THE INTERNATIONAL JOURNAL OF HUMANITIES & SOCIAL STUDIES

Application of Lotka's Law in the Field of Women's Studies Literature in India

Asha Sharma
Deputy Librarian, Central Library, Banasthali Vidyapith, Rajasthan, India
Dr. M. S. Rana
Librarian, HNB, Garhwal University, Uttarakhand, India

Abstract:
The purpose of this study is to know the author's productivity as measured by the application of Lotka's Law and its validity using 'Kolmogorov-Smirnov test'. The required data (5530 articles) has been collected from 314 different journals published during 1945- 2008.Lotka's law is applied to know the number of authors contributing single paper, two papers and three papers. It has been found that 3126 (76.23%) authors have contributed only single articles. The average value of authors' productivity has been observed to be 2.61, which is higher to the inverse square value (i.e. 2) of Lotka's law. By using b = 2.61, the expected frequencies of authors were calculated. The maximum deviation (D) using Kolmogorov-Smirnov statistics test was found to be 0.01; the level of significance was calculated to be 0.0254. Here, D < 0.0254, upholds the applicability of Lotka's Law to the data of the present study.

Keywords: *Bibliometric study, Bradford's law, Kolmogorov- Smirnov test', Lotka's law, productivity of authors, Women's Studies and Zipf's law*

1. Introduction

We are living in the age of knowledge and information revolution. In this age of literature explosion, information is growing very fast in all the fields of knowledge and Women's Studies are no exception. Due to accelerated research and development activities in the area of Women's Studies, the literature has proliferated enormously. In the age of information explosion, it is very essential to organize the information resources for effective and efficient use. Bibliometric study tries to portray a certain kind of human behavior with respect to the flow of information. It reveals how authors in a particular subject organize and communicate their intellectual efforts. **B**azerman (1981) [1] has quite aptly put it in these words, "from the shape of things, one can better understand how things happen" Similarly, **F**rohmann (1982) [2] has observed that, "just as a command of geography aids navigation, a good representation of the structural characteristics of a discipline help both researcher and scholar to identify issues and problems in the field".

1.1. Bibliometrics

Bibliometrics is the application of mathematical and statistical method for measuring quantitative and qualitative changes in the production of literature in a given subject or area of specialization. The word measurement means the application of mathematical and statistical technique to find out the growth of documents, scattering of literature in various type of documents, publication of documents by an author and impact of documents and so on. The term 'Bibliometric' was coined by the British scientist Alan Pritchard in 1969. This study highlights the genesis of bibliometrics, and also explains the different analogous terms of bibliometrics like 'Librametrics', 'Infomietrics', 'Scientometrics', 'Cybermetrics' and more recent terms like 'Webometrics' and 'Altmetrics'. Bibliometrics as a technique has extensive applications in identifying the research trends in a subject, trends in authorship and collaboration in research, core periodicals, obsolescence and dispersion of literature useful in estimating the comprehensiveness of periodicals, studying the productivity, characteristics of subject literature including structure of knowledge and helpful in formation of need based collection development policy, weeding and stacking policy and many others.

1.2. Women's Studies

Women's Studies is an important field for research in this country and a subject of great interest in countries abroad today. According to **M**cFadden (2005) [3] "It has moved around the world as an idea, a concept, a practice, and finally a field. By the early twentieth century, this area of women's studies was recognized in higher education from India to Indonesia, United States to Uganda, China to Canada, Austria to Australia, England to Egypt, and South Africa to South Korea."**J**ain and **R**ajput[4]have been discuss in their book "Narrative from the women's studies family, recreating knowledge" that Women's studies are a new field which is being considered by Indian university in recent years. It is not only for academic and intellectual ideas but also for bringing change. It is not an ivory tower exercise but much more than this. The Indian Council of Social Science Research (ICSSR) and later the National Policy of

Education (NPE 1986, which was updated in 1992) identified Women's Studies as a "critical instrument of social and educational development'. The policy, stressing the need for education of women, laid down that 'education will be used as an agent of basic change in the status of women.

1.3. Laws of Bibliometrics

This is one of the oldest areas of bibliometrics. Quantitative relationships based on the empirical work related to the patterns in productivity of journals, productivity of individual scientists, or word frequencies in the literature fall under this category. These relationships go by the names of Bradford's law; Zipf's law and Lotka's law are the best known laws dealing with important phenomena or "regularities" found in science communication. These bibliometric laws are empirically based distributions, and are not innate natural laws but essentially behaviour patterns of the users of the scientific literature. All of them can be used either as integrative research methods or as analytical tools to discover the deeper segments of a research field and to trace trends and developments that are not directly visible. By measuring the flow of textual materials in scientific communication these laws can help to monitor the primary literature in a scientific domain, as well as to evaluate the impact and quality of the information sources available.

These laws play a significant role in modern library resource management, because with decreasing budgets and the rapid growth of periodicals, library managers are desperately looking for effective ways to manage their libraries.

The three fundamental laws which laid the formation of bibliometrics are:

(a) Lotka's Inverse Square Law of Scientific Productivity: In 1926, Alfred J. **L**otka [5] proposed this inverse square law correlating contributions of scientific papers to their number of contributions. His law provided fundamental theoretical base for bibliometric studies involving authorships.

(b) Bradford's Law of Scattering or Scientific Papers: Samuel Clement **B**radford [6] noticed that in each subject there are a few very productive sources, and a large number of sources which are moderately productive and still a large number of sources of constantly diminishing productivity.

(c) **Z**ipf's Law of Word Occurrence: Zipf [7] developed and extended an empirical law governing the relation between rank of a word and frequency of its appearance in a long term.

2. Review of Literature

Radhakrishnan and Kernizan (1979)[8] conducted tests to verify the satisfaction of Lotka's Law, with the papers published in the area of Computer Science. They stated that in their first experiment, the assumption was made that "an author publishes exclusively through one scientific journal", but it was found that this assumption was not valid. In the second experiment the authors considered all the papers published by the authors irrespective of the journal. A random selection of authors was made using the cumulative author index of Computer and Control abstracts and was repeated with JACM. The deviation from Lotka's Law was high.

Gupta (1993)[9] also studied author productivity trends by analyzing a cumulative index of Geophysics for the two important journals *Geophysics* and *Geophysical Prospecting* for the period 1936-1985. Author productivity trends were tested by applying K- S statistical test. Two files were generated out of this database: one of the period 1936-1985 and the other for the period 1936-1976 to test the time sensitivity of Lotka's Law. Lotka's Law did not apply as inverse square law but could apply satisfactorily with exponent value of 2.1 on author productivity distribution patterns of both the files.

Gupta (1998)[10] conducted a study on the growth of Indian and World Physics Literature from 1900-1950. He explored the applicability of selected technology diffusion models to the growth of literature in Indian and world physics. He focused on the applicability and validity of two forms of Lotka's Law and negative binomial distribution model to the cumulative author productivity of data on Indian physics, looked at the linkages between inequality/concentration measures and development of Indian physics as a discipline. He explored the relevance and applicability of two well-known generalizations, price square root law and 80/20 rule to the cumulative author productivity data on Indian physics.

Suresh (2003) [11] found that there exists an inverse relationship between productivity and its producers. The relationship is summarized in a power- law relationship, referred to as Lotka's Law. According to it, the fraction of authors in a given field who publish x contribution is proportional to l/(x-'1) where x is the number of organizations contributions and a is an index number. He counted the number of individual authors and their contributions in the Decennial Index of Chemical Abstract, 1907-16.

3. Objective of the Study

The Objective of this study is to know the author's productivity as measured by the application of Lotka's Law in the field of Women's Studies literature in India. This study also examines the validity using 'Kolmogorov- Smirnov test'.

4. Methodology

For the purpose of this study a computerized database has been designed and created for compilation of Women's Studies literature. The required data (5530 articles) has been collected from 314 different journals published during 1945- 2008. For this database each article is scanned and checked, then the data is entered in the computer using MS Office-Excel software and the results are obtained in tabular form by using SPSS (Statistical Package for Social Sciences) software.

4.1. Application of Lotka's Law

Several formal analytical and predictive models have been published in literature describing the phenomena of author productivity in the different fields. The number of publications and research papers, of an individual scientist is considered as a measure of his/ her scientific productivity.The productivity of authors in terms of scientific productivity was determined by Alfred Lotka in 1926, as a quantitative measure of the frequency distribution of scientific productivity of chemists and physicists after analyzing the number of publications listed in *Chemical Abstracts* (1907- 1916). The contribution of physicists listed from Auerbach's Geschichtstafeln der Physik. On the basis of this data Lotka summarized a statement popularly known as Lotka's Inverse Square Law of Scientific Productivity.[12] Hereafter this equation is recognized in the field of Library and Information Science as Lotka's Equation. This equation is tested by many prominent scientists from Library and Information Science with different types of data from various subjects.

In the present study, the trend of scientific productivity of authors in Women's Study literature has been attempted using Lotka's pattern of identification. It was observed that 4101 authors have contributed 6276 papers and single authorship is very common in this field, 3126 authors (76.23) have contributed one paper. Lotka's law has been applied to know the number of authors contributing two papers, three papers and four papers respectively.

Mathematically, the equation is formulated as

$a_n=c/n^b$ (1)

where a_n is authors' productivity (number of authors contributing n paper each),

c is constant (the number of authors contributing only one paper each),

n is number of papers (1,2,3…..n)

b is characteristic exponent value of the distribution for a set of a data)

Lotka's equation is determined by three parameters which are stated below:

(i) The number of authors with minimal productivity (authors with single paper each)

(ii) The maximal productivity of authors (n_{max})

(iii) The characteristic exponent

Therefore, a sample size to know the pattern of productivity of authors has been considered from 1945- 2008. To apply Lotka's law, the author index has been generated from the computerized database on Women's Study literature. The frequency of authors as number of contributions has been calculated and placed against it respectively. As proposed by Pao[1] in 1985, the following testing procedure for fitting of Lotka's inverse- power rule given by above equation has been followed:

1) A frequency table consisting of pairs (n, a_n) where n represents the number of contributions and a_n is the number of authors writing n papers each is created in table no. 1.

No. of Contribution	No. of Authors	% of Authors	Total Contribution	% of Contribution
1	3126	76.23	3126	49.81
2	571	13.92	1142	18.20
3	179	4.36	537	8.56
4	90	2.19	360	5.74
5	50	1.22	250	3.98
6	26	0.63	156	2.49
7	14	0.34	98	1.56
8	10	0.24	80	1.27
9	7	0.17	63	1.00
10	2	0.05	20	0.32
11	7	0.17	77	1.23
12	3	0.07	36	0.57
13	1	0.02	13	0.21
14	4	0.10	56	0.89
16	2	0.05	32	0.51
17	2	0.05	34	0.54
18	2	0.05	36	0.57
19	2	0.05	38	0.61
23	1	0.02	23	0.37
49	1	0.02	49	0.78
50	1	0.02	50	0.80
Total	4101	100	6276	100.00

Table 1: Contribution of Authors in Women's Study Literature

The above table reveals that if the frequency of the authors increases, the number of contributions decreases, whereas if the frequency of the authors decreases, the number of contributions increases.

2) Adopted Lotka's inverse- power rule, as given in equation (1).
3) The parameter b is expected by the least- square method in the simple regression model:

$\text{Log } a_n = \text{Log } c/ n^b$
$\text{Log } a_n = \text{Log } n^{-b} + \text{Log } c$
$\text{Log } a_n = - b \text{ Log } n + \text{Log } c$
$b \text{ Log } n = \text{Log } c - \text{Log } a_n$
$b = \frac{\text{Log } c - \text{Log } a_n}{\text{Log } n}$ (2)

No. of Contributions/Author	Value of b	Expected frequency of Authors	% of Authors
1		3126	77.32
2	2.45	509	12.58
3	2.60	176	4.35
4	2.56	83	2.05
5	2.57	46	1.14
6	2.67	29	0.71
7	2.78	19	0.47
8	2.76	13	0.33
9	2.78	10	0.24
10	3.19	8	0.19
11	2.54	6	0.14
12	2.80	5	0.12
13	3.14	4	0.09
14	2.52	3	0.08
16	2.65	2	0.05
17	2.60	2	0.05
18	2.54	2	0.04
19	2.50	1	0.03
23	2.57	1	0.02
49	2.07	0	0.00
50	2.06	0	0.00
Total		4043	100.00

Table 2: Frequency of Expected Author
Average value of b = 2.61

The above Table 2 gives the value of b and expected frequency of authors by using equation (2). The average value of b has been found to be 2.61. By using b = 2.61, the expected frequencies of authors hase calculated.

4) The parameter c is equal to number of authors with minimal productivity (i.e. 3126 number of authors)
5) Lastly the Kolmogorov- Smirnov test has been used to test the maximum deviation is : Dmax= Maximum value of the D or $[F_o (x) - S_n (x)]$

Where $F_o(x)$ is the expected relative frequency and $S_n(x)$ is the observed relative frequency of a sample of total number of authors (4101). Since the Kolmogorov- Smirnov D max of 0.01 given in Table 5.3 is smaller than the level of significance (p = 0.01) of 0.0254, Lotka's law has been found to be applicable to this sample of authors of Women's Study literature.

		Observed			Expected			
No. of Contribution	**Frequency of Authors**	**Cumulative Frequency of Authors**	**Relative Frequency $S_n(x)$**	**Frequency of Authors**	**Cumulative Frequency of Authors**	**Relative Frequency $F_o(x)$**	**Deviation D = Fo(x) - $S_n(x)$**	**Dmax**
1	3126	3126	0.76	3126	3126	0.77	0.01	0.01
2	571	3697	0.90	509	3635	0.90	0.00	
3	179	3876	0.95	176	3810	0.94	-0.01	
4	90	3966	0.97	83	3893	0.96	-0.01	
5	50	4016	0.98	46	3939	0.97	-0.01	
6	26	4042	0.99	29	3968	0.98	-0.01	
7	14	4056	0.99	19	3987	0.99	0.00	
8	10	4066	0.99	13	4000	0.99	0.00	
9	7	4073	0.99	10	4010	0.99	0.00	
10	2	4075	0.99	8	4018	0.99	0.00	
11	7	4082	1.00	6	4023	1.00	0.00	
12	3	4085	1.00	5	4028	1.00	0.00	
13	1	4086	1.00	4	4032	1.00	0.00	
14	4	4090	1.00	3	4035	1.00	0.00	
16	2	4092	1.00	2	4037	1.00	0.00	
17	2	4094	1.00	2	4039	1.00	0.00	
18	2	4096	1.00	2	4041	1.00	0.00	
19	2	4098	1.00	1	4042	1.00	0.00	
23	1	4099	1.00	1	4043	1.00	0.00	
49	1	4100	1.00	0	4043	1.00	0.00	
50	1	4101	1.00	0	4043	1.00	0.00	
Total	**4101**			**4043**				

Table 3: Frequency of Observed and Expected Number of Authors (Kolmogorov- Smirnov Test)

Total number of authors = 4101

Dmax= Maximum value of the D [F_o (x) - S_n (x)] = 0.01

At the 0.01 level of significance,

K-S statistics = $1.63/\sqrt{4101}$

= 1.63/64.039

= 0.0254

D < 0.0254

In the graphical form, Lotka's Law has been plotted taking both observed and expected values in figure 1 as number of articles against the number of authors contributing the articles on a logarithmic scale. The curves follow Lotka's pattern with slight deviation.

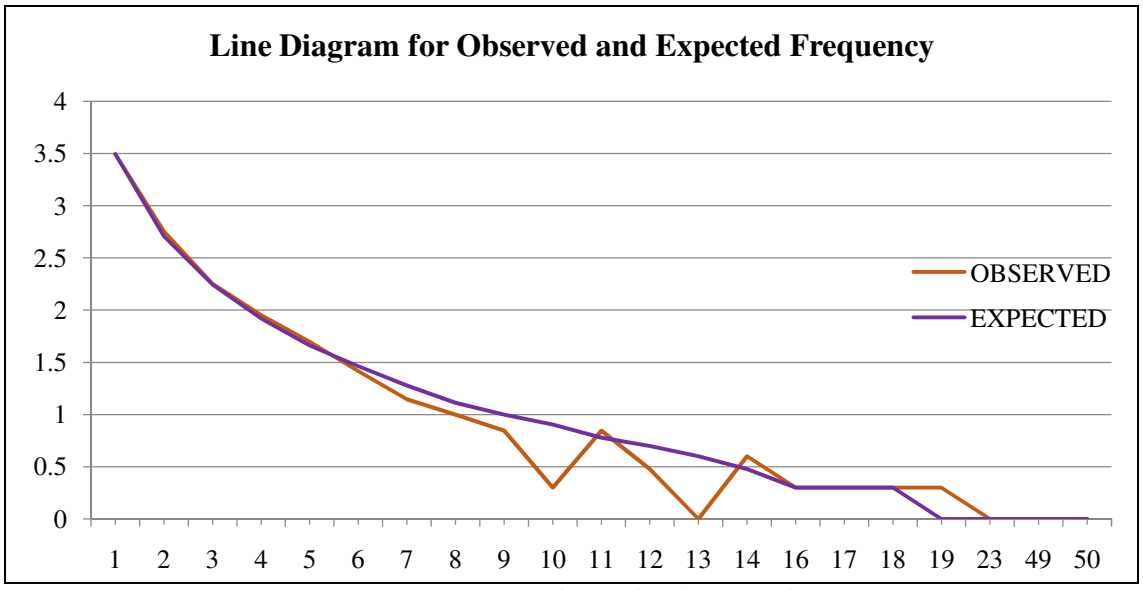

Figure 1: Line Diagram for Observed and Expected Frequency

4.2. Conclusions

In this study the collected data has been tested by Lotka's law and its validity using 'Kolmogorov- Smirnov test' and the results have been obtained. The required data (5530 articles) has been collected from 314 different journals published during 1945- 2008. It is observed that 4101 authors have contributed 6276 papers and single authorship is very common in this field; 3126 authors (76.23) have contributed one paper. Lotka's law is applied to know the number of authors contributing two papers, three papers and four papers respectively. The average value of author's productivity has been observed to be 2.61, which is higher to the inverse square value (i.e. 2) of Lotka's law. The maximum deviation (D) using Kolmogorov- Smirnov statistics test has been found to be 0.01; the level of significance has been calculated to be 0.0254. It is also observed that Lotka's law is applicable to the data of Women's Studies literature.

5. References

i. Bazerman, C. "What Written Knowledge Dose: Three Examples of Academic Discourse." Philosophy of Social Science. 11 (1981): 367.
ii. Forhmann, B. "A Bibliography Analysis of the Literature of Cataloguing and Classification." Library Research. 4 (1982): 265.
iii. McFadden, Margaret, "Women's Studies." New Dictionary of the History of Ideas, 2005. http://www.encyclopedia.com.
iv. Jain, Devaki and Pam Rajput. Narrative from the Women's Studies Family: Recreating Knowledge. New Delhi: Sage Publication, 2003.17.
v. Lotka, Alfred J. "Frequency Distribution of Scientific Productivity. Journal of the Washington Academy of Sciences. 16 (1926): 317-323.
vi. Bradford, S. C. "Sources of Information on Specific Subject." Engineering. 137 (1934): 85- 86.
vii. Zipf, G. K. Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology. Cambridge, Mass Addison: Wesley, 1949.
viii. Radhakrishnan, T and R. Kernizan. "Lotka,s Law and Computer Science Literature." Journal of American Society of Information Science. 30.1(1979): 51
ix. Gupta, D. K. "Collaborative Research Trends in Exploration Geographysics." Scientometrics. 28. (1993):287- 296.
x. Gupta, B. M. "Distribution of Productivity among Authors in Potato Research 1900- 1980.Library Science with Slant to Documentation and Information Studies. 33. 3 (1998): 127- 134.
xi. Suresh Kumar. "Lotka's Law and Author Productivity in the Field of Computer Science in India." Library Herald. 41.2(2003):90- 97.
xii. Lotka, Alfred J. "Frequency Distribution of Scientific Productivity. Journal of the Washington Academy of Sciences. 16 (1926): 317-323.
xiii. Pao, M. L. "Lotka's Law: A Testing Procedure." Information Processing and Management. 21. 4 (1985): 305- 320.