

UNA METODOLOGÍA FLEXIBLE PARA LA MODELIZACIÓN DE LA DISTRIBUCIÓN DE LA RENTA.

Rafael Herrerías Pleguezuelo
Federico Palacios González
Antonio R. Ramos Rodríguez

Facultad de Ciencias Económicas y Empresariales.
Universidad de Granada

RESUMEN.

Este trabajo presenta una metodología flexible para la modelización de la distribución personal de la renta, utilizando los métodos de estimación no paramétricos para obtener un patrón con el que comparar y seleccionar el grado de una función generadora polinómica previamente estimada por el método de los momentos aprovechando que, para dicha familia, coincide con la estimación máximo-verosímil. A partir de la generadora seleccionada se reconstruye la densidad correspondiente, obteniéndose una forma funcional que modeliza la distribución de la renta.

PALABRAS CLAVE: Distribución de la renta, estimación no paramétrica, método de los momentos, función generadora.

1.- INTRODUCCIÓN.

Es indudable la importancia de la modelización de la distribución de la renta en el estudio de la desigualdad y la pobreza en nuestras sociedades, como punto de partida en el diseño y ejecución de las políticas impositivas y de redistribución de la riqueza.

Según Dagum (1981): “Históricamente, los estudios sobre la distribución del ingreso han seguido dos caminos bien definidos de investigación. El primero reconoce a David Ricardo (1817) como jefe de escuela y estudia la distribución funcional del ingreso;...El segundo camino de investigación reconoce como jefe de escuela a V. Pareto (1897) y se ocupa de la distribución personal del ingreso”.

La aportación de Pareto es fundamentalmente econométrica y constituye sin duda el embrión que dará lugar a la posterior aparición de multitud de modelos para explicar la distribución de la renta. Dagum (1981), en un intento por clasificar la infinidad de modelos aparecidos, especifica y analiza las propiedades esenciales de tres sistemas básicos que generan la casi totalidad de los modelos de distribución desarrollados hasta ese momento. A saber, el sistema de D’Addario, el sistema de Pearson y un nuevo sistema generador de modelos de distribución de la renta propuesto por el mismo autor.

Posteriormente, se han desarrollado otros modelos para el estudio de la distribución de la renta. Así, por ejemplo, Esteban (1986) propone una función gamma generalizada, McDonald (1984) presenta dos distribuciones beta generalizadas tetraparamétricas y McDonald y Xu (1992) introducen una distribución beta generalizada pentaparamétrica.

La existencia de tantos modelos sugiere que el problema de la distribución de la renta adquiere forma propia en cada momento y en cada país o región. En consecuencia, parece lógico pensar más en una metodología flexible, que en encontrar un modelo paramétrico estándar para ajustar en cada caso. Esto sugiere utilizar los métodos no paramétricos en una primera instancia y la selección del modelo que más se adecua en una segunda.

Trabajos actuales apuntan en este sentido, así, por ejemplo, Del Oro (1992), realiza una estimación no paramétrica de la densidad de la renta en España; Brachmann (1995), estima los modelos paramétricos más comunes de la distribución de la renta con datos no agrupados de la economía alemana, y los compara mediante un test de bondad de ajuste con la densidad estimada por métodos no paramétricos, concluyendo que la mayoría de los modelos comúnmente utilizados no describen con fidelidad la distribución empírica de la renta.

Los métodos no paramétricos, aunque son más flexibles para explotar la información muestral, no proporcionan una forma funcional con la que trabajar, por ello en este trabajo se pretende abordar el problema basándose en el novedoso concepto de función generadora, inspirada en la expresión del sistema de Pearson y desarrollado por Callejón (1995), donde se sientan las bases de una nueva metodología de generación de distribuciones de probabilidad que abre un amplio abanico de posibilidades en el estudio de las mismas, ya que la familia de distribuciones abarcadas es inmensa.

A grandes rasgos, se pretende, a partir de una muestra, realizar una estimación no paramétrica de la generadora, que se utilizará como patrón para comparar y seleccionar el grado de una generadora polinómica que previamente se habrá estimado por el método de los momentos aprovechando que, para esta familia, coincide con la estimación máximo-verosímil [Callejón-Santos (1994)]. Posteriormente se reconstruye la densidad correspondiente al polinomio seleccionado y se obtiene una forma funcional que modeliza la distribución a partir de una generadora tan flexible como un polinomio.

Para ilustrar esta metodología se realiza una simulación con una muestra de tamaño 200 procedente de una población Gamma (5,10), utilizando para las estimaciones 26 puntos equidistantes del rango muestral y posteriormente se aplica a datos reales correspondientes a ingresos de hogares andaluces recogidos en la Encuesta de Presupuestos Familiares elaborada por el INE para el periodo 1990-91

2.- CONCEPTO DE FUNCIÓN GENERADORA.

Es conocido Callejón (1995), que bajo unas condiciones mínimas de continuidad y existencia de la derivada de la función de densidad, la función generadora de una distribución univariante continua, para los puntos en que $f(x) > 0$, viene dada por:

$$g(x) = \frac{f'(x)}{f(x)}$$

Para las distribuciones univariantes continuas más comunes, las funciones generadoras tienen expresiones analíticas más simples (en ocasiones lineales en los parámetros) que las correspondientes densidades. Esto permite una forma más sencilla, precisa y sistemática de seleccionar el modelo más adecuado utilizando una estimación previa por métodos no paramétricos. Esta idea es la que induce a pensar que la función generadora puede aportar una nueva vía para el problema de la modelización de la distribución de la renta.

3.- ESTIMACIÓN NO PARAMÉTRICA DE LA FUNCIÓN GENERADORA.

La estimación no paramétrica de la generadora se puede obtener por cociente de las estimaciones no paramétricas de la densidad y de su derivada respectivamente. Ambas estimaciones se realizan por el método kernel con el núcleo óptimo de Epanechnikov. La estimación de la densidad viene dada por la siguiente expresión [Härdle (1991); Silverman (1986)]:

$$f^*(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - x_i}{h}\right)$$

donde:

n = tamaño muestral
 x_i = observación i -ésima
 h = parámetro suavizador o ventana
 $k(u)$ = núcleo

En este trabajo se ha optado por el núcleo de Epanechnikov que responde a la expresión:

$$k(u) = \begin{cases} \frac{3}{4}(1-u^2) & \text{si } |u| \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

La elección del parámetro ventana resulta de vital importancia, ya que controla el grado de suavizamiento de la estimación, así valores muy pequeños dan lugar a estimaciones muy rugosas y valores elevados una gráfica demasiado suavizada. Para calibrar dicho parámetro h , se utilizaron métodos de validación cruzada de la forma de máxima verosimilitud (Härdle (1991)). Por tanto, se elegirá h de forma que maximice la expresión:

$$CV_{kl}(h) = n^{-1} \sum_{i=1}^n \log[f_{h,i}^*(x_i)] = n^{-1} \sum_{i=1}^n \log \left[\sum_{j \neq i} k \left(\frac{x_i - x_j}{h} \right) \right] - \log[(n-1)h]$$

Para la realización de estos cálculos es necesario su implementación en módulos de Visual Basic, lo que conlleva un elevado tiempo de computación ya que se realizan cálculos del orden de $n \times n$ en cada uno de los valores de la horquilla propuesta para h .

Como es conocido, el estimador no paramétrico de la densidad hereda las propiedades de continuidad y diferenciabilidad del núcleo [Härdle(1991)], por lo que para obtener la estimación no paramétrica de la derivada de la misma se procedió a la derivación del núcleo y la consideración del mismo parámetro ventana h .

Con las estimaciones no paramétricas de la densidad y su derivada, se construye por cociente la estimación de la generadora:

$$g^*(x) = \frac{f'^*(x)}{f^*(x)}$$

La **Figura I** representa la estimación de $g(x)$ obtenida para la muestra simulada y la generadora de la distribución que proporcionó las observaciones.

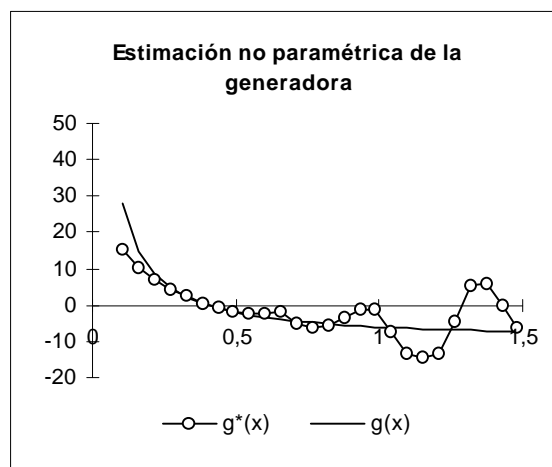


Figura I. Estimación no paramétrica de la generadora ($g^*(x)$) para la muestra simulada. Población Gamma (5,10), $n=200$, $h=0,154$.

4.- ESTIMACIÓN POR EL MÉTODO DE LOS MOMENTOS DE LA GENERADORA POLINÓMICA: EQUIVALENCIA CON EL DE MÁXIMA-VEROSIMILITUD.

El presente trabajo se centra en la familia de generadoras polinómicas, para la que se demuestra que la estimación de los parámetros por el método de los momentos coincide con el de máxima verosimilitud (Callejón-Santos (1994)). Dentro de la familia anterior, resultan especialmente adecuadas para la modelización de la distribución de la renta, las polinómicas de grado impar, que generan densidades con forma campanoide.

Para obtener la relación recurrente de momentos, podemos establecer (Johnson (1970)):

sea $g(x) = \sum_{i=0}^r a_i x^i$ la generadora polinómica propuesta. Por definición:

$$f'(x) = f(x) g(x) = \sum_{i=0}^r a_i f(x) x^i$$

si se multiplica por x^t en ambos miembros, con $t=0..r$, y se integra en el dominio de definición de la variable se obtiene:

$$x^t f'(x) = x^t f(x) g(x) = \sum_{i=0}^r a_i f(x) x^{i+t}$$

$$\int_{-\infty}^{+\infty} x^t f'(x) dx = \sum_{i=0}^r a_i \int_{-\infty}^{+\infty} x^{i+t} f(x) dx = \sum_{i=0}^r a_i m_{i+t}$$

e, integrando por partes en el primer miembro, se obtiene la relación de momentos

$$-t m_{t-1} = \sum_{i=0}^r a_i m_{i+t} \quad \text{para } t=0..r$$

Sustituyendo los momentos poblacionales por sus correspondientes muestrales, se obtiene el sistema que nos permite estimar los coeficientes del polinomio generador y que además, como se apuntó anteriormente, coinciden con las estimaciones máximo-verosímiles de los mismos, por lo que se obtiene estimadores con las propiedades deseables.

5.- SELECCIÓN DEL GRADO DEL POLINOMIO.

Siguiendo la metodología del epígrafe anterior se estiman polinomios generadores de grado 1 a 7 y se comparan con la estimación no paramétrica obtenida según se indica en el epígrafe 3 (**Figura II**), seleccionando el que minimice la diferencia cuadrático media entre ambas en los puntos de estimación (**Tabla I**); en la simulación llevada a cabo es el polinomio de grado 3 el que lo consigue.

generadora	dcm
grado 1	60,4457681
grado 3	30,9385055
grado 5	37,5149122
grado 7	59,3355869

Tabla I. Diferencia cuadrático media entre la estimación por método de los momentos y no paramétrica en los puntos de estimación para la muestra simulada.

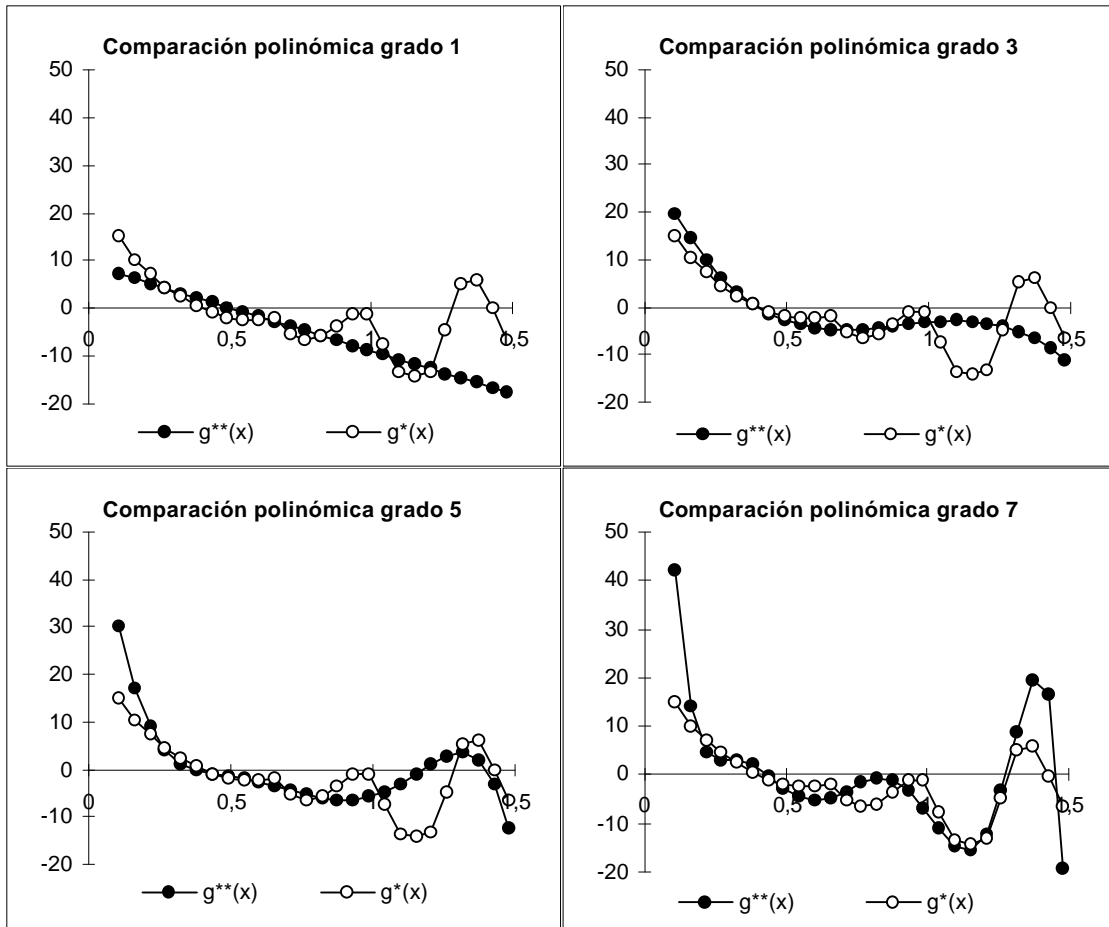


Figura II. Comparación estimaciones por el método de los momentos ($g^{**}(x)$) y mediante kernel ($g^*(x)$) de la generadora para la muestra simulada.

Una vez seleccionado el grado del polinomio generador se puede reconstruir la densidad correspondiente sin más que calcular la constante k con la condición de normalización, esto es:

$$f(x) = K e^{\int_{-\infty}^{+\infty} g(x) dx}$$

e imponiendo la condición de área se obtiene:

$$k = \left[\int_{-\infty}^{+\infty} e^{\int_{-\infty}^{+\infty} g(x) dx} dx \right]^{-1}$$

En la práctica, la constante k de normalización se obtiene por métodos numéricos mediante su implementación en un módulo de Visual Basic. La **Figura III** recoge los resultados obtenidos para la muestra simulada.

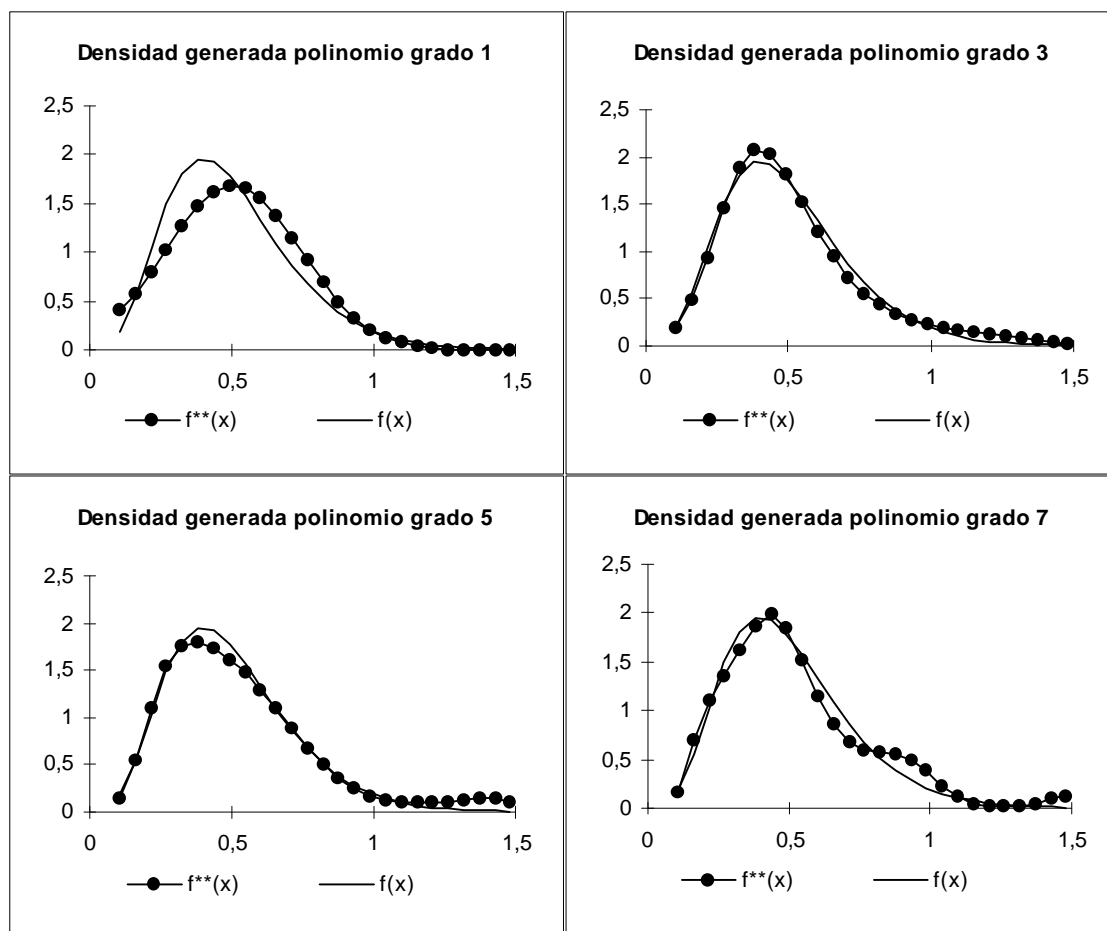


Figura III. Comparación de las densidades obtenidas por generadoras polinómicas con la correspondiente a la población que proporcionó la muestra.

6.- RESULTADOS CON DATOS REALES.

A partir de una muestra de 3600 observaciones del total de ingresos, correspondientes a los hogares andaluces de la Encuesta de Presupuestos Familiares elaborada por el INE para el periodo 1990-91, se procede a la estimación no paramétrica y por el método de los momentos de la generadora con la metodología indicada en los apartados 3 y 4.

En la **Figura IV** se comparan la estimación no paramétrica y la obtenida por el método de los momentos para generadoras polinómicas de grados 1,3,5 y 7. La **Tabla II** recoge, asimismo, la diferencia cuadrática media entre ambas curvas en los puntos de estimación. La generadora polinómica de grado 3 es la que hace mínima dicha cantidad.

generadora	dcm
grado 1	11,32515584
grado 3	9,482603495
grado 5	10,55597479
grado 7	12,65114073

Tabla II. Diferencia cuadrático media entre la estimación por método de los momentos y no paramétrica en los puntos de estimación.

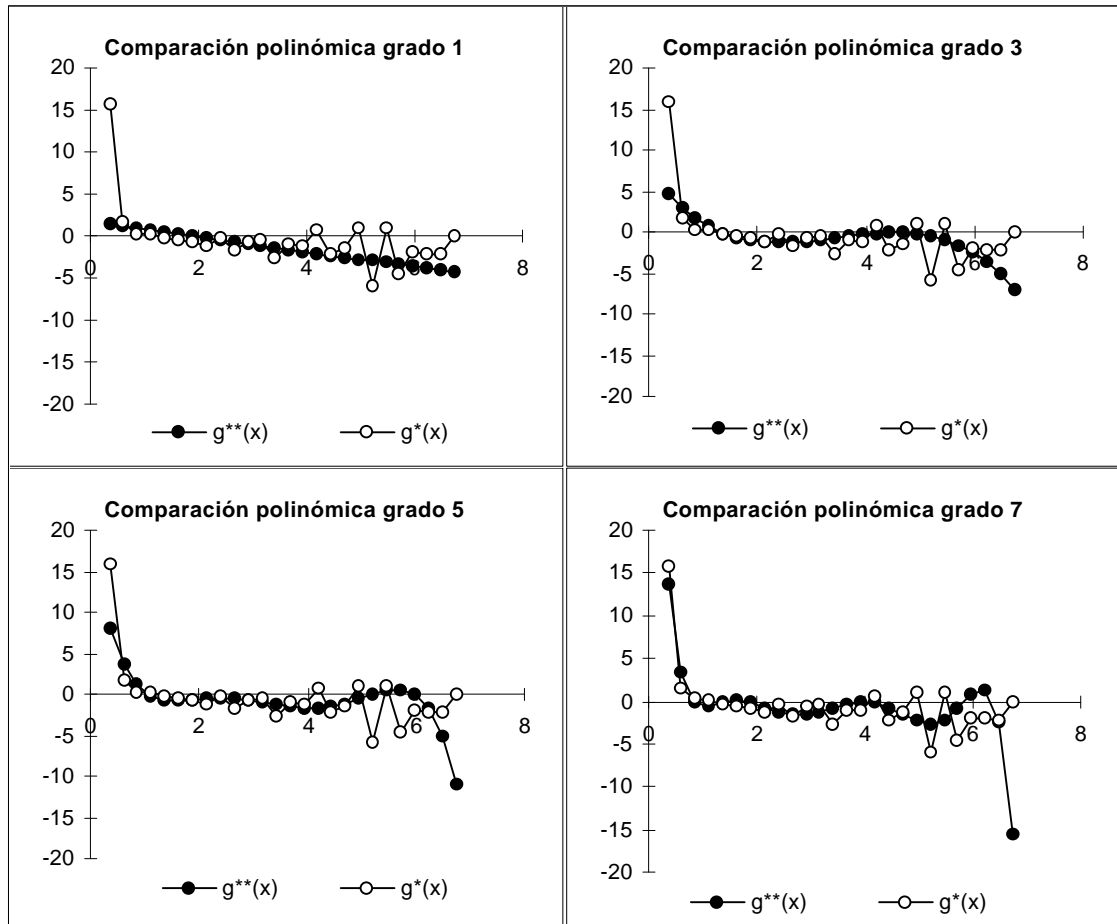


Figura IV. Comparación estimaciones por método de los momentos ($g^{**}(x)$) y mediante kernel ($g^*(x)$) de la generadora. $n=3600$, $h=0,15$

Posteriormente, se reconstruye la densidad como se indica en el apartado 5, observándose como la generadora por el polinomio de grado 3 modeliza adecuadamente la distribución empírica de la renta, (**Figura V**). La densidad obtenida para dicha generadora responde a la siguiente expresión:

$$f(x)=0.0099 e^{x(7.3529-4.3766 x+0.9256 x^2-0.0665 x^3)}$$

7. CONCLUSIONES.

El presente trabajo propone una metodología para modelizar la distribución de la renta que, por un lado, aprovecha la flexibilidad de los métodos de estimación no paramétricos y, por otro, la conveniencia de poseer una forma funcional que modelice la distribución observada. Para ello se utiliza la familia de

generadoras de tipo polinómico de grado impar que proporciona densidades (necesariamente) campanoides para la modelización de este tipo de fenómenos.

La estimación no paramétrica se utiliza como patrón para comparar y seleccionar el grado de la generadora polinómica que hace mínima la distancia cuadrático media entre dicha estimación y la obtenida por el método de los momentos que, además, coincide con la de máxima verosimilitud. Posteriormente se reconstruye la densidad correspondiente y se obtiene la forma funcional que modeliza la distribución observada.

La flexibilidad de la metodología presentada radica en que permite obtener una función de densidad de la renta adaptable a las múltiples y cambiantes formas que ésta adopta, siempre que no se pretenda “encorsertar” los datos muestrales en un intervalo finito y se permita que sea la propia densidad estimada la que delimite la zona en la cual se encuentra el “grueso” de la masa de renta dejando relativamente indefinido las rentas extremas, es decir, las colas de la distribución, siempre conflictivas si tratamos de acotarlas con unos valores determinados de complicada estimación.

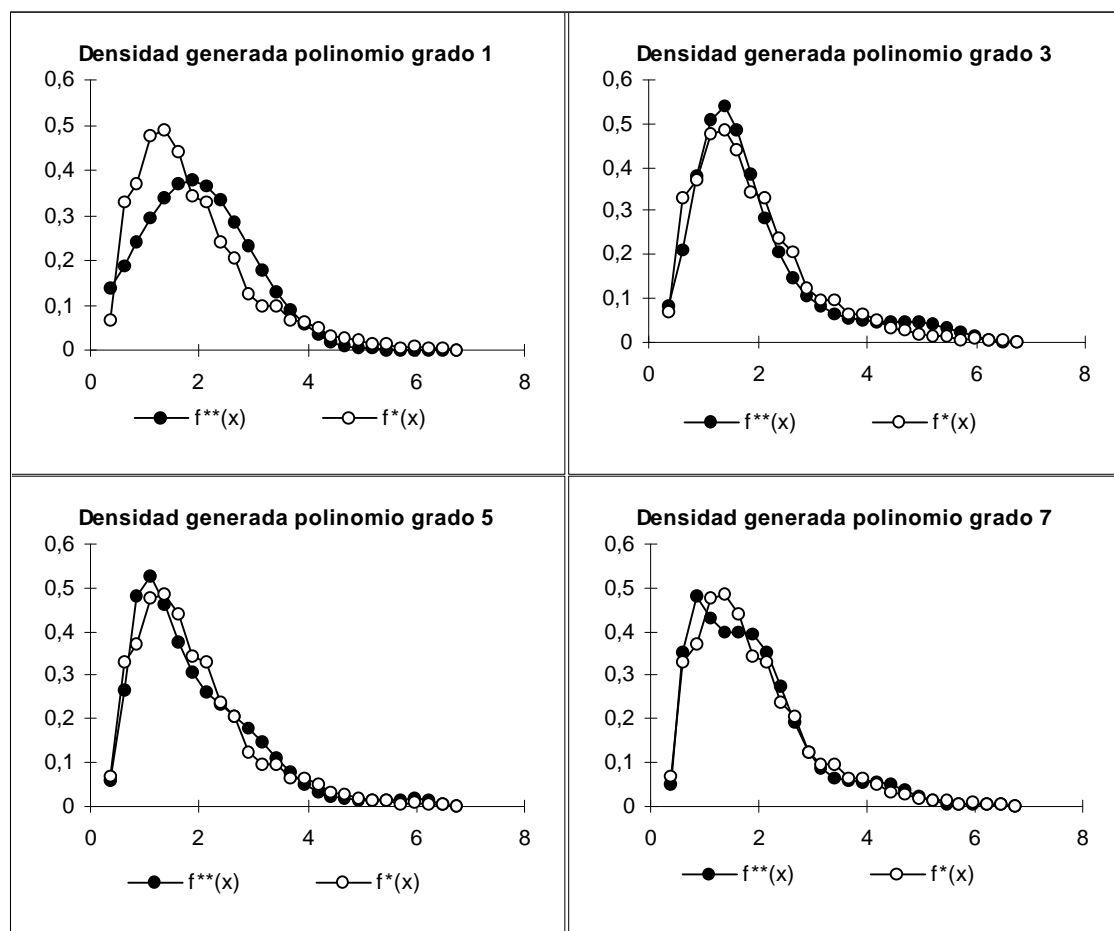


Figura V. Comparación de las densidades obtenidas por generadoras polinómicas ($f^{**}(x)$) con la estimación no paramétrica ($f^*(x)$).

AGRADECIMIENTOS:

Agradecemos la amabilidad de los Catedráticos de Economía Aplicada, D. Bernardo J. Pena Traperro, D. José M. Casas Sánchez y D. Francisco Javier Callealta Barroso de la Universidad de Alcalá de Henares, por habernos facilitado los datos originales utilizados en la elaboración de su trabajo: “Estudio de la Distribución Personal de la Renta en España”, (pendiente de publicación).

BIBLIOGRAFIA

Brachmann K., Stich A. y Trede M. (1995). Evaluating Parametric Income Distribution Models. Discussion Papers in Statistics and Econometrics. Seminar of Economic and Social Statistics University of Cologne. No. 6/95. Internet. <http://www.uni-koeln.de/wiso-fak/wisostatsem/papers/koelse9506.ps>

Callejón, J. (1995). Un nuevo método para generar distribuciones de probabilidad. Problemas asociados y aplicaciones. Tesis Doctoral. Universidad de Granada. ETD Micropublicaciones S.L.

Callejón J. y Santos M. (1994). Comparación de dos métodos de estimación: Estimación máximo verosímil y método de los momentos. VIII Reunión Anual de ASEPELT-España, Volumen I. Pág. 245-249. Universitat de les Illes Balears.

Dagum C. (1981). Sistemas Generadores de Distribución del Ingreso y la Ley de Pareto. Journal of the Inter-American Statistical Institute 125, pág. 143-183.

Del Oro, C.P. y Presedo M.A. (1992). Una Aproximación no Paramétrica a la Distribución de la Renta. Estudios de Economía Aplicada. Volumen I. Pag. 45-50. Universidad de Granada.

Esteban, J. (1986). Income-share Elasticity and the Size Distribution of Income. International Economics Review 27, pág. 439-444.

Härdle, W. (1991). Smoothing techniques with implementation in S. Springer-Verlag. New York Inc.

Johnson, N.L. y Kotz S. (1970). Continuous Univariate Distributions-1. John Wiley & Sons, New York.

McDonald, J.B. (1984). Some Generalized Functions for the Size Distribution of Income. Econometrica 52, pág. 674-663.

McDonald, J.B. y Xu, Y.J. (1992). A generalization of the Beta of the First and Second Kind with an Application. American Statistical Association: Proceedings of the Business and Economic Statistics Section, pág. 155-160.

Silverman, B.W. (1986). Density Estimation for Statistical and Data Analysis. Chapman and Hall.