

Modelando la distribución del número de co-autores por artículo

Rubén Urbizagástegui Alvarado *
Cristina Restrepo Arango **

Artículo recibido:
7 de diciembre de 2010.
Artículo aceptado:
14 de marzo de 2011.

RESUMEN

Este artículo explora la modelación del número de autores que colaboran en la publicación de un artículo. Con ese fin se analizaron la distribución geométrica, la distribución Poisson truncada, la distribución Poisson lognormal y la distribución Gauss Poisson inversa generalizada en la literatura producida sobre la Ley de Lotka desde 1922 hasta junio de 2010. Se encontró que la distribución Poisson lognormal estimó más cercanamente el valor total de los documentos observados seguida del modelo Poisson truncado, la distribución geométrica y finalmente el modelo Gauss Poisson inversa generalizada.

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

** El Colegio de México, México. crestrepoarango@yahoo.com

Palabras-clave: Ley de Lotka, Colaboración; Distribución geométrica; Poisson Lognormal; Gauss Poisson Inverso Generalizado; Poisson Truncada; Bibliometría; Cienciometría; Infometría.

ABSTRACT

Modeling the distribution of co-authorships by paper

Rubén Urbizagástegui Alvarado and Cristina Restrepo Arango

In the literature produced about Lotka's law from 1922 to June 2010, the geometric distribution, truncated Poisson distribution, Poisson lognormal distribution, and the generalized inverse Gaussian Poisson distribution are studied to statistically model the number the authors who collaborate in the publication of an article. It was found that the Poisson lognormal distribution more closely estimated the total value of documents, followed by the truncated Poisson model, the geometric distribution, and finally the generalized inverse Gaussian Poisson distribution.

Key words: Lotka's law, Collaboration; Generalized Inverse Gaussian Poisson; Poisson lognormal; Geometric Distribution; Truncated Poisson; Bibliometrics; Cienciometrics; Infometrics.

INTRODUCCIÓN

A través de los años la bibliometría ha desarrollado diversos modelos estadísticos para probar fenómenos sociales recurrentes observados en el campo de la bibliotecología y la ciencia de la información. Por ejemplo, la ley de Lotka describe y predice la distribución de los autores que contribuyen con artículos en un campo determinado (Lotka, 1926); la ley de Bradford se refiere a la productividad de artículos en las revistas en un campo determinado (Bradford, 1934); la ley de Zipf, basada en la ley del mínimo esfuerzo, describe la distribución de ocurrencias de palabras en un texto (Zipf, 1935, 1949). Aunque para calificarlas muchas veces se usa la palabra “ley” en la literatura, en realidad son apenas “regularidades estadísticas” o “funciones estadísticas” que “modelan” el fenómeno social observado. Modelar un fenómeno es

un método cognoscitivo en el que el objeto en estudio es reemplazado con otro, llamado modelo, que cumple con relación al primero unas condiciones de analogía. Después de reemplazar el original por su modelo, el modelo es estudiado y analizado. Las conclusiones y elementos que surjan de este análisis son transferidos al objeto original con base en los criterios y condiciones de analogía y semejanza antes mencionadas (Domínguez Calle, 2010).

Existe otro grupo de distribuciones o regularidades estadísticas no tan populares como las anteriores, pero que forman parte de los estudios bibliométricos. Por ejemplo, el crecimiento de la literatura que comenzó a ser investigada a fines del siglo XIX por Houzeau & Lancaster (1880, citados por Jaschek, 1989:164) y continuaron siendo analizados en el siglo XX por Tamiya (1931), por Stevens (1932), por Wilson & Fred (1935) y otros. Los estudios de la obsolescencia de la literatura iniciada por Gosnell (1943, 1944), Burton & Kebler (1960) y muchos otros más. La circulación de la literatura y la regla 80/20 iniciada por Trueswell, R. L. (1964, 1965) y Trueswell, R. W. (1966, 1968, 1969). El análisis de las citas iniciado por Gross & Gross (1927), Allen (1929), McNelly & Crosno (1930). El acoplamiento bibliográfico propiciado por los trabajos de Fano (1956), Kessler (1960, 1962, 1963), Kessler & Heart (1962). El análisis de las co-citas propuesto por Small (1973), Marshakova (1973). La teoría epidémica de la literatura científica propuesta por Goffman (1964, 1966, 1969) y Goffman & Newill (1964, 1967). La co-autoría en la publicación de artículos y las redes sociales asociadas a la co-autoría, la formación de colegios invisibles, el impacto y visibilidad de la producción científica, entre otros y solo para citar a los iniciadores.

Dentro de ese espectro de preocupaciones de la bibliometría, uno de los fenómenos a los que se ha prestado poca atención es al número de autores que colaboran en una investigación, es decir, el número de autores múltiples participantes en la producción y publicación de un documento. Se ha observado que esta forma de autoría múltiple viene incrementándose en la producción científica desde hace tiempo hasta el punto que “el fenómeno de la autoría múltiple ha llamado la atención y se ha convertido en un asunto importante en la sociología de la ciencia” (Rousseau, 1994:134). Sin embargo, la modelación estadística sobre este asunto en el campo de la ciencia de la información “todavía está en su infancia y se debería poner atención en los modelos más simples” (Rousseau, 1994:137).

Según Ajiferuke (1991: 279),

así como el número de artículos publicados en revistas académicas es un indicador de la productividad, los artículos publicados en co-autoría en una revista son frecuentemente tomados como colaboración en la investigación.

Por esa razón, los investigadores del campo de la bibliometría están interesados en desarrollar modelos estadísticos que puedan describir el número de autores existentes por artículo en diferentes campos temáticos. Ese modelo puede después ser usado para estimar el número de entradas en un índice de autores de una base de datos bibliográfica automatizada y luego

el valor de los parámetros pueden ser usados para estimar el número de N artículos, el número de artículos con un autor, con dos autores, con tres autores y así sucesivamente (Ajiferuke, 1991:284).

Por las razones apuntadas anteriormente y por disponer de los datos pertinentes, el objetivo de este trabajo es explorar algunas distribuciones estadísticas simples en la literatura producida sobre la Ley de Lotka desde 1922 hasta Junio de 2010, con el propósito de modelar el número de autores que colaboran en una investigación; es decir, el número de autores que participan en la producción y publicación de un documento. Para lograr el objetivo establecido este trabajo está organizado en cuatro partes: la primera presenta el marco teórico de la investigación con base en una revisión de la literatura sobre otros trabajos que han aplicado diferentes distribuciones estadísticas sobre este asunto; la segunda describe la metodología que se usará en el desarrollo de este artículo; la tercera explica los resultados encontrados y las conclusiones, y por último se presenta la bibliografía que fue consultada para elaborar esta investigación.

MARCO TEÓRICO

Uno de los primeros trabajos sobre este asunto parece haber sido el de Price & Beaver (1966), quienes usando memorandos que circulaban entre los miembros de un colegio invisible postularon el modelo de Poisson para dar cuenta de la distribución de las co-autorías. Estos autores estudiaron 533 artículos producidos por 1,239 autores en el campo de fosforilación oxidativa y el transporte del electrón terminal. Goffman & Warren (1969) observaron que para la literatura de eschistosomiasis, el número de publicaciones con un autor era cerca de $2/3$ del total de la literatura publicada, aquellos con dos autores equivalía a $2/3$ de los restantes, aquellos con tres autores $2/3$ de la literatura restante, y así sucesivamente. También Worthen (1978) estudiando la literatura publicada sobre dos productos farmacéuticos (Cephalexin y Flurazepam) encontró que de un total de 630 artículos, 30% de la literatura publicada sobre Cephalexin eran de autoría de un único autor. Similarmente, 36% de la

literatura de Flurazepam eran de la autoría de un único autor. Es decir, en el caso de la literatura de Cephalexin, el 70% había sido de autoría de dos o más autores con una media de 2.7 autores por artículo. En el caso de Flurazepam, el 64% fue producido por dos o más autores, con una media de 2.46 autores por artículo. Comparando estos resultados con los obtenidos por Goffman & Warren (1969) afirmaba que

el número de artículos con un autor hacen cerca de 1/3 de la literatura, aquellos artículos con dos autores cerca de 1/3 del resto, y aquellos con tres autores 1/3 de los restantes, y así sucesivamente (Worthen, 1978:193).

Ese comportamiento observado sugería que la distribución geométrica daría cuenta adecuada de la distribución de las co-autorías observadas.

Por su parte, Ajiferuke (1991) estudió un gran número de distribuciones estadísticas para derivar un modelo probabilístico que describa apropiadamente y se ajuste a la distribución de las co-autorías. Estudió datos recopilados en los campos de ingeniería, ciencias médicas, ciencias físicas, matemáticas, ciencias sociales y humanidades, en las cuales probó las distribuciones de Zipf, Mandelbrot, geométrica, Poisson modificado, Poisson generalizado modificado, logarítmica, Borel-Tanner, Yule modificado, Waring generalizado modificado, Gauss-Poisson inversa modificada, beta-binomial modificada, binomial negativa modificada, binomial negativo generalizado modificado y Waring modificada. Encontró que el modelo Gauss Poisson, inversa modificada, describía mejor la distribución de las autorías observadas al igual que la distribución de Waring modificada, aunque esta última no se ajustó bien a los datos de los campos de ingeniería, ciencias médicas y físicas. Sugirió que el modelo

Gauss Poisson inversa modificada puede ser usado para estimar el número de entradas en un índice de autorías y para determinar el número máximo de autores por artículo incluidos en un índice de ese tipo (Ajiferuke, 1991: 285).

Rousseau (1994) analizó distribuciones estadísticas tradicionales de un parámetro en el número de autores de artículos publicados de 1981 a 1990 en el campo de la bibliotecología y ciencia de la información obtenidos de 5 revistas especializadas: *Journal of Documentation*, *Journal of the American Society for Information Science*, *Scientometrics*, *Online Review* y *Libri*. A los datos extraídos de este grupo agregó los autores listados en la bibliografía de su libro *Introduction to Informetrics* (Egghe & Rousseau, 1990). Estudió luego la distribución de Lotka, la distribución geométrica y la distribución cero

truncada de Poisson, y encontró que la distribución geométrica y la distribución de Poisson cero truncada describen adecuadamente estos conjuntos de datos de autoría, pero no la distribución de Lotka.

Gupta, Kumar & Rousseau (1998) usaron una comprehensiva bibliografía sobre genética que cubre un largo periodo, desde 1870 hasta 1980. En esa bibliografía identificaron once revistas especializadas. Después seleccionaron los nombres de los autores que habían publicado juntos uno, dos, tres, ... n artículos que cubrían el periodo 1976 a 1980. En esa población usaron la distribución de Lotka, la distribución geométrica, la binomial y la distribución binomial negativa, y encontraron que la distribución de Poisson y la distribución geométrica describen y se ajustan mejor a los datos de la distribución del número de autores. Nuevamente Kumar & Gupta (2004) estudiaron dos extensas bibliografías en ecología cuantitativa (1920-1974) y en análisis multivariado (1902-1961). Exploraron tres distribuciones estadísticas: la distribución binomial negativa, la distribución geométrica y la distribución Poisson truncada, y encontraron que la distribución binomial negativa y la distribución geométrica se ajustaron adecuadamente a los datos observados de autoría múltiple en ambas áreas estudiadas. Después de 2004 no se han publicado otros trabajos sobre este asunto que sean del conocimiento de los autores de este trabajo.

METODOLOGÍA

Como unidades de análisis usaremos los autores productores de artículos sobre la ley de Lotka desde 1922 hasta junio del 2010. Para identificar a los autores que contribuyeron con artículos en esta área, se hizo una búsqueda en todas las bases de datos de DIALOG, con los términos Lotka?(5n)Law? como estrategia de búsqueda, y se obtuvieron 50 bases de datos que por lo menos contenían un artículo sobre el asunto investigado. Esta estrategia de búsqueda produjo un total de 475 registros que, tras la depuración de los duplicados y falsas recuperaciones, fueron acumulados en un total de 386 referencias bibliográficas. Estas 386 referencias fueron después trasladadas a Pro-Cite 5.0 para elaborar una base de datos específica sobre el asunto. Posteriormente se realizó una minuciosa lectura de cada uno de los artículos identificados en la búsqueda, dedicándole especial atención a cada cita efectuada en el documento leído. Después cada cita referida a la Ley de Lotka fue confrontada con la base de datos específica en Pro-Cite 5.0 e incorporada a ella si no hubiera sido identificada anteriormente en la búsqueda en DIALOG. Con esta lectura minuciosa se produjo una bibliografía analítica sobre la Ley de Lotka

que lista un total de 651 referencias bibliográficas que contienen artículos de periódicos, monografías, capítulos de libros, comunicaciones presentadas en congresos, y literatura gris. Esta forma de recolección de los datos ya fue explicada en una serie de trabajos publicados por Urbizagástegui Alvarado (2008, 2009a, 2009b).

Como la variable número de autores por artículo publicado sólo puede tomar valores discretos se ha supuesto que una distribución discreta puede describir la distribución del número de co-autores en los artículos publicados en ese periodo. Y como en las experiencias anteriores la distribución geométrica, Poisson truncada, Gauss Poisson inversa generalizada y Poisson lognormal fueron las que han dado mejores resultados, en este trabajo usaremos esos modelos. Para el conteo de los autores productores de artículos se optó por el sistema de conteo completo. Esto significa que los múltiples autores de un único artículo fueron contados y atribuidos como autores contribuyentes a la producción de documentos identificados en el levantamiento bibliográfico.

La distribución geométrica es representada de la siguiente manera:

$$P_G(k) = pq^{k-1}$$

donde,

$$q = 1 - p$$

la media es igual a $\frac{1}{p}$

y la varianza es igual a $\frac{q}{p^2}$

El parámetro p es igual a $1/\bar{x}$, donde \bar{x} es la media de la muestra. Por lo tanto p fue estimado de la inversa de la media de la muestra (Gupta, Kumar & Rousseau, 1998:328).

La distribución Poisson truncada es representada de la siguiente manera

$$P_r = \frac{\lambda^r e^{-\lambda}}{r! (1 - e^{-\lambda})} \quad (r = 1, 2, 3, \dots, n)$$

donde

λ = media aritmética de las ocurrencias

$e = 2.71828$ (la base de los logaritmos naturales)

r = valores 1, 2, ..., n de la variable discreta r

Para resolver esta ecuación y calcular los valores esperados o teóricos Plackett (1953) desarrolló un método bastante eficiente para estimar (Lambda). La distribución Gauss Poisson inversa generalizada fue desarrollada por Sichel (1982, 1985, 1986, 1992) y ahora este modelo es conocido como la “distribución de Sichel”. Este modelo de tres parámetros puede ser representado como:

$$\phi(r | t) = \frac{(1 - \theta_t)^{\frac{\gamma}{2}}}{K_{\gamma} \left\{ \alpha_t (1 - \theta_t)^{\frac{1}{2}} \right\}} \frac{\left(\frac{\alpha_t \theta_t}{2} \right)^r}{r!} K_{r + \gamma}(\alpha_t)$$

donde,

$$\alpha_t = \alpha [1 + (t - 1)\theta]^{\frac{1}{2}}$$

y

$$\theta_t = \frac{t\theta}{1 + (t - 1)\theta}$$

Las variaciones y características de dependencia del tiempo y otras propiedades de este modelo son proporcionadas por Sichel (1982). Las probabilidades individuales se obtienen de la ecuación de recurrencia general,

$$\hat{\phi}(r) = \left(\frac{r + \gamma - 1}{r} \right) \theta_t \phi(r-1) + \frac{\alpha_t^2 \theta_t^2}{4r(r-1)}$$

Esta ecuación puede ser simplificada haciendo que \square tenga un valor de medio integral. Entonces si se supone que $\gamma = -\frac{1}{2}$ a priori, entonces la distribución discreta es,

$$\phi(r) = \left(\frac{2\alpha}{\pi} \right)^{\frac{1}{2}} \exp\left\{ \alpha (1 - \theta)^{\frac{1}{2}} \right\} \frac{\left(\frac{\alpha\theta}{2} \right)^r}{r!} K_{r-\frac{1}{2}}(\alpha)$$

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 d de los científicos para publicar un artículo sigue una distribución lognormal. Dada la propensión subyacente d 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 las cosas así, la distribución de los valores observados de todos los autores que tengan el mismo valor δ tendrán una distribución con una media y una varianza δ . En una muestra de autores cuyos logaritmos δ estén normalmente distribuidos con una media μ y un desvío padrón σ , las P_x de la muestra total serán 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 $x = 0, 1, 2, 3, \dots$

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). Los datos de la distribución geométrica, Poisson truncado y Gauss-Poisson Inversa Generalizada fueron evaluados usando el software estadístico SSPS 17.0 versión para Windows y Mathematica 5.0. Las pruebas de ajuste usadas para los cuatro modelos estudiados son el chi-cuadrado (a un nivel de significancia 0,05) y el K-S (a un nivel de significancia de 0,01). 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 el chi-cuadrado. Para la aplicación de la prueba del chi-cuadrado ninguno de los valores de las frecuencias observadas deben ser inferiores a cinco, por lo tanto, las células observadas con autores menores a 5 fueron acumulados con las células adyacentes para producir autores agrupados iguales o mayores a 5.

RESULTADOS

La *Tabla 1* muestra la distribución del número de autores observados frente al número de documentos producidos en el periodo de la investigación. Es notoria la concentración de documentos producidos por autores únicos (53,3% sin colaboración) y 46,7% de los documentos fueron producidos en colaboración por dos o más autores.

Tabla 1: Distribución de autores según el número de documentos producidos

| Núm. de autores | Núm. de documentos | Porcentaje % |
|-----------------|--------------------|--------------|
| 1 | 347 | 53,3 |
| 2 | 181 | 27,8 |
| 3 | 81 | 12,4 |
| 4 | 29 | 4,5 |
| 5 | 9 | 1,4 |
| 6 | 2 | 0,30 |
| 7 | 1 | 0,02 |
| 9 | 1 | 0,02 |
| Total | 651 | 100,0 |

La media de autores por artículo fue de 1,753 con una varianza de 0,57055. El 28% de los documentos fueron producidos por dos autores, 12% por tres autores y así sucesivamente. Trece documentos (2%) fueron producidos en colaboración por 5 o más autores y el máximo número de autores que colaboraron en un único artículo fueron 9. En estos resultados son notorios los artículos producidos por autores únicos, lo cual parece estar de acuerdo con las afirmaciones de Yitzhaki y Ben-Tamar (1990), quienes sostienen que en las ciencias sociales y humanidades la colaboración es menos notoria que en la ciencias duras, ya que en esta última la ejecución de una investigación involucra más recursos humanos, económicos y tecnológicos, lo cual hace necesaria la participación de varios investigadores en la elaboración de un artículo. Éste no parece ser el caso de la bibliometría y en especial del sub-campo productividad científica de los autores.

La *Tabla 2* presenta el número de documentos observados y estimados, de acuerdo con los modelos estudiados a través del método de la máxima probabilidad. En relación al total de documentos producidos y observados (651), la distribución geométrica (DG) estimó -0,32 documentos; la distribución Poisson truncada (PT) estimó +0,10 documentos; el modelo del poder inverso generalizado (PIG) estimó +5,34 documentos; la distribución Poisson lognormal (PL) estimó +0,06 documentos; la distribución Gauss Poisson inversa generalizada (GPIG)

estimó 3,9 documentos. El modelo PL es el que estimó más cercanamente el valor total de los documentos observados seguido del modelo PT, DG y finalmente el modelo GPIG.

Tabla 2: Valores observados y estimados según los modelos

| Núm. de autores | Núm. de docs. | DG | PT | PL | GPIG |
|-----------------|---------------|---------------|--------------|---------------|--------------|
| 1 | 347 | 371,43 | 326,5 | 345,79 | 380,6 |
| 2 | 181 | 159,51 | 204,32 | 184,36 | 132,4 |
| 3 | 81 | 68,5 | 85,24 | 77,40 | 61,4 |
| 4 | 29 | 29,42 | 26,67 | 28,63 | 32,1 |
| 5 | 9 | 12,63 | 6,68 | 9,89 | 17,9 |
| 6 | 2 | 5,43 | 1,39 | 3,30 | 10,4 |
| 7 | 1 | 2,33 | 0,25 | 1,09 | 6,2 |
| 8 | 0 | 1,00 | 0,04 | 0,36 | 3,8 |
| 9 | 1 | 0,43 | 0,01 | 0,12 | 2,3 |
| Total | 651 | 650,68 | 651,1 | 650,94 | 647,1 |

DG $p = 0,57055$

$q = 0,42945$

DPT $\lambda = 1,2515$

PL $\mu = 1,7527$

$\delta = 1,0312$

GPIG $\Theta = 0,695952$

Si se define la co-autoría como aquellos autores que produjeron conjuntamente 2 o más artículos, la distribución geométrica estimó 379,25 (43%) artículos producidos en colaboración por dos o más autores. La distribución Poisson truncada estimó 324,6 (50%) artículos producidos en colaboración por dos o más autores. El modelo Poisson lognormal estimó 305,15 (47%) documentos producidos en colaboración por dos o más autores. La distribución Gauss Poisson inversa generalizada estimó 380,6 (58,8%) documentos producidos en colaboración por dos o más autores. El mejor resultado es el proporcionado por el modelo PL con una diferencia de +1,15 documentos, seguido del modelo PT con una diferencia de + 20,6 documentos, y finalmente la DG y el modelo GPIG.

Si se define la co-autoría como aquellos autores que produjeron conjuntamente 5 o más documentos, el mejor resultado es proporcionado por el modelo PL con una predicción de 14,76 (2,27%) documentos estimados frente a 13 (2%) documentos observados. Luego viene el modelo PT con una predicción de 8,37 (1,3%) documentos estimados frente a 13 (2%) documentos observados. En seguida la DG con una predicción de 21,82 (3,35%) documentos

estimados frente a 13 (2%) documentos observados. Finalmente viene la distribución GPIG con una predicción de 40,6 (6,27%) documentos estimados frente a los 13 (2%) documentos observados.

La *Tabla 3* presenta los valores estimados y críticos de las pruebas estadísticas chi-cuadrado y K-S utilizados para probar la bondad del ajuste de los modelos a los datos analizados en este trabajo. El chi-cuadrado se estimó para GPIG con 2 grados de libertad; para DPT con 3 grados de libertad; para PL con 2 grados de libertad y para DG con 3 grados de libertad. De acuerdo a la prueba chi-cuadrado, los modelos DG y GPIG no se ajustan a los datos observados, pero sí lo hacen los modelos PT y PL. Cuando se usa la prueba K-S todos los modelos se ajustan a los datos observados, pero el mejor ajuste es proporcionado por el modelo PL, seguido del PT, la DG y finalmente GPIG.

Tabla 3: Valores estimados y críticos de las pruebas estadísticas

| Modelos | Chi-Cuadrado | | Kolmogorov-Smirnov | |
|---------|--------------|---------------|--------------------|---------------|
| | Estimado | Valor Crítico | Dmx | Valor crítico |
| DG | 10,3543 | 5,99147 | 0.0378 | 0,0639 |
| PT | 6,9244 | 7,81473 | 0.0316 | 0,0639 |
| PL | 0,4476 | 5,99147 | 0,0034 | 0,0635 |
| GPIG | 46,1245 | 3,84143 | 0,0551 | 0,0639 |

La *Figura 1* muestra la distribución del número de autores que colaboraron en la producción de documentos según los modelos analizados. Se puede observar la cercanía de los modelos estimados a los datos observados.

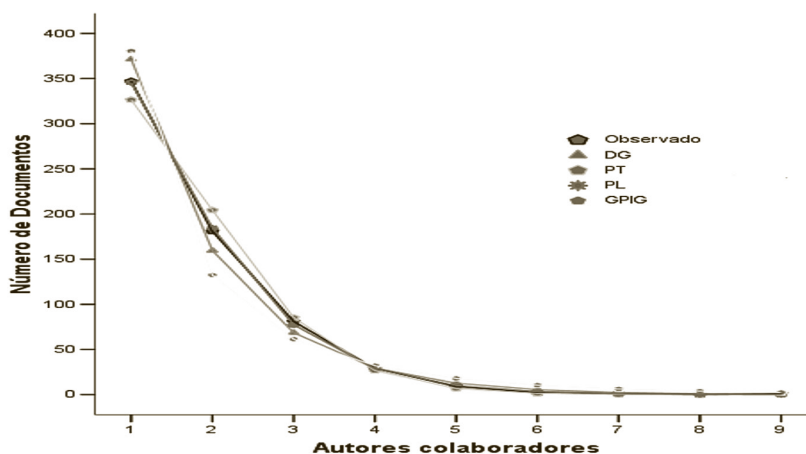


Fig. 1: Colaboración de los autores según los modelos estudiados

CONCLUSIONES

Este trabajo exploró la modelación del número de autores que colaboran en una investigación a través de la aplicación de la distribución geométrica, la distribución Poisson truncada, la distribución Poisson lognormal y la distribución Gauss Poisson inversa generalizada a la literatura sobre la productividad científica de los autores, conocida también como la ley de Lotka, desde 1922 hasta junio de 2010. Se encontró que la distribución Poisson lognormal estimó más cercanamente el valor total de los documentos publicados en co-autoría, seguido del modelo Poisson truncado, la distribución geométrica y finalmente el modelo Gauss Poisson inverso generalizado. La prueba estadística K-S a un nivel de significación de 0,01 confirmó el ajuste en ese mismo orden; sin embargo, la prueba chi-cuadrado a un nivel de significancia de 0,05 rechazó el ajuste de la distribución geométrica y Gauss Poisson inversa generalizada.

Cuando se considera el ciclo de desarrollo de cualquier disciplina, al inicio la colaboración entre los autores (la co-autoría) es nula o baja pero la proporción de autores únicos es alta. Conforme la disciplina se desarrolla comienza la profesionalización, la disciplina se abre a mayores exploraciones y se inicia la especialización. Entonces, los autores comienzan a colaborar y la proporción de autores únicos comienza a decrecer, y al contrario, las publicaciones en colaboración comienzan a crecer. En el caso de la literatura sobre la productividad científica de los autores, estudiada en este trabajo, la proporción de co-autorías (46,7%) es casi similar a la proporción de la producción de los autores únicos (53,3%). Esas cifras indicarían un desarrollo de la bibliometría en dirección a la especialización.

Como se muestra en la literatura revisada, la exploración de la modelación del número de autores que publican en co-autoría no ha sido muy frecuente en la literatura de bibliometría o la ciencia de la información. Sería interesante que se estudien otros modelos en otras áreas del conocimiento para poder comparar las diferencias y/o semejanzas entre los diversos campos del conocimiento.

BIBLIOGRAFÍA

- Allen, Edgard S., "Periodicals for mathematicians", en *Science*, 70(1825):592-594, Dec. 20, 1929.
- Ajiferuke, Isola, "A probabilistic model for the distribution of authorships", en *Journal of the American Society for Information Science*, 42(4):279-289, 1991.

- Bradford, S. C. , “Sources of information on specific subjects”, en *Engineering*, 137:85-86, January 26, 1934.
- Burton, R.E. & Kebler, R.W., “The half-life of some scientific and technical literatures”, en *American Documentation*, 2(1):18-22, January 1960.
- Domínguez Calle, Efraín Antonio, *Modelación matemática : una introducción al método*, disponible en <http://www.mathmodelling.org>, consultado el 25 de Octubre del 2010.
- Egghe, Leo & Rousseau, Ronald, *Introduction to Informetrics: quantitative methods in library, documentation and information science*, New York : Elsevier Science Publishers, 1990.
- Fano, R. M., “Information theory and the retrieval of recorded information”, en *Documentation in action*, New York : Reinhold Publishing Co., 1956; pp. 238-244.
- Goffman, William & Newill, Vaun A., “Generalization of epidemic theory: an application to the transmission of ideas”, en *Nature*, 204(4955),225-228, October 17, 1964.
- Goffman, William, “Mathematical approach to the spread of scientific ideas: the history of mast cell research”, en *Nature*, 212(5061): 449-452, oct. 1966a.
- _____, “Mathematical approach to the spread of scientific ideas: the history of Mast cell research”, en *Nature*, 212(5061):449-452, October 19, 1966b.
- _____, “Stability of epidemic processes”, en *Nature*, 210(5038): 786-787, May 21, 1966c.
- _____, “An application of epidemic theory to the growth of science (symbolic logic from Boole to Gödel)”, en *International Congress of Cybernetics (1st : 1969 : London). Progress of cybernetics: proceedings of the first international congress of cybernetics*, London 1969, London, New York, Gordon and Breach Science Publ., 1970, pp. 971-984.
- Goffman, W. & Newill, Vaun A., “Generalization of epidemic theory: An application to the transmission of ideas”, en *Nature*, 204(4955), 225–228, 1964.
- Goffman, William & Newill, Vaun A., “Communication and epidemic processes”, en *Proceedings of the Royal Society*, 298:316-334, May, 1967.
- Goffman, W. & Warren, K. S., “A mathematical analysis of a medical literature: Schistosomiasis, 1862-1962”, en K. Cheshire (Ed.), *A Symposium: Information in the Health Science: Working to the Future*, Cleveland, Medical Library Association of the Health Sciences Library, Dec. 3-4, 1969. p. 130-153
- Gosnell, Charles F., *The rate of obsolescence in college library book collections as determined by the analysis of three selected list of books for college libraries*, Dissertation: New York University, 1943.

- _____, *Obsolescence of books in college libraries*, College & Research Libraries, 4:115-125, March 1944.
- Gross, P. L. K. & Gross, F. M., "College libraries and chemical education", en *Science*, 66:386-389, 1927.
- Gupta B. M.; Kumar, S & Rousseau, R., "Applicability of selected probability distributions to the number of authors per article in theoretical population genetics", *Scientometrics*, 42(3): 325-334, Jul.-Ago., 1998.
- Jaschek, Carlos, *Data in astronomy*, Cambridge, New York: Cambridge University Press, 1989.
- Kessler, M. M., *An experimental communication center for scientific and technical information*, Lexington, Mass. : Massachusetts Institute of Technology, Lincoln Laboratory, 1960.
- _____, *Analysis of bibliographic sources in a group of physics-related journals*. Lexington, Mass. : Massachusetts Institute of Technology, Lincoln Laboratory, 1962.
- _____, "Bibliographic coupling between scientific papers", en *American Documentation*, 14(1):10-20, 1963a.
- _____, "An experimental study of bibliographic coupling between technical papers", en *IEEE Transactions on Information Theory*, 9(1):49-51, 1963b.
- _____, "Bibliographic coupling extended in time: ten case histories", en *Information Storage Retrieval*, 1(?):169-187, 1963c.
- Kessler, M. M. & Heart, F. E., "Analysis of bibliographic sources", en *Physical Review* (vol. 77, 1950 to vol. 112, 1958). Ft. Belvoir : Defense Technical Information Center, 1962.
- Kumar, Suresh & Gupta, B. M., "Applications of statistical distributions in collaborations data profile of research specialties: case study in quantitative ecology and multivariate analysis", en *WIS-2004: International Workshop on Webometrics, Informetrics and Scientometrics*, 2-5 March 2004 / Editors, Hildrum Kretschmer, Yogendra Singh, Ramesh Kundra. Roorkee, India : Central Library, Indian Institute of Technology Roorkee, 2004. p. 201-209.
- Lotka, Alfred J., "The frequency distribution of scientific productivity", en *Journal of the Washington Academy of Sciences*, 16(12):317-323, June 19, 1926.
- Marshakova, I. V., *Bibliographic coupling system based on cited references*, Nauchno Tekhnicheskaya Informatsiya Seriya, 2:3-8, 1973.
- McNeely, J. K. & Crosno, C. D., "Periodicals for electrical engineers", en *Science*, 72(1856):8, July. 1930.
- Plackett, R. L., "The truncated Poisson distribution", en *Biometrics*, 9(4):485-497, Dec. 1953.
- Price, J. D. de S. & Beaver, D. de B., "Collaboration in an invisible college", en *American Psychologist*, 21:1011-1018, 1966.
- Rousseau, Ronald, "The number of authors per article in library and information science can often be described by a simple probability distribution", en *Journal of Documentation*, 50(2):134-141, June 1994.

- Sichel, H. S., "Asymtotic efficiencies of the three methods of estimation for the inverse Gaussian Poisson distribution", en *Biometrika*, 69:467-472, 1982.
- Sichel, H. S., "A bibliometric distribution which really works", en *Journal of the American Society for Information Science*, 36(5):314-321, 1985.
- _____, "The GIGP distribution model with applications to physics literature", en *Czechoslovak Journal of Physics*, B36(1):133-137, 1986.
- _____, "Anatomy of the generalized inverse Gaussian-Poisson distribution with special applications to bibliometric studies", en *Information Processing and Management*, 28(1):5-17, 1992.
- Small, Henry, "Co-citation in the scientific literature: a new measure of the relationship between two documents", en *Journal of the American Society for Information Science*, 24(4):265-269, July-Aug. 1973.
- Stevens, Neil E., "The fad as a factor in botanical publication", en *Science*, 75(1950):499-504, may 13, 1932.
- Steward, John A., "The Poisson-lognormal model for bibliometric/scientometrics distributions", en *Information Processing Management*, 30(2):239-251, 1994.
- Steward, John A., Comunicación personal, 2005.
- Tamiya, Hiroshi, "Eine mathematische Betrachtung über die Zahlenverhältnisse der in der Bibliographie von Aspergillus zusammengestellten Publikationen", en *The Botanical Magazine*, 45(530):62-71, 1931.
- Trueswell, Richard L., "Two characteristics of circulation and their effect on the implementation of mechanized circulation control systems", en *College and Research Libraries*, 25 (4):285-291, July 1964.
- _____, "A quantitative measure of user circulation requirements and its possible effect on stack thinning and multiple copy determination", en *American Documentation*, 16(1):20-25, Jan. 1965.
- Trueswell, Richard W., "Determining the optimal number of volumes for a library's core collection", en *Libri*, 16(1):49-60, 1966.
- _____, "Some circulation data from a research library", en *College and Research Libraries*, 29 (6):493-495, Nov. 1968.
- Richard .L. Trueswell, "Library Users: The 80/20 Rule," en *Wilson Library Bulletin*, 43(5):458-461, Jan. 1969.
- Urbizagástegui Alvarado, Rubén, "A produtividade dos autores sobre a lei de Lotka", en *Ciência da Informação*, 37(2): 87-102, Maio-Ago, 2008.
- _____, "Elitismo na literatura sobre a produtividade dos autores", en *Ciência da Informação*, 38(2):69-79, maio/ago. 2009a.
- _____, "A frente de pesquisa na literatura sobre a produtividade dos autores", en *Encontros Bibli*, 14(28):38-56, 2009b.

- Wilson, P. W. & Fred, E. B., "The growth curve of a scientific literature: nitrogen fixation by plants", en *The Scientific Monthly*, 41(3):240-250, sept. 1935.
- Worthen, D. B., "Short-lived technical literatures: a bibliometric analysis", en *Meth Inform Med*, 17(3):190-198, 1978.
- Yitzhaki, Hoshe y David Ben-Tamar, "Multiple authorship in Biochemistry and other fields; a case study of the Journal of Biological Chemistry throughout 1905-1988", en *INFORMETRICS 89/90*, L. Egghe and R. Rousseau (Editors) Elsevier Science Publishers B.V., 1990, disponible en:
<http://doclib.uhasselt.be/dspace/bitstream/1942/871/1/yitzhaki373.pdf>, consultado el 9 de octubre de 2010.
- Zipf, George Kingsley, *The psycho-biology of language: an introduction to dynamic philology*, Boston : Houghton Mifflin Co., 1935.
- _____, *Human behavior and the principle of least effort; an introduction to human ecology*, Cambridge, Mass. : Addison-Wesley Press, 1949.

