

## **CURVAS FUNCIONALES DE LORENZ: ANÁLISIS DATUAL E INFERENCIAS**

**Emilio José Chaves<sup>1</sup>**

### **RESUMEN**

Enfoca las ideas pioneras de Pareto y Lorenz sobre desigualdad de rentas desde la perspectiva de su ordenamiento datual. De allí plantea un método no-ortodoxo para inferir curvas funcionales de Lorenz (CL) *no-paramétricas* a partir de datos empíricos puntuales o agrupados, en base a dos premisas principales: 1) Ordenamiento datual descendente *rico-a-pobre* (RaP); 2) Análisis datual logarítmico para inferir un Vector-Función estructural, clave por su presencia tanto en la forma funcional de la CL, como en la Función de Distribución Acumulativa (FDA) derivada. Aparentemente esta área estadística requiere una revisión teórica. Por su significado en metodología de investigación para diversas disciplinas, las ideas aquí consignadas quizás podrían ayudar en el futuro para estructurar una nueva teoría general, otro manejo datual práctico y otra enseñanza de las funciones en el sector de distribuciones univariadas continuas, dentro del contexto de las CL.

**PALABRAS CLAVES:** Curvas Funcionales de Lorenz, Métodos No-Paramétricos de Inferencia Estadística, Distribuciones Continuas Univariadas, Ordenamiento Descendente. .

**JEL CLASSIFICATION:** C14, C46, D63, N01

### **ABSTRACT**

It focuses Vilfredo Pareto and Otto Lorenz contributions on income inequality around their data ordering perspective. As a result, the paper elaborates an heterodox method to infer *non-parametric* functional Lorenz Curves (LC) from scarce empirical quantils and/or grouped data, based in 2 main points: 1) Descending data ordering; 2) Logarithmic data analysis to infer a structural *functional-vector*,  $F(x)$ , key component present in

the obtained functional LC, and in the functional Cumulative Distribution Function (CDF) derived curve. Apparently, it is time to recommend a complementary revision of this sector of statistics. Given its significance for research methods in several fields, ideas here advanced might be of help in the future to structuring possible new trends in general theory, practices and teaching procedures of this sector of continuous univariate distributions, inside the conceptual framework of the Lorenz Curves.

**KEY WORDS:** Functional Fitting Lorenz Curves, Statistical Inference Methods, Continuous Univariate Distributions, Descending Ordering.

**JEL CLASIFICACION:** C14, C46, D63, N01

## INTRODUCCIÓN

La Curva de Lorenz (CL) es una de las gráficas más poderosas en el campo estadístico; cautiva por su sencillez, por el poder de síntesis de su imagen intuitiva, y por su potencial analítico. En esencia, la CL grafica la fracción acumulada de una variable aleatoria versus la fracción acumulada de población receptora de esa variable repartida. Normalmente el investigador en distribución de ingresos cuenta con pocos datos (entre cuatro y quince puntos) para su curva de Lorenz, y todos ellos caben dentro de un gráfico cuadrado de uno por uno, esbozando una curva creciente que arranca del punto (0,0) y llega al punto (1,1) luego de atravesar los puntos tatúales<sup>2</sup> intermedios. En ese momento, el investigador necesita expresar en lenguaje matemático una curva continua que pase por todos esos puntos, que sea creciente, y sea portadora de un nivel satisfactorio de precisión e isomorfismo respecto al fenómeno distributivo real. De esa forma funcional de ajuste depende fuertemente su interpretación posterior, los cálculos de indicadores recomendados, y la visión integral de la distribución empírica estudiada. De no lograrlo, aún le queda la opción de trabajar con los miles de datos de la encuesta original, lo cual exige ordenadores más poderosos, más tiempo, grupos de especialistas mayores, otras técnicas estadísticas, otros costos y otros riesgos diferentes que también pueden afectar la calidad y precisión de los resultados.

Entonces ya se hace fácil introducir criterios para determinar qué sectores de la población están incluidos o excluidos del bienestar o malestar asociados a la variable, así como responder a otras preguntas de interés propias de cada campo de investigación. Aunque esta técnica es aplicable a diversos campos de la ciencia, aquí es referida preferentemente a entornos de distribución socio-económica, razón por la cual términos como *pobre, rico, ingreso, ...* son extrapolados aquí de manera general aunque toda la teoría es aplicable a campos

donde no se emplean esas palabras. Por todo ello, es preciso pedir la comprensión y disculpas anticipadas del lector.

La comunidad de estudiosos y conocedores de la estadística ha propuesto numerosas familias de funciones de ajuste a las CL, en especial para el tema del reparto de ingresos. Esos algoritmos suelen presentar diversas limitaciones: algunos son poco exactos, otros contradicen las premisas lógicas mínimas de las CL; otros impregnan con su forma implícita el mismo fenómeno real que buscan describir, deformándolo en ciertos sectores; otros son demasiado complejos de manejar aún por especialistas; otros solo funcionan con un espectro muy reducido de distribuciones y series datuales.

Una característica virtualmente común es que esos algoritmos están diseñados en su inmensa mayoría para representar el ordenamiento de la población en sentido ascendente, *pobre a rico*, PaR, caso en el cual la CL crece por debajo de la diagonal del cuadrado. El método aquí propuesto es el inverso: ordenar primero los datos de rico a pobre y trabajar el tema solo con ese enfoque descendente. Al avanzar la exposición se harán visibles las razones que justifican esa decisión. Es preciso enfatizar que toda la investigación resultante descansa en ese cambio de perspectiva: Si desde mi orilla del río no puedo ver el bosque que tengo junto a mí, no tiene sentido culpar a los árboles vecinos que me impiden verlo con claridad; más logro si decido atravesar el río para ver el bosque desde la otra orilla. Cambiar de perspectiva complementa la visión global e impide que predomine una visión unilateral. En este caso de análisis estadístico la perspectiva hace la diferencia y favorece no solo a quienes investigan la distribución de ingresos, sino a aquellos que investigan temas análogos en otros ámbitos, desde otras ciencias y disciplinas.

En el campo internacional, el tema aquí tratado es de actualidad cuando se mira el sistema-mundo económico y político surcado por muchas desigualdades y exclusiones. Desde hace décadas, organismos como el Banco Mundial, la ONU y su PNUD, el FMI, la CEPAL y otras entidades internacionales financian batallones de especialistas y equipos regionales dedicados a estudiar el tema de la desigualdad, la pobreza, los indicadores del desarrollo humano, el crecimiento, el desarrollo económico y la dinámica del progreso socio-económico y cultural, la relación entre crecimiento y desigualdad, etc.. Dentro de los servicios que ofrece el Banco Mundial se encuentra el programa POVCAL, el cual significa literalmente *calculadora de pobreza* (Poverty Calculation). Si el usuario lo consulta a través de Internet, puede descargarlo, introducir sus datos de distribución de ingresos según las instrucciones, y obtener resultados y gráficos. El único problema es que los algoritmos implícitos en su diseño no son los mejores para obtener resultados precisos y no siempre cumplen

las condiciones lógicas mínimas exigidas, razón por la cual POVCAL ha sido objeto de duras críticas por parte de estudiosos del asunto (Minoui y Reddy; 2008).

Con el enfoque propuesto, cualquier investigador que sepa usar una hoja de cómputo electrónico con cierta destreza mínima, puede aplicar con toda confianza los cálculos y gráficos aquí descritos a los datos propios de su investigación. Sólo requiere ser muy consciente y claro sobre las inevitables premisas que adopta y sobre sus implicaciones, en cuantos componentes normales de los procesos de inferencia. Hace algunos años publiqué un par de ensayos sobre el tema (Chaves, 2003), donde proponía mirar el asunto desde el ordenamiento RaP y cuestionaba el POVCAL. Lo que parecía una idea apenas útil en el campo de la economía-aplicada se convirtió gradualmente en el eje de una propuesta de metodología investigativa portadora de un enfoque alternativo en el planteamiento teórico de algunos subtemas de estadística matemática sobre distribuciones continuas de variables aleatorias positivas. Me refiero en particular al análisis datual y a los procesos de inferencia a partir de datos escasos, bien sea en forma de algunos cuantiles completos –el caso más sencillo-, o a partir de histogramas agrupados de frecuencias simples –el caso más complejo, donde no se conoce el valor medio de la muestra, y tan solo conocemos los límites inferior y superior de cada grupo y su respectiva frecuencia de ocurrencia-.

La propuesta ofrecida en el artículo es un desarrollo más general y menos restrictivo que la planteada en 2003. Contiene las bases para considerar una eventual renovación teórica de la estructura de las distribuciones continuas univariadas, de modo que puedan ser entendidas desde el contexto unificador de las CL, donde se trabaja con fracciones abstractas, en unidades adimensionales, tales como las empleadas cuando manejamos números fraccionarios puros. La primera sección justifica el cambio de perspectiva analítica. La segunda parte explica la fundamentación matemática y datual de la función estructural particular, una especie de vector portador de las claves que le dan su forma final a cada CL particular; a ese vector lo llamamos la función  $F(x)$  y está presente en todas las fórmulas y análisis derivados, así como en la elaboración de distribuciones estadísticas artificiales, tema importante que se esboza rápidamente. En otra sección analizamos de manera muy rápida ejemplos concretos de inferencia y aspectos prácticos al elaborarlas.

Dado que casi la totalidad de la literatura del último siglo sobre el tema está diseñado y redactado desde la óptica *PaR*, no son muchas las referencias bibliográficas útiles para apoyar el punto de vista aquí sustentado. Por esa razón los lectores acostumbrados al planteamiento tradicional pueden encontrar sorprendente que este enfoque no haya sido explorado por la comunidad de estadísticos matemáticos, a pesar de que Vilfredo Pareto, el pionero de estos estudios, sí lo hiciera en sus primeros estudios de 1897, así como el famoso economista

polonés Oskar Lange en 1958 en las quince páginas de su libro de introducción a la econometría que dedica a los aportes de Pareto.

## 1. FUNDAMENTACIÓN DE LA PROPUESTA

**Tabla I. Datos de Pareto con Ordenamiento RaP  
Distribución Cumulativa Inversa GB-Irlanda 1893-1894**  
Nota inferior: Ingreso en Libras ( $x$ ) y número de ingresos *mayores* de  $x(N)$

Table 1  
Reverse Cumulative Distribution of Income in Great Britain  
and Ireland for Years 1893 to 1894

| Income (£) |          | Great Britain |          | Ireland |          |
|------------|----------|---------------|----------|---------|----------|
| $x$        | $\log x$ | $N$           | $\log N$ | $N$     | $\log N$ |
| 150        | 2.18     | 400,648       | 5.60     | 17,717  | 4.25     |
| 200        | 2.30     | 234,985       | 5.37     | 9,365   | 3.97     |
| 300        | 2.48     | 121,996       | 5.09     | 4,592   | 3.66     |
| 400        | 2.60     | 74,041        | 4.87     | 2,684   | 3.43     |
| 500        | 2.70     | 54,419        | 4.74     | 1,898   | 3.28     |
| 600        | 2.78     | 42,072        | 4.62     | 1,428   | 3.15     |
| 700        | 2.85     | 34,269        | 4.53     | 1,104   | 3.04     |
| 800        | 2.90     | 29,311        | 4.47     | 940     | 2.97     |
| 900        | 2.95     | 25,033        | 4.40     | 771     | 2.89     |
| 1,000      | 3.00     | 22,896        | 4.36     | 684     | 2.84     |
| 2,000      | 3.30     | 9,880         | 3.99     | 271     | 2.43     |
| 3,000      | 3.48     | 6,069         | 3.78     | 142     | 2.15     |
| 4,000      | 3.60     | 4,161         | 3.62     | 88      | 1.94     |
| 5,000      | 3.70     | 3,081         | 3.49     | 68      | 1.83     |
| 10,000     | 4.00     | 1,104         | 3.04     | 22      | 1.34     |

Note: Income in £ ( $x$ ) and number of incomes greater than  $x$  ( $N$ ).  
Source: Data from Pareto (1897, §958).

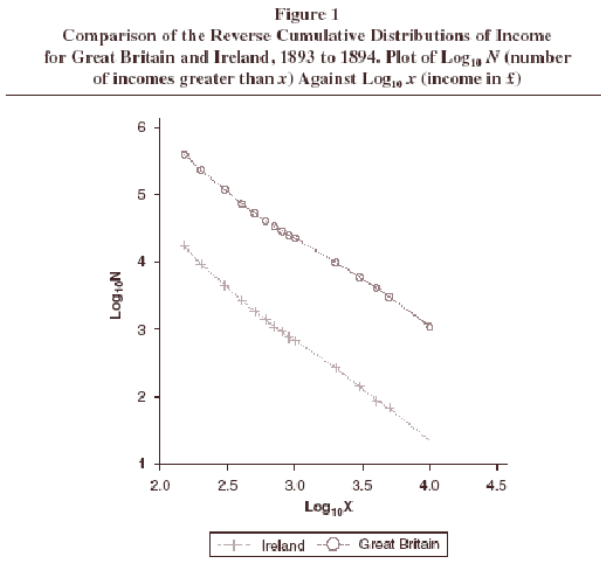
FUENTE: Nielsen, François. Module 8 – Vilfredo Pareto. 2007.  
<http://www.unc.edu/~nielsen/soci250/m5/soci250m5.pdf>

### 1.1 ORDENAMIENTO DE LOS DATOS EN LOS TRABAJOS DE PARETO Y DE LORENZ

Fue el economista italiano Vilfredo Pareto, quien propuso las primeras gráficas sobre la relación estructural entre el logaritmo del ingreso mínimo y el logaritmo de la población acumulada que supera ese ingreso, lo cual requiere datos previamente ordenados RaP. En la Tabla I, podemos ver una fotocopia de un documento del propio Pareto, en 1897, sobre datos tomados del Reino Unido para 1893-94, unos ocho años antes de que el norteamericano Otto Lorenz propusiera hacia 1905 un gráfico que con el tiempo llamaría la atención, sería estudiada, transformada y bautizada con el nombre de Curva de Lorenz (CL). La Tabla I contiene los datos de ingresos de los sectores más afluentes del dúo Inglaterra-Irlanda, así como el manejo logarítmico y la interpretación que hizo Pareto de ellos (Nielsen; 2007). En la nota de pie de página de la Tabla I puede leerse al traducir del inglés la frase “ingreso en £( $x$ ) y número de ingresos *mayores que*  $x(N)$ ”, frase que únicamente

puede invocarse cuando se trabaja desde un contexto de datos previamente ordenados de Rico-a-Pobre, o sea de datos clasificados en orden descendente.

**Gráfico I. Fococopia de Original de Pareto  
(Datos de Cuadro I)  
"Distribución Reversa Cumulativa de Ingreso" Log 10 N vs. Log 10  
de Ingreso  $\geq K_0$  (UK-Irlanda 1893-1894)**



Source: Redrawn from Pareto (1897, §958, Figure 47). See Table 1 for data.

FUENTE: Nielsen, François. Module 8 – Vilfredo Pareto. 2007.  
<http://www.unc.edu/~nielsen/soci250/m5/soci250m5.pdf>

**Notas de autor:** 1) Pareto usa datos RaP de la distribución acumulativa de ingresos *mayores a  $K_0$* , 2) le llama *Reversa* de la distribución acumulativa porque los ingresos  $K$  van en el eje horizontal  $X$ , y la población  $N$  acumulada en el eje vertical  $Y$ . 3) Al girar el gráfico de derecha a izquierda 180 grados como hoja de un libro, y luego 90 grados en sentido del reloj se da la gráfica normal  $K(N) \geq$  ordenada RaP.

Y es sensato que así ocurriera, ya que era más fácil entonces reunir los datos de altos ingresos de gente poco numerosa y relativamente conocida, antes que reunir una gran masa de datos de bajos ingresos entre los sectores más pobres, y además ordenarlos de pobre a rico. En los datos empleados por Pareto no aparece el sector de bajos ingresos con

el mismo nivel de detalle que el sector opuesto de altos ingresos, el cual superaba las 150 Libras de ingreso anual. Hoy sabemos que un gran sector popular vivía con ingresos muy inferiores a ese nivel durante la época de los datos: 1893-1894.

Hace medio siglo, el economista polaco Oskar Lange hizo un excelente análisis crítico sobre *la curva de Pareto* (Lange, O.; 1958:: 147-163); allí menciona su origen aristócrata, sus simpatías hacia el fascismo italiano y anota que *desde 1897, esto es, antes de la aparición de los estudios econométricos en el estricto sentido del término, [...] Vilfredo Pareto (1848-1923) [...] encontró cierta regularidad en la distribución de los ingresos* de países tanto capitalistas como otros. Resume el método empleado así<sup>3</sup>:

Con base en estadísticas de varios países, Pareto elaboró distribuciones acumulativas de frecuencias que indicaban que muchas personas tenían *ingresos no menores* de ciertas sumas indicadas en la serie. Luego trazó los diagramas de esas distribuciones de frecuencias, marcando sobre el eje de las abscisas los ingresos  $x$  y sobre el eje de las ordenadas el número de personas con ingresos *de  $x$  o mayores que  $x$* .

NOTAS:

1) Cursivas de autor de artículo para resaltar el ordenamiento.

2) Lange cita: <sup>26</sup>V. Pareto: *Cours d'economie politique*, Lausanne, 1897, vol. 2, parte I, cap. 1.

En el libro mencionado de Lange no aparecen referencias a la Curva de Lorenz. Es muy posible que no la conociera, pues parece que Lorenz se limitó a sugerirla en su artículo, pero no se mencionan textos suyos donde elaborara la idea posteriormente. También parece que el estadístico italiano Conrado Gini, quien estudió el tema y propuso un índice para medir la desigualdad desde la teoría de las estadísticas discretas, hizo su trabajo sin mencionar a Lorenz. Diversas referencias concuerdan en que solo hasta después de la segunda guerra mundial la CL se convirtió en centro de atención e investigación.

Paul A. Samuelson, galardonado con el premio Nobel hace algunos años, publica en 1948 su famoso libro de introducción a la economía, traducido a varios idiomas, reeditado varias veces, y usado por décadas como texto en varias universidades privadas del mundo. En la 16ava versión española de 1968, menciona tanto a Pareto como a Max Otto Lorenz y su CL, presenta unas Tablas de Datos de Estados Unidos-1962, de distribución de ingresos acumulados ordenados en sentido PaR. Luego presenta los correspondientes gráficos, y a continuación se pregunta: “¿Qué porcentaje de la totalidad de la renta va a parar al diez por ciento más bajo de la población? .. ¿Y al 95 por ciento más bajo? Y así sucesivamente. Es fácil deducir tales datos de las columnas (4) y (5) de la Tabla 2 .. [..]. En la llamada Curva de Lorenz marcamos estos datos, con el porcentaje de la población en el eje horizontal, **empezando por los más pobres**, y el porcentaje de renta que reciben en el eje vertical.” (Samuelson, P.; 1968; 128-134). El pasaje deja implícito el ordenamiento PaR, ascendente, como premisa metodológica, sin informar que hay otra forma de ordenamiento posible. Esa

premisa única predominó más tarde desde los años 1970s, en numerosos estudios académicos que se publicaron al aumentar el interés mundial sobre las *Curvas de Lorenz*.

En ningún momento menciona Samuelson la posibilidad de emplear otro ordenamiento diferente, ni a los colegas que usaron esa otra opción en las seis décadas que precedieron sus palabras.

Y en una nota corta de pie de página comenta el trabajo de Pareto en 1897, medio siglo después, así (Samuelson, P.; 1968; 132-133):

*Hay todavía otros medios de medir el grado de desigualdad de la renta. Mencionaremos aquí, aunque no entremos a examinarlo en detalle, uno de los más interesantes. El italiano Vilfredo Pareto .. [...] ... sirviéndose de un gráfico logarítmico, .. [...] ... descubrió que el extremo superior de las cifras de distribución de la renta, en muchos y distintos países .. [...]. y épocas, formaba líneas rectas de inclinación casi idéntica, llegando a creer que se trataba de una ley fundamental de la Naturaleza. .. [...] .. Estudios más minuciosos llevados a cabo durante los últimos 50 años han proyectado dudas sobre la universalidad y sobre la inevitabilidad de la ley de Pareto.*

*NOTA: cursiva de autor para resaltar ordenamiento*

Se observa que Samuelson omite los autores de esos estudios minuciosos sobre Pareto, no informa que los datos originales estaban ordenados de rico-a-pobre –lo que sí hace Lange- y emplea el eufemismo de *extremo superior de las cifras de distribución de la renta* para aludir al sector de super-ricos que estudió Pareto, dentro del orden obvio de RaP. Por su parte, el socialista Lange critica el apoyo de Pareto al fascismo – ya derrotado- pero hace un elogio elegante de Pareto y del efecto innovador de su propuesta, dentro de su contexto y su época. Luego acompaña esas palabras con unos análisis matemáticos profundos y bien sustentados sobre las muy especiales características de la distribución implícita usada por Pareto y sobre las fallas de la llamada *ley de Pareto*.

La valoración que hace Paul Samuelson sobre Pareto, Lorenz y la distribución de ingresos deja muchos interrogantes. Ante un debate que no profundiza, propone un nuevo escenario: el de las CL, un tema nuevo para la época, aún sin desarrollar, en el cual dicta las premisas fundantes y pasa luego a contrastar la desigualdad entre Estados Unidos e Inglaterra, entre los países subdesarrollados y avanzados, o el ingreso medio entre hombres y mujeres, burócratas soviéticos y trabajadores soviéticos, entre las curvas de ingreso y las de capital. Por su parte, Oskar Lange, maneja la parte matemática con desenvoltura, respeta las premisas de



Pareto y desde ellas lo evalúa y corrige, pero no menciona la existencia de las CL. Puede decirse que Lange analiza mejor a Pareto pero desconoce a Lorenz, mientras Samuelson divulga la agenda futura en la investigación de la desigualdad, y ayuda a imponer el sentido de las premisas analíticas para el manejo futuro de las curvas de Lorenz.

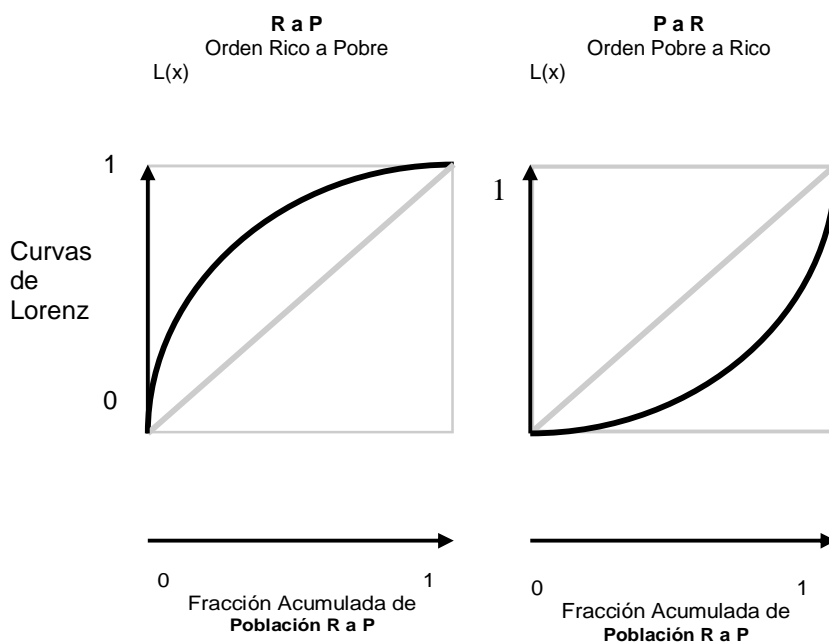
La consecuencia para las generaciones posteriores fue importante en cuanto estos eventos se juntaron para imponer una perspectiva casi única que ha regido en la investigación y la interpretación (epistemología) del tema de la distribución de ingresos desde entonces. Si bien aquí no podemos profundizar en el origen histórico de la CL, hay publicaciones (Xu, Xuan; 2004)<sup>4</sup> que precisan el papel de las primeras estadísticas inglesas de ingresos y de sus analistas de fines del siglo XIX en la generación de las ideas básicas<sup>5</sup> (Chiozza Money; 1905;42). También (Kleiber, C; 2005: 4) observa que en el artículo original de Lorenz la propuesta presenta la fracción de ingreso acumulada en el eje horizontal del gráfico con la fracción de población acumulada en el eje vertical, y muestra que las gráficas modernas de Lorenz son muy distintas ya que ahora traen la fracción de población acumulada en el eje horizontal. Aunque Kleiber también trabaja con el ordenamiento PaR, esta observación es sorprendente porque quizás indica que el trabajo de Lorenz solo trascendió cuando su gráfico fue modificado<sup>6</sup>. Afortunadamente, es posible retomar el hilo perdido y encontrar de nuevo la ruta de salida del laberinto, en parte gracias a que la valiosa CL fue impulsada después de los años 1960, por los mismos analistas que impusieron la perspectiva Pobre-a-Rico como estándar en la copiosa literatura axiomática que ha dominado el tema desde la post-guerra.

Si pasamos del gráfico de Pareto (1897) (ver Gráfico I) al de las Curva de Lorenz modernas (Ver Gráfico II), el sentido del ordenamiento es clave por una razón central: cuando los datos se ordenan PaR, la CL une los puntos [0;0] y [1;1] viajando por debajo de la diagonal del cuadro. Dado que esta gráfica es la tradicional en los estudios de distribución y en la literatura axiomática elaborada desde 1960, se entiende que los nuevos análisis se distanciaron del manejo original de Pareto en sus gráficas, que eran al inverso: ordenadas RaP, tal como se aprecia en la Tabla I, y en el Gráfico I.

Esta decisión ha dificultado el trabajo con la CL para las investigaciones sobre desigualdad de ingresos: ya fue mencionado que no es un ejercicio fácil encontrar una ecuación matemática para una curva de Lorenz que pase debajo de la diagonal del cuadro típica del ordenamiento PaR y cumpla a la vez con las condiciones lógicas de pasar por los vértices [0;0] y [1;1] sin salirse del cuadro. Al olvidar la otra opción, se renunció al camino más sencillo, el del ordenamiento RaP; la opción más compleja y difícil resultó consagrada por la comunidad de especialistas para estudiar el tema durante casi un siglo, se acumuló un voluminoso cuerpo

teórico de tipo axiomático sobre esa premisa, y se acostumbró a las academias y estudiantes a mirar el tema solo en un sentido y a hacer las gráficas con esa

**Gráfico II. Formas de las Curvas de Lorenz**

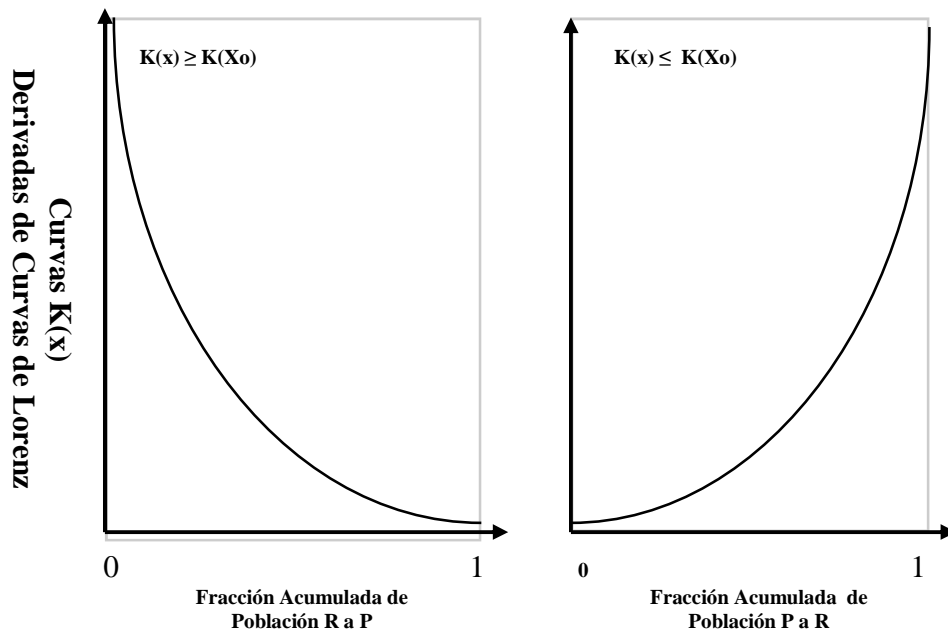


perspectiva. Este artículo aspira a rescatar las matemáticas básicas ignoradas del enfoque Rico a Pobre (RaP), e invita al lector investigador a experimentar este método mediante algunos ejemplos. Para quienes prefieren la óptica ortodoxa PaR, se respeta su elección como válida y se deja abierta una opción compatible con su preferencia: basta una simple transformación del resultado RaP y se obtiene la función de ajuste complementaria, porque las dos miradas poseen una simetría particular que las liga entre sí y permite traducir los resultados de un enfoque al otro de manera expedita aplicando las fórmulas correspondientes.

Para no usar demasiadas palabras en la explicación, los Gráficos II, III y la Tabla II resumen las analogías y diferencias existentes entre las dos perspectivas, así como las que se dan entre las curvas claves y otros análisis derivados. Al reflexionar sobre las gráficas se hace posible entender que son dos maneras distintas, complementarias y simétricas de describir el mismo fenómeno. A su vez, los comentarios finales de la Tabla II resumen la visión estadística que aportaría el cambio de perspectiva.

El trabajo logarítmico de Pareto sobre los datos es importante porque ayuda a separar el factor de escala (el ingreso medio en unidades reales), del factor estructural de la distribución, tal como lo planteó Pareto con agudeza. Por otra parte, nos legó la idea central de la famosa distribución de Pareto, muy empleada en ciertos campos. En lo que no acertó fue en su idea de que todas las distribuciones nacionales de ingresos exhibían estructuras logarítmicas de rectas paralelas como las que creyó encontrar para el dúo Inglaterra-Irlanda en los datos y gráficos aquí presentados y en otras épocas y culturas. Justamente, la investigación posterior a Pareto se encargó de demostrar que lo raro es encontrar dos estructuras distributivas de ingresos con curvas logarítmicas paralelas.

**Gráfico III Curvas K (FDA Funciones de Distribución Acumulativa)  
Según ordenamientos RaP y PaR**



**TABLA II. RESUMEN DEL FUNDAMENTO TEORICO**

**Transformaciones para obtener curvas de Lorenz  $L(z)_{PaR}$  a partir de curvas  $L(x)_{RaP}$  :**

Los ordenamientos RaP y PaR se complementan con una simetría especial ..

1)  $L(z)_{PaR} = 1 - L(1-z)_{RaP}$

2)  $L'(z)_{PaR} = -L'(1-z)_{RaP}$

3) Derivadas de CL:

$L'(z)_{PaR} = \text{Función Cumulativa de Probabilidad } K(z) \leq K_0$

$L'(x)_{RaP} = \text{Función Cumulativa de Probabilidad } K(z) \geq K_0$

4) La segunda derivada de las 2 Curvas de Lorenz son de signo diferente:

$L''(x)_{RaP} \leq 0 \dots \text{Para } (0 \leq x \leq 1)$

$L''(z)_{PaR} \geq 0 \dots \text{Para } (0 \leq z \leq 1)$

5) En las CL (Curvas de Lorenz) modernas el eje vertical se usa siempre para la variable distribuída, y el eje horizontal para la fracción de población acumulada, sea RaP o PaR. En estadística matemática convencional el eje horizontal se usa para la variable distribuída.

6) La metodología alterna requiere: 1) Datos ordenados RaP. 2) Inferir las formas funcionales de ajuste a las Curvas de Lorenz RaP. 3) Derivar la curva  $K \geq$ . 4) Si hace falta, obtener las curvas  $CL_{-PaR}$  y  $K_{-PaR} \leq K_0$  mediante las transformadas correspondientes.

7) En las CL no se necesita pensar en términos de Funciones de Densidad de Probabilidad, *fdp*. De hecho, las *fdp* vienen de dividir las frecuencias de ocurrencia por el valor medio de la variable en unidades particulares, lo cual las hace poco útiles para el análisis abstracto y adimensional.

8) Esta propuesta aún requiere ser formalizada en lenguaje matemático riguroso. Tal vez sea susceptible de presentarse en forma axiomática.

## 1.2 PREMISAS MATEMÁTICAS Y MÉTODO DESARROLLADO

### 1.2.1 Premisa Inicial: mirar el tema desde la perspectiva Pobre a Rico (PaR)

Para entender la relación entre el ordenamiento RaP y su complemento el ordenamiento tradicional, PaR, es preciso recordar que la probabilidad de pertenecer al x% más rico de una población es igual a:

$$P_{RaP} = 1 - P_{PaR} \quad [\text{uno menos la probabilidad de pertenecer al sector más pobre}]$$

Lo cual equivale también a decir que si conocemos un ordenamiento, podemos inferir fácilmente el otro, por su relación de complementariedad y simetría. Dicha relación incide también en la forma de las Curvas de Lorenz, así como en sus derivadas. Todo esto conduce a dos proposiciones importantes que expresamos sin prueba:

1) La derivada de la CL-RaP respecto a la población acumulada  $X_{RaP}$  produce la Función  $K \geq K_o$  de ingreso mínimo vs.  $X_{RaP}$ , también llamada Función Acumulativa de Distribución Mayor-o-Igual, donde  $X_{RaP}$  es la probabilidad de recibir un ingreso superior o igual a cierto valor predefinido  $K_o$ .

2) La derivada de CL-PaR respecto a la población acumulada  $X_{PaR}$  produce la Función  $K$  de ingreso máximo vs.  $X_{PaR}$ , también llamada Función Acumulativa de Distribución Menor-o-Igual. En este caso  $X_{PaR}$  es la probabilidad de recibir un ingreso inferior o igual a cierto valor predefinido  $K_o$ . Este es el enfoque tradicional.

Lo anterior nos permite afirmar que cuando se trata de datos ordenados según su tamaño, si hablamos de ingreso mínimo –o de  $K \geq K_o$ -, necesariamente nos referimos a un sistema ordenado RaP. A su vez, cuando hablamos de ingreso máximo –o de  $K \leq K_o$  -, necesariamente estamos inmersos en un sistema ordenado PaR. Dentro de este contexto también es correcto decir que:

$$P(K \geq K_o) = 1 - P(K \leq K_o)$$

*La probabilidad de que una variable sea mayor o igual a  $K_o$  es igual a uno menos la probabilidad de que la variable sea menor a  $K_o$ .*

Como ya comentamos antes, cuando Vilfredo Pareto elaboró sus famosas gráficas del logaritmo decimal de la población ( $\log_{10} N$ ) versus el logaritmo decimal del ingreso mínimo ( $\log_{10} X$ ), según lo expuesto, se refería a datos ordenados de Rico a Pobre, RaP, y datos de la relación  $K \geq K_o$  –o función de probabilidad acumulativa para  $K(x) \geq K_o$ . Ver Cuadro I, fotocopia de la Tabla de Datos elaborada por el mismo Pareto.

No hay razón alguna para preferir una perspectiva sobre la otra. Si cierto problema puede ser investigado desde dos o más perspectivas, sería valioso hacerlo desde todas las diferentes opciones disponibles. Eso es precisamente lo que se intenta en la investigación adelantada, sin excluir el otro enfoque porque son complementarios, como se insinúa desde la teoría simple de probabilidades.

Por lo tanto, nuestra primera premisa de cambio metodológico es: *1) Investigar la CL desde el enfoque PaR.*

### **1.2.2. Premisa de Función General de Lorenz**

Es el momento de recordar que en las Curvas de Lorenz modernas la variable horizontal  $X$ , *no* representa el ingreso –como es común en Pareto, o en la estadística tradicional, o en las fdp (funciones de densidad de probabilidad) –. En adelante en el eje horizontal  $X$  va la *fracción acumulada de población bien sea ordenada RaP o PaR*. A su vez, en el eje vertical  $Y$  va la *fracción acumulada de la variable repartida*, tal como el ingreso, o el capital, o la tierra en los casos socio-económicos.

Además, proponemos el empleo de la familia general de funciones de Lorenz de tipo potencial-exponencial:

$$L_{\text{RaP}}(x) = X^{F(x)} \quad [2.1]$$

.... Donde  $(0 \leq X \leq 1)$ ;  $(0 \leq L_{\text{RaP}}(x) \leq 1)$ ;  $(0 \leq F(x) \leq 1)$

Esta propuesta es coherente con la observación ya mencionada sobre la mayor facilidad de encontrar una expresión matemática para el ordenamiento RaP dentro del cuadrado de una unidad de lado que contiene las CL. Además, puede verificarse fácilmente que la función cumple los valores esperados en los dos vértices del cuadro.<sup>7</sup>

La idea es entonces buscar una curva específica de la forma [2.1] que se ajuste a la serie de datos reales y ordenados RaP, los cuales pueden venir en forma de puntos datuales de la CL, o en forma de histograma que presenta las ocurrencias versus diversos rangos de valores –o ingresos- dentro de los cuales ocurre y toma su valor el fenómeno real. Por simplicidad trabajamos con el caso más sencillo en el cual el investigador recibe una serie de cuantiles de la CL, los cuales deben ser ordenados en el sentido RaP antes de llevar a cabo el próximo análisis.

### 1.2.3 Análisis datual para generar vector estructural $F(x)$

El siguiente punto es conocer la forma de la función continua  $F(x)$  que da la forma particular a la CL a partir de los datos de cuantiles. Si sacamos los logaritmos de la expresión [2.1], llegamos a:

$$\ln(L) = \ln(x) \cdot F(x) \rightarrow F(x) = \ln(L) / \ln(x) \quad [2.2]$$

Donde  $(0 < X < 1)$  y  $(0 \leq F(x) \leq 1)$

En la práctica, si tenemos  $n$  parejas de datos empíricos de la  $CL_{RaP}$  de tipo  $(x_i, L_i)$  basta calcular para cada pareja esta relación logarítmica y obtenemos directamente  $n-1$  parejas internas de puntos  $(x_i, F_i)$  debido a que en los extremos calcular [2.2] se torna indefinido. Así las cosas, el foco del problema se transfiere al análisis de los puntos internos que definen el exponente  $F(x)$  de la expresión [2.1] y basta obtener una función continuas de ajuste de  $F(x)$  para conocer el exponente (o vector funcional-estructural) que define una de las muchas funciones posibles de ajuste de la CL. Dicha función particular  $F(x)$  [2.2] es ensamblada luego en la familia general de CL [2.1] para originar la CL particular que corresponde a la serie de datos empíricos del caso. Esta relación logarítmica directa y plena de sentido entre los datos ordenados RaP, el vector funcional estructural  $F(x)$  particular y la estructura matemática de la familia de curvas CL-RaP propuesta, permite destacar la coherencia teórica y empírica del método aquí planteado.

#### **1.2.4 Curva continua de ajuste a puntos datuales de la función F exponencial**

El siguiente paso es conseguir una curva suave y continua que se ajuste con la precisión estipulada según criterios del investigador a los puntos hallados en el anterior numeral. Para este procedimiento se sugiere hacerlo por etapas con ayuda de un programa convencional de Hoja Electrónica en un computador personal. En nuestro caso se empleó el programa Excel de Microsoft <sup>TM</sup>. El proceso intermedio sería así:

1) Graficar los ene puntos  $(x_i, F_i)$

2) Usar la función Tendencia <sup>TM</sup> -o su equivalente en otros programas- hasta hallar una función de ajuste a los puntos  $(x_i, F_i)$  con una correlación  $R^2$  (O Chi cuadrada,  $\chi^2$ , superior a 0.99 en lo posible). El programa devuelve la fórmula de la regresión -o tendencia- y el valor de la correlación  $\chi^2$ . En lo posible, conviene operar las fórmulas y sus parámetros con al menos 6 cifras decimales debido a que la CL es muy sensible al valor del exponente.

3) Elaborar una tabla que calcule los valores de  $F(x)$  para los ene puntos  $x_i$  de los datos empíricos de la función F. Calcular el promedio de los errores absolutos o el error cuadrático estándar, o cualquier índice deseado de bondad de ajuste según los criterios del analista. Si la aproximación no es satisfactoria repetir el proceso del punto 2 anterior, hasta lograr otra curva -quizás más compleja- de correlación que cumpla de mejor manera los criterios predeterminados por parte del investigador. Es importante aclarar que para un conjunto de datos puede haber infinidad de curvas matemáticas de ajuste  $F(x)$ , y en algunos casos el investigador puede acudir a la opción de emplear dos o más funciones segmentadas en lugar de una única función para todo el rango  $[0; 1]$

de población acumulada  $X$ . En este sentido, el investigador es quien decide la clase de función  $F(x)$  que prefiere, así como los criterios sobre los cuales toma esas decisiones que lo lleven a escoger la función  $F(x)$  finalmente optada. Se trata de entender y asumir el papel de la subjetividad autocontrolada del investigador en el proceso para lograr un nivel *adecuado* de objetividad.

4) Debido a que la función exponencial tiene la misma base  $X$ , basta con lograr una buena función  $F(x)$  que ajuste a los datos, e inmediatamente el ajuste de la CL queda asegurado, de modo que se puede ensamblar la función real. Basta poner  $F(x)$  como exponente de la base  $X$ , en la expresión  $L(x) = x^{F(x)}$  y tenemos asegurada la curva funcional de Lorenz para los datos del caso, dentro de las premisas del ordenamiento Rico-a-Pobre de la variable  $X$ .

### 1.2.5 Condiciones lógicas y premisas complementarias

La función  $F(x)$  debe cumplir una serie de criterios matemáticos para garantizar que la CL resultante sea apropiada a las condiciones lógicas que debe cumplir, y que la curva  $K \geq K_0$  derivada de la anterior resulte compatible con datos ordenados de Rico-a-Pobre.

En la práctica, quienes trabajan el tema encuentran situaciones inesperadas cuando extrapolan funciones principalmente en la cola baja de la distribución. Por ejemplo, la CL toma un valor mayor a uno, o en el caso de las curvas  $K \geq K_0$ , que deben ser estrictamente decrecientes dentro del contexto RaP, de pronto inician un camino ascendente inesperado. En términos muy generales esto se debe a que no cumplen ciertas reglas lógicas que debe seguir el vector-función-estructural  $F(x)$  durante su rango  $[0 \leq x \leq 1]$ . Esto puede ocurrir cuando la función  $F(x)$  de ajuste usa una curva polinómica de grado muy alto, y  $F(x)$  puede contaminar con su forma sinuosa ciertos intervalos de la distribución, haciendo inviable la propuesta aunque haya excelente precisión en el ajuste. A manera de proposición se recomienda seguir las siguientes reglas muy generales, cuya comprobación matemática no incluimos, y queda para cuando todo el conjunto de la propuesta sea formalizado de manera apropiada.

1. Para que la  $CL_{\text{RaP}}$  esté inscrita en el triángulo superior del cuadro es preciso que

$$0 \leq F(x) < 1$$

2. Para que sea positivo el ingreso,  $K \geq 0$ , es preciso que  $F'(x) / F(x) \leq [-1/(x \cdot \ln x)]$ .



Esto significa que si el vector  $F(x)$  es estrictamente decreciente cumple la condición. Si el vector  $F(x)$  tiene zonas crecientes debe verificarse su cumplimiento.

3. Para que la función  $K(x) \geq K_0$  sea decreciente como corresponde en el contexto RaP, y no tenga ningún sector donde crezca al aumentar  $x$ , lo que violaría la premisa central del ordenamiento RaP, es necesario verificar en todo el rango de  $X$  que se cumpla la relación:

$$-\frac{\left[ \frac{2 \cdot \dot{F}}{x} - \frac{F}{x^2} + \ln x \cdot \ddot{F} \right]}{\left[ \frac{F}{x} + \ln x \cdot \dot{F} \right]^2} \geq 1$$

(Nota: un punto sobre la función  $F$  significa primera derivada respecto a  $x$ , dos puntos significa segunda derivada respecto a  $x$ ).

En el momento en que esta relación pase de ser mayor que uno, a ser menor que uno, empieza a crecer el valor del ingreso  $K(x) \geq K_0$ . Esta es una señal de alarma que exige revisar la función  $F(x)$ , sus derivadas y los cálculos para detectar la causa del problema y corregirla. A veces la solución consiste en usar una función de ajuste  $F(x)$  diferente –por lo general más sencilla, como suele ocurrir cuando  $F(x)$  es una función polinómica de grado muy alto, caso en el cual basta muchas veces con buscar otra  $F(x)$  polinómica de grado menor-.

Sea cual sea su elección, al tomar esa decisión, el investigador está asumiendo de manera implícita que entre los puntos internos se interpola la función y de igual manera, en las zonas de las colas en los extremos siempre hay una extrapolación implícita portadora de riesgos lógicos no siempre conocidos ni prevenidos. Dentro de este marco, cabe la hipótesis de que la argumentación lógica que sustenta el programa POVCAL puede ser defectuosa, incompleta o simplemente errónea, de acuerdo a las críticas sobre su comportamiento, y podrían deberse a causas parecidas a las mencionadas, pero vistas desde el ordenamiento Pobre-a-Rico en el cual está concebido el programa empleado por el Banco Mundial.

### 1.2.6 Cálculo de Area de CL, Índice de Gini y Derivación de la FDA, $K \geq K_0$

Obtenida la expresión matemática de ajuste de la CL para los datos de cada caso, conviene elaborar una tabla con incrementos pequeños que contenga entre 100 y 150 puntos en el intervalo  $0 \leq X \leq 1$ . Esto hace

posible calcular varias columnas con variables claves como son (F, F', F'', CL,  $K \geq K_0$ ), verificar que se cumplen las condiciones lógicas del caso en todo el rango de las funciones, y estimar el área A bajo la CL mediante la sumatoria mecánica de elementos trapezoidales. Obtenida el área bajo la CL se obtiene el Índice de Gini con la expresión válida para el ordenamiento RaP:

$$G = 2 \cdot A - 1 \quad [2.3]$$

Resumiendo: Conocida F(x), en seguida se calcula  $L(x) = x^{F(x)}$ , y se grafica la curva de Lorenz obtenida. También conviene calcular las derivadas F'(x) y F''(x), para verificar que se cumplan las condiciones explicadas en el anterior numeral 1.2.5.

El análisis de distribución requiere normalmente conocer la curva K derivada de la correspondiente CL, en este caso obtendríamos una función de distribución acumulativa  $K \geq K_0$ , debido a que usamos la opción RaP. Para la derivación, ayuda tener en cuenta que la función general de la CL también puede expresarse como una expresión exponencial de la forma:

$$L(x_{RaP}) = e^{\ln(x) \cdot F(x)} \quad [2.4]$$

La derivada de esta función existe siempre y cuando la función F(x) sea derivable y cumpla ciertas condiciones lógicas propias de su ordenamiento RaP. La expresión de la función derivada es:

$$\partial L(x_{RaP}) / \partial x_{RaP} = K(x_{RaP} \mid K \geq K_0) \text{ o Función de Distribución Acumulativa }_{RaP} (FDA_{RaP})$$

Desarrollando la expresión se llega a

$$K(x_{RaP} \mid K \geq K_0) = [ F(x) / x + \ln(x) \cdot F'(x) ] \cdot e^{\ln(x) \cdot F(x)} \quad [2.5]$$

Para el cálculo en forma de tablas de Hoja Electrónica conviene usar la forma alterna:

$$K(x_{RaP} \mid K \geq K_0) = [ F(x) / x + \ln(x) \cdot F'(x) ] \cdot L(x) \quad [2.6]$$

Es muy interesante observar la presencia del vector-funcional F(x) en todas las funciones de interés. Cabe señalar que la función cumulativa  $K \geq K_0$  es muy sensible tanto al valor de F(x), como al de su derivada F'(x).

La función  $K \geq K_0$  posee además una propiedad muy importante que puede usarse para facilitar otros procesos de inferencia: si se calcula su valor para  $x=1$ , produce el ingreso mínimo de la distribución, y puede verificarse que:

$$K_{\geq}(x=1) = F(x=1) \quad [2.7]$$

La expresión [2.7] nos dice que el ingreso mínimo de toda la distribución está implícito en el extremo derecho del vector-funcional  $F(x)$ . Por eso, cuando adoptamos mecánicamente la premisa de extrapolar la función  $F(x)$  en el extremo bajo de la distribución, permitimos que la función predefina el valor del ingreso mínimo, así como la forma de la distribución en este sector de bajos ingresos; esto es importante porque es allí donde se calculan los índices de pobreza absoluta y pobreza extrema. Conviene por lo tanto que las agencias oficiales de encuestas y estadísticas informen al público el valor del ingreso mínimo real, al igual que el valor del ingreso medio real, en el momento de darle los cuantiles a los analistas. Otra opción consiste en sugerir que entreguen siempre al menos dos puntos datales que disten menos de un 2% de las colas alta y baja de la distribución, o sea para  $X=0.02$  y  $x=0.98$ , de modo que se reduzca la posibilidad de asumir premisas muy arbitrarias y erróneas en estos dos sectores claves de la distribución.

Una vez calculadas las tablas correspondientes, es fácil graficarlas y analizarlas. Hay puntos de interés general que pueden determinarse rápidamente, tales como: 1) el valor de la mediana, 2) la probabilidad de ganar más de un ingreso medio, la cual ocurre en el punto de la CL con pendiente igual a uno, 3) El ingreso mínimo de la muestra, el cual ocurre cuando  $x=1$ .

Tal como se explicó en el numeral anterior, el investigador debe tener presente que tanto la función  $F(x)$  como la función  $K \geq K_0$  vs  $X_{RaP}$ , deben cumplir ciertas normas lógicas en todo su rango, cuyo cumplimiento es de exclusiva responsabilidad del analista que propone las curvas funcionales. Se recomienda vigilar que la curva  $K \geq K_0$  sea:

- a) Positiva porque no se aceptan ingresos negativos, o valores de la variable aleatoria menores de cero.
- b) Estrictamente decreciente, sin pendientes positivas, porque de lo contrario implicaría que la premisa fundante sobre el ordenamiento RaP no se cumple.

**NOTAS IMPORTANTES:**

1) Es destacable que el poder de síntesis de la Curva de Lorenz se debe a que sólo emplea fracciones acumuladas, lo cual nos permite dejar de lado el factor de escala de los datos reales, como cuando trabajamos números fraccionarios abstractos sin pensar en la dimensión real de la unidad, ni en las unidades empleadas para medirla. Una vez hallada la distribución acumulativa  $K$  (Ver [2.5] y [2.6]), el valor de las unidades arrojadas para la función  $K \geq K_0$  viene dado en *unidades medias* de la muestra, o sea en unidades adimensionales. Esto significa que si se trata de una distribución de ingresos, con ingreso medio per cápita de 500 dólares mensuales, cuando  $K \geq 2$ , por ejemplo, nos referimos a un ingreso mensual per cápita igual o superior a  $2 \cdot 500 = 1000$  dólares. O sea,  $K$  viene en unidades adimensionales abstractas, lo cual facilita enormemente el análisis y la <sup>8</sup> comparación de distribuciones diversas que pueden venir en distintas unidades, con distintos valores reales de las *unidades medias*.

Fundamentar la parte teórica del método exige necesariamente el uso de un lenguaje matemático como el usado en esta sección. Sin embargo, el método aún requiere ser vertido en un lenguaje formalizado más riguroso. En el siguiente aparte abandonamos el análisis matemático e ilustramos con ejemplos concretos el procedimiento.

2) Dentro del contexto RaP, derivar la llamada Ley de Van der Wijk es relativamente fácil. Esta “ley” es explicada por (Cowell, Frank. A.; 2000: 81-83) desde la perspectiva PaR, y en esencia se refiere a una particularidad de la distribución de Pareto referida a la relación entre un ingreso  $K_0$  y el ingreso promedio de quienes reciben una suma *mayor o igual* a tal  $K_0$ . (Se observa que esta ley emplea términos del contexto RaP).<sup>9</sup>

## **2. EJEMPLO CON DATOS DE CUANTILES DE CURVA DE LORENZ**

Para ilustrar el método se usan dos series de cuantiles tomados de la base de datos de distribución de ingresos de la UNU-WIID2C 2009 sobre distribución de ingresos de China-2002 y Estados Unidos-2000. Este es el ejemplo más sencillo de manejar.

La Tabla III muestra los datos empleados. Estados Unidos entrega una serie de 10 deciles ordenados PaR, mientras China hace igual pero agrega los cuantiles del 5% más pobre y del 5% más rico, para un total de una serie de 12 cuantiles.

| <b>Tabla III. Series de Cuantiles China-2002 y EE.UU.-2000 Porcentajes PaR</b> |     |     |     |     |     |      |      |      |      |      |     |      |
|--|-----|-----|-----|-----|-----|------|------|------|------|------|-----|------|
|  | 1   | 2   | 3   | 4   | 5   | 6    | 7    | 8    | 9    | 10   | 15  | 95   |
| <b>China</b>   | .82 | .38 | .71 | .96 | .69 | 0.07 | 3.34 | 7.95 | 2.10 | 2.00 | .68 | 9.91 |
| <b>E.EUU</b>   | .81 | .53 | .79 | .04 | .32 | .67  | 0.30 | 2.45 | 6.07 | 9.03 |     |      |

FUENTES:  
 China: Li Shi, 2005. (Sample Survey by the Economics Institute of the Chinese Academy of Social Sciences)  
 USA: Luxembourg Income Study.  
 Base de datos UNU-WIID2C 2009 (Series 956 China, 5000 USA)  
 Cifras redondeadas en la Tabla. Los cálculos emplean 6 cifras significativas disponibles.

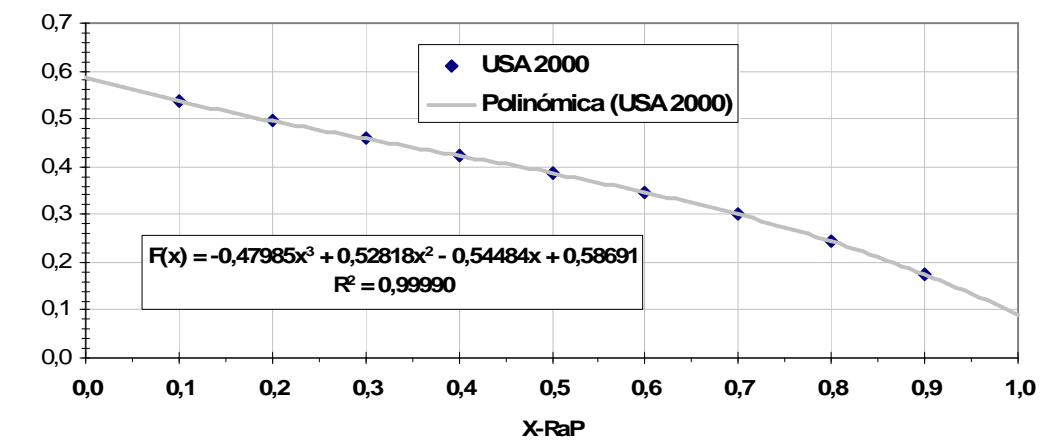
## 2.1 ORDENAMIENTO RAP Y PROCESO DATUAL PRELIMINAR

El siguiente paso reordena los datos de la Tabla III, los presenta en forma de cuantiles acumulados RaP, y efectúa los cálculos logarítmicos necesarios para graficar los puntos de la función  $F(x)$ , la cual debe insertarse en el exponente de la Función General de CL. La Tabla IV muestra los resultados de este procesamiento intermedio para los datos de Estados Unidos, año 2000.

| <b>Tabla IV. Ejemplo de Proceso Intermedio de Datos.</b> |                             |                                    |                          |
|--|-----------------------------|------------------------------------|--------------------------|
| <b>Cuantiles de EEUU-2000</b>                            |                             |                                    |                          |
| Fracción $X_i$<br>Acumulada. RaP                         | Fracción Y<br>Simple<br>RaP | Fracción $Y_i$<br>Acumulada<br>RaP | $\ln Y / \ln X$<br>$F_i$ |
| 1  | .2903                       | .2903                              | .5372                    |

|  |       |       |       |
|--|-------|-------|-------|
| 2  | ,1607 | ,4510 | ,4948 |
| 3  | ,1245 | ,5755 | ,4589 |
| 4  | ,1030 | ,6785 | ,4233 |
| 5  | ,0867 | ,7652 | ,3862 |
| 6  | ,0732 | ,8384 | ,3451 |
| 7  | ,0604 | ,8988 | ,2993 |
| 8  | ,0479 | ,9466 | ,2459 |
| 9  | ,0353 | ,9819 | ,1735 |
|  | ,0181 |       |       |
| Elaboración propia con datos de EEUU de la Tabla 2 |       |       |       |
| Cifras redondeadas                                 |       |       |       |

**Gráfico IV, Ln(L)/Ln(X) vs X Acumulada y Línea de Tendencia**



## 2.2 INFERENCIA DE CURVA ESTRUCTURAL F(X)

La Tabla IV contiene diez puntos de la Curva de Lorenz, RaP, y 9 puntos internos de la función  $F(x)$ . En este momento nuestro interés se enfoca en el Gráfico IV, el cual muestra los 9 puntos datuales de  $F(x)$ , los que

están unidos por una línea de tendencia cuya fórmula matemática es entregada por el computador, así como su correlación R2.

El Gráfico IV es portador de una valiosa información que merece varios comentarios detallados:

1) Los 9 puntos  $(x_i; L_i)$  sugieren una curva descendente. Al usar la función Línea de Tendencia del programa Excel de Microsoft ©, luego de varios ensayos desde la línea recta, la cual da un ajuste relativamente bueno, hasta otras opciones más complejas, se escogió el polinomio de grado tres que muestra el Gráfico 5. Puede observarse su alta correlación R2 (Chi-cuadrado).

2) El polinomio hallado tiene la expresión:

$$F(x) = -0,47985x^3 + 0,52818x^2 - 0,54484x + 0,58691 \text{ [2.7.a]}$$

$$R2 = 0,99990$$

Dicho polinomio arrojó un Error Absoluto Promedio igual a 0.3 % para los 9 puntos datuales disponibles de F(x), considerado satisfactorio para los criterios de este ejemplo.

3) El gráfico muestra la curva extrapolada en las dos colas, es decir para valores de  $X_{RaP}$  entre 0 y 0.1, a la izquierda, y para valores entre 0.9 y 1.0 a la derecha. El investigador no posee datos para justificar esta decisión, y se acepta que fue tomada de manera unilateral, por la simplicidad que supone trabajar con una sola función de buen ajuste intermedio. Por otra parte, se acepta de manera implícita la forma del polinomio entre los puntos intermedios como una premisa *razonable* de trabajo, sin pensar en la infinidad de expresiones matemáticas posibles para el mismo propósito particular de obtener una curva generadora F(x) satisfactoria para los datos del ejemplo.

### 2.3 ENSAMBLE DE F(X) PARTICULAR EN CL-GENERAL PRODUCE CL-RAP PARTICULAR

Así las cosas, basta reemplazar el polinomio F(x) en el exponente de la CL teórica general  $L(x)_{RaP} = x^{F(x)}$  y queda ensamblada la CL funcional para el caso particular de los EEUU-2000, inferida desde los datos del ejemplo:

$$L(x)_{RaP} = x^{(-0,47985x^3 + 0,52818x^2 - 0,54484x + 0,58691)} \dots\dots[2.7]$$

Al comparar los valores muestrales de los 9 puntos disponibles de la CL con los valores estimados por el modelo a partir de la expresión [2.7], encontramos que el Error Absoluto Medio (EAM) es de tan solo 0.09 %, considerado muy bueno para estos casos, y muy útil para mostrar que basta una precisión promedio del 0.3% de la curva F(X), para lograr una exactitud aún mayor en la CL.

## 2.4 DESARROLLOS COMPLEMENTARIOS: CL-POBRE A RICO, CURVA $K \geq K_0$ , INDICE DE GINI

Tanto la función F(x) (ver [2.7]) como su derivada son necesarias para calcular la función de distribución acumulativa  $K(x_{RaP} \mid K \geq K_0)$  (Ver [2.6]).

$$F'(x) = -0,47985 \cdot 3 \cdot x^2 + 0,52818 \cdot 2 \cdot x - 0,54484 \dots \quad [2.8]$$

Recordando que la función [2.6] tiene la forma:

$$K(x_{RaP} \mid K \geq K_0) = [ F(x) / x + \ln(x) \cdot F'(x) ] \cdot L(x) \quad [2.6]$$

Basta entonces reemplazar para cada x el valor L(x) de la ecuación [2.7], el valor de F(x) tomado de la expresión [2.7.a], y el valor de F'(x) de la expresión [2.8], en la expresión anterior [2.6] y obtenemos la curva K, o función de distribución acumulativa.

Para resolver este cálculo particular de la econometría, es suficiente usar una hoja de cálculo de un ordenador personal, disponer de datos de cuantiles, dominar el procedimiento, y ser consciente de los criterios de precisión satisfactoria así como de las premisas admisibles en los extremos desconocidos de la distribución.

Normalmente se calculan en su orden las funciones F, F', F'', la CL, su área, el índice de Gini, la Curva  $K \geq K_0$ , índices de error, y luego se grafican las curvas esenciales y se realizan otros estimativos propios del campo de la distribución estadística y de los estudios de desigualdad y pobreza. La verificación de las condiciones lógicas debe aplicarse cuando se detectan anomalías en las curvas.



Para calcular el área como paso previo a la estimación del índice de Gini, lo normal es que no se dispone de una integral precisa, por lo cual se emplea el método numérico mecánico. Esto se obtiene con facilidad en una hoja de cálculo moderna, haciendo la sumatoria manual de unos 100-120 segmentos trapezoidales sobre el rango de la abscisa de cero a uno. En nuestro caso empleamos 20 incrementos de 0.001 entre  $x=0$  y  $x=0.02$  y 98 incrementos de 0.01 para el resto de la variable  $X$  entre 0.02 y 1.00. Esto se hace porque la zona inicial influye bastante en el área, ya que allí se concentran los valores altos (los super-millonarios de la distribución). Puede verse que conviene acordar unas normas básicas para reglamentar este procedimiento, de modo que se facilite comparar diversos estudios análogos.

Con los datos del ejemplo, podemos también construir las curvas complementarias para quienes prefieren trabajar con el ordenamiento ortodoxo PaR. La CL tendría la forma ya explicada cuando se trató la simetría complementaria que liga los ordenamientos PaR y RaP. Se requieren dos expresiones para las transformadas: una para la frecuencia acumulada de población, y otra para la CL:

$$z = 1 - x \quad [2.9]$$

$$L(z)_{PaR} = 1 - (1 - z)^{F(1-z)} \quad [2.10]$$

En nuestro ejemplo particular, la función  $F$  particular tomaría la forma

$$F(1-z) = -0,47985(1-z)^3 + 0,52818(1-z)^2 - 0,54484(1-z) + 0,58691 \quad [2.11]$$

Y la CL-PaR sería así:

$$L(z)_{PaR} = 1 - (1 - z)^{[-0,47985(1-z)^3 + 0,52818(1-z)^2 - 0,54484(1-z) + 0,58691]} \quad [2.12]$$

Obsérvese, que al igual que en numerosas propuestas teóricas de CL convencionales, la expresión  $(1-z)$  está repetida numerosas veces. Este es un síntoma de que las curvas  $CL_{PaR}$  son el resultado de aplicar una transformada del tipo  $z = 1 - x$  (para valores de  $x$  entre cero y uno) provenientes del ordenamiento RaP. (Pascual, M.; 2004: Pág. 101)

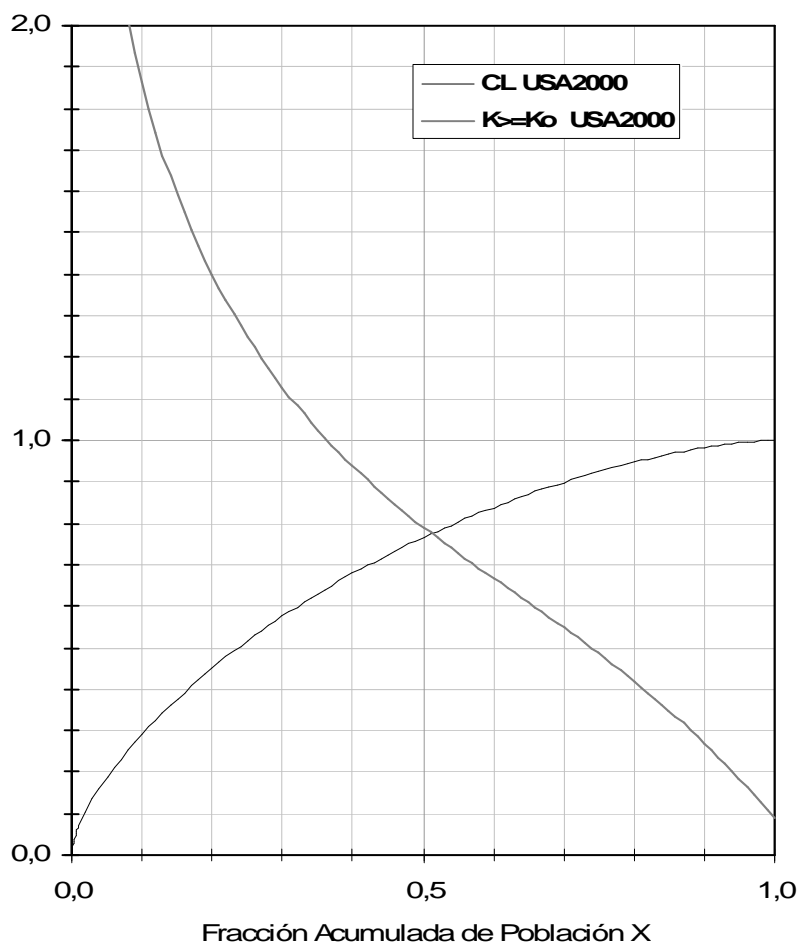
## 2.5 RESULTADOS DE EJEMPLO EEUU-2000 Y CHINA-2002

Los datos del ejemplo fueron tomados de la Base de Datos WIID de la Universidad de las Naciones Unidas (UNU-WIID2C) los cuales proporcionan no sólo los cuantiles oficiales de encuestas nacionales de ingresos de numerosos años y fuentes, sino otros elementos claves que sirven como referentes para determinar la precisión de los estimativos, tales como el Ingreso Promedio, la Mediana, las Unidades de Ingreso, el Índice de Gini, y otros relativos a los datos, sus fuentes, sus unidades reales y su calificación en cuanto a calidad y nivel de confiabilidad. El Gráfico V. entrega las dos curvas principales que nos interesan desde la perspectiva RaP: 1) La Curva de Lorenz, y 2) La función de distribución acumulativa  $K(x) \geq K_0$ . Esta última viene en unidades medias, lo que significa que los valores del ingreso mínimo están divididos por el valor promedio esperado de la distribución. Conviene aclarar que en el Gráfico se presentan simultáneamente la Curva de Lorenz RaP y la curva  $K(x) \geq K_0$  para valores bajos, menores de dos ingresos medios. Esto se hace por simplicidad, y porque en el caso de los ingresos, esa es la zona en la cual se dan las mediciones de porcentajes de pobreza y exclusión. El Gráfico VI muestra que el ingreso mínimo es aproximadamente una décima parte del ingreso medio. A su vez, indica que aproximadamente el 37% de la población ganaba más del ingreso promedio, y el resto, un 63% de la población, no alcanza el ingreso promedio en el año 2000.

La Tabla V. presenta un resumen comparativo entre los datos oficiales y los obtenidos a través del método explicado. Predecir el valor de la mediana es particularmente difícil debido a que implica un ajuste muy bueno a los datos de la CL, y es muy sensible a los cambios en la derivada de la CL. Predecir el índice de Gini es mucho más sencillo ya que éste es un índice relativamente tosco, y no aporta mayor información sobre la calidad de un modelo de ajuste, en especial cuando no se exigen precisiones superiores al 0.1 %. Sería conveniente diseñar un conjunto de criterios y pruebas muy exigentes para comparar los resultados de las diversas curvas propuestas para el ajuste, de modo que puedan clasificarse según la capacidad de predicción de la curva K, en lugar de la CL, o del índice de Gini.

**Gráfico V. Curva Funcional de Lorenz y  
Curva  $K \geq K_0$   
EEUU-2000 Ordenamiento Rico a Pobre**

1)CL 2..



**Tabla V. Resultados versus datos oficiales China-2002 y EEUU-2000**

|   | Índice de Gini |        |              | Mediana en Unidades Medias |                      |             |        |              |
|---|----------------|--------|--------------|----------------------------|----------------------|-------------|--------|--------------|
|   | Oficial        | Modelo | Error Abs. % | Ingreso Medio Real Oficial | Mediana Real-Oficial | Oficial / A | Modelo | Error Abs. % |
| China 2002                                  | 45356          | 45359  | 008%         | 912<br>Yuan/Cápita         | 986<br>Yuan/Cápita   | 6742        | 6649   | 94%          |
| EEU 2000                                    | 39420          | 39456  | 09%          | 9567.74<br>Dólares/cápita  | 5567.0<br>Dól./Cáp   | 7955        | 7911   | 56%          |
| Elaboración propia sobre datos de Tabla III |                |        |              |                            |                      |             |        |              |

### 3. APLICACIONES EN EL ANÁLISIS ESTRUCTURAL Y TEORÍA DE LAS DISTRIBUCIONES

Un aspecto muy prometedor desde el punto de vista teórico de la estadística matemática es el estudio de la estructura de las distribuciones continuas de variables aleatorias positivas a través de la función-vector-estructural  $F(x)_{RaP}$ . Basta identificar la ecuación de ajuste a  $F(x)$  inferida desde datos empíricos reales, o postulada desde datos matemáticos teóricos, y se logra una aproximación valiosa de las CL, de la Curva  $K \geq$  que permite análisis matemáticos y gráficos de alta calidad relativa. En este punto, la lógica inferencial da paso a los análisis deductivos y al lenguaje axiomático, tarea que aún está por cumplirse en el enfoque propuesto. Por ahora, es suficiente con esbozar las generalidades del tema.

En el enfoque PaR es frecuente leer estudios dedicados a comparar CL que presentan pequeñas diferencias entre sí. Los estudios de “Dominancia” por ejemplo, trabajan gráficos de CL que no se entrecruzan, y en base a ellos determinan que una distribución es más equitativa que la otra en todo el espectro de  $X$ . Sin embargo, no hay análisis frecuentes para interpretar los casos en los cuales las CL se entrecruzan una o más veces. Aquí

ofrecemos una opción de este tipo de análisis basado en los vectores estructurales  $F(x)$ , cuyas gráficas son más nítidas y directas.

El Gráfico VI proporciona ayuda para explicar el tema. Este es un gráfico que contiene dos distribuciones de elaboración artificial, diseñadas a partir de dos vectores  $F(x)$  diferentes. El primer vector  $F_1(x)=0.5$  es una línea horizontal de valor constante, típica de la distribución simple de Pareto, dentro de nuestro contexto RaP. El segundo vector es una línea recta descendente,  $F_2(x) = 0.6 - 0.2714 \cdot x$  que une los puntos  $[0;0.6]$  y  $[1; 0.3286]$  dentro del cuadro de las CL. Los dos vectores lineales se cortan en el gráfico; basta resolver la ecuación:

$$F_1(x)=F_2(x) \text{ y se obtiene el punto de corte: } \dots x = 0.36846$$

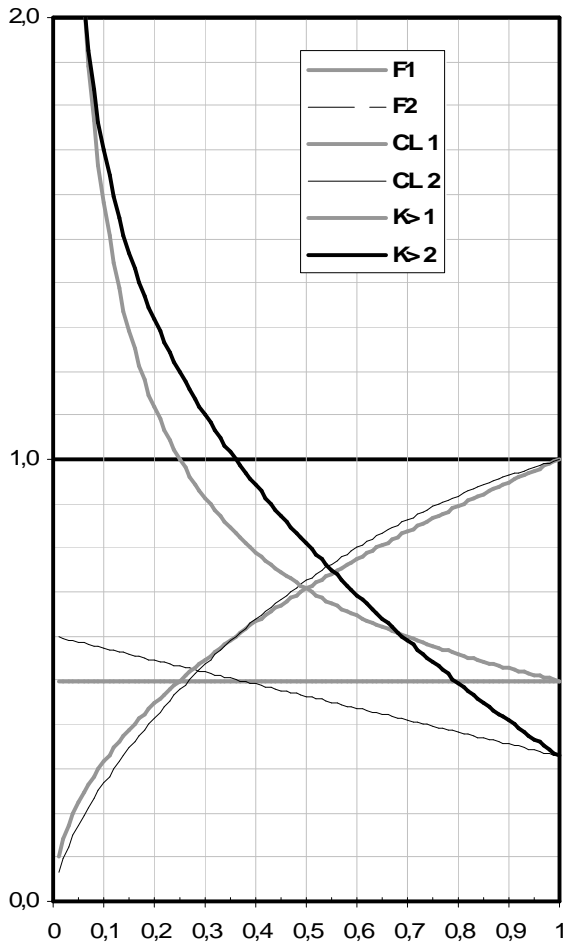
Si se observa el Gráfico VI, las Curvas de Lorenz también se cruzan encima del corte antes descrito, aunque es menos clara la imagen debido a que las dos CL son bastante parecidas. En general, dentro de nuestro contexto RaP, entre más alta la posición del vector  $F(x)$  dentro del cuadro de la CL, más equitativa la distribución. En el límite, cuando el vector se convierte en una línea horizontal  $F(x)=I$ , la distribución se hace absolutamente equitativa y la CL toma la forma  $y=x$ , con Índice de Gini igual a cero. A su vez, entre más baja sea la posición del vector  $F(x)$ , más inequitativa será la distribución y más alto su índice de Gini. En el límite, si  $F(0+)=0$ , ocurre la inequidad perfecta, un solo poblador es dueño del total, los demás son excluidos del reparto, y el Gini toma un valor de uno. Por otra parte, cuando el vector es una línea inclinada, puede afirmarse que entre más alto su valor en los extremos izquierdo y derecho, más equitativa es la distribución, menor el Gini, y más alto el ingreso mínimo.

Cuando se trata de dos líneas que se entrecruzan una o más veces es difícil predecir cuál posee un índice de Gini más alto. Sin embargo, las funciones  $F(x)$  poseen una propiedad muy especial: el valor de la función en el extremo derecho, cuando  $x=1$ , coincide con el ingreso mínimo de toda la distribución. Esta es una propiedad interesantísima que puede probarse fácilmente y es de gran ayuda en los procesos de inferencia. En el caso de nuestro Gráfico 7, puede decirse, sin trazar las curvas CL y  $K_{\geq}$ , que los más pobres reciben 0.3286 ingresos medios en el caso  $F_2$ , mientras que en el caso  $F_1$  reciben la mitad (0.5) del ingreso promedio. Este tipo de inferencias simplifican el trabajo analítico y hacen más sencilla y atractiva la enseñanza de la estadística aplicada. Sin embargo, encierran una advertencia: cuando se trabaja con deciles, con frecuencia es necesario asumir la forma de las dos colas en zonas desconocidas, y en el caso del decil más pobre, extrapolar o asumir a priori la forma de la distribución en ese sector más pobre, equivale a predefinir sin suficiente conocimiento el

valor del ingreso mínimo. Pese a todo esto, puede considerarse un privilegio tener que enfrentar este tipo de dilemas a la hora de inferir curvas distributivas a partir de datos normalmente escasos cerca a los extremos.

Conviene agregar que los dos vectores estructurales  $F(x)$  fueron prediseñados de manera que tuvieran casi el mismo índice de Gini igual a 0.3333 (en realidad una pequeña diferencia aparece en la quinta cifra decimal). Desde hace mucho tiempo es conocido el hecho de que dos distribuciones con formas bastante diferentes pueden tener un índice de Gini virtualmente igual. Si se observan las Curvas de Lorenz de los dos casos analizados, una domina en el sector de ingresos altos, la otra en el sector de ingresos bajos. Si se observan las 2 curvas  $K \geq$  las diferencias aumentan, estas se entrecruzan dos veces: una en el sector de altos ingresos alrededor de  $x=0.07$ , y otra en el sector de bajos ingresos entre  $x=0.68$  y  $x=0.69$ . Llama la atención que entre esos dos puntos de la zona media, la clase media de  $F_2$  posee mejor ingreso que la clase media de  $F_1$ , sin embargo en las dos colas, tanto el ingreso de los más ricos, como el de los más pobres es superior para el vector funcional  $F_1$ . Surge una pregunta que tiene implicaciones éticas: ¿Cuál distribución es más justa en este caso... la que permite un mayor ingreso mínimo al 30% más pobre al precio de darle más ingreso al 7% más rico, o la que garantiza mayor ingreso relativo a las clases medias que son en este caso el 63%? Desde luego, la respuesta depende de que el valor real del ingreso medio de la distribución  $F_2$  pueda garantizar un nivel de vida digno a los más pobres de su curva, los cuales sólo ganan el 32.86% del ingreso medio, mientras que en la distribución  $F_1$  los más pobres están relativamente mejor porque ganan el 50% del ingreso medio. Mientras el ingreso medio de  $F_2$  no sea mayor en un 52% al de  $F_1$ , siempre será preferible la distribución  $F_1$ , a pesar de que ambas poseen el mismo índice de Gini, si se acepta como prioridad ética la capacidad de consumo de los más pobres. Sin embargo, es posible que las clases medias de la opción  $F_2$  consideren que ellos merecen vivir mejor porque son el 63%, pueden hacer valer su mayoría electoralmente en una contienda democrática, y creen que los más pobres deben consumir menos por no tener virtudes similares a las del sector medio. Aunque es sólo un ejemplo hipotético, no es muy diferente al debate que puede darse en diversas sociedades modernas.

Gráfico VI. Curvas Iso-Ginis  
 $F1=0,5$   $F2=(0,6 - 0,2714X)$   $Gini=0,3333$



El ejemplo anterior, antes que resolver la pregunta sobre ética y economía política, nos debe llevar a concluir que el índice de Gini es un índice muy precario para entender la realidad de la estructura distributiva en profundidad y detalle.

De los ejemplos y desarrollos explicados hay las siguientes proposiciones relacionadas con el tema:

1. Es posible diseñar y construir encuestas y estadísticas artificiales para el consumo público y hacerlas pasar como realidades *científicas y técnicamente elaboradas*. La actitud ética basada en la práctica de la verdad, la buena voluntad y la transparencia informativa no tiene sustitutos que puedan autojustificarse detrás

de calificativos como *científico, técnico, matemático, moderno, positivo, estadístico, objetivo, cuantitativo, computarizado, actualizado según los últimos estándares, certificado por tal entidad* y otros de corte similar.

2. Aparte de la influencia de la calidad de los datos en la inferencia estadística, esta es una disciplina que se basa tanto en técnicas y matemáticas especializadas acumuladas culturalmente, como en premisas asumidas en base a criterios no siempre bien aclarados, ni informados, ni suficientemente regulados por la comunidad de expertos en el tema. La intuición apoyada en esos conocimientos acumulados culturalmente puede tener un rol parcial importante para acertar en las premisas que deben adoptarse sobre zonas desconocidas de la distribución, sin las cuales sería imposible completar la tarea de proponer una mirada global a cada fenómeno distributivo en condiciones de información limitada de los datos. Dado que los datos siempre serán parciales y no suelen dar información correcta sobre la forma de las colas, no hay sino dos opciones: 1) exigir la entrega de datos de calidad cercanos a las colas que disten uno o máximo dos por ciento de los dos extremos posibles de cero y uno, aparte de los datos intermedios normales; o 2) adoptar premisas sobre las colas, justificarlas y aceptar su componente subjetivo –tal como ocurre con cualquier curva funcional propuesta para ajustar las CL- y presentar el resultado como el producto de un diseño responsable que aspira a presentar una imagen integral del fenómeno estudiado dentro del marco limitado de la información datual disponible.

3. De toda distribución continua de una sola variable aleatoria no negativa, cuyos datos sean de origen empírico o teórico, y cumpla con las normas lógicas y matemáticas mínimas que la hacen posible dentro del ordenamiento Rico a Pobre, *RaP*, puede inferirse al menos un vector-funcional  $F(x)$  del cual se pueden derivar versiones funcionales de Curva de Lorenz y Curva de Distribución Acumulativa  $K_{\geq}$ , esenciales para otros análisis y estimativos derivados. Este enunciado quizás puede servir como base para fundamentar la renovación teórica y la enseñanza de este sector de la estadística matemática, de fundamental importancia en la investigación académica, científica, social y aplicada en general.

4. El índice de Gini es un indicativo de baja sensibilidad y poca capacidad para describir y comparar fenómenos distributivos complejos como la estructura de las desigualdades. Por lo pronto se sugiere buscar otro índice que podría basarse en el vector-funcional-estructural  $F(x)$  por su facilidad de inferencia a partir de los datos, por su claridad gráfica y por sus ventajas interpretativas en estudios de la dominancia.

5. Aunque las funciones de densidad de probabilidad (f.d.p) han desempeñado un papel importante en la evolución de la estadística matemática, no son entidades importantes dentro del enfoque *RaP* de este estudio, salvo como fuentes de datos para estudiar a fondo las funciones de distribución acumulativa (f.d.a.) que



acompañan las tablas de distribuciones famosas en los textos (Gauss-Normal, Lognormal, Weibull, Pareto, etc.). Por una parte, su estructura puede ser estudiada con el método propuesto; por otra parte, las *f.d.p.* no son adimensionales y pueden ser sustituidas con histogramas obtenidos de particiones muy finas de la curva  $K_{\geq}$ , con forma de datos agrupados muy detallados. Este tema necesita más investigación y reflexión en el futuro.

#### **4. INFERENCIA A PARTIR DE DATOS AGRUPADOS CON INGRESOS MEDIOS TOTALES Y GRUPALES DESCONOCIDOS**

La inferencia obtenida en el ejemplo anterior del numeral 2. Puede considerarse como un ejercicio trivial cuando se compara con el nivel de dificultad del ejemplo que explicamos a continuación de manera muy general. Debido a que el proceso es largo y requiere usar ordenador personal, lo indicado es dejar la explicación práctica detallada para otro artículo, o para un aula de clase, con bastante tiempo disponible. Es frecuente que los datos obtenidos vengan en forma de histograma con dos columnas datuales: la primera columna presenta los límites superior e inferior de ingreso, y la segunda columna presenta el número de casos registrados en la muestra (o de frecuencias simples). La Tabla VI resume los datos oficiales de Ingresos por Hogar de Canadá, año 2006.

En la Tabla VI se observa que los únicos datos de la variable repartida son los del ingreso menor y del ingreso mayor que sirven de límite a cada grupo. No conocemos ningún valor medio, sea el de la distribución total, sea el de cualquiera de los grupos. En la literatura estadística consultada en Internet sobre este tipo de inferencias aplicadas al ordenamiento convencional PaR, son pocos los investigadores del tema que consideran posible resolver este tipo de inferencia con datos escasos (Owang, T; 2003; 422). Sin embargo, dentro del método descrito RaP, es posible y relativamente sencillo obtener muy buenas aproximaciones al problema. La dificultad principal consiste en que el proceso requiere asumir simultáneamente varias premisas para llenar los principales vacíos existentes, concretamente en: 1) El ingreso medio total; 2) Una distribución teórica inicial que tenga máximo dos parámetros, 3) Un criterio de medición del error de ajuste.

Para resolver estos tres puntos el procedimiento seguido fue: 1) Arrancar de un valor de Can. Dólares \$ 80000, límite inferior del grupo 6. 2) Usar un vector función lineal del tipo

$F(x)=2-A-Bx$ , con valores iniciales  $A=1.5$   $B=0.3$ , lo cual significa que el ingreso mínimo asumido es  $K_{min}=0.2$  unidades medias. 3) Al dividir cada ingreso mínimo grupal por el valor medio asumido (\$80 000), encontramos un referente para medir el error respecto a los valores teóricos calculados para cada valor datual

de ingreso mínimo grupal (en unidades promedio). El analista puede escoger otro conjunto de valores para los dos primeros puntos, si lo cree conveniente. Como hay 18 puntos datuales internos, calculamos para cada uno el Error Absoluto puntual, y calculamos el Error Absoluto Estándar (*SAE* en inglés), o promedio de los anteriores. Esta es la función de control usada para minimizar el error a medida que se cambian los valores del ingreso medio, y de los parámetros A,B. Todo el proceso se adelantó con un programa de hoja de cómputo electrónico en el ordenador personal. Para una persona con experiencia y destreza en el manejo del tema, el proceso desde que se copian los datos iniciales, se reordenan, se plantean las simulaciones lineales, y se pasa a la etapa de la función  $F(x)$  no lineal, hasta llegar a un punto aproximado y aceptable, incluidos los cálculos y gráficos finales, puede tomar unas cuatro horas para un conjunto de datos como los del ejemplo de histograma de la Tabla 5.

**Tabla VI. Inferencia desde Datos Agrupados**

Histograma de Ingresos por Hogar Canada 2006

Reordenado Rico a Pobre (RaP)

| Grupos de Ingreso |                 |                 | Número    |
|-------------------|-----------------|-----------------|-----------|
| Grupo.<br>número  | Desde<br>Can.\$ | Hasta<br>Can.\$ | Hogares   |
|                   | 250.000         | y más           | 209.710   |
|                   | 200.000         | 250.000         | 158.440   |
|                   | 150.000         | 200.000         | 456.690   |
|                   | 100.000         | 150.000         | 1.386.150 |
|                   | 90.000          | 100.000         | 473.690   |
|                   | 80.000          | 90.000          | 544.990   |
|                   | 75.000          | 80.000          | 297.050   |
|                   | 70.000          | 75.000          | 311.480   |
|                   | 60.000          | 70.000          | 657.350   |
|                   | 50.000          | 60.000          | 683.570   |
|                   | 45.000          | 50.000          | 345.160   |
|                   | 40.000          | 45.000          | 344.930   |
|                   | 35.000          | 40.000          | 341.920   |
|                   | 30.000          | 35.000          | 341.740   |
|                   | 25.000          | 30.000          | 334.250   |

|       |        |        |           |
|-------|--------|--------|-----------|
| 5     | 20.000 | 25.000 | 238.220   |
| 7     | 15.000 | 20.000 | 166.310   |
| 8     | 10.000 | 15.000 | 120.250   |
| 9     | 0      | 10.000 | 217.430   |
| SUMA= |        |        | 7.629.330 |

Statistics Canadá, CANSIM, Table

FUENTE: 111-0012

<http://cansim2.statcan.ca/>

Después de varias iteraciones y pruebas de ensayo y error, se llegó en este caso a un error SAE del orden del 8% para la función  $F(x)$  de tipo lineal, con parámetros A,B. En ese punto, utilizando las propiedades del modelo en ese punto del proceso, se calcularon diversas proporciones del ingreso para cada grupo, las cuales se usaron luego para recalculer tanto el ingreso medio, como los ingresos acumulados de CL provisionales, más la respectiva función provisional  $F(x)$ .

Cuando se estanca el progreso con la función lineal, se puede pasar a otro tipo de función no lineal. Entonces se grafica  $F(x)$  y se hace un análisis datual sencillo que produce una función de ajuste no lineal, de tipo polinómico que puede ser de grado entre 2 y 5, de modo que se logre un coeficiente chi cuadrado ( $R^2$  o  $\chi^2$ ) superior a 0.999. Con los parámetros de la función de ajuste se recalcula, y se observa que el error SAE disminuye gradualmente de forma lenta, hasta llegar en este ejemplo a un 2% promedio para los 18 puntos internos.

En cierto punto, después de varios intentos, se observa que la convergencia se estanca: ni avanza, ni empeora. En ese momento se aceptan los valores de ingreso medio y de función  $F(x)$  paramétrica obtenidos. El resto consiste en graficar, calcular índices deseados, y presentar los resultados de la inferencia, para que sean cotejados con otros estudios y datos disponibles. Es de anotar que entre más riqueza datual haya (más datos grupales repartidos), mejor será la inferencia, y menor será el efecto de extrapolación en las dos colas.

Los resultados obtenidos se presentan de dos maneras: por una parte la Tabla VII incluye los valores principales obtenidos en el proceso de inferencia; por otra parte, el Gráfico VII resume las tres curvas fundamentales de la distribución: la curva F (o vector estructural), la curva CL (o curva de Lorenz) y la curva

$K_{\geq}$  (o curva de la Función Acumulativa de Distribución), todas elaboradas dentro del contexto RaP. A su vez, se presentan los datos de ingresos mínimos grupales resultantes de dividir sus valores reales por el ingreso promedio. Puede observarse que el ajuste es satisfactorio y tiene coherencia interna, aunque ello no significa que sea de alta precisión respecto a los datos reales, porque estas inferencias dependen de premisas y son siempre aproximaciones.

El nivel de aproximación solo puede determinarse una vez se conozcan los cuantiles exactos de la distribución. En nuestro caso, tanto el histograma como el dato de la mediana los proporciona StatCan, la agencia oficial de estadísticas canadienses. La inferencia fue hecha sin conocer los datos de ingresos medios, ni el índice de Gini, ni otros datos que suelen procesar las agencias estadísticas.

Como es bien conocido, la inferencia consiste en obtener una conclusión general a partir de datos reales particulares. El complemento de la inferencia es la deducción, la cual consiste en obtener una conclusión particular a partir de una teoría general. El Gráfico VII es resultado de deducir las fórmulas principales derivadas en secciones anteriores, usando la fórmula del vector función estructural  $F(x)$  inferido en el proceso. El proceso en su conjunto emplea tanto la lógica inferencial estadística, como la lógica deductiva de la matemática estadística, la cual es relativamente sencilla dentro del enfoque RaP.

**Tabla VII. Resultados de Inferencia con Datos Agrupados de Canadá-2006**

|                   | Ingreso Medio<br>CAN \$<br>por Hogar | Area bajo<br>CL | Indice de<br>Gini | Mediana<br>= $K(x=0,5)$ | Mediana \$<br>= Col. 2 x Col. 5 |
|-------------------|--------------------------------------|-----------------|-------------------|-------------------------|---------------------------------|
| Valores Inferidos | 86608,28                             | 0,69469         | 0,38939           | 0,81821                 | 70864,09                        |
| Valores Oficiales | n/d                                  | n/d             | n/d               | n/d                     | 70400,00                        |
| Error Absoluto    | n/d                                  | n/d             | n/d               | n/d                     | 0,66%                           |

Función F(x) Inferida

$$F(x) = -2,7922182x^5 + 5,4745177x^4 - 3,7479617x^3 + 0,9076429x^2 - 0,3836088x + 0,5680929$$

(Extrapolado para  $x \leq 0,02749$  y para  $x \geq 0,97150$ )

Ingreso Mínimo = 0,02648 Unidades Medias = \$2 292.07

EAM (Err-Abs.Medio) = 1.88% para 18 puntos

**Elaboración propia de autor sobre datos de Tabla VI**

**Datos no disponibles, n/d, requieren consultarse con Stat Can, la agencia oficial estadística de Canadá.**

Cabe señalar que en el Gráfico VII, por razones de espacio fue necesario alterar la escala vertical del gráfico, razón por la cual luce aplastado, y los cuadrados de uno por uno toman forma rectangular. En el gráfico es visible que cerca del 39% gana más de un ingreso medio, lo cual significa que el 61% gana menos del ingreso medio. A su vez, la mediana es un 81% del ingreso medio. Como parte del proceso inferencial, se asumió que el ingreso mínimo sería aproximadamente 1/40avo del ingreso promedio. Este valor, que no es preciso, resultó de la extrapolación en la cola baja de la función F(x) asumida por comodidad, para no usar más de una función F(x). De todas maneras, el último punto datual a la derecha muestra un ingreso bajo apenas superior a 1/10 que cae rápidamente, por lo cual es muy posible que el ingreso mínimo sea realmente muy bajo.

En la literatura consultada sobre inferencias en el contexto tradicional también hay un debate sobre qué tipo de premisas deben tomarse en las dos colas de la distribución y qué clase de análisis datual es el más conveniente para justificar esas premisas. A la luz de esta observación, puede entenderse mejor el valor relativo de las inferencias desarrolladas hasta ahora, pues el problema no es fácil ni para la opción PaR, ni para el caso RaP.

Es posible preguntarse ¿Porqué no intentar la inferencia a partir de datos completos, de modo que haya un referente completo para evaluar el resultado? La pregunta es válida y de hecho sí se realizaron varios cómputos satisfactorios, usando datos completos, con aproximaciones inferiores al 3% de error. El problema radica en que el hecho de disponer de cualquier dato adicional que permita conocer el ingreso promedio genera una información que puede sesgar el proceso, así se afirme que se actúa “como si no se conociera ese dato”.

Gráfico VII Curvas Deducidas de Inferencia

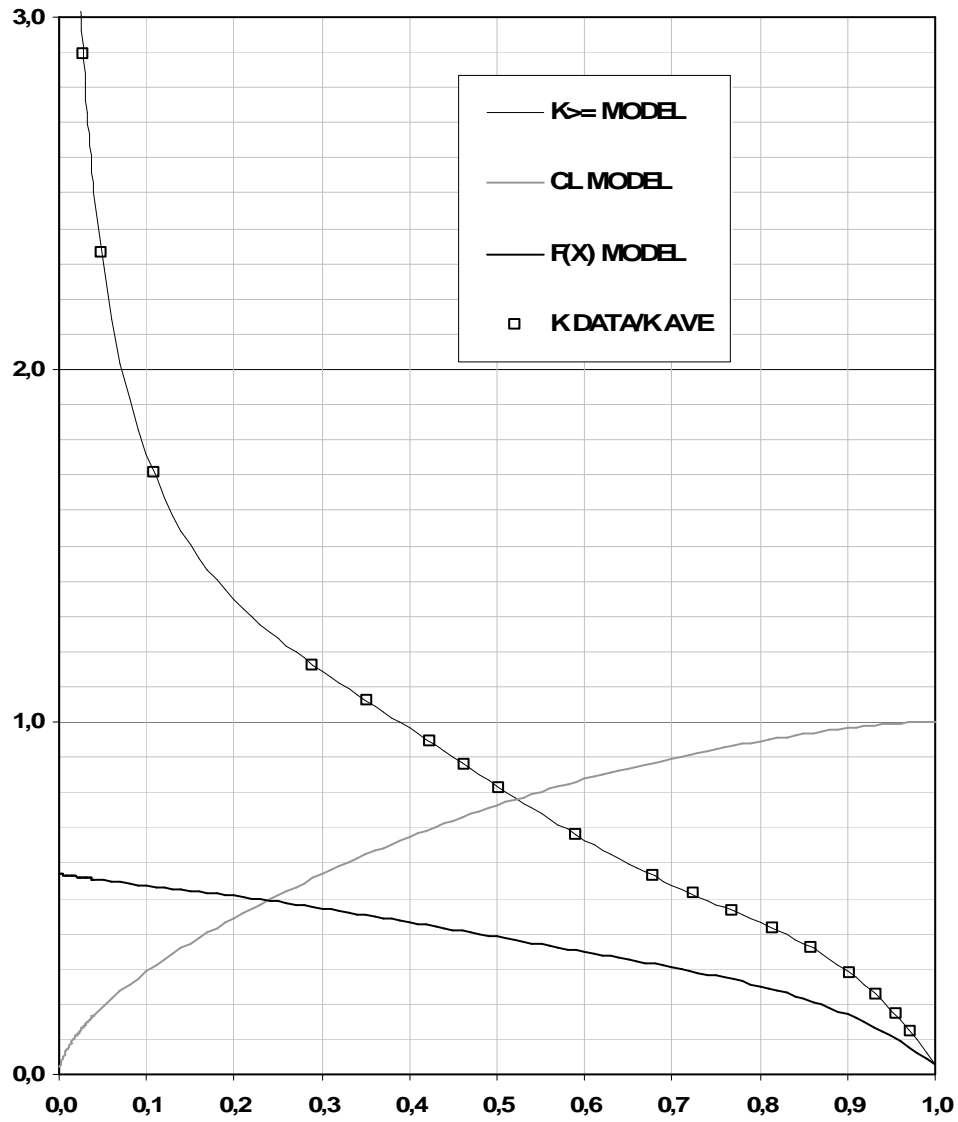
$$F(x) = -1,33444x^4 + 2,03057x^3 - 1,03641x^2 - 0,18635x + 0,57015$$

1) K\$ ave= \$86608,28 2) Mediana=0,8182 (\$70864)

3) Gini=0,3894 4) SAE=1,88%

Datos de Statistics Canada, CANSIM, table 111-0012

Median=\$70400



Por esta razón, y dado que no se explica el método en detalle, se usaron los datos del histograma de frecuencias canadienses para trabajar el tema en las condiciones más difíciles posibles, con la idea de que en el futuro podremos comparar los resultados con los datos completos, y conocer mejor los niveles de aproximación obtenidos. Ya llegará el momento para conocer los datos complementarios que permitan medir la calidad de la inferencia, así como también será necesario escribir un artículo futuro explicando paso a paso este tipo de procesos de inferencia.

Finalmente, y para someter los resultados obtenidos a una comparación aún más rigurosa, se entrega la Tabla VIII con los ventiles deducidos de dicha inferencia y otros valores importantes, de tal modo que los especialistas la comparen con los datos oficiales cuando estos sean publicados. Esos resultados ayudarán a conocer mejor las causas de las diferencias comparando los dos vectores estructurales  $F(x)$ : el obtenido aquí mediante la inferencia, y el valor real calculado a partir de la serie de cuantiles oficiales canadienses cuando sean publicados.

**Tabla VIII. Cuantiles RaP estimados de inferencia**

| <b>Cuantil</b> | <b>Y</b>      | <b>X</b>     | <b>Y</b>     | <b>F aprox</b>                    |
|----------------|---------------|--------------|--------------|-----------------------------------|
| <b>No.</b>     | <b>Grupal</b> | <b>Acum.</b> | <b>Acum.</b> | <b><math>\ln Y / \ln X</math></b> |
| <b>1</b>       | 0,19207       | 0,05         | 0,19207      | 0,5507                            |
| <b>2</b>       | 0,09928       | 0,10         | 0,29135      | 0,5356                            |
| <b>3</b>       | 0,08090       | 0,15         | 0,37225      | 0,5209                            |
| <b>4</b>       | 0,07098       | 0,20         | 0,44323      | 0,5056                            |
| <b>5</b>       | 0,06444       | 0,25         | 0,50767      | 0,4890                            |
| <b>6</b>       | 0,05947       | 0,30         | 0,56714      | 0,4711                            |
| <b>7</b>       | 0,05517       | 0,35         | 0,62231      | 0,4518                            |
| <b>8</b>       | 0,05108       | 0,40         | 0,67339      | 0,4316                            |
| <b>9</b>       | 0,04701       | 0,45         | 0,72040      | 0,4107                            |
| <b>10</b>      | 0,04294       | 0,50         | 0,76334      | 0,3896                            |
| <b>11</b>      | 0,03892       | 0,55         | 0,80226      | 0,3685                            |
| <b>12</b>      | 0,03509       | 0,60         | 0,83735      | 0,3475                            |
| <b>13</b>      | 0,03155       | 0,65         | 0,86891      | 0,3262                            |
| <b>14</b>      | 0,02837       | 0,70         | 0,89727      | 0,3039                            |
| <b>15</b>      | 0,02551       | 0,75         | 0,92279      | 0,2793                            |

|             |         |      |         |        |
|-------------|---------|------|---------|--------|
| <b>16</b>   | 0,02284 | 0,80 | 0,94563 | 0,2505 |
| <b>17</b>   | 0,02005 | 0,85 | 0,96568 | 0,2149 |
| <b>18</b>   | 0,01669 | 0,90 | 0,98237 | 0,1688 |
| <b>19</b>   | 0,01211 | 0,95 | 0,99448 | 0,1079 |
| <b>20</b>   | 0,00552 | 1,00 | 1,00000 |        |
| <b>Suma</b> | 1       |      |         |        |

1) Cálculos estimados de

**NOTAS:** autor

2) Resultados pendientes de verificación con  
datos

oficiales de Canadá, Couple Families, 2006

Statistics Canada, CANSIM, table 111-0012

## CONCLUSIONES

El ordenamiento Rico-a-pobre abre la puerta para una interpretación sencilla, coherente y directa de las distribuciones continuas de una variable no negativa. Es complementario al tradicional ordenamiento Pobre-a-rico que predomina en el manejo del tema. Dentro de este contexto los temas y ejemplos desarrollados permiten concluir los siguientes puntos generales:

1. A partir de algunos datos de cuantiles de la Curva de Lorenz es posible inferir de manera expedita vectores-funcionales-estructurales  $F(x)$  de buena precisión, los cuales son claves para el análisis detallado de distribuciones, para generar CL funcionales, para definir las condiciones lógicas que debe cumplir y para derivar otras funciones, gráficos e índices de uso frecuente, y para profundizar los estudios de dominancia.

2. A partir de histogramas de frecuencias de ingresos agrupados –sin otra información de ingresos que los valores reales de los límites de ingreso de cada grupo- es posible inferir el ingreso medio y la estructura distributiva como en el caso anterior, en procesos graduales más complejos que arrojan resultados de menor precisión.

3. Es posible concebir una teoría estadística del sector de las distribuciones continuas elaborada alrededor de las Curvas de Lorenz, en la cual las funciones particulares de densidad de probabilidad (f.d.p) serían



calculadas dividiendo la frecuencia entre el valor medio que ocurre entre los límites definidos del grupo en *ingresos-medios*.

4. Los planteamientos teórico-prácticos aquí resumidos podrían usarse para rediseñar el programa POVCAL del Banco Mundial. Las aplicaciones de este enfoque también pueden extenderse a otros campos de investigación diferentes a los de economía aplicada.

5. El poder de la Curva de Lorenz estriba en su carácter adimensional, abstracto y sintetizador de realidades complejas y variadas, que pueden ser graficadas dentro de un cuadrado de una unidad por lado –propiedad que no poseen las f.d.p.-. Dentro de tal cuadro caben también el vector-función-estructural  $F(x)$ , la CL, y la función de distribución acumulativa  $K \geq K_0$  para valores menores a una media de la distribución.

6. Aplicando las correspondientes transformaciones dentro de la simetría complementaria que las une, es posible obtener las CL para el enfoque tradicional ascendente, pobre a rico, a partir de los resultados del enfoque RaP aquí desarrollado.

7. Dado que aquí se entregan los elementos y condiciones para diseñar y construir distribuciones de todo tipo, es importante reafirmar que no hay sustitutos para una ética de la verdad, la buena voluntad y la confianza en la práctica de la estadística. Por lo tanto, es preciso pedirle a los funcionarios públicos y comunicadores que no usen a la herramienta estadística y su técnica como argumento suficiente de veracidad en sus informes y noticias.

8. El artículo es un esbozo preliminar de la teoría desarrollada. Aún requiere presentarse en un lenguaje matemático más riguroso y formalizado.

## COMENTARIOS Y NOTAS DE PIE DE PÁGINA

---

<sup>1</sup> Miembro de Grupo de Investigación de Economía Endógena – Udenar

<sup>2</sup> En este artículo empleo el término “datual” como adjetivo referido al sustantivo “dato”, ya que no contamos con una palabra para ello en español, como sí ocurre con los dúos punto-puntual, gesto-gestual, rito-ritual, etc.

<sup>3</sup> En las 17 páginas que dedica al tema, Lange reproduce una Tabla tomada del economista usamericano Harold T. Davis, *The theory of Econometrics*, Bloomington, 1947, p.29, con datos de Estados Unidos-1919 presentados en sentido RaP. También trabaja Lange otra tabla con datos ordenados RaP, para Polonia-1955 en el mismo texto. Esto confirma que el enfoque RaP era usado y mencionado entre 1897-1958. Lange, Oskar, *Introducción a la Econometría*, Fondo de Cultura Económica, México, 1964. (Primera edición en idioma polaco, 1958).

<sup>4</sup> Xuan Xu informa el papel pionero del libro de Chiozza Money (1905) y le asigna la propuesta de la curva de Lorenz. A su vez explica que el italiano Conrado Gini trabajó paralelamente su índice de desigualdad desde la estadística de datos discretos en 1921, para lo cual no le hacía falta conocer el artículo de Lorenz. Xu trabaja dentro de la tradición PaR, pero se diferencia porque

---

menciona y formula los símbolos y definiciones matemáticas para el ordenamiento RaP (pág. 6), reconociendo su existencia. Xu, Xuan. "How Has the Literature on Gini's Index Evolved in the Past 80 Years?" December 2, 2004, Department of Economics, Dalhousie University, Halifax, Nova Scotia, Canada B3H 3J5. Disponible en: <http://economics.dal.ca/RePEc/dal/wparch/howgini.pdf> (Junio 20, 2009).

<sup>5</sup> Varios estudios destacan el papel del súbdito y lord británico Leo George Chiozza Money, (1870-1944), periodista y político de origen genovés, contemporáneo de Lorenz, conocido por sus artículos de análisis estadístico en el diario londinense Daily News. En su obra más famosa, *Riches and Poverty* (1905) presenta y analiza las primeras estadísticas oficiales de distribución en la Gran Bretaña de la década de 1890 (las mismas que usó Pareto), y nos da dos cuantiles graficables como CL: *Entonces, a la conclusión de que cerca de la mitad del ingreso total de la nación es disfrutada por apenas la novena parte de su población, debemos agregar otra aún más destacada: que más de un tercio del ingreso total del Reino Unido es apenas disfrutado por menos de un treceavo de su gente* (Traducción propia). Los datos más usuales entonces eran los ingresos más altos, usados por Pareto; Chiozza-Money, además incluye a los sectores más pobres, lo que Pareto no hace. Parece que Lorenz nunca desarrolló su idea con un nivel comparable al de Chiozza-Money, ni al de Pareto. Chiozza Money, *Riches and Poverty*. 1906. Methuen & Co. 36 Essex Street W.C. London. 3a. Ed. En Internet: <http://www.archive.org/stream/richespoverty00moneiala#page/ix/mode/1up> (Junio, 2009)

<sup>6</sup> Kleibert menciona que "Según Derobert y Thieriot (2003), el término "Lorenz curve" es usado por primera vez en (King, W.; 1912:156:158), un texto de estadística para ciencias económicas y sociales". El texto es de 1911, incluye en la edición de 1920 la Figura 15 titulada "Lorenz Graph. Showing Distribution of Wealth", muy diferente a la que Kleibert atribuye a Lorenz, con la fracción acumulada de ingreso en el eje Y, y la población acumulada en el eje x, con un valor de 100% a la izquierda, que disminuye de izquierda a derecha hasta llegar a cero. El resultado es una gráfica totalmente distorsionada con una curva que desciende de uno a cero por debajo de la diagonal entre (0;1) y (1;0). Sospecho que ni Lorenz, ni King manejaban una idea clara de la curva. Curiosamente, ambos eran contemporáneos, los dos vinculados a la Universidad de Wisconsin, y en su libro King elogia al "Doctor Lorenz" por haber desarrollado la CL que ha "probado ser especialmente aplicable en el estudio de la distribución de la riqueza". King, Willford I. 1911. "The elements of Statistical Method". 1920, MackMillan company, NY. (Pág. 156-158) Kleiber, Christian. "The Lorenz curve in economics and econometrics", 2005. Institut für Wirtschafts- und Sozialstatistik, Universität Dortmund, Dortmund, Germany. Disponible en: [http://www.statistik.tu-dortmund.de/fileadmin/user\\_upload/Lehrstuehle/MSind/SFB\\_475/2005/tr30-05.pdf](http://www.statistik.tu-dortmund.de/fileadmin/user_upload/Lehrstuehle/MSind/SFB_475/2005/tr30-05.pdf)

<sup>7</sup> En 2003 propuse la curva de Lorenz  $L = x^{2-a-b \cdot x}$ , dentro del ordenamiento RaP, como idea embrionaria de la manejada en este artículo. Entonces anoté que la función que va en el exponente podía tomar otras formas no lineales según los datos de cada caso reflejados en su análisis logarítmico, lo cual obligaría a replantear el manejo matemático. Chaves, Emilio J. "Distribución y Coeficiente de Gini, Curva Paramétrica de Lorenz sugerida y Cálculos", Dic. 2003. **Tendencias**. Pág. 97-123, Vol. IV, No.2. Revista de la Facultad de Ciencias Económicas y Administrativas. Universidad de Nariño, San Juan de Pasto, Colombia. Disponible en: <http://www.udenar.edu.co/general/tendencias/contenidos/Vol4.2/emilio.pdf>

<sup>9</sup> En la literatura existe la propuesta de las Curvas Generalizadas de Lorenz y de Ingreso construidas en unidades reales de ingreso, con el fin de comparar los efectos de escala y de estructura distributiva simultáneamente. Desde la visión RaP no es una idea apropiada debido a que se sacrifica la abstracción. En otro momento se hará otro artículo para estudiar estos casos a partir de la forma de integrar regiones diferentes de un mismo país para producir la distribución del conjunto.

<sup>9</sup> La ley de Van der Wijk se prueba así en el contexto RaP. 1) El ingreso medio de los valores arriba de  $K_0$  está dado por:  $K_{ave} = L(x) / x \dots$  2) La función  $K \geq K_0$  se expresa como:  $K(x) = L(x) \cdot [F(x) / x + \ln(x) \cdot F'(x)] \dots$  3) La relación general  $K(x) / K_{ave} = [F(x) + x \cdot \ln(x) \cdot F'(x)] \dots$  4) Sólo la distribución de Pareto tiene  $F(x)$  constante, y  $F'(x) = 0$  en el rango entre cero y uno de  $x$ . 5) Solo en este caso se cumple que  $K(x) / K_{ave} = F(x) = \text{constante}$  para el rango de  $X$  entre cero y uno. QED.

Para expresar la distribución de Pareto en el contexto RaP, conviene usar  $F(x) = 2 - A$  donde  $(1 \leq A \leq 2)$ . Entre mayor sea  $A$ , más desigual la distribución, menor el valor de  $F(x)$  constante, menor el ingreso mínimo  $(2-A)$  y mayor el índice de Gini. Conocido el parámetro  $A$ , la forma de la CL viene dada por  $L(x) = x^{2-A}$ .

Si calculamos  $K(x)$  para  $X_{RaP}=1$ , puede verse que en general –dentro del análisis RaP expuesto– se llega a  $K(1) = \text{Ingreso mínimo} = F(1)$ . Esta expresión se parece a una ley general, la de Van der Wijk es una propiedad particular.

---

## REFERENCIAS BIBLIOGRAFICAS

**COWELL, Frank A.** Measuring Inequality, Mayo, 2000. 3a. Ed. Oxford University Press, U.K.. Disponible en: <http://sticerd.lse.ac.uk/research/frankweb/MeasuringInequality/index.html> (Junio 15/2009)

**CHAVES, Emilio J.** "Distribución y Coeficiente de Gini, Curva Paramétrica de Lorenz sugerida y Cálculos", Dic. 2003. Tendencias. Revista de la Facultad de Ciencias Económicas y Administrativas. Universidad de Nariño, San Juan de Pasto, Colombia. Pág. 97-123, Vol. IV, No.2. Disponible en: [www.alromer.netfirms.com/volumen42/emilio-1.pdf](http://www.alromer.netfirms.com/volumen42/emilio-1.pdf) (Primera Parte) y [www.alromer.netfirms.com/volumen42/emilio-2.pdf](http://www.alromer.netfirms.com/volumen42/emilio-2.pdf) (Segunda Parte).

**CHIOZZA MONEY, L.G.**, Riches and poverty, 3a Ed. 1906 (1a. Ed. 1905 Oct.) Methuen & Co. 36 Essex Street W.C. London. Disponible en: <http://www.archive.org/stream/richespoverty00moneiala#page/ix/mode/1up> (Mayo, 2009)

**KING, Willford I.** The elements of Statistical Method. 1920, MackMillan company, New York. (Primera edición en 1911). Disponible en: <http://www.archive.org/stream/elementsstatist03kinggoog#page/n11/mode/1up>

**KLEIBER, Christian.** "The Lorenz curve in economics and econometrics", 2005. Institut für Wirtschafts- und Sozialstatistik, Universität Dortmund, Dortmund, Germany. Disponible en: [http://www.statistik.tu-dortmund.de/fileadmin/user\\_upload/Lehrstuehle/MSind/SFB\\_475/2005/tr30-05.pdf](http://www.statistik.tu-dortmund.de/fileadmin/user_upload/Lehrstuehle/MSind/SFB_475/2005/tr30-05.pdf)

---

**LANGE, Oskar.** Introducción a la Econometría. 1978. México. Fondo de Cultura Económica, 4ª. Reimpresión, 1978. (1ª. Edición en polaco, 1958).

**LORENZ, M.O.** (1905). Methods of measuring the concentration of wealth. Quarterly Publications of the American Statistical Association, 9 (New Series, No. 70), 209–219. Copia electrónica suministrada por British Library, Referencia:

RZPEJF000001 C~C% EU DSED48 COPYRT|INTERNATIONAL LIBRARY OF CRITICAL WRITIN|GS IN ECONOMICS|2003; VOL 158; VOL 1; PP 133-143|THE ECONOMICS OF POVERTY AND INEQUALITY|`METHODS OF MEASURING THE CONCENTRATION|OF WEALTH'|LORENZ, M. O.|4542.91000 0 1368-6933

**MINOIU, Camelia; REDDY, Sanjay G.** “Estimating Poverty and Inequality from Grouped Data: How Well Do Parametric Methods Perform?”, Junio 25, 2008. (Working paper) Disponible en: [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=925969](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=925969) [Junio 10, 2009]

Nielsen, François; “Module 5 – Vilfredo Pareto”, Spring 2007, Sociological Theory, University of North Carolina, Chapel Hill. (Presentación PowerPoint) Disponible en: <http://www.unc.edu/~nielsen/soci250/m5/soci250m5.pdf> [Junio 10, 2009]

**OGWANG, Tomson.** “Bounds of the Gini Index using sparse information on mean incomes”. En Review of Income and Wealth. Serie 49, Número 3, pp. 415-423, Septiembre 2003. Disponible en: <http://www.h.scb.se/scb/Projekt/iariw/program/8Bboundssparce.pdf> (Dic. 2008)

**PASCUAL SÁEZ, Marta.** "Modelización estocástica de distribuciones de renta a partir de familias paramétricas". Tesis Doctoral, 2004. Departamento de Economía, Universidad de Cantabria. Disponible en: [http://www.ief.es/Publicaciones/Investigaciones/Inves2004\\_05.pdf](http://www.ief.es/Publicaciones/Investigaciones/Inves2004_05.pdf) (Mayo, 2009)

**SAMUELSON, Paul.** A. Curso de Economía Moderna, 1968, Aguilar, Madrid. (*Economics, an introductory Analysis*, 1948, Mc Graw Hill, Ney York – Traducción de 6ª. Edición en inglés por José Luis Sampedro-Madrid).

**UNU-WIID2C** 2009 Base de datos de Distribución de Ingresos de UNU-WIDER World Institute for Development Economics Research, United Nations University. Helsinki, Finlandia. Disponible en: [http://www.wider.unu.edu/research/Database/en\\_GB](http://www.wider.unu.edu/research/Database/en_GB) (Junio 2009)

**XU, XUAN.** “How Has the Literature on Gini’s Index Evolved in the Past 80 Years?” December 2, 2004, Department of Economics, Dalhousie University, Halifax, Nova Scotia, Canada B3H 3J5. Disponible en: <http://economics.dal.ca/RePEc/dal/wparch/howgini.pdf> (Junio 20, 2009).