TENDENCIAS
Revista de la Facultad de Ciencias
Económicas y Administrativas.
Vol. IV. No.2
Diciembre de 2003, páginas 97-123
Universidad de Nariño

DISTRIBUCIÓN Y COEFICIENTE DE GINI, CURVA PARAMÉTRICA DE LORENZ SUGERIDA Y CALCULOS

Por Emilio José Chaves M.*

RESUMEN

Explica paso a paso un método alterno sugerido para ajustar curvas continuas de Lorenz a los escasos cuantiles de distribución, o macro-datos recibidos. Emplea los datos ordenados de rico-a-pobre y determina –como parte del proceso- dos parámetros (*A* y *B*) presentes en la función exponencial que sustenta al modelo. Al aplicar el modelo a los mismos cuantiles (macro-datos) usados en (OGWANG:2002;17), el Indice de Gini estimado (0.4041) estuvo dentro de los topes señalados (0.3883-0.4083), y su valor fue apenas un 0.67% mayor respecto al valor real (0.4014) reportado a partir de los cerca de sesenta mil microdatos de la encuesta original. Desde un punto de vista teórico, el modelo puede ser aplicado para reformular la ley de Pareto y entender el efecto-forma de las diferentes distribuciones, en especial para las distribuciones iso-Ginis, tema relacionado con las llamadas *especificidades regionales* de la distribución, y los análisis de *dominancia*.

PALABRAS CLAVES

Cálculo del Coeficiente de Gini, Curva de Lorenz Paramétrica, Ley de Pareto, Distribución de Ingresos en Colombia.

SUMMARY:

It suggests and explains an alternative Lorenz Curve' fitting method to distribution macro-data. Using data ordered from rich-to-poor two parameters *A* and *B* are determined, which are present in the exponential function that supports the model. When the model was applied to the same macro-data quantils used in (OGWANG: 2002; 17), our estimated Gini's Index (0.4041) was inside their two limits (0.3883-0.4083) and our value was only 0.67% bigger than their reported one (0.4014) from the nearly sixty thousand micro-data of the original survey mentioned by Ogwang. From a theoretical point of view, the model may be helpful to reformulate Pareto's Law and in understanding the shape-effect of different distributions, mainly for *iso-Gini* ones. This topic may be related to other ones like *regional specificities* of distribution and to *dominance* analysis.

KEY WORDS

Gini's Index, Parametric Lorenz Curves, Pareto's Law, Income Distribution in Colombia

1. INTRODUCCION

El tema de la distribución de una variable de interés social es una de las claves para entender cualquier grupo humano desde un punto de vista socio-económico, tanto en las comunidades más pequeñas como en aquellas de dimensión nacional o global. Desde luego, la distribución refleja factores culturales en permanente evolución interactiva como son los de economía-política, clase, etnia, género, religión, mentalidad, medio ambiente, ética y costumbres, los que se expresan de manera diferente en cada grupo, región y nación particulares. Para las ciencias sociales estos estudios, que forman parte de la estadística histórica, permiten conocer mejor nuestro pasado y, desde ese conocimiento, ayudan a diseñar, construir y evaluar caminos colectivos hacia el futuro.

El artículo se centra en el manejo de los datos relativos a la distribución del ingreso, una vez que el investigador en temas socio-económicos los recibe de quien hace las encuestas de campo y las presenta semi-elaboradas. En Colombia esta labor la hace el DANE, institución oficial que realiza periódicamente encuestas de hogares, fundamentalmente de carácter urbano y limitado a trece áreas urbanas, a partir de las cuales publica unos datos semiprocesados sobre el reparto de ciertas variables entre hogares, individuos y otras unidades de población. Una vez obtenidos, los investigadores particulares someten los datos a un proceso de análisis, cálculos e interpretaciones. Parte de éste puede incluir la construcción de algunas gráficas que ilustran la distribución, así como el cálculo del famoso Coeficiente o Indice de Gini. Este último indicador mide el grado de inequidad en el reparto de la variable distribuida dentro del grupo humano objeto de estudio. En un nivel internacional, esos datos semi-elaborados de carácter nacional son compilados, calificados y publicados en Internet y en otros medios por oficinas estadísticas de entidades como la ONU, la CEPAL, la OECD y otras similares enfocadas a ciertas zonas mundiales y grupos de países.

El proceso general comprende a grandes rasgos las siguientes etapas donde participan diferentes entidades y personas:

1) Recolección de micro-datos en el terreno según criterios de calidad adecuados para el proyecto definido.

2) Procesamiento inicial de la información y su presentación en forma de macrodatos (*cuantiles*). Por ejemplo, nos informan que el tercio más rico de la población recibió un 71.92% del ingreso total, el segundo tercio recibió un 20.29%, y el resto más pobre, el último tercio de la población recibió, un 7.79% del ingreso total. Estas cifras, que usaremos en el ejemplo de cálculo, constituyen el conjunto de datos semielaborados a procesar por parte de los analistas.

3) Con ayuda de la teoría estadística –es decir, de un modelo implícito-, aplicada a los datos del segundo punto, se determina y publica el valor del índice de Gini (igual a 0.546 como se verá en el ejemplo), se realizan gráficos que ilustran la situación, y se interpreta el asunto.

Las dos primeras etapas requieren bastante trabajo, pero se cuenta con buena experiencia práctica para resolver muchas de las diversas dificultades que presentan. La tercera etapa se suele realizar en nuestra época con ayuda de programas estadísticos de computador como el POVCAL del Banco Mundial, que puede obtenerse gratis de Internet, para evitar los costos de otras alternativas. Es aquí donde se sitúa este estudio, cuya finalidad está en mostrar un método sencillo y distinto, paso a paso, para realizar la tercera etapa. Basta que los estudiantes universitarios e investigadores sepan manejar una hoja de cálculo y posean la comprensión básica de la estadística relacionada con el tema, y pueden aplicar el modelo a los datos de su interés, mediante un procedimiento comparativamente sencillo en relación a otros que aparecen en la literatura especializada.

En lo que sigue, se aborda brevemente la base conceptual empleada y se presenta el manejo práctico con las cifras del anterior numeral 2), combinado con los comentarios teóricos esenciales. Hacia el final, presentamos resultados de la distribución del ingreso en Colombia, sobre datos de cobertura urbana de las grandes ciudades, para el segundo trimestre de 2003, acompañado de comentarios sobre algunos temas específicos de metodología, interpretación y limitaciones. El tratamiento matemático del tema está simplificado al máximo para facilitar su lectura.

2. BREVE BASE TEORICA DEL METODO

En la última década del siglo XIX El economista italiano Wilfredo Pareto ordenó los ingresos en dinero de rico-a-pobre, *k,* y los graficó contra el número de personas, *x,* que ganaba ese ingreso o más. Como resultado, encontró que había una relación exponencial entre las variables, de la forma $K = x^{C}$, conocida como la ley de Pareto. Investigadores posteriores desarrollaron el tema ordenando la información en sentido inverso, es decir, de pobre-a-rico. También aplicaron a sus datos correlaciones logarítmicas, es decir, para cada punto se obtiene el logaritmo del ingreso *K,* así como el de la población *x*, los que luego se grafican y analizan para observar el comportamiento y obtener conclusiones. La decisión de invertir el ordenamiento fue desafortunada, en mi opinión, porque le restó al tema el valioso aporte conceptual inicial encerrado en la ley de Pareto. Actualmente, casi todos los estudios análogos vienen ordenados de pobre-a-rico, la gente se acostumbró a este formato de gráficas, y con ello, se dificultó el cálculo del índice de Gini de alguna manera.

En un estudio anterior (CHAVES: 1998), se demuestra que a partir de Pareto, la distribución de la curva de Lorenz, ordenada de rico a pobre toma la forma:

$$Y = X^{2-a^*} \quad \textbf{(Ecuación 1)}$$

(*y* es la fracción de la cantidad total distribuida que es percibida por la fracción *x* más rica de la población total; *a** es un parámetro exponencial constante)

El valor del parámetro a* varía entre 1 y 2 para todas las naciones. La Ecuación 1 tiene enormes ventajas para el análisis y permite calcular el índice de Gini y la función cumulativa mediante fórmulas muy sencillas, pero tiene una limitación seria consistente en que al graficar datos reales de naciones de la relación (*2- log Y/ Log X*), se encontró que la supuesta constante *a** resultó ser muy variable y mostraba la forma de una línea ascendente casi recta del tipo:

$a^*(x) = A + B \,.\, x$ (Ecuación 2)

El resultado más interesante al estudiar múltiples datos es que el valor de A es normalmente mayor que la unidad (por lo general superior a 1.3), el valor de B suele estar entre cero y uno, y en todos los casos la suma de A y B es menor a dos. Estas observaciones muy generales permiten decir que *si se trabaja con un ordenamiento de rico-a-pobre*, la Curva de Lorenz se puede re-escribir como:

$$Y = X^{2 - A - B \,.\, x}$$ **(Ecuación 3)**

(X=fracción más rica de población acumulada, Y= fracción de ingreso acumulado que percibe la fracción X de población más rica, A y B son los parámetros propios del caso estudiado y cumplen que A>1 y A+B<2)

Una fase del proceso de cálculo de Gini exige encontrar una curva suave que pase muy cerca de los dos a diez macro-datos (X_i,Y_i) conocidos de la Curva de Lorenz, cumpliendo al mismo tiempo las condiciones matemáticas de validez de ella. En esta etapa la Ecuación 3 resulta útil, y a diferencia de otras propuestas de ajuste, siempre cumple todos los requisitos, además de presentar un ajuste de gran calidad con los datos empíricos. La posible novedad del método aplicado consiste en que el ajuste no se hace sobre la curva de Lorenz directamente, sino sobre el exponente de su fórmula; hecho esto, es fácil proceder con los análisis correspondientes.

Cuando se ordena de pobre-a-rico, la curva de Lorenz es cóncava y ascendente entre los puntos (0,0) y (1,1) de las gráficas. Para tal caso, suelen usarse dos procedimientos de ajuste, aunque existen otros: uno basado en la llamada Ecuación General Cuadrática (EGC) (Villaseñor y Arnold); el otro se llama método BETA, del investigador Kakwuani. Ambos emplean tres parámetros, ninguno de los dos métodos logra siempre pasar por los puntos (0,0) y (1,1) de la curva de Lorenz al tiempo, y sus derivadas no siempre satisfacen ciertas condiciones lógicas exigidas. A su favor puede decirse que constituyen una aproximación aceptable para un problema que no es sencillo. El programa POVCAL ("POVCAL":1998), facilitado por el Banco Mundial, emplea ambos enfoques y luego de hacer correlaciones para las dos curvas, escogen la que mejor se ajuste. El método aquí propuesto es claramente diferente y no presenta esas limitaciones. Es conveniente recomendarlo, a mi juicio, por las siguientes razones:

1) Tiene una base teórica (Pareto) que permite dar sentido explicativo a los parámetros A y B, en especial sobre el tema aún no explorado de las diferentes formas que pueden tomar las curvas de distribución para el mismo Gini (curvas equi-Ginis). También puede ser útil para entender el efecto-forma en los análisis de *dominancia* sobre el tema.

2) Es más sencillo y claro, satisface todas las condiciones de una Curva de Lorenz rico-a-pobre, sólo maneja una función de ajuste, los dos parámetros que usa ayudan a explicar, simular y comparar formas de distribución. El índice de Gini tiene una precisión muy satisfactoria.

3) Trabaja el tema en unidades adimensionales, es decir en fracciones de población, en fracciones de ingreso total, y en número de ingresos medios percibidos. En teoría, esto facilita realizar comparaciones cruzadas entre sociedades y épocas distintas.

4) Conocidos los parámetros A y B, la función Cumulativa de Probabilidad queda definida y de ella salen datos sobre promedios de ingreso de grupos entre dos márgenes, sin necesidad de asumir valores *empíricos*.

5) Mediante transposiciones sencillas se pueden construir las mismas gráficas para el ordenamiento opuesto –el de *pobre-a-rico*-.

3. EJEMPLO NUMERICO DE CALCULO DEL INDICE DE GINI

Paso 1. Supongamos que disponemos de las cifras del ejemplo anterior para tres cuantiles. El siguiente paso es calcular la curva de Lorenz para esos datos, *ordenada de mayor a menor*. Así las cosas, se presentan los resultados en la Tabla 1.

Paso 2. Una vez presentados los datos para una de Curva de Lorenz rico-a-pobre, se procede a calcular los dos parámetros A, y B. Por sencillez, en nuestro ejemplo sólo hay dos puntos datuales internos: P1(0.3333, 0.7192) y P2 (0.6667, 0.9221).

Tabla 1. Datos y Cálculos de las Etapas 1 y 2

Dato	Dato	Cálculo de Datos para Curva de Lorenz de Mayor a Menor	
Cuantil de rico a pobre	% de Ingresos del Cuantil	% Acumulado de Población	% Acumulado de Ingresos
1 33.33%	71.92	33.33	71.92
2 33.33%	20.29	66.67	92.21
3 33.33%	7.79	100.00	100.00

Paso 2.a. Calcular las constantes C1 y C2 de las correlaciones mediante las fórmulas:

C1 = 2 – log(y1)/log(x1) ..... C1= 2 – log(0.7192)/log(0.3333)= 1.7
C2 = 2 – log(y2)/log(x2) ..... C2= 2 – log(0.9221)/log(0.6667)= 1.8

Estos cálculos representan una correlación logarítmica entre ***y***, el ingreso total, y ***x***, la fracción más rica de la población que lo percibe. Como ya dijimos esa correlación se aproxima a una línea recta.

Paso 2.b. Calcular los parámetros A y B resolviendo las ecuaciones simultáneas:

C1 = A + B. X1 y C2 = A + B. X2
1.7 = A + 0.3333 B .... y .....1.8 = A + 0.6667 B

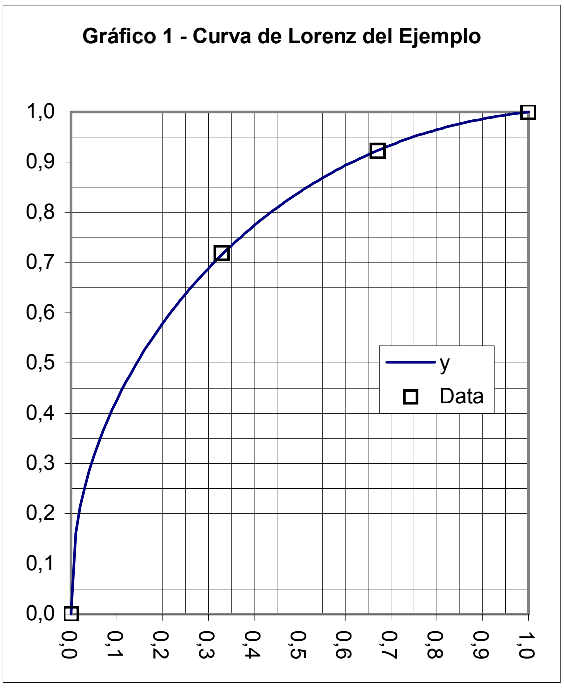
Para estas dos ecuaciones lineales las soluciones halladas son: A=1.6 y B=0.3

En este ejemplo donde sólo hay dos puntos datuales internos los parámetros se derivan de la solución de dos ecuaciones lineales. Normalmente el investigador trabaja con varios más, caso en el cual se efectúa una regresión lineal –como la de mínimos cuadrados- para hallar los parámetros A y B de la recta buscada A+B.x = C. El parámetro A corresponde a la constante de la recta, y el parámetro B a la pendiente. En general, entre más datos puntuales haya, tanto mejor la calidad del resultado.

Paso 2.c. Formulación de la curva de Lorenz ajustada a los datos:

Como ya dijimos, en el modelo la curva de Lorenz ajustada a los datos tiene la forma:

$Y = X^{2-A-B.x}$ **(Ecuación 3)**



Obsérvese que se trata de una función exponencial doble muy particular (la variable X aparece tanto en la base como en el exponente), siempre pasa por los puntos (0,0) y (1,1), y mientras la suma de A+B sea menor a dos, siempre será creciente, tal como lo exige toda curva de Lorenz ordenada de rico a pobre. Cuando el valor B es cero, entonces se obtiene una distribución paretiana uniforme del tipo $y = x^{2-a^*}$, la cual constituye un caso particular muy importante desde el punto de vista teórico, pero poco observada al analizar datos de naciones reales.

Con los valores de A=1.6 y B=0.3 encontrados, la curva de Lorenz del ejemplo sería:

$$\mathbf{Y(x) = x^{2-1.6-0.3X}} \qquad \textbf{(Ecuación 4)}$$

Si calculamos esta función para diferentes valores de x comprendidos entre x=0 y x=1, y la graficamos, entonces podemos visualizar una curva de Lorenz continua, ordenada de *ricos a pobres*, o de *más a menos*, como también se suele llamar. El Gráfico 1 presenta la curva de Lorenz para el ejemplo, así como los puntos datuales iniciales. En el próximo aparte explicaremos el procedimiento para hallar el índice de Gini de esta distribución concreta, el cual arrojó un valor igual a 0.546.

Paso 3. Medición del Indice de Gini

Una vez obtenida una curva satisfactoria, es preciso medir su área para obtener el índice de Gini en un paso posterior. Como la función expuesta parece no tener una integral exacta se aplicó el método de integración por partes hasta su sexto elemento (el procedimiento matemático es algo laborioso). Los resultados fueron confrontados con los obtenidos mediante técnicas numéricas, usando intervalos muy pequeños con ayuda de una hoja electrónica de computador personal; esta contrastación sólo prueba la coherencia interna de los cálculos. Para juzgar su precisión deben usarse datos de una encuesta externa, cuyo cálculo del Gini tenga una precisión incuestionable, y comparar este último con el Gini calculado por el modelo.

Paso 3.a. Medición del Area bajo la curva de Lorenz y del Indice de Gini

Con los valores de los parámetros A y B se calcula el área de la curva de Lorenz (integral entre cero y uno), así como el índice de Gini, tal como se explica y resume en la Tabla 2. Puede observarse en este cuadro que después del cuarto item las contribuciones se hacen cada vez menos significativas; por esta razón se asume que la aproximación es satisfactoria y convergente. La prueba más importante de la calidad del índice de Gini, arrojado por este procedimiento, fue su aplicación a un conjunto de 10 puntos de una distribución de los Estados Unidos *–usado como referente externo*- tomada de (OGWANG:2002;17), en la cual se obtuvo un índice de Gini de (0.4041) que concuerda con los topes señalados por Owang, y es tan sólo un 0.67 % mayor que el Gini real (0.4014) reportado en dicho estudio, luego de procesar los cerca de sesenta mil datos reales de la encuesta. Se aplicaron los mismos diez macro-datos al programa POVCAL, el cual arrojó un valor de (0.3944) para el índice de Gini, también muy cercano al real.

Tabla 2. Cálculo de Area de Curva de Lorenz e Indice de Gini

Ejemplo de Procedimiento de Cálculo

	Datos:	A=	1,6	B=	0,3
	D = Denominador		N = Numerador		Sumandos
tem	Fórmula	Valor V1	Fórmula	Valor V2	V2 / V1
	(3-A)	1,4000	1	1	0,71429
	(3-A)*(4-A)	3,3600	B	0,3	0,08929
	(3-A)*(4-A)*(5-A)	11,4240	$B^2 - B$	-0,21	-0,01838
	(3-A)*(4-A)*(5-A)*(6-A)	50,2656	$B^3 - 3B^2 - B$	-0,597	-0,01188
	(3-A)*(4-A)*(5-A)*(6-A)*(7-A)	271,4342	$B^4 - 6B^3 - B^2$	-0,2439	-0,00311
	(3-A)*(4-A)*(5-A)*(6-A)*(7-A)*(8-A)	1737,1791	$B^5 - 10B^4 - 14B^3 - 5B^2$	0,74943	-0,00107
			L = Sumatoria (Area)		0,772845
			G = 2.L-1 (Indice Gini)		**0.5457**

Como resultado, el índice de Gini del ejemplo es de 0.5457 o sea, de 54.57% cuando se expresa en porcentaje. En esencia, el índice de Gini es el doble del área entre la curva de Lorenz y la diagonal que une los puntos (0,0) y (1,1) para nuestra curva ordenada de rico-a-pobre. Mide qué tan inflada es la curva de Lorenz en proporción a la distribución perfecta y refleja el nivel de inequidad en la distribución. El índice de Gini puede variar entre cero y la unidad; cuando su valor es cero, la distribución es perfecta y la curva de Lorenz se hace igual a la diagonal que une los puntos (0,0) y (1,1), significando que cada poblador percibe el mismo ingreso; cuando su valor es la unidad, la distribución es de máxima inequidad, caso en el cual un solo poblador o familia se apropia de todo el ingreso y los demás no perciben nada, de modo que la curva de Lorenz se transforma en una línea horizontal de altura igual a la unidad. En general, el índice de Gini de las naciones de la época oscila entre 0.3 y 0.7, considerándose que un valor superior a 0.4 es bastante inequitativo.

Cuando se ordena a la inversa, de pobre a rico, la curva de Lorenz toma forma cóncava y va siempre por debajo de la diagonal. Su índice de Gini es el doble del área entre la diagonal y la curva de Lorenz, y resulta idéntico al obtenido con nuestro ordenamiento rico-a-pobre. A partir de la curva de Lorenz rico-a-pobre, o de su fórmula, puede construirse la curva ordenada de pobre-a-rico, e igualmente en sentido inverso.

4. LA FUNCIÓN K DE PROBABILIDAD CUMULATIVA MAYOR/IGUAL

Hay una pregunta frecuente en esto de la distribución de cualquier variable: ¿Cuál es el ingreso que iguala o supera el sector X más rico de una población? O a la inversa, en términos de probabilidad, ¿Dado un ingreso de valor K veces el ingreso medio del grupo, cuál es la probabilidad de que alguien gane una suma igual o mayor a K veces ese ingreso medio grupal?

El modelo propuesto de Curva de Lorenz permite responder rápidamente a esas preguntas. Ocurre que la derivada de la Curva de Lorenz es igual a la función que da el valor *K* (medido en ingresos medios de toda la muestra) para cada valor de X (porcentaje o fracción de la población más rica que gana una cantidad igual o mayor a esa cantidad *K*). A esta función se la conoce también como Función Cumulativa de Probabilidad Mayor o Igual a K, a la que llamaremos *Función K* para simplificar.

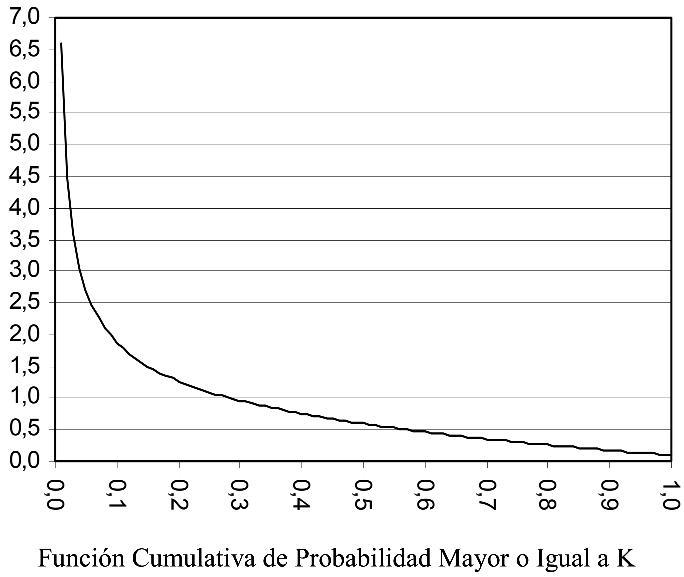
La derivada de la curva paramétrica de Lorenz (Ecuación 3) en el modelo es:

$$K_{(x,\ /\geq\ )} = dy/dx = x^{\,2 - A - B\,.\,x}\,.\,[\,(2 - A - B\,.\,x\,)\,/\,x\ -\ B\,.\,\mathrm{logn}(x)\,]$$
(Ecuación 5)

Cuando esta función es calculada para los diferentes valores de X entre cero y la unidad, y luego tabulada o graficada, entonces podemos responder a la pregunta anterior, así como lograr cierta comprensión visual del fenómeno. Veamos la gráfica resultante para los valores A=1.6 y B=0.3 de nuestro ejemplo, usando intervalos de 0.01 (1%) para X:

Si la pregunta fuera ¿Cuál es la probabilidad de ganar más de un ingreso medio?, de la gráfica podemos observar que es cercana al 28%. Esto indica que en nuestro ejemplo particular, el resto de la población, o sea el 72%, gana menos del ingreso medio del grupo. El ingreso medio del grupo siempre es la unidad, cuando se mide en ingresos medios muestrales, y corresponde al ingreso per cápita de la muestra en otras unidades como dinero, salarios mínimos, etc. Por esta razón, las encuestas hechas así deben informar sobre el IPC de la muestra en su ficha técnica. Si se desea hacer la gráfica en unidades reales (en circulante, o en Salarios Mínimos), es preciso conocer cuánto es el IPCm en esas unidades, y multiplicar el valor K (en ingresos medios del grupo) por el factor de conversión para obtener una gráfica en unidades reales.

Gráfico 2. Curva K en Ingresos Medios
A=1,6 B=0,3



4.a. La Ley de Pareto reformulada

Anteriormente habíamos dicho que la ley de Pareto expresa la distribución cumulativa con la forma: $K = x^{C}$ (Ecuación 1). Pareto estaba muy bien orientado pero no disponía de tanta información como nosotros ahora, en un contexto muy diferente, cien años después. En realidad las cosas son algo más complejas y, como homenaje a Wilfredo Pareto -gran pionero de los estudios sobre distribución-, desde la Ecuación 5 del modelo propuesto podemos decir que en realidad la ley de Pareto adquiere la forma paramétrica:

$K = f(x) \cdot x^{g(x)}$ (Ecuación 6) donde

$F(x) = [\,(2 - A - B \cdot x) / x \;-\; B \cdot \log_e x\,]$ ... y ... $G(x) = 2 - A - B \cdot x$
(A y B son parámetros específicos de cada grupo, $1 \leq A \leq 2$, $0 \leq B \leq 1$, $1 \leq A+B < 2$)

Esta sería una versión más actualizada de la ley de Pareto, ya que las cosas no son tan simples como él lo expuso.

NOTA: Es posible que la expresión G(x) adquiera otra forma no lineal de acuerdo al tipo de correlación logarítmica empleada para analizar los datos. En ese caso habrá otros y tal vez más parámetros, otra será la expresión F(x) y otra la forma de calcular el Indice de Gini. Pero eso no altera la esencia de la Ecuación 6.

5. COMENTARIOS GENERALES SOBRE LA DISTRIBUCIÓN

Conviene anotar que por lo general, en las encuestas de las naciones el ingreso promedio de las muestras, IPCm, suele ser bastante inferior al IPC derivado de las cuentas nacionales y de la población nacional. Esto depende de cómo se midan tanto el ingreso en las encuestas como los grandes agregados macroeconómicos en las cuentas nacionales. Es un tema importante cuyo tratamiento no puede abordarse en este artículo.

Si para esta distribución del ejemplo, y de acuerdo al ingreso medio per cápita en moneda corriente, definiéramos el umbral de pobreza como medio ingreso promedio, y el umbral de miseria como un cuarto del ingreso promedio, entonces, del Gráfico 2 o de la tabla que lo origina, se observaría que el 43% de la población se halla en la pobreza, y que un 19% se puede considerar en condición de miseria. Esta es una definición *relativa*, por cuanto usa proporciones.

Hay otras formas de definir los umbrales de pobreza y miseria; una de las más conocidas es la que define el umbral de pobreza como un ingreso de dos dólares per cápita diarios, y el de miseria como un dólar per cápita diario. Estas definiciones son funcionales como estándar internacional; sin embargo, han sido cuestionadas debido a que el poder de compra de un dólar cambia de país a país, de región a región, y de mes a mes; también suele cambiar la composición de la canasta de bienes básicos de una región con el paso del tiempo. Esta clase de definición es conocida como *absoluta*, por cuanto emplea unidades reales. De todas formas, basta observar en cada lugar –a partir de la experiencia cotidiana- lo que puede hacer una familia de cuatro personas para vivir todos con ocho dólares diarios, y el nivel de vida que esto les permite.

Es de señalar que si el Gráfico 2 incluyera al uno por millón más rico, el ingreso mínimo de esta élite minúscula se hace mil seiscientas veces mayor que el ingreso medio, y necesitaríamos una gráfica 250 veces más alta que la mostrada. Es decir, los super-ricos existen, no son muchos, y hacen lucir como enanos a quienes ganan, digamos unos cinco ingresos medios, que sí alcanzan a aparecer en las encuestas y en el gráfico. Este punto es crucial para entender un gran problema práctico que afecta la recolección de los datos primarios de las encuestas, causado por el ocultamiento de ingresos, bien sea porque la gente los oculta y su efecto es mayor cuando lo hacen los más ricos, o bien sea porque las encuestas no incluyen generalmente a quienes ganan más de treinta ingresos medios. Se trata de una práctica que disminuye el ingreso medio del cuantil más rico y el de toda la muestra, deforma las curvas, rebaja el Gini y en general, le resta credibilidad a las estadísticas oficiales de casi todos los países del mundo (PARENTI:2000).

La Estructura de las Clases Sociales

La curva de Lorenz de la distribución del ingreso nos puede ayudar a definir el concepto de clase en una sociedad dada. Precisar donde termina la clase alta y donde comienza la clase media es tan difícil como saber dónde termina la clase media y dónde empieza la clase baja. La gente percibe ingresos sobre un amplio rango continuo de valores entre un mínimo muy reducido y un máximo muy alto y se hace preciso cuantificar el tema de alguna manera con cifras estadísticas, ya que no basta con hacer una encuesta preguntándole a la gente cuál es su clase social. Por esta razón proponemos la siguiente definición como posible guía general:

La fracción de población más rica que gana un tercio del producto total es clase alta. La fracción que gana el segundo tercio es clase media, y la que gana el tercio restante es clase baja.

Bajo esta definición, si miramos el Gráfico 1, encontraríamos que la clase alta es menos de un 5% de la población, la clase media es alrededor de un 23%, y la clase baja alrededor de un 72%, dentro de una sociedad como la del ejemplo trabajado. Estas proporciones varían según el país y la época estudiada.

Diferencias Regionales en el IPC y la Distribución

Es posible que un país muy pobre posea la misma estructura de clase y el mismo índice de Gini que un país muy rico, a pesar de sus claramente diferentes ingresos per cápita medidos en una moneda común. Es de sentido común reconocer que si un país no produce lo suficiente para alimentar a su población, por más equitativa que sea la distribución del ingreso no se supera la necesidad sin mejorar la producción de satisfactores. En este caso particular, aparte de las medidas de urgencia para atender a los más necesitados en el corto plazo, producir más para consumir la cantidad apropiada es la respuesta adecuada. Pero si una nación produce los bienes básicos suficientes y genera excedentes, nada justifica las carencias que enfrentan los sectores que están por debajo de la línea de pobreza, y todo se reduce a un problema práctico en el campo ético-político de la distribución.

Es común que dentro de una nación existan varias regiones con ingresos per cápita diferentes, es también normal que cada región arroje su propia curva de Lorenz diferente a las demás. En estos casos, es necesario realizar un procedimiento de cálculo especial para obtener los resultados consolidados de la nación integrando los datos regionales. Su cálculo combina tres elementos: 1) el IPC de la encuesta de cada región; 2) las curvas de distribución de cada región, cada una con sus parámetros propios A y B; y 3) la fracción de población de cada región respecto al conjunto. En base a estas ideas muy resumidas, es posible realizar este cálculo para integrar varias regiones, pero su descripción detallada desborda los propósitos de este artículo. En la práctica, se hacen encuestas regionales por separado y se integran posteriormente, en lugar de realizar una encuesta nacional, lo que resultaría más complejo de organizar.

Cuando la oficina estadística entrega datos de cuantiles que conducen directamente a la Curva de Lorenz, tal como vimos en el ejemplo trabajado, es muy sencillo calcular el índice de Gini y las gráficas esenciales, porque implican que el proveedor de datos maneja y entrega los valores globales de la muestra y los subgrupos. En este formato se presentaban los resultados de Colombia hasta hace unos dos años. Sin embargo, a comienzos del año 2002 dicho formato fue cambiado, y ahora se llama Encuesta Continua de Hogares, ECH, la cual se realiza sobre 13 áreas para las principales ciudades del país, cada trimestre. Este tipo de datos informa el número de unidades de ingreso o de gasto en grupos que perciben entre ciertos márgenes inferior y superior de salarios mínimos legales (SM), es decir, aporta datos para la función K. Por ejemplo, la Tabla 3, referida a la ECH de abril a junio del año 2003, resume cómo vienen ahora los datos de distribución en los reportes del DANE.

6. DATOS POR GRUPOS ENTRE MARGENES DE INGRESO
Colombia, grandes áreas urbanas, 2003

Tabla 3 - Formato de la Encuesta Contínua de Hogares, ECH.		
Grupos de Ingresos en		Número de Personas
MAS DE	HASTA	
0	0,5	4.582.977
0,5	1	3.302.326
1	1,5	3.094.907
1,5	2	1.036.470
2	4	1.202.658
4		553.130
	Total	13.772.468

Trabajar con este formato de datos es más difícil que con los de curvas de Lorenz, ya que entregan unos cuantos puntos de la curva *K* de distribución acumulada, sin dar información del ingreso medio de cada sub-grupo, ni del conjunto de la muestra, a diferencia de encuestas análogas en otros países. Bajo estas condiciones de información incompleta y pocos puntos datuales, necesariamente deben asumirse ciertas premisas sobre la forma de la curva, pues de otra manera no se logra resultado alguno. Sin embargo, con ayuda del modelo paramétrico descrito se puede estimar una aproximación de las curvas, del índice de Gini, y de la relación Ingreso Medio de la Muestra/Salario Mínimo. Como consecuencia de este formato incompleto de datos, los resultados pueden ser algo inexactos y variables de acuerdo a los criterios que se empleen para escoger las premisas faltantes. En tales condiciones, los resultados y Ginis obtenidos no suelen ser aprobados por las bases de datos internacionales, como el WIID (World International Income Database) a cargo de la Universidad de las Naciones Unidas, entidad que forma parte de la ONU. (ONU, WIDER: 2000)

Debido a lo anterior, no se justifica detallar aquí el procedimiento algo complejo empleado para el caso en cuestión. En esencia, el método emplea las propiedades del modelo paramétrico sobre las relaciones que se dan entre los valores máximo, mínimo y medio dentro de cada sub-grupo, y mediante el método de prueba y error se ensayan diversas combinaciones de valores de A y B, hasta encontrar un ajuste *relativamente aceptable*. Debido a que no se dispone del ingreso medio de la muestra, muchas parejas *A,B*, -dentro de un rango continuo pero pequeño de combinaciones- producen muy buenas correlaciones, pero no hay de donde elaborar criterios para definir cual de ellos es la mejor solución de las muchas posibles, debido a que falta información. En este caso especial, la existencia de una buena correlación no basta como garantía de precisión, ni puede aducirse como prueba de calidad metodológica.

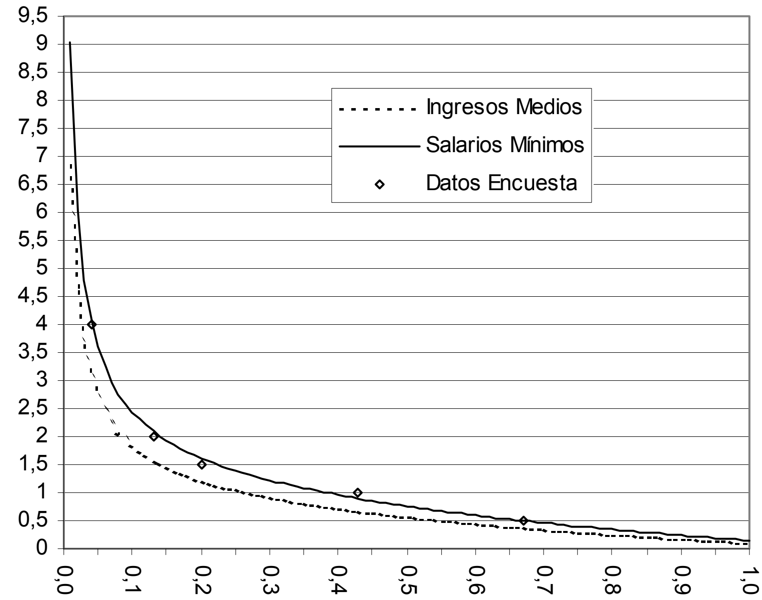
Tabla 4. Resultados Estimados según ENH-2003

Parámetro A	1,6180
Parámetro B	0,2700
Indice de Gini	0,5543
Salario Mínimo Col. $ / mes labor, 2003	369500
Salario Mínimo en US $ / día-labor	4,3273
Ingreso Medio de Muestra Col.$/mes	489111
Relación Ingreso Medio / Salario Mínimo	1,3237
Fuente de Datos:	
DANE, ENCUESTA CONTINUA DE HOGARES - ABRIL A JUNIO DE 2003	
CUADRO NO. 63 (TS-ECH-TAB-63)	
POBLACION OCUPADA POR RANGOS DE INGRESOS LABORALES MENSUALES	
SEGUN SEXO Y RAMA DE ACTIVIDAD ECONOMICA	
*Estimativos hechos por el autor. **Pesos/US Dólar a Mayo 15/2003= 2821,08	

Los valores generales de esta distribución obtenidos con el modelo, y bajo las limitaciones expuestas, se resumen en la Tabla 4, con la advertencia adicional de que debido a la cobertura básicamente urbana y limitada de la encuesta, ésta no es representativa de una verdadera distribución nacional de ingresos, ni debe usarse para comparaciones con resultados provenientes de otras encuestas, sin antes verificar que cumplan las exigentes normas que lo permiten.

Podemos ver en la misma Tabla 4 que se obtuvo un índice de Gini de 0.554, valor inferior al obtenido de la Encuesta Nacional de Hogares de 1994-1995 del DANE que sí entregaba los datos completos, para la cual encontramos un índice de Gini algo mayor a 0.6. Dejando de lado el hecho de que las dos encuestas no son comparables, bajo cualquier punto de vista ambas cifras son muy altas y reflejan una gran inequidad social en Colombia, semejante a la de Brasil. Con los valores obtenidos de los parámetros A y B, se construyó la curva de distribución *K vs. X* del Gráfico 5. Este muestra a

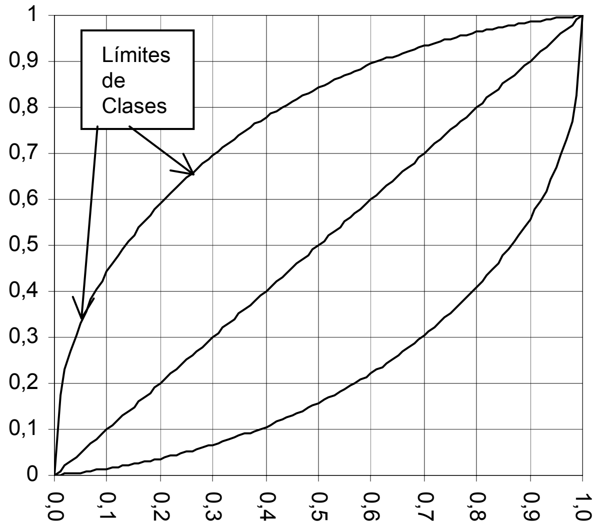
Gráfico 5. Distribución Cumulativa
(Rico a Pobre, en SM e IM) A=1.6180 B=0.27



manera de ejemplo el ajuste logrado entre la curva de distribución K y los 5 puntos datuales, el cual puede resultar satisfactorio a la vista, pero como ya

fue comentado, debe tenerse en cuenta que se trata de una aproximación. En dicho Gráfico se presentan dos curvas K : la superior, expresada en unidades de Salarios Mínimos (SM), tal como vienen los datos, mientras la segunda curva inferior está en ingresos medios muestrales (IM).

Gráfico 6. Curvas de Lorenz
A=1,618 B=0,2700 Gini=0,554



•FUENTE: DANE, Encuesta Continua de Hogares, 2o. Trimestre de 2003, 13 Ciudades, CUADRO NO. 63 (TS-ECH-TAB-63), sobre Poblacion Ocupada Por Rangos De Ingresos Laborales Mensuales

** Elaborada por autor sobre Datos de Fuente

NOTA: La curva superior va ordenada de Rico a Pobre, la inferior de Pobre a Rico

Para las cifras del ejemplo se encontró que el Ingreso Medio es 1.35 veces el Salario Mínimo. Si se observa el Gráfico 5, cuando x es cercano a 0.38, la curva superior tiene un valor aproximado a 1 Salario Mínimo, lo cual nos indica que el 62% de los hogares de la muestra tuvieron un ingreso que no sobrepasó el salario mínimo de 4.33 dólares por hogar al día.

Usando también los valores de A y B, se construyeron las Curvas de Lorenz para nuestro ordenamiento de rico-a-pobre y para el opuesto, de pobre-a-rico, más empleado en la literatura sobre el tema. Pueden verse en el Gráfico 6.

Al observar el Gráfico 6, es interesante resaltar que para el universo de la muestra, la cual sólo abarca 13 grandes áreas urbanas del país, aproximadamente un 5% de la población conforma la clase alta, un 22% la clase media, y el restante 73% la clase baja, de acuerdo a la definición sugerida de clases sociales, donde cada una de ellas controla un tercio del ingreso total de la muestra. Dentro de cada clase social pueden existir diferencias considerables entre el ingreso máximo y el mínimo de ellas, ya que operan sobre un espectro variable de valores reales de ingreso. La teoría estadística busca simular esa situación mediante las curvas ajustadas, aunque en la realidad el fenómeno ocurre en forma de valores discretos, a menudo discontinuos, pero al observarse para grandes números de población adquieren un aspecto similar a las curvas exhibidas.

Como dato final para estudios más avanzados, damos la fórmula de la segunda derivada de la Curva de Lorenz para el modelo, la cual debe ser siempre negativa y monotónicamente creciente cuando el ordenamiento es de rico-a-pobre.

$$\ddot{Y} = dK/dx = \{ [(2-A)/x - B(1+ \log x)]^2 - (2-A)/x^2 - B/x\} \; x^{2 - A - Bx}$$

(Ecuación 7)

CONCLUSIONES

Cualquier modelo es una abstracción elaborada sobre datos tomados de la realidad. A su vez, para seleccionar y medir esos datos es necesario asumir ciertas premisas abstractas que reglamentan el método empleado para producir eso que llamamos *datos*. Al final, un modelo no es más que una abstracción construida sobre datos que a su vez son el resultado de otras abstracciones. Eso no significa que construir modelos sea una tarea estéril e inútil a pesar de los elementos tautológicos y sesgos subjetivos que forman parte normal de ellos. Por el contrario, ellos han jugado un importante papel en el avance del conocimiento, la ciencia, la técnica aplicada y el mismo arte de construir modelos. Detrás de los modelos hay un esfuerzo continuo de interacciones entre teorías y prácticas, sujeto a permanentes revisiones por parte de quienes los diseñan e imponen y quienes los disfrutan o padecen. En el caso de la distribución de variables de interés socio-económico, el último referente de todo modelo llevado a la práctica es la gente y su experiencia vital. Por ello estos modelos necesariamente se prestan para el debate sobre temas de economía, sociología, ética y política; y como es de esperar, las diversas interpretaciones difícilmente pueden ser neutrales o ajenas a intereses concretos de clase, etnia, género, región, cultura, etc..

*Aquí hemos mostrado ejemplos limitados al ingreso, pero en general, el mismo método puede extenderse a muchas otras variables de interés social.

* No sólo es importante la cobertura de población para obtener una encuesta representativa. Es fundamental tener claridad sobre los criterios empleados para definir la población y la medición de la variable objeto de reparto. La forma como se definen afecta al índice de Gini medido, así como a la posibilidad de comparar dos encuestas separadas en el espacio o en el tiempo.

* Las encuestas oficiales disponibles en el caso de Colombia sólo cubren los grandes centros urbanos y sus vecindarios rurales, dejando por fuera las periferias más apartadas, de modo que son muy pocas las encuestas representativas en cuanto a cobertura nacional.

* Es preciso subrayar la responsabilidad del investigador a la hora de realizar los estudios de distribución, y en especial, cuando se hacen análisis comparativos entre dos estadísticas diferentes de distribución. Conviene tener una idea tan clara como resulte posible sobre los criterios y conceptos básicos que fundan las encuestas, los datos y los métodos de análisis empleados.

* Es posible realizar en este campo investigaciones muy valiosas de carácter local o regional que suelen abarcar desde la encuesta hasta su procesamiento posterior. En este caso donde lo normal es que el investigador conozca en detalle los microdatos, el total repartido y la población de la muestra, no debe haber problemas para desarrollar datos compatibles con la Curva de Lorenz, de modo que se puede aplicar el método explicado con toda confianza.

*Cuando se publiquen datos de muestras estadísticas de distribución, sea en forma de cuantiles, o ya sea como grupos entre márgenes de la variable estudiada, se recomienda informar siempre el valor promedio de toda la muestra, y en lo posible, el de por lo menos un subgrupo. Cuando se conoce el ingreso medio de al menos un subgrupo, con el modelo explicado puede estimarse inmediatamente el ingreso medio de la muestra, así como calcular los datos para la Curva de Lorenz, de modo que se puede usar el método básico explicado con toda confianza.

* Cuando nos dan información de algunos puntos en la Curva K, sin darnos los promedios de los subgrupos, ni el promedio total, se puede aplicar un método aproximado, pero no hay garantía de suficiente precisión por información incompleta.

* Es necesario sugerirle al DANE que incluya el valor promedio muestral en todo reporte de distribución que publique, así sea un estimativo provisional, especialmente en los datos de distribución que forman parte de las recientes Encuestas Continuas de Hogares, ya que -desde su implantación en el año 2002- no fue posible encontrar esa información en ninguna de las tablas de distribución consultadas.

* Cinco temas apenas mencionados requieren una exposición más profunda por separado:

1) El procedimiento cuando los datos entregan algunos puntos de la curva K de distribución cumulativa y tan sólo el ingreso promedio de un subgrupo.

2) El procedimiento para integrar datos de regiones diferentes en su IPC y en su curva de Lorenz, de modo que se obtenga un resultado consolidado.

3) La forma de las Curvas de Lorenz y su relación con los parámetros A y B, en especial para curvas diferentes con idéntico valor del Indice de Gini (isoGinis).

4) La discrepancia entre el Ingreso Medio obtenido en las encuestas de distribución, y el IPC derivado de las cuentas nacionales oficiales.

5) La distorsión causada por el ocultamiento de ingresos en las encuestas, en especial por los más afluentes, sobre las curvas y el índice de Gini calculado.

BIBLIOGRAFIA

ATKINSON, Anthony B., BRANDOLINI, Van Der Laan, SMEEDING. **"Producing Time Series Data for Income Distribution: Sources,**

Methods, and Techniques”. Canberra Group Meeting, Luxemburg, Mayo 15-17, 2000. En Internet:
http://www.lisproject.org/links/canberra/luxembourg/chapter10.pdf

BOURGUIGNON, FRANCOIS. **“The Poverty-Growth-Inequality Triangle”.** Banco Mundial, 2004. En Internet:
http://econ.worldbank.org/files/33634_PovertyInequalityGrowthTriangleFeb24.pdf

BOURGUIGNON, FRANCOIS **“The Growth Elasticity of Poverty Reduction: Explaining Heterogeneity Across Countries and Time Periods”**. Banco Mundial. 2003. En Internet:
http://econ.worldbank.org/files/32322_growth_elasticity.pdf

CHAVES, EMILIO JOSÉ, **"Pareto and Income Distribution".** Social Sciences Research Network, 1996. En Internet:
http://ssrn.com/abstract=221368

DATT, GAURAV. **“Computational Tools For Poverty Measurement and Analysis”**, International Food Policy Research Institute, FCND DISCUSSION PAPER NO. 50, Oct.1998, en Internet: www.ifpri.org/divs/fcnd/dp/papers/dp50.pdf

LITCHFIELD, JULIE A**. “Inequality: Methods and Tools”.** Text for World Bank’s Web Site on Inequality, Poverty, and Socio-economic Performance. En Internet:
http://www.worldbank.org/poverty/inequal/index.htm

MILANOVIC, BRANKO, **“True World Income Distribution, 1988 and 1993: First Calculations, Based on Household Surveys Alone”**. Banco Mundial. 1999. En Internet:
http://econ.worldbank.org/view.php?type=5&id=978

ONU, **“POVCAL”**. En Internet: www.worldbank.org/lsms/tools/povcal/

OGWANG, TOMSON. **“Bounds of the Gini Index Using Sparse Information on Mean Incomes”**, 2002, en Internet:
www.econ.nyu.edu/iariw Pág.14

PARENTI, MICHAEL. **“The Super Rich Are Out of Sight”**, 2000. En Internet:http://www.michaelparenti.org/Superrich.html