Journal of Scientometric Research, 2023; 12(1):197-203.
https://www.jscires.org

Research Article

Redesigning of Lotka's Law with Simpson's 3/8 Rule

Anindya Basu[1], Bidyarthi Dutta[2,*]

[1]Central Library, Maharani Kasiswari College, Kolkata, West Bengal, INDIA.
[2]Department of Library and Information Science, Vidyasagar University, Kolkata, West Bengal, INDIA.

ABSTRACT
The paper examines the possibility of applying higher degree numerical integration method upon Lotka's distribution data. The widely used method applied here is Pao method which precisely calculate the value of the constant C and this is a very crucial and deterministic controlling factor to define the behaviour of the authors' productivity distribution fitted to Lotka's equation. Simpson's 3/8 rule has been used to derive a new equation for the calculation of a much refined constant C and fitted to original Chemical Abstract and Auerbach data.

Keywords: Lotka's Law, Empirical Law, Numerical Integration, Simpson's 3/8 Rule.

Correspondence:
Bidyarthi Dutta
Department of Library and Information Science, Vidyasagar University, Rangamati, West Midnapore, West Bengal, INDIA.
Email: bidyarthi.bhaswati@gmail.com
ORCID ID: 0000-0001-6049-5445

Received: 29-05-2022;
Revised: 22-11-2022;
Accepted: 12-02-2023.

INTRODUCTION

The evolution of Scientometrics got structured with three empirical laws–Bradford's Law, Zipf's Law and Lotka's Law; all of them can be inter-related through common general theory of Information Production Process (IPP). IPP can be associated with a size-frequency function f(k) which is denoted as density of item k. For Lotka's Law, f(K) is a decreasing power law with a certain maximal item density i.e., number of authors [f(k)] with the highest number of papers(k). Under size-frequency functional framework, all the three laws can be interchanged mathematically; but, the volume of the bi-variable distribution data and the nature of items are different. For the exponent greater than 1, both Lotka and Zipf's law are mathematically same, but informetric interpretations are different.[13] The function f(k) indicates the density of sources with items k, k є [1,k_{max}] where items can be of different types e.g., articles, citations, word frequency etc. In the general rule of 80/20 rule, eighty percent of the population controls twenty percent of the resources; on the contrary, twenty percent of the same controls the eighty percent of the total resources. Likewise, under the parlance of standard square law exponent formulation, sixty percent of authors publish only one article in their lifetime. While fitting the real-time data with Lotka's functional form, it's never found 60 percent of authors publishes only one paper; but it is always found that, majority of the authors publish in the same pattern. Empirically, this phenomenon is evident across all domains/ subjects/disciplines. The controlling factors i.e., constant and exponent are largely influenced by coverage of the subject, coverage time, nature of research, volume of research production, collaborating behaviour among authors etc. Due to variations of these factors, the values of the distribution get changed controlled by its constant and exponent.

Fundamentals of Lotka's Law

Lotka's law is such a law which describes the pattern of authors' productivity in terms of frequency of publication in a certain subject field. It states that, the number of authors those publish x number of papers at certain time interval is nearly proportional to that of $1/n^2$ of the making one article.

Lotka's law of inverse productivity indicates that, the number of authors producing certain number of papers in their lifetime is inversely proportional to individual productivity of authors.[1]

$$\varphi(x) = \frac{c}{x^n} \qquad - 1$$

$\varphi(x)$is the number of authors with x number of papers and n is the exponent of the distribution, if n=2, the the distribution becomes a perfect inverse square law. Thus, when n=2, it becomes a special case which is called as inverse-square law which invites Lotka's Law to fit in. Lotka's Law is a special case of Riemann-Zeta Function. Eqn. 2 is only solvable when n=2 and 4 only.

$$\sum_{x=1}^{\infty} \frac{1}{x^n} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots \infty = \frac{\pi^2}{6} \qquad -- 2$$

So, eqn. 1 becomes

$$\varphi(x) = \frac{c}{x^n} = \frac{6}{x^{n} {}_{*\pi^2}} \qquad -- 3$$

It's worth noting that, the exponent value never become a rounded number, so, to calculate the exponent value, MLE method is used –

DOI: 10.5530/jscires.12.1.017

$$n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - (\sum X)^2} \qquad -- 4$$

Here, N = Number of Data Points as pair of x and y values, X= Logarithmic Transformation of x and Y= Logarithmic Transformation of y.

Literature Review

Lotka devised a very simple law on scientific productivity basically directs us to believe that, prolific writers actually capture the maximum share of scientific production in a subject and more or less 60 percent above number of writers actually write only one paper in their lifetime. This phenomenon is true for every subject.[1] There is no proof why this happens as there is no scope to add different latent factors (intellectual capacity of the writer, guidance from seniors, scope of further research following a problem, availability of research problems at hand, lacking of teamwork and collaborative movement etc.) into it. However, the trend of researches in this topic can be broadly divided into two ways-methodology development to apply Lotka's Law to datasets and second one is to analyze scientific data with Lotka's Law.

The researches done so far on Lotka's Law can be segregated into two broad sub-divisions– a. Development of Methods to solve it, building mathematical relation among Bradford, Lotka and Zipf's law; b. Application of the law in different subjects to understand authors' productivity. As Lotka's law shows a power law form, there have been continuous efforts to fit different standard distributions on authors' productivity in different disciplines. In a study, Lotka's law was applied in Humanities discipline using inverse square law and the law did fit on the dataset.[2] Hersh analyzed authors' productivity on researches on *Drosophilia* assuming a simple power law function: $y = bx^k$ and he stated that, prediction on the behaviour of exponential curve is risky due to uncertainty in reaching at the point of inflection at any moment.[3] In another study, a simple method was used to calculate constant value C by putting the first value of number of paper published (here x=1); he also calculated constant value C with Lee Pao Method.[4] His method showed C= 1.485 and Pao method showed the constant C to be 1.682 and he also showed his method by and large follows Sen's Method of deriving C.[5] Nicholls made an empirical analysis on the available standard datasets using inverse power model of Lotka's Law ($Y_x = kx^{-b}$) and estimated constant K using Pao's method and exponent with Maximum Likelihood Estimator (MLE) and subsequently K-S Test was done in order to test the goodness of fit.[6] In another research study, Bookstein made an important analysis on the intrinsic behavior of classic bibliometric distributions as shown in Bradford's Law, Lotka's Law and Zipf's law. He concluded that, "Lotka's law is not sensitive to how we count articles, so that two people testing the law for a single population, but different count methods, will very likely to come up with the same law".[7] In a short communication to the editor explaining a study on randomly chosen sample and he observed that, Lotka's law is biased towards the distribution of relatively more prolific authors and he also concluded that, the dataset must prepared taking authors' publication over a long time period; then only Lotka's law can be best fitted on the dataset. He fitted exponential distribution form over his collected data, but it did not decrease as fast as Lotka's distribution.[8] All the empirical laws get reduced to simple form of hyperbolic law and its probability density function revolves around k/x^2 and operates upon certain intervals of X-axis. Bradford's law shows identical behaviour with other similar empirical laws (Lotka, Price and Zipf).[9] In a ground-breaking work done by Pao used general inverse power law $x^n y = c$ and estimated the exponent value using linear least-square method. Constant C was approximated by numerical integration method using trapezoidal rule. As compared with the actual value of $\pi^2/6$, Pao method produced error less than 1/110,000; for $\pi^4/90$, the error is less than 1/25,000,000.[4] In another study, Lotka's law was validated on 15 classical datasets after ground breaking work of Pao and he proposed two modification on Pao method. First proposal of modification was while calculating the exponent value, the data need to be truncated and Maximum Likelihood Estimator should be solved by numerical iterative method. Second suggestion for modification was to also include multi-authorship fractional credit counts.[6] Nicholls also investigated the validity of Price's Law against the back-drop of the Lotka's law and he tried to prove consistency with respect to its theoretical and empirical behaviour between the two. But, the value of the constant k is always dependent on the exponent value b of the distribution and empirically X_{max} (No. of Papers produced by most prolific authors) is not infinite and also doesn't follow a limiting value; in some datasets–number of authors with single publication vary considerably across different subjects and this actually disobeys Price's conjecture due to problem in indicating cut-off point between prolific and non-prolific authors. His observation is "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."[10] In a similar type of investigation, Pao method was applied on 70 datasets and he observed that, in 90% of the cases followed generalized Lotka's method.[11] Bailon-Moreno *et al.* deduced a Unified Scientometric model by unifying all three classical scientometric laws and their variant forms through concept of Fractal theory and accumulated advantage models. Through the use of Index of fractality, they also showed with the difference of Fractality Index, how different forms of distributions(Zipf-Mandelbrot, Lotka, Leimkuhler Distribution form of Bradford's Law, Booth-Federowicz Zipf Distribution, Condon-Zipf Distribution, Brooke's Law for Aging of Science, Price's Law of Exponential Growth, Generalized Model of Aging-Viability etc.) are created as well as some of them changed their equation forms.[12] Egghe explained two dimensional Information Production Process(IPP) with size-frequency function and size-frequency functions and through approach a new domain of subject was created by "Lotkaian Informetrics".[13,14]

After literature review, it was realized that, no efforts being made so far to refine/research new alternate methods apart from trapezoidal rule and that is probably due to minimization of error approximation.

METHODOLOGY AND MATHEMATICAL FORMULATION

The Integral form of timpson's 3/8 rule is -

It is a four point Newton-Cotes formula in the interval [a,b]

Here, n=3 and $h = \frac{b-a}{3}$

$$I = (b-a)\sum_{r=0}^{3} f(x_r)\, k_r^{(3)}$$

$$= (b-a)\left[f(x_0)k_0^{(3)} + f(x_1)k_1^{(3)} + f(x_2)k_2^{(3)} + f(x_3)k_3^{(3)}\right]$$

$$= \frac{(b-a)}{8}\left[y_0 + 3y_1 + 3y_2 + y_3\right] \qquad -5$$

The error is - $E_{3/8} = \frac{-3(b-a)^5}{80} f^{IV}(\varepsilon)\, a < \varepsilon < b$ -6

The degree of precision

$$\int_a^b f(x)dx - \frac{b-a}{8}\left[f(a) + f\left(\frac{2a+b}{3}\right) + f\left(\frac{a+2b}{3}\right) + f(b)\right] < \frac{-3M}{80(b-a)^6} \qquad -7$$

Error Approximation can be used as M here -

$$M = \frac{3n(n+1)(n+2)(n+3)(n+4)}{80 * x^{(n+6)}}$$

So, within the interval of x and x+1, the values can be replaced with the values of a and b, eqn.7 becomes –

$$\therefore \int_x^{x+1} \frac{1}{x^n}dx - \frac{1}{8}\left[\frac{1}{x^n} + \frac{3}{\left(x+\frac{1}{3}\right)^n} + \frac{3}{\left(x+\frac{2}{3}\right)^n} + \frac{1}{(x+1)^n}\right] < \frac{3n(n+1)(n+2)(n+3)(n+4)}{80*x^{(n+6)}} \qquad -8$$

$$\Rightarrow 0 < \frac{3n(n+1)(n+2)(n+3)(n+4)}{80*x^{(n+6)}} - \left[\int_x^{x+1} \frac{1}{x^n}dx - \frac{1}{8}\left[\frac{1}{x^n} + \frac{3}{\left(x+\frac{1}{3}\right)^n} + \frac{3}{\left(x+\frac{2}{3}\right)^n} + \frac{1}{(x+1)^n}\right]\right] \qquad -9$$

$$\Rightarrow 0 < \frac{1}{8}\left[\frac{1}{x^n} + \frac{3}{\left(x+\frac{1}{3}\right)^n} + \frac{3}{\left(x+\frac{2}{3}\right)^n} + \frac{1}{(x+1)^n}\right] - \int_x^{x+1}\frac{1}{x^n}dx < \frac{3n(n+1)(n+2)(n+3)(n+4)}{80*x^{(n+6)}} \qquad -10$$

X= P, P+1, P+2, P+3.........∞, summing the inequality form -

$$\Rightarrow 0 < \sum_p^{\infty}\frac{1}{8}\left[\frac{1}{p^n} + \frac{3}{\left(p+\frac{1}{3}\right)^n} + \frac{3}{\left(p+\frac{2}{3}\right)^n} + \frac{1}{(p+1)^n}\right] - \int_x^{x+1}\frac{dx}{x^n} < \sum_p^{\infty}\frac{3n(n+1)(n+2)(n+3)(n+4)}{80 * x^{(n+6)}}$$

$$\Rightarrow 0 < [(\frac{1}{8p^n} + \frac{3}{8\left(p+\frac{1}{3}\right)^n} + \frac{3}{8\left(p+\frac{2}{3}\right)^n} + \frac{1}{8(p+1)^n}) + (\frac{1}{8(p+1)^n} + \frac{3}{8\left(p+\frac{4}{3}\right)^n} + \frac{3}{8\left(p+\frac{5}{3}\right)^n} + \frac{1}{8(p+2)^n}) + ...]$$
$$- \int_P^{\infty}\frac{dx}{x^n} < \sum_P^{\infty}\frac{3n(n+1)(n+2)(n+3)(n+4)}{80*x^{(n+6)}}$$

$$\Rightarrow 0 < \frac{1}{8p^n} + \frac{1}{8}\left[\sum_P^{\infty}\frac{3}{\left(x+\frac{1}{3}\right)^n} + \sum_P^{\infty}\frac{3}{\left(x+\frac{2}{3}\right)^n} + 2\sum_{p+1}^{\infty}\frac{1}{x^n}\right] - \int_P^{\infty}\frac{dx}{x^n} < \sum_P^{\infty}\frac{3n(n+1)(n+2)(n+3)(n+4)}{80*x^{(n+6)}} \qquad -11$$

This may be rewritten as-

$$\int_P^{\infty}\frac{dx}{x^n} - \frac{1}{8p^n} < \frac{3}{8}\sum_p^{\infty}\frac{1}{\left(x+\frac{1}{3}\right)^n} + \frac{3}{8}\sum_p^{\infty}\frac{1}{\left(x+\frac{2}{3}\right)^n} + \frac{1}{4}\sum_{p+1}^{\infty}\frac{1}{x^n} < \int_P^{\infty}\frac{dx}{x^n} - \frac{1}{8p^n} + \sum_P^{\infty}\frac{3n(n+1)(n+2)(n+3)(n+4)}{80*x^{(n+6)}} \qquad -12$$

$$\sum_P^{\infty}\frac{1}{x^n} < \int_{P-1}^{\infty}\frac{dx}{x^n} \qquad -13$$

Estimation of $\sum_P^{\infty}\frac{3n(n+1)(n+2)(n+3)(n+4)}{80*x^{(n+6)}}$

$$\sum_P^{\infty}\frac{3n(n+1)(n+2)(n+3)(n+4)}{80 * x^{(n+6)}} < \frac{3n(n+1)(n+2)(n+3)(n+4)}{80}\int_{P-1}^{\infty}\frac{dx}{x^{(n+6)}}$$

$$= \frac{3n(n+1)(n+2)(n+3)(n+4)}{80}\left[\frac{x^{-(n+5)}}{(n+5)}\right]_{P-1}^{\infty}$$

$$= \frac{\frac{3n(n+1)(n+2)(n+3)(n+4)}{80} * 1}{(n+5)} \qquad -14$$

Estimation of $\sum_p^{\infty}\frac{1}{\left(x+\frac{1}{3}\right)^n} < \int_{p-1}^{\infty}\frac{1}{\left(x+\frac{1}{3}\right)^n}dx$

$$= \left[\frac{1}{-(1-n)\left(x+\frac{1}{3}\right)^{(n-1)}}\right]_{P-1}^{\infty}$$

$$= \frac{1}{(n-1)\left(p-\frac{2}{3}\right)^{(n-1)}} \qquad -15$$

Estimation of $\sum_p^{\infty}\frac{1}{\left(x+\frac{2}{3}\right)^n} < \int_{p-1}^{\infty}\frac{1}{\left(x+\frac{2}{3}\right)^n}dx$

$$= \left[\frac{1}{-(1-n)\left(x+\frac{2}{3}\right)^{(n-1)}}\right]_{P-1}^{\infty}$$

$$= \frac{1}{(n-1)\left(p-\frac{1}{3}\right)^{(n-1)}} \qquad -16$$

Substituting the value of 12, 15 into 16-

$$\int_P^{\infty}\frac{dx}{x^n} - \frac{1}{8p^n} - \frac{3}{8(n-1)\left(p-\frac{2}{3}\right)^{(n-1)}} - \frac{3}{8(n-1)\left(p-\frac{1}{3}\right)^{(n-1)}} < \frac{1}{4}\sum_{p+1}^{\infty}\frac{1}{x^n}$$
$$< \int_P^{\infty}\frac{dx}{x^n} - \frac{1}{8p^n} - \frac{3}{8(n-1)\left(p-\frac{2}{3}\right)^{(n-1)}} - \frac{3}{8(n-1)\left(p-\frac{1}{3}\right)^{(n-1)}}$$
$$+ \frac{3n(n+1)(n+2)(n+3)(n+4)}{80(n+5)(p-1)^{(n+5)}}$$

$$\Rightarrow \frac{4}{(n-1)P^{(n-1)}} - \frac{1}{2p^n} - \frac{3}{2(n-1)\left(p-\frac{2}{3}\right)^{(n-1)}} - \frac{3}{2(n-1)\left(p-\frac{1}{3}\right)^{(n-1)}} < \sum_{p+1}^{\infty}\frac{1}{x^n} < \frac{4}{(n-1)P^{(n-1)}} - \frac{1}{2p^n} - \frac{3}{2(n-1)\left(p-\frac{2}{3}\right)^{(n-1)}} - \frac{3}{2(n-1)\left(p-\frac{1}{3}\right)^{(n-1)}} + \frac{3n(n+1)(n+2)(n+3)(n+4)}{20(n+5)(P-1)^{(n+5)}} \qquad -17$$

$$\text{So, } \sum_1^{\infty}\frac{1}{x^n} = \sum_1^{P}\frac{1}{x^n} + \frac{4}{(n-1)P^{(n-1)}} - \frac{1}{2p^n} - \frac{3}{2(n-1)\left(p-\frac{2}{3}\right)^{(n-1)}} - \frac{3}{2(n-1)\left(p-\frac{1}{3}\right)^{(n-1)}} + \frac{3n(n+1)(n+2)(n+3)(n+4)}{20(n+5)(P-1)^{(n+5)}}$$

$$= \sum_1^{P-1}\frac{1}{x^n} + \frac{1}{2P^n} + \frac{4}{(n-1)P^{(n-1)}} - \frac{3}{2(n-1)}\left[\frac{1}{\left(p-\frac{2}{3}\right)^{(n-1)}} + \frac{1}{\left(p-\frac{1}{3}\right)^{(n-1)}}\right] + \frac{3n(n+1)(n+2)(n+3)(n+4)}{20(n+5)(P-1)^{(n+5)}} \qquad -18$$

So, from Lotka's law,

$$C = \frac{1}{\sum_1^{\infty}\frac{1}{x^n}}$$

$$= 1/\sum_1^{P-1}\frac{1}{x^n} + \frac{1}{2P^n} + \frac{4}{(n-1)P^{(n-1)}} - \frac{3}{2(n-1)}\left[\frac{1}{\left(p-\frac{2}{3}\right)^{(n-1)}} + \frac{1}{\left(p-\frac{1}{3}\right)^{(n-1)}}\right] + \frac{3n(n+1)(n+2)(n+3)(n+4)}{20(n+5)(P-1)^{(n+5)}} - \qquad 19$$

Calculation of Constant C(For Chemical Abstract Data):-

$$C = \frac{1}{\sum_1^\infty \frac{1}{x^n}} = 1/\sum_1^{P-1} \frac{1}{x^n} + \frac{1}{2P^n} + \frac{4}{(n-1)P^{(n-1)}} - \frac{3}{2(n-1)}\left[\frac{1}{\left(p-\frac{2}{3}\right)^{(n-1)}} + \frac{1}{\left(p-\frac{1}{3}\right)^{(n-1)}}\right] + \frac{3n(n+1)(n+2)(n+3)(n+4)}{20(n+5)(P-1)^{(n+5)}}$$

$$= 1/\ 1.764142$$

$$= 0.5668478$$

So, the Lotka's equation becomes $-\varphi(x) = \frac{c}{x^n} = \varphi(x) = \frac{0.5668478}{x^{1.8878}}$

Analysis of Data

The collected data are analyzed and presented in Table 1 and Table 2.

Critical Value is $= \frac{1.63}{\sqrt{\sum y_x}} = \frac{1.63}{\sqrt{6891}} = 0.0196357$ (At 0.01 Significance Level) -17

And Maximum Difference (D_{max}) is - 0.018117

Here, we can see that, the maximum difference is less than the critical value at 0.01 significant level and thus null hypothesis is accepted. That means, $\varphi(x) = \frac{0.5668478}{x^{1.8878}}$ is fair enough to fit the observed values in Chemical Abstract Data through Simpson's 3/8 rule.

$$C = \frac{1}{\sum_1^\infty \frac{1}{x^n}} = 1/\sum_1^{P-1} \frac{1}{x^n} + \frac{1}{2P^n} + \frac{4}{(n-1)P^{(n-1)}} - \frac{3}{2(n-1)}\left[\frac{1}{\left(p-\frac{2}{3}\right)^{(n-1)}} + \frac{1}{\left(p-\frac{1}{3}\right)^{(n-1)}}\right] + \frac{3n(n+1)(n+2)(n+3)(n+4)}{20(n+5)(P-1)^{(n+5)}}$$

$$= 1/\ 1.622413$$

$$= 0.6163658$$

So, the Lotka's equation becomes -

$$\varphi(x) = \frac{c}{x^n} = \varphi(x) = \frac{0.6163658}{x^{2.021}}$$

Critical Value is $= \frac{1.63}{\sqrt{\sum y_x}} = \frac{1.63}{\sqrt{1325}} = 0.04477954$ (At 0.01 Significance Level)

Table 1: KS Test of the Observed and Expected Values of Authors' Productivity Distribution of Chemical Abstract Data.

No. of Articles (x)	No. of Contributing Authors (y)	Propotions of Authors	Cumulative of Column C	Fitted Value with Eq. 19	Cumulative of Column E	Difference Between Column D&F
1	3991	0.579161	0.579161	0.568390	0.568390	0.010771
2	1059	0.153679	0.732840	0.153590	0.721980	0.010860
3	493	0.071543	0.804383	0.071439	0.793419	0.010964
4	287	0.041649	0.846032	0.041503	0.834922	0.011110
5	184	0.026701	0.872733	0.027235	0.862157	0.010576
6	131	0.019010	0.891743	0.019304	0.881461	0.010282
7	113	0.016398	0.908141	0.014430	0.895891	0.012250
8	85	0.012335	0.920476	0.011215	0.907106	0.013370
9	64	0.009287	0.929763	0.008979	0.916085	0.013678
10	65	0.009433	0.939196	0.007359	0.923444	0.015752
11	41	0.005950	0.945146	0.006148	0.929592	0.015554
12	47	0.006820	0.951966	0.005216	0.934808	0.017158
13	32	0.004644	0.956610	0.004485	0.939293	0.017317
14	28	0.004063	0.960673	0.003899	0.943192	0.017481
15	21	0.003047	0.963720	0.003423	0.946615	0.017105
16	24	0.003483	0.967203	0.003030	0.949645	0.017558
17	18	0.002612	0.969815	0.002703	0.952348	0.017467
18	19	0.002757	0.972572	0.002426	0.954774	0.017798
19	17	0.002467	0.975039	0.002191	0.956965	0.018074
20	14	0.002032	0.977071	0.001989	0.958954	0.018117
21	9	0.001306	0.978377	0.001814	0.960768	0.017609
22	11	0.001596	0.979973	0.001661	0.962429	0.017544
23	8	0.001161	0.981134	0.001527	0.963956	0.017178
24	8	0.001161	0.982295	0.001410	0.965366	0.016929
25	9	0.001306	0.983601	0.001305	0.966671	0.016930

No. of Articles (x)	No. of Contributing Authors (y)	Propotions of Authors	Cumulative of Column C	Fitted Value with Eq. 19	Cumulative of Column E	Difference Between Column D&F
26	9	0.001306	0.984907	0.001212	0.967883	0.017024
27	8	0.001161	0.986068	0.001129	0.969012	0.017056
28	10	0.001451	0.987519	0.001054	0.970066	0.017453
29	8	0.001161	0.988680	0.000986	0.971052	0.017628
30	7	0.001016	0.989696	0.000925	0.971977	0.017719
31	3	0.000435	0.990131	0.000869	0.972846	0.017285
32	3	0.000435	0.990566	0.000819	0.973665	0.016901
33	6	0.000871	0.991437	0.000773	0.974438	0.016999
34	4	0.000580	0.992017	0.000730	0.975168	0.016849
36	1	0.000145	0.992162	0.000656	0.975824	0.016338
37	1	0.000145	0.992307	0.000623	0.976447	0.015860
38	4	0.000580	0.992887	0.000592	0.977039	0.015848
39	3	0.000435	0.993322	0.000564	0.977603	0.015719
40	2	0.000290	0.993612	0.000537	0.978140	0.015472
41	1	0.000145	0.993757	0.000513	0.978653	0.015104
42	2	0.000290	0.994047	0.000490	0.979143	0.014904
44	3	0.000435	0.994482	0.000449	0.979592	0.014890
45	4	0.000580	0.995062	0.000430	0.980022	0.015040
46	2	0.000290	0.995352	0.000413	0.980435	0.014917
47	3	0.000435	0.995787	0.000396	0.980831	0.014956
49	1	0.000145	0.995932	0.000366	0.981197	0.014735
50	2	0.000290	0.996222	0.000353	0.981550	0.014672
51	1	0.000145	0.996367	0.000340	0.981890	0.014477
52	2	0.000290	0.996657	0.000327	0.982217	0.014440
53	2	0.000290	0.996947	0.000316	0.982533	0.014414
54	2	0.000290	0.997237	0.000305	0.982838	0.014399
55	3	0.000435	0.997672	0.000295	0.983133	0.014539
57	1	0.000145	0.997817	0.000275	0.983408	0.014409
58	1	0.000145	0.997962	0.000266	0.983674	0.014288
61	2	0.000290	0.998252	0.000242	0.983916	0.014336
66	1	0.000145	0.998397	0.000209	0.984125	0.014272
68	2	0.000290	0.998687	0.000197	0.984322	0.014365
73	1	0.000145	0.998832	0.000173	0.984495	0.014337
78	1	0.000145	0.998977	0.000152	0.984647	0.014330
80	1	0.000145	0.999122	0.000145	0.984792	0.014330
84	1	0.000145	0.999267	0.000132	0.984924	0.014343
95	1	0.000145	0.999412	0.000105	0.985029	0.014383
107	1	0.000145	0.999557	0.000084	0.985113	0.014444
109	1	0.000145	0.999702	0.000081	0.985194	0.014508
114	1	0.000145	0.999847	0.000074	0.985268	0.014579
346	1	0.000145	0.999992	0.000009	0.985277	0.014715

Table 2: KS Test of the Observed and Expected Values of Authors' Productivity Distribution of Auerbach Data.

Articles_x	Authors_y	Propotions ofAuthors	Cumulative of Column C	Fitted Value with Eqn. 19	Cumulative of Column E	Difference Between Column D&F
1	784	0.591698	0.591698	0.616366	0.616366	-0.024668
2	204	0.153962	0.745660	0.151865	0.768231	-0.022571
3	127	0.095849	0.841509	0.066923	0.835154	0.006355
4	50	0.037736	0.879245	0.037418	0.872572	0.006673
5	33	0.024906	0.904151	0.023835	0.896407	0.007744
6	28	0.021132	0.925283	0.016489	0.912896	0.012387
7	19	0.014340	0.939623	0.012075	0.924971	0.014652
8	19	0.014340	0.953963	0.009219	0.934190	0.019773
9	6	0.004528	0.958491	0.007266	0.941456	0.017035
10	7	0.005283	0.963774	0.005873	0.947329	0.016445
11	6	0.004528	0.968302	0.004844	0.952173	0.016129
12	7	0.005283	0.973585	0.004063	0.956236	0.017349
13	4	0.003019	0.976604	0.003456	0.959692	0.016912
14	4	0.003019	0.979623	0.002975	0.962667	0.016956
15	5	0.003774	0.983397	0.002588	0.965255	0.018142
16	3	0.002264	0.985661	0.002271	0.967526	0.018135
17	3	0.002264	0.987925	0.002010	0.969536	0.018389
18	1	0.000755	0.988680	0.001790	0.971326	0.017354
21	1	0.000755	0.989435	0.001311	0.972637	0.016798
22	3	0.002264	0.991699	0.001193	0.973830	0.017869
24	3	0.002264	0.993963	0.001001	0.974831	0.019132
25	2	0.001509	0.995472	0.000922	0.975753	0.019719
27	1	0.000755	0.996227	0.000789	0.976542	0.019685
30	1	0.000755	0.996982	0.000638	0.977180	0.019802
34	1	0.000755	0.997737	0.000495	0.977675	0.020062
37	1	0.000755	0.998492	0.000417	0.978092	0.020400
48	2	0.001509	1.000001	0.000247	0.978339	0.021662

And Maximum Difference(D_{max}) is - 0.021662

Here, we can see that, the maximum difference is less than the critical value at 0.01 significant level and thus null hypothesis is accepted. That means, $\varphi(x) = \frac{0.6163658}{x^{2.021}}$ is fair enough to fit the observed values in Chemical Abstract Data through Simpson's 3/8 rule.

Interpretation and Conclusion:- Potter advised to devise a more sophisticated model to fit authors' productivity in a better way.[15] The basic difference between Pao method and our method is the use of higher degree of Newton-Cotes formula i.e., cubic interpolation instead of quadratic interpolation or trapezoid interpolation. In this method, two intermediate points are connected between two points (a,b) and it's more effective than other methods as it contains more functional values. As our method covers more area under the curve, the constant value gets adjusted by the functional values/terms of the derived equation using Simpson's 3/8 rule. This method is robust in its design and fits the data. This paper may trigger to research on application of other higher degree interpolation forms.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Lotka AJ. The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences. 1926;16(12):317-23.
2. Murphy LJ. Lotka's law in the humanities?. Journal of the American Society for Information Science. 1973;24(6):461-2.
3. Hersh AH. *Drosophilia* and the course of research. Ohio Journal of Science. 1942;42(5):198-200.
4. Pao ML. An emphirical examination of Lotka's Law. Journal of the American Society for Information Science. 1986;37(1):26-33.
5. Sen B K. Lotka's Law: A View Point, Annals of Library and Information Studies. 2010;57:166-7.
6. Nicholls PT. Empirical validation of Lotka's law. Information Processing and Management. 1986; 22(5):417-9.
7. Bookstein A. The bibliometric distributions. The Library Quarterly. 1976;46(4):416-23.
8. Krisciunas K. Lotka's Law-Year by Year. Journal of the American Society for Information Science. 1977;28(1):65.
9. Brookes BC. Ranking techniques and the empirical log law. Information Processing and Management. 1984;20(1-2):37-46.
10. Nicholls PT. Price's square root law: Empirical validity and relation to Lotka's law. Information Processing and Management. 1988;24(4):469-77.
11. Nicholls PT. Bibliometric modeling processes and the empirical validity of Lotka's law. Journal of the American Society for Information Science. 1989;40(6):379-85.
12. Bailón-Moreno R, Jurado-Alameda E, Ruiz-Baños R, Courtial JP. The unified scientometric model. Fractality and Transfractality. Scientometrics. 2005;63(2):231-57.
13. Egghe L. Zipfian and lotkaian continuous concentration theory. Journal of the American Society for Information Science and Technology. 2005;56(9):935-45.
14. Egghe L. Power laws in the information production process: Lotkaian informetrics. Elsevier Ltd., 2005.
15. Potter WG. Lotka's law revisited. 1981;31:21-39.

Cite this article: Basu A, Dutta B. Redesigning of Lotka's Law with Simpson's 3/8 Rule. J Scientometric Res. 2023;12(1):197-203.