Lotka's Law and Authorship distribution pattern in Global Synthetic Biology Literature

NAHEEM KT Mr
Pondicherry University, naheemkt@gmail.com

Sivaraman P Dr
Annamali University, psraman.p@gmail.com

Saravanan G Dr
French Institute of Pondicherry, saravanan.g@ifpindia.org

Lotka's Law and Authorship distribution pattern in Global Synthetic Biology Literature

Naheem K.T.,[1]Research Scholar, Dept. of Library & Information Science, Annamalai University, Annamalai Nagar – 608002.; [1]Professional Assistant, Central Library, Pondicherry University, Kalapet, Puducherry-605014. Email: naheemkt@gmail.com;

Dr P. Sivaraman[2]., Professor and Coordinator, Library & Information Science Wing, DDE, Annamalai University, Annamalai Nagar – 608002. Email: psraman.p@gmail.com

Dr G. Saravanan[3]., Librarian, Library and Publication Division, French Institute of Pondicherry. Email: saravanan.g@ifpindia.org

ABSTARCT

An attempt is made to examine the authorship distribution in Synthetic Biology (SB) literature and to validate Lotka's law of author productivity. Authors obtained data for this study from the WOS database. A total of 12012 papers with 33151 unique authors has identified , and used for further analysis. Authors calculated the exponents "n" and "c". Researchers employed Kolmogorov-Smirnov (K-S) test of goodness-of-fit to verify the validity of Lotka's Law in SB literature. The results of this study proved that Lotka's Law of author productivity does fit with SB literature based on the calculated values n = -2.45 and c= 0.74.

Keywords: Lotka's law, Synthetic biology, Authorship distribution, Kolmogorov-Smirnov test, Bibliometrics**,** Scientometrics.

1. INTRODUCTION

Quantitative evaluation of research contributions is one of the most significant acts among the academic community in the modern world. Such evaluative studies help to identifies key and trendy areas in different areas as well as it helps to spot out the key players like productive authors, institutions and countries in various scientific fields. Bibliometric or scientometric methodologies had used for research evaluation. Lotka's Law is one of the prominent and well-known law of Bibliometric studies, which focused on the frequency of publication by authors in any specified field[1]. It is the most frequently used bibliometric Law used to determine the authorship distribution patterns in diverse scientific fields. In 1926 Alfred

J. Lotka[2] proposed a law, i.e., "inverse square law" indicating that there is an inverse relationship between the number of publications and the number of authors producing these publications. Lotka's Law describes the frequency of publication by authors in a given subject field. It stated that:

> "the number of authors making n contributions is about 1/of those making one, and the proportion of all contributors that make a single contribution is about 60 per cent. Which means that out of all the authors in a given subject field, about 60% publish only one article, 15% publish two papers, 7% publish three, and so on. According to Lotka's Law, only 6% of authors in a subject field produce more than ten articles. We can express the generalised form of Lotka's Law as "$x^n y = c$",where y is the number of authors with x articles, the exponent n and constant c are parameters to be estimated from a given set of author productivity data".

Several researchers have studied the validity of Lotka's Law in various subjects. Most notable of them by Pao[3] and Nicholls[4], who found that the Lotka model fitted the majority of the data sets studied. The present study is an effort to verify and test the authorship distribution in the literature on Synthetic Biology using Lotka's Law based on the publication data obtained WOS database during 2005-2019.

2. LITERATURE REVIEW

So far numerous studies are available on author productivity of diverse subjects. Due to the availability of these multiple studies, here present some of the recent studies, mostly concentrating application and validation of Lotka's Law.

Shilpa Dhoble and Sudhir Kumar[5] applied Lotka's Law to Mustard research output of India. Results of this analysis proved that Lotka's Law is admissible to these data set. Similarly, Aidi Ahmi et al, [6] examined the legality of Lotka's Law in "Xtensible Business Reporting Language" (XBRL) literature. Results of this study also exposed that Lotka's Law well fit the literature on XBRL studies.

Zabed and Rahman[7] inspected the legitimacy of Lotka's Law to authorship distribution in the literature on Nutrition Research in Bangladesh. The results of this study showed that Lotka's generalised inverse square law does not apply to Nutrition Research of Bangladesh. But after removing most productive authors Lotka's Law was found to fit to sample data. Sudheer[8] verified the applicability of Lotka's Law to physics research output and reported that Lotka's Law doesn't find fit to physics literature.

Askew[9] selected Library and Information Studies literature for testing the applicability of Lotka's Law and reported that Lotka's law does fit with in this field . Similarly, Suresh Kumar[10] verified the authorship frequency distribution on 2106 publications references of the Journal of Documentation published during 2003-2015 using K-S test and concluded that Lotka's Law fit to the data set of LIS literature. Kumar and Senthilkumar[11] selected Astronomy & Astrophysics literature and tested applicability of Lotka's Law. Reported that Lotka's Law doesn't fit to this area. In another study Sevukan and Sharma[12] found that the Lotka's Law fits the data of Biotechnology literature. Naqvi and Fatima[13] used Lotka's Law to test the authorship distribution in international business literature and found that Lotka's distribution applies to international business literature. In a recent paper Kumar[14] examined the distribution of author productivity in the Human-Computer Interaction (HCI) literature and reported that the conformity of Lotka's Law in this field. From the above-reported studies, it had found that Lotka's Law applies to some subjects in its modified form and for some topics, it's not applicable.

3. OBJECTIVES

1. To examine the validity of Lotka's Law in modified form using the complete count of authors; and
2. To apply Kolmogorov- Smirnov(K-S) goodness-of-fit test for the conformity of Lotka's Law.

3. METHODOLOGY

The following methodology suggested by Pao[3] is adopted to test the applicability of Lotka's Law to Synthetic Biology literature.

1. Data collection;
2. Frequency distribution of authors and contributions;
3. Calculation of ‘n’;

4. Calculation of ‘c’;

5. Kolmogorov-Smirnov (K-S) test of goodness-of-fit of Lotka’s theoretical distribution; and

6. Calculation of critical value.

3.1 Data collection

Authors retrieved data for this study from Clarivate analytics WOS Science Citation Index Expanded database. All records related to Synthetic Biology published during the last 15 years (2005-2019) had identified using a well structured advanced search strategy. The same records have been retrieved and used for further analysis. The data set consists of a total of 12012 records contributed by 33151 unique authors. Researchers considered the complete count of authors for the present study.

3.2 Frequency distribution of authors and contributions

Authorship distribution data along with number of publications should be arranged in a table with first two columns containing values of x (publications) and y (frequency of authors) in increasing order of productivity (Table 1).

3.3 Calculation of exponent *“n”*

The value of *n* is calculated by the least square method using the following formula:

$$n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - \left(\sum X\right)^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(1)}$$

Where N = number of pairs of data; *X* = logarithm of x, i. e. number of articles*; Y* = logarithm of y, i. e. number of authors.

3.4 *Calculation of “c”*

The value of *c* is estimated using the following formula:

$$\sum \frac{1}{x^n} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2)}$$

3.5 Kolmogorov-Smirnov (K-S) test

K-S test, a goodness-of-fit statistical test is used to find out the significant difference between the observed and theoretical distribution of authors. Formula:

$$D = \max\left|F_0(x) - S_n(x)\right| \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(3)}$$

Where $F_0(x)$ = theoretical cumulative frequency and $S_n(x)$ = observed cumulative frequency

3.6 Calculation of critical value.

The critical value at the 0.01 level of significance has calculated using the following equation suggested by Nichols [3]:

$$=1.63 /[\sum yx +(\sum yx/10)^{1/2}]^{1/2} \quad \ldots\ldots (4)$$

$\sum yx$ = the total population under study

5. RESULTS AND DISCUSSIONS

Table 1 shows the frequency distribution of author productivity in the field of SB. The number of articles contributed by each author is arranged in increasing order of productivity. It is evident from the table that a total of 33151 unique authors had made contributions in SB during the period under study. Of the 33151 individual authors, 24716 (74.56 %) produced one article each, 4353 (13.13 %) had two papers each, 1720 (5.19%) produced three articles each, and so on. While comparing to this, only a few authors contributed ten or more items.

Table 1 : Frequency distribution of author productivity in SB Literature

No. of pairs	No. of articles	No. of authors observed	% of authors observed
1	1	24716	74.56
2	2	4353	13.13
3	3	1720	5.19
4	4	763	2.30
5	5	446	1.35
6	6	280	0.84
7	7	175	0.53

8	8	128	0.39
9	9	118	0.36
10	10	78	0.24
11	11	53	0.16
12	12	50	0.15
13	13	32	0.10
14	14	34	0.10
15	15	22	0.07
16	16	27	0.08
17	17	16	0.05
18	18	19	0.06
19	19	13	0.04
20	20	10	0.03
21	21	9	0.03
22	22	11	0.03
23	23	4	0.01
24	24	8	0.02
25	25	1	0.00
26	26	3	0.01
27	27	3	0.01
28	28	3	0.01
29	29	3	0.01
30	30	3	0.01
31	31	4	0.01
32	32	6	0.02
33	33	3	0.01
34	34	4	0.01
35	36	1	0.00
36	37	4	0.01
37	38	1	0.00
38	39	3	0.01
39	40	2	0.01
40	41	3	0.01
41	42	1	0.00
42	43	2	0.01
43	44	2	0.01
44	45	1	0.00
45	48	3	0.01
46	50	2	0.01
47	52	1	0.00
48	53	2	0.01

49	54	1	0.00
50	57	1	0.00
51	62	1	0.00
52	80	1	0.00
53	118	1	0.00
		33151	100.00

Calculation of "n" and "c"

To determine the value of n, the Linear Least Square (LLS) method is used. Authors used the data given in column 1 to 7 of Table 2 for calculating the value of exponent "*n*". By substituting the values in the equation (1), the value of *n* has calculated as:

$$n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - \left(\sum X\right)^2} = \frac{53(49.6410) - (70.60987)(53.2537)}{53(102.7957) - (70.60987)^2} = -2.45$$

n = -2.45

The next step in the application of Lotka's Law is to determine the value of 'c'. For which data given in column 9 in Table 2 is used. By replacing the values in the equation (2), value of "*c*" is obtained as:

$c = \sum 1/ x^n$
$= \sum 1/ 1.35917 = 0.735744 = 0.74$
$C = 0.74$

Table 2. Calculation of "n" and "c"

No. of pairs	x	y	X=log x	Y=Log y	XY	X^2	x^n	$1/x^n$
Col.1	*Col.2*	*Col.3*	*Col.4*	*Col.5*	*Col.6*	*Col.7*	*Col.8*	*Col.9*
1	1	24716	0.0000	4.3930	0.0000	0.0000	1.0000	1.00000

2	2	4353	0.3010	3.6388	1.0954	0.0906	5.4642	0.18301
3	3	1720	0.4771	3.2355	1.5437	0.2276	14.7553	0.06777
4	4	763	0.6021	2.8825	1.7355	0.3625	29.8571	0.03349
5	5	446	0.6990	2.6493	1.8518	0.4886	51.5794	0.01939
6	6	280	0.7782	2.4472	1.9043	0.6055	80.6252	0.01240
7	7	175	0.8451	2.2430	1.8956	0.7142	117.6224	0.00850
8	8	128	0.9031	2.1072	1.9030	0.8156	163.1438	0.00613
9	9	118	0.9542	2.0719	1.9771	0.9106	217.7179	0.00459
10	10	78	1.0000	1.8921	1.8921	1.0000	281.8383	0.00355
11	11	53	1.0414	1.7243	1.7956	1.0845	355.9689	0.00281
12	12	50	1.0792	1.6990	1.8335	1.1646	440.5488	0.00227
13	13	32	1.1139	1.5051	1.6767	1.2409	535.9956	0.00187
14	14	34	1.1461	1.5315	1.7553	1.3136	642.7080	0.00156
15	15	22	1.1761	1.3424	1.5788	1.3832	761.0682	0.00131
16	16	27	1.2041	1.4314	1.7235	1.4499	891.4438	0.00112
17	17	16	1.2304	1.2041	1.4816	1.5140	1034.1889	0.00097
18	18	19	1.2553	1.2788	1.6052	1.5757	1189.6457	0.00084
19	19	13	1.2788	1.1139	1.4245	1.6352	1358.1456	0.00074
20	20	10	1.3010	1.0000	1.3010	1.6927	1540.0098	0.00065
21	21	9	1.3222	0.9542	1.2617	1.7483	1735.5506	0.00058
22	22	11	1.3424	1.0414	1.3980	1.8021	1945.0715	0.00051
23	23	4	1.3617	0.6021	0.8198	1.8543	2168.8684	0.00046
24	24	8	1.3802	0.9031	1.2465	1.9050	2407.2298	0.00042
25	25	1	1.3979	0.0000	0.0000	1.9542	2660.4373	0.00038
26	26	3	1.4150	0.4771	0.6751	2.0021	2928.7662	0.00034
27	27	3	1.4314	0.4771	0.6829	2.0488	3212.4858	0.00031
28	28	3	1.4472	0.4771	0.6905	2.0943	3511.8599	0.00028
29	29	3	1.4624	0.4771	0.6977	2.1386	3827.1466	0.00026
30	30	3	1.4771	0.4771	0.7048	2.1819	4158.5993	0.00024
31	31	4	1.4914	0.6021	0.8979	2.2242	4506.4665	0.00022
32	32	6	1.5051	0.7782	1.1712	2.2655	4870.9923	0.00021
33	33	3	1.5185	0.4771	0.7245	2.3059	5252.4164	0.00019
34	34	4	1.5315	0.6021	0.9220	2.3454	5650.9745	0.00018
35	36	1	1.5563	0.0000	0.0000	2.4221	6500.4156	0.00015
36	37	4	1.5682	0.6021	0.9442	2.4593	6951.7513	0.00014
37	38	1	1.5798	0.0000	0.0000	2.4957	7421.1261	0.00013
38	39	3	1.5911	0.4771	0.7591	2.5315	7908.7580	0.00013
39	40	2	1.6021	0.3010	0.4823	2.5666	8414.8616	0.00012
40	41	3	1.6128	0.4771	0.7695	2.6011	8939.6486	0.00011
41	42	1	1.6232	0.0000	0.0000	2.6349	9483.3278	0.00011
42	43	2	1.6335	0.3010	0.4917	2.6682	10046.1051	0.00010

43	44	2	1.6435	0.3010	0.4947	2.7009	10628.1838	0.00009
44	45	1	1.6532	0.0000	0.0000	2.7331	11229.7648	0.00009
45	48	3	1.6812	0.4771	0.8022	2.8266	13153.4910	0.00008
46	50	2	1.6990	0.3010	0.5114	2.8865	14537.0576	0.00007
47	52	1	1.7160	0.0000	0.0000	2.9447	16003.2501	0.00006
48	53	2	1.7243	0.3010	0.5191	2.9731	16767.7921	0.00006
49	54	1	1.7324	0.0000	0.0000	3.0012	17553.5399	0.00006
50	57	1	1.7559	0.0000	0.0000	3.0831	20039.8006	0.00005
51	62	1	1.7924	0.0000	0.0000	3.2127	24624.0588	0.00004
52	80	1	1.9031	0.0000	0.0000	3.6218	45980.1589	0.00002
53	118	1	2.0719	0.0000	0.0000	4.2927	119154.6288	0.00001
	Total	**33151**	**70.60987**	**53.27537**	**49.6410**	**102.7957**	**433887.9122**	**1.35917**

Application of K-S Statistical test

To examine the conformity of the observed author distribution versus Lotka's distribution, the K-S test had employed. For applying K-S test, convert the observed and expected number of authors into fractional values, and take the difference between cumulative fractional values of the observed and expected number of authors (equation number.3). Table 3 shows the K-S test with n = 2.45 c = 0.74 in the theoretical distribution.

Table 3. Kolmogorov –Smirnov test of observed and expected distributions of authors in SB literature

		Observed		**Expected**		
x	***yx***	***yx/Σyx (% of authors)***	***Σ(yx/Σyx) (Cumulative % of authors)***	***fe = C (1/xn) (% of expected authors)***	***Σfe (Cumulative % expected authors***	***Dmax***
Col.1	***Col.2***	***Col.3***	***Col.4***	***Col.5***	***Col.6***	***(Col.4-Col.6)***
1	24716	0.7456	0.7456	0.7400	0.7400	0.0056
2	4353	0.1313	0.8769	0.1354	0.8754	0.0014
3	1720	0.0519	0.9287	0.0502	0.9256	0.0032
4	763	0.0230	0.9518	0.0248	0.9504	0.0014
5	446	0.0135	0.9652	0.0143	0.9647	0.0005
6	280	0.0084	0.9737	0.0092	0.9739	-0.0002
7	175	0.0053	0.9789	0.0063	0.9802	-0.0012
8	128	0.0039	0.9828	0.0045	0.9847	-0.0019
9	118	0.0036	0.9864	0.0034	0.9881	-0.0018

10	78	0.0024	0.9887	0.0026	0.9907	-0.0020
11	53	0.0016	0.9903	0.0021	0.9928	-0.0025
12	50	0.0015	0.9918	0.0017	0.9945	-0.0027
13	32	0.0010	0.9928	0.0014	0.9959	-0.0031
14	34	0.0010	0.9938	0.0012	0.9970	-0.0032
15	22	0.0007	0.9945	0.0010	0.9980	-0.0035
16	27	0.0008	0.9953	0.0008	0.9988	-0.0035
17	16	0.0005	0.9958	0.0007	0.9995	-0.0038
18	19	0.0006	0.9963	0.0006	1.0002	-0.0038
19	13	0.0004	0.9967	0.0005	1.0007	-0.0040
20	10	0.0003	0.9970	0.0005	1.0012	-0.0042
21	9	0.0003	0.9973	0.0004	1.0016	-0.0043
22	11	0.0003	0.9976	0.0004	1.0020	-0.0044
23	4	0.0001	0.9978	0.0003	1.0023	-0.0046
24	8	0.0002	0.9980	0.0003	1.0027	-0.0047
25	1	0.0000	0.9980	0.0003	1.0029	-0.0049
26	3	0.0001	0.9981	0.0003	1.0032	-0.0051
27	3	0.0001	0.9982	0.0002	1.0034	-0.0052
28	3	0.0001	0.9983	0.0002	1.0036	-0.0053
29	3	0.0001	0.9984	0.0002	1.0038	-0.0054
30	3	0.0001	0.9985	0.0002	1.0040	-0.0055
31	4	0.0001	0.9986	0.0002	1.0042	-0.0056
32	6	0.0002	0.9988	0.0002	1.0043	-0.0055
33	3	0.0001	0.9989	0.0001	1.0045	-0.0056
34	4	0.0001	0.9990	0.0001	1.0046	-0.0056
36	1	0.0000	0.9990	0.0001	1.0047	-0.0057
37	4	0.0001	0.9991	0.0001	1.0048	-0.0057
38	1	0.0000	0.9992	0.0001	1.0049	-0.0057
39	3	0.0001	0.9993	0.0001	1.0050	-0.0057
40	2	0.0001	0.9993	0.0001	1.0051	-0.0058
41	3	0.0001	0.9994	0.0001	1.0052	-0.0057
42	1	0.0000	0.9994	0.0001	1.0052	-0.0058
43	2	0.0001	0.9995	0.0001	1.0053	-0.0058
44	2	0.0001	0.9996	0.0001	1.0054	-0.0058
45	1	0.0000	0.9996	0.0001	1.0055	-0.0059
48	3	0.0001	0.9997	0.0001	1.0055	-0.0058
50	2	0.0001	0.9998	0.0001	1.0056	-0.0058
52	1	0.0000	0.9998	0.0000	1.0056	-0.0058
53	2	0.0001	0.9998	0.0000	1.0057	-0.0058
54	1	0.0000	0.9999	0.0000	1.0057	-0.0058
57	1	0.0000	0.9999	0.0000	1.0057	-0.0058

62	1	0.0000	0.9999	0.0000	1.0058	-0.0058
80	1	0.0000	1.0000	0.0000	1.0058	-0.0058
118	1	0.0000	1.0000	0.0000	1.0058	-0.0058
Total	**33151**				**Dmax**	**0.0056**

$$D = \max|F_0(x) - S_n(x)| = 0.0056$$

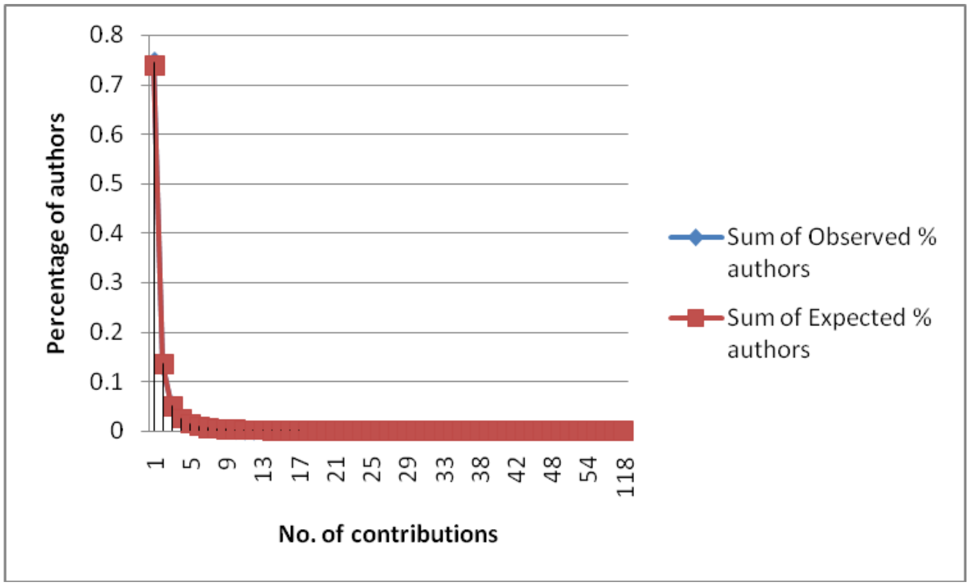

Figure 1. The fraction of Observed and Expected authors

Critical value at 0.01 level of significance had calculated using the equation (5), by substituting the values in this equation, the critical value has obtained as follows:

$$= \frac{1.63}{(33151 + (33151/10)^{1/2})^{1/2}}$$

$_{c\ v\ =}$ 0.0089

It is evident from Table 3 that the maximum difference (D max) obtained as (0.0056), which is lower than the calculated critical value (0.0089) at the 0.01 level of significance. Therefore, the null hypothesis that the set of data conforms to Lotka's Law has accepted at 0.01 level of significance. We can, therefore conclude that author productivity in SB literature does fit the Lotka's Law. Figure 1, shows the Lotka's plot showcasing the fraction of observed and expected number authors which depicts that both lines are moving almost similar way.

Conclusion

Examination of author distribution pattern is one of the core area of research in library and information science field. Generally, authorship distributions have been evaluating by using the famous Lotka's inverse square law of author productivity. Academicians had treated this Law as one of the classical laws of bibliometrics. The present analysis exposed that the Lotka's Law in its modified form does fit the author productivity distribution pattern in SB literature based on full-count authors with the calculated values of 'n' (2.45) and 'c' (0.74). This study is a lead up attempt to study the authorship distribution in the global SB literature. Hence the outcome of this study could elicit further such inclusive studies of similar kind in diverse subject areas in future.

References

1. Kiran Savanur, S. Gayathri Devi & P. V. Konnur (2014) Lotka's Law and Authorship Distribution in the Journal of 'Columbia Law Review', *COLLNET Journal of Scientometrics and Information Management*, 8:1, 193-208, DOI: 10.1080/09737766.2014.947840

2. Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington academy of sciences*, *16*(12), 317-323.

3. Pao, M. L. (1986). An empirical examination of Lotka's law. *Journal of the American Society for Information Science*, *37*(1), 26-33.

4.Nicholls, P. T. (1989). Bibliometric modeling processes and the empirical validity of Lotka's law. *Journal of the American Society for Information Science*, *40*(6), 379-385.

5. Dhoble, S., & Kumar, S. (2017). Applicability of Lotka's law in mustard research publications in India–a scientometric study. *SSARSC International Journal of Library, Information Networks and Knowledge*, *2*(1), 1-10.

6. Ahmi, A., Saidin, S. Z., Mohd Nasir, M. H., & Ismail, Z. (2020). Applicability of Lotka's Law in eXtensible Business Reporting Language (XBRL) Studies. *International Journal of Advanced Science and Technology*, *29*(6), 282-289.

7. AhmedI, S. Z., & Rahman, M. A. (2009). Lotka's law and authorship distribution innutrition research in Bangladesh. *Annals of Library and Information Studies*, *56*, 95-102.

8. Pillai Sudhier, K. G. (2013). Lotka's Law and Pattern of Author Productivity in the Area of Physics Research. *DESIDOC Journal of Library & Information Technology*, *33*(6), 457-464.

9. Askew, C. A. (2008) An Examination of Lotka's law in the Field of Library and Information Studies. *FIU ElectronicTheses and Dissertations*,182.

10. Suresh Kumar, P.K. (2018). Author productivity and the application of Lotka's Law in LIS publications. *Annals of Library and Information Studies (ALIS)*, *64*(4), 234-241.

11. Kumar, S., & Senthilkumar, R. (2019). Applicability of Lotka's Law in Astronomy & Astrophysics Research of India. *Library Philosophy and Practice*, 1-13..http://digitalcommons.unl.edu/libphilprac/2129

12. Sevukan, R., & Sharma, J. (2008). Bibliometric analysis of research output of biotechnology faculties in some Indian central universities. *DESIDOC Journal of Library & Information Technology*, *28*(6), 11-20.

13. Naqvi, S. H., & Fatima, N. (2018). Authorship patterns in international business literature: applicability of Lotka's Law. *Annals of Library and Information Studies (ALIS)*, *64*(4), 253-259.

14. Kumar, S. (2015). Author productivity in the field Human Computer Interaction (HCI) research. *Annals of Library and Information Studies (ALIS)*, *61*(4), 273-285.