

# Las publicaciones sobre el péndulo: un análisis comparativo

Cristina Restrepo Arango \*

Rubén Urbizagástegui Alvarado \*\*

*Artículo recibido:  
18 de enero de 2012.*

*Artículo aceptado:  
30 de agosto de 2012.*

## RESUMEN

Este trabajo re-evalúa la productividad de los autores sobre la literatura del péndulo publicada desde 1629 a 1885. Replica los datos del trabajo de Moreno-Cabo & Solaz-Portolés (2008). Se aplican los modelos del cuadrado inverso, poder inverso generalizado, Poisson compuesto y Poisson lognormal por el método de la máxima probabilidad. Los resultados obtenidos muestran que el modelo Poisson lognormal y el modelo del poder inverso generalizado son los que estiman los autores observados más coherentemente. El test estadístico Kolgomorov-Smirnov confirma el ajuste de los modelos del poder inverso generalizado, Poisson lognormal y Poisson compuesto y rechaza el ajuste del modelo del cuadrado inverso.

\* El Colegio de México, México. lrestrepo@colmex.mx

\*\* Universidad de California en Riverside, USA. ruben@ucr.edu

**Palabras Clave:** Péndulo; ley de Lotka; Poisson log-normal; Poder inverso generalizado; Poisson compuesto; Bibliometría; Cienciometría; Infometría.

## ABSTRACT

### **Publications on the Pendulum: a comparative analysis**

*Cristina Restrepo Arango and Rubén Urbizagástegui Alvarado*

This study examines the production of the literature on the pendulum published from 1629 to 1885. The findings of Moreno-Cabo & Solaz-Portolés (2008) are replicated. The data are analyzed through the following statistical methods: inverse square, generalized inverse power, and compound Poisson and Poisson log-normal by maximum likelihood models. The results show that the Poisson lognormal and generalized inverse power are those that estimate the observed authors most consistently. The statistical test Kolgomorov-Smirnov confirms the fit of the generalized inverse power, Poisson lognormal, and compound Poisson, while rejecting the inverse square model.

**Keywords:** Pendulum; Lotka's law; Poisson lognormal; Generalized Inverse Power, Compound Poisson; Bibliometrics; Cienciometrics; Infometrics.

## INTRODUCCIÓN

Por la importancia que tiene el péndulo en el desarrollo de la ciencia, en España se realizó

un estudio bibliométrico del conjunto de publicaciones proporcionadas por Wolf (1889a) en su *Bibliographie du pendule*. Ese estudio pretendía mostrar un caso particular [...] del modo de crecimiento de la ciencia en el período comprendido entre 1629 y 1885, así como ciertos aspectos de la forma en que sus actores trabajaban: productividad de los autores, colaboración en los trabajos y lenguas que utilizaban” (Moreno-Cabo & Solaz-Portolés, 2008a, p. 640).

Además de la bibliografía mencionada, los autores incorporaron al estudio los datos bibliográficos contenidos en el Tomo IV de la *Collection de Mémoires relatifs a la Physique*, publicadas por La Société Française de Physique. Esa bibliografía

recoge artículos, libros, informes de instituciones científicas y disertaciones, y proporciona el nombre de los autores, el título del trabajo y un breve resumen del mismo (Moreno-Cabo & Solaz-Portolés, 2008a, p. 641).

Con esos datos los autores analizaron la ley de Lotka a través del modelo del cuadrado inverso.

En los resultados de la investigación mencionada anteriormente los autores afirman que:

el análisis de regresión lineal entre el logaritmo del número de autores y el logaritmo del número de publicaciones, nos proporciona un coeficiente de correlación lineal de  $r = 0,998$  y una pendiente de la recta de  $-11,43$ . Por lo tanto, la curva se ajusta de manera aproximada a la ecuación  $y = k/x^{11}$  (Moreno-Cabo & Solaz-Portolés, 2008a, pp. 642-643).

Según esos datos el modelo del cuadrado inverso de la ley de Lotka tendría el valor del exponente  $n = 11,43$ , hecho no reportado hasta la fecha por ningún trabajo publicado que sea del conocimiento de los autores del presente artículo. Moreno-Cabo & Solaz-Portolés (2008a), concluyen su artículo afirmando que los resultados no se ajustan a la ley del cuadrado inverso, sino a la ecuación  $y = k/x$ .<sup>11</sup> A ésta misma conclusión llegan en otro trabajo sobre el mismo asunto (Solaz-Portolés, Moreno-Cabo y Sanjosé López, 2008b).

Intrigados por esos resultados decidimos replicar este estudio. Nos proponemos re-evaluar la productividad de los autores sobre el péndulo con los mismos datos que nos fueron amablemente proporcionados por uno de los autores vía correo electrónico. Inicialmente pensamos también replicar los datos del trabajo de Parreiras; Silva; Matheus y Brandão (2006), sobre las redes de colaboración y producción científica en ciencias de la información en el Brasil, donde informan haber encontrado un  $\beta = 24,017$ ; es decir,  $n = 24,017$ , para la aplicación de la ley de Lotka por el modelo del cuadrado inverso. Según estos autores, los datos fueron probados con el test Kolgomorov-Smirnov e indican un resultado estadísticamente significativo (Parreiras; Silva; Matheus; Brandão, 2006, p. 314). Tenemos fuertes dudas sobre la bondad del ajuste de esos resultados. Sin embargo debido a que los datos observados no aparecen en el artículo publicado fue necesario solicitar esta

información. Lamentablemente, a pesar de nuestra insistencia, los autores no nos proporcionaron los datos usados en su estudio. Para despejar dudas sobre los resultados de esa investigación sería conveniente replicar el análisis de los datos. Por lo tanto este estudio se concentra solamente en los datos de Moreno-Cabo & Solaz-Portolés (2008a).

Ya es conocido que cuando se usa una prueba estadística, sea el chi-cuadrado o Kolmogorov-Smirnov, el modelo del cuadrado inverso de Lotka (1926) no produce un ajuste adecuado de la distribución de la productividad de los autores. Para superar esa deficiencia se han sugerido otros modelos más adecuados como el modelo del poder inverso generalizado, el modelo Gauss Poisson inverso generalizado, el modelo binomial negativo, el Poisson lognormal, etc. Esos modelos están en exploración pero han sido poco estudiados y/o experimentados en América Latina. Por esa razón el objetivo de este trabajo es replicar la aplicación del modelo del cuadrado inverso de Lotka con los datos de Moreno-Cabo & Solaz-Portolés (2008a). Paralelamente experimentaremos otros modelos de ajuste que producen mejores resultados, como el modelo Poisson compuesto, el modelo Poisson lognormal y el modelo del poder inverso generalizado, analizados a través del método de la máxima probabilidad. Esto nos permitirá comparar la bondad del ajuste de cuatro modelos distintos que actualmente son explorados y evaluados en este sub-campo de la Bibliometría.

Para alcanzar los objetivos propuestos este trabajo está organizado de la siguiente manera: en la primera parte se hace una somera introducción al asunto y se plantea el problema de investigación; luego se hace una revisión exhaustiva de la literatura sobre la aplicación de los modelos en estudio y los resultados obtenidos en diferentes campos del conocimiento; se intenta entonces demostrar que las investigaciones donde se aplica el modelo del cuadrado inverso han resultado en experiencias fallidas o dudosas. Por el contrario, cuando se usan los otros modelos aquí experimentados, los resultados generalmente resultan exitosos. Posteriormente se detalla la metodología con énfasis en su formulación matemática y en la forma de medición de los datos. A continuación se describen los resultados obtenidos y se ofrecen las conclusiones. Finalmente se presenta la bibliografía de la literatura revisada en la redacción de esta investigación.

## REVISIÓN DE LITERATURA

El primer trabajo que examina la productividad científica de los autores fue publicado por Dresden (1922). Este autor analizó los documentos presentados

en la Reunión de Chicago de la Asociación Americana de Matemáticas desde 1896 a 1922 por 278 autores. Encontró 1,102 documentos, de los cuales sólo analizó 649 documentos pertenecientes a aritmética, álgebra, geometría, mecánica, historia, pedagogía y filosofía de las matemáticas. En estos 649 documentos encontró que 113 autores sólo presentaron un documento, mientras que 329 documentos fueron producidos por 10 autores, al igual que un autor tiene alrededor de un total de 70 documentos; es decir, un 6% del total (Dresden, 1922). Posteriormente Alfred Lotka (1926) propuso la ley de la productividad científica de los autores. Contando los autores que aparecían en el índice del *Chemical Abstracts* en el periodo de 1907 a 1916, sólo tuvo en cuenta los nombres de las letras A y B. Pero también analizó los autores que publicaron trabajos sobre física en el *Auerbach's Geschichtstafeln der Physik* de 1900 a 1910. Los datos que obtuvo fueron similares en ambas fuentes. A partir de estos datos concluyó que la proporción de todos los autores que contribuyen con un único trabajo podrían estar sobre el 60%, mientras que los que hacen dos contribuciones son cerca de 1/4 de aquellos que hacen una, el número de autores que hacen tres es cerca de 1/9, etcétera (Lotka, 1926). A partir de estos datos formuló la ley del cuadrado inverso. Es a este modelo que se denomina como la “ley de Lotka”, en homenaje a su formulador.

Seis años después Carr (1932), extrajo una muestra aleatoria de 1,000 nombres del índice alfabético de 1916 de la Oficina de Patentes de EE. UU. y recolectó datos sobre el número de las patentes registradas por cada individuo durante diez años hasta 1925, e hizo lo mismo para los registros de 1930. La muestra se componía de 904 inventores de los Estados Unidos y 96 de países extranjeros, principalmente de la Comunidad Británica, Alemania, Francia, Suecia y Suiza, todos ellos residentes en los Estados Unidos. En diez años los 1.000 titulares de patentes habían producido 3.457 patentes, con un promedio de 1.144 patentes por persona. Menos de la séptima parte de los inventores habían producido el 57,5% de las patentes, mientras que el 53,2% sólo produjeron una patente. El autor afirma que la producción de patentes *parece* seguir la distribución de Lotka. Posteriormente Dufrenoy (1938), examinó el comportamiento de publicación de los biólogos en dos conjuntos de datos: documentos sobre la patología de plantas en el *Review of Applied Mycology* de 1935, y el tomo 120 del *Comptus Rendus de la Societe de Biologie* de 1935. En los datos del primer grupo 1.527 autores contribuyeron con 2.231 artículos; en el segundo conjunto de datos 500 autores contribuyeron con 677 artículos. El autor sugiere que este dato refleja la ley de Lotka, pero sólo para los pequeños valores de las  $n$  contribuciones.

Cuatro años después Hersh (1942), usando una exhaustiva bibliografía publicada en 1939 por H. J. Muller sobre la genética de la *Drosophila*, analizó el

número de artículos y autores publicados cada año con el propósito de trazar el curso de la productividad científica y sus relaciones cuantitativas. Cuando los datos del número de artículos publicados cada año desde 1910 hasta 1938 se trazaron semi-logarítmicamente contra el tiempo, encontró que los puntos caían dentro de una banda estrecha, con bordes rectos y paralelos; es decir, que los datos se ajustaban más o menos a la relación exponencial simple llamada ley del interés compuesto. Cuando los logaritmos del número de autores se trazaron en relación a los logaritmos del número de documentos acreditados a los autores, los primeros diez puntos que incluían alrededor del 90% de todos los autores estaban cerca de una línea recta de pendiente negativa. Es decir, parecían seguir la ley de Lotka. Williams (1944) discutió los datos de Dufrenoy (1938) y recogió un nuevo conjunto de datos para ser analizados. En sus datos incluyó los datos de Dufrenoy y los complementó con el conteo de la productividad de los biólogos del volumen 1 de 1913 y el tomo 24 de 1936 de la revista *Review of Applied Mycology*. Los datos de 1913 produjeron 411 autores con 656 contribuciones, y los datos de 1936 produjeron 2.379 autores con 1.534 contribuciones. Sin embargo, los resultados fueron similares a los de Dufrenoy (*parecían* seguir la ley de Lotka) con -0,40 y -0,31 como estimaciones de las pendientes para ambos conjuntos de datos respectivamente.

Este modelo del cuadrado inverso ha sido analizado por numerosos autores pero cuando a los datos observados se le aplica una prueba estadística, ya sea el chi-cuadrado o el Kolmogorov-Smirnov, el modelo no se ajusta a los datos esperados. Esta falta de ajuste se ha observado en la física (Vlachý, 1970; Inönü, 1971; Turkeli, 1973). Por ejemplo, González Blasco (1975), estudió la producción científica española en el periodo de 1965 a 1970 y encontró que de una muestra de 85 autores el 70% publicaron un único artículo y 16% dos artículos. Este autor concluyó que los datos sobre la producción científica española solamente *se aproximan* a ley del cuadrado inverso de Lotka. También en el campo de la psicología Carpintero; Peiró & Quintanilla (1977) analizaron 75 artículos producidos por 76 autores desde 1969 hasta 1974 en la revista *Anuario de Psicología* publicada por la Universidad Central de Barcelona. Encontraron que casi 1/3 de los artículos fueron producidos por el 10% de los autores, pero que los datos *no se ajustaban* a la ley del cuadrado inverso de Lotka. Tampoco en el campo de la sociología el modelo del cuadrado inverso parece ajustarse a la distribución de los autores, como es el caso del trabajo de Martin & Berry (1977), quienes informan que las tasas de productividad difieren significativamente de lo previsto por la ley de Lotka. Igualmente no se ajusta a los datos analizados en el campo de la ciencia de la computación, pues, Radhakrishnan y Kernizan (1979) informan que realizaron dos experimentos para verificar la satisfacción del modelo del cuadrado

inverso en esa área con resultados negativos. En el primer experimento utilizaron documentos publicados entre 1968-1972 en las revistas *Communications of the Association for Computing Machinery* (CACM) y *Journal of the Association for Computing Machinery* (JACM). Fueron encontrados 715 autores en CACM y 382 en JACM. En el primer experimento se supone que un autor publica exclusivamente a través de una revista y la ley de Lotka no se ajustó a esta distribución, pero la predicción del cubo inverso es más cercana. En el segundo experimento seleccionaron una muestra aleatoria de 300 autores utilizando el índice acumulativo de autores del *Computer and Control Abstracts*; en este caso la desviación de la ley de Lotka fue considerablemente alta.

Budd (1988) analizó 569 autores principales en el campo de la educación superior. Estos autores fueron examinados con el modelo del cuadrado inverso de Lotka, pero al probar el ajuste de los datos con el test estadístico de Kolgomorov-Smirnov el resultado indicó que estos datos no se ajustaron al modelo del cuadrado inverso de Lotka. Sen, Taid y Hassan (1996) estudiaron la productividad científica en la literatura de bibliotecología y ciencia de la información indizada en LISA en el periodo de 1992 a 1993. Hallaron 7.624 autores en el año de 1992 y 6.423 autores en el año de 1993. Aplicaron la ley del cuadrado inverso de Lotka para los datos obtenidos en ambos años y compararon los resultados con diferentes valores de  $n$ . Por ejemplo, con un valor de  $n = 2$  y con  $n = 3,23$  para 1992 y con  $n = 3,1$  para 1993, y concluyeron que los valores estimados con un valor diferente a  $n = 2$  se aproximan más a los valores observados, no así los valores estimados con  $n = 2$ . Díaz Mujica (2007) analizó los artículos originales de la revista *Anales venezolanos de nutrición* desde septiembre de 1990 hasta junio de 2003 aplicando la ley del cuadrado inverso de Lotka a 3,045 autores de 904 artículos originales publicados en esta revista, y encontró que 55,8% autores sólo han contribuido con un único trabajo en la revista. La autora no incluyó los datos estimados ni evidencias del ajuste estadístico de los datos observados.

Ahmed y Rahman (2008) estudiaron la literatura sobre nutrición de Bangladesh publicada de 1972 a 2006. Aplicaron la ley del cuadrado inverso de Lotka a 998 autores que produjeron 636 artículos con un promedio de 1,57 autores por artículo. Estos autores concluyeron que la ley del cuadrado inverso se ajusta a los datos sobre nutrición; sin embargo, no aplicaron ninguna prueba estadística de ajuste de los datos observados, sino que usaron el método del “ojímetro” para llegar a esa conclusión. Luor, Johanson, Lu y Wu (2008) analizaron la literatura indizadas en 29 bases de datos sobre aprendizaje asistido por computador en el periodo de 1996 a 2008. Ellos aplicaron el modelo del cuadrado inverso de Lotka a los autores de 536 artículos, probaron el ajuste de los datos con el test estadístico chi-cuadrado y concluyeron que

los datos estimados *no se ajustan* a este modelo, pero no presentaron los datos estimados ni observados.

Pao (1986) afirma que los datos estimados que se obtienen a partir de la ley del cuadrado inverso propuesto por Lotka no se justan a las pruebas estadísticas del Chi- cuadrado ni Kolmogrov-Smirnov. Por esa razón propone que el valor de  $c$  y  $n$  no sean constantes, sino que varíen en función de los datos que se observan. El valor de  $c$  y  $n$  se hallan a partir de una regresión lineal, ya sea a través del método de la máxima probabilidad o el método de los mínimos cuadrados. Esta autora propone el modelo del poder inverso generalizado por el método de los mínimos cuadrados. Gupta (1989) analizó la aplicabilidad de este modelo del poder inverso generalizado a la literatura biomédica nigeriana en el periodo de 1970 a 1984. Estudió 500 documentos y dividió los datos de los autores en cuatro grupos. Estos cuatro grupos fueron: la comunidad completa de autores, los autores que aparecen solamente como primeros; los autores individuales; y los autores que aparecen solamente como coautores y nunca como primeros autores. Encontró que los datos observados y estimados de la comunidad completa de autores se ajustan al modelo del poder inverso generalizado, según el test de Kolgomorov-Smirnov (K-S) a un nivel de significancia de 0,01. Urbizagástegui y Cortés (2002) analizaron los artículos recolectados de la *Revista Geológica de Chile* desde 1974 a 1997, abarcando un período de 24 años. Para la recolección de los datos adoptaron el criterio de conteo directo y por lo tanto sólo los autores principales fueron acreditados con plena contribución a la producción de un artículo. En el periodo objeto de investigación fueron identificados 178 autores con un total de 320 artículos publicados. El 72% de ellos produjo sólo un artículo. Usando el modelo del poder inverso generalizado estimaron los parámetros  $c$  y  $n$  por el método de los mínimos cuadrados. La prueba Kolmogorov-Smirnov fue usada para ajustar la distribución observada a los valores esperados. A un 0,01 nivel de significación el valor crítico encontrado fue 0,1222 y la desviación máxima fue igual a 0,1063. Concluyeron que esta literatura se ajusta a la ley de Lotka. Bonnevie (2003) estudió la visibilidad, la demografía de las autorías, la internacionalización y comportamiento de las auto-citas y el impacto de la revista *Journal of Information Science* en el periodo de 1979 a 2001. En relación a los investigadores que publicaron en la revista analizó 1.326 autores a través del modelo del poder inverso generalizado; con el valor estimado de  $n = 2,618$  y  $c = 0,977$ , encontró que los datos se ajustan a este modelo.

Patra y Chand (2006), examinaron la productividad en bibliotecología y ciencia de la información de la India recuperada en la base de datos LISA en el periodo de 1967 a 2004. Encontraron 2.732 autores que publicaron 3.396



trabajos, cerca de 1,24 artículos por autor. De estos autores sólo 37 han publicado más de 10 artículos. Estos datos son analizados con el modelo del poder inverso generalizado de la ley de Lotka y probados con el test de Kolmogorov-Smirnov ajustándose a los datos observados a un nivel de significancia de 0,01. Sevukan y Sharma (2008) aplicaron la ley del cuadrado inverso generalizado de Lotka a través del método de los mínimos cuadrados a 613 autores que publicaron 329 documentos sobre biotecnología en la India. Encontraron que a través de la aplicación del test K-S los valores observados se ajustan al modelo de Lotka.

La distribución Poisson lognormal (PL) ha sido estudiada por Steward (1994), que la describe como una herramienta que proporciona un ajuste aceptable a una variedad de distribuciones bibliométricas y cienciométricas. Aplica este modelo a dos conjuntos de datos: el primer grupo compuesto de químicos que recibieron su doctorado entre 1955-1961. En este conjunto fueron contados los coautores y también fueron incluidos los autores con productividad de cero documentos. El segundo grupo, compuesto de entomólogos de Nigeria, encontró que el modelo PL se ajusta muy bien a ambos conjuntos de datos y confirma la ley de Lotka con una producción estratificada en ambas áreas estudiadas. Kuperman (2006) examinó los patrones de productividad en las listas de correo del Internet, también conocido como listas de discusión o grupos de discusión. Recopiló los datos de los archivos electrónicos de dos listas de correo: *The LINGUIST* and *The History of the English Language*. En ambos grupos de datos el modelo PL mostró excelentes resultados y los datos se ajustaron a la ley de Lotka. También Urbizagástegui (2007a, 2007b), analizó la productividad de los autores sobre plantas medicinales del Perú y en plantas usadas como colorantes naturales. En ambos casos la distribución PL a través de las pruebas estadísticas del chi-cuadrado y K-S al 0,01 nivel de significancia confirmaron el ajuste de la ley de Lotka a los datos observados. Este mismo autor Urbizagástegui (2008), utilizando la distribución Poisson lognormal y Gauss Poisson inversa generalizada analizó la literatura estratificada producida sobre la ley de Lotka desde 1922 hasta 2003. Para este caso utilizó como formas de recuento de la literatura el conteo directo, el conteo completo y el conteo fraccionado. Los datos fueron evaluados con la prueba estadística del chi-cuadrado al 0,05 nivel de significación. Encontró que ambos modelos se ajustan muy bien a la distribución de la literatura producida, pero el modelo Gauss Poisson inverso generalizado produjo un chi-cuadrado menor y predijo mejor el número total de autores que la distribución Poisson lognormal.

Como se puede concluir por la literatura revisada, el modelo del cuadrado inverso no produce un ajuste estadístico adecuado de los datos observados

sobre la productividad de los autores. Sin embargo los otros modelos (poder inverso generalizado, Poisson lognormal y Poisson compuesta) parecen ajustarse mejor a los datos sobre la productividad de los autores. Contrariamente a las afirmaciones de uno de los revisores de este artículo que sustenta que “el tema de la productividad de los autores ha sido ampliamente tratado en la literatura científica” (Dictamen no. 2, p. 2), reafirmamos que los modelos aquí estudiados, y en especial los modelos Poisson lognormal y Poisson compuesto, no han sido suficientemente explorados ni en América Latina ni en otras latitudes.

## METODOLOGÍA

Como en este trabajo se van a evaluar y comparar cuatro modelos diferentes, cada modelo va a ser descrito por separado, especificando las ecuaciones y los parámetros necesarios para calcular los valores esperados de la distribución.

Con relación al modelo del cuadrado inverso considerando el simple volumen de la producción científica, Alfred Lotka (1926) realizó el conteo del número de veces que un nombre fue referenciado en dos fuentes: el *Chemical Abstracts*, 1907-1916 (Letras A y B) y el *Auerbach's Geschichtsatafeln*, 1900-1910 y estableció que la fórmula general para la relación existente entre la frecuencia de y personas haciendo x contribuciones, sería:

$$X^n Y = C$$

donde

X es el número de trabajos publicados

n es el exponente del número de trabajos publicados por los autores

Y es el número de autores que publicaron x trabajos

C es una constante

Esta propuesta es llamada del *cuadrado inverso* porque en la ecuación anterior el valor de n es igual a 2 siempre. La solución de esa ecuación es simple y el cálculo de los valores esperados se hizo usando el software estadístico SPSS versión 17,0 para Windows.

Como en la mayoría de los casos el modelo del cuadrado inverso no se ajusta a las pruebas del chi-cuadrado o a la prueba Kolmogorov-Smirnov, ahora es conocida como la ley de Lotka una forma más general llamada *poder inverso generalizado* que se expresa como:

$$y_x = k x^{-b}, \quad x = 1, 2, \dots, x_{\max}$$

donde

$y_x$  es la probabilidad de que un autor haga  $x$  contribuciones sobre un asunto

$k$  y  $b$  son los dos parámetros que deben ser estimados de los datos observados.

En este modelo del poder inverso generalizado, el valor del parámetro  $k$  y  $b$  puede ser estimado por dos técnicas: los mínimos cuadrados y la máxima probabilidad. En este trabajo para estimar el parámetro  $k$  se usó la función Zeta inversa de Riemann, para el que Pao (1985) proporciona una fórmula de aproximación exacta. La estimación del parámetro  $b$  se hizo a través del método de la máxima probabilidad y usando los valores de la tabla proporcionada por Johnson & Kotz (1969). Esta Tabla ha sido recalculada y reproducida tanto por Rousseau (1993) como por Nicholls (1987). Los valores esperados fueron estimados usando el software NLREG (Non-Linear Regression).

La distribución Poisson compuesta fue propuesta por Sichel (1971) como una familia de distribuciones discretas que proceden de una mezcla de la distribución de Poisson con el parámetro  $\lambda$ . La distribución Poisson compuesta es expresada como:

$$\varphi(r) = \frac{\left((1-\theta)^{\frac{1}{2}}\right)^y}{K_y \left\{ \alpha(1-\theta)^{\frac{1}{2}} \right\}} \frac{\left(\frac{\alpha\theta}{2}\right)^r}{r!} K_{r+y}(\alpha)$$

donde  $r \geq 0$ .

La aplicación del modelo Poisson compuesto fue realizado siguiendo las especificaciones proporcionadas por Sichel (1971, 1974, 1975). En esos documentos Sichel proporciona las ecuaciones pertinentes para calcular los parámetros  $\alpha$ ,  $\theta$  y  $\gamma$  por el método de la máxima probabilidad. Una vez conocidos

esos valores es simple estimar los valores esperados o teóricos de la distribución de la productividad de los autores, que siempre tienen la forma de una J invertida de larga cola. Los valores esperados de este modelo fueron estimados usando el software Mathematica versión 5.2 para Windows.

La distribución Poisson lognormal fue propuesta por Steward (1994). Según este modelo los datos recolectados para el análisis de la productividad de los autores generalmente producen una forma J inversa cero truncada, con una larga cola de grandes productores. Eso hace que el modelo Poisson lognormal sea un candidato ideal para probar este tipo de distribuciones discretas. El modelo es descrito como una distribución compuesta, donde la propensión subyacente  $\delta$  de los científicos para publicar un artículo sigue una distribución lognormal. Dado la propensión subyacente  $\delta$  específica de un científico, su probabilidad  $P_x$  de publicar  $x$  artículos, sigue un simple modelo Poisson:

$$P_x = \frac{\delta^x e^{-\delta}}{x!} \quad \text{para } x = 0, 1, 2, 3, \dots$$

Siendo así las cosas, la distribución de los valores observados de todos los autores que tengan el mismo valor  $\delta$  tendrán una distribución con una media y una varianza  $\delta$ . En una muestra de autores cuyos logaritmos  $\delta$  están normalmente distribuidos con una media  $\mu$  y un desvío padrón  $\sigma$ , las  $P_x$  de la muestra total son proporcionadas por la siguiente ecuación:

$$P_x = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x!} \int_0^\infty e^{-\delta} \delta^{x-1} \exp\left\{-\frac{(\ln \delta - \mu)^2}{2\sigma^2}\right\} d\delta$$

para  $x = 0, 1, 2, 3, \dots, n$

Los valores esperados de este modelo Poisson lognormal fueron estimados con la ayuda de un software gratuitamente proporcionado por el profesor John A. Steward (2005).

La prueba de ajuste usada para los cuatro modelos es el Kolmogorov-Smirnov (K-S), que es aplicado al conjunto de valores observados y esperados a un 0,01 nivel de significancia. La prueba K-S compara la función de densidad de los valores observados con la de los valores esperados o calculados. Una de sus ventajas es que trabaja muy bien con pequeñas muestras, no pierde información con la agrupación de los datos en clases y es más poderosa que la prueba chi-cuadrada.

## RESULTADOS

La *Tabla Núm. 1* muestra los datos recolectados por Moreno-Cabo & Solaz-Portolés (2008a). El 74% de los autores produjeron un solo documento y 8 autores (1%) produjeron 10 o más documentos (6,6%). La producción media de los autores es de 1,6 documentos, pero esta tasa de productividad aumenta a 11,4 documentos para los autores productores de 10 o más documentos. Esta distribución tiene una varianza de 2,38, un desvío padrón de 1,54 y un índice de dispersión igual a 1,48.

**Tabla Núm. 1:** Número de autores productores de documentos sobre el péndulo

Núm. de docs.	Núm. de autores	
x	y	xy
1	638	638
2	123	246
3	41	123
4	22	88
5	10	50
6	6	36
7	7	49
8	5	40
9	3	27
10	5	50
13	1	13
14	2	28
Total	863	1388

La *Tabla Núm. 2* muestra las contribuciones de documentos según los autores observados y los valores estimados (esperados) de acuerdo a los modelos utilizados: cuadrado inverso (CI), Poisson compuesto (PC), Poisson lognormal (PLN) y Poder inverso generalizado (PIG). Si se observa el total de los autores estimados, el modelo Poisson lognormal (con una diferencia de -4) y el modelo del poder inverso generalizado (con una diferencia de +5,4) son los modelos que estiman más coherentemente el total de los autores observados. El modelo del cuadrado inverso produce una diferencia de +142 autores, demasiado alejado de los 863 autores observados. El modelo Poisson compuesto produce una diferencia de -25 autores, es decir, sub-estima los 863 autores observados.

Tabla Núm. 2: Valores observados y estimados según los modelos

Contribuciones	Autores observados	Autores estimados	Autores estimados	Autores estimados	Autores estimados
x	y	CI	PC	PLN	PIG
1	638	638,0	612,53	634,60	638,259
2	123	159,5	120,79	127,93	119,679
3	41	70,9	47,63	44,39	44,953
4	22	39,9	23,48	20,39	22,441
5	10	25,5	12,96	11,03	13,092
6	6	17,7	7,67	6,64	8,429
7	7	13,0	4,75	4,30	5,809
8	5	10,0	3,04	2,95	4,208
9	3	7,9	2,00	2,11	3,166
10	5	6,4	1,34	1,56	2,455
11	0	5,3	0,91	1,18	1,950
12	0	4,4	0,63	0,92	1,58
13	1	3,8	0,44	0,73	1,303
14	2	3,3	0,31	0,58	1,09
Total	863	1005,5	838,50	859,31	868,414

Si observamos los valores de cada una de las filas (en las columnas) estimadas de las contribuciones de los autores, los valores de los modelos Poisson lognormal (PLN) y Poder inverso generalizado (PIG) están más próximas a las contribuciones observadas. También en este aspecto el modelo del cuadrado inverso produce una inflación de la mayoría de los valores esperados.

Para el caso del modelo del cuadrado inverso (CI), los valores fueron estimados con los parámetros  $n = 2$ , y  $C = 638$ . Si se define a los autores productivos como aquellos que produjeron dos o más documentos hay una relación de 225 (26.1%) autores observados contra 368 (36.6%) esperados. Si los autores más productivos son definidos como aquellos que escribieron seis documentos o más el resultado es de 29 autores (3,4%) observados frente a un número esperado de 72 (7,2%) autores. En otras palabras, los valores observados y esperados muestran grandes discrepancias. Estas grandes discrepancias hacen que no soporten ninguna prueba estadística.

Para el modelo Poisson compuesto (PC), los valores estimados fueron calculados con los parámetros  $\alpha = 2.44$ ,  $\theta = 0.87$  e  $\gamma = 1.751$ . Si se define a los autores productivos como aquellos que produjeron dos o más documentos hay una relación de 225 (26.1%) autores observados contra 226 (27%) esperados. Si los autores más productivos son definidos como aquellos que escribieron seis documentos o más el resultado es de 21 (2.5%) de autores esperados ligeramente inferior a los 29 (3,4%) autores observados. En este modelo las discrepancias entre los valores observados y esperados son mínimas.

En el caso de la distribución Poisson lognormal (PLN), los valores estimados fueron calculados con los parámetros  $\beta = -3.6283$  y  $\alpha = 1.8514$ . En este caso si los autores más productivos son aquellos que produjeron dos o más documentos, hay una relación de 225 (26.1%) autores observados contra 225 (26.1%) autores esperados. Si los autores más productivos son aquellos que produjeron seis documentos o más, el resultado es de 29 autores (3,4%) observados frente a un número esperado de 21 (2,42%) autores. En este caso, las diferencias entre los valores observados y esperados de la distribución de la productividad de los autores también están dentro de los límites aceptables.

Para la distribución del Poder inverso generalizado (PIG), los valores estimados fueron calculados con los parámetros  $C = 0.7339$  y  $n = 2.42$ . Si esos mismos criterios son adoptados para la esta distribución PIG, se obtiene una relación de 225 (26.1%) autores observados frente a 230 (26.5%) autores esperados que produjeron más de dos documentos. Si los autores más productivos son definidos como aquellos que escribieron seis documentos o más el resultado es de 21 (2.5%) de autores observados ligeramente inferior a los 30 (3,4%) autores observados. Como se ve, para este modelo, las diferencias entre los valores observados y esperados también están dentro de los límites aceptables.

La *Tabla Núm. 3* muestra el valor crítico y la desviación máxima estimada para los cuatro modelos en estudio usando la prueba estadística Kolmogorov-Smirnov (K-S) con un nivel de 0.01 de significancia.

**Tabla Núm. 3.** Valor crítico y desviación máxima

Modelo	Desviación máxima	Valor crítico	Conclusión
Cuadrado inverso	0.104834	0.055486	No se ajusta
Poisson compuesto	0.01065	0.055486	Se ajusta
Poisson lognormal	0.00998	0.055486	Se ajusta
Poder inverso	0.00903	0.055486	Se ajusta

La prueba de ajuste Kolmogorov-Smirnov (K-S) es un método no-paramétrico para probar si hay diferencias significativas entre las frecuencias observadas y las frecuencias teóricas o calculadas de una distribución. Es una medida de la bondad del ajuste de una distribución de frecuencias similar al chi-cuadrado. Pero esta prueba K-S, es más poderosa que el chi-cuadrado, más fácil de usar y no necesita que los datos estén agrupados en frecuencias inferiores a 5 como lo exige la prueba chi-cuadrada. Es particularmente útil para juzgar cuán próximas están las frecuencias observadas de las frecuencias estimadas. Si el valor crítico de la distribución es menor que su desviación máxima, la distribución no se ajusta al modelo propuesto. En otras palabras, para

que un modelo se ajuste a la distribución observada, el valor crítico de la distribución debe ser siempre mayor que su desviación máxima. Como se pudo observar en la *Tabla Núm. 3*, el único modelo que no se ajusta a la distribución de Lotka es el modelo del cuadrado inverso, puesto que, la desviación máxima (0.104834) es mayor que el valor crítico (0.055486) estimado a un nivel de significación del 0,01. Sin embargo, la mejor predicción es dada por el modelo del poder inverso generalizado (su desviación máxima es el menor entre todos los modelos), luego por el modelo Poisson lognormal (desviación máxima de 0.00998) y finalmente el modelo Poisson compuesto (desviación máxima de 0.01065). Estos tres últimos modelos se ajustan a la distribución de los autores productores de literatura sobre el péndulo. No así el modelo del cuadrado inverso.

La *Figura 1* muestra la aproximación entre los valores observados y estimados según los modelos analizados en este trabajo.

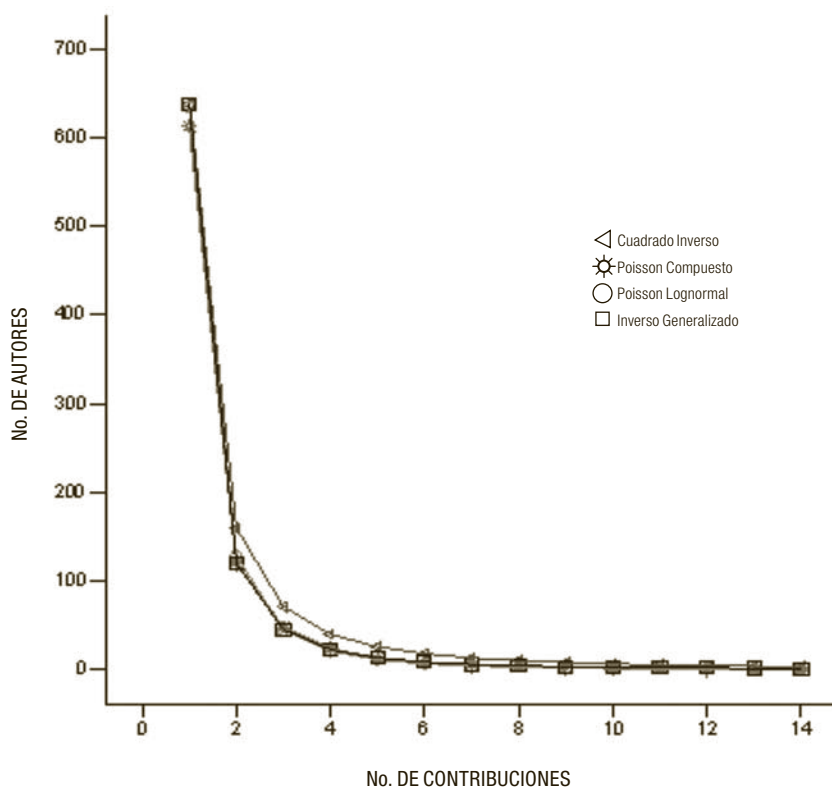


Figura 1: Valores observados y estimados según los modelos



## CONCLUSIÓN

Es conocido que la distribución de la producción científica de los autores en la forma de documentos publicados es altamente sesgada. Este comportamiento es denominado en la literatura como el fenómeno “éxito-genera-éxito”; es decir, es más probable que un autor quien ya publicó un artículo, publique un segundo artículo, un tercer artículo, etcétera, que alguien que aún no ha publicado ningún artículo publique su primer artículo. Este fenómeno se explicita en la forma de un pequeño número de autores dominando el campo en el que actúan con su alta tasa de productividad de documentos. Éste es también el caso de las publicaciones sobre el péndulo donde se observó que el 7% de los autores publicaron el 25% de los documentos, lo que indica además que apenas el 12% de los autores fueron responsables de la producción del 36% del total de documentos publicados sobre este asunto.

Tradicionalmente esta producción estratificada ha sido evaluada a través del modelo del cuadrado inverso propuesto por Lotka (1926), pero ha sido suficientemente probado y comprobado que este modelo no soporta una prueba estadística adecuada. Esta falta de ajuste a una prueba estadística se comprobó en la distribución de autores productores de literatura sobre el péndulo entre 1629 y 1885. Las diferencias con los resultados del artículo de Moreno-Cabo & Solaz-Portolés (2008) puede deberse principalmente a que los autores no estimaron los parámetros del modelo de los datos observados ni probaron la bondad del ajuste de los datos estimados con ningún test estadístico. Este puede ser también el caso del trabajo de Parreiras; Silva; Mathews y Brandao (2006), para encontrar en su estudio un  $n = 24.017$ . En este artículo que replica los datos de la literatura sobre el péndulo, se encontró un  $C = 0.7339$  y  $n = 2.42$  estimados con el modelo del poder inverso generalizado. Estos resultados son diferentes a los hallazgos de los autores mencionados y están más próximos a los valores que se esperan encontrar en una distribución Lotkiana.

Cuando se estudia la productividad de los autores en cualquier campo del conocimiento se deben estimar siempre los parámetros del modelo de los datos observados y luego probar la bondad del ajuste de los datos, ya sea con la prueba chi-cuadrada o la Kolmogorov-Smirnov. Esa prueba es una práctica recomendable y su ausencia puede llevarnos a conclusiones apresuradas.

También es conveniente abandonar el uso del modelo del cuadrado inverso. Hay pruebas suficientes para aseverar que este modelo no se ajusta a los datos observados cuando se usa una prueba estadística. Es hora de abandonarla en favor de modelos más adecuados y que producen mejores resultados. En este trabajo se experimentaron tres modelos alternativos, Poisson

compuesto, Poisson lognormal y poder inverso generalizado, que estiman adecuadamente la productividad de los autores, y usamos la prueba estadística K-S al 0.01 nivel de significancia para tener certeza de la bondad del ajuste de los datos observados de la distribución de frecuencias de la productividad de los autores. En este caso el modelo Poisson compuesto se ajusta a la distribución de autores sobre el péndulo pero en mayor medida lo hacen los modelos poder inverso generalizado y Poisson lognormal.

Existen muchos otros modelos que, contrariamente a las afirmaciones de uno de los dictaminadores de este artículo (Dictamen Núm. 2), no han sido suficientemente explorados: la distribución de Yule, la distribución de Pareto, la distribución binomial negativa, la distribución lognormal, la distribución de Waring, la distribución Poisson compuesta, la distribución Gauss Poisson inversa generalizada y la distribución Poisson lognormal, han sido escasamente estudiadas en América Latina y en otras latitudes. Estos modelos deberían ser explorados en el campo de la bibliotecología y ciencias de la información por las potencialidades que ofrecen para analizar la productividad de los autores. Esta es una tarea que queda por hacer.

## BIBLIOGRAFÍA

- Ahmed, S. M. Z. & Rahman, A. R., "Nutrition literatura of Bangladesh: a bibliometric study", en *Malaysian Journal of Library & Information Science*, 2008, vol. 13, núm. 1, pp. 35-43.
- Bonnevie, E., "Science a multifaceted portrait of a library and information science journal: the case of the Journal of Information Science", en *Journal of Information Science*, 2003, vol. 29, núm. 1, pp. 11-23.
- Budd, J. M., "A bibliometric analysis of higher education literature author(s)", en *Research in Higher Education*, 1988, vol. 28, núm. 2, pp. 180-190.
- Carpintero, H.; Peiró, J. M. & Quintanilla, I., "El Anuario de Psicología, (1969-1974): un estudio estadístico y bibliométrico", en *Anuario de Psicología*, 1977, vol. 16, núm. 1, pp. 22-34.
- Carr, L. J., "The patenting performance of 1,000 inventors during ten years", en *The American Journal of Sociology*, 1932, vol. 37, núm. 4, pp. 569-580.
- Díaz Mujica, D., "Análisis bibliométrico de la revista Archivos Latinoamericanos de Nutrición", *Anales venezolanos de Nutrición*, 2007, vol. 20, núm. 1, pp. 22-29.
- Dictamen Núm.2, de la revista *Investigación Bibliotecológica*, Comunicación por mail recibida el 24 de agosto del 2012.

- Dresden, A., "A report on the scientific work of the Chicago Section, 1897-1922", en *Bulletin of The American Mathematical Society*, 1922, vol. 28, pp. 303-307.
- Dufrenoy, J., "The publishing behavior of biologists", *Quarterly Review of Biology*, 1938, vol. 13, núm. 1, pp. 207-210.
- González Blasco, P., "La producción científica española de 1965 a 1970: un estudio comparado", en *Revista Mexicana de Sociología*, 1975, vol. 37, núm. 1, pp. 217-244.
- Gupta, D. K., "Scientometric study of biochemical literature of Nigeria, 1970-1984: application Lotka's law and the 80/20 rule", en *Scientometrics*, 1989, vol. 15, núm. 3-4, pp. 171-179.
- Hersh, A. H., "Drosophila and the course of research", en *The Ohio Journal of Science*, 1942, vol. 42, núm. 5, pp. 198-200.
- Inönü, E., *A Bibliography of research papers in physics published in the period 1923-1966 by Turkish or Foreign scientists working in Turkey and by scientists of Turkish origin working in foreign countries, accompanied by some observations on the research output of Turkey in physics*, Ankara : Orta Dogu Teknik Universitesi, 1971.
- Johnson, N. L. & Kotz, S., *Discrete distributions*, Boston: Houghton Mifflin, 1969.
- Kuperman, V., "Productivity in the Internet mailing lists: a bibliometric analysis", en *Journal of the American Society for Information Science and Technology*, 2006, vol. 57, núm. 1, pp. 51-59.
- Lotka, A., "The frequency distribution of scientific productivity", en *Journal of the Washington Academy of Sciences*, 1926, vol. 16, núm. 12, pp. 317-323.
- Luor, T.; Johanson, R. E.; Lu, H. P. & Wu, L. I., "Trends and lacunae for future computer assisted learning (CAL) research: an assessment of the literature in SSCI journals from 1998-2006", en *Journal of the American Society for Information Science and Technology*, 2008, vol. 59, núm. 8, pp. 1313-1320.
- Martin, T. W. & Berry, K. J., "Lotka's inverse square law of scientific productivity: a new look at the relationship between professional productivity and institutional status in sociology", en [S. I.] : *Mid-west Sociological Society* (MSS), 1977.
- Moreno-Cabo, M. & Solaz-Portolés, J. J., "Estudio bibliométrico de las publicaciones relacionadas con el péndulo entre los años 1629 y 1885", en *Revista Española de Documentación Científica*, 2008, vol. 31, núm. 4, pp. 639-645.
- Nicholls, P. T., "Price's square root law: empirical validity and relation to Lotka's law", en *Information Processing Management*, 1988, vol. 24, núm. 4, pp. 469-477.
- Pao, M. L., "An empirical examination of Lotka's Law", en *Journal of the American Society for Information Science*, 1986, vol. 37, núm. 1, pp. 26-33.
- \_\_\_\_\_, "Lotka's law: a testing procedure", en *Information Processing & Management*, 1985, vol. 21, núm. 4, pp. 305-320.

- Parreiras, F. S.; Silva, A. B. de O.; Matheus, R. F. & Brandão, W. C. "Redeci: colaboração científica em ciência da informação no Brasil", en *Perspectivas em Ciência da Informação*, 2006, vol. 13, núm. 3, pp. 1-6.
- Patra, S. K. & Chand, P., "Library and information science research in India: A bibliometric study", en *Annals of Library and Information Studies*, 2006, vol. 53, pp. 219-223.
- Radhakrishnan, T. & Kernizan, R., "Lotka's Law and computer science literature", en *Journal of the American Society for Information Science*, 1979, vol. 30, núm. 1, pp. 51-54.
- Rousseau, R., "A table for estimating the exponent in Lotka's law", en *Journal of Documentation*, 1993, vol. 49, núm. 4, pp. 409-412.
- Sen, B. K., Taid, C. A. & Hassan, M. F., "Library and information science literature and Lotka's Law", en *Malaysian Journal of Information Science*, 1996, vol. 1, núm. 2, pp. 89-93.
- Sevukan, R. & Sharma, J., "Bibliometric analysis of research output of biotechnology faculties in some Indian central universities", en *DESIDOC Journal of Library & Information Technology*, 2008, vol. 28, núm. 6, pp. 11-20.
- Sichel, H. S., "On a family of discrete distributions particularly suited to represent long-tailed frequency data", en Laubscher, F. F. (ed.), *Proceedings of the Third Symposium on Mathematical Statistics*, Pretoria, South Africa: C.S.I.R., 1971. pp. 51-97.
- \_\_\_\_\_, "On a distribution representing sentence-length in written prose", en *Journal of the Royal Statistical Society, Serie A (General)*, 1974, vol. 137, núm. 1, pp. 25-34.
- \_\_\_\_\_, "On the distribution law of word frequencies", en *Journal of the American Statistical Association*, 1975, vol. 70, núm. 351, pp. 542-547.
- Solaz-Portolés, J. J.; Moreno-Cabo, M. & Sanjosé López, V., "Aprendiendo cómo se construye la ciencia: el caso del péndulo", en *Latin American journal physics education*, 2008b, núm. 2, vol. 1, pp. 47-50.
- Steward, J. A., "The Poisson -lognormal model for bibliometric/scientometrics distributions", en *Information Processing Management*, 1994, vol. 30, núm. 2, pp. 239-251.
- \_\_\_\_\_, Comunicación personal, 2005.
- Turkeli, A., "Doctoral training environments and post-doctorate productivity of Turkish physicists", en *Hacettepe Bulletin of Social Sciences and Humanities*, 1973, vol. 5, núm. 1, pp. 91-100.
- Urbizagástegui Alvarado, R. & Cortés, M. T., "La productividad de autores en la Revista Geológica de Chile", *Ciencias de la Información*, 2002, vol. 33, núm. 2, pp. 15-25.
- Urbizagástegui Alvarado, R., "El crecimiento de la literatura sobre plantas usadas como colorantes naturales y la productividad de sus autores", en *Revista AIBDA*, 2007, vol. 28, núm. 1, pp 69-110.

- \_\_\_\_\_, “La Productividad de los autores sobre plantas medicinales del Perú”, *Revista ACB: Biblioteconomia em Santa Catarina*, Florianópolis, Brasil, 2007, vol. 12, núm. 2, pp. 235-253.
- \_\_\_\_\_, “A. produtividade dos autores sobre a lei de Lotka”, en *Ciência da Informação*, 2008, vol. 37, núm. 2, pp. 87-102.
- Vlachý, J., “On publication characteristics of research establishments”, en *Czechoslovak Journal of Physics*, 1970, vol. B20, pp. 1149-1155.
- Williams, C. B., “The number of publications written by biologists”, en *Annals of Eugenics*, 1944, vol. 12, núm. 2, pp. 143-146.



