LOTKA'S FREQUENCY DISTRIBUTION
OF SCIENTIFIC PRODUCTIVITY

Russell C. Coile

Professional Paper No. 216
February 1978

The ideas expressed in this paper are those of the author. The paper does not necessarily represent the views of the Center for Naval Analyses.

CENTER FOR NAVAL ANALYSES
1401 Wilson Boulevard
Arlington, Virginia 22209

DISTRIBUTION STATEMENT A
Approved for public release; Distribution Unlimited
LOTKA'S FREQUENCY DISTRIBUTION OF SCIENTIFIC PRODUCTIVITY

Russell C. Coile

Center for Naval Analyses
University of Rochester
Arlington, Virginia

Centre for Information Science
The City University
London, England

(To be published in the Journal of the American Society for Information Science)
ABSTRACT

In 1926, Alfred Lotka examined the frequency distribution of scientific productivity of chemists and physicists. After analyzing the number of publications of chemists listed in Chemical Abstracts 1907-1916 and the contributions of physicists listed in Auerbach's Geschichtstafeln der Physik, he observed that the number of persons making n contributions is about $\frac{1}{2n+1}$ of those making one and the proportion of all contributors that make a single contribution is about 50 percent. Recently, investigators studying the applicability of Lotka's Law to the humanities and to map librarianship may have misinterpreted Lotka's law and have concluded erroneously that the law applies to these fields. Corrected calculations indicate that Lotka's law does not apply.
INTRODUCTION

Among various 1976 events, Washingtonians might well celebrate on June 19, 1976, the 50th anniversary of a famous bibliometric formula now known as Lotka's Law. Alfred J. Lotka (1) published a paper in the *Journal of the Washington Academy of Sciences* in 1926 on the frequency distribution of scientific productivity. His purpose was "to determine the part which men of different caliber contribute to the progress of science." He first considered simple volume of production of papers on chemistry by counting the number of names in the decennial index of Chemical Abstracts 1907-1916, against which appeared 1, 2, 3, etc., entries. He tabulated data for 6,981 names beginning with the letters A and B. Lotka said in a footnote that, "Joint contributions have in all cases been credited to the senior author only." Names of firms were omitted.

In a similar fashion, Lotka looked at the name index of *Auerbach's Geschichtstafeln der Physik* which covered contributions by 1,325 physicists throughout history up to 1900. He felt that
in this case, not merely volume of productivity was being measured but that some account had been taken of quality, since only the outstanding contributions were listed by Auerbach.

Table I - Frequency Distribution of Scientific Productivity presents the highlights of Lotka's data. He also gave data for letters A and B separately and his table continued up to 346 contributions.

Figure 1 illustrates how Lotka presented his data graphically. As he put it, "On plotting the frequencies of persons having made 1, 2, 3... contributions, against these numbers 1, 2, 3... of contributions, both variables on a logarithmic scale, it is found that in each case the points are rather closely scattered about an essentially straight line having a slope of approximately two to one." Lotka found the slope of the curve to the first 17 points of Auerbach's data to be $2.021 \pm 0.017$ as determined by least squares. Similarly, the slope for his data from Chemical Abstracts, letters A and B jointly, as determined from the first thirty points, was $1.888 \pm 0.007$.

DEVELOPMENT OF "LOTKA'S LAW"

Lotka then stated, "The general formula for the relation thus found to exist between the frequency $y$ of persons making $x$ contributions is"
### Table 1

**FREQUENCY DISTRIBUTION OF SCIENTIFIC PRODUCTIVITY**

<table>
<thead>
<tr>
<th>Number of contributions (n)</th>
<th>Number of persons making stated number of contributions</th>
<th>Percent of total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chemical abstracts (A + B)</td>
<td>Auerbach's tables</td>
</tr>
<tr>
<td></td>
<td>Observed</td>
<td>Computed*</td>
</tr>
<tr>
<td>1</td>
<td>3,991</td>
<td>784</td>
</tr>
<tr>
<td>2</td>
<td>1,059</td>
<td>204</td>
</tr>
<tr>
<td>3</td>
<td>493</td>
<td>127</td>
</tr>
<tr>
<td>4</td>
<td>287</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>184</td>
<td>33</td>
</tr>
<tr>
<td>6</td>
<td>131</td>
<td>28</td>
</tr>
<tr>
<td>7</td>
<td>113</td>
<td>19</td>
</tr>
<tr>
<td>8</td>
<td>85</td>
<td>19</td>
</tr>
<tr>
<td>9</td>
<td>64</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>65</td>
<td>7</td>
</tr>
</tbody>
</table>

---

*According to \(f = 56.69/n^{1.888}\).*

**According to \(f = 600/\pi^2n^2\).*
FIG. 1: Lotka's logarithmic frequency diagram showing percent of total authors mentioned once, twice, etc. Chemical Abstracts data is shown by o's and Auerbach's data by x's. Dashed line corresponds to exponent 1.89, and solid line to exponent 2.
\[ x^n y = \text{const} \quad (1) \]

"For the special case that \( n = 2 \) (inverse square law of scientific productivity) the value of the constant in (1) is found as follows:

\[ y_1 = \frac{c}{2} \quad (2) \]

\[ y_2 = \frac{c}{2^2} \quad (3) \]

\[ y_n = \frac{c}{2^n} \quad (4) \]

\[ \sum_{1}^{\infty} y = c \left( \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \ldots \right) \quad (5) \]

\[ = c \sum_{1}^{\infty} \frac{1}{x^2} \quad (6) \]

\[ \sum_{1}^{\infty} y = c \frac{\pi^2}{6} \quad (7) \]

\[ c = \frac{6}{\pi^2} \sum_{1}^{\infty} y \quad (8) \]

"But since \( y \) is a frequency, the summation \( \sum_{1}^{\infty} y \) gives unity.

Then finally"
\[ C = \frac{6}{\pi^2} \quad (9) \]
\[ = \frac{6}{9.87} \quad (10) \]
\[ = 0.6079 \text{ or } 60.79 \text{ percent} \quad (11) \]

Lotka's footnote after equation (7) said, "See, for example K. Knopp, *Theorie und Anwendung der unendlichen Reihen*: 239, 1924 or J. L. Coolidge, *Mathematical Theory of Probability*: 22, 1925. For method of summation when exponent is fractional, see Whittaker and Robinson, *Calculus of Observations*: 136, 1924. Exponent 1.888 thus gives the value \( c = 0.5669 \)" (appearing in Table 1).

Lotka added, "Thus, according to the inverse square law, the proportion of all contributors who contribute a single item should be just over 60 percent. In the cases here examined, the actual proportion of this class to the whole was 59.2 percent in Auerbach's data (1,325 contributors), 57.7 percent in the Chemical Abstracts under initial A (1,543 contributors), 57.98 under letter B (5,348 contributors) and 57.9 under letters A and B jointly (6,891 contributors)."

Lotka concluded by saying, "Frequency distributions of the general type (1) have a wide range of applicability to a variety of phenomena, and the mere form of such a distribution throws
little or no light on the underlying physical relations. The fact that the exponent has, in the examples shown, approximately the value 2 enables us to state the result in the following simple form:

"In the cases examined it is found that the number of persons making 2 contributions is about one-fourth of those making one; the number making 3 contributions is about one-ninth, etc; the number making \( n \) contributions is about \( \frac{1}{n^2} \) of those making one; and the proportion of all contributors, that make a single contribution, is about 60 percent."

Note that Lotka's own use of his inverse square law of scientific productivity in footnote 2 of his Table 1 gave it as:

\[
f = \frac{600}{\pi^2 n^2}
\]

This formula has two parts: an exponent with value of 2 in the inverse square \( \left( \frac{1}{n^2} \right) \) portion of the expression, and a constant of \( \left( \frac{600}{\pi^2} \right) \) or \( \left( \frac{600}{\pi^2} \right) \) in percent, determined by the value, 2, of the exponent.

LOTKA'S LAW IN THE HUMANITIES?

The detailed presentation of extracts from Lotka's original paper given above has been thought necessary background to
clarify discussion of several recent papers which have examined the possible application of Lotka's Law to fields other than chemistry and physics. The first of these papers was a Brief Communication in this journal entitled, "Lotka's Law in the Humanities?" by Murphy (2). Murphy stressed that Lotka's Law (the inverse square law of scientific productivity) was originally only applicable in physical science, specifically chemistry and physics. He bemoaned the more recent general application of Lotka's Law in nonphysical science without appropriate new tests of validity. A plea was made for more "spot checks" of so called general "Laws" applied in information science.

Murphy questioned whether or not Lotka's Law could be applied successfully, predictively to nonscientific productivity. He tested its general application by examining the productivity of scholars publishing in the first decade of Technology and Culture, a quarterly specializing in the history of technology. His interpretation of Lotka's Law gave him a tabulation of observed and "ideal" number of authors which he stated "essentially verifies that Lotka's Law applies to this field within the humanities."

However, it would appear that Lotka's Law may have been misinterpreted. Murphy did not use Lotka's formula which has both a constant and the inverse square exponent. Instead, he
calculated "ideal" numbers on the basis that one could predict that for every 100 authors of a single article there would be 25 authors with two articles \( (n = 2 ; \frac{100}{n^2} = \frac{100}{4} = 25) \), about 11 authors with three articles \( (n = 3 ; \frac{100}{n^2} = \frac{100}{9} = 11+) \), about 6 authors with four articles \( (n = 4 ; \frac{100}{n^2} = \frac{100}{16} = 6+) \), 4 authors with five articles \( (n = 5 ; \frac{100}{n^2} = \frac{100}{25} = 4) \), etc."

This ignored Lotka's statement that the proportion of all contributors that make a single contribution is about 60 percent. Murphy then used the actual number of single authors of 130 as the basis for calculating the predicted number (rounded off) of authors with two articles, to be 33 which he termed "the ideal." He did not use the total number of authors as his base as Lotka did. His other predictions were 14 authors of three articles, 8 of four, and 5 of five. He did not add up these predictions, which sum to 190 as compared with 170 observed. Murphy plotted the number of authors versus the number of articles using notation of observed values with an x and ideal values with an o. His dashed ideal line is drawn as a straight line on the log-log plot, but is only drawn through two ideal points: the 130 authors of 1 paper and the 8 authors of 4 papers. Lotka plotted the percent of authors which does give a straight line. Murphy's number of authors will not be a straight line if drawn through all of his "ideal" points.
In order to test whether or not Lotka's Law applies to the observed data, it is necessary to eliminate the 6 co-authors mentioned in Murphy's table to make the data consistent with Lotka's senior-author formulation. Murphy's revised data with Lotka's predictions added would be as shown in Table 2 and as illustrated in Figure 2 following Lotka's format.

The Kolmogorov-Smirnov test is used to test where the maximum deviation, \( D \) is

\[
D = \text{maximum} \left| F_o(X) - S_N(X) \right|
\]

where \( F_o(X) \) is the theoretical cumulative frequency distribution function

\( S_N(X) \) is the observed cumulative frequency distribution function of a sample of \( N \) observations.

Since the Kolmogorov-Smirnov \( D_{\text{max}} \) of .1543 given in Table 3 is greater than the \( p = .01 \) level of significance of .1273, it appears that Lotka's Law does not apply to this sample of Technology and Culture authors.

LOTKA'S LAW AND MAP LIBRARIANSHIP

In response to Murphy's plea for more "spot checks" of so-called general "Laws," Schorr (3) published a Brief Communication in this journal reporting on a test in the field of map librarianship. Schorr used data extracted from a bibliography he had published earlier on authors publishing...
TABLE 2

ACTUAL OBSERVED NUMBER OF AUTHORS
WITH NUMBER OF ARTICLES

Murphy's data revised for senior authors only

<table>
<thead>
<tr>
<th>Papers/man</th>
<th>Men</th>
<th>% of men</th>
<th>% of men</th>
<th>Calculated*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>125</td>
<td>76.22</td>
<td>60.79</td>
<td>99.7</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>12.80</td>
<td>15.20</td>
<td>24.9</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>5.49</td>
<td>6.75</td>
<td>11.1</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>4.88</td>
<td>3.80</td>
<td>6.2</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.61</td>
<td>2.43</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>164</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Calculated by \( f = \frac{6}{\pi^2 n^2} \) where \( N \) is 164 total men.
FIG. 2: Murphy's data revised for senior authors only. Solid line is Lotka's Law with exponent 2.
### TABLE 3
KOLMOGOROV-SMIRNOV TEST

<table>
<thead>
<tr>
<th>Papers/man</th>
<th>Lotka's theoretical</th>
<th>Observed</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( F_0 (X) )</td>
<td>( F_0 (X) )</td>
<td>( S(N)(X) )</td>
</tr>
<tr>
<td>1</td>
<td>.6079</td>
<td>.7622</td>
<td>.1543</td>
</tr>
<tr>
<td>2</td>
<td>.1520</td>
<td>.1280</td>
<td>.1303</td>
</tr>
<tr>
<td>3</td>
<td>.0675</td>
<td>.0549</td>
<td>.1177</td>
</tr>
<tr>
<td>4</td>
<td>.0380</td>
<td>.0488</td>
<td>.1285</td>
</tr>
<tr>
<td>5</td>
<td>.0243</td>
<td>.0061</td>
<td>.1103</td>
</tr>
</tbody>
</table>

\( D_{\text{max}} = .1543 \)

\( N = 164 \)

.01 level of significance = \( \frac{1.63}{\sqrt{164}} = 0.163 = 0.1273 \)
between 1921 and 1973 on map librarianship, map libraries, and maps. He followed Murphy's procedures and came to the conclusion that "Lotka's Law holds true for the field of map librarianship." However, this conclusion is not true when the correct data is tested with an appropriate goodness-of-fit test.

It is first necessary to correct Schorr's Table I to eliminate co-authors in order to be consistent with Lotka's senior-author approach. The revised data given in Table 4 shows a total of 306 senior authors. Second, the chi-square test used to analyze entries in the original table through nine papers per contributor is not appropriate since the table entries for five through nine show fewer than 5 contributors in each case. In any event, the chi-square calculations are not given by Schorr.

The percent of contributors calculated by Lotka's inverse square formula is given in Table 4 along with the Kolmogorov-Smirnov calculations. Since the $D_{\text{max}}$ of .1601 is greater than the .01 level of significance of .0932, it would appear that Lotka's Law does not apply to this field of map librarianship.
TABLE 4

Frequencies of observed number of senior authors for MAP Librarianship

<table>
<thead>
<tr>
<th>Papers/ contributors</th>
<th># of contributors</th>
<th>% of contributors</th>
<th>$S_N(X)$</th>
<th>% of contributors</th>
<th>$F_O(X)$</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>235</td>
<td>76.80</td>
<td>.7680</td>
<td>60.79</td>
<td>.6079</td>
<td>.1601</td>
</tr>
<tr>
<td>2</td>
<td>39</td>
<td>12.75</td>
<td>.8955</td>
<td>15.20</td>
<td>.7599</td>
<td>.1356</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>5.23</td>
<td>.9478</td>
<td>6.75</td>
<td>.8274</td>
<td>.1204</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>1.96</td>
<td>.9674</td>
<td>3.80</td>
<td>.8654</td>
<td>.1020</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>1.31</td>
<td>.9805</td>
<td>2.43</td>
<td>.8897</td>
<td>.0908</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>.33</td>
<td>.9838</td>
<td>1.69</td>
<td>.9066</td>
<td>.0772</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>.65</td>
<td>.9903</td>
<td>1.24</td>
<td>.9190</td>
<td>.0713</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>.33</td>
<td>.9936</td>
<td>0.95</td>
<td>.9285</td>
<td>.0651</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>.33</td>
<td>.9969</td>
<td>0.75</td>
<td>.9360</td>
<td>.0609</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>.33</td>
<td>1.0002</td>
<td>0.14</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

D = max |$F_O(X) - S_N(X)$| = .1601

N = 306

.01 level of significance = \( \frac{1.63}{\sqrt{306}} = \frac{1.63}{17.49} = 0.0932 \)
REFERENCES


*Research supported in part by under Office of Naval Research Contract N00014-68-D2730-0017.


PP 126 — Classified.


PP 133 Squires, Michael L., "A Stochastic Model of Regime Change in Latin America," 42 pp., Feb 1975, AD A 007 912.


* CNA Professional Papers with an AD number may be obtained from the National Technical Information Service, U.S. Department of Commerce, Springfield, Virginia 22151. Other papers are available from the author at the Center for Naval Analyses, 1401 Wilson Boulevard, Arlington, Virginia 22209.
CNA Professional Papers — 1973 to Present (Continued)


PP 140 Mahoney, Robert and Druckman, Daniel*, "Simulation, Experimentation, and Context," 36 pp., 1 Sep 1975, (Published in Experimentation & Games, Vol. 6, No. 3, Sep 1975) *Mathematica, Inc.


PP 165 Jondrow, James M., "Effects of Trade Restrictions on Imports of Steel," 67 pp., November 1976, (Delivered at ILAB Conference in Dec 1976)


CNA Professional Papers — 1973 to Present (Continued)


PP 175  Levine, Daniel; Stoloff, Peter and Spruit, Nancy, "Public Drug Treatment and Addict Crime," June 1976, (Published in Journal of Legal Studies, Vol. 5, No. 2)


PP 182  Murray, Russell, 2nd, "The Quest for the Perfect Study or My First 1138 Days at CNA," 57 pp., April 1977


PP 187  Mahoney, Robert B., Jr., "New Technology and Naval Forces in the South Atlantic," 22 pp. (This paper was the basis for a presentation made at the Institute for Foreign Policy Analysts, Cambridge, Mass., 28 April 1977.)


PP 193  Colle, Russell C., "The Quest for the Perfect Study or My First 1138 Days at CNA," 57 pp., April 1977


PP 207

PP 208

PP 209 - Classified.

PP 210

PP 211

PP 213

PP 216