DE LA LEY RANGO-TAMAÑO (RANK-SIZE A LA LEY LOG-NORMAL: LOS PROCESOS ALEATORIOS EN EL CRECIMIENTO DEMOGRÀFICO DE LOS AGREGADOS DE POBLACIÓN

Daniel Devolder
Albert Esteve

247

DE LA LEY RANGO-TAMAÑO (RANK-SIZE A LA LEY LOG-NORMAL: LOS PROCESOS ALEATORIOS EN EL CRECIMIENTO DEMOGRÀFICO DE LOS AGREGADOS DE POBLACIÓN

Daniel Devolder
Albert Esteve

247

Aquesta comunicació es va presentar al VI Congreso de la ADEH,
Granada 1-3 abril 2004

Centre d'Estudis Demogràfics

2004

ÍNDICE

ÍNDICE DE TABLAS

ÍNDICE DE GRÁFICOS

DE LA LEY RANGO-TAMAÑO (RANK-SIZE) A LA LEY LOGNORMAL: LOS PROCESOS ALEATORIOS EN EL CRECIMIENTO DEMOGRÁFICO DE LOS AGREGADOS DE POBLACIÓN (SETTLEMENTS)

RESUMEN

En esta comunicación pretendemos primero repasar las principales interpretaciones que se han hecho para explicar la regularidad de la ley rank-size, en la distribución del tamaño de la población de las ciudades de un país. Esto nos llevará a hablar de los procesos de crecimiento aleatorio y su traducción en términos de distribución lognormal de la población de los municipios o agregados. En segundo lugar presentaremos los resultados de un estudio de las entidades de población de Cataluña en el periodo 1857-1991 y comprobaremos en qué medida la distribución de estas poblaciones se ajusta a la ley de Zipf y a la ley lognormal. Para finalizar discutiremos algunos puntos débiles o problemáticos relacionados con estas leyes, por ejemplo: ¿En qué medida un ajuste *rank-size* perfecto se puede interpretar en términos de un equilibrio o un óptimo? ¿Qué significado tiene una *rank-size* con una pendiente negativa distinta de -1? ¿Por qué la ley *rank-size* no se verifica para poblaciones pequeñas, con un umbral variable según las distribuciones?

Palabras Clave: Rank-size, distribución lognormal, procesos aleatorios, Cataluña

RESUM

En aquesta comunicació, pretenem en primer lloc repassar les principals interpretacions que s'han fet per explicar la regularitat de la llei rank-size, en la distribució de la grandària de població de les ciutats en un país. Això ens portarà a parlar dels processos de creixement aleatori i la seva traducció en termes de distribució lognormal de la població dels municipis o agregats. En segon lloc presentarem els resultats d'un estudi de les entitats de població de Catalunya durant el període 1857-1991 i comprovarem en quina mesura la distribució d'aquestes poblacions s'ajusta a la llei de Zipf i a la llei lognormal. Per acabar, discutirem alguns punts febles o problemàtics relacionats amb aquestes lleis, com per exemple : en

quina mesura un ajustament *rank-size* perfecte es pot interpretar en termes d'equilibri o d'òptim? Quin significat té una *rank-size* amb una pendent negativa de –1? Perquè la llei rank-size no es verifica per poblacions petites, amb un llindar variable segons les distribucions?

Paraules clau: Rank-size, distribució lognormal, processos aleatoris, Catalunya.

Abstrat.

In this communication we try to revise first the principal interpretations that have been done to explain the regularity of the law rank-size, in the distribution of the size of the population of the cities in a country. This will lead us to speaking about the processes of random growth and lognormal distributions about the population of the municipalities or settlements. Secondly we will present the results of a study over the entities of population of Catalonia in the period 1857-1991 and we will verify in what measurement the distribution of these populations it adjusts to Zipf's law and to the law lognormal. To finish we will discuss some weak or problematic points related to these laws, for example:

In what measurement an adjustment rank-size perfect can be interpreted in term of a balance or the ideal one? What meaning has a rank-size with a negative different slope of-1? For what law rank-size it does not happen for small populations, with a changeable threshold according to the distributions?

DE LA LEY RANGO-TAMAÑO (*RANZ-SIZE*) A LA LEY LOG-NORMAL: LOS PROCESOS ALEATORIOS EN EL CRECIMIENTO DEMOGRÁFICO DE LOS AGREGADOS DE POBLACIÓN

1.- Las distintas aproximaciones al estudio de las ciudades y los sistemas de ciudades

El estudio del crecimiento de las ciudades y de los sistemas de ciudad, es decir la evolución de un grupo de ciudades con su red de relaciones, se puede hacer desde distintos puntos de partida. Berry (1964) los clasifica en dos grupos: las construcciones lógicas y las generalizaciones inductivas. Las primeras son intentos de explicar el proceso de distribución y organización de los asentamientos humanos en el territorio desde una perspectiva generativa. El ejemplo más conocido es la teoría de los lugares centrales de Christaller (1933). Al contrario, las generalizaciones inductivas empiezan con la observación empírica a partir de la cual se buscan generalizaciones, patrones o leyes matemáticas que permitan resumir de forma cómoda la información estudiada. El desarrollo más importante en este sentido es la ley rank-size (Zipf 1949). Pero esta distinción no deja de ser artificiosa, ya que una construcción teórica tiene que cotejarse con los datos, para comprobar la existencia de las regularidades previstas por la teoría, y al revés, la aproximación inductiva, si llega al descubrimiento de patrones lleva a la elaboración de explicaciones adhoc. Esto se verifica especialmente con el desarrollo reciente de la geometría fractal (Mandelbrot 1984), en la que se ha intentado sobrepasar esta división entre teoría y descripción o inducción, con la aportación de nuevas herramientas para la descripción, análisis y la elaboración de modelos de los sistemas de poblamiento (Frankhausser 1997), aunque este modelo quedará fuera de nuestra discusión.

1.1.- Construcciones teóricas: la teoría de los lugares centrales

Christaller basa su teoría sobre la idea de una organización regular del espacio en hexágonos sucesivos ocupados por una jerarquía de lugares centrales, cada uno de ellos envueltos por un conjunto de ciudades de rango inmediatamente inferior, estructura que se repite sistemáticamente en el territorio y para distintos niveles de escala. La distribución de los lugares centrales obedece a las leyes del poblamiento basadas en el papel de las ciudades como mercados, centros de transporte y concentradoras de la actividad económica. Como es obvio, se trata de un modelo teórico que no encuentra una aplicabilidad inmediata, puesto que éste supone un territorio homogéneo, sin diferencias algunas, en el cual se desarrolla una estructura puramente jerárquica, cuando lo primero que encontramos es un territorio heterogéneo. Se trata pues de un modelo estático, o unidimensional, que no incorpora otras dimensiones de los sistemas urbanos, deformados por el efecto de otras estructuras territoriales (Le Bras 1993, p.141). No obstante, Shuper, convencido que la teoría de los lugares centrales es suficientemente autónoma, propone una versión relativista que tiene en cuenta las múltiples variaciones que existen dentro de cada una de sus jerarquías (1999, p. 3)[1].

1.2.- Generalizaciones inductivas: las distribuciones *rank-size* y lognormal

1.2.1.- La ley de Zipf: rango-tamaño o *rank-size*

Con la ley *rank-size,* Zipf (1949) mostró la existencia de una relación constante entre el efectivo de población de ciudades y su rango dentro de la distribución ordenada de los mismos efectivos[2]. A partir de múltiples observaciones, tanto de distribución de tamaño o densidad de ciudades, como de otro tipo de distribuciones, Zipf demostró la presencia de regularidades en las que el tamaño de una ciudad era función del tamaño de cualquier otra ciudad de la distribución y del rango de las dos:

$$P(r_1) = P(r_2).\left(\frac{r_1}{r_2}\right)^{-b} \qquad [1]$$

donde *P(r)* es el efectivo de la ciudad de rango de *r* en la distribución ordenada, y *b* es el parámetro principal de la ley. Se puede dar a esta ley una representación que facilita la búsqueda de explicaciones del proceso generador. En efecto, si partimos de la ciudad

[1] La tesis doctoral de Pilar Riera (1988) es un intento de aplicación del modelo de Christaller a la red urbana de Cataluña, donde señala que se le dio demasiado importancia al elemento estructural de la teoría de los lugares centrales en detrimento del esfuerzo del propio Christaller para analizar las variaciones que pueden explicar los demás elementos estructuradores.

[2] Horacio Capel (1972) escribió un articulo sobre la validad de la ley que recogía las principales aportaciones. Se trata de uno de los pocos trabajos publicados en España sobre este tema. Analiza de los problemas con los datos, los relacionados con la escala, la extensión del sistema o la presencia de varios sistemas en un mismo ámbito.

de mayor tamaño, podemos encontrar el tamaño de cualquier ciudad menor con la relación:

$$P(r).r^{b} = P(1) = A \qquad [2]$$

Cuando se representa esta función en una escala doblemente logarítmica, la ley nos indica que tendríamos que encontrar una línea de pendiente menos *b*:

$$\ln(P(r)) = \ln(A) - b.\ln(r) = a - b.\ln(r) \qquad [3]$$

Zipf mostró que muchos fenómenos a priori no relacionados siguen esta ley, como por ejemplo la frecuencia de uso de las palabras, pero que el parámetro *b* tiende en las distribuciones más "perfectas" hacía un valor de 1. Es esta versión de la ley *rank-size* que se conoce como ley de Zipf, en la que la relación entre tamaño de las poblaciones en una situación óptima se simplifica:

$$P(r).r = A \qquad [4]$$

La segunda ciudad tiene una población la mitad de la primera, la tercera tiene una tercera parte de la primera, etc., es decir, que hay una relación sistemática entre el tamaño de todas las ciudades que verifican la ley.

La ley *rank-size* es equivalente a una ley de Pareto. Este último autor estudiaba la distribución de los ingresos de una población, y llegó a la conclusión de que la frecuencia de personas con ingresos superiores a una cifra *S* era proporcional a *1/S*. Esto se puede expresar a partir de la ecuación [2] de la manera siguiente:

$$P(r)^{\beta}.\frac{r}{N} = \frac{A}{N} = a \qquad [5]$$

donde β el exponente de la relación inversa, del rango en función del tamaño de la población, y *N* es el número total de ciudades o individuos que se estudian. Entonces *r/N* es la frecuencia del número de casos con un tamaño superior o igual a *P(r)*:

$$\Pr(P \geq P(r)) = f(r) = \frac{r}{N} = \frac{a}{P(r)^{\beta}} \qquad [6]$$

y para la ley de Zipf, es decir un valor del exponente igual a 1, tenemos:

$$\Pr(P \geq P(r)) = \frac{a}{P(r)} \qquad [7]$$

es decir, la probabilidad (o la frecuencia de casos) de las ciudades con una población superior o igual al tamaño *P(r)* es inversamente proporcional a esta misma cifra de población.

La ley *rank-size* supone un relleno sistemático del espacio por aglomeraciones cada vez menos importantes, sin que haya el efecto de jerarquía de Christaller, en las que ciudades de un mismo nivel tendrían que tener un efectivo similar. No obstante la existencia de esta regularidad ha llevado a muchos autores a formular teorías o modelos susceptibles de generar el orden observado a través de la ley. Así, el propio Zipf sugiere que podría ser la manifestación de la ley del menor esfuerzo o del coste mínimo, que en su versión geográfica postula que las ciudades grandes se crean allí donde se optimiza su localización en relación con los asentamientos de menor rango que las rodean y con las cuales mantiene relaciones diversas. La localización óptima tiene en cuenta el volumen de los intercambios con el resto de los asentamientos:

> "Pour expliquer la répartition des individus entre des agrégats de taille différente, Zipf a invoqué l'effet conjugué de deux forces opposées: une force de diversification selon laquelle un très grand nombre de petites villes se développent, de manière à se trouver le plus près possible des matières premières et des sources d'énergie, elles-mêmes spatialement dispersées, et donc à minimiser les coûts de transport à la production; à l'inverse, une force d'unification tend à faire surgir un petit nombre de très grandes villes pour minimiser les coûts de transport au consommateur." (Pumain 1982, p. 39).

La forma de los sistemas urbanos está ligada al campo de fuerzas que atraen o repulsan la actividad urbana. Una vez creado el sistema de relaciones, el crecimiento puede seguir por una tensión distinta:

> "once inter-spatial population densities set in, then they are reinforced by scale factors in agglomeration effects by such density differentials" (Dendrinos, 1992, p. 164).

Woldenberg (1971) y Coffey (1981) han interpretado esta relación constante como el resultado de la interacción entre las fuerzas de crecimiento alométrico y isométrico. El primer tipo supone que hay asentamientos que crecen de forma más rápida que otros como consecuencia de la competencia para apoderarse de unos recursos limitados. El crecimiento alométrico tiende a modificar el valor del exponente *b* de la ley de Zipf. En cambio el crecimiento isométrico supone que las variaciones son proporcionales, con una tasa igual para todas las ciudades, sin hacer cambiar el valor del exponente de Zipf.

Pero el éxito de la ley *rank-size* no se debe a la argumentación teórica surgida, sino obviamente a su verificación empírica, por imperfecta que esta sea a veces[3]. Según Guérin (1995) la distribución *rank-size* se observa raramente. Berry (1964) señala que se da sobre todo en los casos con un índice de primacía bajo (proporción baja de la población de la primera ciudad en relación con las cinco más pobladas) mientras que para valores altos, el modelo es más del tipo *primate city size distribution*: la población de la ciudad principal difiere del valor esperado por un sobredesarrollo relacionado con el resto de los asentamientos del sistema urbano. Pero a pesar de sus imperfecciones, Guérin recomienda utilizar la aproximación *rank-size* porque proporciona una descripción sencilla de la organización jerárquica de un sistema (1995, p. 553), cuyos parámetros se pueden utilizar como indicadores para describir su evolución.

En este sentido es interesante el análisis de Shuper (1999) que muestra para el caso de Moscú que la capacidad predictiva de la ley *rank-size* se deteriora en la misma medida que mejora el ajuste por un modelo basado sobre la teoría de los lugares centrales. El autor considera la ley de Zipf como un estado anterior a la distribución jerárquica propia a la lógica de los lugares centrales, y esta transformación se hace siguiendo un proceso que él llama de "christallerización".

[3] Desde el punto de vista epistemológico, la descripción matemática de un fenómeno, con independencia de la precisión con la cual la formula describe los datos, no es una explicación (Fonseca 1989).

Otra limitación de la ley *rank-size*, a la cual daremos mucha importancia a continuación, es que la relación constante que se observa entre el logaritmo del rango y el logaritmo del tamaño se pierde por debajo de un umbral de población. Dicho de otra manera, cuando se incorporan todos los asentamientos de un territorio, las propiedades de la *rank-size* dejan de observarse totalmente. En este sentido, el término de sensibilidad se utiliza para describir como se comporta esta ley ante distintos umbrales de población, es decir, el punto a partir del cual se pierde la relación constante, y que varia según las regiones y los periodos. Algunos autores, para defender la ley de Zipf, interpretan este punto de ruptura como la frontera entre población urbana y rural, hipótesis que contrastaremos más adelante.

Finalmente, un tema importante es el significado del valor del exponente *b* de la ley. Para Zipf el valor de 1 parece ser el límite hacía el cual tienden todas la distribuciones que verifican su ley. El ejemplo que da es él de los sistemas urbanos plenamente desarrollados, como el norte americano, para el cual una sociedad de libre mercado permite el juego de las fuerzas del menor esfuerzo y de optimización del acceso a los recursos, a diferencia del sistema soviético, para el cual el valor del parámetro de la ley esta muy alejado de 1. Pero para algunos investigadores no es una preocupación que este exponente no se aproxime o no tienda hacía 1 (Fonseca 1989).

¿Pero qué sentido se puede dar a un valor límite del exponente de Zipf? Le Bras (1994) rechaza el hecho de hablar en término de óptimo. El autor considera que es una falacia hablar de poblamiento equilibrado en los casos en los que se observa un ajuste perfecto, con un valor del exponente de 1. Para él se trata de una ley empírica, que no se asocia a un proceso social, y por lo tanto no se le puede atribuir un significado normativo a un valor determinado del exponente de Zipf.

1.2.2.- La ley lognormal

Distintos autores (Pumain 1982, De Cola 1985) han observado que la distribución lognormal es especialmente adecuada para estudiar los sistemas de poblamiento. Se define como una variable aleatoria cuyo valor exponenciado sigue una distribución normal. La clase de fenómenos que se asocian a la lognormal se basan en el efecto proporcional o ley de Gibrat, es decir, los procesos proporcionalmente aleatorios en los que una variable depende de su valor anterior por una función multiplicativa cuyo parámetro sigue una normal (Aitchison y Brown 1966, p 22).

Para fijar las ideas, tomamos el caso hipotético de un conjunto inicial de asentamientos, por ejemplo 1 millón, con una cifra de población igual, pongamos 100 personas. Suponemos que estos asentamientos experimentan un proceso de crecimiento al azar, con una tasa de crecimiento que varia de forma aleatoria, siguiendo una distribución normal de valor medio positivo, por ejemplo de 1% anual, con una desviación estándar también de 1%, durante un periodo largo de por ejemplo 1.000 años. Al final de este periodo la distribución de estos asentamientos seguiría una lognormal, es decir con muchos asentamientos con un tamaño de la población inferior al valor modal, pero con unos pocos con un tamaño muy elevado, superior en algunos casos al millón de personas. En este modelo el tamaño de las poblaciones son el producto de números aleatorios por el valor de la constante inicial, lo que es uno de los procesos generadores de una distribución lognormal (unas cadenas de Markov de tipo multiplicativo).

Hay distintos aspectos de este modelo que es conveniente discutir, siguiendo la presentación moderna de Gabaix (1999). Primero la cifra inicial de las poblaciones no tiene que ser igual para todos. De hecho los asentamientos pueden tener una distribución inicial cualquiera, ya que lo que realmente importa es que el proceso multiplicativo o de crecimiento sea independiente del tamaño de las poblaciones, es decir la tasa de crecimiento no dependa del efectivo de la población del asentamiento. Segundo, el tiempo de convergencia hacía una distribución lognormal es de hecho bastante más corto que 1.000 años, y Gabaix indica que sus simulaciones de Monte Carlo sugieren valores de 100 años o menos como suficientes para la convergencia. Tercero está el problema de saber que hacemos con las poblaciones cuyo efectivo tienda hacía valores pequeños o incluso negativos. Si los incluimos, la distribución límite de la población de los asentamientos tiende hacía la lognormal, en cambio si establecemos un valor inferior por debajo del cual se eliminan los asentamientos pequeños, la distribución límite es la de una función potencia de tipo Pareto o *rank-size*.

El significado de esto es que cuando se representan las variables de una distribución lognormal en función de su rango en un gráfico de escala doblemente logarítmica, se obtiene una recta para los valores más elevados, hasta un umbral a partir del cual se trata de un curva con una convexidad orientada hacía arriba. Dicho de otra forma, la ley *rank-size* equivale a una distribución lognormal truncada a un determinado valor umbral. Las dos representaciones ofrecen el mismo resultado para valores elevados, pero cuando se incorporan la totalidad de los asentamientos de un territorio, la distribución lognormal ofrece un mejor ajuste de los datos (Pumain 1982, p. 28).

En el contexto de este modelo de crecimiento aleatorio, se puede construir una explicación de porque la distribución de las ciudades tiende hacía una ley de Zipf, es decir porque tiende hacía una ley *rank-size* con un exponente igual a 1. Primero es fácil convencerse de que la distribución tiene que aproximarse a una ley *rank-size*, puesto que el hecho de suponer que el crecimiento es en probabilidad según una tasa igual para todos, remite a una distribución que es igual para cualquier escala, y se puede mostrar que este tipo de distribución que es invariante para la escala se obtiene con una ley potencia, equivalente a la ley *rank-size*. La explicación de porque el exponente de la función potencia tiene que ser 1 de nuevo remite al tipo de crecimiento de las ciudades, que por hipótesis es aleatorio, y igual para todas. En estas condiciones la esperanza matemática del crecimiento para cada periodo es igual para todas las ciudades, con lo cual la probabilidad de crecer más que la media por un factor multiplicador m es siempre igual a $1/m$ de la probabilidad de crecer menos que la media por un factor multiplicador $1/m$. Es como si hubiese un mecanismo de compensación que garantiza que el peso relativo de la población de las ciudades grandes es el mismo que él de las ciudades más pequeñas.

Otro punto interesante en este contexto es que el hecho de que la ley *rank-size* sea equivalente a una ley de Pareto o función potencia que rige para muchos fenómenos de tipo aleatorio que dan lugar a patrones que parecen ordenados, tener una cierta regularidad. Estos fenómenos han sido estudiados por el padre de la geometría fractal, lo que hace que estas leyes potencia sean a veces conocidas también como ley de Mandelbrot (1984). Esto plantea una pregunta importante acerca de saber qué importa más en el crecimiento urbano, si el papel de los factores económicos o bien el juego del azar, pregunta que remite obviamente al significado normativo o no de estas leyes.

2.- Aplicación de las leyes *rank-size* y lognormal al sistema de poblamiento en Cataluña

Utilizaremos aquí la ley *rank-size* y la distribución lognormal para explorar la presencia de procesos aleatorios en el crecimiento de la población de las entidades de población de Cataluña durante el periodo 1887-1998. Verificaremos en primer tiempo el ajuste a nuestros datos de la ley *rank-size*, pero siguiendo la argumentación de Le Bras (1994), no interpretaremos las discrepancias entre distribución observada y esperada en termino de desequilibrio poblacional, sino de aproximación a una distribución aleatoria de los crecimientos, es decir, aplicaremos en todo momento el marco interpretativo del ajuste por una lognormal. En este contexto, las discrepancias se explicarían por la perdida del carácter aleatorio, y la explicación del tamaño de ciertos asentamientos por factores sistemáticos de carácter por ejemplo político.

Esta interpretación de las desviaciones en términos de pérdida del carácter aleatorio se verá refrendada por el uso de distribuciones lognormal que, como lo indicábamos en la sección anterior, parecen ajustar mejor los datos, especialmente si se consideran todos los asentamientos incluidos los más pequeños.

2.1.- El uso de las entidades de población como unidad de análisis

Este trabajo presenta como elementos novedosos i) el uso de las entidades de población como unidad de análisis ii) la inclusión de la practica totalidad de las mismas y iii) la profundidad histórica.

La entidad de población es la unidad geográfica principal de nuestro análisis, que se puede estudiar a partir de la explotación de los Nomenclátores, publicación que acompaña los Censos de Población en España desde el primer censo moderno de 1857[4]. A diferencia de los municipios u otras unidades territoriales administrativas, la entidad es una unidad de tipo estadístico establecida según un criterio morfológico, que permite una mejor aproximación, al margen de los problemas que pueda tener la fuente, a la distribución del poblamiento. Por definición las unidades estadísticas son comparables entre sí, porque se establecen a partir de criterios de aplicación homogéneos en todo el territorio. En cambio, no se puede decir lo mismo de las unidades administrativas, aunque tienen otras ventajas: permanencia en el tiempo, fronteras más precisas y disponibilidad de muchas otras fuentes de datos a esta misma escala. De todas formas esto se tiene que matizar, ya que por ejemplo el mapa municipal de Cataluña cambió de forma significativa en el tiempo, y el nombre de municipios se redujo de cerca de 140

[4] Un análisis más exhaustivo se puede encontrar en la tesis doctoral del primer autor de este trabajo (Esteve 2003).

entre 1857 y 1998, lo que obliga a penosas correcciones si se quiere homogeneizar las divisiones municipales en el tiempo.

En cambio, la utilización de la entidad como unidad de análisis evita en parte los problemas de los cambios en la división municipal. Bien es cierto que, indirectamente, la estructura municipal sigue condicionando el listado de entidades, por la sencilla razón de que cada entidad pertenece a un solo municipio. Hay un número de casos significativo de entidades que de hecho sobrepasan o superan la superficie municipal, pero la fuente no permite captarlas porque esta condicionada por la estructura municipal, que lleva a contabilizar tantas unidades como municipios sobre los cuales se extienden estas macro entidades. Este problema se manifiesta sobre todo en las zonas urbanas, como el área metropolitana de Barcelona. Pero a pesar de los problemas de la fuente, las entidades son las mejores unidades estadísticas para una aproximación detallada y fidedigna al estudio del poblamiento durante un periodo de tiempo muy amplio.

Utilizar las entidades permite trabajar con unidades que contienen una población reducida, y hace posible observar la bondad del ajuste con varios tipos de distribución, aquí la *rank size* y la lognormal cuando no se excluyen ningún tipo de asentamiento humano en un territorio. De hecho hay muy pocos ejemplos parecidos al nuestro, y en la mayoría de los trabajos parecidos se estudian las poblaciones de municipios o ciudades que sobrepasen un tamaño en general elevado (superior a 5.000 o 10.000 habitantes). Esta laguna se explica o bien por la disponibilidad de los datos, o bien por una elección consciente de los investigadores. En este último caso, estudiar las ciudades claramente urbanas facilita de forma considerable el análisis y el ajuste de funciones matemáticas o distribuciones de probabilidad. No obstante, una ventaja muy significativa de considerar la totalidad de los asentamientos es la posibilidad de estudiar los umbrales a partir de los cuales cambia la forma de los patrones estudiados, lo que permite estudiar la sensibilidad de un sistema en el que varían de forma significativa las propiedades del modelo cuando se incorporan entidades cada vez más pequeñas.

En cuanto a la dimensión histórica, no existe que sepamos ningún trabajo que analice el sistema de poblamiento en Cataluña, a esta escala y para un periodo tan amplio. Eso es posible gracias a la existencia del Nomenclátor. Pero a pesar de que esta fuente remonte a 1857, acompañando la publicación de los censos, es solamente a partir de 1887 que se publican la cifra de población de las entidades, de manera continuada. El periodo abarcado por nuestro estudio va de 1887 a 1998. Se trata de un intervalo de 110 años, en el que el poblamiento de Cataluña experimentó transformaciones profundas.

No es la primera vez que se ha llevado a cabo un análisis *rank-size* en el ámbito de Cataluña. Hay experiencias anteriores, aunque en ninguna de ellas se utilizó la entidad como unidad de análisis para un periodo tan amplio. El antecedente más reciente corresponde al *Atlas Socio-Econòmic de Catalunya* (Vidal Bendito 1980), en el que se analizan la distribución de la población de las entidades en el año 1970 utilizando este método, pero limitándolo a las de más de 2.000 habitantes. Es un estudio metodológicamente muy interesante, ya que compara los resultados obtenidos con los correspondientes a las entidades, municipios y aglomeraciones a partir de los datos de 1860. Exceptuando este estudio, en el resto de los casos, se ha utilizado siempre el municipio como unidad principal (Miralles, Rosés y Armet 1982, Camps 1990, Oliveras 1994).

2.2.- Presentación de los datos

Los datos utilizados aquí provienen de forma exclusiva del Nomenclátor de los Censos de Población. No hemos utilizados los datos de los años 1857 y 1940 debido a la definición de entidad, incompatible con la de otros años, y tampoco hemos utilizado los datos de los años 1860 y 1873 esta vez por la ausencia de cifras de población de las entidades. No hemos incluido las entidades con menos de 50 habitantes, con el objetivo principal de mejorar la consistencia temporal de las series. No hemos solucionado el problema de las aglomeraciones urbanas que superan los límites municipales y que tendrían que constar como una entidad única, pero es importante tener en cuenta este aspecto a la hora de comentar los resultados. Este es un problema que afecta sobre todo a la ciudad de Barcelona y su área metropolitana y también ciudades que acaban juntándose como son las de Sabadell, Barbera del Vallés, Sardañola del Vallés y Ripollet entre otras.

2.3.- Ajuste de las distribuciones *rank-size* y lognormal

El ajuste de la *rank-size* se puede obtener por el método clásico de la regresión lineal, una vez aplicado una transformación logarítmica a los datos. Se obtienen el valor de 2 parámetros, el primero la constante que indica la cifra estimada de la población mayor y el segundo es el coeficiente de regresión que indica la pendiente de la relación *rank-*

size, lo que corresponde al valor del exponente *b*. Este exponente hace referencia al grado de diferencia de la población de las unidades analizadas. Estas diferencias serán mayores para valores más elevados de *b*, y más pequeñas para valores más bajos de *b*. Si observamos que determinadas poblaciones se alejan de la recta de regresión, esto significa un sobredimensionamiento para valores superiores al valor esperado, al revés un subdimensionamiento para valores inferiores al esperado.

Para comprobar si la distribución de la población de las entidades corresponde con una lognormal, hemos procedido a una agrupación según el efectivo. Hemos utilizado los intervalos de la clasificación propuesta por Eurostat y las Naciones Unidas (1998), con algunas modificaciones. Se trata de agrupar las entidades calculando el logaritmo del efectivo de sus poblaciones. La clasificación utilizada presenta la ventaja de presentar intervalos fáciles de interpretar, y a la vez que definen intervalos aproximadamente constantes en término de su valor logarítmico, lo que es apropiado para este ejercicio. Hemos agrupado todas las entidades con una población mayor de 50.000 habitantes, lo que se podría criticar, pero que corresponde a la realidad del poblamiento catalán, con muy pocos casos superiores a este umbral.

Cuadro 1. Clasificación de las entidades según su población para el ajuste lognormal

Categoria	Entidades mayores de......habitantes
1	49
2	100
3	250
4	500
5	1000
6	2500
7	5000
8	10000
9	25000
10	50000

Para calcular las frecuencias esperadas en cada categoría según una distribución lognormal y compararlas con las observadas, hemos utilizado el método propuesto por De Cola (1995). La formula para estimar las frecuencias absolutas acumuladas para distintos límites de la población es:

$$n^k = n^1[1-F(Z^k)],$$

Donde n^k es la frecuencia acumulada hasta el límite *k*, n^1 es el número total de entidades y $F(Z^k)$ es la función e probabilidad normal, cuyos parámetros μ y σ tienen una interpretación similar a los de la *rank*-size. La concentración o desconcentración de la población de las entidades esta expresada por la desviación estándar σ y el crecimiento del sistema urbano en conjunto esta representado por el valor medio μ.

2.4.- Resultados

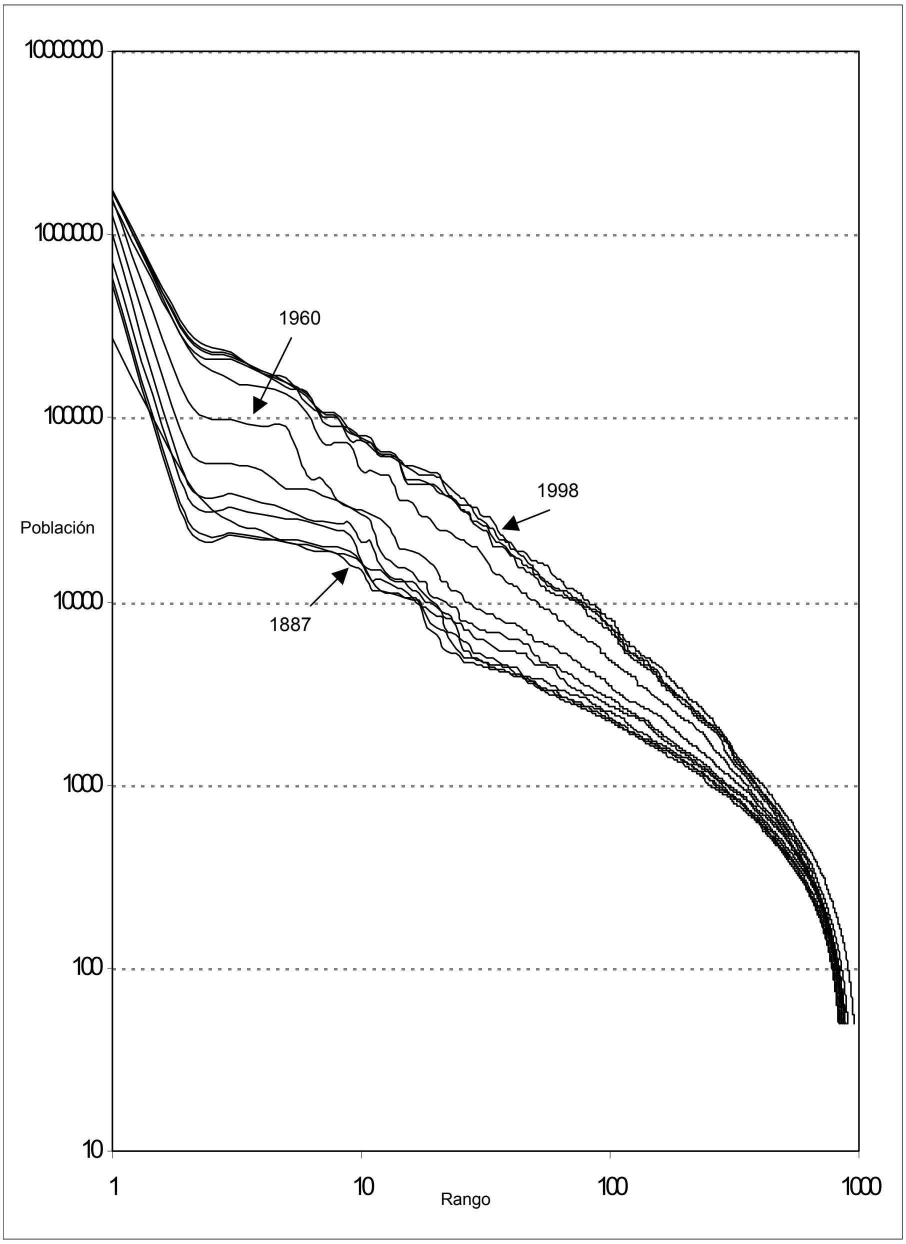
2.4.1.- La aproximación rank-size

La población de Cataluña no se reparte de forma equitativa entre las distintas entidades, ni al inicio ni al final del periodo estudiado. Las transformaciones en el poblamiento han sido intensas en el periodo analizado, fruto del crecimiento intenso en algunas áreas y el despoblamiento en otras. La cifra total de población era de 1.841.000 habitantes en el año 1887 y de 6.144.000 para el año 1998, y este crecimiento neto de 4.303.000 no se ha repartido de forma igualitaria entre las entidades, ni ha sido constante en el tiempo.

El gráfico 1 muestra las curvas *rank-size* de las entidades de Cataluña desde 1887 hasta 1998. El predominio de la entidad más grande, de Barcelona, es evidente en todos estos años, y lo sería aún más si se considerase la continuidad urbana, agregándola con la población de Hospitalet del Llobregat, Badalona, etc. Este sobredimensionamiento no solamente afecta a Cataluña, sino a distribuciones de otros países como por ejemplo es el caso de Paris para Francia.

A nivel de la evolución histórica constatamos que el crecimiento demográfico es generalizado para todas las entidades, puesto que la curva de 1998 esta por encima de la curva de 1887 para todos los rangos. Observamos que para las entidades comprendidas entre la segunda y la número cien, se observa la relación lineal esperada por la *rank-size*, pero para las entidades más pequeñas se observa una ruptura de la linealidad a favor de una curva convexa. El punto de inflexión o de ruptura se sitúa entre 500 y 1.000 habitantes. Algunos autores interpretan este punto como la frontera entre el ámbito rural y urbano, pero para el caso de Cataluña la cifra de 1.000 habitantes es inferior a la que utiliza el INE en el censo de 1991 para establecer la diferencia entre urbano y rural, es decir, una cifra de 2.000 habitantes.

Gráfico 1. Curvas rank-size, Cataluña 1887-1998

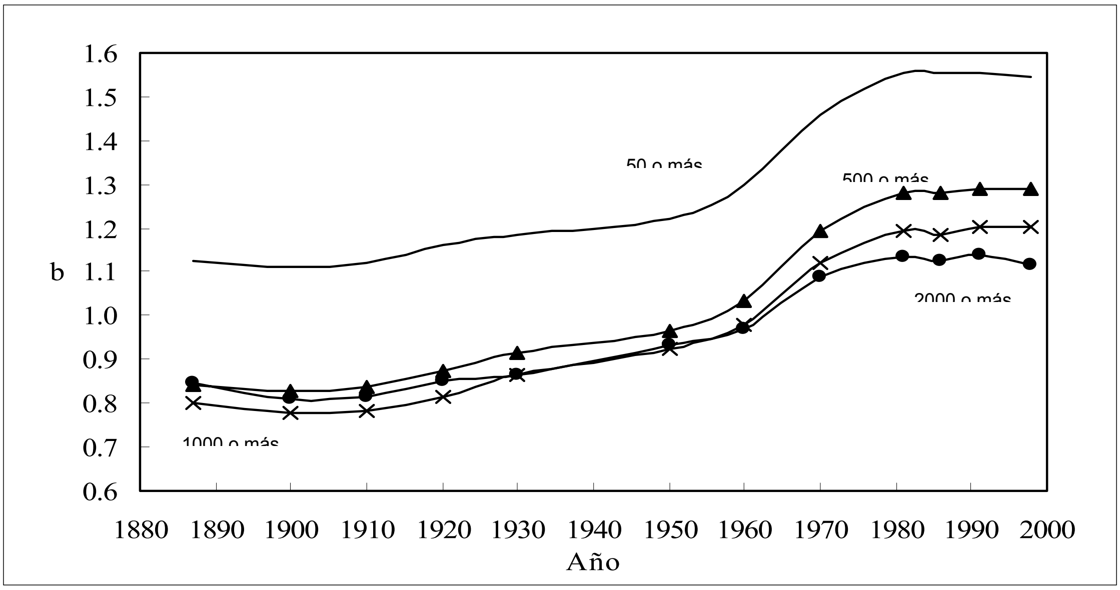


Fuente: Elaboración propia a partir de los datos del Nomenclátor.

Hemos ajustado unas rectas de regresión por el método de los mínimos cuadrados, para distintos grupos de entidades: todas, las mayores de 500, las mayores de 1.000 y las mayores de 2.000 habitantes. Esto permite apreciar la sensibilidad del método a la variación del umbral de población. Los resultados figuran en el gráfico 2 que muestra la evolución del parámetro *b* de pendiente de la recta (equivalente el exponente *B* de la ley *rank-size*) para los distintos umbrales.

La principal conclusión para la disposición de las curvas en el gráfico es que la inclusión de todos los asentamientos tiene efectos en el ajuste de la distribución *rank-size* que se manifiesta en el nivel, pero no en la forma de las curvas. El no incluir las entidades menores de 500 habitantes representa una diferencia significativa por lo que hace el nivel del parámetro *b*, si se compara con la curva que incorpora todas las entidades. De todas formas, a partir de este umbral no se observan diferencias significativas por lo que hace el nivel de *b*. Esto quiere decir que incluir o no las entidades menores de 1.000 o de 2.000 habitantes no cambia de forma significativa los resultados cuando se han eliminado previamente las entidades menores de 500 habitantes. Esto significa que esta cifra de 500 marca el punto de ruptura para el ajuste de la ley, ya que la pendiente cambia de forma significativa cuando se incluyen estas entidades con poca población.

Gráfico 2. Evolución del parámetro *b* para diferentes umbrales de la población, 1887 – 1998



Fuente: Elaboración propia a partir de los datos del Nomenclátor.

La distribución *rank-size* perfecta, es decir con una pendiente de –1 se observa solamente en el año 1960 para las curvas que no incluyen los asentamientos menores de 500 habitantes (gráfico 2). Para todos los años considerados, el valor del parámetro *b* que se obtiene cuando se consideran todos los asentamientos es siempre superior a 1, y en este caso se puede hablar de alometría positiva, es decir que la distribución de los crecimientos beneficia a las entidades más grandes.

Cuadro 2. Coeficientes de correlación (R) y de regresión (*b)* para distintos umbrales de población, 1887 – 1998

Nomenclátor	Umbral	Pendiente (b)	R
1887	50	-1.123	-0.942
	500	-0.844	-0.993
	1000	-0.800	-0.991
	2000	-0.847	-0.983
1900	50	-1.112	-0.941
	500	-0.830	-0.988
	1000	-0.779	-0.981
	2000	-0.812	-0.962
1910	50	-1.122	-0.941
	500	-0.838	-0.988
	1000	-0.785	-0.982
	2000	-0.815	-0.963
1920	50	-1.163	-0.944
	500	-0.875	-0.988
	1000	-0.815	-0.985
	2000	-0.852	-0.969
1930	50	-1.185	-0.949
	500	-0.916	-0.990
	1000	-0.863	-0.986
	2000	-0.864	-0.973
1950	50	-1.220	-0.952
	500	-0.966	-0.993
	1000	-0.925	-0.990
	2000	-0.931	-0.980
1960	50	-1.297	-0.956
	500	-1.034	-0.993
	1000	-0.980	-0.993
	2000	-0.972	-0.987
1970	50	-1.458	-0.963
	500	-1.193	-0.995
	1000	-1.123	-0.997
	2000	-1.089	-0.998
1981	50	-1.555	-0.964
	500	-1.282	-0.992
	1000	-1.193	-0.995
	2000	-1.132	-0.997
1986	50	-1.555	-0.963
	500	-1.278	-0.991
	1000	-1.183	-0.995
	2000	-1.124	-0.997
1991	50	-1.553	-0.965
	500	-1.292	-0.992
	1000	-1.204	-0.994
	2000	-1.137	-0.996
1998	50	-1.547	-0.966
	500	-1.290	-0.990
	1000	-1.203	-0.992
	2000	-1.116	-0.995

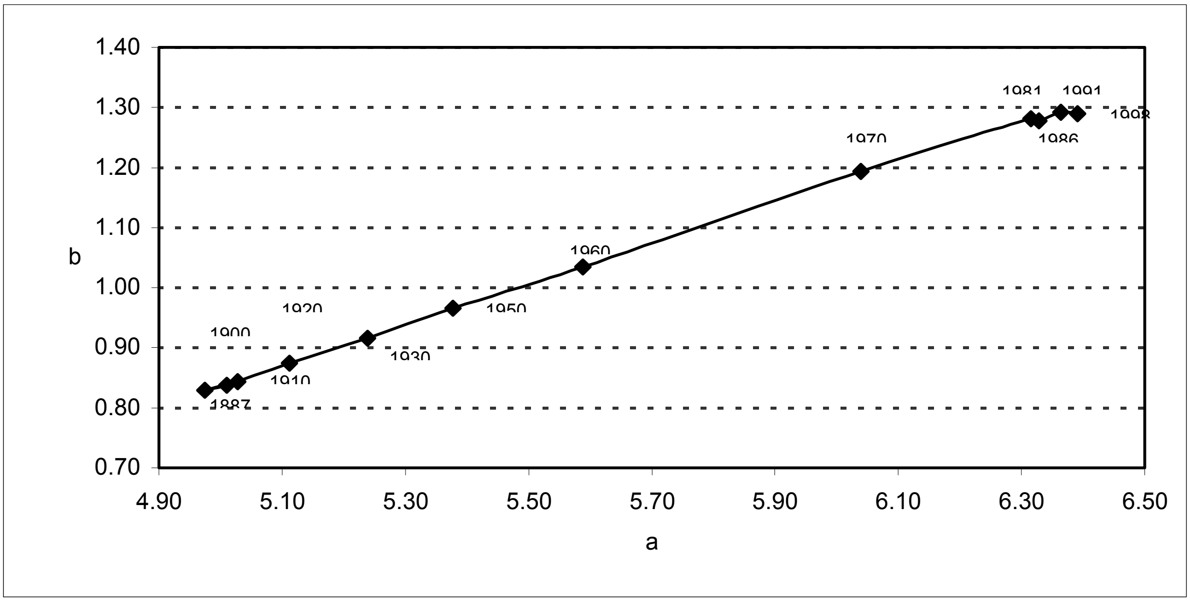
Fuente: Elaboración propia a partir de los datos del Nomenclátor.

En el año 1960 también hay una separación para los umbrales 500, 1.000 y 2.000 para los cuales se pasa del crecimiento alométrico positivo a negativo, lo que quiere decir que se pasa de una situación en la cual se benefician las entidades menos pobladas a una otra que beneficia a la más pobladas. Cuantas menos entidades haya, las desigualdades son menores, bien es cierto que a partir de un cierto umbral, en este caso de 500 habitantes, las diferencias no son tan importantes. Pero a partir de 1960 el fuerte crecimiento desigual de la población hace que las diferencias entre incluir o no las entidades de más de 500, 1.000 o 2.000 habitantes empiecen a ser significativas. De todas formas, la forma de las curvas es siempre la misma: estabilidad hasta 1910, aumento lento de las desigualdades hasta 1960, aceleración de las desigualdades entre 1960 y 1981, y finalmente estabilidad con tendencia a un decrecimiento de estas desigualdades entre 1981 y 1998. Esto quiere decir que, cualquier sea el nivel del umbral utilizado, se pueden extraer las mismas conclusiones sobre la cronología y la tendencia del poblamiento de Cataluña.

En resumen, el aumento del valor de la pendiente b más allá del valor de 1 muestra que hay factores no aleatorios que hacen que determinadas entidades crecen más que las otras, y por lo tanto su crecimiento no obedece a una hipótesis de tasa de crecimiento igual en probabilidad para todos los asentamientos. La desigualdad en la distribución de la población de las entidades de Cataluña se ha incrementado más de lo que se podría esperar en base a las hipótesis planteadas.

El cuadro 2 presenta los coeficientes de correlación (R) y de regresión (b) del ajuste entre Log de la población y Log de los rangos. Es importante señalar, como lo notamos anteriormente, la no inclusión de las entidades menores de 500 habitantes mejora la calidad del ajuste, medido por R, pero a partir del umbral siguiente, la mejora significativa no es sistemática. Hasta 1960 el mejor ajuste es él que incluye las entidades de más de 500 habitantes, cuando a partir de este momento el mejor ajuste es para las entidades mayores de 2000 habitantes. Esto quiere decir que el umbral a partir del cual hay un cierto deterioro va aumentando con el tiempo, constatación que juega a favor de la hipótesis del umbral variable introducida por ejemplo por Mendizábal (1994) en su artículo sobre el fenómeno urbano a Cataluña. La explicación que se puede dar al hecho de que el umbral no sea constante es que no representa lo mismo un asentamiento de 1.000 habitantes a principios del siglo XX que al final de este mismo siglo.

Gráfico 3. Evolución de los parámetros *b* y *a* (entidades >= 500 habitantes), 1887-1998



Fuente: Elaboración propia a partir de los datos del Nomenclátor.

La representación conjunta de los parámetros *a* y *b* de la ley rank-size (gráfico 3) para las entidades mayores de 500 habitantes muestra que el sistema de entidades de Cataluña ha pasado de un estado a otro en el intervalo de 70 años, de 1910 a 1981. El parámetro *a* es un indicador del crecimiento del sistema urbano en conjunto y *b* de la desigualdad de este crecimiento. La relación entre los dos parámetros es estable, y cuando uno se modifica el otro también lo hace de forma coherente. El aumento de las desigualdades va acompañado de un aumento generalizado del sistema que acentúa la situación de desigualdad inicial. A juzgar por la distribución de los puntos, la transformación más intensa del poblamiento se dio entre 1960 y 1970, lo que coincide con la década en la que Cataluña experimentó el mayor crecimiento demográfico, básicamente debido a la inmigración.

A tenor de los resultados obtenidos en este ejercicio, la conclusión es evidente: la distribución de la población por entidades de Cataluña no obedece a una ley *rank-size* perfecta, que se define como la que tiene una pendiente negativa igual a 1, es decir una ley de Zipf. También se tiene que valorar el hecho de que el ajuste varía en función del número de asentamientos que se consideran. En este sentido, el hecho de considerar las entidades de más de 50 habitantes lleva a un ajuste siempre peor que cuando se limita el sistema a las entidades de más de 500 habitantes. Para el conjunto del periodo, el umbral de 500 habitantes aparece como el punto de ruptura evidente, ya que a partir de este límite las diferencias de ajuste no son muy significativas. De todas formas, para obtener una pendiente negativa más cercana a 1, se hubiese tenido que limitar aún más las entidades, utilizando un valor del umbral más elevado.

En referencia a la tendencia y cronología de las variaciones en esta distribución, se observan resultados parecidos para distintos umbrales. Se trata de un periodo en él que las desigualdades aumentan por encima de lo que tendrían que haber aumentado, si el único factor de variación hubiese sido el aleatorio. Las principales transformaciones se produjeron entre 1950 y 1981, y el aumento de las desigualdades acompañó el aumento del conjunto del sistema urbano.

El sobredimensionamiento de Barcelona respecto del resto de entidades, y la incorporación de asentamientos más pequeños son las principales razones que explican porque no se observa una distribución *rank-size* perfecta. En el caso de Barcelona existe una situación de dominio evidente, que de hecho sería más acentuado si se considerase la continuidad urbana con las ciudades vecinas. Claramente los factores que determinan el tamaño de Barcelona son de otra clase que los exclusivamente aleatorios. Otra explicación posible podría ser que el ámbito de influencia de Barcelona se extiende más allá de las cuatro provincias catalanas, y por lo tanto estaríamos hablando de un sistema de poblamiento incompleto.

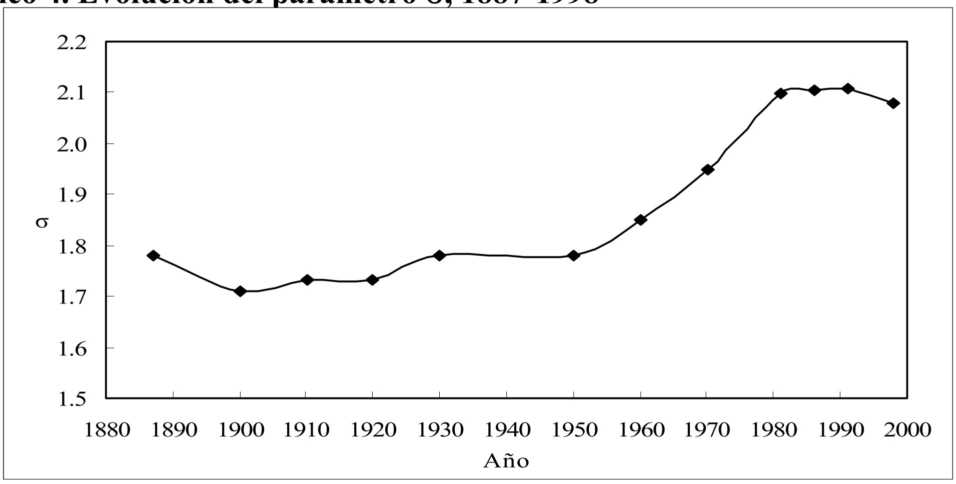
Por lo que respeta a las entidades más pequeñas, se ha visto que existe un claro punto de inflexión en torno a los 500 habitantes. El método *rank-size* requiere limitar las entidades a un determinado umbral para poder observar la relación lineal predecida y poder contrastar la hipótesis de la distribución aleatorio de crecimiento. A pesar de ésto, y con independencia del umbral utilizado, los parámetros obtenidos ofrecen una buena descripción de la intensidad y cronología de los cambios en las estructura de la población por entidad de Cataluña.

2.4.2.- La aproximación lognormal

En la distribución lognormal, la probabilidad de crecimiento se distribuye aleatoriamente entre las entidades. En cifras absolutas, las entidades más pobladas tienden a crecer más, pese a tener probabilidades menores de crecimiento. La distribución lognormal asume con el tiempo una mayor dispersión en el tamaño de las entidades. A diferencia del método *rank size*, la distribución lognormal permite contrastar para el conjunto de las entidades la hipótesis de la existencia de procesos aleatorios en el crecimiento poblacional de las mismas.

El Gráfico 4 muestra la evolución del parámetro σ (la desviación estándar de la normal asociada con la lognormal), calculado para las entidades de población en Cataluña según una distribución lognormal. Se trata de un indicador de concentración, cuya evolución muestra que la población en Cataluña se ha concentrado durante el periodo analizado. Este resultado es coherente con la evolución esperada de una distribución lognormal. El crecimiento más destacado de σ se observa entre 1950 y 1981, manteniéndose estables antes y después de este intervalo. El valor de σ para el año 1887 es superior al de 1900. Este hecho contradice la tendencia dominante y probablemente se explica por la disminución del número de entidades entre estos dos años a raíz de un cambio metodológico en la captación de las mismas. Así mismo, en 1998 se observa una ligera disminución de las desigualdades que, en la lógica anterior, también podría atribuirse a un número menor de entidades, aunque esta vez no es atribuible a ningún cambio metodológico, sino a la propia dinámica del poblamiento.

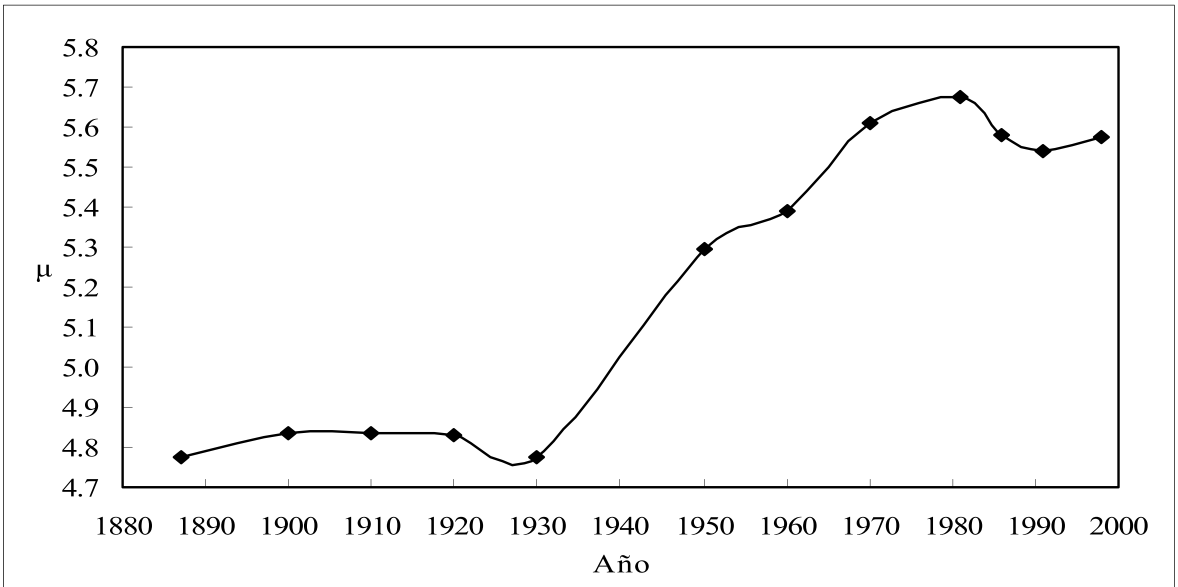
Gráfico 4. Evolución del parámetro σ, 1887-1998



Fuente: Elaboración propia a partir de los datos del Nomenclátor.

En la distribución lognormal, el parámetro μ se interpreta como un indicador de crecimiento del sistema de entidades en su conjunto (Gráfico 5). El crecimiento de μ se concentra básicamente entre 1930 y 1981. A pesar del aumento de las desigualdades en la distribución de la población, μ muestra que el crecimiento de la población afecta a escala a todos los agregados de población del sistema urbano, lo que no significa que crezcan en proporción a su tamaño, sino que lo hacen según una distribución lognormal de las probabilidades de crecimiento.

Gráfico 5. Evolución del parámetro μ, 1887–1998



Fuente: Elaboración propia a partir de los datos del Nomenclátor.

Conocidos los parámetros y mediante el procedimiento indicado anteriormente, se pueden calcular las frecuencias esperadas según una distribución lognormal y compararlas con las observadas. A guisa de ejemplo, el Cuadro 3 muestra los resultados de este ejercicio para el año 1960 y el Gráfico 6 compara gráficamente las frecuencias observadas y esperadas para este mismo año. Las diferencias no son visualmente importantes, aunque la escala del gráfico esconde diferencias relativas mayores en los intervalos de menor frecuencia, que agrupan a las entidades más pobladas.

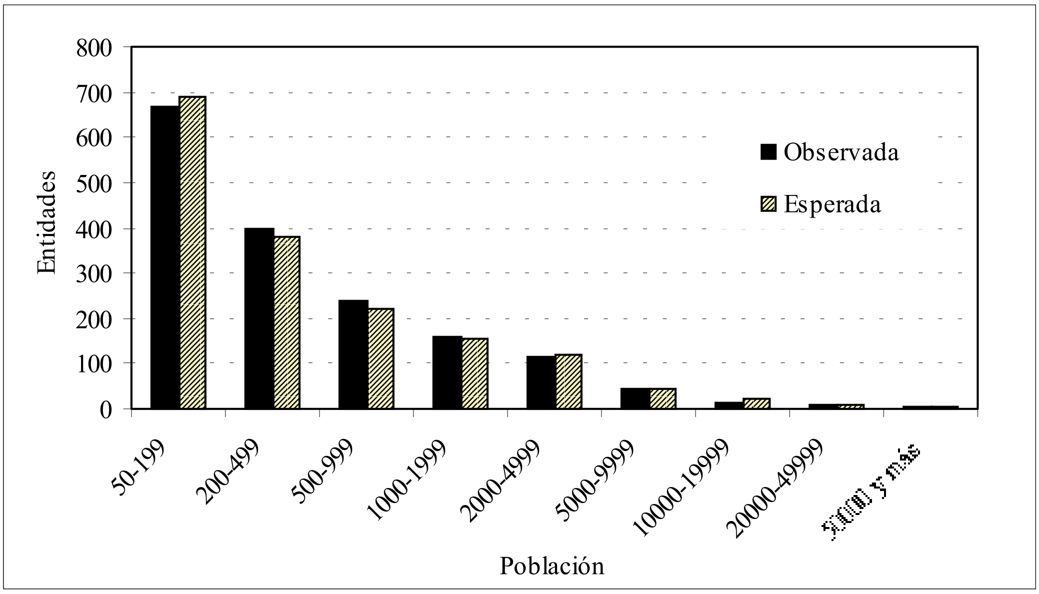
Más allá de la demostración visual, el estadístico chi cuadrado (X^2) y el Kolmogorov-Smirnov (Daniel 1978, p. 288) son utilizados para contrastar estadísticamente la hipótesis de la distribución lognormal de las entidades de población y, consecuentemente, la existencia de procesos aleatorios en el crecimiento de la población por entidades.

Cuadro 3. Frecuencia observada y esperada, 1960 (basado en $\sigma = 1.85$ y $\mu = 5.39$)

Intervalo	Límite inferior (k)	Ln (k)	Frecuencia observada acumulada	Frecuencia observada	Frecuencia esperada
1	49	3.89	1643	669	690
2	200	5.30	974	396	380
3	500	6.21	578	237	221
4	1000	6.91	341	157	156
5	2000	7.60	184	116	120
6	5000	8.52	68	42	43
7	10000	9.21	26	13	20
8	20000	9.90	13	8	9
9	50000	10.82	5	5	3

Fuente: Elaboración propia a partir de los datos del Nomenclátor.

Grafico 6. Frecuencias observadas y esperadas por intervalos de población, 1960



Fuente: Elaboración propia a partir de los datos del Nomenclátor.

En el caso de X2, los resultados indican que sólo 4 de las 12 distribuciones no son de tipo lognormal a un nivel de significación de 0.025 (Cuadro 4). Tres de ellas se encuentran entre 1887 y 1920, años de relativa estabilidad según se desprende del análisis de los parámetros μ y σ, y la cuarta en 1998. Sin embargo, los resultados del test Kolmogorov-Smirnov muestra que todas las distribuciones se ajustan a una distribución de tipo lognormal.

Cuadro 4. Resultados según X^2, 1887-1998

Nomenclátor	X^2	prob.	Ns < 0.025	Ns < 0.05
1887	19.33	0.0132	NO	NO
1900	15.83	0.0705	SI	SI
1910	20.68	0.0141	NO	NO
1920	19.36	0.0223	NO	NO
1930	11.26	0.2581	SI	SI
1950	18.18	0.0331	SI	NO
1960	8.29	0.5049	SI	SI
1970	11.49	0.2436	SI	SI
1981	10.56	0.3073	SI	SI
1986	8.67	0.3707	SI	SI
1991	10.72	0.2182	SI	SI
1998	20.14	0.0098	NO	NO

Fuente: Elaboración propia.

Cuadro 5. Resultados según Kolmogorov-Smirnov, 1887-1998

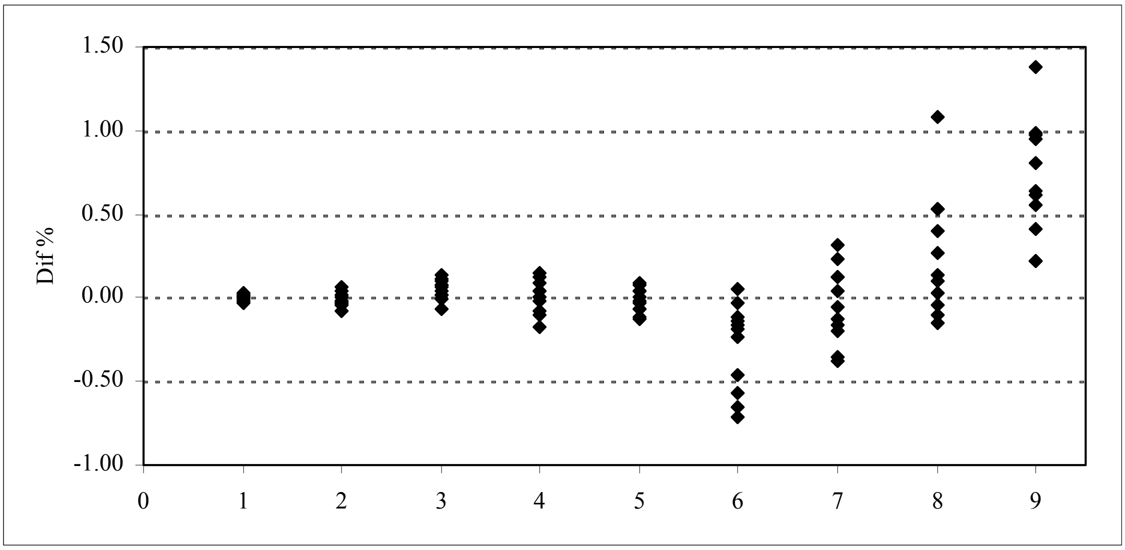
Nomenclátor	ks	Ns < 0.025	Ns < 0.05
1887	0.0094	SI	SI
1900	0.0050	SI	SI
1910	0.0067	SI	SI
1920	0.0182	SI	SI
1930	0.0074	SI	SI
1950	0.0109	SI	SI
1960	0.0128	SI	SI
1970	0.0131	SI	SI
1981	0.0156	SI	SI
1986	0.0160	SI	SI
1991	0.0180	SI	SI
1998	0.0220	SI	SI

Fuente: Elaboración propia.

X^2 es más sensible a las diferencias relativas importantes, que normalmente se dan en intervalos con una frecuencia de casos baja. Un error de predicción de una sola entidad en una categoría con tres entidades observadas tiene un efecto mayo que en una desviación de 50 entidades en una categoría con 1.000 observaciones. Por esta razón, no es aconsejable utilizar el X^2 en intervalos con menos de 5 observaciones. En cambio, el Kolmogorov-Smirnov trabaja con las frecuencias acumuladas, razón por la cual los errores de predicción en los últimos intervalos, que son los que tienen las frecuencias más bajas, no afectan en demasía los resultados finales.

El análisis de los residuos, las diferencias entre frecuencias observadas y acumuladas, informa de la desviación de cada categoría respecto a la distribución lognormal (Gráfico 8). Los valores positivos indican que la frecuencia observada es mayor que la esperada y los negativos al revés. Las diferencias relativas son mayores a medida que aumenta la población o disminuye el total d'entidades por categoría. En términos absolutos, las principales diferencias se dan en los grupos con más entidades. Para la categoría de más de 50.000 habitantes, el número de entidades observadas es mayor que el de las esperadas. Así pues, las variables independientes que condicionan la distribución de la población por entidades no han actuado del mismo modo. Este hecho ha generado una distribución de la población por entidades que ha beneficiado a las más pobladas en una proporción mayor de la que deberíamos esperar de una distribución de las probabilidades de crecimiento aleatorias.

Gráfico 8. Diferencia por intervalos entre frecuencias observadas y esperadas



Fuente: Elaboración propia a partir de los datos del Nomenclátor.

En cambio, las entidades entre 5.000 y 10.000 habitantes son menos de las esperadas. En el resto de categorías, las diferencias observadas no definen ningún patrón, aunque si que la dispersión aumenta conforme las categorías tienen una frecuencia menor.

3.- Rank-size versus lognormal

El estudio de la ciudades y sistema de ciudades se puede aproximar desde distintas perspectivas. En este trabajo se ha hecho a partir de la comparación con las distribuciones *rank-size* y lognormal, concebidas como la manifestación de un mismo proceso: la distribución aleatoria de las tasas de crecimiento entre todas las poblaciones. La distribución lognormal puede generar una distribución *rank-size* si nos limitamos a las unidades a partir de un mínimo de población. Hemos constatado de todas formas que cualquier sea el umbral elegido, prácticamente en ningún caso obtenemos el ajuste de una *rank-size* con pendiente igual a –1, es decir la ley de Zipf, que interpretaríamos aquí como una distribución que se explica solamente por el juego de factores aleatorios en el crecimiento. La única vez que se ha conseguido el ajuste por la ley de Zipf ha sido para la distribución de 1960, tomando las entidades mayores de 500 habitantes. A nivel de la interpretación, el modelo de crecimiento aleatorio utilizado en este trabajo nos lleva a descartar la idea de convergencia hacia una distribución óptima o equilibrada, interpretación clásica de la ley de Zipf. Al contrario, la idea es comparar el valor de la pendiente observada con el valor –1. Cuanta más pequeña sea la diferencia con respecto de –1, mayor será el juego de los factores aleatorios para la explicación del crecimiento de los asentamientos. En cambio, las diferencias con respecto de –1, tanto a nivel de la curva global como a nivel local, para determinadas ciudades o intervalos de ciudades, se pueden interpretar como la presencia de factores sistemáticos o no aleatorios.

Pero si la interpretación correcta de la ley de Zipf es la de ser una representación gráfica de la parte alta de una distribución lognormal truncada a un umbral pequeño, entonces también surge la pregunta de porqué limitarse a una sola parte de la distribución y porqué no descartar la representación *rank-size* en beneficio del ajuste por la lognormal. En efecto, esta distribución ofrece una mejor aproximación al sistema de entidades, sin necesidad de establecer ningún límite o umbral mínimo.

La conclusión principal de este trabajo es que la existencia de procesos aleatorios en el crecimiento de la población en el territorio de Cataluña entre 1887 y 1998 ha sido contrastada por esta aplicación. A la luz de los resultados, la presencia de estos procesos no se puede negar, pero es preciso notar el sobredesarrollo de la entidad de Barcelona en el contexto del ajuste *rank-size*, así como en general de las entidades de más de 50.000 habitantes en el contexto del ajuste lognormal, lo que demuestra que la población de Cataluña se ha concentrado durante este periodo por encima de lo que sería de esperar si solamente hubiesen entrado en juego los factores aleatorios del crecimiento.

Podemos acabar este trabajo con una paradoja, el hecho de que los trabajos clásicos sobre la dinámica del poblamiento partan de la hipótesis de que existe una relación necesaria entre el tamaño de la población de distintas ciudades. Para la teoría de los lugares centrales, la jerarquía requiere ciudades más grandes en relación con un grupo de ciudades de tamaño parecido distribuidas en el espacio según unas reglas de simetría. La interpretación clásica de la ley de Zipf reencuentra esta idea de jerarquía, pero esta vez a favor de una regla parecida a la que podría existir en un sistema físico con fuerzas de atracción y de repulsión: las ciudades pequeñas tienen que tener una población en una relación sistemática con la población de las ciudades grandes. Si este no es el caso, el sistema no está equilibrado. La interpretación probabilítisca del crecimiento rompe con esta idea de la existencia de una relación necesaria entre la población de las ciudades: las tasas de crecimiento no dependen ni del tamaño de las ciudades, ni de sus vecinas. Se pierde la idea de un sistema de relaciones fijas entre ciudades, y sí hay efectos de compensación, que se manifiestan por el hecho de que a una ciudad con mayor crecimiento que la media le corresponde un número determinado de ciudades con un crecimiento menor que la media, y ésto se produce por una distribución al azar de la distancias geográficas entre todas estas ciudades. La interpretación aleatoria proporciona curiosamente una visión mucho más dinámica y cambiante, más cerca de la realidad, que la visión rígida de la teoría de los lugares centrales o de la interpretación clásica de la ley de Zipf.

Bibliografía

AITCHISON, J. Y BROWN, J.A.C. (1966), *The lognormal distribution,* Cambridge, University Press.

BERRY, J.L. (1964), "Cities as systems within systems of cities" a Friedman, J. i Alonso, W. (eds.), *Regional Development and Planning: a Reader*, Cambridge, MIT Press ,pag. 138-151.

CAMPS, E. (1990), "Urbanización y migraciones internas durante la transición al sistema fabril: el caso catalán", *Boletín de la Asociación de Demografía Histórica, VIII*, 2, pag. 73 - 95.

CAPEL, H. (1972), "La validez del modelo rank-size", *Revista de Geografía,* VI, 1, pag. 121 - 138.

CHRISTALLER, W. (1933), *Die zentrallen Orte in Suddeutschland,* Jena, G.Fisher.

COFFEY, W.J. (1981), *Geography: Towards a General Spatial Systems Approach,* London, Methuen anc Co. Ltd.

DE COLA, L. (1985), "Lognormal estimates of macroregional city-size distributions, 1950-1970", *Environment and Planning, A,*17, pag. 1637-1652.

DENDRINOS, D.S. (1992), *The dynamics of cities: ecological determinism, dualism and chaos,* New York, Routledge.

ESTEVE, A. (2003) *El Nomenclàtor com a font per a l'estudi territorial de la població a Catalunya. Aplicacions*, 1857-1998.Tesi Doctoral. Departament de Geografia, Universitat Autònoma de Barcelona.

FONSECA, J.W. (1989), *Urban size hierarchy: a mathematical intepretation,* Zannesville, Institute of Mathematical Geography.

FRANKHAUSER, P. (1997), "L'approche fractale", *Population,* 4, pag. 1005-1040.

GABAIX, X., (1999). "Zipf's law for cities: an explanation", *Quarterly Journal of Economics*, 114, pag. 739-767.

GUÉRIN, F. (1995), "Rank-size distribution and the process of urban growth", *Urban Studies,*32, 3, pàg. 551-562.

LE BRAS, H. (1994), *Les limites de la planète,* París, Champs Flammarion.

LE BRAS, H. (1996), *Le peuplement de l'Europe,* París, Datar.

MANDELBROT, B.B. (1984), *The fractal geometry of nature,* San Francisco, W.H. Freeman & Co.

MENDIZÀBAL, E. (1994), "El fenómeno urbano en Cataluña, 1717- 1991", *Papers de Demografia*, 83.

MIRALLES, F., ROSÉS, J. Y ARMET, J. (1982), "El sistema urbano catalán" a *Reconeixement Territorial de Catalunya*, Vol. 21 Barcelona, Departament de Política Territorial i Obres Públiques. Generalitat de Catalunya ,pag. 209 - 412.

OLIVERAS, J. (1994), *La formació dels desequilibris territorials a Catalunya en el segle XIX,* Barcelona, Direcció General de Planificació i Acció Territorial de la Generalitat de Catalunya.

RIERA, P. (1988), *Les àrees funcionals de Catalunya*, Tesi Doctoral, Departament de Geografia de la Universitat Autònoma de Barcelona.

PARETO, V (1896), *Cours d'Économie Politique*, Geneva, Droz.

PUMAIN, D. (1982), *La dynamique des villes,* Paris, Economica.

SHUPER, V. (1999), "La théorie des lieux centraux et les phénomènes d'évolution", *Cybergeo* 87, pàg. 1-17.

UNITED NATIONS y EUROSTAT (1998), *Population and Housing Censuses,* vol. Statistical Standards and Studies 49 Ginebra, United Nations Economic Commission for Europe.

VIDAL BENDITO, T. (1980), *Atlas sòcio-econòmic de Catalunya,* vol. II Barcelona, Caixa d'Estalvis de Catalunya, Ahorrobank i Servei d'Estudis a Barcelona del Banco Urquijo.

WOLDENBERG, M.J. (1971), *Allometric growth in social systems. Harvard Papers in Theroretical Geography.,* Cambridge, Harvard University Graduate School of Design.

ZIPZ, G.K. (1949), *Human behavior and the principle of least effort,* Cambridge, Addison-Wesley Press.