ESTADÍSTICA ESPAÑOLA
Vol. 45, Núm. 152, 2003, págs. 22 a 54

# Sobre concentración económica: Índice E para colectivos discretos

por
JOSÉ MARÍA MONTERO LORENZO
Facultad de Ciencias Jurídicas y Sociales de Toledo
Universidad de Castilla-La Mancha

## RESUMEN

En este trabajo se demuestra la exactitud del Indice de Gini a la hora de medir la concentración existente en el reparto de la masa total de una variable entre los elementos de un colectivo discreto (errores de agrupamiento aparte en distribuciones por intervalos), poniendo de manifiesto la no necesidad de nuevos "índices exactos" de concentración puesto que el legado por Gini ya cumple esta característica. También apuntaremos la inconveniencia de la utilización de expresiones aproximatorias cuando de colectivos discretos con frecuencias no unitarias se trata.

Dada la exactitud del Indice de Gini, lo que sí merece la pena es abordar nuevas versiones del mismo que reduzcan significativamente su coste operativo, ya que viene expresado en términos del número de elementos del colectivo y éste, generalmente, es muy elevado. En este sentido se presenta el Índice E, exacto y de un coste operativo mínimo, que, además, constituye un estimador insesgado del índice de Gini para distribuciones continuas.

*Palabras clave*: Índice de Gini, índice E, área de concentración, área de máxima concentración, curva de Lorenz.

*Clasificación AMS*. 6201

## 1. INTRODUCCIÓN

La cuestión relativa a la medición de la concentración existente en el reparto de la masa total de una magnitud de carácter económico(1) es importante a muchos efectos y, por consiguiente, no es extraño que sobre ella se haya escrito con profusión.

Desde una de las perspectivas más populares, la medición de la concentración existente en el reparto de una magnitud es una cuestión, en principio, simple: A partir de la famosa curva de Lorenz, se trata de obtener la relación por cociente entre el área de concentración (delimitada por la recta de equidistribución y la curva de concentración) y el área de máxima concentración, que no es otra que la comprendida entre la línea de equidistribución y la supuesta curva de concentración correspondiente al caso en que un sólo elemento del colectivo acapare la totalidad de la masa de la variable objeto de reparto.

Muchos son los índices o expresiones que tratan de reflejar la relación anterior, si bien el más utilizado (con multitud de versiones que aparentan ser distintos índices) es el ya clásico razón entre la desviación media simple de Gini y el doble de la media aritmética(2). Sin embargo, para lograr una mejor comprensión de lo que representa el índice de concentración, suele dársele otra forma, conocida por todos, puesto que es la que contemplan mayoritariamente los textos de Estadística Descriptiva que abordan esta cuestión(3):

---

(1) Que denotaremos por $MTV = \sum_{i=1}^{N} x_i\, n_i = \sum_{i=1}^{k} x_i\, n_i$ siendo N el número de elementos del colectivo y k el número de valores distintos de la variable.

(2) La mayoría de los autores plantean el Índice de Gini para colectivos discretos como la diferencia media con repetición dividida por el doble de la media aritmética. Esto es conceptualmente un error, si bien en la práctica carece de importancia cuando el tamaño del colectivo es elevado.

(3) Vid. GINI, C. (1953), pág. 213. Vid. también GINI, C. (1921).

$$IG = \frac{\sum_{i=1}^{N-1} \left( P_i - Q_i \right)}{\sum_{i=1}^{N-1} P_i} = 1 - \frac{\sum_{i=1}^{N-1} Q_i}{\sum_{i=1}^{N-1} P_i} \qquad [1]$$

siendo N el número de elementos del colectivo, $P_i$ el porcentaje que suponen respecto del total los i elementos del colectivo con menor intensidad de la variable, y $Q_i$ el porcentaje de masa total de la variable que acumulan dichos elementos. Por consiguiente, en dicha expresión, en el caso clásico de distribución de rentas, se comparan proporciones de rentistas con proporciones de renta acumulada por ellos.

Pues bien, la expresión [1] es exactamente igual a la diferencia media simple de Gini dividida por el doble de la media aritmética, si bien al considerar el número de elementos del colectivo en vez del número de valores distintos de la variable cuya masa es objeto de reparto, su cálculo puede resultar tedioso(4). Por ello, en este trabajo se plantea una reformulación de la expresión [1] en términos de los distintos valores de la variable. Dicha reformulación será bautizada con el nombre de Índice E (IE)(5), y su expresión es mucho más sencilla y rápida de calcular que cualquiera de las versiones supuestamente distintas de la relación original propuesta por Gini y, por supuesto, aglutina a todas ellas sea como sea el tipo de distribución.

Además, dada su simplicidad y reducido coste operativo, permitirá evitar aproximaciones que, en muchos casos, proporcionan resultados que distan mucho de ser satisfactorios. La más popular aproxima la relación entre el área de concentración y el área de máxima concentración mediante la expresión(6):

$$IG = \frac{\sum_{i=1}^{k-1} \left( P_i - Q_i \right)}{\sum_{i=1}^{k-1} P_i} = 1 - \frac{\sum_{i=1}^{k-1} Q_i}{\sum_{i=1}^{k-1} P_i} \qquad [2]$$

donde k es el número de valores distintos de la variable, suponiendo que

(4) Más tedioso aún resulta el cálculo de la diferencia media de Gini dividida por el doble de la media aritmética.

(5) En homenaje póstumo al Ingeniero de Telecomunicaciones D. Emilio Lorenzo Lorenzo.

(6) Muy frecuente, sobre todo, en los manuales de Estadística Descriptiva. Algunos pueden verse citados en FERREIRA, E.; GARÍN, A. (1997), pág 110.

$$1-\frac{\sum_{i=1}^{k-1} Q_i}{\sum_{i=1}^{k-1} P_i} \cong 1-\frac{\sum_{i=1}^{N-1} Q_i}{\sum_{i=1}^{N-1} P_i} \qquad [3]$$

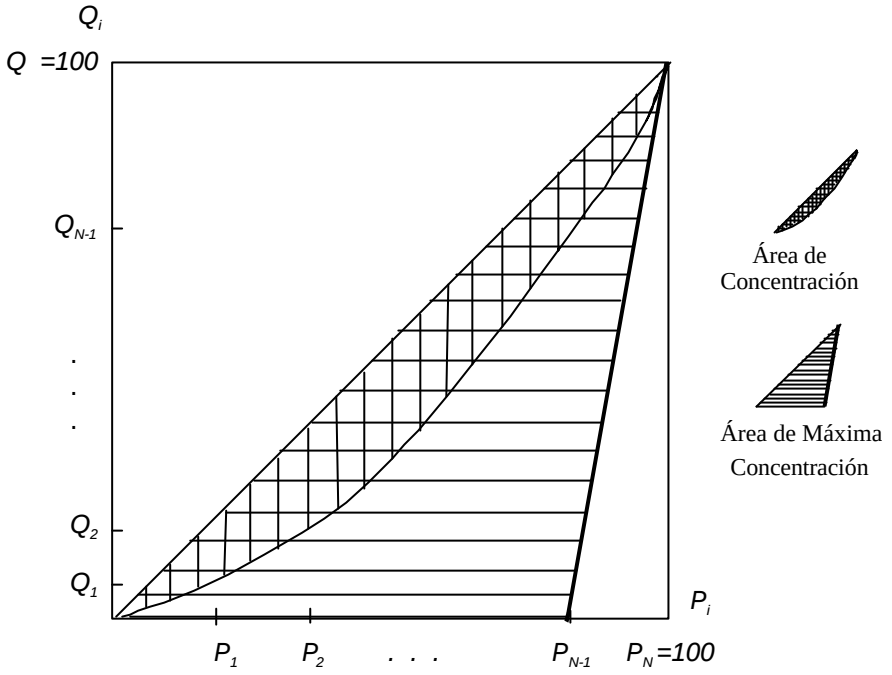
Dicho lo anterior, el artículo se estructura en varias partes. En la primera de ellas demostraremos cómo la expresión [1] mide exactamente la concentración existente en el reparto de la masa total de una variable entre los elementos del colectivo. En la segunda abordaremos la inconveniencia de la utilización de expresiones aproximatorias del legado de Gini (en concreto la expuesta anteriormente, por ser de uso generalizado en el ámbito académico) para los casos en que la distribución de frecuencias no es unitaria, así como pondremos de manifiesto nuestra sorpresa ante algunas afirmaciones atribuidas al propio Gini relativas a la exactitud (en realidad, a la no exactitud) de su indicador de concentración. Ahora bien, dado que el Índice de Gini exige tomar los elementos del colectivo uno a uno, no resultará apropiado -en el sentido de un excesivo coste operativo- para el caso de frecuencias no unitarias por lo que, en la tercera parte, formularemos una versión de dicho índice en términos de los valores distintos de la variable (evidentemente muchos menos) que medirá de forma exacta la relación entre las áreas de concentración y máxima concentración(7) y exigirá un coste operativo mínimo. Dicha versión ha sido bautizada con el nombre de Índice E (IE). Finalmente, demostraremos que el IE es un estimador insesgado del Índice de Gini para distribuciones continuas.

## 2. EL LEGADO DE CORRADO GINI: UN ÍNDICE PERFECTO

A partir de la curva de concentración se deduce de forma inmediata que cuanto mayor sea el área de concentración respecto del área de máxima concentración, mayor será la desigualdad en el reparto de la masa total de la variable. Por consiguiente, de manera natural, las medidas o índices de concentración se obtendrán por cociente entre las dos áreas mencionadas.

(7) En el caso de distribuciones agrupadas por intervalos el índice proporcionará una aproximación debido a los inevitables errores de agrupamiento. Esta matización deberá ser tenida en cuenta a lo largo del texto.

**Gráfico 1**




La más utilizada de las medidas de concentración es la razón de concentración, más conocida en la literatura estadística al uso como Índice de Gini. Su fundamento es el siguiente: A partir de la curva de concentración (curva de Lorenz), y siendo N el número de elementos del colectivo, pueden establecerse N-1 desigualdades entre $P_i$ y $Q_i$ (recuérdese que $P_N$=$Q_N$=100 necesariamente) y de la amplitud de las mismas dependerá el mayor o menor nivel de desigualdad en el reparto de la masa total de la variable(8).

Cuanto más grandes son tales diferencias mayor es la concentración en el reparto de la masa total de la variable entre los elementos de la distribución. Por consiguiente, una primera expresión para cuantificar el nivel de concentración es $\sum_{i=1}^{N-1}(P_i - Q_i)$.

(8) Recuérdese que $P_i$ es el porcentaje que suponen los i elementos de la distribución con menor valor de la característica cuya concentración se pretende medir, respecto del tamaño total del colectivo. $Q_i$ es el porcentaje de la característica cuya concentración se pretende medir, acumulado por los i elementos de la distribución con menor valor o intensidad de dicha característica.

Ahora bien, queda claro que en caso de equidistribución la expresión anterior vale 0 (todos los $P_i$ coinciden con los respectivos $Q_i$ para los N-1 primeros elementos de la distribución) pero, sin embargo, su límite superior dependerá de la distribución en cuestión, lo que implica que, para determinar si la distribución está poco o muy concentrada, deberemos calcular dicho límite en cada caso. En consecuencia, para mayor facilidad de interpretación y comparación, modificaremos la expresión anterior de tal forma que, manteniendo su fundamento, su límite inferior siga siendo cero pero su límite superior valga siempre 1. Ello se consigue sin más que dividirla por $\sum_{i=1}^{N-1} P_i$ .

La expresión resultante es el índice propuesto por Gini para la cuantificación de la concentración existente en el reparto de la masa total de una variable entre los elementos del colectivo:

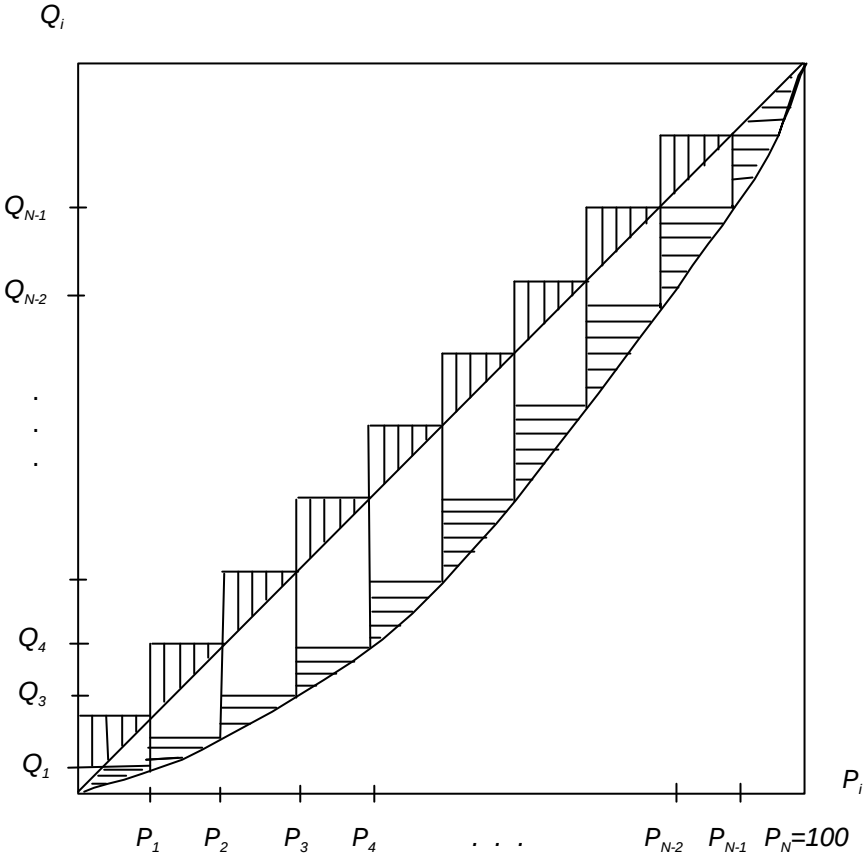
$$IG = \frac{\sum_{i=1}^{N-1} (P_i - Q_i)}{\sum_{i=1}^{N-1} P_i} = 1 - \frac{\sum_{i=1}^{N-1} Q_i}{\sum_{i=1}^{N-1} P_i} \qquad [4]$$

y su interpretación es sobradamente conocida.

En términos geométricos, si se multiplican las diferencias ($P_i$-$Q_i$) por la diferencia constante 100/N entre dos valores sucesivos de $P_i$ se tiene que la suma de dichos productos para todos los elementos de la distribución menos el último proporcionará el valor exacto del área de concentración. Esto parece, en principio, un poco extraño a la vista del Gráfico Nº 2. Es más, resulta sorprendente que el propio Gini, en su Curso de Estadística (versión española con traducción y adaptación de la versión italiana de 1946-47 de Jorge Stecher Navarra), señale que la suma de los productos anteriormente mencionados *"representa un valor aproximado del área de concentración cuya aproximación será mayor cuanto mayor sea N"*. Incluso en dicho texto se llega a afirmar que *"para un N muy grande la suma de estos productos será prácticamente igual al área de concentración"*.

Sin embargo, es sencillo demostrar que el área comprendida entre la línea de equidistribución y la curva de concentración (área de concentración) coincide exactamente con el área de los N-1 rectángulos de base ($P_i$-$P_{i-1}$=1/100) y de altura ($P_i$-$Q_i$), incluso aunque el tamaño del colectivo sea N=2. Para ello ha de ocurrir que el área rayada verticalmente en el Gráfico 2 coincida con el área de rayado horizontal.

**Gráfico 2**




Veamos que efectivamente es así: El área rayada verticalmente está compuesta por un trapecio y N-2 rectángulos, siendo el área del trapecio $\frac{P_1^2}{2} - \frac{Q_1^2}{2}$ y la de cada uno de los N-2 rectángulos $\frac{(P_i - P_{i-1})^2}{2}$ con $1 < i \leq N-1$. En consecuencia, el valor del área rayada verticalmente es:

$$A_V = \frac{P_1^2}{2} - \frac{Q_1^2}{2} + \sum_{i=2}^{N-1} \frac{(P_i - P_{i-1})^2}{2} \qquad [5]$$

y como $P_1 = \frac{1}{N}100$ y $(P_i - P_{i-1}) = \frac{1}{N}100$, entonces se tiene que el valor del área rayada verticalmente no es otro que:

$$A_V = \frac{10.000}{2N^2} + \sum_{i=2}^{N-1} \frac{10.000}{2N^2} - \frac{Q_1^2}{2} = 5.000\frac{N-1}{N^2} - \frac{Q_1^2}{2} \qquad [6]$$

El área rayada horizontalmente está formada por N triángulos. Considerándolos de izquierda a derecha se tiene que:

– El área del primero es $\frac{P_1 Q_1}{2} - \frac{Q_1^2}{2}$.

– El área del último vale $(P_N - P_{N-1})(P_{N-1} - Q_{N-1}) + \frac{1}{2}(P_N - P_{N-1})^2 - \frac{1}{2}(P_N - P_{N-1})(P_N - Q_{N-1})$.

– El área de cada uno de los N-2 triángulos intermedios es $\frac{(P_i - P_{i-1})(Q_i - Q_{i-1})}{2}$ $1 < i \leq N-1$.

En consecuencia, el valor del área rayada horizontalmente es:

$$A_H = \frac{P_1 Q_1}{2} - \frac{Q_1^2}{2} + \frac{1}{2}\sum_{i=2}^{N-1}(P_i - P_{i-1})(Q_i - Q_{i-1}) + (P_N - P_{N-1})(P_{N-1} - Q_{N-1}) +$$

$$+ \frac{1}{2}(P_N - P_{N-1})^2 - \frac{1}{2}(P_N - P_{N-1})(P_N - Q_{N-1})$$

y como $P_1 = \frac{1}{N}100$ ; $P_i - P_{i-1} = \frac{1}{N}100$ ; $P_N = 100$ ; $P_{N-1} = \frac{N-1}{N}100$, entonces

$$A_H = \frac{100Q_1}{2N} - \frac{Q_1^2}{2} + \frac{100}{2N}\left[Q_2 - Q_1 + Q_3 - Q_2 + Q_4 - Q_3 + \ldots\ldots + Q_{N-1} - Q_{N-2}\right] +$$

$$+ \frac{100}{N}\left[\frac{100(N-1)}{N} - Q_{N-1}\right] + \frac{10.000}{2N^2} - \frac{100}{2N}\left[100 - Q_{N-1}\right]$$

$$= \frac{100Q_1}{2N} - \frac{Q_1^2}{2} + \frac{100}{2N}\left[Q_{N-1} - Q_1\right] + \frac{10.000}{N} - \frac{10.000}{N^2} - \frac{100Q_{N-1}}{N} +$$

$$+ \frac{10.000}{2N^2} - \frac{10.000}{2N} + \frac{100Q_{N-1}}{2N} \qquad [7]$$

$$= -\frac{Q_1^2}{2} + \frac{10.000}{2N} - \frac{10.000}{2N^2}$$

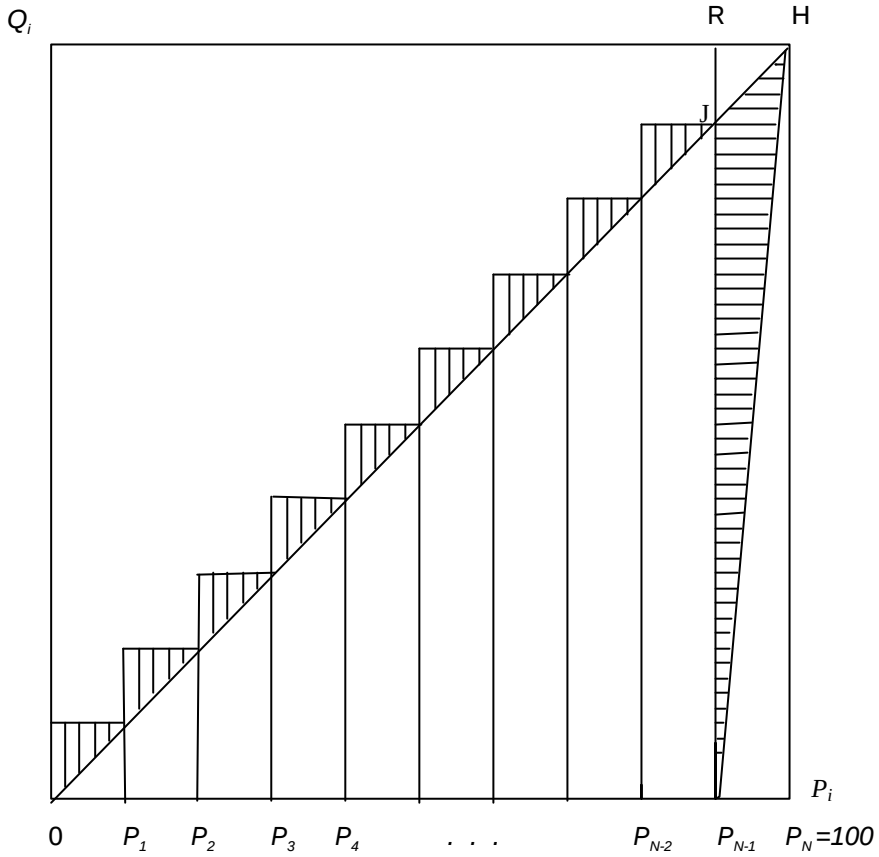
$$= 5.000\frac{N-1}{N^2} - \frac{Q_1^2}{2}$$

con lo que queda demostrado que el valor exacto del área de concentración es

$$AC = \frac{100}{N}\sum_{i=1}^{N-1}\left(P_i - Q_i\right) \qquad [8]$$

El área de máxima concentración (AMC) refleja la situación en la que un único individuo del colectivo acapara toda la masa de la variable, por lo que vendrá dada por el área del triángulo $OP_{N-1}H$ (Gráfico 3).

**Gráfico 3**




En la expresión propuesta por Corrado Gini tal área viene dada por la expresión $\frac{100}{N}\sum_{i=1}^{N-1} P_i$ que, dado que $P_i = \frac{100i}{N}$, no representa sino la suma de las áreas de los N-1 rectángulos del Gráfico 3, que, aparentemente, no parece coincidir con el área del triángulo $OP_{N-1}H$. Sin embargo, vamos a ver que sí coinciden(9). Para ello bastará con demostrar de nuevo que el área rayada verticalmente en el Gráfico 3 coincide exactamente con el área rayada en horizontal:

(9) Según la versión española del texto de Gini su expresión es también una aproximación al área de máxima concentración.

El área rayada verticalmente vale

$$A_V = \frac{P_1^2}{2} + \frac{P_2(P_2 - P_1)}{2} + \frac{P_3(P_3 - P_2)}{2} + \ldots\ldots + \frac{P_{N-1}(P_{N-1} - P_{N-2})}{2}$$

y como $P_1 = \frac{100}{N}$ y $P_i - P_{i-1} = \frac{100}{N}$, entonces

$$A_V = \frac{N-1}{2}\frac{10.000}{N^2} = 5.000\frac{N-1}{N^2} \qquad [9]$$

El valor del área rayada horizontalmente es la diferencia entre las áreas de los triángulos $P_{N-1}RH$ y JRH. Pero, como $H - R = P_N - P_{N-1} = \frac{100}{N}$, entonces

$$A_H = \frac{P_N(P_N - P_{N-1})}{2} - \frac{(P_N - P_{N-1})^2}{2} = \frac{10.000}{2N} - \frac{10.000}{2N^2} = 5.000\frac{N-1}{N^2} \qquad [10]$$

con lo que queda demostrado que

$$AMC = \frac{100}{N}\sum_{i=1}^{N-1} P_i \qquad [11]$$

y que la expresión legada por Corrado Gini

$$IG = \frac{\frac{100}{N}\sum_{i=1}^{N-1}(P_i - Q_i)}{\frac{100}{N}\sum_{i=1}^{N-1} P_i} = \frac{\sum_{i=1}^{N-1}(P_i - Q_i)}{\sum_{i=1}^{N-1} P_i} = 1 - \frac{\sum_{i=1}^{N-1} Q_i}{\sum_{i=1}^{N-1} P_i} \qquad [12]$$

refleja de forma exacta (y no aproximada) la relación por cociente entre las áreas de concentración y de máxima concentración.

Ahora bien, es necesario hacer una matización respecto al valor exacto del área de máxima concentración, y es que, en colectivos discretos, no alcanza nunca el valor ½ ó 5.000 (dependiendo de si se opera en tantos por uno o tantos por ciento). En concreto su valor exacto, cuando se trabaja en tantos por ciento, es el siguiente(10):

(10) Si se trabaja en tantos por uno su valor exacto es (N-1)/2N.

$$AMC=\frac{100}{N}\sum_{i=1}^{N-1}P_i=\frac{100}{N}\sum_{i=1}^{N-1}\frac{i}{N}100=\frac{10.000}{N^2}\sum_{i=1}^{N-1}i=\frac{10.000}{N^2}\left[\sum_{i=1}^{N}i-N\right]$$

$$=\frac{10.000}{N^2}\left[\frac{(N+1)N}{2}-N\right]=5.000\frac{N-1}{N} \qquad [13]$$

cuestión de absoluta relevancia, sobre todo, cuando el tamaño del colectivo es pequeño.

En consecuencia, a partir de la expresión anterior se tiene que

$$\sum_{i=1}^{N-1}P_i=\frac{AMC}{\frac{100}{N}}=\frac{5.000\frac{N-1}{N}}{\frac{100}{N}}=50(N-1) \qquad [14]$$

con lo que la expresión tradicional de Gini también puede escribirse como

$$I_G=1-\frac{\sum_{i=1}^{N-1}Q_i}{50(N-1)} \qquad [15]$$

Por consiguiente, cuando se consideran los elementos del colectivo uno a uno, la expresión [15], idéntica a la [1], es una medida exacta de la relación por cociente entre el área de concentración y el área de máxima concentración. En consecuencia, no son necesarios nuevos procedimientos de “medición exacta de áreas” mediante mecanismos geométricos, formulaciones matriciales, etc. Lo único necesario es reducir el coste operativo de la expresión anterior, puesto que trabaja con el número de elementos del colectivo (N) y, si éste es muy elevado, el coste operativo puede ser realmente importante.

## 3. LA INCONVENIENCIA DE LA UTILIZACIÓN DE EXPRESIONES APROXIMATORIAS AL ÍNDICE DE GINI EN DISTRIBUCIONES DISCRETAS CON FRECUENCIAS NO UNITARIAS

En este apartado trataremos de poner de manifiesto dos cuestiones:

1) La inconveniencia de las expresiones aproximatorias del tipo de la [2] en el caso de distribuciones discretas con frecuencias no unitarias.

2) Nuestra sorpresa ante algunas afirmaciones contenidas en la versión española de 1953 del "Curso de Estadística" de Gini (de indudable importancia por haber constituido un referente importante de los estudiosos españoles sobre la cuestión) que parecen dar a entender que su índice es una aproximación a la relación entre las áreas de concentración y máxima concentración, tanto mejor cuanto mayor sea el tamaño del colectivo.

En cuanto a la primera de las cuestiones, hemos demostrado en el punto anterior que la expresión

$$IG=\frac{\frac{100}{N}\sum_{i=1}^{N-1}(P_i-Q_i)}{\frac{100}{N}\sum_{i=1}^{N-1}P_i}=\frac{\sum_{i=1}^{N-1}(P_i-Q_i)}{\sum_{i=1}^{N-1}P_i}=1-\frac{\sum_{i=1}^{N-1}Q_i}{\sum_{i=1}^{N-1}P_i}=1-\frac{\sum_{i=1}^{N-1}Q_i}{50(N-1)} \qquad [16]$$

es una medición absolutamente exacta de la relación por cociente entre el área de concentración y el área de máxima concentración.

Ahora bien, en muchas ocasiones, y generalmente a modo de aproximación, cuando se propone la expresión anterior para una distribución de frecuencias no unitarias, se hace en términos del número de valores de la distribución, es decir, bajo la forma

$$IG=\frac{\sum_{i=1}^{k-1}(P_i-Q_i)}{\sum_{i=1}^{k-1}P_i}=1-\frac{\sum_{i=1}^{k-1}Q_i}{\sum_{i=1}^{k-1}P_i} \qquad [17]$$

que no refleja la relación por cociente entre las dos áreas involucradas en el índice. Y ello por dos razones:

1) Las bases de los rectángulos que integrarán las áreas de concentración o máxima concentración en cada estrato o valor de la variable son generalmente distintas y se están tomando como iguales.

2) No se están considerando las áreas rectangulares correspondientes al último valor de la variable, lo cual puede llevar a mayúsculos distanciamientos del verdadero valor del índice.

Algunos autores(11) apuntan que la expresión anterior es "*una mala aproximación en colectivos con pocos valores diferentes (k pequeño)*". Sin embargo, a nuestro juicio, se trata más bien de una generalización poco afortunada de la expresión inicial de Gini. Además, es muy frecuente achacar el carácter de "*mala aproximación*" de la expresión [17] a la aproximación de las áreas de concentración y máxima concentración mediante rectángulos, cuando hemos demostrado que si en vez de trabajar con el número de valores distintos de la variable se trabaja con los elementos del colectivo tomados uno a uno, la misma expresión proporciona una medición exacta de ambas áreas y, en consecuencia, de la concentración existente en el reparto de la masa total de la variable entre los integrantes de dicho colectivo.

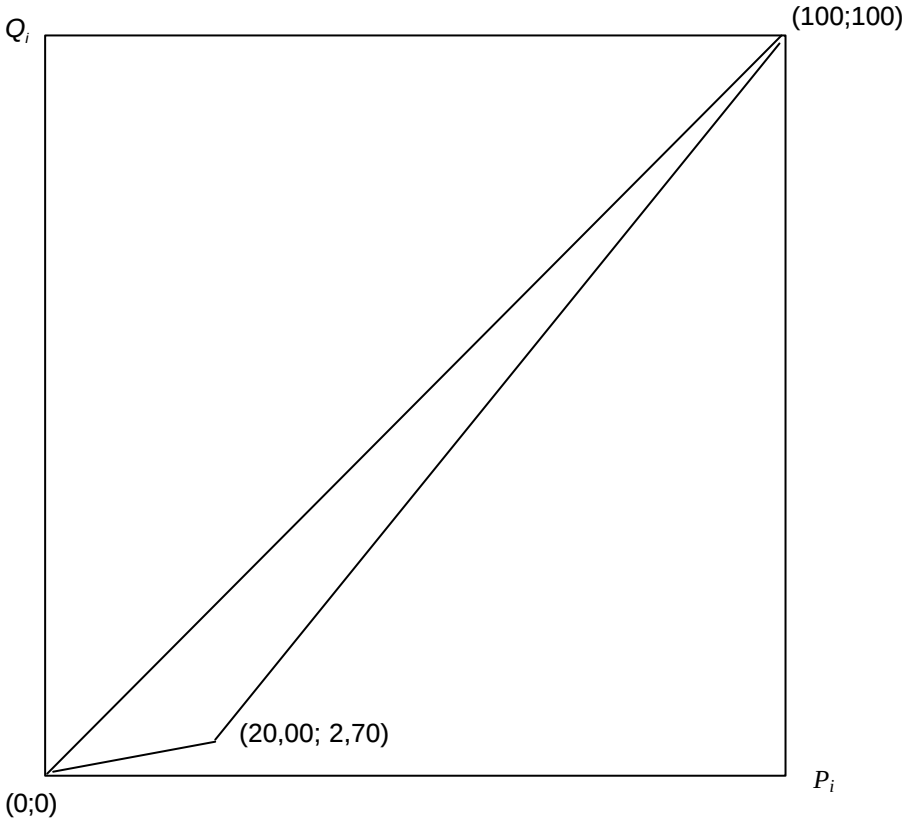
A modo de ejemplo, considérese la siguiente situación. Se trata de calcular el Índice de Gini para la siguiente distribución de frecuencias:

| $x_i$ | $n_i$ | $P_i$ | $Q_i$ | $P_i$-$Q_i$ |
|---|---|---|---|---|
| 800 | 20 | 20,00 | 2,70 | 17,30 |
| 7.200 | 80 | 100,00 | 100 | 0,00 |

para la cual la curva de concentración es la siguiente:

(11) Vid., por ejemplo, FERREIRA, E.; GARÍN, A. (1997), pág 210.

**Gráfico 4**




Resulta evidente, a simple vista, que la distribución de la masa total de la variable está bastante próxima a la equidistribución y, sin embargo, la solución que proporciona la expresión aproximatoria [2] es

$$IG=\frac{\sum_{i=1}^{k-1}(P_i - Q_i)}{\sum_{i=1}^{k-1}P_i}=\frac{17,30}{20}=0,8648 \qquad [18]$$

que indicaría justo lo contrario (la verdadera razón de áreas es 0,1747).

Es solo un ejemplo de las distorsiones que pueden tener lugar al calcular la concentración existente en distribuciones con frecuencias no unitarias mediante la

expresión aproximatoria anterior, trabajando con el número de valores distintos de la variable (k) en vez de con el número de elementos de colectivo (N).

Para finalizar con la primera de las cuestiones planteadas en este apartado, hemos de poner de manifiesto que Gini, en su "Curso de Estadística con apéndice matemático de Luigi Galvani", parte de distribuciones de frecuencias unitarias y, en este sentido, para explicar su archiconocido índice señala: "*Tomemos n cantidades* $a_1, a_2, ...., a_i, ...., a_n$, *que midan la intensidad de un carácter determinado en n casos, por ejemplo la renta de n individuos.....*". Parece claro que el que fue profesor y decano de la Facultad de Ciencias Estadísticas, Demográficas y Actuariales de la Universidad de Roma considera que las frecuencias son unitarias y nada dice sobre la expresión del índice cuando éstas no revisten dicho carácter. Es más, en la página 214 del mencionado Curso de Estadística señala que "*A menudo no son conocidos los n valores de* $p_i$ *y* $q_i$ (en nuestra nomenclatura $P_i$ y $Q_i$ ), *sino sólo algunos de ellos, puesto que las tablas estadísticas no indican la intensidad del fenómeno en todos los n casos, sino una clasificación de dichas intensidades en un número reducido de clases*", y el traductor y adaptador de la obra, Jorge Stecher Navarra, lo ilustra con una distribución agrupada por intervalos relativa a la clasificación de los municipios de España y su población respectiva en 13 intervalos, según datos del Censo de 1940. Parece estar introduciendo las bases para elaborar una expresión aproximatoria de su índice de concentración en el caso en que la distribución esté agrupada por intervalos y las frecuencias no sean unitarias; en este caso resultan inevitables los errores de agrupamiento, pero lo importante, a nuestro juicio, es que parece confirmar la no adecuación de la popular expresión aproximatoria [2] al caso en que se trabaje con distribuciones de frecuencias no unitarias. Sin embargo, esta situación es muy frecuente, sobre todo en el ámbito académico.

En cuanto a la segunda de las cuestiones planteadas al principio de este epígrafe, resulta curioso leer en la página 213 de la versión española de 1953 del Curso de Estadística de Corrado Gini que(12)"$\frac{1}{n}\sum_{i=1}^{n-1}(p_i - q_i)$ *representa un valor aproximado del área de concentración cuya aproximación será tanto mayor cuanto mayor sea n. Por lo tanto, para n muy grande la suma de estos productos será prácticamente igual al área de concentración. Con análoga aproximación, la suma*

(12) En la lectura de las citas hemos de tener en cuenta que:

1.- Gini trabaja en tantos por uno y nosotros en tantos por ciento, por lo cual utiliza un factor 1/n en vez de 100/n.

2.- Que al tamaño del colectivo lo llama n en vez de N.

3.- Que denomina $p_i$ y $q_i$ a nuestros $P_i$ y $Q_i$ , respectivamente.

$\frac{1}{n}\sum_{i=1}^{n-1} p_i$ *es igual al área del triángulo de máxima concentración; por lo tanto, R* (se refiere al cociente de ambas) *representa la razón existente entre el área delimitada por la recta de equidistribución y la curva de concentración, y el área correspondiente al caso de máxima concentración* (y señala un triángulo de area ½)".

De la lectura de este párrafo se deducen varias cuestiones:

1. Que no parece considerarse que $\frac{1}{n}\sum_{i=1}^{n-1}(p_i - q_i)$ , ó $\frac{100}{N}\sum_{i=1}^{N-1}(P_i - Q_i)$ en nuestra terminología, sea una medición exacta del valor del área de concentración, tal y como hemos demostrado en el punto anterior.

2. Igualmente, no parece considerarse que $\frac{1}{n}\sum_{i=1}^{n-1} p_i$ , ó $\frac{100}{N}\sum_{i=1}^{N-1} P_i$ en nuestra notación, represente exactamente el valor del área de máxima concentración.

3. Por consiguiente, de lo anterior no parece calificarse la expresión [1] como una medición exacta de la concentración existente en el reparto de la masa de una variable entre los elementos de un colectivo discreto, sino como una aproximación (y así lo reflejan los autores españoles de mediados del siglo pasado, quizás influenciados por esta obra).

4. Gini supone que el valor del área de máxima concentración es ½ , ó 50 si se trabaja en tantos por ciento, cuando, como hemos demostrado anteriormente, es ½(N-1) ( si se trabaja en tantos por uno) ó 50(N-1) (si se opera en tantos por ciento). Además, resulta curioso que previamente, en la misma página, exponga que "*obtendríamos la concentración máxima cuando la intensidad del carácter fuera igual a 0 en n-1 casos, e igual a* $A_n$ (se refiere a la totalidad de la masa de la variable) *en un solo caso, esto es, cuando, por ejemplo, toda la renta fuese poseída por una sola persona*". Queda claro que, entonces, el área de máxima concentración es, en tantos por ciento, 50(N-1).

Además, en este sentido, hemos de decir que la medición correcta del área de máxima concentración es una cuestión importantísima cuando el tamaño del colectivo es pequeño y que las consecuencias de una mala medición ya fueron puestas de manifiesto por Theil(13) en 1967.

5. Por otra parte, se plantea lo que hoy es conocido por índice geométrico como una expresión adecuada para calcular, "*aunque sea aproximadamente*", la razón de concentración en el caso de distribuciones agrupadas por intervalos y con frecuen-

(13) Vid. THEIL, H. (1967), pág.. 92, punto (2).

cias no unitarias. En este punto surge una duda: Que el término "*aproximadamente*" haga referencia a que la curva de concentración resultante sea interior a la curva real de concentración por adoptarse el supuesto de uniformidad en el reparto de los valores a lo largo del intervalo. En este caso, ninguna objeción salvo la relativa a la medición incorrecta del área de concentración y que, como propondremos posteriormente, dicho cálculo se puede llevar a cabo con un menor coste operativo. La otra posibilidad, es que el hoy denominado índice geométrico "aproxime" su expresión inicial (índice de Gini) y que la no utilice porque el tamaño del colectivo sea muy grande (de hecho el traductor pone un ejemplo con una población de 25.877.971 habitantes) y el coste operativo de trabajar con los integrantes del colectivo uno a uno sea prohibitivo. Esta última tesitura nos llevaría a pensar que (según el texto citado) no era consciente de que, con el área de máxima concentración correctamente medida, ambos índices son idénticos.

6.- Finalmente, se apunta que $R = \frac{\Delta}{2\bar{x}}$, donde R representa la razón de concentración en su nomenclatura y la diferencia media simple. Ahora bien, dicha igualdad es cierta en el caso de que

$$R = \frac{\sum_{i=1}^{n-1}(p_i - q_i)}{\sum_{i=1}^{n-1} p_i} \qquad [19]$$

pero no cuando

$$R = 1 - \sum_{i=1}^{n}(p_i - p_{i-1})(q_i + q_{i-1}) \qquad [20]$$

pues en este último caso está incorrectamente medida el área de máxima concentración.

La verdadera relación entre dichas expresiones (utilizando ya nuestra nomenclatura) es

$$\frac{\sum_{i=1}^{N-1}(P_i - Q_i)}{\sum_{i=1}^{N-1} P_i} = \frac{N}{N-1}\left[1 - \frac{\sum_{i=1}^{N}(P_i - P_{i-1})(Q_i + Q_{i-1})}{10.000}\right] = \frac{\Delta}{2\bar{x}} \qquad [21]$$

## 4. REFORMULACIÓN DEL ÍNDICE DE GINI: INDICE E DE CONCENTRACIÓN

De lo anteriormente expuesto se deducen tres cosas:

1. El índice originalmente propuesto por Gini -expresión [1]- es un indicador exacto de la concentración existente en el reparto de la masa total de una variable entre los elementos del colectivo (errores de agrupamiento en distribuciones por intervalos aparte). Lástima que el hecho de representar las áreas de concentración y máxima concentración mediante rectángulos haya provocado algunos equívocos y haya dirigido algunas investigaciones a la búsqueda de indicadores "exactos", cuando estamos en presencia de uno.

2. La obtención del índice de Gini exige tomar uno a uno los elementos del colectivo, lo cual, evidentemente, hace muy tedioso su cálculo a poco que el tamaño del colectivo sea mínimamente elevado.

3. Para evitar dicho tedio no resulta conveniente sustituir en la expresión de Gini, N-1 (todos los elementos del colectivo menos el que tiene mayor intensidad de la característica objeto de reparto) por k-1 (todos los valores distintos de la característica menos el mayor). Tal es el caso de la expresión aproximatoria más popular.

Pues bien, en lo que sigue proponemos una reformulación de la expresión original de Corrado Gini, indicador exacto de la relación existente entre las áreas de concentración y máxima concentración, que operará con el número de valores distintos de la variable (k) en vez de con el número de elementos del colectivo (N), con lo cual su cálculo será inmediato. Dicha reformulación ha sido bautizada con el nombre de Índice E (IE).

Dado que, como hemos visto anteriormente:

$$I_G = \frac{\sum_{i=1}^{N-1}(P_i - Q_i)}{\sum_{i=1}^{N-1} P_i} = 1 - \frac{\sum_{i=1}^{N-1} Q_i}{50(N-1)} \qquad [22]$$

la única simplificación posible es la de la expresión $\sum_{i=1}^{N-1} Q_i$. Para llevar a cabo la simplificación de la expresión anterior nos apoyaremos en la tabla de construcción de la curva de concentración.

Sea la siguiente distribución de frecuencias:

| *Distribución de frecuencias* | |
|---|---|
| $x_1$ | $n_1$ |
| $x_2$ | $n_2$ |
| . | . |
| . | . |
| $x_i$ | $n_i$ |
| . | . |
| . | . |
| $x_{k-1}$ | $n_{k-1}$ |
| $x_k$ | $n_k$ |

A continuación, unitarizamos las frecuencias absolutas y procedemos al cálculo de los $Q_i$ (las tres primeras columnas):

**Tabla de construcción del Índice E**

| $x_i$ | $n_i$ | $Q_i$ | $M_i$ |
|---|---|---|---|
| $x_1$ | 1 | $(x_1/mtv)100$ | |
| $x_1$ | 1 | $(2x_1/mtv)100$ | |
| . | . | . | $M_1=$ |
| . | . | . | $(x_1/mtv)100\ [1+2+...n_1]$ |
| $x_1$ | 1 | $(n_1x_1/mtv)100$ | |
| $x_2$ | 1 | $(n_1x_1+x_2/mtv)100$ | |
| $x_2$ | 1 | $(n_1x_1+2x_2/mtv)100$ | |
| . | . | . | $M_2=$ |
| . | . | . | $n_2\ [(n_1x_1)/mtv)]100+(x_2/mtv)100\ [1+2+...n_2]$ |
| $x_2$ | 1 | $(n_1x_1+n_2x_2/mtv)100$ | |
| $x_3$ | 1 | $(n_1x_1+n_2x_2+x_3/mtv)100$ | |
| $x_3$ | 1 | $(n_1x_1+n_2x_2+2x_3/mtv)100$ | |
| . | . | . | $M_3=$ |
| . | . | . | $n_3[(n_1x_1+n_2x_2)/mtv]100+(x_3/mtv)100[1+2+...n_3]$ |
| $x_3$ | 1 | $(n_1x_1+n_2x_2+n_3x_3/mtv)100$ | |
| . | . | . | . |
| . | . | . | . |
| $x_i$ | 1 | $(n_1x_1+..+n_{i-1}x_{i-1}+x_i/mtv)100$ | |
| $x_i$ | 1 | $(n_1x_1+..+n_{i-1}x_{i-1}+2x_i/mtv)100$ | |
| . | . | . | |
| . | . | . | $M_i=$ |
| $x_i$ | 1 | $(n_1x_1+..+n_{i-1}x_{i-1}+n_ix_i/mtv)100$ | $n_i\ [(n_1x_1+..+n_{i-1}x_{i-1})/mtv]100+(x_i/mtv)100\ [1+2+...n_i]$ |
| . | . | . | . |
| . | . | . | . |
| $x_k$ | 1 | $(n_1x_1+..+n_{k-1}x_{k-1}+x_k/mtv)100$ | |
| $x_k$ | 1 | $(n_1x_1+..+n_{i-1}x_{k-1}+2x_k/mtv)100$ | |
| . | . | . | $M_k=$ |
| . | . | . | $n_k\ [(n_1x_1+..+n_{k-1}x_{k-1})/mtv]100+(x_k/mtv)100\ [1+2+...+n_k]$ |
| $x_k$ | 1 | $(n_1x_1+..+n_{k-1}x_{k-1}+(n_k-1)x_k/mtv)100$ | |
| $x_k$ | 1 | $(n_1x_1+..+n_kx_k/mtv)100 = 100$ | |

Seguidamente, en la cuarta columna, se obtienen fácilmente las sumas parciales de los $Q_i$ (una para cada valor de la variable), que denominaremos $M_i$:

$$M_i = n_i \frac{\sum_{j=1}^{i-1} n_j \; x_j}{MTV} 100 + \frac{x_i}{MTV} \sum_{i=1}^{n_i} i \, 100 \quad i = 1,2,\ldots k \qquad [23]$$

donde:

– $n_i$ son las frecuencias absolutas de los valores de la variable.

– MTV es la masa total de la variable, es decir, $MTV = \sum_{i=1}^{k} x_i \; n_i$.

En este momento hemos dado el paso fundamental: el paso de elementos del colectivo (N) a valores de la variable (k, que, evidentemente, son muchos menos).

Pues bien, denominando(14):

– $Q_i$ a la participación porcentual en la masa total de la variable de los elementos cuyo valor de la característica X es como máximo el i-ésimo (ahora i=1, 2, ...., k).

– $prop_i$ a la participación porcentual en la masa total de la variable de un elemento con el i-ésimo valor de la característica objeto de estudio.

Y sabiendo que $\sum_{i=1}^{n_i} i = \frac{n_i (n_i + 1)}{2}$, se llega a la expresión:

$$M_i = n_i Q_{i-1} + \frac{n_i \; (n_i + 1)}{2} prop_i \quad i = 1,2,\ldots\ldots k \qquad [24]$$

Si a partir de la tabla con frecuencias unitarias se calcula el índice de Gini -con la expresión [15]- tenemos que calcular y sumar, para su numerador, los (N-1) primeros $Q_i$. Ahora bien:

(14) Por tanto, en lo que sigue, $P_i$ y $Q_i$ se referirán a acumulaciones hasta el valor i-ésimo de la variable si se trabaja con valores de la variable y a acumulaciones hasta el i-ésimo elemento del colectivo si se trabaja con la totalidad de éstos.

$$\sum_{i=1}^{N-1} Q_i = \sum_{i=1}^{n_1} Q_i + \sum_{i=n_1+1}^{n_2} Q_i + \sum_{i=n_2+1}^{n_3} Q_i + \ldots\ldots + \sum_{i=n_{k-1}+1}^{n_k-1} Q_i$$

[25]

$$= M_1 + M_2 + M_3 + \ldots\ldots + M_k - 100 = \sum_{i=1}^{k} M_i - 100$$

con lo que,

$$IG = 1 - \frac{\sum_{i=1}^{N-1} Q_i}{50(N-1)} = 1 - \frac{\sum_{i=1}^{k} M_i - 100}{50(N-1)}$$

$$= \frac{50(N-1)}{50(N-1)} - \frac{\sum_{i=1}^{k} M_i}{50(N-1)} + \frac{100}{50(N-1)}$$

$$= \frac{N+1}{N-1} - \frac{\sum_{i=1}^{k} M_i}{50(N-1)}$$

siendo esta formulación del índice de Gini la que hemos dado en denominar Indice E, el índice que proponemos para la medición exacta y rápida de la concentración existente en el reparto de la masa total de una variable entre los elementos de un colectivo discreto.

En consecuencia, el IE no sólo es un medidor de concentración exacto(15), sino que, adicionalmente, presenta como ventaja tener un coste operativo mínimo (inferior, en todo caso, a las versiones del índice de Gini propuestas en la literatura estadística).

A modo de ejemplo, sea la siguiente distribución de frecuencias, a partir de la cual se genera la tabla de construcción del IE:

(15) Yitzhaki diría que el IE incorpora una corrección para distribuciones discretas con valores finitos. Vid. YITZHAKI, S. (1998), pág 23.

| $x_i$ | $n_i$ | $x_i\, n_i$ | $q_i$ | $Q_i$ | $prop_i$ | $M_i$ |
|---|---|---|---|---|---|---|
| 10 | 25 | 250 | 0,5236 | 0,5236 | 0,0209 | 6,8070 |
| 99 | 5 | 495 | 1,0468 | 1,5604 | 0,2073 | 5,7283 |
| 100 | 470 | 47000 | 98,4496 | 100 | 0,2094 | 23915,9074 |
| | | | | | **Total:** | 23928,4428 |

con $q_i = \frac{x_i\, n_i}{MTV} \cdot 100$ , donde

$$IE = \frac{N+1}{N-1} - \frac{\sum_{i=1}^{k} M_i}{50\,(N-1)} = \frac{501}{499} - \frac{23928{,}4428}{50 \cdot 499} = 0{,}045 \qquad [27]$$

mientras que si hubiésemos utilizado la expresión

$$IG = \frac{\sum_{i=1}^{k-1} (P_i - Q_i)}{\sum_{i=1}^{k-1} P_i} = 1 - \frac{\sum_{i=1}^{k-1} Q_i}{\sum_{i=1}^{k-1} P_i} \qquad [28]$$

que no recomendamos por ser la distribución no unitaria, el resultado hubiese sido 0,81. Es decir, en un caso claro de no concentración nos indica una fortísima concentración.

## 5. EL ÍNDICE E COMO ESTIMADOR INSESGADO DEL ÍNDICE DE GINI PARA DISTRIBUCIONES CONTINUAS

El Índice de Gini para distribuciones continuas se puede interpretar como la co-varianza entre la variable objeto de estudio (X) y su función de distribución F(x)(16):

(16) Vid. a estos efectos LERMAN, R; YITZHAKI, S. (1984). También YITZHAKI, S. (1998).

$$I_G = 2\frac{Cov\left(X;F(x)\right)}{E(X)} \qquad [29]$$

Pues bien, un estimador insesgado de la anterior expresión, a partir de una muestra de tamaño N que contiene k valores distintos de la variable X, es el siguiente:

$$\hat{I}_G = 2\frac{\frac{N}{N-1}\sum_{i=1}^{k}(x_i-\bar{x})(\hat{F}_i-\bar{F})\,p_i}{\bar{x}} \qquad [30]$$

donde, trabajando en tantos por uno,

$$p_i = \frac{n_i}{N}$$

$$\bar{x} = \sum_{i=1}^{k} x_i\ p_i$$

[31]

$$\hat{F}_i = \sum_{j=1}^{i-1} p_j + \frac{p_i}{2}$$

$$\bar{F} = \sum_{i=1}^{k}\hat{F}_i\ p_i = \sum_{i=1}^{k}\left(\sum_{j=1}^{i-1} p_j + \frac{p_i}{2}\right)p_i = \frac{1}{2}$$

por lo que el estimador insesgado del Índice de Gini no es sino la cuasicovarianza muestral entre X y F(x) dividida por la mitad de la media aritmética de X. Ahora bien, esta expresión no es otra que el denominado IE.

A continuación pasamos a demostrar lo expuesto, señalando previamente que, por motivos de comparación con los escritos sobre la cuestión, en lo que sigue $P_i$, $Q_i$, $p_i$ y $q_i$ están expresados en tantos por uno. Por tanto:

$$p_i = \frac{n_i}{N} \qquad P_i = \sum_{j=1}^{i} p_j = \sum_{j=1}^{i} \frac{n_j}{N}$$

$$q_i = \frac{x_i n_i}{MTV} \qquad Q_i = \sum_{j=1}^{i} q_j = \sum_{j=1}^{i} \frac{x_j n_j}{MTV} \qquad [32]$$

Hemos demostrado que

$$IG = \frac{\sum_{i=1}^{N-1} (P_i - Q_i)}{\sum_{i=1}^{N-1} P_i} = 1 - \frac{\sum_{i=1}^{N-1} Q_i}{50(N-1)} \qquad [33]$$

es un medidor exacto de la relación por cociente entre las áreas de concentración y de máxima concentración que, si se trabaja en tantos por uno, adopta la expresión

$$IG = \frac{\sum_{i=1}^{N-1} (P_i - Q_i)}{\sum_{i=1}^{N-1} P_i} = 1 - \frac{2\sum_{i=1}^{N-1} Q_i}{(N-1)} \qquad [34]$$

Por otra parte, tenemos que:

a) Su versión en términos de valores distintos de la variable es el denominado IE.

$$IE = \frac{N+1}{N-1} - \frac{\sum_{i=1}^{k} M_i}{50(N-1)} \qquad [35]$$

que, en tantos por uno, se escribiría como

$$IE = \frac{N+1}{N-1} - \frac{2\sum_{i=1}^{k} M_i}{(N-1)} \qquad [36]$$

b) El $I_G$ coincide con los medidores geométricos de concentración a través de trapecios(17).

$$I_G = \frac{N}{N-1}\left[1-\sum_{i=1}^{k}(P_i - P_{i-1})(Q_i + Q_{i+1})\right] = \frac{N}{N-1}\left[1-\left(\sum_{i=1}^{k} p_i \sum_{j=1}^{i-1} 2q_j + q_i\right)\right] \qquad [37]$$

c) En consecuencia, cuando se trabaja en tantos por uno

(17) La primera de las expresiones puede verse en MONTERO LORENZO, J.M. (2000), págs. 84 y 85. La segunda en FERREIRA, E.; GARÍN, A. (1997).

Resulta interesante, a efectos académicos, la siguiente demostración:

$$R = \frac{N}{N-1}\left[1-\sum_{i=1}^{N}(P_i - P_{i-1})(Q_i + Q_{i-1})\right] = \frac{N}{N-1}\left[1-\sum_{i=1}^{N}\frac{1}{N}(Q_i + Q_{i-1})\right]$$

$$= \frac{N}{N-1}\left[1-\frac{\sum_{i=1}^{N}(Q_i + Q_{i-1})}{N}\right] = \frac{N}{N-1}\left[1-\frac{\sum_{i=1}^{N}Q_i + \sum_{i=1}^{N}Q_{i-1}}{N}\right]$$

$$= \frac{N}{N-1}\left[1-\frac{\sum_{i=1}^{N-1}Q_i + 1 + \sum_{i=1}^{N-1}Q_i}{N}\right] = \frac{N}{N-1}\left[1-\frac{2\sum_{i=1}^{N-1}Q_i + 1}{N}\right]$$

$$= \frac{N}{N-1} - \frac{2\sum_{i=1}^{N-1}Q_i}{(N-1)} - \frac{1}{N-1} = 1 - \frac{2\sum_{i=1}^{N-1}Q_i}{(N-1)}$$

por lo que el índice de Gini coincide con el comúnmente denominado “índice geométrico”.

$$IE=\frac{N-1}{N+1}-\frac{2\sum_{i=1}^{k}M_i}{(N-1)}=\frac{N}{N-1}\left[1-\left(\sum_{i=1}^{k}p_i\sum_{j=1}^{i-1}2q_j+q_i\right)\right] \qquad [38]$$

Pues bien, expresando el numerador de la expresión [30] en función de los momentos bidimensionales respecto del origen se tiene que

$$\hat{I}_G=2\frac{\frac{N}{N-1}\sum_{i=1}^{k}(x_i-\bar{x})(\hat{F}_i-\bar{F})p_i}{\bar{x}}=2\frac{\frac{N}{N-1}\left(\sum_{i=1}^{k}x_i\hat{F}_ip_i-\bar{x}\bar{F}\right)}{\bar{x}}$$

$$=2\frac{\frac{N}{N-1}\left(\sum_{i=1}^{k}x_i\hat{F}_ip_i-\frac{\bar{x}}{2}\right)}{\bar{x}}=\frac{2}{\bar{x}}\frac{N}{N-1}\left(\sum_{i=1}^{k}x_i\hat{F}_ip_i-\frac{\bar{x}}{2}\right) \qquad [39]$$

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

y sustituyendo las estimaciones de $F_i$ por su valor se llega a

$$\frac{N}{N-1}\left[\left(\frac{2}{\bar{x}}\sum_{i=1}^{k} x_i \hat{F}_i p_i\right)-1\right]=\frac{N}{N-1}\left[\left(\frac{2}{\bar{x}}\sum_{i=1}^{k} x_i\, p_i\left(\sum_{j=1}^{i-1} p_j+\frac{p_i}{2}\right)\right)-1\right]=$$

$$\frac{N}{N-1}\left[\frac{2}{\bar{x}}\left(x_1 p_1\frac{p_1}{2}+x_2 p_2\left(p_1+\frac{p_2}{2}\right)+x_3\, p_3\left(p_1+p_2+\frac{p_3}{2}\right)+\ldots+x_k\, p_k\left(p_1+p_2+\ldots+\frac{p_k}{2}\right)\right)-1\right]=$$

$$\frac{N}{N-1}\left[2\left(\frac{x_1\frac{n_1}{N}}{\frac{MTV}{N}}\frac{p_1}{2}+\frac{x_2\frac{n_2}{N}}{\frac{MTV}{N}}\left(p_1+\frac{p_2}{2}\right)+\frac{x_3\frac{n_3}{N}}{\frac{MTV}{N}}\left(p_1+p_2+\frac{p_3}{2}\right)+\ldots+\frac{x_k\frac{n_k}{N}}{\frac{MTV}{N}}\left(p_1+p_2+\ldots+\frac{p_k}{2}\right)\right)-1\right]=$$

$$\frac{N}{N-1}\left[2\left[q_1\frac{p_1}{2}+q_2\left(p_1+\frac{p_2}{2}\right)+q_3\left(p_1+p_2+\frac{p_3}{2}\right)+\ldots+q_k\left(p_1+p_2\ldots+\frac{p_k}{2}\right)\right]-1\right]=$$

$$\frac{N}{N-1}\left[2\left[p_1\left(\frac{q_1}{2}+q_2+\ldots q_k\right)+p_2\left(\frac{q_2}{2}+q_3+\ldots q_k\right)+\ldots\ldots p_k\frac{q_k}{2}\right]-1\right]=$$

$$\frac{N}{N-1}\left[2\left(\sum_{i=1}^{k} p_i\left(\sum_{j=i+1}^{k} q_i+\frac{q_i}{2}\right)\right)-1\right]=$$

$$\frac{N}{N-1}\left[2\left(\sum_{i=1}^{k} p_i\left(1-\left(\sum_{j=1}^{i-1} q_j+\frac{q_i}{2}\right)\right)\right)-1\right]$$

$$\frac{N}{N-1}\left[2\sum_{i=1}^{k} p_i-2\sum_{i=1}^{k} p_i\left(\sum_{j=1}^{i-1} q_j+\frac{q_i}{2}\right)-1\right]=$$

$$\frac{N}{N-1}\left[2-2\sum_{i=1}^{k} p_i\left(\sum_{j=1}^{i-1} q_j+\frac{q_i}{2}\right)-1\right]=$$

$$\frac{N}{N-1}\left[1-\sum_{i=1}^{k} p_i\left(\sum_{j=1}^{i-1} 2q_j+q_i\right)\right]$$

que, por ser un medidor exacto de la relación entre las áreas de concentración y máxima concentración, coincide con el IE. En consecuencia,

$$IE = \frac{N+1}{N-1} - \frac{2\sum_{i=1}^{k} M_i}{N-1} = 2\,\frac{\frac{N}{N-1}\sum_{i=1}^{k}(x_i - \bar{x})(\hat{F}_i - \bar{F})\,p_i}{\bar{x}}$$

y, por consiguiente, constituye un estimador insesgado del Índice de Gini para distribuciones continuas.

## 6. CONCLUSIONES

A modo de resumen, en este trabajo se ha tratado de poner de manifiesto la exactitud del índice originalmente formulado por Corrado Gini para la medición de la concentración existente en el reparto de la masa total de una variable entre los elementos de un colectivo discreto (errores de agrupamiento aparte en distribuciones por intervalos), si bien sorprende que, según la versión española de 1953 de su Curso de Estadística (de amplia influencia en los estudiosos españoles de la cuestión), él mismo lo considerase una aproximación. Sin embargo, por venir expresada dicha formulación en términos del número de elementos del colectivo, su cálculo se hace muy tedioso si el tamaño de éste es relativamente elevado. Para reducir el coste operativo de dicho cálculo no nos parece una buena solución la utilización de la expresión aproximatoria [2] y hemos propuesto otra, denominada Índice E que, manteniendo la característica de la exactitud, tiene un coste operativo mínimo y constituye un estimador insesgado del Índice de Gini para distribuciones continuas.

## REFERENCIAS

BERREBI, S.M.; SILBER, J. (1985): «Income Inequality Indices and Deprivation: A Generalization», *Quarterly Journal of Economics*, 99, 807-810.

BERREBI, S.M.; SILBER, J. (1987): «Regional Diferences and The Components of Growth and Inequality Change», *Economics Letters*, 25, 295-298.

DAGUM, C. (1997): «A New Decomposition of the Gini Income Inequality Ratio», *Empirical Economics* 22, 515-531.

DONALDSON, D.; WEYMARK, J.A. (1980): «A Single Parameter, Generalization of the Gini Indices of Inequality», *Journal of Economic Theory*, 22, 67-87.

FERREIRA, E.; GARÍN, A. (1997): «Una nota sobre el cálculo del índice de Gini», *Estadística Española*, Vol. 39, Nº 142, 207-218.

GINI, C. (1912): «Variabilitá e mutabilitá, contributio allo studio delle distribuzioni e relazioni statistiche», *Estudi Economico-Giuridici dell'Universiti di Cagliari*, 3, parte 2, 1-158.

GINI, C. (1921): «Measurement of Inequality of Incomes». *The Economic Journal*, March, 124-126.

GINI, C. (1939): «Memorie di metodologia statistica: vol.I, Variabilitá e concentrazione». *Giuffrè*. Milan.

GINI, C. (1953): «Curso de Estadística. (2ª ed)». Traducción y adaptación de la edición italiana de 1946-47. *Editorial Labor S.A.* Barcelona.

KENDALL, M.G.; STUART, A. (1958): «The Advanced Theory of Statistics». Vol. I. *Charles Griffin & Company limited*. London.

LERMAN, R.I.; YITZHAKI, S. (1989): «A note on the calculation and interpretation of the Gini coefficient», *Economics Letters*, 15, 363-368.

LERMAN, R.I.; YITZHAKI, S. (1989): «Improving the accuracy of estimates of the Gini coefficient», *Journal of Econometrics*, 42, 43-47.

LORENZ, M.O. (1905): «Methods for Measuring Concentration of Wealth». *Journal of American Statistical Association*, 9, (New Series Nº 70), June, 209-219.

MILANOVIC, B. (1997): «A simple way to calculate the Gini coefficient and some implications», *Economic Letters*, 56, 45-49.

MONTERO LORENZO, J.M. (2000): «Estadística para Relaciones Laborales». *Alfa Centauro*. Madrid.

MONTERO LORENZO, J.M. (2002): «E-Index for Measuring Economic Concentration». *International Advances in Economic Research*, Vol. 08, Núm 4.

PYATT, G. (1976): «On the Interpretation and the Dissaggregation of Gini Coefficients», *The Economic Journal*, 86, June, 243-255.

SEN, A (1973): «On Economic Inequality». *Oxford University Press*. London.

SILBER, J. (1989): «Factor Components, Population Subgroups and the Computation of the Gini Index of Inequality», *The Review of Economics and Statistics*, Vol. LXXI, Nº 1, 107-115.

THEIL, H. (1966): «Economics and Information Theory». *North-Holland Publishing Company*. Amsterdam.

YITZHAKI, S. (1998): «More than a dozen ways of spelling Gini», *Research on Economic Inequality,* 8, 31-38.

## ON ECONOMIC CONCENTRATION: E-INDEX FOR DISCRETE COLLECTIVES


### SUMMARY

The primary aim of this article is to demonstrate the Gini index accuracy at the time of measuring concentration when we deal with discrete collectives (grouping errors not considered in interval distributions). In that way it is pointed out that new "exact indexes" for measuring concentration are not needed due to the fact that the Gini index verifies such a property. In addition, the article also reveals the inconvenience of using approximatory expressions of Gini index when applied to discrete collectives with non unitary frequencies.

What is the worthwile in the fields of concentration is to appraise new versions of the Gini index that allow us to reduce to a significant extent its operating cost, which in time is proportional to the elements in the collective. In that sense, this article presents the E index, that verifies a minimum opperating cost, being un unbiased estimator of the Gini index for continuous distributions.



*Key Words*: Gini index, E-index, concentration area, maximum concentration area, Lorenz curve.

*Clasificación AMS*: 6201