la desviación típica de los 1000 estimadores. Esto nos dará, a su vez, una estimación de la desviación típica del estimador.

En el siguiente gráfico representamos, para cada valor de $\lambda$, las desviaciones típicas del estimador $\lambda$-media correspondiente a los diferentes valores del parámetro r de la familia generadora.




*— media, - - - mediana, …… G-media, -.-.- $G^c$-media*

Se observa como para valores pequeños del parámetro r, curtosis pequeña, los estimadores G-media y media, mejoran a los estimadores $G^c$-media y mediana, pues las desviaciones al parámetro son menores en los primeros casos que en los

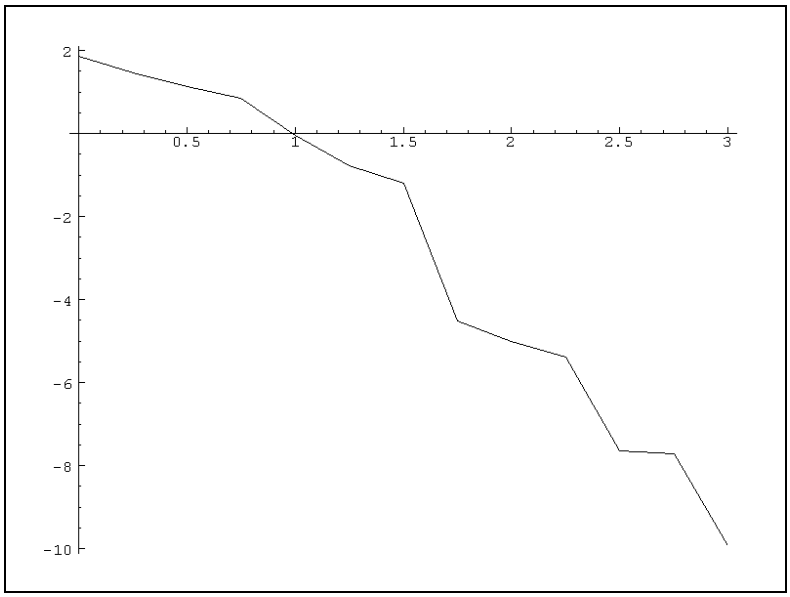
segundos. Para valores grandes del parámetro r la ordenación de la bondad de los estimadores queda invertida.

Esto pone de manifiesto que, en curtosis elevada es favorable para la estimación del parámetro central la infraponderación de los valores extremos, mientras que en curtosis pequeña resulta más adecuado la sobreponderación de dichos valores extremos.

La idea que se obtiene del gráfico precedente es que, para cada valor de r, y por tanto para cada valor de curtosis, exista una λ-media óptima de la familia. Para constatar este hecho, hemos calculado, para una serie de valores de r, que λ-media tiene menor desviación típica respecto al parámetro, usando en cada caso mil muestras de tamaño 100.

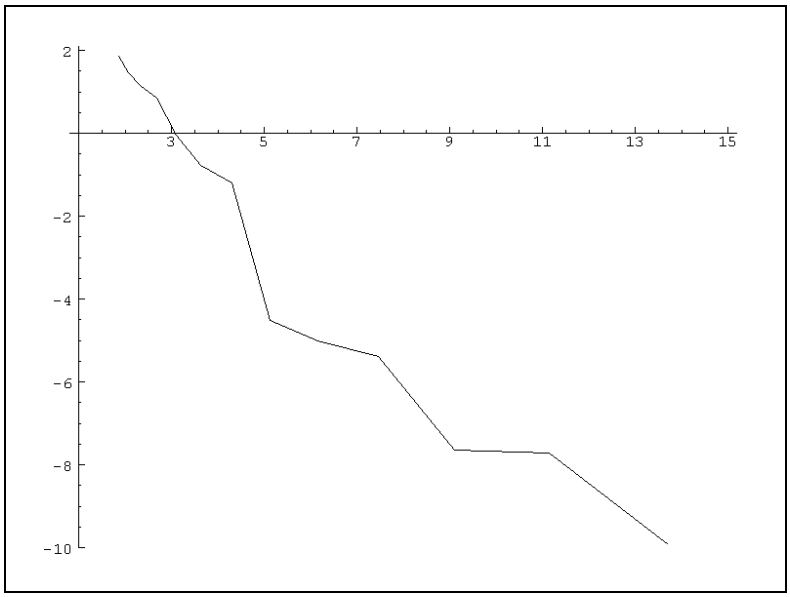
Para hallar la λ óptima, calculamos las desviaciones típicas de valores λ = –20, –19.75, –19.5,…,3.75, 4. El óptimo lo hemos calculado mediante una aproximación parabólica según el mejor valor observado, el anterior y el posterior.

En el gráfico representamos el λ óptimo calculado para cada valor del parámetro "r".




*λ óptimo según r*

En el siguiente gráfico aparecen esos óptimos en función de la curtosis implicada por cada valor "r" según la expresión analítica aproximada expuesta anteriormente.




*λ óptimo según curtosis*

## 5. Conclusiones y futuras ampliaciones

Una media ponderada es un estimador adecuado en función de la relación de ponderación entre los valores centrales y los valores extremos según la curtosis del modelo poblacional de los datos. En concreto, para la familia Hr, simétrica, comprobamos que dentro de la familia de las λ-medias, hay unas ciertas ponderaciones que resultan óptimas para una curtosis dada, favoreciendose la infraponderación de valores extremos a medida que la leptocurtosis aumenta y, recíprocamente, al disminuir la platicurtosis, el óptimo se obtiene al sobreponderar los extremos.

En el presente estudio se ha trabajado con muestras de tamaño 100. Una ampliación sería comprobar los resultados en muestras de menor y de mayor tamaño.

Se puede, igualmente, estudiar la generalización de estos resultados para otras familias generadoras, a fin de poder establecer una relación más definitiva entre curtosis y ponderaciones.

## 6. 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

5. Romero, J.E., Gamero, J., Basulto, J. "Medidas de localización y asimetría basadas en el índice de Gini". *ASEPELT-2004*.