5-2019

Applicability of Lotka's Law in Parasitology research output of India

NAHEEM KT Mr
Pondicherry University, naheemkt@gmail.com

Sivaraman P. Dr.
Dept. of Library & Information Science, Annamalai University, psraman.p@gmail.com

Saravanan G. Dr.
Library and Publication Division, French Institute of Pondicherry, saravanan.g@ifpindia.org

Applicability of Lotka's Law in Parasitology research output of India

Naheem K.T.,[1]Research Scholar, Dept. of Library & Information Science, Annamalai University, Annamalai Nagar – 608002.; [1]Professional Assistant, Central Library, Pondicherry University, Kalapet, Puducherry-605014. Email: naheemkt@gmail.com;

Dr. P. Sivaraman[2]., Associate Professor, Dept. of Library & Information Science, Annamalai University, Annamalai Nagar – 608002. Email: psraman.p@gmail.com

Dr. G. Saravanan[3]., Librarian, Library and Publication Division, French Institute of Pondicherry. Email: saravanan.g@ifpindia.org

Abstracts

This paper examines the conformity of Lotka's law to authorship distribution in the field of parasitology using Scopus during 2007-2016. Totally, 5792 articles produced by 3473 unique first authors, was compiled for analysis. Lotka's law was tested using both generalized and modified forms by using the formula: $x^n y = c$, the values of the exponent *n* and the constant *c* were computed; and Kolmogorov-Smirnov (K-S) and Chi-square tests were applied. The results showed that the Lotka's law fit to the author productivity distribution pattern in parasitology literature.

Keywords: Lotka's law, Author productivity, Parasitology, Kolmogorov-Smirnov (K-S) statistical test, Chi-square test

Introduction

Lotka law of author productivity is one of the prominent laws of bibliometric research. In 1926, Lotka[1] proposed his inverse squire law. It illustrates the frequency of publication by authors in any specified field or subject concerned. According to the law, "the number of authors making *n* contributions is about $1/n^2$ of those making one; and the proportion of all contributors, that make a single contribution, is about 60 percent". Lotka's Law is often called "inverse square law" indicating that there is an inverse relation between the number of publications and the number of authors producing these publications. The generalized form of Lotka's law can be expressed as "$x^n y = c$", where *y* is the number of authors with *x* articles, the exponent *n* and constant *c* are parameters to be estimated from a given set of author productivity data. The applicability of Lotka's law in different disciplines has been studied. The present study is an attempt to apply

Lotka's law and test its conformity on the authorship distribution in the field of parasitology literature emanated from India using Scopus data during 2007-2016.

Review of literature

There have been a good number of studies on the application and validation of Lotka low is available in literature. Sudheer[2] applied Lotka low to physics literature and reported that the distribution author productivity doesn't fit Lotka's low. Askew[3] examined the conformity of Lotka's law in the Field of Library and Information Studies. Results of this study confirmed the validity of Lotka's law for predicting author productivity in the field of library and information studies. Kumar and Senthilkumar[4] applied and tested the conformity of Lotka's law in Astronomy & Astrophysics literature. Reported that, Lotka's law doesn't fits to the literature of Astronomy & Astrophysics. Suresh Kumar[5] verified the authorship frequency distribution on 2106 publications references of the Journal of Documentation published during 2003-2015 using K-S test and concluded that Lotka's law fit to the data set of LIS literature. Sevukan and Sharma[6] studied the biotechnology research in central universities of India and found that the Lotka's law fits to the data of Biotechnology literature. Naqvi and Fatima[7] studied applicability of Lotka's law in international business literature and found that Lotka's distribution is applicable to international business literature. Kumar[8] analysed the distribution of author productivity in the Human-Computer Interaction (HCI) research output and reported the conformity of Lotka's law to HCI literature.

Objectives

1. To analyse the author productivity patterns in the field of parasitology literature
2. To examine the validity of Lotka's law using straight count method of authors
3. To apply Kolmogorov-Smirnov (K-S) goodness-of-fit test and Chi-square test and for the conformity of Lotka's law.

Methodology

The data for this study were retrived from Scopus. Following key words ("Parasitol*', 'Parasitology', 'Medical Parasitology', 'Parasites', 'Parasitic Diseases') have been used in the Scopus search window of title, abstract and keywords to retrieve the records. In addition to this,

'India' is used as a country of affiliation of the author. A total of 5792 records under the document category articles were retrieved. As stated in the objective section, straight count of first author was made and identified 3473 unique authors. Analysis was performed on these 5792 research articles along with 3473 unique authors using MS-Excel.

Analysis and Results

Distribution of research contributions

Table 1 shows frequency distribution of author productivity in the field of parasitology research in India. Of the 3473 unique author names, 2525 (72.70%) produced one article, 492 (14.17%) produced two articles, 202 (5.82%) produced three articles and so on. It is evident from table 1 that the number of authors who produced > 10 articles is quite small.

Table 1: Frequency distribution of research contributions

No. of articles	No. of authors observed	Percentage of authors	Total no. of contributions
1	2525	72.70	2525
2	492	14.17	984
3	202	5.82	606
4	94	2.71	376
5	51	1.47	255
6	33	0.95	198
7	21	0.60	147
8	14	0.40	112
9	8	0.23	72
10	4	0.12	40
11	6	0.17	66
12	2	0.06	24
13	6	0.17	78
14	1	0.03	14
15	1	0.03	15
16	1	0.03	16
18	2	0.06	36

19	3	0.09	57
21	3	0.09	63
22	1	0.03	22
23	1	0.03	23
29	1	0.03	29
34	1	0.03	34
Total	**3473**	**100**	**5792**

Calculation of the exponent '*n*'

The first step in the application of Lotka's law is to determine the value of exponent 'n' which is to be determined by the Linear Least Square (LLS) method by using the following formula.

$$n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - \left(\sum X\right)^2}$$

N = number of pairs of data
X = logarithm of x, i. e. number of publications
Y = logarithm of y, i. e. number of authors

To compute the parameter 'n' data given in Table 2 is used. By substituting the values in the above equation , the value of 'n' is calculated as:

$$n = \frac{23 * 12.711 - 22.875 * 20.685}{23 * 25.951 - (22.875)2}$$

$$n = \frac{292.353 - 473.169}{596.873 - 523.265}$$

$$n = \frac{-180.816}{73.608}$$

$$n = -2.46$$

Table 2: Calculation of n-Straight count method

No. of pairs	***x***	***yx***	***X=lg x***	***Y=lg y***	***X²***	***XY***
1	1	2525	0.000	3.402	0.000	0.000

2	2	492	0.301	2.692	0.091	0.810
3	3	202	0.477	2.305	0.228	1.100
4	4	94	0.602	1.973	0.362	1.188
5	5	51	0.699	1.708	0.489	1.194
6	6	33	0.778	1.519	0.606	1.182
7	7	21	0.845	1.322	0.714	1.117
8	8	14	0.903	1.146	0.816	1.035
9	9	8	0.954	0.903	0.911	0.862
10	10	4	1.000	0.602	1.000	0.602
11	11	6	1.041	0.778	1.084	0.810
12	12	2	1.079	0.301	1.165	0.325
13	13	6	1.114	0.778	1.241	0.867
14	14	1	1.146	0.000	1.314	0.000
15	15	1	1.176	0.000	1.383	0.000
16	16	1	1.204	0.000	1.450	0.000
17	18	2	1.255	0.301	1.576	0.378
18	19	3	1.279	0.477	1.635	0.610
19	21	3	1.322	0.477	1.748	0.631
20	22	1	1.342	0.000	1.802	0.000
21	23	1	1.362	0.000	1.854	0.000
22	29	1	1.462	0.000	2.139	0.000
23	34	1	1.531	0.000	2.345	0.000
	Total	**3473**	**22.875**	**20.685**	**25.951**	**12.711**

Estimation of value 'c'

The value of parameter *n* is calculated as, *n*= -2.46. Substituting the given value of *n*, the value of *c* is estimated from the Table of exponents given by Rousseau[9] as, *c* = 0.73

Goodness-of-Fit Tests

There are several statistics available for goodnessof- fit tests. Among those tests, the Chi-square test and Kolmogorov-Smirnov (K-S) test commonly used as goodness-of-fit tool.

K-S Statistical Test

To test the applicability of Lotka's law, Coile[10] recommends the K-S statistical test. For applying K-S test , the observed and expected number of authors were converted into fractional values and take the difference between cumulative fractional values of observed and expected number of authors. The K-S test is performed to examine the conformity of the observed author

distribution versus Lotka's distribution. The obtained data are tested against the modified Lotka's Law with the exponent *n* being 2.46 and presented in Table 3.

The maximum difference value, *D*max, representing the maximum deviation is identified as 0.020. The critical value, to centrality level a = 0.01, is calculated according to Nicholls[11] as:

$cv = 1.63 / \sqrt{\sum yx} + \sqrt{\sum yx}/10$

$cv = 1.63 / \sqrt{3473} + \sqrt{3473}/10$

$cv = 1.63 / \sqrt{3473} + 18.6$

$cv = 1.63 / \sqrt{3491.6}$

$cv = 1.63/59.08$

$cv = 0.027$

The critical value is obtained as 0.027, thus while comparing the actual value *D max*, 0.020 with critical value 0.027 , it is found that the actual value of *D max* is lower than the critical value (0.027) at the 0.01 level of significance. Therefore, therefore these data fit modified Lotka's law with the value *n* = 2.46.

Table 3. K-S test on observed and expected distribution of authors

x	*y*	*yx/∑yx*	*∑(yx/∑yx)*	*1/x^n*	*fe = C (1/x^n)*	*∑fe*	*Dmax*
1	2525	0.727	0.727	1.000	0.730	0.730	-0.003
2	492	0.142	0.869	0.182	0.133	0.863	0.006
3	202	0.058	0.927	0.067	0.049	0.912	0.015
4	94	0.027	0.954	0.033	0.024	0.936	0.018
5	51	0.015	0.969	0.019	0.014	0.950	0.019
6	33	0.010	0.978	0.012	0.009	0.959	**0.020**
7	21	0.006	0.984	0.008	0.006	0.965	**0.020**
8	14	0.004	0.988	0.006	0.004	0.969	0.019
9	8	0.002	0.990	0.004	0.003	0.972	0.018
10	4	0.001	0.992	0.003	0.003	0.975	0.017
11	6	0.002	0.993	0.003	0.002	0.977	0.017
12	2	0.001	0.994	0.002	0.002	0.978	0.016
13	6	0.002	0.996	0.002	0.001	0.980	0.016
14	1	0.000	0.996	0.002	0.001	0.981	0.015

15	1	0.000	0.996	0.001	0.001	0.982	0.014
16	1	0.000	0.997	0.001	0.001	0.983	0.014
18	2	0.001	0.997	0.001	0.001	0.983	0.014
19	3	0.001	0.998	0.001	0.001	0.984	0.014
21	3	0.001	0.999	0.001	0.000	0.984	0.015
22	1	0.000	0.999	0.000	0.000	0.984	0.015
23	1	0.000	0.999	0.000	0.000	0.985	0.015
29	1	0.000	1.000	0.000	0.000	0.985	0.015
34	1	0.000	1.000	0.000	0.000	0.985	0.015

*n = 2.46, c = 0.73, $D\ max$ = 0.020

x = Frequency of papers
yx=Relative frequency of authors with X research papers,
$yx/\sum yx$ = Fraction of observed number of authors,
$\sum(yx/\sum yx)$= Cumulative fraction of observed number of authors,
$fe = C\ (1/x^n)$= Fraction of expected number of authors,
$\sum fe$ = Cumulative of theoretical value of authors,
$D\ max$ = Difference of the observed and expected cumulative value of authors.

Chi square test on author's productivity

To check whether the author productivity distribution follows the Lotka's law or not, the Chi-square test is applied to the data. The results of the analysis are tabulated in Table 4.

Table 4: Chi square test

x	***fo***	***fe***	***fo-fe***	***(fo-fe)²***	***Chi***
1	2525	2526	-1	1	0.00
2	492	459	33	1089	2.37
3	202	169	33	1089	6.43
4	94	83	11	121	1.45
5	51	48	3	9	0.18
6	33	31	2	4	0.13
7	21	21	0	0	0.00
8	14	15	-1	1	0.07
9	8	11	-3	9	0.79
10	4	9	-5	25	2.85

11	6	7	-1	1	0.14
12	2	6	-4	16	2.86
13	6	5	1	1	0.22
14	1	4	-3	9	2.35
15	1	3	-2	4	1.24
16	1	3	-2	4	1.48
18	2	2	0	0	0.00
19	3	2	1	1	0.55
21	3	1	2	4	2.49
22	1	1	0	0	0.00
23	1	1	0	0	0.00
29	1	1	0	0	0.00
34	1	0	1	1	3.07
				Chi	**28.69**

The calculated value of chi-square obtained is 28.69 and the critical value at 5% level of significance is 35.17. On comparing, it is found that the calculated value of Chi-square is less than the critical value of chi-square. Thus, again it is conclude that the Lotka's law fit to the data under study.

The statistical tests showed the Lotka's law fit to the author distribution pattern in parasitology literature. We can therefore conclude that author productivity in this hypothetical research area fits Lotka's law.

Conclusion

Analysis of author distribution pattern is one of the core area of research in library and information field. Generally authorship distributions are checked by using the famous Lotka's inverse square law of author productivity. This law is considered as one of the prominent law in bibliometric studies. The present analysis reveals that Lotka's law is relevant to parasitology research output in India. Chi-square test and K-S test are applied to verify the conformity of Lotka's law and found that this law fit to the author distribution pattern using straight count

method. This is a primary study on authorship distributions in the field of parasitology research of India and this study may stimulate more such studies in various subject fields.

References

1. Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington academy of sciences*, *16*(12), 317-323.

2. Sudhier, K. P. (2013). Lotka's law and pattern of author productivity in the area of physics research. *DESIDOC Journal of Library & Information Technology*, *33*(6). 457-464.

3. Askew, C. A. (2008). An Examination of Lotka's law in the Field of Library and Information Studies. *FIU ElectronicTheses and Dissertations*. 182.

4. Kumar, S., & Senthilkumar, R. (2019). Applicability of Lotka's Law in Astronomy & Astrophysics Research of India. *Library Philosophy & Practice*. 2129. http://digitalcommons.unl.edu/libphilprac/2129

5. Suresh Kumar, P.K. (2017). Author productivity and the application of Lotka's Law in LIS publications. *Annals of Library and Information Studies (ALIS)*, *64*(4), 234-241.

6. Sevukan, R., & Sharma, J. (2008). Bibliometric analysis of research output of biotechnology faculties in some Indian central universities. *DESIDOC Journal of Library & Information Technology*, *28*(6), 11-20.

7. Naqvi, S. H., & Fatima, N. (2018). Authorship patterns in international business literature: applicability of Lotka's Law. *Annals of Library and Information Studies (ALIS)*, *64*(4), 253-259.

8. Kumar, S. (2015). Author productivity in the field Human Computer Interaction (HCI) research. *Annals of Library and Information Studies (ALIS)*, *61*(4), 273-285.

9. Rousseau, R. (1993). A table for estimating the exponent in Lotka's law. *Journal of documentation*, *49*(4), 409-412.

10. Coile, R. C. (1977). Lotka's frequency distribution of scientific productivity. *Journal of the american society for information science*, *28*(6), 366-370.

11. Nicholls, P. T. (1986). Empirical validation of Lotka's law. *Information Processing & Management, 22*(5), 417-419.