Winter 12-12-2020

Author Productivity and the Application of Lotka's Law in the Field of Horticulture

SANTOSH KUMAR TUNGA
RISHI BANKIM CHANDRA EVENING COLLEGE, WEST BENGAL, INDIA, tungask@rediffmail.com

AUTHOR PRODUCTIVITY AND THE APPLICATION OF LOTKA'S LAW IN THE FIELD OF HORTICULTURE

Dr. Santosh Kumar Tunga
Librarian (Sl. Grade), Rishi Bankim Chandra Evening College
Naihati-743165, North 24 Parganas, West Bengal, India
E-mail: tungask@rediffmail.com

Abstract

Citation study of 10,845 citations appended to 80 doctoral dissertations in the field of horticulture awarded by Bidhan Chandra Krishi Viswavidyalaya (BCKV), Mohanpur and Uttar Banga Krishi Viswavidyalaya (UBKV), Cooch Bihar, West Bengal has been carried out to determine the authorship pattern and productivity to cited articles during 1991-2010. The study revealed that researchers are mainly used journal articles 8437 (77.796%). Generally Loka's law describes the frequency of publications by authors in a given subject/ discipline. In this paper, an attempt has been made to study the applicability of the Lotka's law to the publications of horticulture scientists in BCKV and UBKV. A Kolmogorov-Smirnv (K-S) test has been conducted to find out as to what extent, the author productivity conforms to the Lotka's law. But this test is applied for the fitness of Lotka's law does not fit to the horticulture literature.

Keywords

Author Productivity, Doctoral Dissertations, Horticulture, Journal Articles, Lotka's Law

1. Introduction

One of the main areas in bibliometric research concerns the application of bibliometric laws. The three commonly used laws in bibliometrics are: Lotka's law of scientific productivity, Bradford's law of scatter, and Zipf's law of word occurrence. Bibliometric laws are statistical expression by which seek to describe the working of science by mathematical means.

Lotka's law is considered as the earliest and most widely applied study in measuring the scientific productivity of an author. A. J. Lotka claims that a large proportion of the literature is produced by a small number of authors and it is distributed so as the number of people producing 'n' paper is approximately proportional to $1/n^2$. Lotka in his classic paper published on frequency distribution of scientific productivity presented an analysis of the number of Publications listed in Chemical Abstracts from 1907 to 1916 with the frequency of publications of the authors and proposed an inverse square law of scientific productivity.

The thrust of this paper is employing the authorship pattern and author productivity and test has been conducted to find out the applicability to Lotka's law to author productivity of both universities.

2. Objectives of the Study

- To find out the bibliographic forms used by the researchers
- To identify the authorship pattern of researchers
- To ascertain the author productivity of researchers
- To find out as to what extent, the author productivity conforms to the lotka's law

3. Previous Studies

According to the extent papers, numerous authors attempted to apply the Lotka's law to different fields. It was applied the Lotka's law, at the humanities field, concluding that the law came true, not applying any statistics test to check the significance degree (Murphy, 1972).

He published three articles in which he presented other Lotka's law applications: the librarianship and the libraries organization. He found a law which was quadruple (x -4), instead of the inverse quadratic Lotka (x-2), in other experience about libraries organization (Schorr, 1974). The authors' productivity studied at the information science field, between 1966 and 1970, and compared the results with the Lotka's observation (n=2), and discovered that the distribution of authors adjusted itself so well to a new constant equals to x-3.5. The percentage of authors with only one work, obtained by Voos, was 88 percent, instead of the 60 percent obtained by Lotka (Voos, 1974).

The productivity studied in the history of legal medicine and applying the test x2, he discovers that the authors with multiple works were very below from the expected according to the Lotka's law (< 60%), concluding that this law was not the most appropriate to this subject (Schorr, 1975). After applying the test x2 he concluded that this discipline adjusted itself to the Lotka's law (Schorr, 1975).

After presenting the Lotka's law, extracted from the original work, examined and checked the data from the article of Murphy, in humanities, and the ones of score, in libraries organization, using the test of Kolmogorov-Smirnov (K-S), concluding that in no one of the cases it was accomplished the Lotka's law (Coile, 1977).

A bibliography presented about Lotka and related work, among them about Bradford and Zipf, as well as distribution of frequencies and of bibliometrics (Vlachy, 1978). In previous work he had found discrepancies among the empiric data and the inverse square law, that is, the exponent value of Lotka's law was variable (Vlachy, 1976).

The law of Lotka did not apply appropriately to the data about publications in informatics, observing that it was nearest to a law x-3. These authors assumed that when a work had many authors, to each one of them belonged the complete work (normal count). To prove this hypothesis, it was examined sample from this field, registering only one author for each work, and without applying any statistics test, concluded that the data adjusted themselves to the law of Lotka (Radhakrishnan and Kerdizen, 1979).

At a study about entomology of Nigeria, analyzed and studied productivity models of authors and checked the applicability of Lotka's law to four different groups of data. It is showed that Lotka's law, on its original shape, as inverse quadratic was not applicable to any of the four groups of data (Gupta, 1987).

Four different FICHEIROS created in the biochemical field from Nigeria, one with all the authors, another with only the first ones, with the non collaborators and one fourth only with the co-authors, it was checked that the Lotka's law could be applied at the four cases, but with distinct values at the exponent. To check the adjustment it was used the test Kolmogorov-Smirnov, to a significant level of 0.01 (Gupta, 1989). The Lotka's law is applied and its application to author productive distribution of psychological literature of Africa for the period of 1966-1975(Gupta, 1989).

It tried to estimate the lotka's law at the domain of information science and checked that it is applicable to this field (Sen et al., 1996).

The authors' productivity analyzed at the field of librarianship and documentation in Spain, concluded that the Lotka's law described fairly well the data distribution (Jimenez and Anegon,1997).

It was presented a detailed analysis of research performance of biotechnology faculties in Central Universities of India from 1997-2006. The results indicated that the growth of literature in biotechnology has steadily increased from 15 articles in 1997 to 43 articles in 2006; two-authored publications predominate amongst the pattern of authorship; applicability of Lotka's law is validated from the values n=2.12, C=0.669, and D=0.027 obtained using least square method (Sevukan and Sharma, 2008).

An application of Lotka's law introduced at whole of authors with publication in the field of information science between 1996 and 2007. The results showed the data: one pending equal a '2,75', the obtained it is lower in the work of Voos (1974), as in the Sen et al. (1996), in this camp; a percentage of authors, executors of one work only, it is equal a 79 percent and a excellent adjust of the Lotka's law, to be application at the Kolmogorov-smirnov (Sobrino et al., 2008).

The applicability of Lotka's law examined as a general inverse power ($\alpha \neq 2$) and as an inverse square power relationship ($\alpha = 2$) to the distribution of the research productivity in CSIR, India. Two data sets of the research papers (6076 and 17681) contributed by CSIR's scientists during the period of 1988-1992 and 2004-2008were collected from SCI CD-ROM and Web of science respectively. AK-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 α=2. It was found that the inverse square law of Lotka did not confirm as such (Kumar, 2010).

It was provided an insight into the citation analysis of research publications of the National Metallurgical Laboratory (NML) during the period 1972-2007. To validate Lotka's law, they showed log-log plot of number of authors and number of citations in figure and concluded the low and medium productive cited authors are not a good fit but the high productive authors can be said as a good fit the original Lotka's law (Mishra, 2010).

The meaning of author productivity and research productivity discussed and showed the difference between the two. He demonstrated that how simply the value of c and a pertaining to the equation of lotka's law can be calculated. The value of a obtained according to the method described in the paper seems to be equally good, if not better than the value obtained through Pao's method. The method is much simpler compared to Pao's method (Sen, 2010).

It was studied e-commerce technology trends and forecasts using bibliometric analysis from 1989 to 2009 in SSCI database. The paper performed K-S test to verify the reliability of Lotka's law. After checked by K-S test, the distribution of frequency indexes of author productivity is suitable for lotka's law (Tsai and Chiang, 2011).

4. Methodology

Dissertations awarded in the field of horticulture at the BCKV and UBKV were examined during 1991-2010. Title pages and reference sections were photocopied from each of the 80 dissertations. Information extracted from each dissertation for determining the bibliographic forms, authorship pattern and productive pattern of authors of cited journal articles. Author

productivity is reflected by the respective number of published articles in journals. The data on author productivity provide the basis for the application of Lotka's law. Only personal authors were considered for analysis. Authors were given full credit for every publication in which his or her name appears. The number of authors contributing one, two, or more articles each was counted manually.

5. Lotka's Law

It was proposed an inverse square law relating to scientific papers to the number of contributions made by each author. Lotka's law describes the frequency of publication by authors in a given field. It states that the number of authors making n contributions is about $1/n^a$ of those making one contribution; and the proportion of all contributors, that make a single contribution, is about 60 percent. This means that out of all the authors in a given field, 60 percent will have just one publication, and 15 percent will have two publications ($1/2^2$ times× 60), 7 percent of authors will have three publications ($1/3^2$ times × 60), and so on. This law can be expressed as:

$$X^n \times y = c, \quad \text{or} \quad y = c/x^n, \quad \text{or} \quad y = c \times x^n \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

Where, x is the number of publications of interest (1,2, etc); n is an exponent that is constant for a given set of data; y is the expected percentage of authors with frequency x of publications, and c is a constant (Lotka, 1926).

This study followed the recommendation conducted analysis which calculated the slope n value and the constant c value by using the Kolmogorov-Smirnv (K-S) examination to confirm whether the horticulture literatures consistent with the Lotka's law or not. It can be expressed as (Pao, 1986):

$$n = (N\sum XY - \sum X \sum Y) / (N\sum X^2 - (\sum X)^2) \text{ ....(2)}$$

where, N is the number of data pairs considered; X is the logarithm of x (x is the number of articles) and Y is the logarithm of y (y is the number of authors). Then get the constant c value by following equation:

$$c = 1/ \sum_1^{p-1} 1/X^n + 1/(n-1) (p^{n-1}) + 1/2p^n + \ldots\ldots\ldots + n/24(p-1)^{n-1}$$

$$\text{or, } c = 1/ \sum(1/x^n) \text{...(3)}$$

To verify that the observed distribution of author productivity fits the estimated distribution, he suggested applying the non-parametric Kolmolgorov-Smirnov (K-S) goodness-of-fit test. To this end, the maximum difference between the real and estimated accumulated frequencies was calculated, and this value was then compared with the critical value (c.v.) obtained from the following equation:

$$\text{c.v.} = 1.63/ \{ \sum y_x + (\sum y_x/10) \}^{1/2} \text{...(4)}$$

D = Dmax = Differences between the columns of the observed and expected cumulative frequencies

$$= \sum f(x) - \sum(y_x/\sum y_x)$$

6. Data Analysis and Interpretation

6.1. Bibliographic Forms Used by the Researchers

The following table presents data on different types of documents cited by the researchers in their doctoral dissertations.

Table 1: Form wise Distribution of Bibliographic Forms

Sl. No.	Bibliographic Forms	No of Citations	% of Citations	Cum. Citations	% of Cum. Citations
1	Journal Articles	8437	77.796	8437	77.796
2	Books	1327	12.236	9764	90.032
3	Conference Proceedings	512	4.721	10276	94.753
4	Theses /Dissertations	158	1.458	10434	96.211
5	Bulletins	122	1.125	10556	97.336
6	Reports	120	1.107	10676	98.443
7	Yearbooks	42	0.387	10718	98.830
8	News Letters	39	0.360	10757	99.190
9	Web Resources	16	0.147	10773	99.337
10	Others	72	0.633	10845	100.000
	Total	10845	100.000	10845	100.000

The analysis of data in table 1 shows that both journal articles (77.796%) and books (12.236%) dominated the list as source of information for researchers in Horticulture. Together, both journal articles and books constituted nearly 90 percent of the total items cited. Conference Proceedings and Theses & Dissertations occupied the third and fourth places with 4.721 percent and 1.458 percent respectively.

Citations to bibliographic forms that are accounted for less than 0.100 percent are grouped under ‘Others’ category. This category includes citation to Course Materials, Manuals, Leaflets,

Working Papers, Abstracts, Magazines, Reviews, Souvenir, Pamphlets, Patents, and Standards. This category constitutes only 0.047 percent of the total citations cited by the researchers.

6.2. Authorship Patterns of Articles

Table 2 gives the distribution of the cited articles with respect to the number of authors. Out of 8437 references cited 1763 (20.695%) are single authored journals.

Table 2: Year wise distribution of articles

No of Authors / Year	1	2	3	4	5	6	7	8	9 & <	An.	No of Articles
1991	154	277	179	78	17	7	2	1	2	8	725
1992	58	116	42	15	3	-	-	-	-	-	234
1993	80	94	73	18	4	-	1	-	1	-	271
1994	27	44	35	11	1	1	-	-	-	-	119
1995	66	93	50	9	2	-	-	-	-	-	220
1996	24	32	12	2	-	-	-	-	-	-	70
1997	65	91	71	21	10	3	1	-	-	-	262
1998	30	28	4	2	4	-	-	-	-	-	68
1999	25	62	44	11	5	1	-	1	-	-	149

2000	146	274	175	59	9	6	2	4	1	-	676
2001	118	146	108	43	13	2	-	-	1	-	431
2002	145	267	163	75	11	10	-	2	1	-	674
2003	47	86	52	14	7	6	2	1	2	-	217
2004	35	54	37	28	4	3	-	1	-	-	162
2005	200	348	224	108	36	16	3	7	5	1	948
2006	101	271	204	104	32	4	4	1	1	-	722
2007	147	226	190	101	30	9	3	1	3	-	710
2008	125	234	159	88	29	12	5	4	3	-	659
2009	109	230	181	95	30	17	7	1	1	-	671
2010	61	152	116	74	26	11	7	1	1	-	449
Total	1763	3125	2119	956	273	108	37	25	22	9	8437

An.= Anonymous, D= Dissertations

6691(79.305%) are multi-authored journals. Citations to single author contributions are more in number in the year 2005 with 200 citations followed by 154 citations in 1991 and 147 citations in 2007. The lowest citations to single author publications are 27 in 1994. Among the multi-author articles, the share of two author contributions is found to be more i.e. 3125 citations (37.039%), followed by 2119 citations (25.116%) of three author contributions and 956 citations

(11.331%) of four author contributions. The study reveals that team research is on the increase in the field of Horticulture.

6.3. Author Productivity

The study has analyzed the citations by number of authors to assess the pattern of authorship in the literature of horticulture. It is clear that two authors (37.079%) are the highest in the cited journals followed by three authors (25.142%), and single author (20.917%).

Table 3: Distribution of author productivity

No. of Publication (x)	No. of Authors (y)	% y	xy	∑xy	% of ∑xy	∑y	% of ∑y
15	1	0.012	15	15	0.072	1	0.012
14	3	0.036	42	57	0.273	4	0.048
13	1	0.012	13	70	0.335	5	0.060
12	4	0.047	48	118	0.564	9	0.107
11	2	0.024	22	140	0.670	11	0.131
10	10	0.119	100	240	1.147	21	0.250
9	1	0.012	9	249	1.191	22	0.262
8	25	0.297	200	449	2.147	47	0.559
7	37	0.439	259	708	3.385	84	0.998
6	108	1.281	648	1356	6.483	192	2.279
5	273	3.240	1365	2721	13.010	465	5.519

4	956	11.343	3824	6545	31.293	1421	16.862
3	2119	25.142	6357	12902	61.688	3540	42.004
2	3125	37.079	6250	19152	91.571	6665	79.083
1	1763	20.917	1763	20915	100.000	8428	100.000
Total	8428		20915	20915	100.000	8428	100.000

From Table 2 it is also observed that would be a total of 20915 articles with 8428 authors with an average of 0.403 authors for each article. The result indicates that the literatures of data were usually generated by multi-authors.

6.4. Calculation of the Exponent 'n' for Horticulture

Table 4: Calculation for the exponent 'n'

No. of Publication (x)	No. of Authors (y)	X= Log x	Y= Log y	XY	X^2
15	1	1.176	0.000	0.000	1.383
14	3	1.146	0.477	0.547	1.314
13	1	1.114	0.000	0.000	1.241
12	4	1.079	0.602	0.650	1.165
11	2	1.041	0.301	0.313	1.084
10	10	1.000	1.000	1.000	1.000
9	1	0.954	0.000	0.000	0.911

8	25	0.903	1.398	1.262	0.816
7	37	0.845	1.568	1.325	0.714
6	108	0.778	2.033	1.602	0.606
5	273	0.669	2.436	1.703	0.489
4	956	0.602	2.980	1.793	0.362
3	2119	0.477	3.326	1.586	0.228
2	3125	0.301	3.125	1.052	0.091
1	1763	0.000	3.246	0.000	0.000
Total	8428	12.115	22.492	12.833	11.404

By the result of calculation on Table 4, it could bring into the equation of the Lotka's law as below to calculate n value:

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

$$= (15\times12.833 - 12.115\times22.492)/ \{15\times11.404 - (12.115)^2\} = -3.294$$

6.5. Application of Lotka's Law

Table 5: The K-S test for horticulture

No. of Publications (x)	No. of Authors (y)	$y_x/\sum y_x$	$\sum (y_x/\sum y_x)$ (Observed)	$1/x^n$	$F_x = C(1/x^n)$	$\sum f_x$ (Expected)	D=
1	1763	0.209	0.209	1.000	0.896	0.896	0.687*

2	3125	0.371	0.580	0.085	0.076	0.972	0.392
3	2119	0.251	0.831	0.020	0.018	0.990	0.159
4	956	0.114	0.945	0.007	0.006	0.996	0.051
5	273	0.032	0.977	0.003	0.002	0.998	0.021
6	108	0.013	0.990	0.001	0.001	0.999	0.009
7	37	0.004	0.994	-	-	-	-
8	25	0.003	0.997	-	-	-	-
9	1	0.000	0.997	-	-	-	-
10	10	0.001	0.998	-	-	-	-
11	2	0.000	0.998	-	-	-	-
12	4	0.001	0.999	-	-	-	-
13	1	0.000	0.999	-	-	-	-
14	3	0.001	1.000	-	-	-	-
15	1	0.000	1.000	-	-	-	-
Total	8428	1.000	1.000	1.116	-	-	-

Value c is calculated by using following formula:

$$c = 1/\sum_1^{p-1} 1/X^n + 1/(n-1)\,(p^{n-1}) + 1/2p^n + \ldots\ldots + n/24(p-1)^{n-1}$$

$$\text{or, } c = 1/\sum(1/x^n) = 1/1.116 = 0.896$$

when we get n = -3.294 and c = 0.896, it explored:

$$f(x) = c\,(1/x^n) = 0.896\,(1/x^{-3.294}) = 0.896 \times x^{3.294}$$

From Table 5, we can find D ($D = \max| \sum f(x) - \sum(y_x/\sum y_x) | = 0.687$

According to K-S test, the critical value (c.v.) is

$$c.v. = 1.63/ \sqrt{\{\sum y_x + \sqrt{(\sum y_x/10)}\}} = 0.0178$$

Here, Dmax is greater than the K-S test critical value. Therefore, this result indicated that the distribution of author productivity is not match by the lotka's law. The consequence means the Lotka's law is not suitable for the literature author productivity distribution in horticulture research.

7. Findings

- The horticulture researchers are mainly used journal articles 8437 (77.796%) citations for collecting the required information
- Citations to single author and contributions are more in number in the year 2005 with 200 citations followed by 154 citations in 1991.
- Citations to multi-author contributions are more in number in the year 2005 with 348 citations followed by 277 citations in 1991.
- Two authors (37.039%) are the highest in the cited articles followed by three authors (25.116%) and single author (20.896%)
- The frequency distribution regarding author productivity did not match Lotka's law

8. Conclusion

The results obtained in this study do not follow the inverse square law of Lotka, however it should be stressed that Lotka's inverse square law is a general, theoretical estimate of productivity, and is not a precise statistical measurement (Potter, 1981). Regarding the validity of Lotka's law, it has analyzed about 15 studies conducted during 1973 and 1986, and observed that the result of these studies are conflicting and, in brief they did not provide the clear-cut validation of the law (Nicholls, 1986). It has also described in his studies based on CSIR samples that the inverse square law of Lotka did not conform of CSIR productivity distribution (Kumar, 2010).

9. References

Coile, R. C. (1977), "Lotka's frequency distribution of scientific productivity", *Journal of the American Society for Information Science*, Vol. 28 No. 6, pp. 366-70.

Gupta, D. K. (1987), "Lotka's law and productive patterns of entomological research in Nigeria for the period 1900-1973", S*cientometrics*, Vol. 12 No. 1-2, pp. 33-46.

Gupta, D. K. (1989), "Lotka's law and its application to author productivity distribution of psychological literature of Africa for the period 1966-1975-partII", *Herald of Library Science*, Vol. 38 No. 4, pp. 315-26.

Gupta, D. K. (1989), "Scientometric study of biochemical literature of Nigeria, 1970-1984; application of Lotka's law and the 80/20 rule", *Scientometrics*, Vol. 15 No. 3-4, pp. 171-79.

Jimenez, E. C., & Anegon, F. M. D. (1997), "Analisis de la autoria en revistas espanolas de Biblioteconomia y Documentacion: 1975-19995", *Revista Espanola de documentacion Cientifica*, Vol. 20 No. 3, pp. 252-66.

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

Lotka, A. J. (1926), "The Frequency distribution of scientific productivity", *Journal of Washington Academy of Science*, Vol. 16, pp. 317-23.

Mishra P. N., Panda, K. C. and Goswami, N. G. (2010), "Citation analysis and research impact of National Metallurgical Laboratory, India during 1972-2007: A case study", *Malaysian Journal of Library & Information Science*, Vol. 15 No. 1, pp. 91-113.

Murphy, L. J. (1973), "Lotka's law in the humanities"? *Journal of the American Society for Information Science*, Vol. 24 No. 6, pp. 461-62.

Nicholls, P. T. (1986), "Bibliometric modeling processes and empirical validity of Lotka's law", *Journal of American Society for Information Science*, Vol. 40 No. 6, pp. 379-85.

Pao, M. L. (1986), "An Empirical examination of lotka's law", *Journal American Society for Information Science*, Vol. 37, pp. 26-33.

Potter, W. G. (1981), "Lotka's law revisited", Library Trends, Vol. 30 No. 1, pp. 21-39.

Radhakrishnan, T. and Kerdizan, R. (1979), "Lotka's law and computer science literature", *Journal of the American Society for Information Science*, Vol. 30 No. 1, pp. 51-54.

Schorr, A. E. (1974), "Lotka's law and library science", *Reference Quarterly*, Vol. 14 No. 1, pp. 32-33.

Schorr, A. E. (1975a), "Lotka's law and the history of legal medicine", *Research in Librarianship*, 30(5): 205-209.

Schorr, A.E. (1975b), "Lotka's law and map librarianship", *Journal of the American Society for Information Science*, Vol. 26No. 3, pp.189-90.

Sen, B. K. (2010) "Lotka's law: a viewpoint", *Annals of library and Information Studies*, Vol. 57, pp.166-67.

Sen, B. K., Taib, C. A. and Hassan, M. F. (1996), "Library and information science literature and Lotka's law", *Malaysian Journal of Library & Information Science*, Vol. 1 No. 2, pp. 89-93.

Sevukan, R. and Sharma, J. (2008), "Bibliometric analysis of research output of biotechnology faculties in some Indian Central Universities", *DESIDOC Journal of Library & Information Technology*, Vol. 28 No.6, pp. 11-20.

Sobrino, M. I. M., Caldes, A. I. P. and Guerrero, A. P. (2008), "Lotka law applied to the scientific production of information science area", *Brazilian Journal of Information Science*, Vol. 2 No. 1, pp. 16-30.

Tsai, Hsu-Hao and Chiang, Johannes K. (2011), "E-Commerce research trend forecasting: A study of bibliometric methodology", *International Journal of Digital content tech and its applications*, Vol. 5 No. 1, pp. 101-10.

Vlachy, J. (1976), "Time factor in Lotka's law", *Probleme de Informare si documentare*, Vol. 10 No. 2, pp. 44-87.

Vlachy, J. (1978), "Frequency distribution of scientific performance: A bibliography of Lotka's law and related phenomena", *Scientometrics*, Vol. 1 No. 1, pp. 109-130.

Voos, H. (1974), "Lotka and information science", *Journal of the American Society for Information Science*, Vol. 25 No. 4, pp. 270-72.