

EL COEFICIENTE GINI-INTERVALO; LA FORMA DE LA CONCENTRACIÓN A TRAVÉS DE UNA DESCOMPOSICIÓN DEL ÍNDICE DE GINI

FERNANDO CORTÉS Y ROSA MA. RUBALCAVA
El Colegio de México

1. Introducción

ES CONVENIENTE que antes de entrar de lleno al problema que nos interesa tratar en este artículo, dediquemos espacio a perfilar las diferencias centrales entre dos ideas que aparecen, con frecuencia, estrechamente relacionadas a la noción de desigualdad: nivel de la variable y forma de la desigualdad.

Desde el punto de vista estadístico, se entiende por “concentración” la manera como se reparte el total de una variable entre un conjunto de unidades individuales o agregadas. Se define en términos de la distribución de un monto total entre las n unidades. Es frecuente para referirse a esta noción usar como símil el modo como se distribuye un pastel entre el conjunto de comensales.

La estadística ha desarrollado una serie de medidas¹ que permiten evaluar el grado de desigualdad de una distribución de frecuencias; entre ellas se destaca el coeficiente de concentración de Gini.²

Al monto total a repartir lo denominaremos el *nivel de la variable*. La distinción entre el nivel y la manera como se reparte una variable tiene importancia porque remite a diferentes dominios del análisis estadístico los cuales entregan elementos clave para interpretar con precisión los resultados que arrojan las medidas estadísticas.

¹ Véase por ejemplo: Sen Amartya. “On economic inequality”; Clarendon Press, Oxford, 1973.

² Gini, Corrado “Curso de Estadística”; Editorial Labor, 1935, págs. 173 a 187.

El hecho de que al aplicar el coeficiente de Gini a la distribución de la variable (por ejemplo del ingreso personal) en dos momentos del tiempo entregue como resultado un mayor grado de desigualdad, no nos autoriza para realizar aseveraciones sobre los cambios experimentados por el monto de la variable en manos de los individuos o grupos, según sea el caso. Si un coeficiente estadístico de concentración marca grados mayores de desigualdad no debe interpretarse como que ahora los ricos tienen más que antes y los pobres menos, de manera que los primeros tienen a su disposición más dinero, los segundos menos y que en comparación con la situación original los ricos son más ricos y los pobres más pobres.

Estas derivaciones pueden reflejar o no lo acontecido en el lapso bajo consideración, pero lo que no corresponde es basarlas en la pura evidencia que nos proporciona el coeficiente de desigualdad. Para saber si un grupo cualquiera mejoró o empeoró respecto de su situación original, es necesario recurrir a la comparación de alguna medida de tendencia central.³ Tendríamos que ver, por ejemplo, si el ingreso medio del estrato más bajo es mayor o menor que antes, aun cuando resultaron perjudicados en el reparto. Un empeoramiento relativo, no necesariamente es contradictorio con una mejoría absoluta. Esta situación se podría presentar si en el período aumentó el tamaño del pastel y a ese estrato le correspondió una proporción del incremento menor que la que tenía originalmente.

Un coeficiente de concentración sólo nos dice cómo se repartió el total de una variable. Si deseamos conocer los cambios de las cantidades en manos de cada agregado (u observaciones individuales) debemos usar medidas de tendencia central.

Ahora bien, supongamos que el nivel de la variable se mantuvo constante y que la medición de la desigualdad arroja como resultado $I_1 = I_2$.⁴ ¿Qué podemos decir respecto a la fluctuación temporal de la concentración?

A primera vista pareciera que podríamos sostener que nada aconteció con las participaciones relativas.

Como el nivel de la variable se mantuvo y no cambió el grado de la desigualdad pareciera lógico concluir que la repartición de la variable es la misma en uno y otro momento en el tiempo. El mismo problema pero mirado desde otro ángulo nos lleva a plantear que dado un nivel constante de la variable, que no se modifiquen las propor-

³ El uso simultáneo de medidas de tendencia central y concentración puede verse en: Cortés, F. y Joclevsky, R., "La Distribución del Ingreso en el Gobierno de la Unidad Popular"; Demografía y Economía No. 43, México, 1981.

⁴ I: simboliza un coeficiente cualquiera de concentración.

ciones implica no sólo un nivel de desigualdad fijo, sino también que no se alterarán las cantidades absolutas.

Sin embargo, el hecho de que no cambien los niveles de la variable ni el grado de desigualdad existente en la distribución, no permite concluir que se mantuvieron las participaciones relativas y absolutas. El indicador del grado de desigualdad que nos dice que $I_1 = I_2$, entrega una medida resumen que se puede conformar de diversas maneras para arrojar un mismo resultado. Cada uno de los caminos para obtenerlo es producto de las *diversas formas que conducen a un mismo grado de concentración*. Esta situación es análoga a la que se presenta entre las medidas de posición y de dispersión en la estadística descriptiva. Podríamos pensar que el indicador I juega el papel de una medida de tendencia central y que a la *forma de la desigualdad* le corresponde un rol similar al de las medidas de dispersión.

Incorporar el análisis de la forma de la desigualdad puede resultar básico para disponer de una imagen relativamente acabada de los procesos que han afectado la manera como se distribuye una variable. Para ilustrar el papel que juega el estudio de la forma de la desigualdad, tomemos a manera de ejemplo, el caso en que nos interese seguir la pista de la evolución de la distribución del ingreso personal clasificado en seis categorías: obreros, empleados, trabajadores por cuenta propia, empleados domésticos, fuerzas armadas y empleadores. Además supóngase que se constata que la variable ingreso personal no ha sufrido cambios en su nivel y que a través de un coeficiente de concentración establecemos que $I_1 = I_2$. La situación descrita es perfectamente consistente con el hecho de que, por ejemplo, los obreros hayan mejorado su situación tanto relativa cuanto absoluta y que los empleados domésticos la hayan deteriorado desde ambos puntos de vista en el lapso bajo consideración. El valor del indicador en el tiempo 2 resultará igual al que asumió en el tiempo 1, debido a que los dos movimientos de sentido contrario se cancelaron dejando inalterado el valor global de la medida.

Si nos quedamos sólo en el nivel de la información que nos proporciona el coeficiente de concentración, únicamente estaremos en condiciones de afirmar que el grado de desigualdad no ha variado en el período. Formularse interrogantes respecto a la forma de la concentración permite dirigir nuestra atención hacia las participaciones relativas, cuestión que puede ser de importancia en algunos casos concretos. Supongamos que en nuestro ejemplo las medidas económicas se orientan hacia una redistribución del ingreso que permita sentar bases para obtener el apoyo político de las clases desposeídas de la sociedad y que al observar que $I_2 = I_1$ se concluye que la polí-

tica ha sido inefectiva. Pero hemos visto que a pesar de este resultado puede haber tenido éxito en la medida que ha beneficiado al sector obrero. Lo probablemente inconsistente, sería que forme parte del mismo propósito perjudicar a los trabajadores domésticos. Estos cambios pueden o no estar de acuerdo con los objetivos políticos propuestos, hecho que no es motivo de preocupación en este trabajo. Lo que interesa destacar es que a partir de la noción *forma de la desigualdad* surgen preguntas que dirigen el análisis a una dimensión del problema que no se agota en el simple estudio de la variación experimentada por el grado de desigualdad.

Este trabajo está dedicado a presentar una descomposición del índice de Gini para datos agrupados cuya utilidad radica en que constituye un *procedimiento numérico* que permite analizar la *forma en sí y en relación al grado de concentración*.

El valor del coeficiente de Gini expresa, en un intervalo limitado por 0 y 1, el grado de desigualdad en la repartición de una variable. La curva de Lorenz permite apreciar *geoméricamente* la forma de la concentración. El coeficiente Gini-intervalo es un instrumento que nos capacita para analizar *numéricamente la forma de la concentración* y a través de su relación con Gini global *estudiar el vínculo entre la forma y el grado de desigualdad*.

2. El coeficiente Gini-intervalo

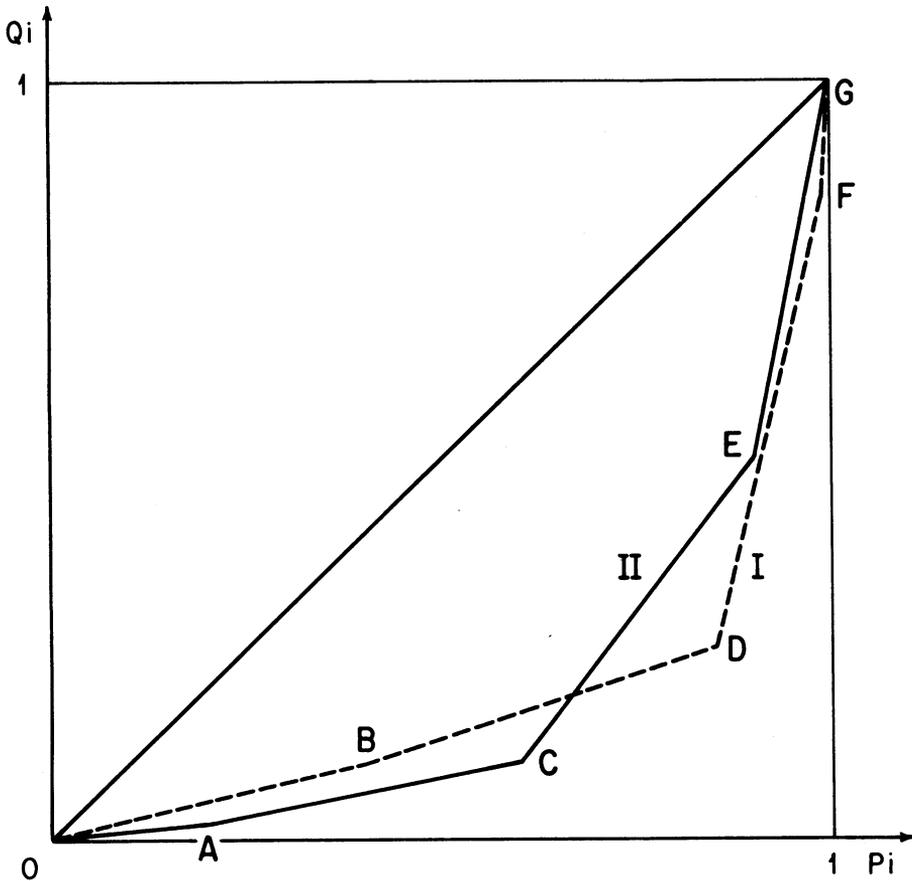
En la introducción hemos visto que dos o más conjuntos de datos pueden presentar el mismo nivel de desigualdad global lo que no significa que la forma de la concentración coincida en ambos. En términos gráficos este hecho se expresa en distribuciones que originan dos poligonales (curvas) de Lorenz distintas que delimitan áreas de concentración idénticas. A pesar de la diferencia en la forma de la desigualdad los índices de Gini asumirán el mismo valor.

El problema que nos interesa consiste en buscar procedimientos numéricos que nos permitan estudiar la forma de la desigualdad. La construcción de un diagrama de concentración es el camino inmediato para percatarse de la manera como se constituye el nivel de desigualdad global.

El área delimitada por la figura OACEG es igual a la que se encuentra en el interior de la OBDFG, de manera que el coeficiente de Gini será el mismo para la distribución I que para la II. Ahora bien, el aporte a la desigualdad de los tramos bajos de II es mayor que el que hacen las categorías bajas de I y como el área total es igual en ambos casos, debe cumplirse que la contribución de los tramos superiores de I debe ser mayor que la que hacen los superiores de II. Es

Gráfica 1

FORMA DE CONCENTRACIÓN



decir, en las distribuciones representadas por la gráfica 1 vemos que es distinta la forma que conduce a un mismo nivel global de desigualdad.

¿Qué significan estas formas distintas? *Grosso modo* se puede interpretar la situación representada en la figura como que en la distribución I las unidades más pobres (vale decir, a las que les corresponde una menor proporción de la variable) reciben relativamente más que las más desfavorecidas por el reparto que refleja la distribución II y que las más ricas en I tienen una participación en el total, menor que en II. El mismo nivel de desigualdad se conforma en el caso de la distribución I por un “perjuicio” menor que en II de las

unidades pobres y un “beneficio” también menor para las unidades ricas.

Lo anterior es sólo la resultante del hecho de que se debe repartir un pastel de tamaño dado y en tanto a algunas unidades les corresponda una tajada grande, otras necesariamente verán la suya disminuida. Por ello si las observaciones más pobres de una distribución son menos perjudicadas que las más pobres de otra, se sigue inmediatamente que las más ricas no serán tan beneficiadas como las más favorecidas de la otra distribución.

Aun cuando baste construir el diagrama de concentración para distinguir la forma de la desigualdad, resulta conveniente abordar su estudio desde el punto de vista numérico.

Para ello aplicamos a cada segmento de la poligonal de Lorenz, la idea básica sobre la cual se construye el índice de Gini. En lugar de tomar toda el área de concentración (A) la descomponemos en una serie de subáreas (A_1, A_2, \dots, A_m) tales que sumadas conformen

el área total ($\sum_{i=1}^m A_i = A$). A partir de cada una de las A_i se propone

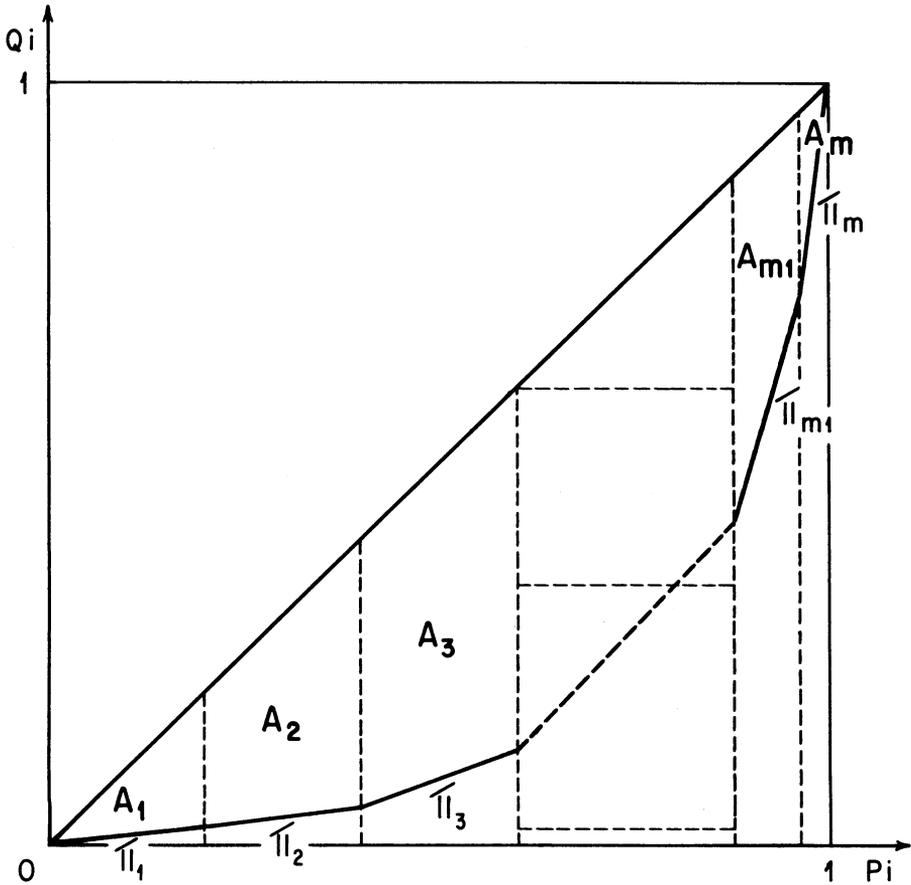
un coeficiente de concentración que nos ayuda en el estudio de la forma de la desigualdad. Para ello establecemos la relación entre cada A_i y su correspondiente subárea de concentración máxima \mathfrak{A}_i :

$$(1) \quad G_i = \frac{A_i}{\mathfrak{A}_i} \quad (i = 1, 2, \dots, m)$$

obteniéndose así el coeficiente de desigualdad Gini-intervalo (G_i).

Cada coeficiente G_i puede asumir valores entre cero y uno, ambos extremos incluidos. El valor cero se obtiene cuando $A_i = 0$, es decir, si el segmento de la poligonal de Lorenz se confunde con la línea de equidistribución. G_i será igual a 1 cuando $A_i = \mathfrak{A}_i$, es decir, cuando el segmento de la poligonal de Lorenz coincide con el eje de las abscisas. Hay que notar que como las observaciones, o los intervalos de clase, están ordenadas (en nuestro caso de menor a mayor), basta que $G_1 = 0$ para que todos los restantes G_i sean también iguales a cero. Esto quiere decir que si $G_1 = 0$, entonces $G_2 = G_3 = \dots = G_m = 0$. Ello es consecuencia de que si los más perjudicados (los que se encuentran en el primer intervalo) tienen una participación en el total de la variable igual a la que les correspondería por la aplicación de la norma democrática, entonces no habrá categorías estadísticas beneficiadas.

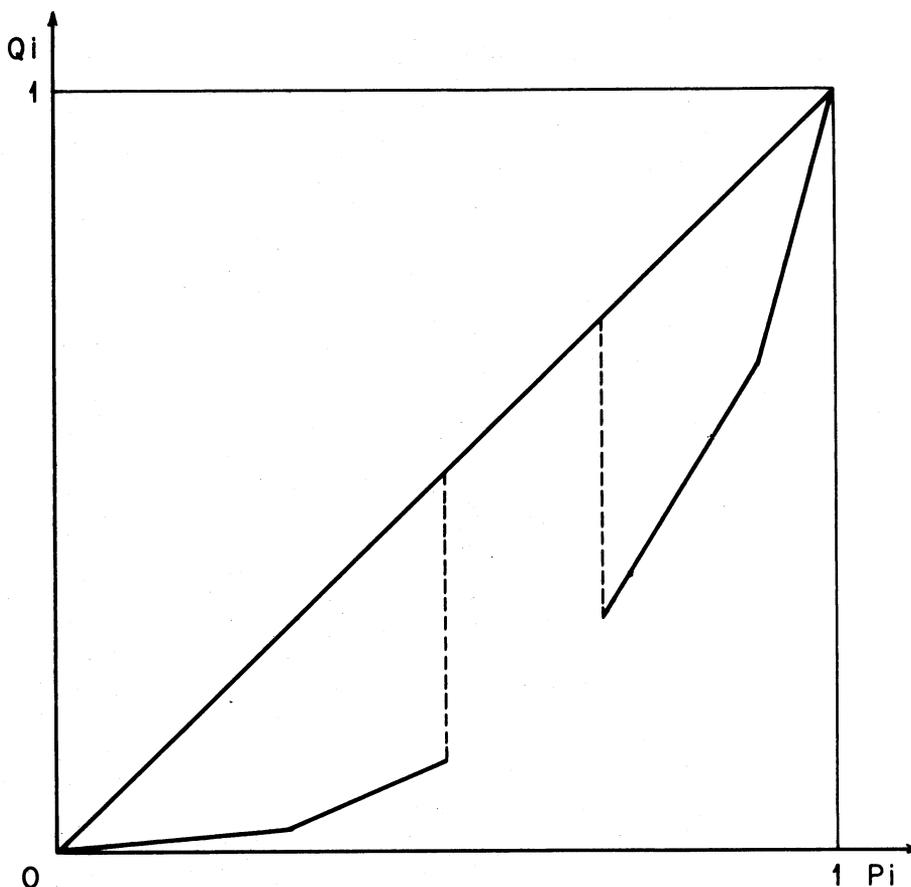
Gráfica 2
COEFICIENTE GINI-INTERVALO



Por otra parte, el hecho de que si a un intervalo de clase le corresponde menos que la norma democrática, a otro le tendrá que corresponder necesariamente más, unido a la ordenación de menor a mayor participación en la distribución de la variable, implica que entre dos Gini-intervalo distintos de cero ninguno de los intermedios será nulo.

Esta condición significa que si G_1 es distinto de cero, ningún segmento de la poligonal coincidirá con la línea de equidistribución. Lo anterior implica que en términos gráficos, no se puede presentar una situación como la siguiente.

Gráfica 3



El orden de las observaciones conjuntamente con el hecho de que en la repartición de un total hay perjudicados y beneficiados, se reflejan en que las pendientes de la poligonal (o curva) de Lorenz serán siempre crecientes. En otros términos, en la medida que nos movemos de izquierda a derecha en el diagrama de concentración las razones de ventaja⁵ serán siempre crecientes.

Sabemos que hay un punto en que la razón de ventaja es igual a la unidad, y éste divide a las observaciones en dos regiones: una en

⁵ Geométricamente las razones de ventaja son las tangentes de los segmentos de la poligonal de Lorenz y se interpretan como el porcentaje de la variable que corresponde a cada uno por ciento de las observaciones. Para más detalles véase Alker Hayward, *El uso de las matemáticas en el análisis político*, Amorrortu editores, 1969, págs. 56 y 57.

que se encuentran todas las unidades perjudicadas por la repartición y otra en que se ubican las beneficiadas. Al ordenar los intervalos de clase de menor a mayor, la región a la izquierda del punto contendrá a los estratos perjudicados y la región de la derecha a los beneficiados. En otros términos a la derecha de la razón de ventaja unitaria, las pendientes serán siempre mayores que uno y en la medida que nos movemos de ese punto hacia la derecha, las razones de ventaja (y por lo tanto sus equivalentes las pendientes) serán cada vez mayores que uno y cuando nos alejamos hacia la izquierda serán cada vez menores. En consecuencia, en una gráfica de concentración, las pendientes de la curva de Lorenz necesariamente serán crecientes.

En el apéndice se demuestra que al aplicar la definición de

$$G_i = \frac{A_i}{P_i} \quad (i = 1, 2, \dots, m)$$

a la gráfica de concentración y operar algebraicamente se llega al coeficiente Gini-intervalo:

$$(2) \quad G_i = 1 - \frac{Q_i + Q_{i-1}}{P_i + P_{i-1}} \quad (i = 1, 2, \dots, m)$$

Si $Q_i = Q_{i-1} = 0$, es decir, si la participación relativa acumulada de los intervalos i e $(i-1)$ es nula, entonces de acuerdo con (2) $G_i = 1$; estaríamos en presencia del máximo de concentración posible. Tal vez sería conveniente señalar que como las observaciones se han ordenado de menor a mayor, entonces también se debe cumplir que $Q_{i-k} = 0$, en que $k = 2, \dots, (i-1)$.

Si por otra parte, $Q_i + Q_{i-1} = P_i + P_{i-1}$, entonces conforme a (2) $G_i = 0$. Esta condición se cumple únicamente sobre la línea de equidistribución, ya que:

(i) Como consecuencia de la ordenación de las observaciones:

$$P_{i-1} \geq Q_{i-1}$$

$$P_i \geq Q_i$$

Al sumar término a término estas dos inecuaciones se llega a:

$$P_i + P_{i-1} \geq Q_i + Q_{i-1}$$

y por tanto,

$$(P_i + P_{i-1}) - (Q_i + Q_{i-1}) \geq 0$$

Que en el caso estudiado debe reducirse sólo a la igualdad,

$$(P_i + P_{i-1}) - (Q_i + Q_{i-1}) = 0$$

(ii) Además, por tratarse de frecuencias relativas acumuladas se cumplirá siempre que,

$$P_i \geq P_{i-1}$$

$$Q_i \geq Q_{i-1}$$

Las dos últimas desigualdades y la igualdad del inciso anterior generan un sistema que sólo se satisface en el caso que:

$$P_{i-1} = Q_{i-1}$$

y

$$P_i = Q_i$$

en cualquier otra situación entra en contradicción. Este resultado se encuentra dentro de lo esperado, en la medida que el índice se ha definido como la relación entre las áreas A_i y $\mathbb{1}_i$ el valor unitario sólo ocurre cuando ambas son iguales, es decir, cuando el tramo de poligonal correspondiente al intervalo coincide con el eje de abscisas.

Aun cuando desde el punto de vista aritmético cada G_i puede tomar valores entre 0 y 1 (ambos extremos incluidos) sabemos que si G_1 es igual a cero entonces los restantes Gini-intervalo necesariamente serán también nulos. Además vimos que por la jerarquización de las observaciones involucradas en Gini, ningún G_i puede tomar el valor cero si alguno anterior o posterior es distinto de cero. Estas características tienen una explicación general en el hecho de que la serie de los coeficientes Gini-intervalo para una distribución dada es decreciente.

Para examinar la secuencia de los coeficientes, consideremos la expresión:

$$Q_1 = a P_1$$

Que representa la ecuación de la recta con pendiente a , que une el origen con el punto de coordenadas (P_1, Q_1) . De acuerdo con (2), G_1 será:

$$G_1 = 1 - \frac{Q_1}{P_1}$$

y al remplazar Q_1 obtendremos:

$$G_1 = 1 - a$$

es decir, el coeficiente Gini para el primer intervalo es igual al complemento a uno de la pendiente.

Como las inclinaciones de los segmentos que conforman la curva de Lorenz son no-decrecientes, entonces la ecuación de la recta que une el origen con el punto de coordenadas (P_2, Q_2) se puede expresar como:

$$Q_2 = (a + \epsilon_2)P_2$$

Donde ϵ_2 es una cantidad no-negativa que surge de la diferencia entre las inclinaciones de las rectas que pasan por el origen y los dos puntos ubicados sobre la poligonal de Lorenz.

Al aplicar (2) para el caso en que i es igual a 2 tenemos:

$$G_2 = 1 - \frac{Q_2 + Q_1}{P_2 + P_1}$$

Al remplazar en esta igualdad la relación que establecimos entre Q_2 y P_2 se llega a:

$$G_2 = 1 - a - \frac{\epsilon_2}{1 + P_1/P_2}$$

Como $G_1 = 1 - a$, entonces

$$G_2 = G_1 - \frac{\epsilon_2}{1 + P_1/P_2}$$

Si ϵ_2 fuese cero, es decir si el punto (P_2, Q_2) estuviese sobre la recta que une el origen con el punto (P_1, Q_1) , entonces G_2 sería igual a G_1 . En cualquier otro caso G_2 será menor que G_1 .

Del mismo modo, si representamos la recta que une al origen con el punto de coordenadas (P_3, Q_3) por:

$$Q_3 = (a + \epsilon_2 + \epsilon_3) P_3$$

donde ϵ_3 se define de manera análoga a ϵ_2 y refleja la diferencia entre las inclinaciones del segundo y tercer segmentos de la poligonal. La sustitución de esta ecuación en la fórmula de G_3 conduce a:

$$G_3 = 1 - a - \epsilon_2 - \frac{\epsilon_3}{1 + P_2/P_3}$$

Si $\epsilon_2 = \epsilon_3 = 0$, entonces $G_1 = G_2 = G_3$

El análisis de esta expresión⁶ permite afirmar que G_3 es menor que G_2 . En consecuencia tenemos que:

$$G_1 \geq G_2 \geq G_3$$

En general, para el i -ésimo segmento de la poligonal tendremos que la recta que une al origen con (P_i, Q_i) responderá a:

$$Q_i = (a + \epsilon_2 + \epsilon_3 + \dots + \epsilon_i) P_i$$

y la sustitución originará:

$$G_i = 1 - a - \sum_{j=2}^{i-1} \epsilon_j - \frac{\epsilon_i}{1 + \frac{P_{i-1}}{P_i}}$$

El estudio⁷ de esta igualdad permite afirmar que:

$$G_1 \geq G_2 \geq G_3 \geq \dots \geq G_i$$

⁶ La forma de analizarla se presentará más adelante.

⁷ Para analizar la relación entre dos Gini-intervalo consecutivos considérense las expresiones generales:

$$G_{i-1} = 1 - a - \sum_{j=2}^{i-2} \epsilon_j - \epsilon_{i-1} \frac{P_{i-1}}{P_{i-1} + P_{i-2}}$$

y

$$G_i = 1 - a - \sum_{j=2}^{i-1} \epsilon_j - \epsilon_i \frac{P_i}{P_i + P_{i-1}}$$

Nótese que si hay m estratos y se cumple que los primeros $(m-1)$ Gini-intervalo son iguales entre sí y no nulos, el último necesariamente tendrá que ser menor.

Al comparar los valores de G_i en dos o más distribuciones para un mismo intervalo i , tendremos un indicador numérico de la diferencia en la forma de la concentración. Pero antes de seguir el análisis en esta dirección nos preocuparemos por indagar respecto a la vinculación entre los valores G_i y el del G global. Para determinarla nos basaremos en la definición geométrica de Gini: relación entre áreas.

En el apéndice se presenta el desarrollo de esta idea y se demuestra que al ponderar cada G_i por $\frac{A_i}{A}$, es decir, por la importancia relativa de cada subárea de máxima concentración con respecto al área de máxima desigualdad se llega a:

$$(3) \quad G = G_1 \frac{A_1}{A} + G_2 \frac{A_2}{A} + \dots + G_m \frac{A_m}{A}$$

como

$$G_i = \frac{A_i}{A} \quad (i = 1, 2, \dots, m)$$

entonces,

$$G = \frac{A_1}{A} \times \frac{A_1}{A} + \frac{A_2}{A} \times \frac{A_2}{A} + \dots + \frac{A_m}{A} \times \frac{A_m}{A}$$

Esta última expresión se puede escribir como:

$$G_i = \left[1 - a \sum_{j=2}^{i-2} \epsilon_j - (\epsilon_{i-1}) \frac{P_{i-1}}{P_{i-1} + P_{i-2}} \right] + (\epsilon_{i-1}) \frac{P_{i-1}}{P_{i-1} + P_{i-2}} - \epsilon_{i-1} - (\epsilon_i) \frac{P_i}{P_i + P_{i-1}}$$

Los términos entre paréntesis son iguales a G_{i-1} , por lo tanto:

$$G_i = G_{i-1} - \epsilon_{i-1} \left(1 - \frac{P_{i-1}}{P_{i-1} + P_{i-2}} \right) - \epsilon_i \left(\frac{P_i}{P_i + P_{i-1}} \right)$$

Como los términos que multiplican a ϵ_{i-1} y ϵ_i son menores que uno pero mayores que cero, se sigue que:

$$G_i \leq G_{i-1}$$

La igualdad sólo se cumple cuando $\epsilon_{i-1} = \epsilon_i = 0$.

Al simplificar:

$$G = \frac{A_1 + A_2 + \dots + A_m}{\mathcal{A}}$$

y por tanto

$$G = \frac{A}{\mathcal{A}}$$

Esta expresión demuestra que el sistema de ponderaciones aplicado sobre los coeficientes Gini-intervalo permite obtener el valor de G global.

Los desarrollos geométricos y algebraicos realizados en el apéndice llevan a establecer que se cumple:

$$\frac{\mathcal{A}_i}{\mathcal{A}} = (P_i - P_{i-1}) (P_i + P_{i-1})$$

en consecuencia:

$$G = \sum_{i=1}^m G_i (P_i - P_{i-1}) (P_i + P_{i-1})$$

Como

$$P_i - P_{i-1} = P_i$$

entonces esta igualdad puede escribirse de la siguiente manera:

$$(4) \quad G = \sum_{i=1}^m G_i P_i (P_i + P_{i-1}) = \sum_{i=1}^m G_i \frac{\mathcal{A}_i}{\mathcal{A}}$$

Este resultado nos indica que el coeficiente de concentración de Gini es un promedio ponderado de los coeficientes Gini-intervalo en que los pesos son las correspondientes proporciones entre la subárea y el área de máxima concentración.

Pareciera que cada término de la ecuación (3) sólo incorpora información de un intervalo de clase y si así fuera, los sumandos reflejarían la contribución específica de los estratos a la conformación del nivel de desigualdad global. Entonces, $G_i \frac{\mathcal{A}_i}{\mathcal{A}}$ mediría la cuota con que concurre el primer tramo a la formación de G. La del segundo in-

tervalo se evaluaría a través de $G_2 \frac{q_2}{q_1}$ y así sucesivamente. Sin embargo, esta interpretación es lícita sólo en apariencia, puesto que tanto los G_i como sus correspondientes ponderaciones dependen de las frecuencias acumuladas P_i , P_{i-1} , Q_i y Q_{i-1} que por definición contienen los p_i y q_i referidos a los estratos del primero al i -ésimo (o según el caso, hasta el $(i-1)$ -ésimo).⁸ Como el valor de cada término del G global depende de la información no sólo de la clase que le corresponde sino también de todas las anteriores, se concluye que no se debe interpretar como la aportación propia de ese estrato.

También hay que notar que el coeficiente Gini-intervalo ejerce una especie de control por el tamaño de los intervalos de clase. En efecto, al aumentar el tamaño de los estratos naturalmente tienden también a acrecentarse (entre otras cosas) las frecuencias relativas. Si incrementamos el intervalo i lo mismo ocurrirá con $(P_i - P_{i-1})$, es decir, con la distancia a que se encuentran las frecuencias relativas acumuladas de las unidades observadas (ver gráfica 4). Una consecuencia directa del aumento del tamaño del intervalo⁹ es el crecimiento del área de máxima concentración: pasamos del trapecoide BACD al BAEF. Pero al mismo tiempo aumenta el área de concentración desde BKND al BKMF de modo que al definir el coeficiente como una relación entre la subárea de concentración y la de máxima desigualdad se tienden a compensar los incrementos: el aumento en el tamaño de un intervalo de clase se refleja tanto en el numerador como en el denominador G_i .

A partir de esta gráfica es posible apreciar que la unión de los intervalos ocasiona una caída en el valor del G global que se explica

⁸ Nótese que G_i puede escribirse como:

$$G_i = 1 - \frac{2 \sum_{j=1}^{i-1} q_j + q_i}{2 \sum_{j=1}^{i-1} p_j + p_i}$$

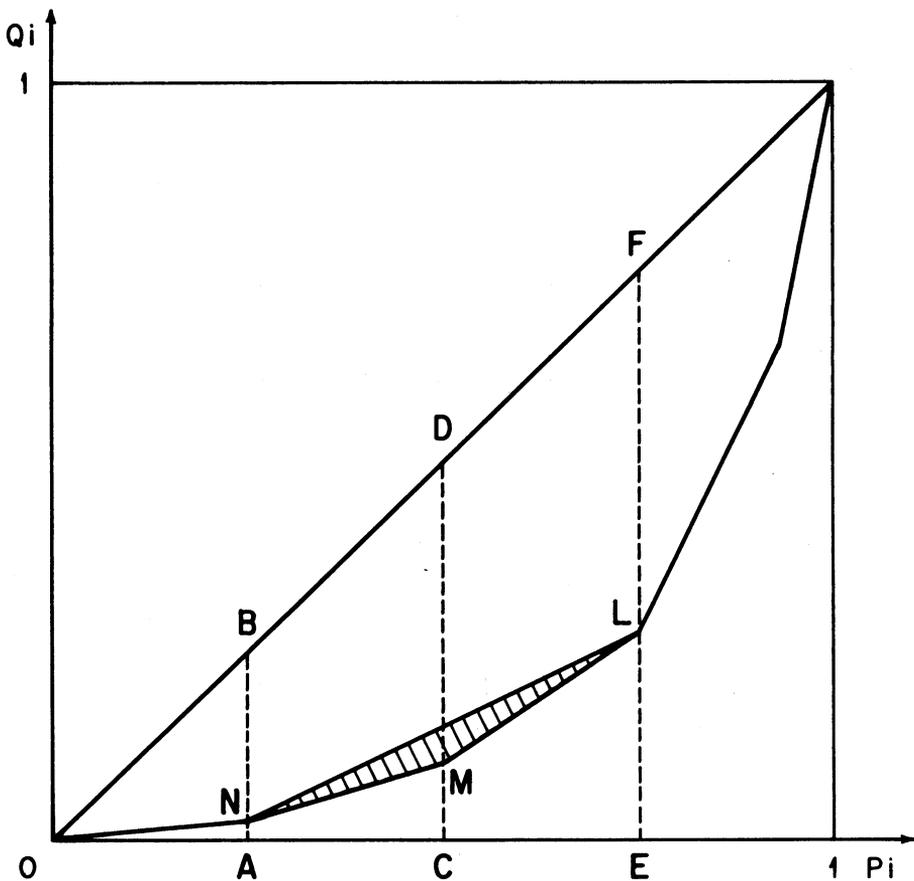
y el peso como $q_i / q_1 = p_i (2 \sum_{j=1}^{i-1} p_j + p_i)$

⁹ Para simplificar hemos puesto como ejemplo un caso en que el cambio en el tamaño del intervalo resulta de la unión de dos consecutivos.

por la eliminación de la parte KNM del área de desigualdad. Esta disminución también puede analizarse a través de (4) que descompone el valor de G en función de los Gini-intervalo. Su variación se deberá básicamente a las alteraciones que experimentan los pesos, puesto que como sabemos, los valores de G_i tenderán a no modificarse.

Gráfica 4

TAMAÑO DE LAS CLASES Y EL VALOR DEL COEFICIENTE
GINI-INTERVALO



Para usar el coeficiente Gini-intervalo en el estudio de la forma de la concentración consideraremos dos distribuciones de frecuencias, lo que nos lleva directamente a preguntarnos por el conjunto mínimo

de requisitos que permiten la comparación. El análisis que se desea puede llevar a establecer el paralelo entre dos distribuciones de una misma variable ubicadas en diferentes puntos del tiempo o a buscar las semejanzas en la forma como se reparten dos variables distintas entre el conjunto de unidades. En este último caso el interés radica casi siempre en determinar a través de las formas, la relación entre las variables. Pero si así fuera, sería más conveniente recurrir al análisis de regresión o de correlación. Por ello nuestra atención se depositará en el cambio que experimenta la distribución de una variable en el tiempo.

La fórmula 4 nos dice que para comparar el valor de dos Gini-intervalo es necesario que ambos tengan el mismo número de sumandos, es decir, que el total de estratos sea igual. Por lo tanto, si decidimos agregar dos o más clases en una de las distribuciones, una operación similar deberá realizarse en la otra a fin de que no se viole este principio. En la gráfica 4 se presentó el efecto que tiene la agregación de dos categorías, sobre la forma de la poligonal. En consecuencia, podríamos alterarla a voluntad a través de la manipulación de este recurso.

Al contrastar la desigualdad en dos distribuciones de frecuencias con diferente forma y el mismo número de intervalos hay que vigilar que las proporciones de las observaciones (p_i) sean las mismas, porque en caso contrario los segmentos de las poligonales tendrían diferente longitud (y la forma de la desigualdad cambiaría de inmediato).

Ahora bien, tomemos un ejemplo numérico para ilustrar los conceptos y desarrollos precedentes. Consideremos que disponemos de dos distribuciones de frecuencias que se caracterizan por presentar el mismo valor del coeficiente de Gini global, pero que difieren en cuanto a la forma de la concentración (véase cuadro 1).

Como en este ejemplo hemos controlado por los tamaños de los grupos (los p_i son los mismos en las dos distribuciones) entonces las contribuciones relativas a la concentración total reflejan directamente los impactos de los G_i , es decir, de *la forma* que asume la concentración en ambos casos.

Al comparar los G_i del cuadro vemos que el aporte a la desigualdad que realiza el primer estrato es mayor en la distribución II que en la I. Es decir, aun cuando el nivel de desigualdad global es el mismo, las unidades más desposeídas han resultado más perjudicadas en el reparto del pastel reflejada por la distribución de frecuencias II que en la implícita en I.

Por la naturaleza de las medidas de desigualdad los valores de G_i en los tramos altos tenderán a ser menores en II que en I. Esta es la

Como era de esperarse, acorde con los desarrollos que hemos presentado, al agrupar el segundo y tercer intervalos disminuyó el valor del coeficiente. Ya sabíamos que en la medida que mayor es el nivel de agregación, menor tiende a ser el valor de G . Puesto en otros términos, mientras más grandes son los intervalos de clase, mayor es el tamaño de la “subestimación” con respecto al índice de Gini calculado sobre la base de los datos originales.

Examinemos ahora los G_i antes y después de la agregación de categorías. Hay que notar que en el cuadro 2, $G_2 = 0.375$, mientras que en el caso I de la 1, $G_2 = 0.500$ y $G_3 = 0.367$. Sin embargo, tal como se había establecido el G_2 de la tabla modificada no aumenta en la misma proporción que el tamaño del nuevo estrato, aún más, aparece un valor intermedio entre G_2 y G_3 pero más cercano a G_3 .

Este resultado nos muestra numéricamente que el coeficiente Gini-intervalo es poco sensible a las variaciones en el tamaño de los intervalos de clase y que el G_i resultante de la agregación de dos o más estratos no es igual a un promedio de los G_i de cada uno de ellos. Esta última característica se encontraba implícita cuando examinábamos las consecuencias de la agregación en la gráfica 4 y concluíamos que se perdía la parte KNM del área de concentración.

Como conclusión podemos decir que para estudiar la forma en que se constituye el nivel de desigualdad es posible recurrir al coeficiente Gini-intervalo. Si queremos comparar la *forma* de la concentración entre dos o más distribuciones será de utilidad examinar tanto las poligonales de Lorenz cuanto los G_i .

3. Conclusiones

En este trabajo planteamos que cada coeficiente Gini-intervalo (G_i) puede ser útil para estudiar numéricamente la forma de la concentración. En este sentido es análogo a la curva de Lorenz.

Al aplicarle un sistema de ponderaciones hemos logrado reconstituir el valor del índice global (G), por lo que estamos en condiciones de evaluar los tamaños relativos de cada uno de sus componentes. Al observar la descomposición pareciera, a primera vista, que los sumandos serían indicadores de la contribución de cada intervalo a la desigualdad general. Sin embargo hemos mostrado que esta interpretación sería errónea por cuanto los valores de G_i y de sus ponderaciones son función de las frecuencias acumuladas. Habría que investigar la posibilidad de obtener una descomposición de G con tales características.

Tal vez la mayor utilidad que puede obtenerse al separar los elementos de G , radique en su aplicación al análisis del cambio en la forma de la desigualdad. La comparación de los valores correspondientes de G_i en dos distribuciones con la misma proporción de observaciones por intervalo, permite detectar los puntos que originaron el cambio de forma.

APENDICE

Gini-intervalo

El centro de nuestro interés radica en obtener una expresión matemática que nos permita estudiar numéricamente la forma de la desigualdad y reconstituir el índice de Gini. Para alcanzar este propósito se recurre a la idea básica que da sustento y origina este coeficiente: relacionar el área de concentración con la de máxima desigualdad. Que adecuada a cada intervalo asumiría la forma: relacionar la sub-área de concentración (A_i) con la de máxima desigualdad en el intervalo (véase gráfica 5).

La expresión geométrica de esta idea nos lleva a definir el coeficiente Gini-intervalo para la i -ésima clase como:

$$G_i = \frac{A_i}{\mathcal{A}_i}$$

Es relativamente simple calcular las áreas ($\mathcal{A}_i - A_i$) que se encuentran delimitadas por el eje de las abscisas y por la curva de Lorenz. En efecto, tenemos que:

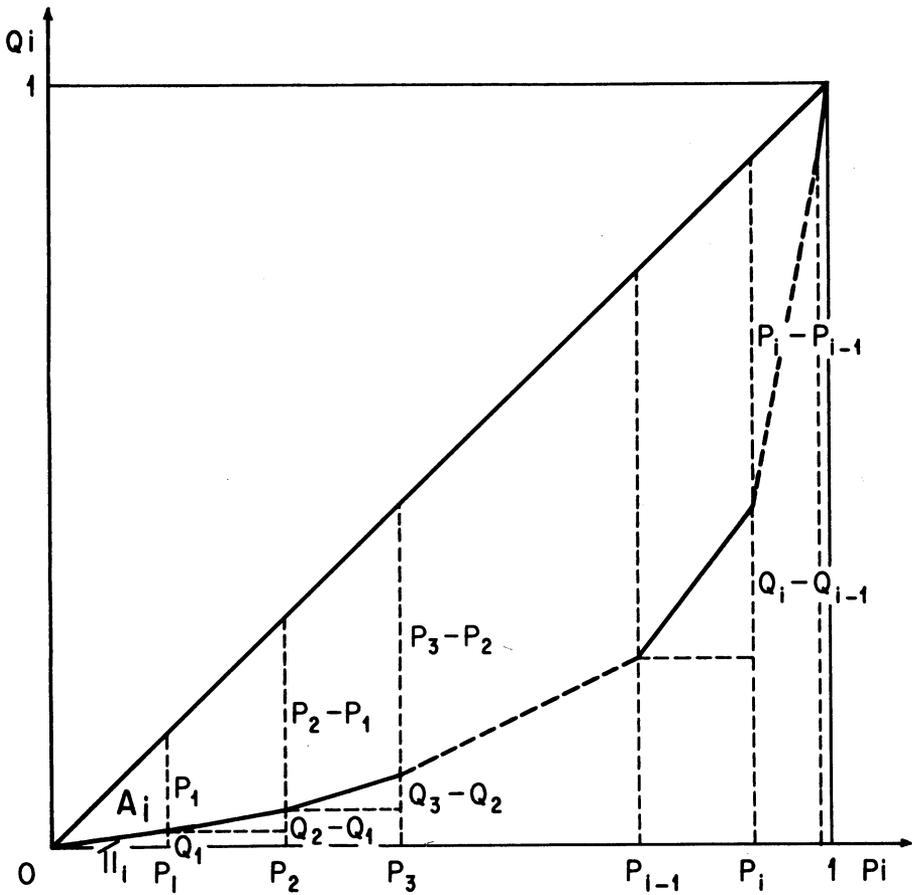
$$(\mathcal{A}_1 - A_1) = \frac{P_1 Q_1}{2}$$

$$(\mathcal{A}_2 - A_2) = (P_2 - P_1) (Q_1) + \frac{1}{2}(P_2 - P_1) (Q_2 - Q_1) =$$

$$\frac{(P_2 - P_1) (Q_1 + Q_2)}{2} = \frac{p_2 (Q_1 + Q_2)}{2}$$

$$(\mathcal{A}_3 - A_3) = (P_3 - P_2) Q_2 + \frac{(P_3 - P_2) (Q_3 - Q_2)}{2} =$$

Gráfica 5



$$\frac{(P_3 - P_2)(Q_3 + Q_2)}{2} = \frac{p_2(Q_3 + Q_2)}{2}$$

.....

$$(A_i - A_{i-1}) = (P_i - P_{i-1}) Q_{i-1} + \frac{(P_i - P_{i-1})(Q_i - Q_{i-1})}{2} =$$

$$\frac{(P_i - P_{i-1})(Q_i + Q_{i-1})}{2} = \frac{p_i(Q_i + Q_{i-1})}{2}$$

Por otra parte, los valores de las subáreas de máxima concentración simbolizadas por las letras \mathbb{N} serán iguales a:

$$\mathbb{N}_1 = \frac{P_1 P_1}{2}$$

$$\mathbb{N}_2 = (P_2 - P_1) P_1 + \frac{(P_2 - P_1)(P_2 - P_1)}{2} = \frac{(P_2 - P_1)(P_2 + P_1)}{2} =$$

$$\frac{p_2 (P_2 + P_1)}{2}$$

$$\mathbb{N}_3 = (P_3 - P_2) P_2 + \frac{(P_3 - P_2)(P_3 - P_2)}{2} = \frac{(P_3 - P_2)(P_3 + P_2)}{2} =$$

$$\frac{p_3 (P_3 + P_2)}{2}$$

.....

$$\mathbb{N}_i = (P_i - P_{i-1}) P_{i-1} + \frac{(P_i - P_{i-1})(P_i - P_{i-1})}{2} = \frac{(P_i - P_{i-1})(P_i + P_{i-1})}{2} =$$

$$p_i \frac{(P_i + P_{i-1})}{2}$$

A partir de estos dos conjuntos de expresiones se puede obtener una expresión aritmética que permita evaluar cualquier subárea de concentración (A_i). Así por ejemplo, si nos interesa determinar A_i procedemos a:

$$A_i = \mathbb{N}_i - (\mathbb{N}_i - A_i) = \frac{p_i (P_i + P_{i-1})}{2} - \frac{p_i (Q_i + Q_{i-1})}{2}$$

Luego:

$$A_i = \frac{p_i}{2} \left[(P_i + P_{i-1}) - (Q_i + Q_{i-1}) \right]$$

Al aplicar la definición de Gini-intervalo para el intervalo i

$$G_i = \frac{A_i}{\mathbb{1}_i} = \frac{\frac{P_i}{2} \left[(P_i + P_{i-1}) - (Q_i + Q_{i-1}) \right]}{\frac{P_i}{2} [P_i + P_{i-1}]}$$

concluimos que:

$$G_i = 1 - \frac{Q_i + Q_{i-1}}{P_i + P_{i-1}}$$

Tenemos un caso particular cuando i asume el valor 1. El coeficiente Gini-intervalo para el primer estrato resulta ser:

$$G_1 = 1 - \frac{Q_1 + Q_0}{P_1 + P_0}$$

Si suponemos que:

$$Q_0 = P_0 = 0$$

llegamos a:

$$G_1 = 1 - \frac{Q_1}{P_1} = 1 - \frac{q_1}{p_1}$$

Expresión que se verifica si obtenemos directamente la expresión correspondiente a G_1 .

El otro caso extremo se presenta si $i = m$. El coeficiente en el último estrato asume la forma

$$G_m = 1 - \frac{1 + Q_{m-1}}{1 + P_{m-1}}$$

Si simbolizamos el área total de máxima concentración por \mathcal{Q} entonces se cumple:

$$\mathcal{Q} = \sum_{i=1}^m \mathcal{Q}_i$$

Del mismo modo si denotamos por A el área de concentración entonces:

$$A = \sum_{i=1}^m A_i$$

Sabemos que el índice de Gini se define como:

$$G = \frac{A}{\mathcal{Q}} = \frac{\sum A_i}{\sum \mathcal{Q}_i}$$

Además

$$G_i = \frac{A_i}{\mathcal{Q}_i}$$

Por lo tanto:

$$\sum_{i=1}^m G_i \left(\frac{\mathcal{Q}_i}{\mathcal{Q}} \right) = \sum_{i=1}^m \frac{A_i}{\mathcal{Q}_i} \frac{\mathcal{Q}_i}{\mathcal{Q}} = \frac{\sum A_i}{\mathcal{Q}} = \frac{\sum A_i}{\sum \mathcal{Q}_i}$$

Esta ecuación nos dice que si ponderamos los coeficientes Gini-intervalo por su correspondiente relación entre la subárea de máxima concentración (\mathcal{Q}_i) y el área de máxima concentración (\mathcal{Q}), se obtiene el Gini global.

Sabemos que:

$$\mathcal{Q}_i = \frac{(P_i - P_{i-1})(P_i + P_{i-1})}{2}$$

y que:

$$q = \frac{1}{2}$$

porque se trata del área de un triángulo rectángulo de catetos unitarios.

Luego la relación:

$$W_i = \frac{q_i}{q} = (P_i - P_{i-1})(P_i + P_{i-1}) = P_i^2 - P_{i-1}^2$$

y

$$\begin{aligned} \sum_{i=1}^m W_i &= \sum_{i=1}^m P_i^2 - \sum_{i=1}^m P_{i-1}^2 \\ &= P_1^2 + P_2^2 + \dots + P_m^2 - P_1^2 - P_2^2 - \dots - P_{m-1}^2 \\ &= P_m^2 = 1 \end{aligned}$$

$$\sum_{i=1}^m W_i = 1$$

Al aplicar estas ponderaciones a G_i obtenemos:

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

y al desarrollar:

$$\sum_{i=1}^m G_i W_i = \sum_{i=1}^m (P_i - P_{i-1}) - \sum_{i=1}^m (P_i - P_{i-1})(Q_i + Q_{i-1})$$

$$\sum_{i=1}^m G_i W_i = \sum_{i=1}^m p_i - \sum_{i=1}^m p_i (Q_i + Q_{i-1})$$

$$\sum_{i=1}^m G_i W_i = 1 - \sum_{i=1}^m p_i (Q_i + Q_{i-1})$$

El lado derecho de esta expresión es la fórmula de cálculo para el G global, por lo que:

$$G = \sum_{i=1}^m G_i W_i$$