Estimación del coeficiente de concentración de Gini a partir de la curva estimada de Lorenz

Rolly Buccioni Vadulli

Departamento de Estadística y Econometría, FAE, Universidad Tecnológica Metropolitana, Santiago de Chile. E-mail: rolly.buccioni@utem.cl.

Resumen

El objetivo del trabajo es generar el índice de Gini a partir de una estimación de la curva de Lorenz, que se basa en una función exponencial continua con un parámetro α por determinar, y así realizar comparaciones con su cálculo tradicional.

Abstract

The aim of this work is to generate the Gini index from an estimate of the Lorenz curve, which is based on a continuous exponential function with a parameter α to be determined, and by this way to make comparisons with the traditional calculation.

Clasificación JEL: C02, C18, I39.

El Coeficiente de Concentración de Gini creado por el estadístico italiano Corrado Gini (1884-1965), quien en su obra "Variabilità e mutabilità" publicada en 1912, desarrolló un método para medir la desigualdad de una distribución. En ella introdujo el valor de 0 para expresar la igualdad total y el valor de 1 para la máxima desigualdad. Este método se aplica en el estudio de la distribución de desigualdad en ciencias de la salud, ingeniería, ecología, química, transporte, etc. Pero quizá donde tiene su uso más característico es en el estudio de la desigualdad de los ingresos que se realiza en Economía. También se aplica a la descripción de unidades económicas según el tamaño (empresas, ventas, producción, etc.), pero en general se puede aplicar para medir el nivel de concentración o desigualdad para cualquier tipo de distribución.

La metodología propuesta por Gini considera la obtención de porcentajes acumulados de totales y ser comparados, mediante cuociente, con porcentaje acumulado de observaciones. Esta información genera una curva denominada "Curva de Concentración" o "Curva de Lorenz". El objetivo del trabajo pretende encontrar dicho índice realizando una estimación de la curva de Lorenz definiendo una función matemática, que depende de un parámetro α, el que es estimado a partir de la información. La función definida aproxima dicha curva para luego realizar las comparaciones y evidenciar empíricamente el grado cercanía entre el resultado real de la curva y el obtenido a partir de la estimación mediante la diferencia cuadrática de los residuos.

Introducción:

Las medidas o índices de concentración tienen como objetivo fundamental cuantificar el grado de desigualdad en el reparto o distribución de una magnitud económica las que pueden ser ingresos, rentas, beneficios, etc. entre un número determinado de "unidades" tales como individuos, familias, empresas, etc.

La forma de la distribución de una magnitud, representada por una variable estadística, ya se ha estudiado a través de diversas medidas de posición, dispersión, asimetría y apuntamiento. Lo que ahora nos interesa es la mayor o menor equidad en el reparto del recurso, expresada como la suma total observada entre los integrantes del conjunto que conforman las unidades. Para ello, deberemos recoger de cada elemento perteneciente al conjunto, la información de la cuantía individual recibida en el reparto. La dificultad reside en que, en muchas ocasiones, esa información viene agrupada en clases y, por tanto, el estudio de la concentración no se podrá hacer con la precisión debida.

Es evidente que las dos situaciones extremas que podemos considerar, respecto a la equidad en el reparto, son:

- Mínima concentración o máxima igualdad: Esto ocurre cuando a todos los integrantes del conjunto se les asigna la misma cantidad en el reparto del monto total.

- Máxima concentración o mínima igualdad: Esto se da cuando un único sujeto recibe la suma total a repartir y los demás no perciben parte alguna del recurso.

Estas dos situaciones deberán estar claramente identificadas por las medidas de concentración que se definan y que, asimismo, deberán graduar las situaciones intermedias, entre las que se encontrará la mayoría de los casos estudiados en la práctica.

Indice de Gini

Gini propuso su índice mediante la metodología que a continuación se describe: Los datos obtenidos deben ser ordenados previamente de menor a mayor cuantía. Si el número de sujetos es N y representamos por Vi el valor que le corresponde al i-ésimo sujeto, la ordenación queda representada como se muestra a continuación:

$$V_1 \leq V_2 \leq V_3 \leq \ldots \leq V_i \leq \ldots \leq V_N$$

Luego se ordenan las cantidades acumuladas del número de sujetos y de la cantidad del recurso que recibe cada sujeto. Con dichas cantidades se obtienen finalmente tanto las proporciones acumuladas de sujetos como las proporciones acumuladas del recurso.

Todo lo anterior se puede exponer en el siguiente cuadro:

Sujetos	Valor obtenido (V)	Nº acumulado de sujetos (i)	Valor acumulado del Recurso (U)	$p_i = \frac{i}{N} \cdot 100$	$q_i = \frac{U_i}{U_N} \cdot 100$
1º	V_1	1	$U_1 = V_1$	p_1	q_1
2º	V_2	2	$U_2 = V_1 + V_2$	p_2	q_2
…	…	…	…	…	…
Nº	V_N	N	$U_N = \sum_{j=1}^{N} V_j$	p_N	q_N

Las dos últimas columnas entregan la información de cómo se ha distribuido el recurso entre las unidades en estudio. Si el reparto fuese equitativo, es decir, todos recibiendo la misma cantidad del recurso tendríamos que $p_i = q_i \quad \forall\, i = 1, 2, \ldots, N-1$.
En el caso que un solo sujeto reciba la totalidad del recurso, lo que implica la máxima concentración, se tiene que $q_i = 0 \quad \forall\, i = 1, 2, \ldots, N-1$. En cualquier otro caso se está en presencia de una situación intermedia.
Cuanto mayor sea la diferencia (pi - qi) mayor será la concentración producida en el reparto, pudiendo, así, identificar en qué sectores o grupos del conjunto de sujetos se da mayor concentración, por el mayor valor de esas diferencias.
El índice de concentración de Gini se define como:

$$I_G = \frac{\sum_{i=1}^{N-1}(p_i - q_i)}{\sum_{i=1}^{N-1} p_i} = 1 - \frac{\sum_{i=1}^{N-1} q_i}{\sum_{i=1}^{N-1} p_i}$$

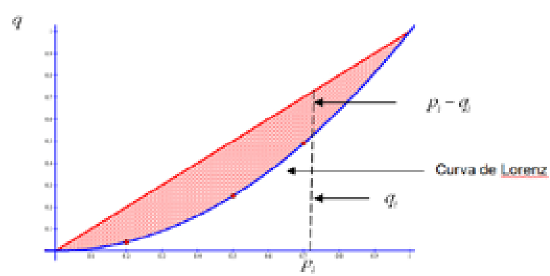
Si estamos en presencia de la máxima equidad en el reparto del recurso, ocurriría que $I_G = 0$
Si estamos en presencia de la mínima equidad en el reparto del recurso, ocurriría que $I_G = 1$

$$\therefore\ 0 \le I_G \le 1$$

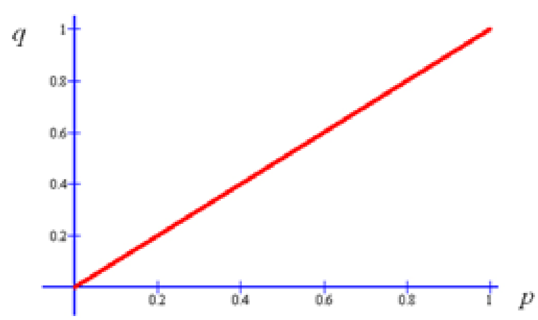
Curva de Lorenz

La curva de Lorenz o curva de concentración es una gráfica que se deduce a partir de la información suministrada para el cálculo del índice de Gini y que, por tanto, refleja la mayor o menor concentración en la distribución de una magnitud. Como expondremos a continuación, existe una relación directa entre el índice de Gini y la forma de la curva de Lorenz, suponiendo esta última una información adicional muy interesante sobre la forma en que se ha llevado a cabo el reparto de la cuantía total.

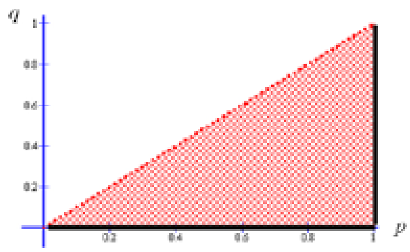
Para su representación gráfica utilizamos un sistema de ejes coordenados donde ubicamos en el eje de las abscisas, la proporción acumulada de sujetos, mientras que en el eje de las ordenadas está la proporción acumulada del recurso, de modo que un punto (p; q) representará la proporción acumulada de sujetos (p), los cuales reciben un porcentaje q del recurso.



En caso de mínima concentración, la gráfica de la curva de Lorenz concuerda con la diagonal, como lo muestra la siguiente figura:



En caso de máxima concentración, la gráfica de la curva de Lorenz está compuesta por una línea que concuerda con el eje horizontal y en p=1 nace una perpendicular, como lo muestra la siguiente figura:



Considerando la forma de la curva de Lorenz y el dominio de la función, podemos darnos cuenta que su estructura funcional obedece a funciones de la forma $q = p^{\alpha}$ para valores de $\alpha \geq 1$. Tomando en consideración esta propuesta, lo que debemos hacer ahora es la estimación del parámetro mediante el método de los mínimos cuadrados ordinarios.

Indice de Gini Mediante el Modelamiento de la Curva de Lorenz

Como vemos, existe una relación entre el índice de Gini y la curva de Lorenz: a mayor alejamiento de la diagonal, por parte de la curva de Lorenz, mayor valor tomará el índice de concentración de Gini. Si por el contrario, la curva de Lorenz tiende a la diagonal, el índice de Gini tomará menor valor, acercándose a cero.

Mientras que el índice de Gini nos da un valor indicativo del nivel de concentración producido en la distribución del recurso, la curva de Lorenz nos describe gráficamente ese fenómeno, pudiendo identificar para que grupos de sujetos se acentúa la concentración y para cuáles de ellos se aminora.

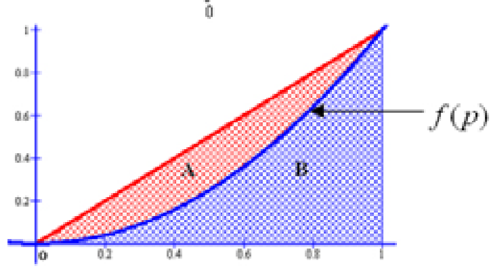
Si le damos un carácter continuo a los puntos, la expresión que define al índice de Gini queda expresada como:

$$\lim_{\Delta p \to 0} I_G = \frac{\lim_{\Delta p \to 0} \sum (p_i - q_i)\Delta p}{\lim_{\Delta p \to 0} \sum p_i \Delta p}$$

Esta expresión, en el límite, representa geométricamente el siguiente cuociente:

$$I_G = \frac{\acute{A}rea\ A}{\acute{A}rea\ A + \acute{A}rea\ B} = \frac{\frac{1}{2} - \int_0^1 f(p)dp}{\frac{1}{2}}$$

$$I_G = 1 - 2\int_0^1 f(p)dp$$



Estimación de α

El modelo funcional está dado por:

$$q = f(p) = p^{\alpha} \cdot \varepsilon$$

Al linealizar este modelo, se obtiene la expresión:

$$Ln\ q = \alpha \cdot Ln\ p + Ln\varepsilon$$

El estimador mínimo cuadrático del parámetro α está dado por:

$$\hat{\alpha} = \frac{\sum Ln\ p \cdot Ln\ q}{\sum (Ln\ p)^2}$$

Esta expresión permite encontrar una estimación del exponente de la función propuesta de Lorenz y a partir de ella realizar los cálculos para encontrar el índice, el que en este caso particular queda definido como:

$$I_G = 1 - 2\int_0^1 p^{\alpha} dp = \frac{\alpha - 1}{\alpha + 1}$$

La varianza del estimador del parámetro α está dado por:

$$Var(\hat{\alpha}) = \frac{S_e^2}{\sqrt{\sum_{i=1}^{N} (Lnp)^2}}$$

A partir de estos datos, se puede construir el intervalo para el parámetro, el que viene dado por:

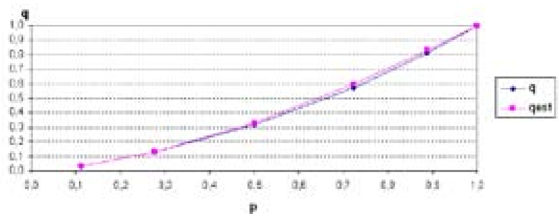
$$IC(\alpha)_{1-\alpha} : \hat{\alpha} \pm t_{(1-\frac{\alpha}{2};N-1)} \cdot S(\hat{\alpha})$$

Diferencia Cuadrática de los Residuos Entre la Curva Tradicional de Lorenz Versus Curva Propuesta. Indice de Gini

A continuación se mostrará distintas distribuciones de una variable ficticia en que se ha graficado la curva de Lorenz y el cálculo del índice de Gini bajo la metodología tradicional y el comportamiento observado bajo la metodología propuesta. Además se ha calculado la suma cuadrática de los residuos.

Caso 1: Distribución simétrica

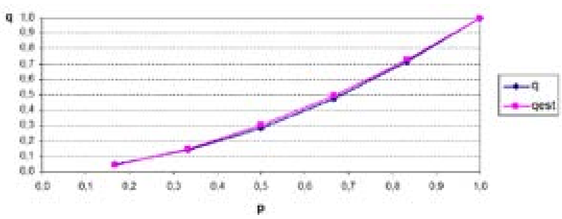
Curvas de Lorenz (tradicional y propuesta)



Indice de Gini (trad)	Indice de Gini (propuesto)	SC Residual	SC Total
0,26	0,22	0,0013	0,74

Caso 2: Distribución uniforme:

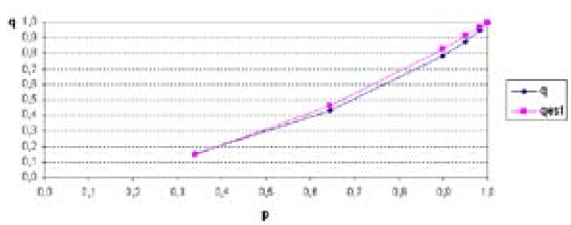
Curvas de Lorenz (tradicional y propuesta)



Indice de Gini (trad)	Indice de Gini (propuesto)	SC Residual	SC Total
0,24	0,27	0,0009	0,65

Caso 3: Distribución asimétrica con sesgo positivo.

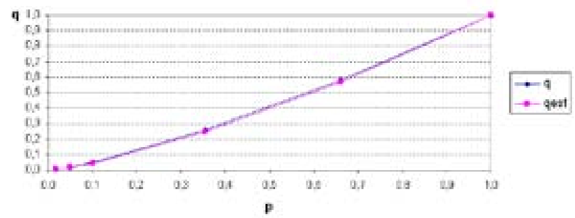
Curvas de Lorenz (tradicional y propuesta)



Indice de Gini (trad)	Indice de Gini (propuesto)	SC Residual	SC Total
0,13	0,27	0,0045	0,559

Caso 4: Distribución asimétrica con sesgo negativo.

Curvas de Lorenz (tradicional y propuesta)



Indice de Gini (trad)	Indice de Gini (propuesto)	SC Residual	SC Total
0,13	0,15	0,00035	0,797

Conclusiones

De acuerdo al desarrollo realizado, las principales conclusiones son las siguientes:

1.- Desde el punto de vista de los fundamentos matemáticos y estadísticos, la propuesta de encontrar la expresión que describe la curva de Lorenz está bien definida.

2.- La obtención del índice de Gini, mediante la estimación del parámetro α está sustentada por los fundamentos del cálculo diferencial e integral.

3.- La proximidad entre la curva de Lorenz obtenida mediante el procedimiento tradicional no tiene diferencia respecto de la curva estimada según el nuevo procedimiento.

4.- La diferencia entre el índice de Gini obtenida por el método tradicional y el propuesto se debe a que este último es calculado sobre una curva continua de los puntos (p,q).

Bibliografía:

1.- Alfonso Novales: Econometría. 2ª edición. Editorial McGraw-Hill
2.- Damodar N. Gujarati: Econometría. 4ª edición. Editorial McGraw-Hill