# Empirical examination of Lotka"s Law for Applied mathematics

Azadeh Soltani Torbati [1], Abdolreza Noroozi Chakoli [2]

[1] Department of library and Information science. University of Mysore , Mysore,India
[2.] The Head of Scientometrics and Knowledge and Information Science Department & Manager of Educational Planning Office, Shahed University, Tehran, Iran

**Abstract:** In this study of testing the fitness of Lotka's law on the literature of applied mathematics Mysore university, the needed data is collected from Web of Science data base 1975-2011. The analysis of the collected data leads to the following findings.
[Azadeh Soltani Torbati, Abdolreza Noroozi Chakoli. **Empirical examination of Lotka"s Law for Applied mathematics**. *Life Sci J* 2013;10(5s):601-607] (ISSN:1097-8135). http://www.lifesciencesite.com. 104

**Keywords:** Lotka"s Law; Scientometrics

## 1. Introduction

Information is a national resource most valuable of all resources since it can control physical resources of every kind. Both developed and developing nations recognize that the scientific and technological information constitute most valuable instrument for economic and social change. The progress in science and technology, the popularization of education in all sections of the society, development in the field of communication technology, the change in research pattern, the competition in trade and industry among various nations, revolutionary movements. World wars etc make this age an 'age of information explosion'. As a result of this, libraries have shown considerable degree of momentum in the collection building and readership programmed activities, involving considerable expenditure of funds. The rapid development of libraries has generated several 'evaluator studies on the usefulness of libraries to society. Until recently evaluations were subjective based purely on opinions expressed by an individual or a group of individuals. However, developments in the field of mathematics, Statistics, Operations Research, management studies economics Systems Analysis have led to a minimisation of this subjectivity and the identification and application of appropriate quantitative techniques for the evaluation

## 2. Definitions

Alan Pritchard, who coined the term Bibliometrics, described it as the "application of mathematical and statistical methods to books and other media of communication". This was proposed by Robert A. Fairthorneas "quantitative treatment of the properties of recorded A discourse and behaviour appertaining ot it". In one of his later works 'Bibliometrics and information transfer Pritchard explained Bibliometrics as the "metrology of the information transfer process and its purpose in analysis and control of the process". The British Standard Glossary of Documentation, of terms explained Bibliometrics as the study of the use of documents and pattern of publication in which mathematical and statistical methods have been applied

## 3. Scientometrics

The term Scientomerics or Naukometrija originated from U.S.S.R. and was practised in East European countries. Perhaps Dobrov and Korennoi^ were the first to coin the term scientometrics. They define it as the measurement of information processes. According to Winkler10 "It is a scientific discipline devoted to all quantitative aspects of science and scientific research". It is used for the quantification of science at individual, institutional national and even international level. Sengupta11 has outlined the objective of scientometrics as "to evaluate quantitatively the recent growth of any basic scientific discipline and the factors responsible for the steady increase in research activity in the area of knowledge".

## 4. Lotka's Law of Author Productivity

Lotka's law, named after Alfred J. Lotka, is one of a variety of special applications of Zipf's law. It describes the frequency of publication by authors in any given field. It states that the number of authors making n contributions is about 1 / na of those making one contribution, where a nearly always equals two. More plainly, the number of authors publishing a certain number of articles is a fixed ratio to the number of authors publishing a single article. As the number of articles published increases, authors producing that many publications become less frequent. There are 1/4 as many authors publishing two articles within a specified time period as there are single-publication authors, 1/9 as many publishing three articles, 1/16 as many publishing four articles, etc. Though the law itself covers many

disciplines, the actual ratios involved (as a function of 'a') are very discipline-specific.
The general formula says:

$$X^nY = C$$

Where X is the number of publications, Y the relative frequency of authors with X publications, and n and C are constants depending on the specific field ( $n \approx 2$ ).

This law is believed to have applications in other fields, for example in the military for fighter pilot kills.

## 5. Studies In The Field

Since the publication of Lotka's original article in 1926, much research has done on author productivity. Many authors like price^ have assumed Lotka's law to have been proved and hence proceeded to discuss why the distribution occurs. But no attempts were made to test the applicability of Lotka's law to other disciplines until Murphy^ in 1973. He tested the applicability of Lotka's law in Humanities. Following his study, many studies are conducted to test its applicability in various fields. But these studies are non-comparable and inconclusive owing to substantial differences in the analytical methods applied. When we study Lotka's law, attention should be drawn to several items of importance. They are:

### *5.1. Counting method:*

To measure the productivity of authors, the number of publication is using as a criterion. But how to count a multiauthored paper is a problem. We use different counting methods such as straight count which counts only the first author, normal count which gives full credit to all the authors and adjusted count which gives a fractional credit to each author. In Lotka's original study, joint contributions have credited only to the first author. But normally6 the names appear in alphabetical order and in the straight count method, the name starts A-F would have advantage over those with G-M and G-M over N-Z. Similarly, senior professionals generally head the authors list. Researchers like Cole and Cole7 and Pao8 are of the opinion that the credit should be given only to the first author Nicholls^ opines that the straight count is quite obviously invalid as an indicator of productivity and the normal (complete) count is presently the most common measure in the modelling of author productivity Here the author who contributes a full paper and the one who contributes only a part of a paper gets equal credit. Some researchers like Lindsay*" argue that because of the increase in multi-authored papers in recent times, it would be erreneous to ignore this factor and give full credit to all the collaborative authors for the same paper and suggest adjusted count. But in this case, an author who contributes 2/3 of the paper may get only 1/2 credit. The researchers like D.K.Gupta-1-* creates different files ie., for the publication of all authors, first author only, only for single authors and for co-authors only and test the fitness of the law for each file separately. Bookstein-1-^ after recognizing the degree of ambiguity in determination of authorship, investigated the problem in detail and concluded that "Lotka's law is not sensitive to how we count articles, so that two people testing the law for a single population, but different count methods, will very likely to come up with the same law". This property is known as robustness property. But Rousseau^ found out that the robustness of Lotka's law breaks down in the case of adjusted count.

### 5.2. *Period Covered*

Another factor which affects the result is period covered in the study. The longer the period, the greater the probability to fit the law. The study conducted on Illinoise library card catalogue-'-'* and Aurenbach figures cover authors from the beginning to the present and in both cases the law fits.

### *5.3. Source of data*

Different types of sources are used to collect data of author productivity. They are:

1. Journal or a group of journals of a definite subject and period.
2. Printed bibliographies and abstracts.
3. Library catalogues.
4. Machine readable catalogue records.
5. Data bases of different disciplines.

The volume of data used for study is also a factor. Lotka used 6819 and 1325 names but other studies used many less except some. Wider the coverage, higher the chance of fitting the law. Murphy took 170 authors ad Schorr 326 authors. Coile*-* applied K-S test to the data of Murphy and Schorr and found out that in both the cases, contrary to the claim of authors, Lotka's law did not apply to their observed data. The same objection can also be applied to many studies.

### *5.4. Community of authors*

Community of authors is another factor which influences the result. There is a universal community1" which can be subdivided by discipline, nation, journal, period, institution etc. All studies are concerned with a subset. The result may change according to the characteristic of each subset. The larger and more representative the subset, the more closely it will resemble the universal community.

*5.5. Interpretation of the law*

The misinterpretation of Lotka's law can lead to the error in the conclusion. In the case of Murphy's study, he

used the actual number of single authors of 130 as the basis for calculating the predicted number of authors with more contribution. He ignored Lotka's statement that the proportion of all contributors that make a single contribution is 60% of total authors and he did not use the total number of authors as his base as Lotka did. Some researchers like Schorr tested Lotka's law in its original form while others determined the value of the constant and tested Lotka's law in its general form.

*5.6. Errors in data*

The investigation of the data used in different studies reveals that the data is not correct and the correction of the data lead to a different conclusion.

5.7. *Goodness of fit test*

In many earlier studies, no statistical tests were used to test the applicability of the law. Some researchers like Schorn used the chi-square test to determine if Lotka's law held. But in X 2 ~ test17 if more than 20% of the expected frequencies are below 5, values in the adjacent categories must be combined. Since in author productivity distribution, there are only a few authors in the higher frequency group, combination is necessary and it reduce the power of the test.

So X -test is not an appropriate test. In order to test the applicability of Lotka's law to a set of data, Coile*® suggested Kolmogorov-Smirnov (K-S) test which was supported by Pao*^ applying the same to her study. Now must of the researchers prefer Kolmogorov-Smirnov test (K-S test) and they consider it as the appropriate test. In this test, the cumulative frequency distribution occurring under the known theoretical distribution is compared with the cumulative observed frequency distribution. It has been suggested that the K-S test is used only when the variable has a continuous expected distribution. If the test is used when the resulting distribution is discontinuous, the error occurring in the resulting probability statement is in the conservating direction. Thus in the event that the null hypothesis is rejected, we can have real confidence that the observed distribution is significantly different from the theoretical distribution. So in all cases where it is applicable, the K-S test is the most powerful test available.

## 6. Procedure for testing

We have seen that different persons use different methods to test the law. But a uniform method should be agreed upon by those attempting such a test. Comparisons and generalizations are possible only if compatible data are available. It points out the need for methodological standardization and co-ordination of research efforts.

Miranda Lee Pao^u suggests a step-by-step procedure for testing the applicability of Lotka's law. With the exception of the test, these procedures are modelled very closely on Lotka's own. This procedure will be explained in detail in the chapter on research design.

Paul Travis Nicholls^-1 proposed two modification to this procedure. Instead of using least square method in the estimation of the slope OC as suggested by Pao, he suggested an alternative approach of maximum likelihood, which provide an estimate with optimal qualities such as unbiasedness, consistency and sufficiently. Secondly the probability measure should take account of all collaborating authors instead of giving credit to only the first author. He is of the opinion that Lotka counted only first authors because multiple authorship was less common at that time and probably it was easier. Today, inquiry in most fields is characterized by extensive and increasing collaboration which is reflected in multiple authorship; measures which are insensitive to this phenomenon are invalid, assuming that we are interested in the distribution of author productivity. The senior author measure cannot be considered to be a sampling strategy either, since the underlying process is probably not random.

## 7. Practical Utility

Lotka's law suggests the relationship between the number of authors and their contributions. His purpose was to estimate the part which men of different calibre contribute to the progress of science. It indicates the pattern of author productivity in a definite field of knowledge. If one considers data for a number of different scientific field corresponding to a given fixed regime, the difference In the estimates of the two fields provides some idea about relative ease of publication between the different fields considered^. The increase of oc is a accompanied by the increase of low productivity scientists. Yoblonsky argues that the larger the parameter <X f the greater is the gap between the productivity of individual group of scientists. In this sense, the oC is considered as a measure of inequality in the distribution of scientific papers. J

The study on library of (congress MARC data^ and Illinois^ Card Catalogue25 were conducted for a practical management problem planning for the implementation of the second edition of the Anglo-American Cataloguing Rules. This is the first known case where Lotka's law has been useful in planning.

## 8. Limitations

Lotka's law is found to fit in most cases. Shockley has pointed out that the power law distribution exists in the case of patents also. However the value of oC was found to vary for different group of scientists. For example: Dobrov and Korennoi found that Lotka's value can be used in measuring productivity in Botany, but not in Electronics. Another problem with Lotka's law is that it totally ignores the potential authors who have not produced any publication so far. Because of these limitations, the empiric nature of these laws is questioned.

**9. Inverse Square law**

The applicability of Lotka's law in its original form as inverse square law on the data set is tested and the results of the analysis are presented in Table 1.

*Step 1: Data collection:*

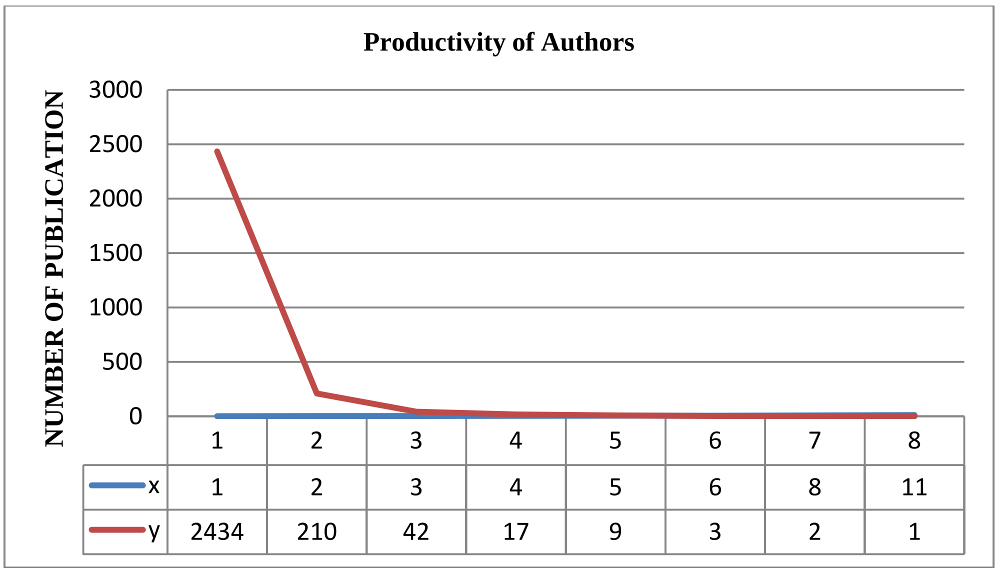
The data on Applied Mathematics was collected from Web of Science database from 1975 to 2011.

*Step 2: Frequency distribution:*

The authors were segregated and number of articles written by each of them have been calculated and stored systematically. Finally the frequency distribution table was developed.

**Table 1: Productivity of Authors**

| X (Number Of Authors) | Y(Number Of Publication) |
|---|---|
| 1 | 2434 |
| 2 | 210 |
| 3 | 42 |
| 4 | 17 |
| 5 | 9 |
| 6 | 3 |
| 8 | 2 |
| 11 | 1 |




**Step 3: Calculation of n:**

| x | y | X=log x | Y=log y | XY | XX |
|---|---|---|---|---|---|
| 1 | 2434 | 0 | 3.386321 | 0 | 0 |
| 2 | 210 | 0.30103 | 2.322219 | 0.699058 | 0.090619 |
| 3 | 42 | 0.477121 | 1.623249 | 0.774487 | 0.227645 |
| 4 | 17 | 0.60206 | 1.230449 | 0.740804 | 0.362476 |
| 5 | 9 | 0.69897 | 0.954243 | 0.666987 | 0.488559 |
| 6 | 3 | 0.778151 | 0.477121 | 0.371273 | 0.605519 |
| 8 | 2 | 0.90309 | 0.30103 | 0.271857 | 0.815572 |
| 11 | 1 | 1.041393 | 0 | 0 | 1.084499 |
|  |  | 4.801815 | 10.294632 | 3.524466 | 3.674889 |

| | |
|---|---|
| $n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - (\sum X)^2}$ | n = -3.3488244 |

**Step 4: Calculation of C**

P=20 x=1, 2, 3….19 in

$$C = \frac{1}{\sum_{1}^{P-1} 1/x^n + 1/(n-1)\, p^{n-1} + 1/2p^n + n/24(p-1)^{n+1}}$$

C = 0.0056601

**Step 5: KS Test of goodness of fit.**

| x | $Y_X$ | fy | ∑fy | f*y | ∑fyx | \|fy-fyx\| |
|---|---|---|---|---|---|---|
| 1 | 2434 | 0.895511 | 0.895511 | 0.005660 | 0.005660 | 0.8898510 |
| 2 | 210 | 0.077263 | 0.972774 | 0.057666 | 0.063326 | 0.9094475 |
| 3 | 42 | 0.015453 | 0.988226 | 0.224192 | 0.287518 | 0.7007081 |
| 4 | 17 | 0.006255 | 0.994481 | 0.587514 | 0.875032 | 0.1194491 |
| 5 | 9 | 0.003311 | 0.997792 | 1.240374 | 2.115405 | 1.1176133 |
| 6 | 3 | 0.001104 | 0.998896 | 2.284108 | 4.399513 | 3.4006174 |
| 8 | 2 | 0.000736 | 0.999632 | 5.985694 | 10.385207 | 9.3855754 |
| 11 | 1 | 0.000368 | 1.000000 | 17.388653 | 27.773861 | 26.7738610 |
| | 2718 | | | | $D_{max}$ | 26.7738610 |

$$\text{Critical value} = 1.63 / \sqrt{\sum Y_X}$$

Critical 0.031265 > 0.01

## 10. Conclusions

Lotka described scientific productivity by establishing a relation between the frequency of authors and, their number of publications. There has been a number of studies on scientific productivity from different points of view. For example: relationship of quantity of publication of individuals and scientific recognition; identification of elites in science on the basis of their scientific output; impact of social change on productivity etc. Since the scientists of each field are very curious to know their growth la productivity to formulate new policies, the studies on scientific productivity attain more attention in this modern era.

## References

1. DRESDEN, Arnold. "A Report on the scientific work of the Chicago Section, 1897-1922". American mathematical Society Bulletin, 28, 303-307, 1922.

2. DUFRENOY, Jean. "The publishing behaviour of Biologists" Quarterly reviews of Biology, 13, 207-210, 1938.
3. DAVIS, Harold.T. The Analysis of Economic Time series, Bloomington, Ind: Principia Press, 1941.
4. WILLIAM, C.B. "The number of publication written by Biologists". Annals of Eugenics 12, 143-146, 1944,
5. ZIPF, George K. "Human Behaviour and the Principle of least effort : An Introduction to Human Ecology". Cambridge Mars: Addison Wesley, 1949.
6. LEAVENS, Dickson H. "Letters to the Editor" Econemetrica, 21, 630-632, 1953.
7. PRICE, Derek de solla. Little Science, Big Science, New York: Columbia University Press, 1963.
8. BOOKSTEIN, Abraham "The Biblimetri Distribution" Library Quarterly, 46, 416-423, 1976.
9. ALLISON, Paul D. et al., "Lotka's Law: A problem in its interpretation and application", Social Studies of Science, 6, 269-270, 1976.
10. SHOCKLEY, William. "On the statistics of individual variation of productivity in research laboratories". Proceedings of the Institute of Radio Engineers, 45, 279-90, 1957.
11. MURPHY, Larry, J. "Lotka's law on the Humanities". JASIS 24(6), 461-462, 1973.
12. HUBERT, John, J. "Letters to the Editor". JASIS 25(4), 270, July-August, 1974.
13. COILE, Russel, C. "Lotka's Freguency distribution of scientific productivity". JASIS 28(6), 366, 1977.
14. VOOS, Henry, "Lotka and Information Science". JASIS 25(4), 270, 1974.
15. COILE, Russel, C. "Letters to the Editor". JASIS 26(2), 133, 1995.
16. SCHORR, Alan Edward. "Lotka's law and Library Science". RQ 14(1), 32, 1974.
17. SCHORR, Alan Edward. "Lotka's law and map 18. librarianship". JASIS 26, 189-190, 1995.
18. BQOKSTEIN, op cit.
19. SCHORR, Alan Edward. "Lotka's law and the History of Legal Medicine". Research Librarianship, 30, 205-209, 1975.
20. ROGGE, A.E. "A look at Academic Anthropology". American Anthropologist, 78, 835, 1976,
21. RADHAKRISHNAN, T. and KERNIZAN, R. "Lotka's law and computer science literature". JASIS 30(1), 51, 1979.
22. SUBRAMANYAN, K. "Lotka's law and the literature of computer Science". IEEE Transactions on professional communication, PC-22(4), 188, 1979.
23. POTTER, W.G. "When names collide: Conflicts in the catalog and AACR II". Library resource and Technical Service, 1980.
24. SUBRAMANYAN, K. "Latka's law and library Science Litrature". Library Reseach, 3, 167, 1981.
25. RICHARDSON, Valerie L. "Lotka's law and the Catalogue" AARL, 12, 185, 1981.
26. McCALLUM, S.H. and GODWIN, J.L. "Statistics of Headings in the MARC File". Network Development Office, L.C.
27. MARQUES DE 0L1VEIRA, Slas. "Application of Lotka's law of Author productivity to the literature of the Jack fruit". Revista de Biblioleconemia de Brasilia, 11(1), 125-130, 1983.
28. MARQUES DE OLIEVIRA, Silas. "Lotka's law of Intellectual productivity further applications" Revista da Escola de Biblioleconimia da UFMG. 13(2), 207-233, 1984.
29. PAO, M.L. "An Emperical examination of Lotka's law". JASIS, 37(1), 26-33, 1986.
30. GUPTA, D.K. "Lotka's law and productivity pattern of entomological research In Nigeria for the period 1900-1973", Scientometrics 12(1-2), 33-46, 1987.
31. GUPTA, Devendra, K. "Lotka's law and its application to author productivity distribution of psychological distribution of Africa for the period 1966-1975". Herald of Library Science, 24(1-2), 11-21, 1989.
32. GUPTA, Devendra, K. "Scientometric Study of Biochemical Literature of Nigeria, 1970-1984. Application of Lotka's law and the 80/20 rules". Scientometrics 15(3-4), 171-179, 1989.
33. KYVIK, S. "Productivity differences, Fields of learning and Lotka's law". Scientometrics 15(3-4), 204-214, 1989.
34. COOK, Kevin, L. "Law of Scattering applied to popular music". JASI 40(4), 277-283, 1989.
35. RAVINDER NATH and JACKSON, Wade, M. "Productivity of Management Information system researchers does Lotka's law apply ?" Information processing and management 27(2/3), 203-209, 1991.
36. GUPTA, D,K. "Author productivity pattern Geo-scientific literature of Nigeria for the period 1904-1979" paper presented at the 3rd international conference on informatrics held at DRTC, Bangalore, August, 9-12, 1991.
37. PARVATHAMMA, N., GUNJAL, S.R. and NINJAGUNAPPA, R."Growth pattern of literature and scientific productivity of authors in Earth Science Literature (1978-1988): A

bibliometric study". Paper presented at the 3rd international conference on informetrics held at DRTC, Bangalore, August, 9-12, 1991.
38. SARADA, B.A. "Bibliometric studies in linguistics and Bebliometric laws". Paper presented at the 3rd international conference on informetrics held at DRTC, Bangalore, August 9-12, 1991.
39. VAISHNAV, A.A. "Literature of computerised library and Information services: A Bibliometric study". Paper presented at the 3rd international conference on Informetrics held at DRTC, Bangalore, August, 9-12, 1991.
40. GUPTA, D.K. and GUPTA, Sushma. "Conference literature: International conferences on Ethiopean studies 1985-1988". INICAE 10(2), 176-184, 1991.
41. LEMOINE, W. " Productivity pattern of men and women scientisits in Venzuela". Scientometrics 24(2), 281-295, 1992.
42. LEMOINE, W. "The frequency distribution of Research papers and patents according to sex. The, case of CSIR India". Scientometrics 24(3), 449-469, 1992.
43. MAHESWARAPPA, B.S. "Bibliometric Analysis of Phytopathology Literature". Ph.D. Thesis, Gulburga University, 494-501, 1992.
44. ROUSSEAU, Ronald. "Breakdown of the robustness property of Lotka's law. The case of Adjusted counts for multiauthorship attribution". JASIS 43(10), 645-647, 1992.
45. GUPTA, B.M. and KARISIDDAPPA, C.R. "Author productivity pattern in theoritical population genetics 1900-1980". Scientometrics, 36(1), 19-41, 1996.
46. SEN, B.K. TAIB, C.Ab. and HASSAN, M.Fb. "Litrary and Information Science Literature and Lotka's law". Malaysian Journal of Library and Information Science, 1(2), 89-93, 1996.
47. GUPTA, B.M. et al. "Distribution of productivity among authors in Potato research 1900-1980" Library Science with slant to documentation and Information Studies, 33, 127-134, 1996.
48. JOSHI, A.N. and MAHESWARAPPA, B.S. "Studies in Scientific productivity: A Review of Literature". International Information Communication and Education 15(2), 161-196, 1996.

3/21/2013