University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal) | Libraries at University of Nebraska-Lincoln

Winter 9-6-2013

LOKTA'S LAW REVISITED IN TOXICOLOGY LITERATURE

Mini Devi B
University of Kerala, drminidevi1968@gmail.com

LOKTA'S LAW REVISITED IN TOXICOLOGY LITERATURE

Dr. B. Mini Devi
Assistant Professor
Department of Library & Information Science
University of Kerala
Thiruvananthapuram-695034
Email: minidevi_b@yahoo.co.in

Abstract

The article highlights the significance of Lotka's Inverse Square Law of Scientific Productivity in today's information age and later studies. The classical law is applied to Toxicology literature collected from the international database, Toxicology Information Online (TOXLINE), and its validity on that data was tested. The data was found unfit for the law Hence a new formula is derived, $f\left(y^2\right) \approx \frac{A}{y^2 + (y-1)} = \beta$, was found fit for the study.

Keywords - Lotka's law, Scientific Productivity, Scientometrics, Toxicology, Toxicology Information Online (TOXLINE)

1. Introduction

In every subject we find some similarities while counting and tabulating readings is known as models, measures or laws. Informetrics deals with the study of library and information dissemination processes by using quantitative treatment of the properties and behaviour of knowledge. From these studies valid laws and theoretical formulations are discovered. The three fundamental classic laws which laid the foundations of informetrics are:

1. Lotka's Inverse Square Law of Scientific Productivity (based on Author Productivity in terms of papers published);

2. Bradford's Law of Scattering of Scientific Papers (based on the distribution of articles over various journals); and

3. Zipf's Law of Word Occurrence (based on ranking of word frequency in a text).

In 1926, Alfred J. Lotka proposed his Inverse Square Law correlating contribution of scientific papers to their number of contributions. The law provided fundamental theoretical base for bibliometric studies involving authorships. He was interested in determining "the part which men of different calibre contribute to the progress of Science" (Lotka 1926). This became known as the Inverse Square Law of Scientific Productivity. For this, he used the decennial index of 'Chemical Abstracts' 1907 - 1916 and counted the number of names against which appeared 1, 2, 3 etc. Then tabulated the data for 6, 8901 names, beginning with letter 'A' and 'B'. In the same way data about 1325 physicists are taken from the Auerbach's Geschietftafeln der physih. Lotka then plotted the graph on a logarithmic scale of authors against the number of contributions made by each author and he found that in each case the points were closely scattered about a straight line, having a slope of approximately two to one. From these data, Lotka deduced a general equation, for the relation between the frequency distribution 'y' of persons making 'x' contributions as

X^n y = constant and for the special case n = 2, the constant is 0.6079. Lotka explained the phenomenon as follows:-

"In the case examined it is found that the number of persons making two contributions is about one-fourth of those making one contribution, the number making 'n' contributions is about $1/n^2$ of those making one and the proportion of all contributions is about 60 percent".

The law was termed Lotka's law in 1949 and attracted the attention of researchers but its applicability to other disciplines was tested only in 1973 in Humanities and found it was fit to this field(Murphy 1973). In 1974 the law was applied in the field of Information Science and found that "the relationship in this field is $\frac{1}{n^{3.5}}$ instead of Lotka's $\frac{1}{n^2}$ "(Voos 1974). In the same year the law was applied to Library Science and showed that Lotka's law does not apply to the field of Library Science (Schorr 1974). In 1975 Schorr studied map librarianship and concluded that the law fits in this field. But later found that Schorr's calculation was wrong as the law did not fit to map librarianship (Coile 1976).

Several studies have assumed the inverse square relation as the basis for testing, and derived the value of constant 'c' form the percentage of single paper contributors which cannot be traced back to Lotka's assumptions (Pao 1982) . "Therefore, a uniform method should be agreed upon by those attempting a test. Comparison and generalisation on author productivity may be possible only if compatible data are available and results are significant".

Many analytical approaches different from Lotka's law for scientific productivity was found. "Scientific talent is highly concentrated in a limited number of individuals"((Narin 1976). There existed a close correlation between quantity of scientific publication and achievement of eminence. The number of elite in science is small compared to total number of scientists and an elite mean is an eminent scientist producing scholarly writing. Later a theoretical model which is a generalised version of Lotka's law, f (x) = $\frac{k}{x^{\infty}}$ where 'k' and α are constants. According to this, the number of authors with 'x' papers is proportional to $\frac{1}{x^{\alpha}}$ (Bookstein 1976).

Pao, Nichollas and Griffith used the version of Lotka's Law by Bookstein, and estimated that the values of 'n' rather than using n = 2. They counted authors and suggested a goodness - of - fit test for the model. Nichollas (1987) found that the generalised version is "surprisingly well fitting and stable" whereas Pao suggested "overwhelming conformity" to this model.

Price (1963) (1971) found that "Half of the scientific papers are contributed by the square root of the total number of scientific authors". This empirical law is later known as Price's Square Root Law of Scientific Productivity. In other words, $N^{½}$ sources yield a fraction ½ of the items and are associated with invisible colleges. This law is sometimes called 'Rousseau's law' since Jean Jacques Rousseau has mentioned the same thing quite clearly in his "Social Contract" about the size of the elite. This law was proved to be invalid both theoretically and empirically by Nicholls (1988), Egghe and Rousseau (1990). This can also be treated as an extension of the success - breeds - success principle originally developed by Simon in 1955.

The problem of crediting authorship to multi-authored paper occurs while applying Lotka's Law. Lotka counted only first author in multi-authored paper. Bookstein (1990) discusses and concluded that "if Lotka's law holds for one accounting method, it will hold for any other one in which the change in the typical amount of credit given to authors per paper may vary from author to author but does not depend strongly on how much the authors published. If this is true, the investigator can give any reasonable system of assigning credit to authors while studying author productivity".

In author productivity studies it is found that the number of single paper producers is more. It is also found that authors who are more productive are having more collaborative studies than single paper producers. Because of the multidisciplinary nature of research topics, there is more scope for multi-authored paper than a single authored paper. Lawani (1980) has shown that "citation rate and quality of paper both correlate positively with the number of authors per paper". In addition to collaboration, individual productivity is affected by working environment, motivation, record system etc.

Lotka's proposition led to a whole gamut of studies on scientific productivity. Such studies conducted during post-second world war period have cultivated in the rise of a new discipline called 'Scientometrics'. Scientometrics is defined as the study of the measurement of Scientific and Technological progress.

Three decades back Yuasa (1962), in a statistical study of scientific achievement in various countries showed that there is a shifting of the world scientific dominance from one country to another. He found out that his dominance shifted from Italy to Britain, then to France, from France to Germany and finally to USA in the 20^{th} century.

Price, who had traced the development of Science since Babylon and plotted the growth of big science from little science had observed that Lotka's law applied equally well to the productivity of scientists in the 17^{th} as well as in the 20^{th} century i.e. the majority of publications emanated from a handful of people. Narin (1976) showed that scientific talent was highly concentrated in a limited number of individuals.

Newby (2003) applies Lotka's law to metadata on open source software development. Lotka's law predicts the proportion of authors at different levels of productivity. Authoring patterns found are comparable to prior studies of Lotka's law for scientific and scholarly publishing, Lotka's law was found to be effective in understanding software development productivity patterns, and offer promise in predicting aggregate behaviour of open source deveoplers. Pao (1985) presents an evaluative framework for comparison of authorship data with Lotka's Law's predictions. Pao suggests the Kolomgrov -Smirnov (K-S), one - sample goodness of fit test for evaluate the statistical significance of results.

2. Relevance of the Study

People are exposed to a great variety of natural and man-made substances. Under certain conditions such exposures cause adverse health effects. These effects range from death to subtle biologic changes. Society's ever - increasing desire to identify and prevent these effects has prompted the dramatic evolution of Toxicology as a study of poisons to the present day complex science. The expansion of the various facets of Toxicology has been outcome of the need of an affluent society to protect itself from harmful chemicals, physical agents, and various industrial and consumer products. The need for Toxicology information on unlimited number of chemicals has had a profound effect upon the development of the Science and profession of Toxicology. Research in Toxicology is carried out in universities, in government and private research laboratories and in certain industrial laboratories. Today Toxicology research is increasingly being focussed on medical, environmental and industrial division as people all around the world are more alert and aware about how widespread the toxins and more particularly over the last century. Many of the themes that are attracting widespread attention and interest are desertification, acid deposition, stratospheric ozone depletion, climatic changes, industrial wastes; drugs are of vital importance to the future of the planet and its people. As the scale of interest of Toxicology research has broadened i.e. from local and regional problems towards global problems, approaches have also progressed from subject-specific disciplinary emphasis towards increasingly multidisciplinary and interdisciplinary research programmes. Increase in research activities results increase in literature. To select relevant literature, the application of scientific techniques is essential. Informetric studies are the widely accepted methods, which enables meticulous selection of literature.

Lack of informetric studies in the field of Toxicology is a major disadvantage pointed out by researchers. Toxicology is a transdisciplinary field which is not only related to traditional subjects like Medicine, Chemistry, Biology, Pharmacology but also to newly emerging subjects like Biotechnology, Environmental Sciences, Food Sciences etc. Being a transdisciplinary subject the results based on Toxicology research may be coming out in a wide variety of documents. Therefore an informetric study of Toxicology literature is an effective tool that can be successfully and wisely used in any library attached to an organization specializing in Toxicology research.

3. Objectives

The main objective of the study is to test the validity of lotka's law in Toxicology literature

4. Methodology

In order to get an idea about similar studies done in informetrics, an exhaustive literature search was carried out. For this many primary periodicals, secondary periodicals like LISA (Library and Information Science Abstract) and its CD-version LISA plus, bibliographies, UGC Infonet E – Journal Consortium, Internet etc were consulted. After collecting the background information, the data from TOXLINE was collected. Collecting, organizing and analysing of data were done on the basis of established informetric methods. The down loaded data was transformed to CDS/ISIS Programme. The data was sorted to prepare tables and figures and informetrically analysed using SPSS, a statistical software programme. Wherever found suitable, the dependence of different variables were tested statistically using formulae in order to prove the validity of hypotheses based on objectives.

4. Analysis &Interpretation

The productivity was measured in terms of the number of times a particular author was cited during 1998 to 2003. Out of the total 9265 citations, minimum number (15) of authors was cited for 10 times and maximum (2935) number of authors were cited only once. From the study it is clear that few authors had been cited more number of times. The number of citations received by the authors is provided in Table 1.

The Lotka's law is applied to author's productivity is presented in the Table 1. From the table it is evident that the observed percentage of authors varied from the expected percentage of authors as predicted by applying Lotka's equation. The Chi-square test was further applied to compare the observed values with the expected value of author's productivity according to Lotka's law. The calculated Chi-square value (228.54) was more than the Table Chi-square value i.e. 18.3, at a degree of freedom of 10, level of significance, □ = 0.05 is shown in the Table 15. Here the Chi-square value was highly significant and Lotka's law was not applicable to this data.

Table 1: Productivity of authors in Toxicology

No. of papers	Observed no. of authors (a_n)	Observed % of authors $\frac{100 \times a_n}{a_1}$	Expected no. of authors $a_n = \frac{a_1}{n^2}$	Expected % of author predicted by Lotka ($100/n^2$)	$\frac{(a_n - p)^2 / p}{p}$
1	2935	100 (71.13)	2935	100	0
2	565	19.25 (13.70)	734	25	38.51
3	218	7.43 (5.28)	326	11.11	35.78
4	122	4.16 (2.96)	183	6.25	20.33
5	70	2.39 (1.70)	117	4.00	18.88
6	62	2.11 (1.50)	82	2.77	4.88

7	30	1.02 (0.73)	60	2.04	15
8	22	0.75 (0.53)	46	1.56	12.52
9	22	0.75 (0.53)	36	1.23	5.44
10	15	0.51 (0.36)	29	1.00	6.76
11	65	2.47 (1.58)	24	0.83	70.04
	4126	100	4572		228.54

Table 2 : Chi-square test on productivity of authors in relation to Lotka's Law

No. of citation s 'n'	**Observed no. of authors (as with 'n' citations (Fi)**	**Expected No. of authors with 'n' citations (Pi)**	**(Fi-Pi)**	**$(Fi\text{-}Pi)^2$**	$\frac{(Fi - Pi)^2}{Pi}$
1	2935	2935	0	0	0
2	565	734	-169	28561	38.91
3	218	326	-108	11664	35.78
4	122	183	-61	3721	20.33
5	70	117	-47	2209	18.88
6	62	82	-20	400	4.88

7	30	60	-30	900	15.00
8	22	46	-24	576	12.52
9	22	36	-14	196	5.44
10	15	29	-14	196	5.44
11	65	24	41	1681	70.04
	4126				X^2=228.54

$X^2 > x_{\alpha}^{1}$

(df. = 10)

$(\alpha = 0.05)$

$((x_{\alpha}^{1} = 18.3)$

Table 3: Author contribution Vs Number of authors

Number of papers	Number of authors	%
1	2935	71.13
2	565	13.69
3	218	5.28
4	122	2.96
5	70	1.70
6	62	1.50
7	30	0.73
8	22	0.53

9	22	0.53
10	15	0.36
11	65	1.58
Total	4126	99.99

The 2935 authors have contributed one paper, 565 authors have two, 218 authors have three, 122 authors have four, 70 authors have five, 62 authors have six, 30 authors have seven, 22 authors have eight, another 22 authors have nine, 15 authors have ten and 65 authors have eleven papers to their credit. According to Lotka's Law,

$$F(y^n) = \frac{A}{y^n} = \alpha$$

Where (F (y^n) stands for the authors productivity, 'y' number of papers, 'A' and '□' are constants, considering the equation i.e.

$$F(y^n) = \frac{A}{y^n}$$

When y = 1

$$F(y^2) = 2935$$

$$2935 = A/1$$

$$A = 2935 \times 1 = 2935$$

Similarly

$$y = 2 \Rightarrow \quad F(y^2) = \frac{2935}{2^2} = \frac{2935}{4} = 734$$

$$y = 3 \Rightarrow \quad F(y^2) = \frac{2935}{3^2} = \frac{2935}{9} = 326$$

$$y = 4 \Rightarrow \quad F(y^2) = \frac{2935}{4^2} = \frac{2935}{16} = 183$$

$$y = 5 \Rightarrow \qquad F(y^2) = \frac{2935}{5^2} = \frac{2935}{25} = 117$$

$$y = 6 \Rightarrow \qquad F(y^2) = \frac{2935}{6^2} = \frac{2935}{36} = 82$$

$$y = 7 \Rightarrow \qquad F(y^2) = \frac{2935}{7^2} = \frac{2935}{49} = 60$$

$$y = 8 \Rightarrow \qquad F(y^2) = \frac{2935}{8^2} = \frac{2935}{64} = 46$$

$$y = 9 \Rightarrow \qquad F(y^2) = \frac{2935}{9^2} = \frac{2935}{81} = 36$$

$$y = 10 \Rightarrow \qquad F(y^2) = \frac{2935}{10^2} = \frac{2935}{100} = 29$$

and $$y = 11 \Rightarrow \qquad F(y^2) = \frac{2935}{11^2} = \frac{2935}{121} = 24$$

Table 4: Values of observed and expected number of authors.

Number of papers	Number of authors (Observed)	No. of authors (expected)
1	2935	2935
2	565	734
3	218	326
4	122	183
5	70	717
6	62	82
7	30	60
8	22	46
9	22	36
10	15	29

11	65	24
Total	4126	4572

From the above table it is evident that the expected values are not close to the observed values up to number of papers. Hence Lotka's Law does not fit for the study.

Thus the law is extended accordingly as shown below.

$$f\left(y^2\right) \approx \frac{A}{y^2 + \left(y-1\right)} = \beta$$

"A" and "β" are constants and "y" is the number of papers.

When the data is applied on the above formula,

When $y = 1 \Rightarrow F(y^2) = \frac{2935}{1^2 + (1-1)} = 2935$

$y = 2 \Rightarrow F(y^2) = \frac{2935}{2^2 + (2-1)} = 587$

$y = 3 \Rightarrow F(y^2) = \frac{2935}{3^2 + (3-1)} = 266$

$y = 4 \Rightarrow F(y^2) = \frac{2935}{4^2 + (4-1)} = 154$

$y = 5 \Rightarrow F(y^2) = \frac{2935}{5^2 + (5-1)} = 101$

$y = 6 \Rightarrow F(y^2) = \frac{2935}{6^2 + (6-1)} = 71$

$y = 7 \Rightarrow F(y^2) = \frac{2935}{7^2 + (7-1)} = 53$

$y = 8 \Rightarrow F(y^2) = \frac{2935}{8^2 + (8-1)} = 41$

$$y = 9 \Rightarrow F(y^2) = \frac{2935}{9^2 + (9-1)} = 33$$

$$y = 10 \Rightarrow F(y^2) = \frac{2935}{10^2 + (10-1)} = 27$$

and $$y = 11 \Rightarrow F(y^2) = \frac{2935}{11^2 + (11-1)} = 22$$

From the above derivation, the expected values are approximately close to the observed values up to number of papers. Hence this law can be fit for the study.

5. Findings & Conclusion

In the case of number of paper of the authors, the expected values were not close to the observed values; hence Lotka's law does not fit for the study.

The law can be extended as

$$f(y^2) \approx \frac{A}{y^2 + (y-1)} = \beta$$

Where 'A' and 'β' are constants and 'y' is a number of papers.

Using benchmarks determined by Lotka's law to establish thresholds for publication activity to guide budgetary decisions is yet another way to equitably reward departments or faculty for their scholarly efforts. Budd and Seavey (1990) acknowledge, this phenomenon, which has come to be referred to as Lotka's law is not intended as an explanation of why some authors are more prolific than others. Because of varying modes of behaviour, patterns of productivity will differ among disciplines . As has been previously noted, the major drawback to previous studies of author publication productivity is that they used varying methods; therefore, their results are not comparable across studies. Using Lotka's law as a departure point allows for some standardization of the procedures used for studies of author publication productivity while allowing researchers to address additional impacting factors. For example, one can examine the characteristics of the more prolific authors in the resulting frequency distribution to discern trends such as gender, geographic location, institution type, etc. The distribution range could also be segmented into tiers (e.g., low, medium, high) and researchers can attempt to discern like characteristics of groups of authors who extend to fall into each tier. Since they would have followed the same basic procedure to achieve the frequency distribution, the studies would be comparable on some level. The hypothesis presented by Narin and Breitzman (1995) that 'scientific and technological creativity and productivity are very highly concentrated in a population and in a relatively small number of highly talented individuals'. Loss of even a few such scientists will substantially erode the scientific productivity and innovation capability of a research organization. The human resource policy of a scientific organization must thus pay a significantly larger attention to this small population of scientists.

References

1. Bookstein, A. (1976). Bibliometric distributions. *Library Quarterly*, 46, 416-423.
2. .Bookstein, A. (1990). Informetric distributions. Pt.II. Resilence to ambiguity. *Journal of American Society for Information Science,* 41, 376-386.
3. Coile, R.C. (1976). Bibliometric studies of scientific productivity. *Proceedings of American Society for Information Science,* 13. 20.
4. Egghe, L. and Rousseau, R. (1990). Introduction to informetrics. Amsterdam: Elsevier, 377.
5. Lawani. S.M. (1980). Quality, collaboration and citation in cancer research. Ph.D. thesis. Florida : State University Press.
6. Lotka, A.J. (1926). The frequency distribution of scientific productivity. *Journal of Washington Academy of Sciences,* 16, 317-325.
7. Murphy, Larry, J. (1973). Lotka's law in the humanities. *Journal of American Society for Information Science*, 24, 461-462.
8. Narin, F. (1976). Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity .New Jersey: Cherry Hill.
9. Narin, F. : Op.Cit.,pp.41-44.
10. Narin, F. and Breitzman, A. (1995). Inventive Productivity. *Research Policy*, 507-519.
11. Newby, Gregory B. (2003). Open source software development and Lotka's law: Bibliometric patterns in programming. *Journal of the American Society for Information Science and Technology,* 54 (2), 169-178.

12. Nicholls, P.T. (1987). Lotka hypothesis and bibliometric methodology. Ph.D. Dissertation. Canada : University of Western Ontario, 175.
13. Nicholls, P.T. (1988). Price's square root law: Empirical validity and relation to Lotka's law. *Information Processing and Management*, 24, 469-477.
14. Pao, H.L. (1982). Lotka's test. Collection Management, 4. 112.
15. Pao, M.L. (1985). Lotka's Law : A testing procedure. *Information Processing and Management,* 21 (4), 305-320.
16. Price, D.J.S. (1963). Little Science, big science. New York: Columbia University Press, 118.
17. Price, D.J.S.(1971). Some remarks on elitism in information and the invisible college phenomenon in science. Journal of American society for Information Science, 22, 74-75.
18. Voos, H. (1974). Lotka and information science. *Journal of American Society for Information Science,* 25. 270-272.
19. Yuasa, M. (1962). Centre of scientific activity, its shift from the 16^{th} to 20^{th} century. *Japanese Studies in the History of Science*, 1, 527.