JOURNAL

OF THE

WASHINGTON ACADEMY OF SCIENCES

VOL. 16 JUNE 19, 1926 No. 12

STATISTICS.—*The frequency distribution of scientific productivity.*

ALFRED J. LOTKA. Metropolitan Life Insurance Company, New York.

It would be of interest to determine, if possible, the part which men of different calibre contribute to the progress of science.

Considering first simple volume of production, a count was made of the number of names, in the decennial index of Chemical Abstracts 1907–1916, against which appeared 1, 2, 3 . . . . entries. Names of firms (e.g. Aktiengesellschaft, etc.) were omitted from reckoning, since they represent the output, not of a single individual, but of an unknown number of persons. The letters A and B of the alphabet only were covered. These were treated both separately and in the aggregate, with the results shown in the table and in figures 1 and 2 below.

A similar process was also applied to the name index of Auerbach's Geschichtstafeln der Physik (J. A. Barth, Leipzig, 1910) which cover the entire range of history up to and including the year 1900. In this case we obtain a measure not merely of volume of productivity, but account is taken, in some degree, also of quality, since only the outstanding contributions find a place in this little volume, with its 110 pages of tabular text. The figures and relations thus obtained are shown in the table and in figures 1 and 2.

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. The approach to this ratio is particularly close in the case of the data taken from Auerbach's

TABLE 1.—FREQUENCY DISTRIBUTION OF SCIENTIFIC PRODUCTIVITY

NUMBER OF CONTRIBUTIONS n	NUMBER OF PERSONS MAKING STATED NUMBER OF CONTRIBUTIONS				PER CENT OF TOTAL					
	Chemical Abstracts			Auerbach's tables entire alphabet	Chemical Abstracts				Auerbach's tables	
					Observed			Computed[1]	Observed	Computed[2]
	Letter A	Letter B	A + B		A	B	A + B	A + B	Entire alphabet	
Total	1,543	5,348	6,891	1,325						
1	890	3,101	3,991	784	57.68	57.98	57.92	56.69	59.17	60.79
2	230	829	1,059	204	14.91	15.50	15.37	15.32	15.40	15.20
3	111	382	493	127	7.19	7.14	7.15	7.12	9.58	6.75
4	58	229	287	50	3.76	4.28	4.16	4.14	3.77	3.80
5	41	143	184	33	2.66	2.67	2.67	2.72	2.49	2.43
6	42	89	131	28	2.72	1.66	1.90	1.92	2.11	1.69
7	20	93	113	19	1.30	1.74	1.64	1.44	1.43	1.24
8	24	61	85	19	1.56	1.14	1.23	1.12	1.43	0.95
9	21	43	64	6	1.36	0.80	0.93	0.90	0.45	0.75
10	15	50	65	7	0.97	0.93	0.94	0.73	0.53	0.61
11	9	32	41	6	0.58	0.60	0.59	0.61	0.45	0.50
12	11	36	47	7	0.71	0.67	0.68	0.52	0.53	0.42
13	6	26	32	4	0.39	0.49	0.46	0.45	0.30	0.36
14	7	21	28	4	0.45	0.39	0.41	0.39	0.30	0.31
15	3	18	21	5	0.19	0.34	0.30	0.34	0.38	0.27
16	4	20	24	3	0.26	0.37	0.35	0.30	0.23	0.24
17	4	14	18	3	0.26	0.26	0.26	0.27	0.23	0.21
18	5	14	19	1	0.32	0.26	0.28	0.24		
19	3	14	17	0	0.19	0.26	0.25	0.22		
20	6	8	14	0	0.39	0.15	0.20	0.20		
21	0	9	9	1	—	0.17	0.13	0.18		
22	2	9	11	3	0.13	0.17	0.16	0.17		
23	4	4	8	0	0.26	0.07	0.12	0.15		
24	4	4	8	3	0.26	0.07	0.12	0.14		
25	0	9	9	2	—	0.17	0.13	0.13		
26	3	6	9	0	0.19	0.11	0.13	0.12		
27	1	7	8	1	0.06	0.13	0.12	0.11		
28	2	8	10	0	0.13	0.15	0.15	0.11		
29	2	6	8	0	0.13	0.11	0.12	0.10		
30	2	5	7	1	0.13	0.09	0.10	0.09		
31	0	3	3	0	—	0.06	0.04			
32	0	3	3	0	—	0.06	0.04			
33	3	3	6	0	0.19	0.06	0.09			
34	1	3	4	1	0.06	0.06	0.06			
35	0	0	0	0	—	—	—			
36	0	1	1	0	—	0.02	0.01			
37	0	1	1	1	—	0.02	0.01			
38	1	3	4	0	0.06	0.06	0.06			
39	0	3	3	0	—	0.06	0.04			
40	1	1	2	0	0.06	0.02	0.03			
41	0	1	1	0	—	0.02	0.01			

TABLE 1—CONTINUED.

NUMBER OF CONTRIBUTIONS n	NUMBER OF PERSONS MAKING STATED NUMBER OF CONTRIBUTIONS				PER CENT OF TOTAL					
	Chemical Abstracts			Auerbach's tables entire alphabet	Chemical Abstracts				Auerbach's tables	
					Observed			Computed[1]	Observed	Computed[2]
	Letter A	Letter B	A + B		A	B	A + B	A + B	Entire alphabet	
42	0	2	2	0	—	0.04	0.03			
43	0	0	0	0	—	—	—			
44	0	3	3	0	—	0.06	0.04			
45	0	4	4	0	—	0.07	0.06			
46	1	1	2	0	0.06	0.02	0.03			
47	0	3	3	0	—	0.06	0.04			
48	0	0	0	2	—	—	—			
49	0	1	1		—	0.02	0.01			
50	1	1	2		0.06	0.02	0.03			
51	0	1	1		—	0.02	0.01			
52	0	2	2		—	0.04	0.03			
53	0	2	2		—	0.04	0.03			
54	0	2	2		—	0.04	0.03			
55	2	1	3		0.13	0.02	0.04			
56	0	0	0		—	—	—			
57	0	1	1		—	0.02	0.01			
58	0	1	1		—	0.02	0.01			
59–60	0	0	0		—	—	—			
61	0	2	2		—	0.04	0.03			
62–65	0	0	0		—	—	—			
66	0	1	1		—	0.02	0.01			
67	0	0	0		—	—	—			
68	0	2	2		—	0.04	0.03			
69–72	0	0	0		—	—	—			
73	0	1	1		—	0.02	0.01			
74–77	0	0	0		—	—	—			
78	0	1	1		—	0.02	0.01			
79	0	0	0		—	—	—			
80	1	0	1		0.06	—	0.01			
81–83	0	0	0		—	—	—			
84	0	1	1		—	0.02	0.01			
85–94	0	0	0		—	—	—			
95	0	1	1		—	0.02	0.01			
96–106	0	0	0		—	—	—			
107	1	0	1		0.06	—	0.01			
108	0	0	0		—	—	—			
109	0	1	1		—	0.02	0.01			
110–113	0	0	0		—	—	—			
114	0	1	1		—	0.02	0.01			
115–345	0	0	0		—	—	—			
346	1	0	1		0.06	—	0.01			

[1] According to $f = 56.69/n^{1.888}$.
[2] According to $f = 600/\pi^2 n^2$.

tables. Determined by least squares, the slope of the curve to Auerbach's data, as determined from the first 17 points,[1] was found to be 2.021 ± 0.017. Similarly, the slope for the data in the Chemical Abstracts, letters A and B jointly, as determined from the first thirty points, came out as 1.888 ± 0.007. The general formula for the relation thus found to exist between the frequency y of persons making x contributions is

$$x^n y = \text{const} \tag{1}$$

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

$$y_1 = \frac{c}{1^2} \tag{2}$$

$$y_2 = \frac{c}{2^2} \tag{3}$$

$$y_n = \frac{c}{n^2} \tag{4}$$

$$\sum_1^\infty y = c\left(\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \dots\dots\right) \tag{5}$$

$$= c \sum_1^\infty \frac{1}{x^2} \tag{6}$$

$$= c\,\frac{\pi^2}{6} \tag{7}$$[2]

$$c = \frac{6}{\pi^2} \sum_1^\infty y \tag{8}$$

But, since y is a frequency, the summation $\sum_1^\infty y$ gives unity. Then finally

$$c = \frac{6}{\pi^2} \tag{9}$$

$$= \frac{6}{9.87} \tag{10}$$

$$= 0.6079 \text{ or } 60.79 \text{ per cent} \tag{11}$$

[1] Beyond this point fluctuations become excessive owing to the limited number of persons in the sample.

[2] 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 at the top of ninth column in Table 1.

Thus, according to the inverse square law, the proportion of all contributors who contribute a single item should be just over 60 per cent. In the cases here examined the actual proportion of this class to the whole was 59.2 per cent in Auerbach's data (1325 contributors), 57.7 per cent in the Chemical Abstracts under initial A (1543 contributors) 57.98 under letter B (5348 contributors) and 57.9 under letters A and B jointly (6891 contributors).

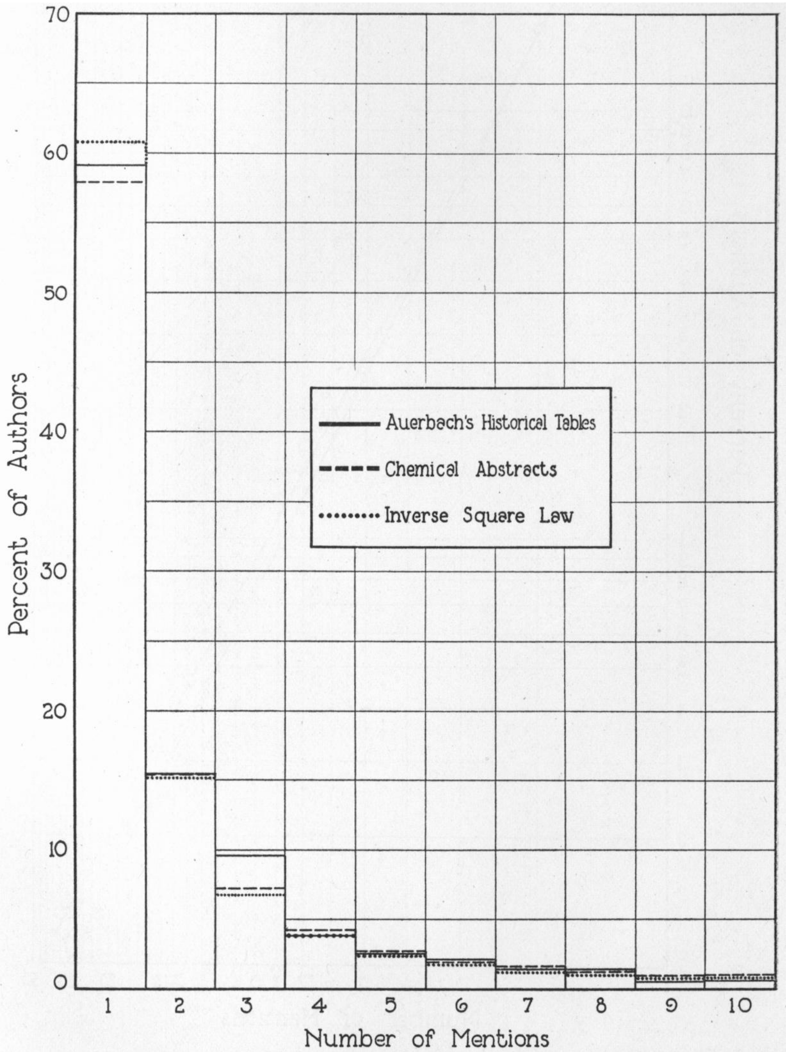

Fig. 1.—Frequency diagram showing per cent of authors mentioned once, twice, etc., in Auerbach's Geschichtstafeln der Physik, entire alphabet, and in the decennial index of Chemical Abstracts 1907–1916, letters A and B. The dotted line indicates frequencies computed according to the inverse square law.

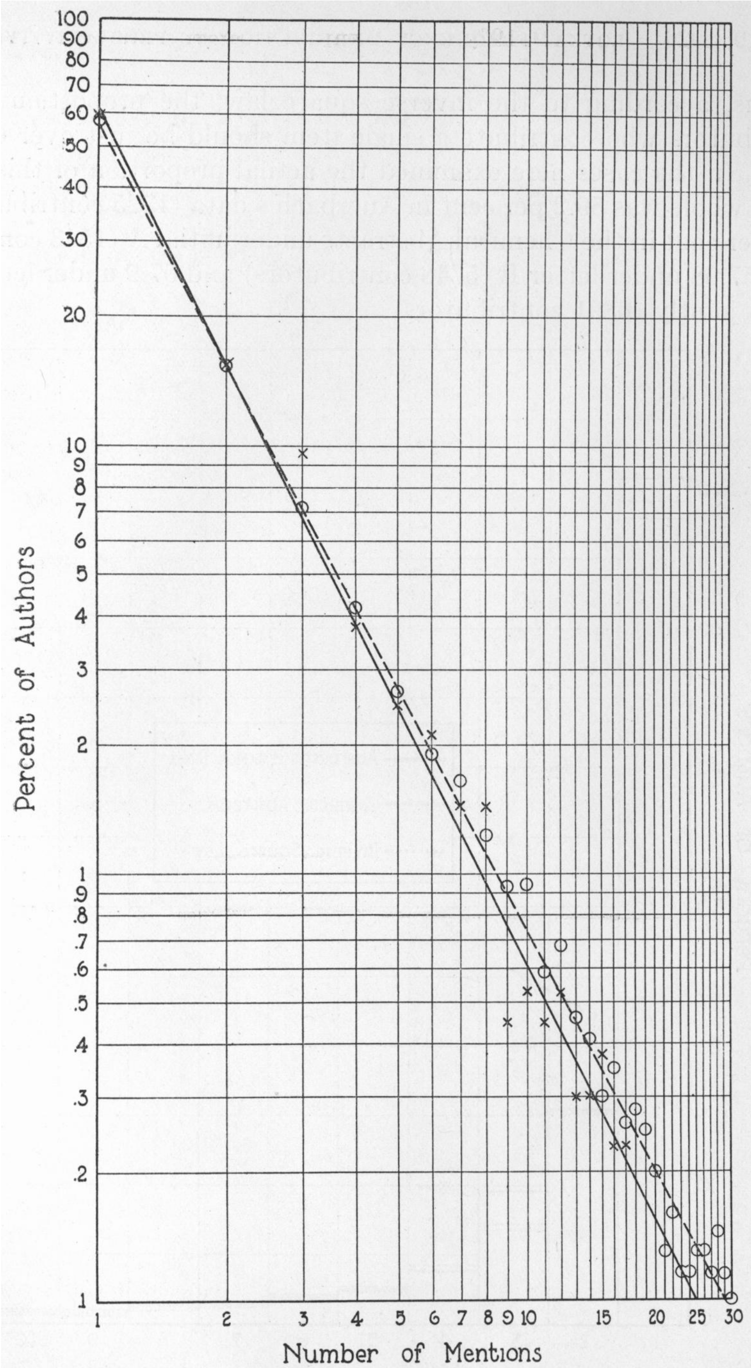

Fig. 2.—Logarithmic frequency diagram showing number of authors mentioned once, twice, etc., in Auerbach's tables (points indicated by crosses), and in Chemical Abstracts, letters A and B (points indicated by circles). The fully drawn line indicates points given by inverse square law, exponent = 2; the line of dashes corresponds to exponent 1.89.

Frequency distributions of the general type (1) have a wide range of applicability to a variety of phenomena,[3] and the mere form of such a distribution throws little or no light on the underlying physical relations.[4] 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;[5] and the proportion, of all contributors, that make a single contribution, is about 60 per cent.

The fact that two such widely different sources as Chemical Abstracts (listing practically all current work in chemistry over a ten year period) and Auerbach's tables (listing selected important contributions only, in physics, for all historical time) give very similar results, seems somewhat remarkable. It would be interesting to extend this study to such a work as Darmstaedter's Handbuch der Geschichte der Naturwissenschaften und der Technik. Unfortunately the index of this work does not indicate multiple entries of the same year under one author's name, but distinguishes only separately dated entries. It would therefore be necessary in each case to refer to the text. On the other hand the work could be abridged by restricting the inquiry to one or two letters of the alphabet, as was here done in the case of the Chemical Abstracts.

[3] Compare especially CORRADO GINI, *Biblioteca dell' Economista*, ser. 5a, **20**: *Indici di concentrazione e di dipendenza*. See also the Report of Commission of Housing and Regional Planning, State of New York, Jan. 11, 1926: 59–73; and *Income in the United States*, by W. I. KING and others; **2**: 344 et seq. 1922.

[4] C. J. WILLIS' conclusions regarding the mechanism of evolution, inferred as they are from the occurrence of curves of this type in the relation between numbers of species and genera, seem for this reason to carry little conviction. See A. J. LOTKA, *Physical Biology:* 311. 1925.

[5] Fortunately, however, there are somewhat more persons of very great productivity than would be expected under this simple law. The very high figures (e.g., Abderhalden, 346 contributions in ten years) should perhaps be considered separately, since they are not the product of one person unassisted. Joint contributions have in all cases been credited to the senior author only.