Information Processing & Management Vol. 24, No. 4, pp. 469–477, 1988
Printed in Great Britain.

0306-4573/88 $3.00 + .00

PRICE'S SQUARE ROOT LAW: EMPIRICAL VALIDITY AND RELATION TO LOTKA'S LAW

PAUL TRAVIS NICHOLLS
School of Library and Information Science, University of Western Ontario,
London, Ontario, Canada N6G 1H1

(*Received* 16 *April* 1987; *accepted in final form* 9 *November* 1987)

Abstract — Price's well-known square root law states that half of the literature on a subject will be contributed by the square root of the total number of authors publishing in that area. Price's contention is treated here as a hypothesis and assessed against the evidence presented by both empirical and simulated author productivity distributions. The results do not support the square root hypothesis. The problem with Price's original claim is traced to its basis in Lotka's law, which is considered as an inverse square law rather than as a generalized model taking variable parameter values. Varying parameter values engender a family of related, but systematically different, distributions in which the nature of inequality in publication productivity, including the size and relative contribution of the most prolific subset of authors, also varies.

1. INTRODUCTION

One of the many enduring contributions of Derek DeSolla Price's classic *Little Science, Big Science* [1] was the statement of a relation that has come to be known as Price's square root law, or simply, the Price law. Basically, Price's contention was that half the published output in a subject field will be contributed by a highly productive subset of authors approximately equal to the square root of the total number of authors publishing in that area. With this, Price introduced to bibliometrics a concept that had long been known in the social sciences as Rousseau's law: "any population of size N contains an effective elite of size $N^{0.5}$" [2]. Price felt that such a phenomenon was directly implied by Lotka's law [3]:

> Let us first examine the nature of the crude inverse-square law of productivity. If one computes the total production of those who write n papers, it emerges that the large number of low producers accounts for about as much of the total as the small number of large producers; in a simple schematic case, symmetry may be shown to a point corresponding to the square root of the number of [authors]. . . . [1]

This claim has been rather uncritically accepted and accorded lawlike status within bibliometrics and scientometrics, without ever really having been subjected to validity testing: "Apart from the fact that his claim is based on poor evidence, it is yet to be tested by statistical techniques" [4]. The purpose of the present study is to assess the validity and generality of the Price law against both empirical and simulated author productivity distributions.

2. LOTKA'S LAW

The Price law is often associated with Lotka's law; indeed, the basis for Price's initial hypothesis was an examination of the predictions of Lotka's theoretical distribution. Although originally proposed as an inverse square model, Lotka's law is now usually defined in the generalized form:

$$g(x) = kx^{-b}; \qquad x = 1,2,\ldots,x_{\max},\ k > 0,\ b > 1, \tag{1}$$

where $g(x)$ represents the probability of an author making x published contributions to a subject field, x_{max} represents the maximum size or value of the productivity variable x, and k and b are parameters. Methods of estimating the parameters are discussed by Nicholls [5,6] and by Tague and Nicholls [7]. Usually, x_{max} is assumed to be infinite. The parameter k, however, which represents the proportion of authors contributing only once to the literature, is of logical necessity given by the inverse of the Riemann zeta function for b:

$$k = \zeta(b)^{-1} = \sum_{x=1}^{\infty} x^{-b} \tag{2}$$

We will denote the total number of authors in a sample or population by t, and the total number of papers they generate by m:

$$t = \sum_{x=1}^{x_{max}} g(x), \tag{3}$$

$$m = \sum_{x=1}^{x_{max}} g(x)x. \tag{4}$$

Recently, an alternative notation for defining Lotka's model has been proposed by Chen and Leimkuhler [8]. This notation explicitly takes account of the fact that empirical values of x will not increase *without gaps* from 1 to x_{max}.

For some time (more than 60 years) the empirical validity of Lotka's law was questionable [9]; recently, however, it began to be clear that the generalized Lotka model (eqn (1)) can provide an adequate description of empirical author productivity distributions [6,10,11]. At the same time, it has been shown that the original inverse square formulation, in which $b = 2$, provides very poor fits to empirical data and may now be viewed as invalid [6,10]. Price based his hypothesis on the inverse square model.

3. THE PRICE LAW

The Price law was originally presented in a verbal formulation. Recently, Glänzel and Schubert [12] have provided a more formal exact definition of the law. These authors adopted a rank–frequency approach; we shall reformulate the definition in size–frequency terms, corresponding to the presentation of Lotka's law in eqn (1). In these terms, the Price law hinges on the definition of a productivity level $x = h$, such that

$$\frac{1}{2}\left[\sum_{x=1}^{x_{max}} g(x)x\right] = \sum_{x=h}^{x_{max}} g(x)x, \tag{5}$$

where h satisfies the requirement that

$$\left[\sum_{x=1}^{x_{max}} g(x)\right]^{1/2} = \sum_{x=h}^{x_{max}} g(x). \tag{6}$$

Allison et al. [13] provided a similar exact formulation of the Price law.

Glänzel and Schubert acknowledge two problems with the definition above. First, the square root of the total number of authors, $t^{0.5}$, is not always integer valued, and therefore the definition of h may be ambiguous. In practice, truncated or rounded values are employed. Second, they note that eqn (5) cannot be characteristic to a probability distribution, since its validity depends also on the sample size. We may add a third problem. For $g(h) > 1$, which is a common occurrence, a further ambiguity arises in defining the cutoff point between the prolific and less prolific author regions. This is because the quantity $t^{0.5}$ will then often fall *within* one of the categories of the frequency distribution, making it necessary to arbitrarily select the cutoff category so that $\sum_{x=h}^{x_{\max}} g(x)$ contains either *no more* than $t^{0.5}$ authors (a narrow definition of the prolific subset) or *at least* this many authors (a broad definition). With very large data sets, the decision may be unnecessary or its effects trivial, since there will be many categories, but with smaller data sets the consequences of the decision will have a more significant effect on analytical results.

Therefore, Glänzel and Schubert formulate Price's hypothesis as a limiting law:

$$\lim_{x_{\max}\to\infty}\left[\frac{\sum_{x=h}^{x_{\max}} g(x)x}{\sum_{x=1}^{x_{\max}} g(x)x}\right] = \frac{1}{2}, \tag{7}$$

where h satisfies the requirement of eqn (6).

Eqn (7) represents an exact definition of the Price law as originally proposed and with which this study is concerned. However, the Price law is clearly a special case of a more general relation. Egghe and Rousseau [14] have provided such a general formulation. Their generalized relation states that t^{α} authors will generate a fraction θ of the total number of papers, and that $\alpha = \theta$. Price's square root law would then be regarded as the special case where $\alpha = 0.5$. Egghe and Rousseau mention also the familiar 80/20 rule, which in turn may be regarded as a particular case of a generalized arithmetic $100x/100\theta$ rule, in which some fraction $100x\%$ of prolific authors will be responsible for producing $100\theta\%$ of the total output. Price's remark [1] that 10% of the authors will produce 90% of the papers is a rule of this type (a "90/10 rule"). The relation of the 80/20 rule to the bibliometric distributions is examined in Egghe [15].

4. VALIDITY OF THE PRICE LAW

Two possible sources of validity for the Price law (eqn (7)) may be identified—empirical and theoretical. If empirical author productivity data sets conformed to Price's hypothesis, then it would possess some degree of empirical validity. If, alternatively, the theoretical values generated by Lotka's model were consistent with Price's hypothesis, then it could be said to have some theoretical validity. Since Lotka's model agrees so well with the empirical case, both avenues of validity testing would be expected to yield consistent results.

Little empirical investigation of the Price law has been carried out to date [4,14]. Glänzel and Schubert [12] have reported some empirical results. They analyzed Lotka's *Chemical Abstracts* data and found that the most prolific $t^{0.5}$ authors contributed less that 20% of the total number of papers. They also refer, but without details, to the examination of "several dozens" of other empirical data sets and conclude that "in the usually studied populations of scientists, even the most productive authors are not productive enough to fulfill the requirements of Price's conjecture" [12]. Some incidental results of scientometric studies suggest that about 15% of the authors will be necessary to generate 50% of the papers [16,17].

To further examine the empirical validity of Price's hypothesis, 50 data sets were collected and analyzed here. The origin and statistical characteristics of these are recorded in Ref. [11]. Included are most of the data sets used in previously published studies of Lotka's law. A wide range of disciplines, time frames, and sample sizes are represented; however, the complete count measurement method is used in all cases; that is, all coauthors are included in the counts. Table 1 illustrates how the data were tabulated for analysis. This is the well-known Dresden data on the productivity of American mathematicians [11]. The first three columns contain the productivity level x; the number of authors having this productivity, $g(x)$; and the number of papers produced, $g(x)x$. The next two columns tabulate

Table 1. Dresden data

x	$g(x)$	$g(x)x$	$\sum g(x)$	$\sum g(x)x$	$\sum (g(x)/t)$	$\sum (g(x)x/m)$
1	133	133	133	133	0.4784	0.1183
2	43	86	176	219	0.6331	0.1948
3	24	72	200	291	0.7194	0.2589
4	12	48	212	339	0.7627	0.3016
5	11	55	223	394	0.8022	0.3505
6	14	84	237	478	0.8525	0.4253
7	5	35	242	513	0.8705	0.4564
8	3	24	245	537	0.8813	0.4778
9	9	81	254	618	0.9137	0.5498
10	1	10	255	628	0.9173	0.5587
11	3	33	258	661	0.9281	0.5881
12	5	60	263	721	0.9460	0.6415
13	1	13	264	734	0.9496	0.6530
14	1	14	265	748	0.9532	0.6655
15	1	15	266	763	0.9568	0.6788
16	2	32	268	795	0.9640	0.7073
19	1	19	269	814	0.9676	0.7242
20	1	20	270	834	0.9712	0.7420
21	1	21	271	855	0.9748	0.7607
24	1	24	272	879	0.9784	0.7820
27	1	27	273	906	0.9820	0.8060
32	1	32	274	938	0.9856	0.8345
35	1	35	275	973	0.9892	0.8657
39	1	39	276	1012	0.9928	0.9004
42	1	42	277	1054	0.9964	0.9377
70	1	70	278	1124	1.0000	1.0000

the cumulative number of authors, $\Sigma g(x)$, and the cumulative number of papers produced, $\Sigma g(x)x$. The last two columns express these in relative frequencies: the cumulative percent of authors, $\Sigma(g(x)/t)$, and of papers, $\Sigma(g(x)x/m)$.

The problem in defining the cutoff point, h, between the prolific and less prolific authors is evident. The Price law states that the square root of the total number of authors constitutes the prolific group. In this case, $t = 278$ and $t^{0.5} = 16.6733$. Since this is not an integer, we round it up to 17. However, we are still faced with the problem of defining the cutoff point, since with $h = 13$ we have only 15 authors, and with $h = 12$ we have 20 in the prolific group. Therefore, if we adopt a narrow definition of the prolific group ($h = 13$), we will have 15 authors ($0.0540t$) producing 403 papers ($0.3585m$); and if adopting a broad definition ($h = 12$), 20 authors ($0.0719t$) producing 463 papers ($0.4119m$). The Price law suggests that half the total of papers, or 562 ($0.5m$), will be contributed by 17 authors ($0.0612t$). Since the productivity of the prolific authors falls short of $0.5m$ even when we define the group broadly and include 20 authors, the Price law does not appear to apply in this case. It is also of interest to know how many prolific authors are *actually* required to produce half the papers in this data set. Table 1 shows that 587 papers ($0.5222m$) were generated by 33 authors ($0.1187t$), about twice as many authors as the Price law would predict.

In the 50 data sets examined here, the contribution of the most prolific $t^{0.5}$ group of authors fell considerably short of the $0.5m$ predicted by Price, whether the point of dissection was defined broadly or narrowly. Furthermore, the actual proportion of all authors necessary to generate at least 50% of the papers was found to be much larger that $t^{0.5}$. Table 2 summarizes these results. In some cases, $h = 1$; that is, half the papers are already contributed at $x = 1$, which is to say that more than half of the total number of papers is generated by those authors contributing only a single paper each. The absolute and relative size of $t^{0.5}$ for various population sizes t is given in Table 3. All the empirical results referred to here are consistent; and, unfortunately, there seems little reason to suppose that further empirical results would offer any support for the Price law.

Price felt that the square root law followed directly from Lotka's law. Allison et al.

Table 2. Empirical author productivity data

	Mean	Standard deviation	Median	Quartile Lower	Quartile Upper	Interquartile range
Percent of papers contributed by prolific authors, broadly defined	.2259	.1050	.2059	.1458	.2881	.1423
Percent of papers contributed by prolific authors, narrowly defined	.1497	.0857	.1328	.0958	.1915	.0957
Actual proportion of authors contributing 50% of the papers	.6266	.3830	.7217	.2759	1.0000	.7241

Table 3. Absolute and relative size of $\sqrt{t}$ for various population sizes t

Population size, t	Number of authors, $\sqrt{t}$	Proportion $\sqrt{t}/t$
10	3.16	.30
50	7.07	.14
100	10.00	.10
500	22.36	.045
1000	31.62	.032
5000	70.71	.014
10,000	100.00	.010
100,000	316.23	.003
500,000	707.11	.001
1,000,000	1000.00	.001

emphasized that "the validity of Price's law does not necessarily depend on the validity of Lotka's law, and hence can be judged on the basis of empirical evidence alone" [13]. Theoretical derivations by these authors suggested that the Price law is implied by Lotka's law only when an additional assumption is made concerning x_{max}, the contribution of the most prolific author. Derivations by Egghe and Rousseau [14] suggested that Lotka's law would not satisfy the Price law, but that it might approximate to some generalized form of the law. Price, Allison et al., and Egghe and Rousseau were all working with the inverse square form of Lotka's law (i.e., with $b = 2$). It is not entirely clear that the Price law follows from Lotka's inverse square law; and it is not at all clear what the implications of the generalized Lotka model, with variable parameter values, might be for the Price law. Therefore, theoretical author productivity distributions were simulated by generating the values predicted by Lotka's model (eqn (1)) with typical parameter values within a plausible range of $b = 1.5, 2.0, 2.5, 3.0$; $x_{max} = 100, 250, 500$. As noted earlier, the parameter k is functionally dependent on b. A population of 10,000 authors was arbitrarily assumed in all cases. Table 4 shows the Lotka distributions resulting with various values of b ($x =$ 1 to 25 only). These data are graphed in Figs. 1 and 2. Again (Table 5), the productivity of the most prolific $t^{0.5}$ authors falls considerably short of 50%, although there is variation according to the particular combination of parameter values. It is interesting to note that the theoretical case comes closest to Price's claim with $b = 2$ and large x_{max}. However, even when $x_{max} = 1000$, the theoretical contribution of $t^{0.5}$ reaches only 0.3793 of the total papers. The actual size of the prolific group producing 50% of the papers is again much larger than $t^{0.5}$ (Table 5). Therefore, the generalized Lotka model does not appear to be consistent with the Price law. We find no theoretical support, therefore, for Price's hypothesis in Lotka's law; in fact, the two are quite inconsistent. As shown in the previous section, the Price law is also inconsistent with empirical author productivity data.

Table 4. Lotka distributions for various values of b

x	b = 1.5	2.0	2.5	3.0
1	0.382800	0.607900	0.745400	0.831900
2	0.135340	0.151975	0.131769	0.103988
3	0.073670	0.067544	0.047817	0.030811
4	0.047850	0.037994	0.023294	0.012998
5	0.034239	0.024316	0.013334	0.006655
6	0.026046	0.016886	0.008453	0.003851
7	0.020669	0.012406	0.005750	0.002425
8	0.016918	0.009498	0.004118	0.001625
9	0.014178	0.007505	0.003067	0.001141
10	0.012105	0.006079	0.002357	0.000832
11	0.010493	0.005024	0.001857	0.000625
12	0.009209	0.004222	0.001494	0.000481
13	0.008167	0.003597	0.001223	0.000379
14	0.007308	0.003102	0.001016	0.000303
15	0.006589	0.002702	0.000855	0.000246
16	0.005981	0.002375	0.000728	0.000203
17	0.005461	0.002103	0.000626	0.000169
18	0.005013	0.001876	0.000542	0.000143
19	0.004622	0.001684	0.000474	0.000121
20	0.004280	0.001520	0.000417	0.000104
21	0.003978	0.001378	0.000369	0.000090
22	0.003710	0.001256	0.000328	0.000078
23	0.003470	0.001149	0.000294	0.000068
24	0.003256	0.001055	0.000264	0.000060
25	0.003062	0.000973	0.000239	0.000053

Table 5. Simulated author productivity data

	Productivity of t $x_{\max}$			Actual proportion producing $0.5m$ $x_{\max}$		
b	100	250	500	100	250	500
1.5	.1314	.1809	.2367	.1434	.0925	.0665
2.0	.0978	.2328	.3295	.0810	.0529	.0369
2.5	.1429	.1681	.1820	.2546	.2546	.2546
3.0	.1047	.1060	.1060	1.0000	1.0000	1.0000

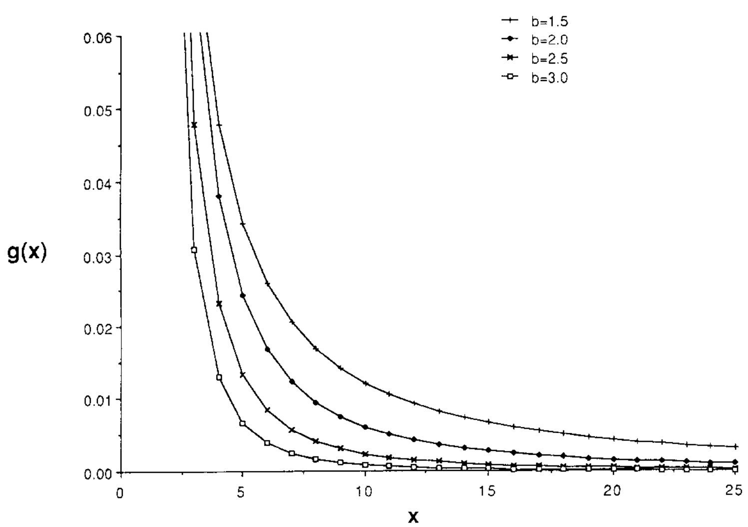

Fig. 1. Lotka curves for various values of b.

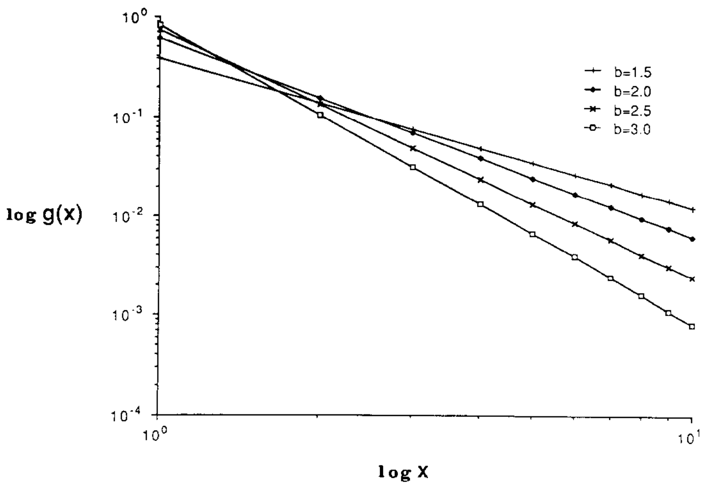

Fig. 2. Linearized Lotka curves for various values of b.

5. INEQUALITY

The Price law, while invalid as stated, represents an attempt to summarize in quantitative terms the inequality that characterizes the distribution of publication productivity. We know that such distributions are characterized by reverse J shape, pronounced positive skewness, and a long straggling tail. The average rate of productivity, m/t, is low whereas variation in productivity is high. Two opposing effects are simultaneously in evidence—dispersal of papers over many low-productivity authors and concentration of papers in a small number of highly prolific authors. The Price law attempts to summarize the nature of inequality under Lotka's inverse square model only and not quite accurately. Inequality is always present in author productivity distributions; however, the particular nature and degree of inequality varies. It is the failure of the Price law to take account of this variation that is its main deficiency.

Yablonsky [18] has stated that inequality is determined in general by the relations among the parameters b, k, and x_{max}. Table 5 clearly illustrates such a dependence. All three factors are determined, in turn, by extrinsic variables related to the discipline, sample size, measurement procedure, and so on. Thus, inequality is not a static phenomenon but varies systematically according to the characteristics of the author population at hand.

Some relations among these characteristics are evident from the values obtained in the simulation described earlier. In general, the relative contribution of $t^{0.5}$ decreases as b grows larger, and varies directly with x_{max}. The proportion of authors making only a single contribution increases rapidly as a nonlinear function of b, and their relative contribution to the total number of papers rises even faster (Table 6). At $b = 3$, more than half of the papers are dispersed over these minimum-productivity authors. Several investiga-

Table 6. Single contributors: relative proportion and contribution for various values of b

b	Proportion of single contributors (k)	Proportion of total papers ($g(1)m/m$)
1.5	.3828	.0231
2.0	.6079	.1472
2.5	.7454	.3397
3.0	.8319	.6108
3.5	.8875	.7461
4.0	.9239	.8321

tors [4,7,8] have noted the importance of the parameter b in characterizing such distributions and have suggested that b may be considered, to some extent, as a summary measure of inequality. Viewed as a type-token relationship, in which m tokens (papers) are generated by t types (authors), variation in the distribution of author productivity has the following character:

> The parameter k represents the number of types with a single token. The parameter b, to some extent, represents the dispersion of the distribution of tokens over types. The larger b the larger the number of types with only one token and the more rapid the decline, with increasing x, of the frequencies g(x). If b is small the numbers g(x) decline very slowly with x, and the distribution has a very long tail. Hence a higher proportion of the tokens are concentrated among a few highly productive types. As b approaches 0, all sizes approach an equiprobable state. [7]

The resulting effects on the average productivity, m/t, are that "high values for average productivity will be found in populations where the exponent is small and where, consequently, many of the tokens are distributed among the highly productive types. Low values of average productivity will be found in populations where the exponent is large and where, consequently, most tokens are distributed among types of low productivity" [7]. Depending on the particular combination of parameter values, therefore, Lotka's model implies not one, but a *family* of related skewed curves (Figs. 1 and 2) all displaying marked inequality, but in varying manners, according to whether dispersal or concentration of papers is the more striking feature.

The interrelationships of these characteristics are summarized in Table 7. Although such a characterization must be tentative at this point, Table 7 agrees well with empirical disciplinary evidence. The mean values of b in the 50 data sets considered here were 2.49 for scientific disciplines, 3.03 for social scientific disciplines, and 2.5 for the humanities. The mean values of k were 0.7244, 0.8160, and 0.7211, respectively. The similarity of natural science and humanities distributions is surprising; a continuum of parameter values corresponding to the "hardness" of the disciplines was perhaps expected. This result may be an artefact of the samples involved, but deserves further investigation. O'Connor and Voos [19] refer to data suggesting that average annual productivity for scientists is on an order of 3.8 articles per year, whereas for social scientists the rate is more like 0.5 papers per year.

6. CONCLUSIONS

Although the validity of the Price law need not depend on that of Lotka's law, the Price law is seen to be inconsistent with the generalized Lotka model. The Price law does not agree with empirical date very well; empirical results do not support the Price hypothesis. Since the empirical validity of Lotka's law has recently been more firmly established, it is not surprising that the empirical and theoretical findings are consistent.

Price articulated his law on the basis of a special case of the Lotka model, where $b = 2$. Inequality in author productivity is a fundamental, but not static, phenomenon—its nature varies according to parameter values that are determined in turn by characteristics of the particular author population at hand. Yablonsky [18] speaks of Lotka's model as a deterministic structure "filled" by a stochastic process. Several investigators [7,12,13] including Price himself [13] have pointed out that it is necessary to regard $x_{\max}$ as a ran-

Table 7. Interrelationships of productivity distribution characteristics

b	k	m/t	Tail	Discipline
small	small	high	longer	science, humanities
large	large	low	shorter	social sciences

dom variable depending on the underlying distribution and the sample size. The structure will be "filled" more evenly, particularly in the region of the tail of prolific contributors, as sample sizes increase. Even then, prolific authors are in some sense truly anomalous, and behavior in the tails of these distributions is likely to remain unreliable, as is the description of statistical patterns in the realm of very small numbers [13]. It is unlikely, therefore, that any convenient rule of thumb similar to the Price law could accurately and reliably summarize the size and relative contribution of the subset of highly prolific authors.

Acknowledgment—The author wishes to thank Dr. Jean Tague for advice on this study, Patti Bennett for generating the graphs reproduced in Figs. 1 and 2, and Elizabeth Walker for research assistance.

REFERENCES

1. Price, D. J. D. Little science, big science. New York: Columbia University Press; 1963.
2. Rescher, N. Scientific progress: A philosophical essay on the economics of research in natural science. Pittsburgh: University of Pittsburgh Press; 1978.
3. Lotka, A. J. The frequency distribution of scientific productivity. Journal of the Washington Academy of Science 16(12): 317-323; 1926.
4. Rao, I. K. R. Quantitative methods for library and information science. New Delhi: Wiley Eastern; 1983.
5. Nicholls, P. T. Empirical validation of Lotka's law. Information Processing & Management 22(5): 417-419; 1986.
6. Nicholls, P. T. Estimation of Zipf parameters. Journal of the American Society for Information Science 38(6): 443-445; 1987.
7. Tague, J. M.; Nicholls, P. T. The maximal value of a Zipf size variable: sampling properties and relationship to other parameters. Information Processing & Management 23(3): 155-157; 1987.
8. Chen, Y. S.; Leimkuhler, F. F. A relationship between Lotka's law, Bradford's law, and Zipf's law. Journal of the American Society for Information Science. 37(5): 307-314; 1986.
9. Potter, W. G. Lotka's law revisited. Library Trends. 31(2): 21-39; 1981.
10. Pao, M. L. An empirical examination of Lotka's law. Journal of the American Society for Information Science. 37(1): 26-33; 1986.
11. Nicholls, P. T. The Lotka hypothesis and bibliometric methodology. Ph.D. dissertation, School of Library and Information Science, University of Western Ontario, London, Canada; 1987.
12. Glänzel, W.; Schubert, A. Price distribution: An exact formulation of Price's "square root law." Scientometrics 7(3-6): 211-219; 1985.
13. Allison, P. D.; Price, D. J. D.; Griffith, D. B.; Moravcsik, M. J.; Stewart, J. A. Lotka's law: A problem in its interpretation and application. Social Studies of Science 6(2): 269-276; 1976.
14. Egghe, L.; Rousseau, R. A characterization of distributions which satisfy Price's law and consequences for the laws of Zipf and Mandelbrot. Journal of Information Science 12: 193-197; 1986.
15. Egghe, L. On the 80/20 rule. Scientometrics 10(1-2): 55-68; 1986.
16. Cole, J. R. Fair science: Women in the scientific community. New York: The Free Press; 1979.
17. Reskin, B. F. Scientific productivity and the reward structure of science. American Sociological Review 42: 491-504; 1977.
18. Yablonsky, A. I. On fundamental regularities of the distribution of scientific productivity. Scientometrics 2(1): 3-34; 1980.
19. O'Connor, D. O.; Voos, H. Empirical laws, theory construction and bibliometrics. Library Trends 31(2): 9-20; 1981.