Winter 12-24-2018

Testing of Lotka's law and its Suitability to research productivity of Annamalai University, A Higher Education Institution, South India

Sadik Batcha M
msbau@rediffmail.com

Testing of Lotka's law and its Suitability to research productivity of Annamalai University, A Higher Education Institution, South India

Sadik Batcha M*
Sivaraman P
Associate Professors, Dept. of Library and Information Science,
Annamalai University, Annamalainagar, Tamil Nadu,
South India -608 002, mob: 9443665624, 9443673113.
*corresponding Author ; Email- batchadlis@gmail.com, *msbau@rediffmail.com*

Abstract

Lotka's Law describes the frequency of publications by authors in a given subject/discipline. In the present study, an attempt has been made to study the suitability of the Lotka's Law to the publications of a Higher Educational institution of a region consisting of academic authors and researchers in various disciplines. Annamalai University has been ranked 20th in the state level and 56th at National level by MHRD. Annamalai University is renowned for its research output performance and has been funded for many major, minor and DST research projects. It has a large network and linkages to academia, R&D organisations and industries. It covers almost all the subjects with 10 Faculties and 49 departments of study. Examines the applicability of Lotka's Law as a general inverse power (a # 2) and as an inverse square power relationship (a = 2) to the distribution of the research productivity Annamalai University, South India. Two datasets of the research papers (936 and 3370) contributed by Annamalai University academic authors and researchers during the period of 2000-2006 and 2011-2017 were collected from Web of Science Database. A K-S Test was applied to measure the degree of agreement between the distribution of the observed set of data against the inverse general power relationship and the theoretical value of a=2. It was found that the inverse square law of Lotka does not show conformity as such.

Key Words: *Scientometrics, Lotka's Law, Lotka's empirical law, Lotka's Exponent value, Kolmogorov-Smirnov test, Annamalai University Research output*

Introduction

Scientometric is an important field in information science as it represents a unique set of techniques and tools for the monitoring and analysis of information resources and for the management of knowledge in social and organizational contexts. The Present study is a Scientometric analysis on Research productivity of academic authors of Annamalai University analyzed with two sets of data downloaded from WOS. An attempt has been made in this study

to find out the suitability of Lotka's Law for the research output of Annamalai University, Tamil Nadu, South India.

There are three basic laws of bibliometrics, one that deals with the frequency distribution of scientific productivity[1], another pertaining to scattering of literature[2] and the third concerning word frequency in the text[3]. These laws, extensively studied by Fairthorne, are found to show some stable regularities[4]. Alfred J. Lotka (1926) in his paper on " Analysis of the number of publications in Chemical Abstracts from 1907 to 1916 on frequency distribution of scientific productivity" proposed an inverse square law of scientific productivity which reads, if X authors contribute exactly one paper each, then a the number of the authors contributing n papers will be expressed in the terms of the equation;

$A_n = X/n^2$ for n = 1, 2, 3, ….etc

Where, A_n= number of authors contributing n papers; x=number of authors contributing one paper; n= 1,2,3,.....i. It is also expressed as $1/ n^2$ where 1= X

Lotka's[5] article was not cited for 15 years until 1941 and his distribution was not termed as Lotka Law until 1949. This law has been tested by many studies. Murphy (1973)[6] mapped in humanities and Schorr (1975)[7] on librarianship, Pao (1986)[8] has also tested this law. Booksteen (1977)[9] gave modified theoretical models for Lotka's Law.

Lotka, based on two sets of data proposed an empirical law of scientific productivity, i.e., Inverse Square Law, to measure the scientific productivity of authors as:

Yx = k/x2 ... (1)

Bookstein(1977)[10] presented a generalized form of Lotka's Inverse Square Law as Lotka's Inverse Power Law as:

Yx = k/xá … (2)

Where, Yx = the relative frequency of authors publishing
x = the number of papers
k & á = constant

The Lotka's law and its modified version have a wide range of applicability to a variety of phenomena and the mere form of such a distribution throws a little or no light on the underlying physical conditions. The fact that exponent has approximately the value of 2 as depicted, enables us to state the results in the following simple form:

Yx = 6/ð2x2 ... (3)

as,

k = 6/ð2 for á =2

The formula has two parts: an exponent with value 2 in the inverse square, i.e., 1/x2 portion of expression and a constant 6/ð2 (or 600/ð2 in percent) determined by the value of the exponent.

Review of literature

The studies on the fitness of Lotka's Law began in a systematic manner that started with the work of Pao[11] using least square method with 48 sets of authors' productivity data. The study found that majority of data sets conform to Lotka's distribution as a general inverse power function (á = 2). Lemoline's study[12] regarding the productivity pattern of the CSIR scientists revealed that the distribution corresponding to the entire authors population does not follow an inverse power relationship – either general or square and showed the existence of a flat distribution amongst those researchers with 10 or less number of articles and of an inverse square power relationship for those who have written more than 10 articles. Radhakrishnan and Kernizan[13] concluded that Lotka's Law in the generalized form seems applicable while considering the publication of authors in one periodical whereas when considering the publications of authors in various journals, the observed values deviate considerably from the value predicted by the law. The findings of Narendra Kumar[14] Amsaveni[15,] Sadik Batcha[16] corroborate with the Lotka's findings and values of 'á' were found to be around two.

Annamalai University, accredited with 'A' Grade by NAAC in 2014, is one of India's largest public residential universities with 10 Faculties and 49 departments of study. "The NIRF-2017**"** by the Ministry of Human Resource Development (MHRD) has ranked the University as 20th in Tamil Nadu and 92nd in India in the Overall Category. In the University Category the ranking is 14th in Tamil Nadu and 56th in India. In the Pharmacy Category the ranking is 2nd in Tamil Nadu and 13th in India. "The Times Higher Education World University Ranking - 2018**"** has ranked Annamalai University in 801 - 1000 for Overall category and 401 - 500 for Life Sciences category. "The CII**",** in its Report based on "Indian Citation Index" Database has ranked Annamalai University second among the Top 50 State Universities in Research Productivity: Ranking based on Articles, Citations and Citations/Paper. As far as the Global Exposure, Indian Science Ascending, a Springer Nature report, done in conjunction with Confederation of Indian Industries, has ranked the University as 11th among the top 20 Indian Institutions in International Collaborations.

Methodology

To test the suitability of Lotka's law in Research productivity of academic authors of Annamalai University, The data for this study was collected from WOS (Web of Science Data Base) for the two different sets of seven year periods, 2000 to 2006 and 2001 to 2017. The data was

downloaded on 19-12-2017. The searches were performed on the name of Annamalai University with all probabilities and bibliographical details of 936 and 3370 research papers collectively contributed by 3017 and 12,827 academic authors published in 302 and 929 scientific periodicals respectively were collected for application of Lotka's Law.

Analysis of Data

The observation is made from the dataset 1 for the period 2000 to 2006 of Table 1 that There have been 501 authors with one, 160 authors with two, 80 authors with three and 42 authors with four papers each accounted to their credit and so on. Similarly, from dataset 2 for the period 2010 to 2017 (Table 1), 2202 authors have one, 598 authors have two, 245 authors have three and 138 authors have four papers each to their credit and so on. The maximum number of papers by an individual is found to be 62 and 161 from DS-1 and DS-2 respectively. Tests have been conducted to find out as to what extent, the author productivity conforms to the Lotka's Law.

Table 1 : Number. of Expected Authors Derived with the value of á = 2

No of Papers	No of Authors observed	Data set 1 expected	No of Authors observed	Data set 2 expected
1	501	888	2202	2202
2	160	222	598	551
3	80	99	308	245
4	42	56	186	138
5	33	36	129	88
6	23	25	94	61
7	20	18	78	45
8	14	14	53	34
9	8	11	42	27
10	7	9	24	22
Total	**888**	**1376**	**3714**	**3413**

Considering the formula of Lotka that for DS-1, 501 and for DS-2, 2202 authors have produced one article each, the values of constant 'k' can be easily derived, by putting the value of Yx, i.e., 1 and á =2 in equation 1. 501 = k/1á and 2202 = k/1á. By taking the value of á as 2 (Lotka's exponent), the expected values as in Table 1 are obtained. It can be observed from the Table 1 that the expected values of these two sets are not exactly close to the observed values except five and six papers' values. The frequency distribution of authorship of Annamalai University research contribution is tested by applying the Lotka's law. An attempt is made to find out the Lokta's exponent value for the Annamalai University data, calculated by using the formula proposed by Pao[17].

$$á = \frac{N Ó XY – Ó X Ó Y}{\qquad\qquad\qquad\qquad} \quad \text{…..(4)}$$

N ÓX2 – (ÓX)2

Where, N = no. of the pairs of data considered
X = the logarithm of x
Y = the logarithm of y

Table 2: Frequency Distribution for Lotka's Exponent of Data Set-1

x	y	X=Log X	Y=Log Y	XY	X2
1	501	0.00000	2.699838	0.000000	0.000000
2	160	0.30103	2.204120	0.663506	0.090619
3	80	0.47712	1.903090	0.908005	0.227645
4	42	0.60206	1.623249	0.977293	0.362476
5	33	0.69897	1.518514	1.061396	0.488559
6	23	0.77815	1.361728	1.059630	0.605519
7	20	0.84510	1.301030	1.099498	0.714191
8	14	0.90309	1.146128	1.035057	0.815572
9	8	0.95424	0.903090	0.861767	0.910579
10	7	1.00000	0.845098	0.845098	1.000000
	888	**6.559763**	**15.505885**	**8.511250**	**5.215159**

$$n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - \left(\sum X\right)^2} = \frac{10(8.511250) - (6.559763)(15.505885)}{10(5.215159) - (6.559763)^2} = -1.82022$$

Theorotical Value of n Value is 1.82 which is matched with the table value of R.ROUSSEAU for getting C.S value -0.5397.

Table 2 presents the values which are determined from the logarithmic table and required for equation no. 4. Based on equation no. 4, the Lotka's exponent is calculated for Annamalai University Data Set 1 as á = 1.82.

Table 3 also presents the values required for Data Set- 2 for calculating the value of á from equation 4. The value of exponent for Annamalai University Data Set- 2 data is calculated as á= 1.83 while theoretical Lotka's value is á = 2.000.

To apply the Kolmogorov-Smirnov test for the fitness of Lotka's Law to two sets of Annamalai University publications, Tables 4 and 5 present the expected value of distribution of authors in Data Set- 1 and Data Set- 2 of Annamalai University publications respectively and theoretical value of Lotka to find out the D-Max values.

Table 3: Frequency Distribution for Lotka's Exponent of Data Set-2

x	y	X=Log X	Y=Log Y	XY	X2
1	2202	0.000000	3.342817	0.000000	0.000000
2	598	0.301030	2.776701	0.835870	0.090619
3	308	0.477121	2.488551	1.187340	0.227645
4	186	0.602060	2.269513	1.366383	0.362476

5	129	0.698970	2.110590	1.475239	0.488559
6	94	0.778151	1.973128	1.535392	0.605519
7	78	0.845098	1.892095	1.599005	0.714191
8	53	0.903090	1.724276	1.557176	0.815572
9	42	0.954243	1.623249	1.548973	0.910579
10	24	1.000000	1.380211	1.380211	1.000000
	3714	**6.559763**	**21.581131**	**12.485591**	**5.215159**

$$n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - \left(\sum X\right)^2} = \frac{10(12.485591) - (6.559763)(21.581131)}{10(5.215159) - (6.559763)^2} = -1.83215$$

Theoretical Value of n Value 1.83 is matched with the table value of R.ROUSSEAU for getting C.S value -0.5438

Table 4: K-S Test of Observed and Expected Distribution of Authors in Annamalai University Publications Data Set- 1

	Present study Value n = 1.82 c = 0.5397						**Lotka's Theoratical Value n = 2, c = 0.6079**		
x	**y_x**	**Observed =yx/∑yx**	**Value = ∑(yx/∑yx)**	**Expected Freq**	**Value of Freq/ Cum**	**Diff(D)**	**Expected Freq**	**Value of Freq/ Cum**	**Diff(D)**
1	501	0.5642	0.5642	0.5397	0.5397	0.0245	0.6079	0.6079	-0.0437
2	160	0.1802	0.7444	0.1529	0.6926	0.0273	0.1520	0.7599	0.0282
3	80	0.0901	0.8345	0.0731	0.7656	0.0170	0.0675	0.8274	0.0225
4	42	0.0473	0.8818	0.0433	0.8089	0.0040	0.0380	0.8654	0.0093
5	33	0.0372	0.9189	0.0288	0.8378	0.0083	0.0243	0.8897	0.0128
6	23	0.0259	0.9448	0.0207	0.8585	0.0052	0.0169	0.9066	0.0090
7	20	0.0225	0.9674	0.0156	0.8741	0.0069	0.0124	0.9190	0.0101
8	14	0.0158	0.9831	0.0123	0.8864	0.0035	0.0095	0.9285	0.0063
9	8	0.0090	0.9921	0.0099	0.8963	-0.0009	0.0075	0.9360	0.0015
10	7	0.0079	1.0000	0.0082	0.9044	-0.0003	0.0061	0.9421	0.0018
	888	Dmax = 0.0273					Dmax = 0.0282		

á = 1.82
Theoretical Value of c = 0.5397
Fe+ = $0.5397(1/X^{1.82})$
D-Max = 0.0273

á = 2
Theoretical Value of c = 0.6079
Fe+ = $0.6079(1/X^{2.00})$
D-Max = 0.0282

Critical Value at .01 level of significance = $1.82/\sqrt{888}$
= 0.0611

Table 5: K-S Test of Observed and Expected Distribution of Authors in Annamalai University Publications Data Set- 2

	Present study Value n = 1.83 c = 0.5438						Lotka's Theoratical Value n = 2, c = 0.6079		
x	y_x	Observed =yx/∑yx	Value = ∑(yx/∑yx)	Expected Freq	Value of Freq/ Cum	Diff(D)	Expected Freq	Value of Freq/ Cum	Diff(D)
1	2202	0.5929	0.5929	0.5438	0.5438	0.0491	0.6079	0.6079	0.0150
2	598	0.1610	0.7539	0.1530	0.6968	0.0081	0.1520	0.7599	0.0090
3	308	0.0829	0.8368	0.0728	0.7696	0.0101	0.0675	0.8274	0.0154
4	186	0.0501	0.8869	0.0430	0.8126	0.0071	0.0380	0.8654	0.0121
5	129	0.0347	0.9217	0.0286	0.8412	0.0061	0.0243	0.8897	0.0104
6	94	0.0253	0.9470	0.0205	0.8617	0.0048	0.0169	0.9066	0.0084
7	78	0.0210	0.9680	0.0154	0.8771	0.0056	0.0124	0.9190	0.0086
8	53	0.0143	0.9822	0.0121	0.8892	0.0022	0.0095	0.9285	0.0048
9	42	0.0113	0.9935	0.0098	0.8990	0.0016	0.0075	0.9360	0.0038
10	24	0.0065	1.0000	0.0080	0.9070	-0.0016	0.0061	0.9421	0.0004
	3714	Dmax = 0.0491					Dmax = 0.0154		

á = 1.83
Theoretical Value of c = 0.5438
Fe+ = $0.5438(1/X^{1.83})$
D-Max = 0.0491

á = 2
Theoretical Value of c = 0.6079
Fe+ = $0.6079(1/X^{2.00})$
D-Max = 0.0154

Critical Value at .01 level of significance = 1.83/ √ 3714
= 0.0300

Results and Discussion

Generally Lotka's Law describes the frequency of publications by authors in a given subject/discipline. In the present study, an attempt has been made to study the suitability of the Lotka's Law to the publications of a Higher Educational institution of a region consisting of academic authors and researchers in various disciplines. Annamalai University has been ranked 20th in the state level and 56th at National level by MHRD. Annamalai University is renowned for its research output performance and has been funded for many major, minor and DST research projects. It has a large network and linkages to academia, R&D organisations and industries. It covers almost all the subjects with 10 Faculties and 49 departments of study.

A K-S test is applied for the fitness of Lotka's law in this study. The obtained results show that the Lotka's Law does not fit to the two sets of Annamlai University data. Result indicates that the value of D-Max, i.e, 0.0273 and 0.0491 determined with Lotka's exponent, á=1.82 and á=1.83 for Data Set- 1 and Data Set- 2 respectively are more close to the D-Max value 0.0282 and 0.0154 determined with the Lotka's exponent á=2 than the critical value determined at the 0.01 level of significance, i.e., 0.0611 and 0.0300. Thus, distribution frequency of the authorship doesn't follow the exact Lotka's Inverse Law with

the exponent á=2. However, the modified form of the inverse square law, i.e., Inverse Power Law with á and c parameters as 1. 82 and 0.5397 for Data Set- 1 and 1.83 and 0.5438 for Data Set- 2, are applicable and appears to provide a good fit.

Conclusion

The results obtained in this study do not follow the inverse square power law of Lotka as such and similarly. Gupta[18] and Narendra[19] have also described in his studies based on CSIR samples that Lotka's formulation is not applicable in case of CSIR productivity distribution. It may be due to longer period of participation in research. Regarding the validity of Lotka's Law, Nicholls[20] has analyzed about 15 studies conducted during 1973 and 1986, and observed that the result of these studies are conflicting and, in brief they do not provide the clear-cut validation of the law. Further, the works of Huber[21-22] describe the statistical problems related to the fitting of the Lotka's formulation and related regularities. Recently, the work of Batcha[23] analyzed about the publications on Robotic Technology is not suitable the Lotka's formulation.

References

1. Lotka, A. J. The frequency distribution of scientific productivity, *Journal of Washington Academy of Sciences*, 1926, **16,** 317-323.
2. Bradford S.C. Documentation. D.C, Public Affairs Press, Washington, 1950.
3. Zipf, G. Human Behaviour and the principle of Least Effort. Addison-Wesley, Cambridge, (MA), 1949.
4. Fairthorne, R. A. Progress in documentation: Empirical hyperbolic distributions (Bradford-Zipf-Mandelbrot) for Bibliometric description and prediction, *Journal of Documentation*, 1969, **25**, 319.
5. Op. Cit. Lotka
6. Murphy, L.J. Lotka's Law in the Humanities, *Journal of American Society for Information Science*, 1973, **24**(6), 461-462.
7. Schorr, A.E. Lotka's Law and map librarianship, *Journal of American Society for Information Science*, 1975, **26** (2), 189-190.
8. Pao, M.L. An empirical examination of Lotka's Law, *Journal of American Society for Information Science*, 1986, **37**, 26-33.
9. Bookstein, A. Pattern of scientific productivity and social change: A discussion of Lotka's Law and bibliometric symmetry, *Journal of American Society for Information Science*, 1977 **28**(4), 206-210.
10. Op. Cit. Bookstein, A.

11. Op. Cit. Pao.

12. Lemoline, W. The frequency distribution of research papers and patents according to sex: The cases of CSIR, India, *Scientometrics*, 1992, **24**(3), 449-469.

13. Radhakrishnan, T. & Kernizen, R. Lotka's Law and computer science literature, *Journal of American Society for Information Science*, 1979, **30**, 51-54.

14. Narendra Kumar. Applicability to Lotka's law to research productivity of Council of Scientific and Industrial Research (CSIR), India, *,Annals of Library and Information Studies*, 2010, **57,** 1-5.

15. Amsaveni, N. & Batcha, Sadik.M. Bibliometric Dimension of Gender Studies in Informatics from G-8 Countries, *Indian Journal of Information Science and Services*, 2009, **1**(3), 65-71.

16. Batcha, Sadik. M. Research Output Analysis on Robotic Technology: A Scientometric Study, *Indian Journal of Information Sources and Services*, 2017, **7** (1), 25-31.

17. Pao, M. L. Lotka's law: A testing procedure, *Information Processing and Management*, 1985, **21**(4), 305-320.

18. Gupta, B.M., Kumar, S., & Aggarwal, B.S. A comparison of productivity of male and female scientists of CSIR. *Scientometics*, 1999, **45** (2), 269-289.

19. Op. Cit.Narendra Kumar….

20. Nicholls, P.T. Bibliometric modeling processes and the empirical validity of Lotka's law. *J. American Soc. Inf. Sci.*, 1989, **40**(6), 379-85.

21. Huber, J.C. A new model for analysing scientific productivity, *Journal of the American Society for Information Science and Technology*, 2001, **52**, 1089–1099.

22. Huber, J.C. A new model that generates Lotka's law, *Journal of the American Society for Information Science and Technology*, 2002, **53**, 209–219.

23. Op. Cit.Batcha, Sadik.M.….