Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht

Scientometrics, Vol. 75, No. 1 (2008) 175–188
DOI: 10.1007/s11192-007-1829-2

Personal name headings in COBIB: Testing Lotka's Law

MARIJA PETEK

Faculty of Arts at the University of Ljubljana
Department of Library and Information Science and Book Studies, Ljubljana (Slovenia)

The purpose of this article is to provide information about author productivity as reflected through the number of occurrences of personal name headings in the Slovenian online catalogue COBIB. Only authors associated with monographs are treated. So, author productivity of monographs that has not been widely researched is empirically examined to determine conformity or nonconformity to Lotka's law. A random sample of 1.600 Slovenian authors is drawn from the authority file CONOR. Next, the authors are searched in COBIB and each attributed the number of monographs. Using the formula: $x^n y = c$, the values of the exponent n and the constant c are computed and the Kolmogorov-Smirnov test is applied. The paper shows that the author productivity distribution predicted by Lotka also holds for the occurrences of personal name headings in COBIB.

Introduction

COBIB is the Cooperative Bibliographic database containing more than 3.085.000 bibliographic records (May 2007) for monographs, serials, non-book materials, electronic resources, component parts and other types of library material. It is part of COBISS, the Slovenian Cooperative Bibliographic System and Services implemented

Received June 15, 2007

Address for correspondence:
MARIJA PETEK
Faculty of Arts at the University of Ljubljana
Department of Library and Information Science and Book Studies
Aškerčeva 2, 1000 Ljubljana, Slovenia
E-mail: marija.petek@ff.uni-lj.si

0138–9130/US $ 20.00

in 1987. COBISS is an online national shared cataloguing system based on a network of about 330 cooperating libraries including two national libraries, public, academic, special and school libraries. The authority file for personal and corporate authors called CONOR was created in 2001 and included into COBISS in 2003. At the time of our study (April 2006) CONOR contained about 380.000 authority records, mostly for personal authors.

The purpose of this article is to provide information about author productivity as reflected through the number of occurrences of personal name headings in COBIB. We are interested only in authors associated with monographs. So, author productivity of monographs that has not been widely researched is empirically examined to determine conformity or nonconformity to Lotka's law. Testing is based on Lotka's methodology.

Lotka's law of scientific productivity

Lotka's law is one of the major laws ob bibliometrics. It is named for demographer and statistician Alfred J. Lotka (1880–1949) who in 1926 published his article "The Frequency Distribution of Scientific Productivity" in the *Journal of the Washington Academy of Sciences*. Examining productivity of authors in chemistry and physics LOTKA [1926] suggests that the number of authors making *n* contributions is about $1/n^2$ of those making one. About 60% of all authors in a given field create one publication, about 15% two publications, about 7% three, etc. A few authors are prolific and account for a relatively large percent of publications; many other authors produce only one or two publications. Lotka's law is also called an inverse square law or sometimes an inverse exponential law indicating that there is an inverse relation between the number of publications produced and the number of authors producing these publications.

Lotka's law is stated by the following formula: $x^n y = c$, where y is the number of authors making x contributions, the exponent n and the constant c are parameters to be estimated from a given set of author productivity data. The frequency of authors y making x contributions will be inversely proportional to some exponential function of x, whereas the inverse square relation in which the value of n equals 2 is viewed only as a special case [PAO, 1985]. But some authors mistakenly think that their data only fit Lotka's law if the exponent is exactly 2 [NEWBY & AL., 2003].

LOTKA (1926) examines journal articles in chemistry and physics of the late nineteenth and early twentieth centuries and derives two different values for n estimated from each empirical distribution. For chemistry he finds the parameter $c = 0.5669$ and $n = 1.888$; for physics $c = 0.6079$ and $n = 2.02$. Lotka collects his physics data from a bibliography; for chemistry he includes only authors whose surnames begin with letters A and B.

The parameters n and c have been conventionally estimated by the least-square method [PAO, 1985]. In some studies the maximum likelihood approach is applied

[NICHOLLS, 1986, 1989; ROUSSEAU & ROUSSEAU, 2000; KRETSCHMER & ROUSSEAU, 2001]. NEWMAN [2000] says that the maximum likelihood is a good method and that there is a tendency for least-squares fits to overestimate the slope of the power law since the statistical fluctuations in the logarithms of the data are greater in the downward direction than in the upward one.

There are three ways of counting multiple authorships: adjusted, complete and straight count [DIODATO, 1994]. The adjusted or fractional count means that if an article is written, e. g. by four authors each receives 1/4 credit. If each co-author receives full credit it is called the complete or normal count meaning that every author is counted fully whenever he or she appears. The straight or senior count means that all secondary authors are ignored; this way is used by Lotka himself, he assigns each article to the senior author only. PAO [1985] says that Lotka excludes all co-authors for unknown reasons. According to POTTER [1981] Lotka uses the senior count because multiple authorship is less common in Lotka's time. The complete count is presently more suitable measure in the author productivity [NICHOLLS, 1989].

Lotka's law is a general law that can be applied to many fields and collections of publications and their authors [DIODATO, 1994].

Review of previous works

There has been a considerable number of studies on the empirical validation of Lotka's law but their results cannot be always meaningfully compared owing to different methods applied.

VLACHY [1978] has compiled a comprehensive bibliography of Lotka's law and related phenomena. VLACHY [1972, 1974, 1976, 1980] presents a large variety of subjects and observes several variables which might influence the distribution of author productivity: time period under study, community of authors involved, types of creative communities, etc. He is interested in the value of the slope being the exponent in Lotka's law and not in the appropriateness of the law.

The applicability of Lotka's law to the authorship of journal articles has been examined in fields other than chemistry and physics. Findings of several studies support Lotka's law. PAO [1985] describes methods for testing the applicability of Lotka's law closely following procedures used by Lotka himself. She suggests how to compute values of the exponent n and the constant c and also how to perform the Kolmogorov-Smirnov test of conformity.

PAO [1986] examines 48 sets of author productivity data; for each she reports in detail number of authors, number of contributions, computed values of n and c, maximum deviation between the cumulative proportions of the observed and theoretical values, etc. Over 80% of the sets conform to Lotka's law. The sets cover twenty subject fields and three large research library catalogues. Lotka's original two sets of data on physics and chemistry are included, too; only the physics data fit the law. All data sets are also tested against the inverse square law in which *n* equals 2 and she finds out that only seven sets fit.

NICHOLLS [1989] investigates main elements in a bibliometric model and proposes a consistent methodology for applying Lotka's law using all author data without truncation and the maximum likelihood approach to estimate parameters. Following Nicholls' methodology ROUSSEAU & ROUSSEAU [2000] present a computer program for fitting the power law distribution called LOTKA; some drawbacks of Pao's methodology are reported such as the fact that the least-square approach gives acceptable results only if author data are truncated.

ROWLANDS [2005] shows that Lotka's law also holds for Emerald authors; 65% publish one article, 15.50% two articles, 6.70% three, etc. Examining author loyalty he tries to answer the question which authors are most likely to return to a particular publisher.

Lotka's law is also used to test productivity of software developers in open source systems [NEWBY & AL., 2003]. Programmers are considered authors and software is considered a publication; findings correspond to prior studies of scientific publishing.

Lotka's law seems to be very resilient feature of intellectual productivity in many different fields. However, a study on high-energy physics is an exception showing that Lotka's law breaks down [KRETSCHMER & ROUSSEAU, 2001]. In this field large groups of authors almost always collaborate and such groups occasionally work together in mega-projects leading to publications with more than 100 authors; the normal count procedure is applied.

Monograph productivity

Some studies have been conducted to examine the frequency of occurrences of name headings in library catalogues. According to POTTER [1980] the frequency of name headings in the catalogue is amenable to Lotka's law meaning that a certain proportion of names (about 60%) in the catalogue is unique representing authors who write only one publication and the rest of the names (about 30%) representing authors of more than one publication.

POTTER [1981] presents an overview of literature on Lotka's law and reports that the majority of studies deals with authors of journal articles; monograph productivity has not been widely investigated. There are only a few researches considering monographs: University of Illinois, University of Wisconsin, Library of Congress, Indiana University Music Library and University of Chicago; they are all based on occurrences of name headings in the catalogue. In all cases but one it is found that about two-thirds of all authors have only one entry in the catalogue.

POTTER [1980] conducts a study of personal authors in two card catalogues and discovers that at the University of Illinois 63.50% of all names occur once and at the University of Wisconsin 69.33%. The University of Illinois data fit Lotka's law, the University of Wisconsin data do not.

MCCALLUM & GODWIN [1981] examine all author headings on MARC tapes of The Library of Congress and find that 65.65% authors appear once in the catalogue; the data do not fit Lotka's law.

PAPAKHIAN [1985] replicates Potter's study in two card catalogues of the Indiana University Music Library: one represents printed materials (monographs, scores and serials) and the other sound recordings (tapes, cassettes, etc.). In the printed materials catalogue he discovers 61.23% of name headings occurring only once, but in the sound recordings catalogue the proportion is relatively small: 47.64%. According to PAPAKHIAN [1985] this might be a result of specific characteristics of sound recordings collections, it might be also the case that performers produce more recordings than authors publish books. The distribution of the printed materials catalogue matches Lotka's distribution, but that of the sound recordings catalogue does not.

Studying the card catalogue of the University of Chicago FULLER [1989] finds that 61.90% authors are responsible for a single entry; this proportion approaches Lotka's law of scientific productivity even closer.

Methodology

The purpose of this study is to replicate Lotka's original calculation. In order to make comparisons with Lotka's work, his method is followed as closely as possible. That is why, the standard testing methodology described by PAO [1985] is used. It implies the senior author count, the most prolific authors are excluded and parameters are estimated by the square-least method. Finally, a statistical test of goodness-of-fit is performed.

The study has four steps: data collection, estimation of the exponent n, calculation of the constant *c* and the Kolmogorov-Smirnov test.

Data collection

In April 2006 a random sample of 1.600 Slovenian personal authors was drawn from the authority file CONOR. Authors whose surnames began with letters A, C, D, H, N, O, R, S, T and Ž were chosen for the sample to provide an alphabetic spread. Then, these authors were searched in the Slovenian online catalogue COBIB to collect data on the number of their occurrences. We are interested only in authors who create monographs in print form, the authors of articles and other types of library materials are not included. The final sample consists of 1.114 personal authors; in case of multiple authorships the senior author count is used. Collected data are arranged into frequency distributions of monographs and authors producing them.

Estimation of the exponent n

The value of *n* is calculated by the least square-method using the following formula [PAO, 1985]:

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

N = number of pairs of data
X = logarithm of x, i. e. number of monographs
Y = logarithm of y, i. e. number of personal name headings

The least-square method is used to estimate the best value for the slope of a regression line which is the exponent *n* for Lotka's law [PAO, 1985]. The slope is usually calculated without data points representing authors of high productivity. Lotka suggests that highly productive authors be considered separately and cuts off his two data sets at the first 17 points for physics and at 30 points for chemistry distribution, his cutoff is determined visually. The best cutoff can be computed also by the formula (Pao, 1985): $\sqrt{\sum y_x}$, where $\sum y_x$ is the total sample.

Since values of the slope change with different number of points for the same set of data, we have made several computations of n. The median or the mean values of *n* can also be identified as the best slope for the observed distribution [PAO, 1985]. Different values of *n* produce different values of the constant c.

Calculation of the constant c

The estimate of the constant *c* is more problematic [PAO, 1985]. If we accept Lotka's conclusion that the proportion of all authors making a single contribution is about 60%, then the value of *c* can be computed by the simple formula $6/\pi^2$. If *n* equals 2, *c* is the inverse of the summation of the infinite series: $\Sigma(1/x^2)$, the limit of each equals to $\pi^{2/}6$. For other non-negative fractional values of n, the summation of the series can be approximated by a function calculating the sum of the first *P* terms. It is found that the error is negligible if *P* is set to 20.

The value of the constant c is calculated using the following formula [PAO, 1985]:

$$c = \frac{1}{\left[\sum_{1}^{P-1} \frac{1}{x^n} + \frac{1}{(n-1)(P^{n-1})} + \frac{1}{2P^n} + \frac{n}{24(P-1)^{n+1}}\right]}$$

$\sum_{1}^{P-1} \frac{1}{x^n}$ = obtained by summing the first 19 terms of $\frac{1}{x^n}$,

with x = 1, 2, 3, … 19

$P = 20$
n = value obtained in step 2
x = number of monographs

Kolmogorov-Smirnov test

PAO [1985] suggests the Kolmogorov-Smirnov test, a goodness-of-fit statistical test to assert that the observed author productivity distribution is not significantly different from a theoretical distribution. The hypothesis concerns a comparison between observed and expected frequencies. The test allows the determination of the associated probability that the observed maximum deviation occurs within the limits of chance. The maximum deviation between the cumulative proportions of the observed and theoretical frequency is determined by the following formula [PAO, 1985]:

$$D = \text{Max} \left| F_o(x) - S_n(x) \right|$$

$F_o(x)$ = theoretical cumulative frequency
$S_n(x)$ = observed cumulative frequency

The test is performed at the 0.05 or at the 0.01 level of significance. When sample size is greater than 35, the critical value of significance is calculated by the following formula [PAO, 1985]:

The critical value at the 0.05 level of significance = $\frac{1.36}{\sqrt{\sum y}}$

The critical value at the 0.01 level of significance = $\frac{1.63}{\sqrt{\sum y}}$

$\sum y$ = the total population under study

If the maximum deviation falls within the critical value the null hypothesis that the data set conforms to Lotka's law can be accepted at a certain level of significance. But if it exceeds the critical value the null hypothesis must be rejected at a certain level of significance and concluded that the observed distribution is significantly different from the theoretical distribution.

Findings and discussion

Findings of our study are grouped into four parts: data collection, estimation of the exponent n, calculation of the constant c, and the Kolmogorov-Smirnov test.

Data collection

The random sample consists of 1.114 Slovenian personal authors. Table 1 shows data on frequency distribution of personal name headings in COBIB according to the number of associated monographs. The senior author count is used.

Table 1. Frequency distribution of personal name headings in COBIB

No. of monographs	No. of personal name headings	% Personal name headings	No. of bibliographic records
1	766	68.76	766
2	172	15.44	344
3	69	6.19	207
4	28	2.51	112
5	19	1.71	95
6	6	0.54	36
7	6	0.54	42
8	7	0.63	56
9	7	0.63	63
10	6	0.54	60
11	1	0.09	11
12	5	0.45	60
13	2	0.18	26
15	1	0.09	15
16	4	0.36	64
17	4	0.36	68
18	1	0.09	18
20	1	0.09	20
21	1	0.09	21
22	1	0.09	22
25	1	0.09	25
27	1	0.09	27
32	1	0.09	32
41	1	0.09	41
45	1	0.09	45
94	1	0.09	94
542	1	0.09	542
Total	1114	100.01	2912

The majority of authors, i. e. 766 or 68.76%, occurs only once in the catalogue meaning that these authors are associated with only one bibliographic record representing monographs in print form. 172 authors or 15.44% appear twice, 69 authors or 6.19% occur three times; 28 authors or 2.51% occur four times, etc. The most prolific author found in the sample is our greatest writer Ivan Cankar who is responsible for 542 bibliographic records in COBIB.

Similar proportions are found on MARC tapes of The Library of Congress: 65.65% authors appear once, 17.22% twice, 6.65% three times, etc. [MCCALLUM & GODWIN, 1981]. In the catalogue of the University of Illinois 63.50% authors appear once, 14.63% twice, 6.82% three times, etc.; in the catalogue of the University of Wisconsin 69.33% authors occur once, 14.95% twice, 6.26% three times [POTTER, 1980]. FULLER [1989] finds a similar share: 61.90% authors in the card catalogue of the University of Chicago are responsible for a single contribution. In the printed materials catalogue PAPAKHIAN [1985] discovers 61.23% of name headings occurring only once, but in the sound recordings catalogue 47.64%.

Estimation of the exponent n

Table 2 shows data for calculation of the exponent *n* for COBIB for 10 pairs of data.

The estimated value of *n* for COBIB is 2.2656 and it is calculated from the observed distribution using the formula:

$$n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - (\sum X)^2} = \frac{10(6.9266) - (6.5598)(13.7093)}{10(5.2152) - (6.5598)^2} = -2.2656$$

Table 2. Calculation of the exponent *n* for COBIB

x	*y*	*X*	*Y*	*XY*	X^2
1	766	0.0000	2.8842	0.0000	0.0000
2	172	0.3010	2.2355	0.6729	0.0906
3	69	0.4771	1.8388	0.8773	0.2276
4	28	0.6021	1.4472	0.8714	0.3625
5	19	0.6990	1.2788	0.8939	0.4886
6	6	0.7782	0.7782	0.6056	0.6056
7	6	0.8451	0.7782	0.6577	0.7142
8	7	0.9031	0.8451	0.7632	0.8156
9	7	0.9542	0.8451	0.8064	0.9105
10	6	1.0000	0.7782	0.7782	1.0000
		6.5598	13.7093	6.9266	5.2152

x = number of monographs; *y* = number of personal name headings; *X* = logarithm of *x*; *Y* = logarithm of *y*

To estimate the best value for *n* several computations are made using the least-square method. Table 3 shows different sample values of *n* for COBIB according to different values of number of pairs. It is seen that the data tend to fluctuate after the 10th point.

Table 3. Sample values of the exponents *n* for COBIB

N	*n*
9	2.3570
10	2.2656
11	2.4321
12	2.3129
13	2.3119
14	2.3525
15	2.9484
16	2.1137
17	2.1359
18	2.1357
19	2.1277
20	2.1142
21	2.0849
22	2.0464
23	1.9896
24	1.9023
25	1.8208
26	1.6265
27	1.1821

N = number of pairs of data

Calculation of the constant c

The calculated value of the constant *c* for COBIB is 0.6890; the value of *n* obtained in step 2 is used:

$$c = \frac{1}{\left[\sum_{1}^{P-1}\frac{1}{x^n} + \frac{1}{(n-1)(P^{n-1})} + \frac{1}{2P^n} + \frac{n}{24(P-1)^{n+1}}\right]}$$

$$\sum\frac{1}{x^{2.2656}} = \sum_{1}^{19}\frac{1}{x^{2.2656}} + \frac{1}{1.2656(20^{1.2656})} + \frac{1}{2(20^{2.2656})} + \frac{2.2656}{24(19^{3.2656})}$$

$$\sum\frac{1}{x^{2.2656}} = 1.4330 + \frac{1}{1.2626(44.3183)} + \frac{1}{2(886.3670)} + \frac{2.2656}{24(14993.3182)}$$

$$\sum\frac{1}{x^{2.2656}} = 1.4330 + \frac{1}{56.0892} + \frac{1}{1772.734} + \frac{2.2656}{359839.6368}$$

$$\sum \frac{1}{x^{2.2656}} = 1.4330 + 0.0178 + 0.0006 + 0.0000$$

$$\sum \frac{1}{x^{2.2656}} = 1.4514$$

$$c = \frac{1}{1.4514} = 0.6889899 = 0.6890$$

Kolmogorov-Smirnov test

The Kolmogorov-Smirnov test is performed to test the conformity of the observed COBIB distribution versus Lotka's distribution. The obtained data are first tested against the inverse square law with the exponent *n* being 2 and presented in Table 4.

Table 4. Kolmogorov-Smirnov test, $n = 2$

x	Lotka	$F_o(x)$	COBIB	$S_n(x)$	D
1	0.6079	0.6079	0.6876	0.6876	0.0797
2	0.1520	0.7599	0.1544	0.8420	0.0821
3	0.0675	0.8274	0.0619	0.9039	0.0765
4	0.0380	0.8654	0.0251	0.9290	0.0636
5	0.0243	0.8897	0.0171	0.9461	0.0564
6	0.0169	0.9066	0.0054	0.9515	0.0449
7	0.0124	0.9190	0.0054	0.9569	0.0379
8	0.0095	0.9285	0.0063	0.9632	0.0347
9	0.0075	0.9360	0.0063	0.9695	0.0335
10	0.0061	0.9421	0.0054	0.9749	0.0328

x = number of monographs; $F_o(x)$ = theoretical cumulative frequency; $S_n(x)$ = observed cumulative frequency; D = maximum deviation

The maximum deviation found is 0.0821 which exceeds the critical value of 0.0488 at the 0.01 level of significance. Therefore, the null hypothesis must be rejected and concluded that the COBIB data do not fit Lotka's law.

$$D = \text{Max} \left| F_o(x) - S_n(x) \right| = 0.0821$$

The critical value at the 0.01 level of significance = $\frac{1.63}{\sqrt{1114}} = \frac{1.63}{33.3766} = 0.0488$

$$D > 0.0488$$

The observed data from COBIB are also tested for conformity with the expected values computed by Lotka's law. The Kolmogorov-Smirnov test is applied with the values of n and c derived from steps 2 and 3. Table 5 presents the test with $n = 2.2656$ and $c = 0.6890$. The maximum deviation found is 0.0144 which falls within the critical value of 0.0488 at the 0.01 level of significance. Therefore, the null hypothesis that the observed distribution is not significantly different from the expected distribution, cannot be rejected suggesting that the COBIB data fit Lotka's law.

Table 5. Kolmogorov-Smirnov test, $n = 2.2656$

x	y	COBIB (observed)	$S_n(x)$	COBIB (expected)	$F_o(x)$	D
1	766	0.6876	0.6876	0.6890	0.6890	0.0014
2	172	0.1544	0.8420	0.1433	0.8323	0.0097
3	69	0.0619	0.9039	0.0572	0.8895	0.0144
4	28	0.0251	0.9290	0.0298	0.9193	0.0097
5	19	0.0171	0.9461	0.0180	0.9373	0.0088
6	6	0.0054	0.9515	0.0119	0.9492	0.0023
7	6	0.0054	0.9569	0.0084	0.9576	0.0007
8	7	0.0063	0.9632	0.0062	0.9638	0.0006
9	7	0.0063	0.9695	0.0048	0.9686	0.0009
10	6	0.0054	0.9749	0.0037	0.9723	0.0026

x = number of monographs; y = number of personal name headings; $S_n(x)$ = observed cumulative frequency; $F_o(x)$ = theoretical cumulative frequency; D = maximum deviation

The column "COBIB (expected)" contains values computed by: $c(\frac{1}{x^n})$

$$D = \text{Max } |F_o(X) - S_n(X)| = 0.0144$$

The critical value at the 0.01 level of significance = $\frac{1.63}{\sqrt{1114}} = \frac{1.63}{33.3766} = 0.0488$

$$D < 0.0488$$

Conclusion

Lotka's law of scientific productivity is regarded one of classical laws of bibliometrics. An important area of bibliometrics is author productivity, which can be measured by a number of publications a particular author has. In a given field only a few authors are prolific and account for a large amount of publications; many other authors representing about 60% create a single contribution. Lotka's law has been used in chemistry, physics and various other fields; not only the authorship of journal articles but also the authorship of monographs and occurrences of author names in library catalogues are treated.

The frequency distribution of personal name headings in COBIB is examined to determine conformity or nonconformity to Lotka's law. A random sample of Slovenian authors is drawn from the authority file CONOR, then these authors are searched in COBIB and each attributed the number of monographs. In case of multiple authorships only senior authors are treated, co-authors are excluded. It is found that 68.76% of personal authors occur only once in COBIB meaning that these authors are associated with only one bibliographic record representing monographs in print form. 15.44% of authors appear twice, 6.19% three times, etc. The most prolific author found in the sample is responsible for 542 monographs.

Following Lotka's methodology the value of the exponent *n* for COBIB is estimated 2.2656 and the constant *c* computed 0.6890. Using the Kolmogorov-Smirnov test it is found that at the 0.01 level of significance the maximum deviation is 0.0144 which falls within the critical value of 0.0488. Therefore, it can be concluded that the data on occurrences of personal name headings in COBIB fit Lotka's law. The observed distribution in COBIB is also tested against the inverse square law in which the exponent *n* equals 2; it is found out that the data do not conform to Lotka's law.

The paper shows that the author productivity distribution predicted by Lotka also holds for the occurrences of personal name headings in COBIB. These findings can be used in designing authority control systems and introducing new cataloguing codes. The name authority file can be reduced by creating name authority records only for headings which occur more than once and for headings which require references. The impact of new rules on changing forms of headings is greater in catalogues with a large proportion of headings appearing more than once.

Future research

It would be interesting to examine our observed set of author data using some variations in methodology, such as the maximum likelihood approach instead of the least-square method or the complete or adjusted author count instead of the senior count, and compare the results.

Besides further examining personal name headings in COBIB it is also necessary to study occurrences of corporate name headings. The study can be repeated for various types of library materials, such as electronic resources, non-book materials, printed music, cartographic materials, etc.

Investigating author productivity can be extended by analysing variables such as chronological and academic age, number and frequency of previous publications, access to research grants, job status, etc. In such a way characteristics of high, medium and low publishing activity of authors can be identified.

References

DIODATO, V. P. (1994), *Dictionary of Bibliometrics*. The Haworth Press, New York.

FULLER, E. E. (1989), Variation in personal names in works represented in the catalog. *Cataloging & Classification Quarterly*, 9 (3) : 75–95.

KRETSCHMER, H., ROUSSEAU, R. (2001), Author inflation leads to a breakdown of Lotka's law. *Journal of the American Society for Information Science and Technology*, 52 (8) : 610–614.

LOTKA, A. J. (1926), The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16 (12) : 317–323.

MCCALLUM, S. H., GODWIN, J. L. (1981), Statistics on headings in the MARC File. *Journal of Library Automation*, 14 (3) : 194–201.

NEWBY, G. B., GREENBERG, J., JONES, P. (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.

NEWMAN, M. E. J. (2000), Comments to the article by Rousseau & Rousseau. *CYBERmetrics*, 4 (1) : [1–2].

NICHOLLS, P. T. (1986), Empirical validation of Lotka's law. *Information Processing and Management*, 22 : 417–419.

NICHOLLS, P. T. (1989), Bibliometric modelling processes and the empirical validity of Lotka's law. *Journal of the American Society for Information Science*, 40 (6) : 379–385.

PAO, M. L. (1985), Lotka's law: A testing procedure. *Information Processing & Management*, 21 (4) : 305–320.

PAO, M. L. (1986), An empirical examination of Lotka's Law. *Journal of the American Society for Information Science*, 37 (1) : 26–33.

PAPAKHIAN, A. R. (1985), The frequency of personal name headings in the Indiana university music library card catalogs. *Library Resources & Technical Services*, 29 : 273–285.

POTTER, W. G. (1980), When names collide: Conflict in the catalog and AACR2. *Library Resources & Technical Services*, Winter 1980 : 3–16.

POTTER, W. G. (1981), Lotka's law revisited. *Library Trends*, 30 (1) : 21–39.

POTTER, W. G. (1988), *Of Making Many Books There is No End: Bibliometrics and Libraries. Libraries & Computing Centers: Issues of Mutual Concern from the Publishers of the Journal of Academic Librarianship, 9.* September 1988, [1–3].

ROUSSEAU, B., ROUSSEAU, R. (2000), LOTKA: A program to fit a power law distribution to observed frequency data. *CYBERmetrics*, 4 (1) paper 4 : [1–6].

ROWLANDS, I. (2005), Emerald authorship data, Lotka's law and research productivity. *Aslib Proceedings: New Information Perspectives*, 57 (1) : 5–10.

SEN, B. K., TAIB, C. A. B., HASSAN, M. F. B. (1996), Library and information science literature and Lotka's law. *Malaysian Journal of Library & Information Science*, 1 (2) : 89–93.

VLACHY, J. (1972), Variable factors in scientific communities (observations on Lotka's law). *Teorie a Metoda*, 4 : 91–120.

VLACHY, J. (1974), *Distribution Patterns in Creative Communities.* World Congress of Sociology, Toronto, August 19–24 1974, 1–20.

VLACHY, J. (1976), Time factor in Lotka's law. *Probleme de Informare si Documentare*, 10 : 44–87.

VLACHY, J. (1978), Frequency distributions of scientific performance: A bibliography of Lotka's law and related phenomena. *Scientometrics*, 1 : 109–130.

VLACHY, J. (1980), Evaluating the distribution of individual performance. *Scientia Yugoslavica*, 6 (1–4) : 267–275.