Proceedings of the National Academy of Sciences
Vol. 66, No. 1, pp. 18–24, May 1970

Statistical Theories of Success

Clarence Zener

CARNEGIE-MELLON UNIVERSITY, PITTSBURGH, PENNSYLVANIA

Communicated February 27, 1970

Abstract. The Pareto and the log-normal distributions are commonly used to describe the statistical distribution of success. A detailed comparison of these distributions is made with Lotka's extensive observations on success as measured by rate of publication. These distributions are found to adequately describe the observed distribution only for low and moderate success. Contrariwise, the flat factor analysis of performance recently developed by the author in these PROCEEDINGS (**59,** 1078 (1968)) is shown to give an excellent agreement over the whole range of success. Lotka's data allows a determination of the number of environmental factors.

The Pareto distribution does give an excellent agreement with the tail of the success distribution where success is defined as income. An interpretation of this distribution is here presented based upon the expected behavior of entrepreneurs.

---

1. **Introduction.** Whereas the response of any individual to his environment is unpredictable, significant conclusions can conceivably be derived from the statistical distribution of responses from a large number of people. It is in this spirit that I have recently developed in this journal a theory[1] to interpret Shockley's data[2] on the statistical distribution of rate of publication. In this theory, success (e.g., rate of publication) is represented by a product of statistically independent factors:

$$S = f_1 \cdot f_2 \cdot \cdot f_n. \tag{1}$$

The probability density $p_i(f_i)$ for each factor f_i is taken as constant from zero up to a critical value, and to be zero thereafter.

$$p(f_i) = \begin{matrix} 1, 0 < f_i < f_i^0 \\ 0, \text{otherwise} \end{matrix}. \tag{2}$$

This theory, with three factors, gave an excellent agreement with Shockley's data. These three factors were interpreted as curiosity, intellectual tools, and drive.

Unfortunately Shockley's samples were too small to decisively differentiate our formulation of success from the more usual Pareto[3] or log-normal[4] formulations. According to Pareto's law the number of people with a success equal to or greater than S is given by

$$N(S) \sim S^{-p}. \tag{3}$$

The Pareto constant p is a characteristic of the group studied. In the log-normal distribution, log S has a gaussian distribution. Thus, a plot of log S on probability paper gives a straight line. The log-normal distribution is in fact asymptotically approached by our flat factor formulation (1)–(2) as the number of factors n becomes large.

I have recently run across Lotka's study[5] of a large sample of 6891 authors, everyone, in fact, listed in *Chemical Abstracts* during 1907–1916 whose names started with either A or B. In section 2 we analyze this data in terms of the formulation (1)–(2), of Pareto's law, and of the log-normal distribution. This data decisively favors our factor formulation. Rice[6] has recently shown that his data on the publication rate of the 1539 members of the American Association of Clinical Chemists is also in agreement with our formulation (1)–(2). In section 2 we contrast this agreement with the disagreement of his data with the Pareto and the log-normal distribution. A comparison of the three sets of data which have now been analyzed—Shockley's, Lotka's, and Rice's—demonstrates the expected increase of n with the heterogeneity of the group studied.

Basic to our theory of success is not only the product function (1) but also the existence of an upper limit, given by (2), which any one of the basic factors can have. In section 3 we show that such an upper limit is consistent with recent psychological research.

Pareto himself proposed his law only for income distributions, specifically for the high income tail. This law, proposed in 1897,[7] was believed valid for all countries. An exhaustive study by the National Bureau of Economic Research[8] for the year 1918 shows this law to be remarkably accurate within the United States over the upper four cycles of ten in income. This high income group is populated primarily by entrepreneurs.[9] In section 4 we show that the expected entrepreneural behavior is such as to lead to the Pareto income distribution.

2. **Success as Measured by Publication Rate.** Davis[3] has used Lotka's data on publication rate as a classic example of Pareto's law. Identifying success S with the number of papers i published in a given interval of time, Davis used Pareto's law in the form

$$f(i) \sim i^{-(p+1)}$$

where $f(i)$ is the number of people who published precisely i papers. Following Davis we attempt to test Pareto's law by a plot of $f(i)$ versus i on log paper. From Figure 1 we see that the first six points lie with remarkable accuracy upon a straight line with a p of 0.89. For values of i above 6, statistical fluctuations appear; above 20, these fluctuations are so great as to render useless any attempt to test Pareto's law by visual inspection of the data.

In this type of data statistical fluctuations will always increase as $f(i)$ becomes small and, hence, as i becomes large. We can, however, ameliorate these fluctuations by working with the cummulative function

$$N(i) = \sum_{i}^{\infty} f(j).$$

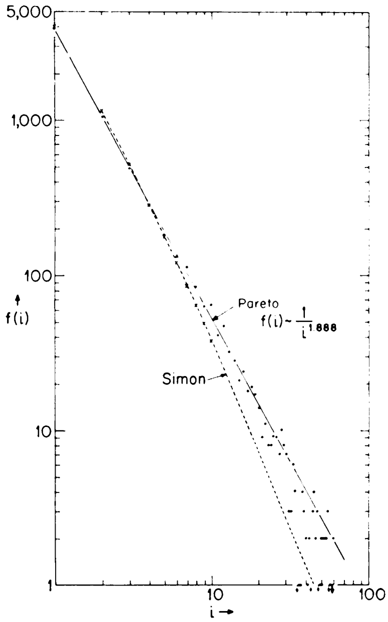

FIG. 1.—Lotka's data on publication rate. Comparison with Pareto's theory. Frequency plot.

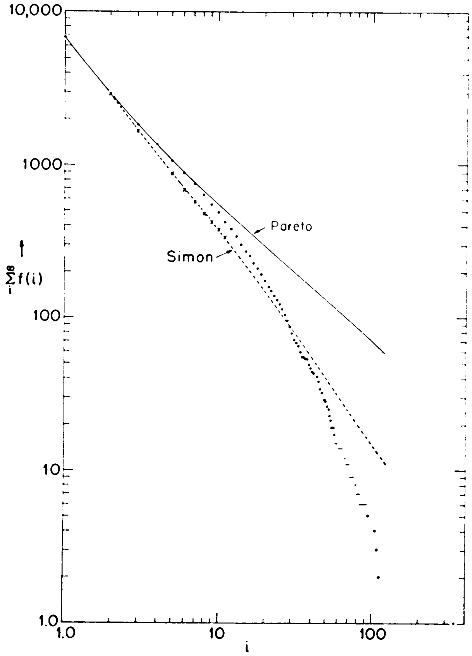

FIG. 2.—Lotka's data on publication rate. Comparison with Pareto's theory. Cumulative plot.

We accordingly plot this function in Figure 2. Whereas according to Pareto's law $N(i)$ should approach the straight line

$$N(i) \sim i^{0.89}$$

for large i's, it is seen that the data rapidly diverge away from this straight line. Simon[10, 11] has recently developed a stochastic theory of success which asymptotically approaches Pareto's law for large i's. He applied his theory to rate of publication, and evaluated those constants in his theory that are appropriate for the Lotka data. His theory, also reproduced in Figure 2, is seen to inadequately represent the observed data.

We now compare Lotka's data with a lognormal distribution. Toward this end we first observe that the authors in Lotka's data have been selected by the criterion of having published at least one paper. The appropriate measure of success must therefore be the number of additional papers. In the log-normal plot of Figure 3 we have accordingly taken the ordinate i as this number of additional papers. The abscissa represents the percentile coordinate x. Thus, the first 58% produced no papers other than that which admitted them into the group. The steps in the lower part of Figure 3 are due to the integer character of our measure of success, namely, numbers of papers. The steps in the upper part of the curve arise from the discrete values of x. Thus the last man, that is, the most successful man, is represented by the top plateau which extends indefinitely to the right.

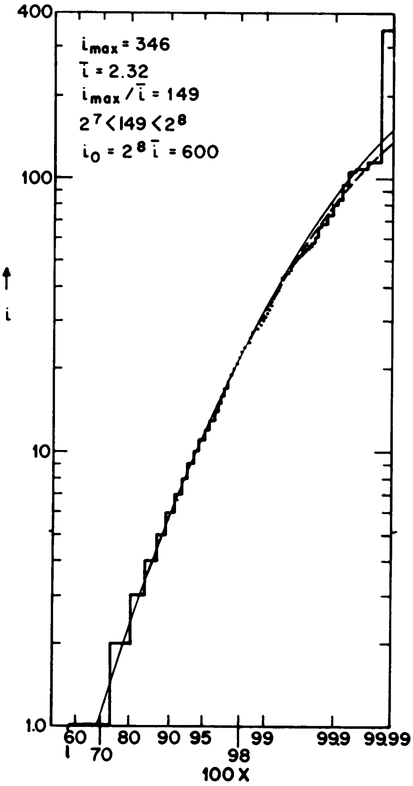

FIG. 3.—Lotka's data on publication rate. Comparison with flat factor analysis. *Solid curve:* $n = 8$. *Dashed curve:* $n = 7.8$.

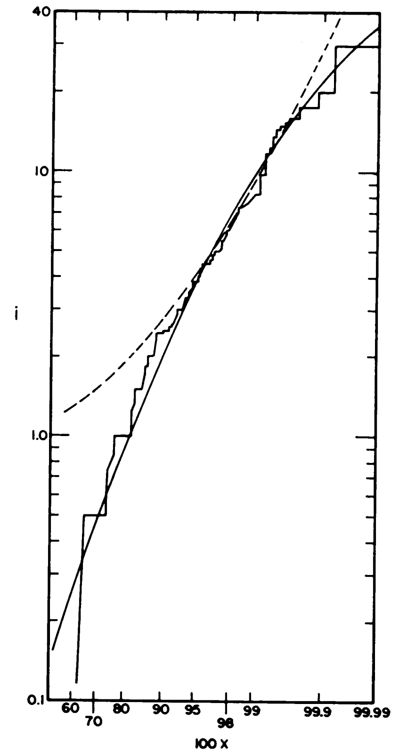

FIG. 4.—Rice's data on publication rate. Comparison with flat factor analysis: *solid curve*, $n = 7$. Comparison with Pareto's theory: *dashed curve.*

Under ideal conditions of a continuous measure of success and of an infinitely large sample, a log-normal distribution would give a straight line on log-normal paper. In our case of a discrete measure of success, and of a finite sample, a log-normal distribution would give a stepped curve. We could, however, pass a straight line through all the steps except perhaps some of the upper steps displaced by fluctuations. In contrast, Lotka's data in Figure 3 show a distinct negative curvature.

We next compare Lotka's data with the flat factor analysis formulated in (1)–(2). Toward this end we must first infer from the data the appropriate number of factors. Now the theoretical maximum S, which we denote by S_0, and the mean S are related by

$$S = (^1/_{2^n})\, S_0. \tag{4}$$

The observed maximum S_{max} is certainly less than S_0. If the group is large enough, however, the observed maximum will be greater than $S_0/2$. The appropriate value of n will then be the smallest n which satisfies

$$2^n > S_{max}/\bar{S}. \tag{5}$$

As before, we identify S with i. Since i_{max} has the value 345, and $\bar{i}$ the value 2.32, we infer that n is the smallest integer which satisfies

$$2^n > 149. \tag{6}$$

This integer is 8.

We next introduce the percentile coordinate x defined as that fraction of the group whose success is less than or equal to S. In my original paper[1] I showed that the flat factor analysis of performance formulated by (1) and (2) gives

$$x\ (S) = Q \int^{\infty} \Gamma_n\ (z)\ dz \tag{7}$$

where Γ_n is the incomplete gamma function, and

$$Q = \ln\ (S_0/S). \tag{8}$$

In Figure 3 the continuous curve gives the theoretical curve (7) with the empirical values for n of 8, and for

$$S_0 = 2^n\, S \tag{9}$$

of 600. This theoretical curve is seen to fit Lotka's data extremely well.

Our parametric equations for success, (7)–(9), contain the two parameters n and $\bar{S}$. Whereas in our derivation of these equations we considered n as an integer, we could obtain a better agreement if n was allowed nonintegral values. Thus, by allowing n to decrease from 8 to 7.8, we remove most of the discrepancies between the empirical data and the parameter solution (7)–(9), as is demonstrated by the dashed curve of Figure 3.

A relaxation of the constraint that n be an integer must reflect a relaxation of some of our basic assumptions. The most suspect of our assumptions is that all the factors have complete statistical independence. The major effects of relaxing this assumption is to lower n. This effect may be readily demonstrated by considering only two factors. Whereas with complete independence $n = 2$, the solution for complete correlation is given approximately by (7)–(9) with $n = 1.59$.

In Figure 4 we contrast the agreement demonstrated by Rice[6] for his data with our distribution based upon (1)–(2), with the marked disagreement of his data with Pareto's law and with the log-normal distribution. Here the constants in Pareto's law have been changed to give agreement with the central part of Rice's data. The log-normal law would of course give a straight line.

All members of Shockley's group worked within the same laboratory. His data for this group required an n of 3. Presumably all three factors referred to the necessary human factors for success in publication, such as curiosity, intellectual tools, and drive. All members of Rice's group belonged to the same professional society. As Rice has suggested, presumably the extra four factors required to represent his data refer to factors in the environment which are necessary for clinical biochemists to be successful as measured by publications. These factors might be, for example, time for research, facilities for research, technical help, and the presence of stimulating colleagues. All members of Lotka's group had published at least one paper which had been listed in *Chemical Abstracts*. Success as measured by reference in *Chemical Abstracts* must involve not only the three personal factors involved in Shockley's group, and the four environmental factors involved in Rice's group, but also an additional factor which measures the overlap of one's field of interest to the fields covered by *Chemical Abstracts*. One thereby arrives in a rational way at the eight factors required by Lotka's data.

3. Psychological Implications. An assumption basic to our theory is the existence of an upper limit to each of the basic factors. There is obviously an upper limit to the degree of overlap of one's research with the field covered by *Chemical Abstracts*. There is also obviously a limit to each of the environmental factors. It comes as a shock to one, like the author, unfamiliar with the working of the human mind, that the human factors also have a limit. The existence of such a limit is, however, in accord with the recent results of experimental psychology. The ability to perform basic mental tasks appears to have a definite upper limit, and this upper limit is not much above the average ability.[12, 13] Great variability is shown only in mental performances which require the consecutive successful performance of many basic mental tasks. As an example, suppose a difficult task requires 20 of the 120 basic mental characteristics postulated by Guilford.[14] An individual whose abilities in these 20 characteristics were uniformly distributed in the upper 50 percentile would be 2800 times as successful as the average individual. There would, however, be only one such individual in 2,000,000.

4. Entrepreneurial Income Distribution. We can readily show that the behavior which we commonly associate with the high income entrepreneurial class automatically gives rise to a Pareto income distribution. Crucial to our theory is that our typical entrepreneur intuitively thinks of money on a logarithmic scale, and will therefore enter a venture which has a 50% chance of increasing his capital by a given factor and a 50% change of decreasing his capital by the same factor. Thus, if this factor is 2, his venture gives a 50% chance of winning 100%, or of losing 50%, of his original capital. This intuitive logarithmic monetary scale will also determine how much he spends on living expenses. He will, in fact, spend a certain fraction of his capital, each year irrespective of its size.

In *Appendix A* we find that an entrepreneurial class behaving as described above will have a Pareto-type capital distribution with

$$p = 2\alpha/(\Delta \ln C)^2. \tag{10}$$

Here α is that fraction of one's capital which is annually absorbed in living expenses and in taxes, our $\Delta \ln C$ is the root mean square of the annual change in the logarithm of the capital due to entrepreneurial enterprise. Only the Pareto constant is known empirically. Neither α nor $\Delta \ln S$ is known individually. The best we can do to check our theory is to see if reasonable values of α and $\Delta \ln C$ satisfy (10). Upon taking the observed value of $^5/_3$ for p, and assuming α to be 0.1, we obtain for $\Delta \ln S$ the value of 0.35. This value of $\Delta \ln S$ corresponds to a 50% chance of a 0.42% gain in capital, a 50% chance of a 0.29% loss in capital. Such chances appear to be sufficiently favorable to entice entrepreneurial behavior.

The concept of income is, of course, inappropriate to our entrepreneur. He has only gains, and losses, of capital. We can take the income as defined by the National Bureau of Economic Research as proportional to the capital. A Pareto distribution of capital would then imply a Pareto distribution of this income.

In developing our theory we have assumed that the entrepreneurial activity does not change the $\overline{\ln C}$ of a group, but only the spread of $\ln C$. Our theory is, however, not sensitive to this assumption. A time rate of change of $\overline{\ln C}$ could be absorbed into our constant. We have demonstrated, however, that such a time rate of change is not necessary to encourage entrepreneurial activity. It is essentially this demonstration which distinguishes our analysis of the Pareto tail from prior stochastic analyses.[15]

Appendix A. Let $n(c)\ d \ln c$ be the number of entrepreneurs whose capital lies within the range of $d \ln c$. Now $n(c)$ will change with time because of two causes. The first is a diffusion due to entrepreneurial enterprise:

$$\left.\frac{dn}{dt}\right|_{D} = D\, d^2 n(c)/d \ln c^2.$$

The second is the change due to living expenses and taxes:

$$\left.\frac{dn}{dt}\right|_{L\&T} = \{-\, n(c) + n(c + \alpha c\delta T\}/\delta T = dn/d \ln c.$$

Equating to zero the sum of these two changes, we obtain the equation

$$(D\ d^2/d \ln c^2 + \alpha\ d/d \ln c)\ n = 0.$$

This equation has the solution

$$n(c) \sim e^{-p \ln c} \sim c^{-p}, \qquad p = \alpha/D.$$

Upon expressing the diffusion coefficient D in terms of the root mean square annual change in capital, $\Delta \ln c$

$$D = (^1/_2)\ (\Delta \ln c)^2$$

we obtain (10).

The author is indebted to Dr. Rice for being shown his manuscript prior to publication.

[1] Zener, C., these PROCEEDINGS, **59**, 1078 (1968).

[2] Shockley, W., *Proc. IRE* (*Inst. Radio Engrs.*), **45**, 279 (1957).

[3] Davis, H. T., in *The Analysis of Economic Time Series* (Bloomington: The Principia Press, 1941), chap. 9.

[4] Aitchison, J., and J. A. C. Brown, *The Lognormal Distribution* (Cambridge, England, Cambridge University Press, 1957).

[5] Lotka, A. J., *Jnl. Wash. Acad. Sci.*, **16**, 317 (1926).

[6] Rice, E. W., submitted to *Clinical Chemistry.*

[7] Pareto, V., *Cours d'economic politque* (Lausanne, 1897).

[8] *Income in the United States,* National Bureau of Economic Research (New York: Harcourt, Brace & Co., 1921), vols. 1 and 2.

[9] *Income in the United States,* National Bureau of Economic Research (New York: Harcourt, Brace & Co., 1921), vol. 2, pp. 376–380.

[10] Simon, H. A., *Biometrika,* **42** (1955).

[11] Simon, H. A., *Models of Man* (New York: John Wiley & Sons, 1957).

[12] Simon, H. A., in *The Sciences of the Artificial* (Cambridge, Massachusetts: The MIT Press, 1969), pp. 31–41.

[13] Miller, G. A., *Psychological Review,* **63,** 81–97 (1956).

[14] Guilford, J. P., *The Nature of Human Intelligence* (New York: McGraw-Hill, 1967).

[15] For review of literature, see: Steindl, J., *Random Processes and the Growth of Firms* (New York: Hufner Publishing Co., 1965).