LUIS RUIZ-MAYA

Sobre la metodología del Indice de Gini

INTRODUCCION

Este trabajo está dedicado a exponer una visión, creemos que personal, de la metodología del Indice de Gini: presentar un procedimiento aproximado del cálculo del Indice aplicable cuando el número de elementos estudiado es grande, analizar el problema de la agregación, y presentar una utilización dinámica del Indice mediante lo que hemos llamado Indices Secuenciales.

El Indice de Gini (o su paralelo gráfico la curva de Lorentz) tiene por finalidad medir el grado de concentración que presenta una variable. El primer problema que se nos plantea es qué entendemos por concentración y nada mejor que las palabras del propio Gini para aclarar este concepto: "Se dice que la riqueza de un país está tanto más concentrada cuanto mayor es la parte de la riqueza total poseída por la parte más rica de la población. Podemos decir también que la concentración de la riqueza es tanto mayor cuanto menor es la parte de esta poseída por el sector más pobre de dicha población"[1].

A fin de estudiar la concentración de una variable, tomamos un conjunto de valores de la característica objeto de análisis y formamos K intervalos de tal manera que en el intervalo i hay n_i elementos y la suma de los valores de la variable incluídos en él es s_i. Llamaremos N y S al número total de elementos y al valor total de la característica, (renta total, masa salarial, superficie total, etc.); por tanto

$$N = \sum_{i=1}^{k} n_i \; ; \; S = \sum_{i=1}^{k} s_i$$

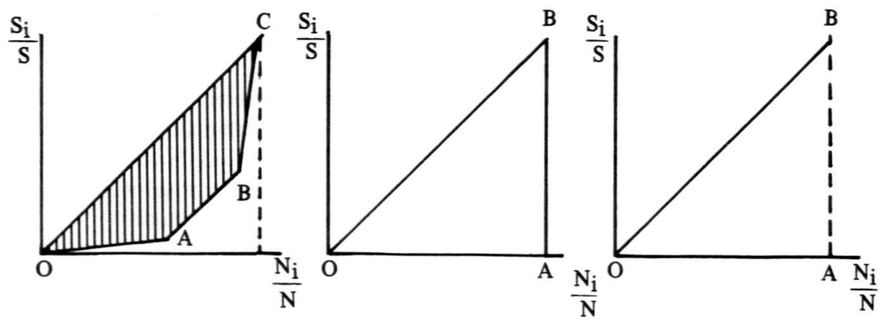
Si no dispusiéramos de los valores individuales o de los totales de cada intervalo (s_i), tendríamos que admitir que los n_i valores del intervalo i serían todos iguales a un valor intermedio, p.e. la marca de clase (c_i), siendo entonces $s_i = n_i c_i$; a los efectos de este trabajo supondremos que conocemos los valores totales s_i.

(1) GINI. C. Curso de Estadística, pág. 211. Ed. Labor 1953. Segunda edición.

Siguiendo las palabras de Gini, en primer lugar determinaremos la distribución porcentual del número y valor de la característica. Estas cifras elocuentes de por sí, ya que nos dicen que el n_i/N% de los elementos poseen el s_i/S% del valor, con lo cual se patentizan las desigualdades, no son lo suficientemente explícitas, siéndolo mucho más los porcentajes acumulados, o mejor las frecuencias relativas acumuladas.

Si $N_i = \sum_{j=1}^{i} n_j$ y $S_i = \sum_{j=1}^{i} s_j$, las respectivas frecuencias relativas acumuladas son N_i/N y S_i/S. La representación gráfica es de una gran ayuda para la comprensión del fenómeno. Para conseguirla llevamos los distintos N_i/N en abscisas y los S_i/S en ordenadas; la poligonal que se obtiene uniendo los K puntos de coordenadas (N_i/N; S_i/S) recibe el nombre de curva de Lorentz. A efectos prácticos admitiremos que el valor de N (número total de elementos) es lo suficientemente grande como para suponer que el error que se comete al tomar la distribución discreta por contínua sea inapreciable.

Si la curva de Lorentz nos representa gráficamente la concentración, su cuantificación se obtiene a través del área comprendida entre la curva y la bisectriz del primer cuadrante. La justificación de este procedimiento de cuantificación es sencilla. La máxima concentración existe cuando un elemento posee todo el valor de la característica. En este caso la curva de Lorentz está compuesta por el eje de abscisas hasta el valor uno (ó 100 según se consideren tantos por uno o por ciento) y la paralela al eje de ordenadas por este valor hasta su intersección con la bisectriz. La concentración mínima, o más exactamente nula, aparecerá cuando todos los elementos tengan el mismo valor, ya que ninguno domina a otro, coincidiendo entonces la curva de Lorentz con la primera bisectriz (llamada por esta razón recta de equidistribución).



En la situación de máxima concentración (la curva de Lorentz es OAB) el área indicada es igual a la del triángulo OAB (segunda figura); si la concentración es mínima (la curva de Lorentz es OB) el área es nula (tercera figura). Tenemos, pues, que la utilización del área OABC (primera figura) queda justificada como elemento que permite cuantificar la concentración. Una vez hecha esta breve introducción pasamos a presentar el cálculo aproximado del Indice de Gini.

CALCULO APROXIMADO DEL INDICE DE GINI

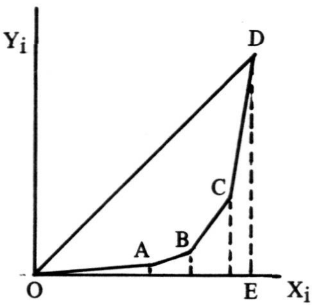
Como hemos dicho en la Introducción partimos de N valores de la característica a estudiar, clasificados en K intervalos de tal forma que en el intervalo i hay n_i elementos, con un valor total de la característica, en él, igual a s_i. Calculamos las cantidades acumuladas

$$N_i = \sum_{j=1}^{i} n_j \quad ; \quad S_i = \sum_{j=1}^{i} s_j$$

y sus cocientes respecto a los totales

$$x_i = \frac{N_i}{N} \quad ; \quad y_i = \frac{S_i}{S}$$

En unos ejes coordenados llevamos los K pares $(x_i; y_i)$ obteniendo la curva de concentración de Lorentz OABCD.



Utilizaremos como medida de la concentración la relación del área (OABCD) a la del triángulo (OED). Para ello calcularemos primero el área (OABCDE). Este área se compone de un triángulo y (K-1) trapecios. Si denominamos el área total (OABCDE) por $\emptyset$.

$$\emptyset = \frac{1}{2} x_1 y_1 + \frac{1}{2}(x_2 - x_1)(y_2 + y_1) + \frac{1}{2}(x_3 - x_2)(y_3 + y_2) + \text{---} + \frac{1}{2}(x_k - s_{k-1})(y_k + y_{k-1})$$

según la definición de x_i e y_i

$$\emptyset = \frac{1}{2}\left[\frac{N_1}{N}\frac{S_1}{S} + \frac{N_2 - N_1}{N}\frac{S_2 + S_1}{S} + \frac{N_3 - N_2}{N}\frac{S_3 + S_2}{S} + \dots + \frac{N_k - N_{k-1}}{N}\frac{S_k + S_{k-1}}{S}\right] =$$

$$= \frac{1}{2NS}\left[n_1 s_1 + n_2(s_2 + 2s_1) + n_3(s_3 + 2s_2 + 2s_1) + \text{---} + n_k(s_k + 2s_{k-1} + 2s_{k-2} + \text{--} + 2s_1)\right]$$

podemos simplificar esta expresión, observando que, p. e.

$$s_3 + 2s_2 + 2s_1 = 2s_3 + 2s_2 + 2s_1 - s_3 = 2\,S_3 - s_3$$

$$\phi = \frac{1}{2NS}\left[n_1 s_1 + n_2 (2s_2 + 2s_1 - s_2) + n_3 (2s_3 + 2s_2 + 2s_1 - s_1) + \cdots + n_k (2s_x + \cdots + 2s_1 - s_k) \right] =$$

$$= \frac{1}{2NS} \sum_{i=1}^{k} n_i \,(2\,S_i - s_i)$$

El área (OABCD) es igual a $\frac{1}{2} - \phi$ y si hacemos la unidad el área del triángulo (OED), obtenemos como resultado el Indice de Gini aproximado mediante una simple regla de tres.

$$I = 1 - 2\,\phi$$

$$I = 1 - \frac{1}{NS} \sum_{i=1}^{k} n_i\,(2\,S_i - s_i) \qquad [1]$$

Antes de continuar creemos conveniente presentar la forma en que Gini efectúa el cálculo del Indice exacto. Primeramente ordena de menor a mayor cada uno de los N elementos, procediendo, a continuación, a obtener las sumas acumuladas. Utilizando su simbología A_i es la suma de los valores de los i primeros elementos, y A_N la suma total. Como indicadores de los desequilibrios en la distribución calcula dos cocientes.

$$q_i = \frac{A_i}{A_N}; \quad p_i = \frac{i}{N}$$

"Esto significa que q_i será la fracción que sobre el total de las N rentas consideradas representa la cuantía de las rentas más bajas y que p_i será la fracción que el número de los i rentistas más pequeños representa sobre el número total N de los rentistas"[2].

Para Gini la diferencia $p_i - q_i$ es la base que le permitirá cuantificar la concentración y lo hará mediante la expresión R, llamada por él "razón de concentración".

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

Como vemos Gini parte, para el cálculo del Indice, del conocimiento individual

(2) Ibidem, pág. 211.

de cada elemento. Esta situación no es la más frecuente mientras que sí lo es el disponer de la distribución de frecuencias, supuesto del que hemos partido. Gini contempla, también, esta posibilidad llegando a una expresión (aproximación de la primera) de la cual la [1] nuestra es una fase más elaborada.

Una vez expuesta la forma mediante la cual Gini realiza el cálculo, vamos a estudiar utilizando nuestra expresión [1] los límites de variación del Indice aproximado.

La finalidad del Indice de Gini es medir la concentración de una característica. Esta concentración, lógicamente, debe estar comprendida entre la máxima, un individuo posee todo y el resto nada, y la mínima, la distribución es perfecta en el sentido de que el x_i% de individuos posee el x_i% de la característica analizada.

Veamos en ambos casos cuales son los valores del Indice que miden tales situaciones.

Supongamos que en el intervalo superior hay un solo elemento, $n_k = 1$ que posee todo el valor de la característica, por la cual $s_k = S$; esto implica que $s_1 = s_2 = \ldots = s_{k-1} = 0$. Podemos expresar [1] de la siguiente forma:

$$I = 1 - \frac{1}{NS} \left[\sum_{i=1}^{k-1} n_i (2 S_i - s_i) + n_i (2 S_k - s_k) \right]$$

Al ser todo $s_i =$ (excepto s_k), todo $S_i = 0$, anulándose el sumatorio del segundo miembro, por lo cual

$$I = 1 - \frac{1}{NS} \left[n_k (2 S_k - s_k) \right]$$

Ahora bien, según vimos $n_k = 1$ y $s_k = S_k = S$, resultando

$$I = 1 - \frac{1}{NS} (2 S - S) = 1 - \frac{1}{N}$$

El límite superior de nuestra aproximación al Indice de Gini depende del número total de elementos considerados. Número que al tender a infinito hace que el límite sea la unidad.

Desde el punto de vista práctico cuando se trabaja con un número suficientemente grande de elementos puede despreciarse el término $1/N$, no así cuando N es pequeño: como situación extrema tenemos que la máxima concentración con dos elementos es 0,5, bastante alejada del usual valor uno en situaciones semejantes (y que se obtendría utilizando el método exacto de Gini), pareciendo indicar el 0,5 que la concentración no es acusada. A esto puede ofrecerse una fácil explicación: únicamente tiene sentido hablar de concentración si la mayor parte de la variable pertenece, en términos relativos, a un número muy reducido de elementos. En el caso de que este número sea dos como situación extrema toda la variable podrá pertenecer a uno, pero este uno no representa un número reducido de elementos sino el 50%. Tenemos, por tanto, que la aplicación de la aproximación del Indice de

Gini solo es plenamente coherente cuando el número de elementos es elevado[3], de tal forma que si solo uno posee toda la característica, este elemento suponga un porcentaje muy bajo respecto del total.

La situación opuesta a la hasta aquí analizada es la de distribución perfecta, que aparece cuando el x_i% de individuos posee el x_i% de la característica, es decir, no existe concentración.

Por definición:

$$x_i = \frac{N_i}{N} \quad ; \quad y_i = \frac{S_i}{S}$$

$$s_i \quad x_i = y_i \quad , \quad \frac{N_i}{N} = \frac{S_i}{S}$$

o lo que es igual $\frac{S}{N} = \frac{S_i}{N_i}$, lo que nos dice que el valor medio total (S/N) debe ser igual al valor medio hasta el intervalo i. Según esto,

$$\frac{S}{N} = \frac{s_1}{n_1} \quad ; \quad \frac{S}{N} = \frac{S_2}{N_2} = \frac{s_2 + s_i}{n_2 + n_i}$$

haciendo operaciones,

$$Sn_2 + Sn_1 = Ns_2 + Ns_1 \text{ ; como } Sn_1 = Ns_1 \text{ ; } Sn_2 = Ns_2$$

$$\frac{S}{N} = \frac{s_2}{n_2}$$

y en general tendremos,

$$\frac{S}{N} = \frac{s_i}{n_i}$$

lo que implica que todos los intervalos poseen la misma densidad de la variable. Partiendo de esto podemos obtener el límite inferior del Indice de Gini. Si $x_i = y_i$ el área $\emptyset$ es igual a

$$\varphi = \frac{1}{2} \left[x_1^2 + (x_2 - x_1)(x_2 + x_1) + (x_3 - x_2)(x_3 + x_2) + \cdots + (x_k - x_{k-1})(x_k + x_{k-1}) \right] =$$

$$= \frac{1}{2} \left[x_1^2 + (x_2^2 - x_1^2) + (x_3^2 - x_2^2) + \cdots + (x_k^2 - x_{k-1}^2) \right] = \frac{x_k^2}{2}$$

(3) Evidentemente el término "elevado" es impreciso, por otra parte no podemos dar ningún valor de N a partir del cual pudiéramos admitir la utilización coherente de nuestra aproximación al Indice de Gini; sin embargo, y sin que esto suponga contradicción con la afirmación anterior, podemos observar que si el número de elementos es de mil el error cometido, en el límite superior, entre el valor exacto (uno) y el aproximado es de una milésima.

Ahora bien $x_k = \frac{N_k}{N} = \frac{N}{N} = 1$ por lo cual

$$\emptyset = \frac{1}{2} \text{ , y el Indice } I = 1 - 2\,\emptyset = 1 - 1 = 0$$

Veamos, pues, que en la situación de distribución óptima el Indice es nulo, y la curva de Lorentz coincide con la bisectriz del primer cuadrante.

En general las frecuencias relativas acumuladas del número de elementos son mayores que las de los valores de la característica, cuando no sucede esto el Indice puede tomar valores negativos, aunque siempre comprendidos en el intervalo, (0, – 1), apareciendo la curva de Lorentz por encima de la bisectriz del primer cuadrante. En la parte de este trabajo dedicada a los Indices Secuenciales veremos como se dá la situación de negatividad del Indice.

A continuación vamos a examinar el comportamiento del Indice aproximado ante cambios de escala proporcionales, y cambios de origen.

Tenemos una distribución cuyo número de elementos es proporcional al de otra, $m_i = an_i$, sucediendo lo mismo con los valores de la variable en cada intervalo, $r_i = bs_i$ manteniéndose el número de intervalos K,

$$x_i^1 = \frac{\sum_{j=1}^{i} m_j}{\sum_{j=1}^{K} m_j} = \frac{M_i}{M} = \frac{a\,N_i}{a\,N} = \frac{N_i}{N} = x_i$$

$$y_i^1 = \frac{\sum_{j=1}^{i} r_j}{\sum_{j=1}^{K} r_j} = \frac{R_i}{R} = \frac{b\,S_i}{bS} = \frac{S_i}{S} = y_i$$

obteniéndose como resultado la invarianza del Indice respecto a cambios proporcionales. No sucede así con las traslaciones. Si consideramos, en general $m_i = an_i + c$ y $r_i = b\,s_{i} + d$.

$$x_i^1 = \frac{M_i}{M} = \frac{a\,N_i + ic}{aN + Kc} \neq \frac{N_i}{N}$$

análogamente,

$$y_i^1 = \frac{R_i}{R} = \frac{b\,S_i + i\,d}{b\,S + kd} \neq \frac{S_i}{S}$$

EL PROBLEMA DE LA AGREGACION EN EL ESTUDIO DE LA CONCENTRACION

El cálculo del Indice de Gini entraña la utilización de las distribuciones del número y valor de la característica estudiada; es lógico plantearse el problema de en qué medida influye la agregación de intervalos en la concentración.

La agregación ha tenido por objetivo englobar en un intervalo varios de los ya existentes. El primer intervalo nuevo está formado por los h_1 primeros intervalos antiguos; su frecuencia es m_1, suma de las correspondientes frecuencias antiguas, $n_1, n_2, -, n_{hi}$. En general el nuevo intervalo g-esimo contiene h_g intervalos antiguos, y su frecuencia es

$$m_g = n_{h1+ - + h_{g-1}+1} + n_{h1+ - + h_{g-1}+2} + - + n_{hi+ - h_g} = \sum_{i=1}^{hg} n_{hi+ - + h_{g-i}+i}$$

El primer paso que vamos a dar es el cálculo de la concentración del intervalo g-esimo, partiendo de los h_g intervalos primitivos que lo integran. Para una mayor comodidad en el cálculo del Indice establecemos la tabla siguiente que corresponde al intervalo g, llamando d a la suma $h_1 + h_2 + - + h_{g-1}$

n_{d+1} ; $s_{d+1} + 2 s_d + 2 s_{d-1}$ + ——————————— $+ 2s_2 + 2s_1$

n_{d+2} ; $s_{d+2} + 2s_{d+1} + 2 s_d + 2 s_{d-1}$ + ——————————— $+ 2s_2 + 2s_1$

n_{d+3} ; $s_{d+3} + 2s_{d+2} + 2 s_{d+1} + 2s_d + 2 s_{d-1}$ + ——————— $+ 2s_2 + 2s_1$

---

n_{d+hg} ; $n_{d+hg} + 2 s_{d+hg-1} + - + 2 s_{d+1} + 2 s_d + 2 s_{d-1} + - + 2 s_2 + 2s_1$

Para el cálculo del Indice en el intervalo g, (I'g), solo hemos de tener en cuenta los valores de la variable hasta s_d ya que éste y los inferiores a él pertenecen a intervalos anteriores al g-esimo.

El Indice de Gini aproximado es

$$I'g = 1 - \frac{1}{m_g S'_g} \sum_{i=1}^{hg} n_{d+i} \, (2\, S'_{d+i} - s_{d+i}) \qquad [2]$$

siendo m_g, como ya vimos, la frecuencia del intervalo g, y siendo la expresión de S'_g y S'_{d+i}

$$S'_g = \sum_{i=1}^{hg} s_{d+i} \quad , \quad S'_{d+i} = \sum_{j=1}^{i} s_{d+j}$$

El Indice global primitivo podemos descomponerlo de la siguiente manera:

$$I = 1 - \frac{1}{NS} \sum_{i=1}^{k} n_i (2S_i - s_i) =$$

$$= 1 - \frac{1}{NS} \Big[\sum_{i=1}^{h1} n_i (2S_i - s_i) + \sum_{i=1}^{h2} n_{h1+i} (2\, S_{h1+i} - s_{h1+i}) +$$

$$+ - + \sum_{i=1}^{hg} n_{h1+ - + h_{g-1}+i} (2S_{h1+ - + h_{g-1}+i} - s_{h1+ - + h_{g-1}+i}) +$$

$$+ - + \sum_{i=1}^{h_e} n_{h1+ - + h_{e-i}+i} (2S_{h1+ - + h_{e-1}+i} - s_{h1+ - + h_{e-1}+i}) \Big] =$$

$$= 1 - \frac{1}{NS} \sum_{g=1}^{e} \sum_{i=1}^{hg} n_{h1+ \_+ h_{g-1}+i} (2S_{h1+ -+ h_{g-1}+i} - s_{h1+ -+ h_{e-1}+i}) =$$

$$= 1 - \frac{1}{NS} \sum_{g=1}^{e} \sum_{i=1}^{hg} n_{d+1} (2 S_{d+i} - s_{d+i}) \qquad [3]$$

Teniendo en cuenta la tabla anterior vamos a expresar el término $2S_{d+i} - s_{d+i}$ de una manera más conveniente:

$$2 S_{d+i} - s_{d+i} = s_{d+i} + 2s_{d+i-1} + - + 2 s_{d+1} + 2s_d + - + 2s_2 + 2s_1 =$$

$$= s_{d+i} + 2S_{d+i-1} + - + 2s_{d+1} + 2 S_d =$$

$$= 2 S'_{d+i} - s_{d+i} + 2 S_d$$

Sustituyendo este segundo miembro en [3]

$$I = 1 - \frac{1}{NS} \sum_{g=1}^{e} \sum_{i=1}^{hg} n_{d+i} (2 S'_{d+i} - s_{d+i} + 2 S_d) =$$

$$= 1 - \frac{1}{NS} \sum_{g=1}^{1} \left[\sum_{i=1}^{hg} n_{d+i} (2S'_{d+i} - s_{d+i}) + \sum_{i=1}^{hg} 2n_{d+i} S_d \right] \qquad [4]$$

Según [2]

$$\sum_{i=1}^{hg} n_{d+i} (2S'_{d+i} - s_{d+i}) = m_g S'_g (1 - I'_g)$$

quedándonos [4]

$$I = 1 - \frac{1}{NS} \sum_{g=1}^{e} \left[m_g S'_g (1 - I'_g) + 2 S_d \sum_{i=1}^{hg} n_{d+i} \right] =$$

$$= 1 - \frac{1}{NS} \sum_{g=1}^{e} \left[m_g S'_g (1 - I'g) + 2 m_g S_d \right] =$$

$$= 1 - \frac{1}{NS} \sum_{g=1}^{e} \left[m_g (2S_d + S'_g) - m_g S'_g I'_g \right] \qquad [5]$$

Por su interés vamos a analizar en detalle el término $m_g (2S_d + S'_g)$; m_g es la frecuencia del nuevo intervalo g-esimo, S'_g el valor total de la variable en ese intervalo, y S_d el valor acumulado de la variable hasta el intervalo g (recordemos que $d = h_1 + --- + h_{g-1}$); por todo ello tenemos que el sumatorio.

$$\sum_{g=1}^{e} m_g (2 S_d + S'_g)$$

es el que corresponde al Indice de Gini aproximado, de los nuevos e intervalos, es decir, si llamamos I_e a este Indice.

$$I_e = 1 - \frac{1}{NS} \sum_{g=1}^{e} m_g (2\,S_d + S'_g)$$

de aquí deducimos que

$$\sum_{g=1}^{e} m_g (2\,S_d + S'_g) = N\,S\,(1 - I_e)$$

por lo cual [5] queda,

$$I = 1 - \frac{1}{NS} \left[NS\,(1 - I_e) - \sum_{g=1}^{e} m_g\,S'_g\,I'_g \right] =$$

$$= I_e + \frac{1}{NS} \sum_{g=1}^{e} m_g\,S'_g\,I'_g \qquad [6]$$

La expresión [6] nos dice que el Indice de Gini (calculado por el método aproximado) obtenido a partir de K intervalos, esto es antes de efectuar la agregación, es igual a la suma del Indice del colectivo agregado más la suma ponderada de los e índices que miden la concentración de los intervalos que integran cada uno de los nuevos. Tomando la simbología del Análisis de la Varianza podemos expresar el valor del Indice, sin agregar, como la suma de la concentración *entre* los nuevos intervalos y la suma ponderada de la concentración *dentro* de los nuevos grupos.

En este punto estamos en condiciones de estudiar el efecto de la agregación sobre el Indice de Gini como medida de la concentración de un colectivo. En general la agregación conduce a Indices menores que los obtenidos antes de agregar, ya que la suma ponderada de los indices dentro de cada nuevo grupo es mayor que cero[4].

INDICES SECUENCIALES

Al utilizar el método aproximado del Indice de Gini para medir la concentración de una variable debemos, de alguna manera, formar una serie de intervalos de magnitud creciente, mediante los cuales clasificamos cada valor de la variable; en otras palabras, hemos formado una distribución de frecuencias. Apoyándonos en este hecho vamos a dar un paso más en el estudio de la concentración mediante el Indice de Gini, analizando el comportamiento dinámico de los distintos tamaños que hemos formado para la obtención de la distribución de frecuencias.

El planteamiento del que partimos para efectuar el análisis dinámico es el siguiente: cuando calculamos el Indice del conjunto, de todos los elementos, hemos cuantificado la concentración del colectivo utilizando los intervalos establecidos, no proporcionando este valor del Indice información alguna sobre la magnitud de la influencia de los distintos grupos. Para conseguir esta información calculamos el

(4) Hemos dicho en general porque puede darse el caso de la aparición de Indices negativos que invertirían la situación.

Indice de los dos primeros tamaños; su valor nos indica en qué medida la concentración (de los dos intervalos) se aparta de la equidistribución. A continuación añadimos el tercer intervalo calculando el Indice de los tres. En general los dos índices no serán iguales, pudiéndose atribuir esta diferencia al hecho de haber incluído el tercer intervalo, o lo que es igual, al hecho de considerar valores de la variable mayores que los anteriores.

Continuando de esta forma obtenemos una serie de índices (uno menos que el número total de intervalos) cuyas diferencias consecutivas nos van indicando el comportamiento de los intervalos que se van introduciendo. (Evidentemente el último índice es igual al global). A este conjunto de índices le hemos dado el nombre de Indices Secuenciales[5].

A continuación analizaremos en detalle las principales características de los Indices Secuenciales.

Tenemos el conjunto de valores de la variable clasificados en una serie de intervalos, y hemos calculado el Indice de Gini de los K primeros tamaños, llamaremos I_k a este Indice, siendo su expresión.

$$I_k = 1 - \frac{1}{NS} \sum_{i=1}^{k+1} n_i (2 S_i - s_i)$$

Incluímos el grupo siguiente: el nuevo Indice será:

$$I_{k+1} = 1 - \frac{1}{N_{k+1} S_{k+1}} \sum_{i=1}^{k+1} n_i (2 S_i - s_i)$$

Vamos a operar en el segundo miembro de I_{k+1}

$$I_{k+1} = 1 - \frac{1}{N_{k+1} S_{k+1}} \left[\sum_{i=1}^{k} n_i (2S_i - s_i) + n_{k+1} (2 S_{k+1} - s_{k+1}) \right] =$$

$$= 1 - \frac{1}{N_{k+1} S_{k+1}} \left[\frac{N_k S_k}{N_k S_k} \sum_{i=1}^{k} n_i (2S_i - s_i) + n_{k+1} (2S_{k+1} - s_{k+1}) \right] =$$

$$= 1 - \frac{1}{N_{k+1} S_{k+1}} \left[N_k S_k (1 - I_k) + n_{k+1} (2 S_{k+1} - s_{k+1}) \right] =$$

$$= 1 - \frac{N_k S_k}{N_{k+1} S_{k+1}} (1 - I_k) - \frac{n_{k+1}}{N_{k+1}} \frac{2 S_{k+1} - s_{k+1}}{S_{k+1}} =$$

(5) El autor ha utilizado el concepto de Indices Secuenciales en el trabajo "ESTUDIO DINAMICO DE LA CONCENTRACION DE LA TIERRA". Agricultura y Sociedad n.° 3. Secretaría General Técnica. Ministerio de Agricultura. Esta parte del presente estudio es la justificación metodológica del trabajo citado.

$$= 1 - \frac{N_k\ S_k}{(N_k + n_{k+1})(S_k + s_{k+1})}(1 - I_k) - \frac{n_{k+1}}{N_k + n_{k+1}} \ \frac{2\,S_k + s^{k+1}}{S_k + s_{k+1}} =$$

$$= 1 - \frac{1}{\left(1+\frac{n_{k+1}}{N_k}\right)\left(1+\frac{s_{k+1}}{S_k}\right)}(1 - I_k) -$$

$$- \frac{\frac{n_{k+1}}{N_K}}{1+\frac{n_{k+1}}{N_K}} \ \frac{2+\frac{s_{k+1}}{S_K}}{1+\frac{s_{k+1}}{S_K}} \qquad [7]$$

Si hacemos $\frac{n_{k+1}}{N_k} = n$ y $\frac{s_{k+1}}{S_k} = s$, [7] queda

$$I_{k+1} = 1 - \frac{1 - I_K}{(1+n)(1+s)} - \frac{n\,(2+s)}{(1+n)(1+s)} = \frac{I_k + s - n}{(1+n)(1+s)}$$

Hemos obtenido una expresión que nos liga el Indice de K intervalos con el Indice de K + 1 intervalos; en ella juegan un papel especial n y s que son en número y valor, respectivamente, la relación (en tanto por uno) de la frecuencia del nuevo intervalo a la frecuencia total de los K intervalos primeros, y análogamente con el valor de la variable. En otras palabras, n y s nos indican la importancia relativa de los elementos que integran el nuevo tamaño respecto a los anteriores. Para valores fijos de I y s es decir, considerando I_{k+1} como función de n exclusivamente, I_{k+1} es función decreciente de n. Por el contrario si se contempla I_{k+1} como función de s es creciente, para cada valor de I_k y n. A vía de ejemplo en el Cuadro 1, presentamos los posibles valores de I_{k+1} si I_k es igual a 0,7 para valores de n y s de 0,1 a 2,0. Para una mayor claridad hemos representado estos valores en dos gráficos. En el Gráfico 1 aparecen en abscisas valores de s y en ordenadas los de los Indices I_{k+1}. Cada curva correspondende a un valor de n, desde 0,1 la superior hasta 2 la inferior, variando n de 0,1 en 0,1. La interpretación del Gráfico es la siguiente: para s = 0,4 (el valor de la característica en el intervalo K + 1 es el 40% del valor total en los K intervalos anteriores) y n = 0,3 (el número de elementos en el intervalo K + 1 es el 30% de los anteriores) el Indice correspondiente a K + 1 intervalos sería 0,41 si el de K intervalos es 0,7, la concentración ha disminuído. (La curva en este caso es la tercera comenzando por la parte superior).

En el Gráfico 2 hemos representado los mismos valores de I_{k+1} variando en este caso n; cada curva corresponde a uno de s, variando desde la parte inferior a la superior; su interpretación es análoga a la del Gráfico 1.

Tanto en el Cuadro como en los Gráficos podemos apreciar el hecho de que la concentración puede disminuir, e incluso hacerse negativa, al incluir nuevos valores

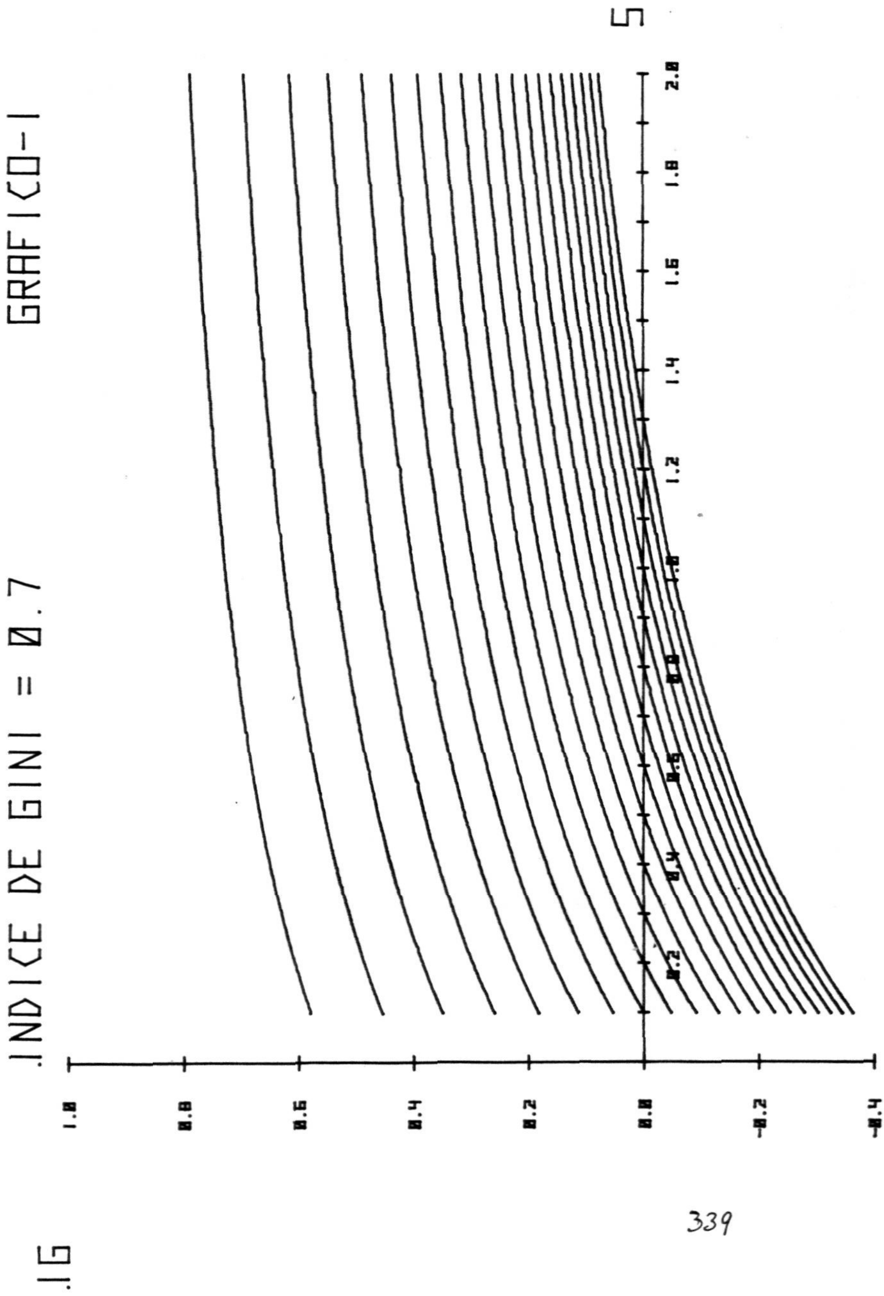


GRAFICO-I
INDICE DE GINI = 0.7
S
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
IG

339

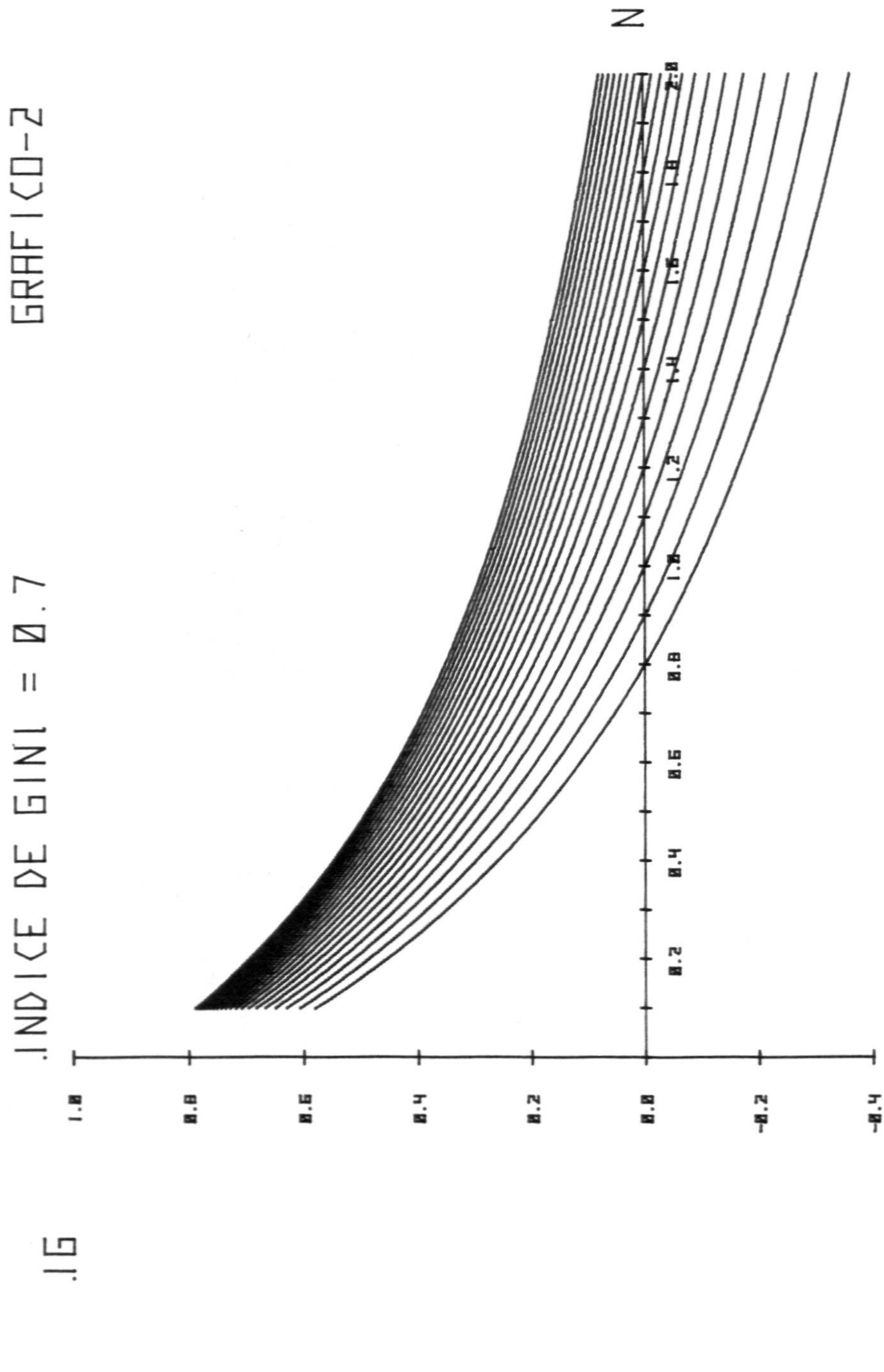


GRAFICO-2
INDICE DE GINI = Ø.7
IG
N
1.Ø
Ø.8
Ø.6
Ø.4
Ø.2
Ø.Ø
-Ø.2
-Ø.4
Ø.2
Ø.4
Ø.6
Ø.8
1.Ø
1.2
1.4
1.6
1.8
2.Ø

de la característica, siempre que se cumplan las condiciones que se exponen a continuación.

La diferencia entre los Indices I_k e I_{k+1} es

$$\Delta I_k = I_{k+1} - I_k = \frac{I_k + s - n}{(1+n)(1+s)} - I_k = \frac{s-n}{(1+n)(1+s)} - \frac{s+n+ns}{(1+n)(1+s)} I_k$$

Vamos a ver bajo que condiciones de n y s la diferencia de los indices es positiva, negativa o nula.

$$\Delta I_k > 0; \quad \frac{s-n}{(1+n)(1+s)} > I_k \frac{s+n+ns}{(1+n)(1+s)}; \quad \frac{s-n}{s+n+ns} > I_k \qquad [8]$$

Si admitimos que la distribución de frecuencia S no puede conducir a Indices negativos[6] de la condición [8] se deduce que si $s > n$ la diferencia también es positiva (aunque la condición [8] es más fuerte). Implicando la condición $s > n$ por la definición de ambos parámetros que

$$\frac{s_{k+1}}{n_{k+1}} > \frac{S_k}{N_k}$$

esto es, que la densidad del intervalo K+ 1 es mayor que la densidad media en los K intervalos anteriores.

De una forma análoga $\Delta I_k < 0$ cuando

$$\frac{s-n}{s+n+ns} < I_k \qquad [9]$$

Las relaciones [8] y [9] nos dicen que los Indices Secuenciales pueden ser crecientes o decrecientes, dependiendo una u otra situación de s y n, de la importancia relativa del nuevo tamaño considerado. Esto puede comprobarse fácilmente en el Cuadro anterior.

La finalidad de los Indices Secuenciales radica en el análisis de la estructura dinámica de la concentración de una población. En una primera aproximación esto se ha logrado mediante las condiciones [8] y [9], sin embargo aún podemos ir más lejos si utilizamos una herramienta tan potente como es el concepto de elasticidad.

Como sabemos la elasticidad de una magnitud con respecto a otra nos mide la valoración relativa de una de ellas (en tanto por ciento) correspondiente a una variación del uno por ciento de la otra. La ventaja de la elasticidad es innegable, elimina el efecto de los valores absolutos, mostrando la respuesta relativa en magnitud y signo. La elasticidad puede obtenerse en un punto o en un intervalo; en nuestro

(6) La condición de no negatividad no parece excesiva a la luz de las distribuciones de las variables económicas más corrientes.

caso nos referiremos a la elasticidad en un intervalo. En la elasticidad de los Indices Secuenciales el problema que surge es la determinación de la magnitud respecto a la cual la calcularemos. La más aconsejable es la "densidad media" de los K intervalos considerados, S_k / N_k (si la variable fuera la masa salarial la densidad sería el salario medio global; si fuera la superficie de explotaciones, la superficie acumulada media, etc.) Las "densidades medias" son en cierta forma un indicador de la presencia de concentración ya que si todos los intervalos tuvieran la misma densidad la concentración sería nula (el Indice cero), por tanto parece razonable relacionar la diferencia de índices con la diferencia de densidades.

Si llamamos ΔM a la variación de densidad tenemos que

$$\Delta M = \frac{S_{k+1}}{N_{k+1}} - \frac{S_k}{N_k} = \frac{S_k + s_{k+1}}{N_k + n_{k+1}} - \frac{S_k}{N_k} = \frac{S_k \left[1 + \frac{s_{k+1}}{S_k} \right]}{N_k \left[1 + \frac{n_{k+1}}{N_k} \right]} - \frac{S_k}{N_k} =$$

$$= \frac{S_k (1+s)}{N_k (1+n)} - \frac{S_k}{N_k} = \frac{S_k}{N_k} \frac{s-n}{1+n}$$

Si $M = \dfrac{S_k}{N_k}$ es la densidad media hasta el intervalo K, la variación relativa será

$$\frac{\Delta M}{M} = \frac{s-n}{1+n}$$

Como vemos esta variación relativa solo depende de s y n

De una manera análoga obtenemos la variación relativa el Indice Secuencial.

$$\frac{\Delta I_k}{I_k} = \frac{s-n}{I_k (1+n)(1+s)} - 1 + \frac{1}{(1+n)(1+s)}$$

Llamamos E a la elasticidad en un intervalo

$$E = \frac{\frac{\Delta I_k}{I_k}}{\frac{\Delta M}{M}} = \frac{1+n}{I_k (1+n)(1+s)} + \frac{1+n}{s-n} + \frac{1+n}{(1+n)(1+s)(s-n)} =$$

$$= \frac{1}{1+s} \left[\frac{1}{I_k} - \frac{s+n+ns}{s-n} \right] \qquad [10]$$

Si es interesante el crecimiento o decrecimiento de los Indices Secuenciales, ya que esto indica que el nuevo tamaño acentúa o aminora la concentración que existía sin él, más importante es el que la elasticidad sea mayor o menor que uno, positiva o negativa. El hecho de ser positiva supone que el Indice crece a la vez que la "densidad media", si fuera negativa tendría lugar un decrecimiento relativo del Indice impulsado por un crecimiento de la densidad. Si la elasticidad, por ejemplo, es mayor que uno esto supone que el crecimiento relativo del Indice es más que proporcional, comparado con el de la densidad, si el valor es menor que la unidad el crecimiento relativo es menos que proporcional.

El que la elasticidad sea creciente o decreciente nos permitirá clasificar la concentración como progresiva o regresiva (neutra si la elasticidad no varía), calificando, de la misma forma, a los intervalos según induzcan unos u otros tipos de elasticidades. Creemos que desde el punto de vista descriptivo y analítico, es más interesante el uso de la elasticidad que el de los Indices Secuenciales, dado que mientras éstos, en general, son crecientes la elasticidad puede no serlo, permitiendo, como ya hemos indicado, la posibilidad de matizar los efectos de los distintos intervalos.

A continuación vamos a establecer las condiciones de variación de la elasticidad.

Elasticidad mayor que cero.

$$\frac{1}{1+s}\left[\frac{I}{I_k} \quad \frac{s+n+ns}{s-n}\right] > 0; \quad \frac{1}{I_k} > \frac{s+n+ns}{s-n}$$

$$\frac{s-n}{s+n+ns} > I_k \qquad [11]$$

que coincide, como es lógico, con la condición [8] que establecía $\Delta I_k > 0$.

La elasticidad será menor que cero cuando se verifique la condición contraria a la [11].

Elasticidad menor que la unidad.

$$\frac{1}{1+s}\left[\frac{1}{I_k} - \frac{s+n+ns}{s-n}\right] < 1 \quad ; \frac{1}{I_k} - \frac{s+n+ns}{s-n} < 1+s$$

$$\frac{1}{I_k} < \frac{s(2+s)}{s-n} \quad ; \quad \frac{s-n}{s(2+s)} < I_k \qquad [12]$$

La elasticidad mayor que uno implica

$$\frac{s-n}{s(2+s)} > I_k \qquad [13]$$

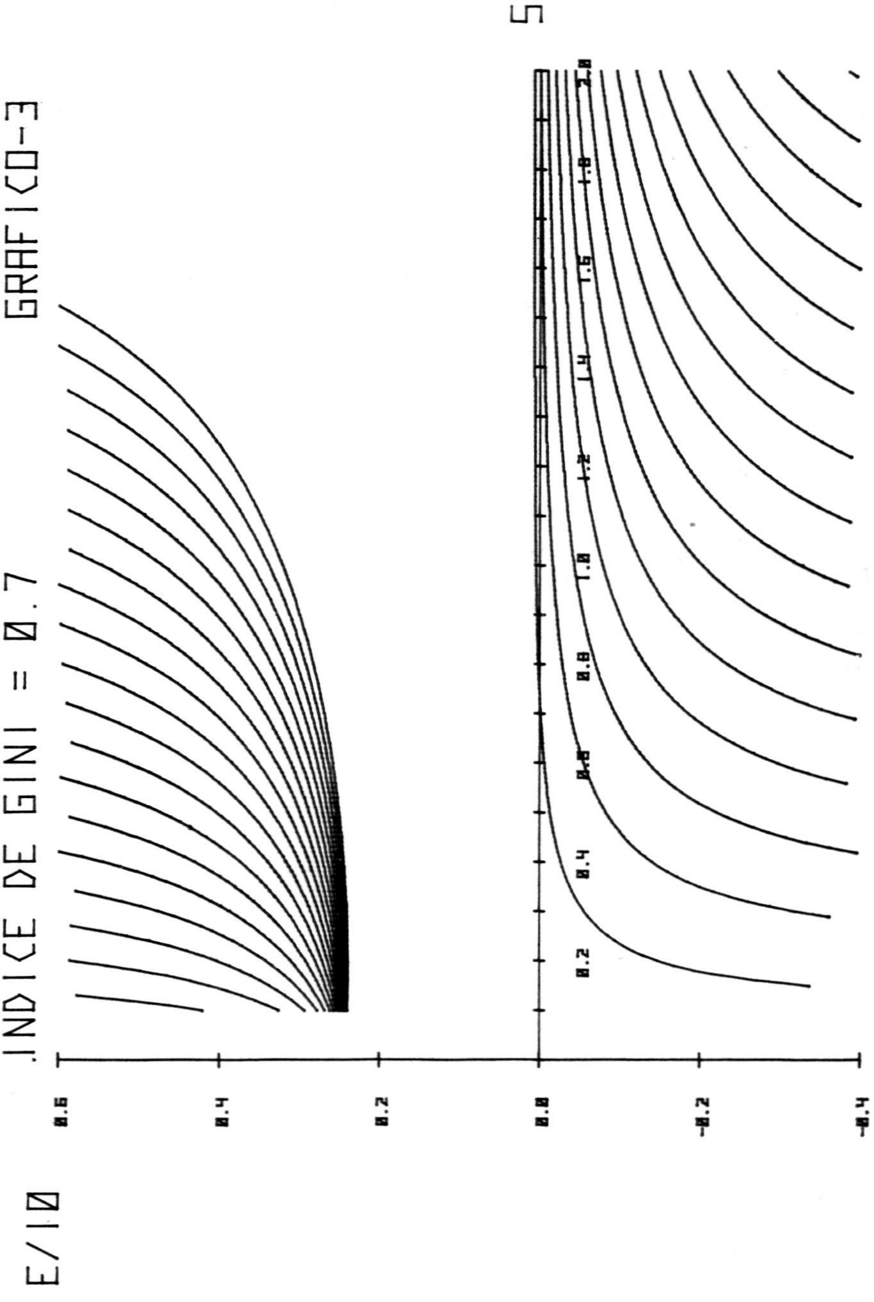
Las condiciones [11], [12], y [13] importan más por el hecho de la existencia de los valores que suponen, que por las condiciones en sí mismas.

En el Cuadro 2 hemos tabulado las elasticidades para los valores del Indice I_K, s y n, análogos a los del Cuadro 1. Los espacios en blanco equivalen a valores de la elasticidad iguales a infinito.

En los Gráficos 3 y 4 tenemos la representación de la elasticidad (dividida por 10 por motivos de escala) para cada valor de I_K y n ó s fijos, respectivamente. A cada curva de la parte positiva del eje de ordenadas le corresponde una rama en la parte negativa. El número de curvas (consideradas las dos ramas) no es igual a 20 debido a los condicionamientos de ejecución de los gráficos, ya que son plausibles valores de s ó n superiores a 2 (que equivalen al 200%).

En el Gráfico 3 podemos contemplar el comportamiento de la elasticidad para valores fijos de I_K y n, y variables de s. La conclusión más importante que podemos extraer es que la elasticidad de los Indices Secuenciales es siempre creciente sea cual sea el valor de s [7]. Esto supone que si la importancia relativa del número de elementos del intervalo K + 1 permanece fija (n), al aumentar la importancia de s el intervalo K + 1 presenta una concentración progresiva. Por el contrario si tomamos n como variable (Gráfico 4) la elasticidad es decreciente. Al movernos a lo largo de una misma línea (s fija), vemos que el incremento de n es decisivo a la hora de disminuir la concentración, haciendo que el intervalo incluído sea regresivo.

(7) Para encontrar elasticidades decrecientes hay que llegar a valores de I_K inferiores a 0,3.


INDICE DE GINI = 0.7
GRAFICO-3
S
E/10
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
0.6
0.4
0.2
0.0
-0.2
-0.4

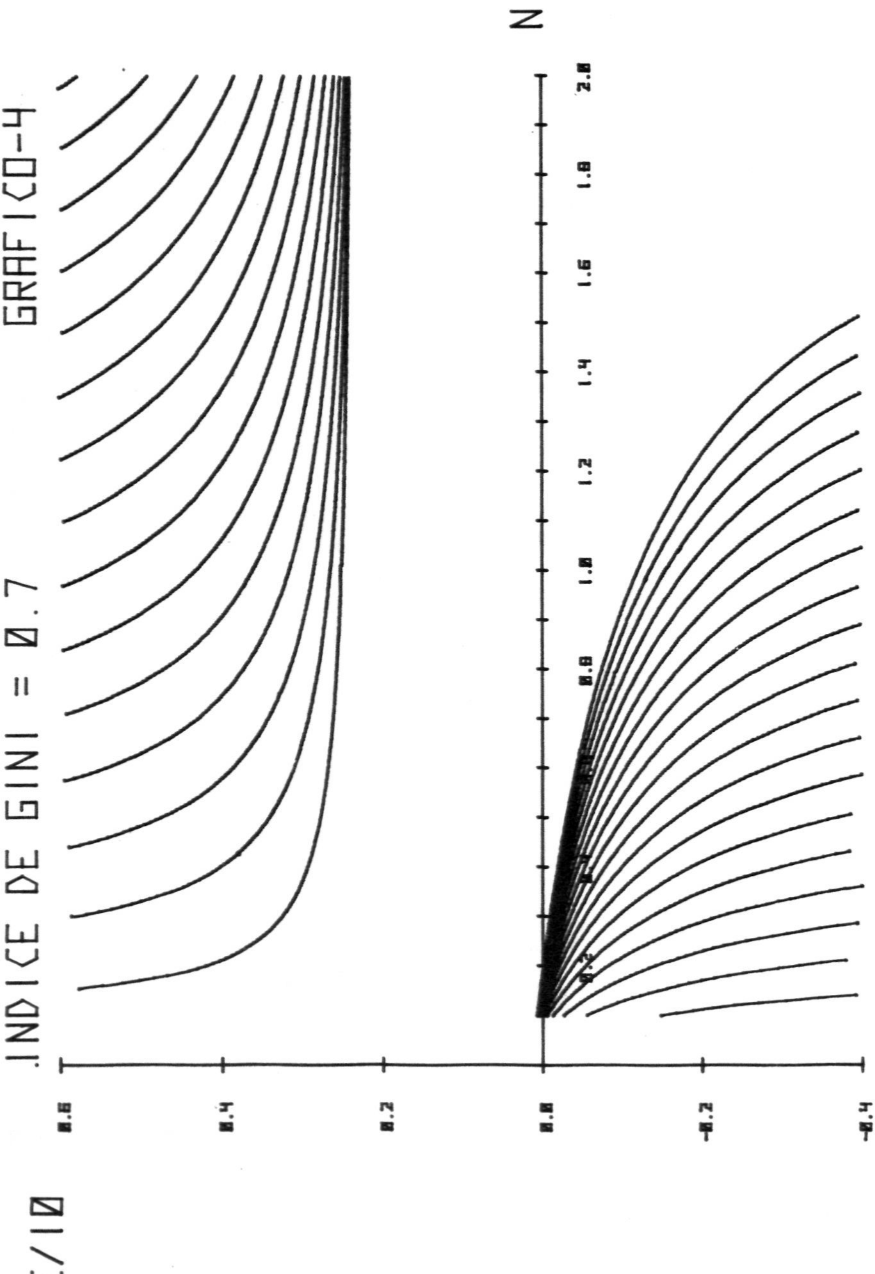


GRAFICO-4
INDICE DE GINI = Ø.7
N
E/1Ø
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4