Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 3380–3383, May 1982
Applied Mathematical Sciences

On 1/*f* noise and other distributions with long tails

(log-normal distribution/Lévy distribution/Pareto distribution/scale-invariant process)

ELLIOTT W. MONTROLL* AND MICHAEL F. SHLESINGER†

*Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742; and †La Jolla Institute, P. O. Box 1434, La Jolla, California 92038

Contributed by Elliott W. Montroll, February 5, 1982

ABSTRACT It is shown, following Shockley [Shockley, W. (1957) *Proc. IRE* 45, 279–290], that, when a population is engaged in tasks whose completion requires the successful conclusion of many independent subtasks, the distribution function for successes in the primary task is log normal. It is also shown that, when the dispersion of the log-normal distribution is large, the distribution is mimicked by a $1/x$ distribution over a wide range of x. This argument provides a generic set of processes that yields the much observed $1/x$ distribution, and will also lead to a $1/f$ noise spectrum. It is commonly found that distributions that seem to be log normal over a broad range (say to the 95th percentile of a population) change to an inverse fractional power (Pareto) distribution for the last few percentile. Annual income distributions are examples with this structure. The very wealthy generally achieve their superwealth through amplification processes that are not available to most. We have introduced a simple amplification model to characterize the transition from a log-normal distribution to an inverse-power Pareto tail.

log-normal distributions and $1/f$ distributions

$1/f$ "noise" has been observed in numerous systems. Whole conferences have been devoted to it, the most recent being that held at the National Bureau of Standards last year. The proceedings (1) of that meeting provide an excellent survey of the state of the field. It has been felt by some that this distribution should be a consequence of some simple generic stochastic process in the same sense that the "central limit theorem" of the theory of the probability states that, under certain rather weak conditions, the probability distribution of a sum of random variables is gaussian (or normal). One of the aims of this paper is to present such a process.

A purely $1/f$ distribution function cannot exist in the range $0 < f < \infty$ since it would not be normalizable; its normalization integral diverges as $\log f$ as $f \to 0$ and $f \to \infty$. Hence, if f is to extend over the positive half line, corrections must exist at both the small and large f extremes. A distribution that satisfies the requirement of being $1/f$-like in the intermediate range and yet remaining normalizable in the full range is the normal distribution of the variable $\log x$:

$$F(\log x) = \frac{\exp[-(\log x - \log \bar{x})^2/2\sigma^2]}{(2\pi\sigma^2)^{1/2}}, \qquad [1]$$

where $\bar{x}$ is the mean and σ^2 is the square of the dispersion of the distribution. Since $d \log x = dx/x$, the probability that the variable $x/\bar{x}$ lies in the interval $d(x/\bar{x})$ at $x/\bar{x}$ is

$$g(x/\bar{x})d(x/\bar{x}) = \frac{\exp[-(\log x/\bar{x})^2/2\sigma^2]}{(2\pi\sigma^2)^{1/2}} \frac{d(x/\bar{x})}{(x/\bar{x})}. \qquad [2]$$

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Generally, a $1/f$ distribution is demonstrated graphically on log-log graph paper by plotting $\log g$ as a function of $\log x$. For this purpose, we write

$$\log g(x/\bar{x}) = -\log(x/\bar{x}) - \{[\log(x/\bar{x})]^2/2\sigma^2\} - \tfrac{1}{2}\log(2\pi\sigma^2), \qquad [3]$$

the last term being a constant. Let us measure the variable x in multiples f of its mean value $\bar{x}$, with

$$x = f\bar{x}. \qquad [4]$$

Then Eq. 3 becomes

$$\log g(f) = -\log f - \tfrac{1}{2}[(\log f)/\sigma]^2 - \tfrac{1}{2}\log(2\pi\sigma^2). \qquad [5]$$

If the distribution $g(f)$ is to be $1/f$, then only the linear term in $\log g(f)$ should remain in Eq. 5, as would be the case as $\sigma \to \infty$. Let σ be large but finite and let f be expressed as a power n of e:

$$f = \exp n, \qquad [6]$$

then

$$\log g(f) = -n - \tfrac{1}{2}(n/\sigma)^2 - \tfrac{1}{2}\log(2\pi\sigma^2). \qquad [7]$$

When σ is large, we can estimate the largest integer value of n that allows Eq. 7 to be regarded as linear in n to within a prescribed precision (always omitting the constant term in the precision estimation). If the middle term on the right-hand side of Eq. 7 is to be less than a fraction θ of the first term, then

$$\tfrac{1}{2}(n/\sigma)^2 \le \theta\,|n| \text{ or } |n| \le 2\theta\sigma^2. \qquad [8]$$

Suppose that $\theta = 0.1$ and $\sigma = 5$. Then, for any $|n| \le 5$, $g(f)$ mimics a $1/f$ distribution to within 10%. This corresponds to 11 integer n values or, from Eq. 8, 11 e-folds, which is equivalent to four orders of magnitude. Generally, the function $g(f)$ mimics a $1/f$ distribution for $(4\theta\sigma^2 + 1)$ e-folds to within a relative error θ. Clearly, the larger σ, the more orders of magnitude the mimicking persists.

We have produced the normalizable log-normal distribution with large dispersion that approximates the $1/f$ distribution over a wide range. Does a simple stochastic model exist to form the basis of our log-normal distribution? The answer is yes, since Shockley (2) has already discussed such a mechanism (or model) in a completely different context.

Shockley was concerned with the measurement of productivity of research scientists, especially in the dispersion in their publication rates. He investigated the output of 88 research staff members of the Brookhaven National Laboratory and found the distribution function of the number of papers published by them to be log normal, thus having a long tail. Generally, he associated long tails with primary achievements that require the successful completion of numerous independent secondary subtasks, with the failure of any one being sufficient to cause the failure of the primary project. The publication of a scientific paper is an example of such a process.

Shockley explained that, to publish a paper, one must (*i*) have the ability to select an appropriate problem for investigation, (*ii*) have competence to work on it, (*iii*) be capable of recognizing a worthwhile result, (*iv*) be able to choose an appropriate stopping point in the research and start to prepare the manuscript, (*v*) have the ability to present the results and conclusions adequately, (*vi*) be able to profit from the criticism of those who share an interest in the work, (*vii*) have the determination to complete and submit a manuscript for publication, and (*viii*) respond positively to referees' criticism.

Given a set of investigators, let p_i be the probability that one of them is able to complete the *i*th step in the process. Then, assuming that the various probabilities are independent of each other, the probability P that our chosen investigator will produce a paper in a given time is the product of the probabilities that he successfully deals with each of the individual items,

$$P = p_1 p_2 \cdots p_8 \qquad [9a]$$

and

$$\log P = \log p_1 + \log p_2 + \ldots + \log p_8. \qquad [9b]$$

When the individual distributions of the log p_i values satisfy certain weak conditions (3) that include the existence of second moments, the central limit theorem is applicable, so that the distribution function of log P (and therefore of the sum) is the normal distribution. Hence, in the Shockley model, the distribution function for the number of papers published per research worker should be log normal, as was observed. Incidentally, this result would remain even if the Shockley list of attributes were not quite correct, or if certain important items were omitted, or new ones were added.

As noted above, the publication of a paper is an example of a complex task whose completion depends on the successful outcome of several independent subtasks. When the number of required subtasks is large, the number of terms in the sum analogous to Eq. **9b** is large, so that the central limit theorem is applicable to the characterization of the probability distribution of successes in a population seeking to complete a required complex task.

Consider a task whose success follows the completion of N subtasks. Then, the sum similar to Eq. **9b** for our process will contain N terms. Since the square of the dispersion, σ^2, of a sum of N independent random variables is the sum of the squares, σ_j^2, of the component random variables, σ^2 should be of order N, with $\sigma^2 = N\bar{\sigma}^2$, where

$$\bar{\sigma}^2 = \frac{1}{N}\sum_{j=1}^{N} \sigma_j^2. \qquad [10]$$

The greater the value of N, the greater the number of *e*-folds or decades over which the distribution function for task successes mimics a $1/f$ distribution.

$1/f$ noise from a $1/f$ distribution

The above analysis leads directly to a $1/f$ distribution. That such a scale-invariant distribution function of relaxation times [$\rho(\tau)d\tau = d\tau/\tau$] of random events leads to a $1/f$ noise spectrum was first pointed out by van der Ziel (4), who was interested in the noise characteristics of semiconductors. van der Ziel's model assumed that (*i*) the electronic motion was energy activated—i.e., $\tau \propto \exp(E/kT)$ with E a random variable—and (*ii*) within a certain energetic range, all energies E appear with equal likelihood. This model has been neglected because too large a range of a constant energy distribution is needed to reproduce the $1/f$ noise data (5). However, in a recent article (6) entitled "Earthquakes, Thunderstorms, and Other $1/f$ Noises," Machlup stressed that the $\rho(\tau)d\tau = d\tau/\tau$ part of van der Ziel's argument is correct, even if the particular physical model leading to this result may not be. Machlup views $1/f$ noise as a reflection that nature "is sufficiently chaotic to possess ... a large ensemble of mechanisms with no prejudice about scale." In this view, the semiconductor noise is just a particular example of the basic message that nature is scale invariant.

The connection between $\rho(\tau)$ and the noise spectrum is [again quoting from Machlup (6)] "... a purely random process is one with an autocorrelation of the form $\exp(-t/\tau)$. The power spectrum of such a process has the Lorentz shape:

$$S(\omega) \propto \text{Fourier transform of } e^{-t/\tau}$$
$$\propto \tau/(1 + \omega^2\tau^2).$$

If we have a large collection of different random processes, each with its own correlation time τ, then the power spectrum of the whole ensemble depends on the statistical distribution $\rho(\tau)$ of these correlation times. If these processes have not been filtered, then our conjecture is that the weighting function is scale-invariant:

$$\rho(\tau)\, d\tau \propto d\tau/\tau.$$

This gives a power spectrum

$$\int_{\tau_1}^{\tau_2} S_\tau(\omega)\rho(\tau)d\tau \propto \int_{\tau_1}^{\tau_2} \frac{\tau}{1 + \omega^2\tau^2}\frac{d\tau}{\tau}$$
$$= \left.\frac{\tan^{-1}\omega\tau}{\omega}\right|_{\tau_1}^{\tau_2}.$$

If the scale invariance extends over many orders of magnitude—i.e., if τ_2/τ_1 is a large ratio—then, the spectrum is $1/\omega$ over a correspondingly large range. For many years, we have been scratching our heads to find special mechanisms that would have that special distribution of time constants over many decades."

We propose that if, in a system of interest, the distribution of relaxation times is determined by a multiplicative process as characterized by Eq. **9**, then that distribution becomes log normal. However, as discussed above, for a considerable range, a log normal $\rho(\tau)$ is mimicked by $1/\tau$, as required by Machlup for his analysis.

Two popular sayings help us distinguish between those processes that should be characterized by normal distributions and those that should be characterized by log-normal distributions. This first is "... foot bone 'tached to the leg bone, leg bone 'tached to the knee bone, knee bone 'tached to the thigh bone, thigh bone 'tached to the hip bone," etc. With some dispersion in the length of member of each type bone in bones selected from a large population, the central limit theorem tells us that heights of individuals should be normally distributed.

Log-normal distributions would be expected in processes whose successful execution follow Franklin's proverb "for the want of a nail the shoe was lost, for the want of a shoe the horse was lost, for the want of a horse the rider was lost," etc.

The statistics of the flooding of the Nile has been a popular example of $1/f$ noise. We would explain this by considering the many stages through which a drop of rain at the source of a river must successfully pass to reach the mouth of the river. First, atmospheric conditions must lead to rain to create the drop and then wind, temperature, and ground porosity conditions must allow the drop to continue downstream at each stage of the river. The resulting log-normal distribution for the flow rate yields, in a natural fashion, a $1/f$ noise in the river level at the mouth.

Pareto and Lévy distributions

One of the earliest investigations of distributions with long tails was made by the social economist Pareto (1897) who collected statistics on the income and wealth of individuals in many countries at various times in history. His data convinced him that (7, 8) "In all places and at all times the distribution of income in a stable economy, when the origin of measurement is at a sufficiently high income level, will be given approximately by the empirical formula $y = ax^{-v}$, where y is the number of people having an income x or greater and v is approximately 1.5." A log–log plot of the data would yield a straight line of slope $-v$ (or -1.5 according to Pareto). As intimated by him, his formula is an asymptotic one valid for high incomes. In more modern times, the diligence of the Internal Revenue Service has given us data over the full income range. The data (9) for 1935/1936 (plotted on probability paper such that a log-normal distribution would yield a straight line; Fig. 1) indicate that the population in the 5–97 percentile range has an excellent fit to a log-normal distribution, while those in the last two or three percentile have an inverse power distribution as proposed by Pareto.

While it would seem that, for the range 5–97 percentile, which includes most of us, the Shockley style model would apply to the manner that we earn money, it is evident that those in the last 2 or 3 percentile operate in a somewhat different mode. They frequently collect their extreme wealth through some amplification process that is not available to the rest of us; that process varies from case to case. At the height of the Beatles' popularity, any new recording by them was soon sold to millions of fans. During periods of prosperity, free money becomes available for speculation—sometimes in real estate, sometimes in stock or silver, or even in tulips. A common characteristic of such times is that the daring may use their easy money to acquire some speculative commodity with a small margin payment (sometimes 10% or less of the value) plus promises to pay the remainder later. If the commodity doubles in price, a 10% margin payment is amplified to a nine-fold

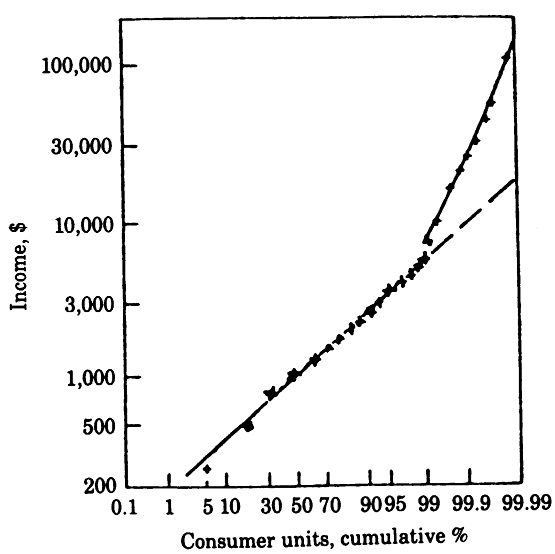

FIG. 1. Distribution of families and single individuals by income level, 1935/1936. Data are from ref. 9. Most of the data follows a log-normal distribution, while the last 1% is governed by a Pareto tail.

profit. Most readers can furnish other examples of amplification in the acquisition of superwealth. As the percentile level approaches 100, the number of remaining examples decreases. The few individuals left in the sample tend to become special cases in the dynamics of their success with increasing diversity of means appearing.

The Pareto distribution applies best to the tail of the income distribution. The diversity of style and performance in the amplification process of the superwealthy suggests that any statistical model used to characterize them should be without well-defined moments. Mandelbrot (10), the devoted champion of Pareto tails, identifies them with the distribution derived by his teacher Paul Lévy.

We can derive the Pareto–Lévy tails from our log-normal distribution by accounting for the process of amplification (and amplification of the amplification, etc.). Let $g(x)$ denote our log-normal distribution, whose mean value we denote by $\bar{x}$. With a probability proportional to λ, let $g(x)$ be amplified such that its mean value is $N\bar{x}$—i.e., $g(x/\bar{x})dx/\bar{x} \rightarrow g(x/N\bar{x})dx/N\bar{x}$. Again, with a probability proportional to λ, let the amplification be amplified so the mean value of the distribution will be $N^2\bar{x}$. The new distribution $G(x)$ that allows for these amplifications is, with x in units of $\bar{x}$,

$$G(x) = \left[\alpha + \frac{\lambda}{1-\lambda}\right]^{-1}\left[\alpha g(x) + \frac{\lambda}{N}g(x/N) + \frac{\lambda^2}{N^2}g(x/N^2) + \ldots\right],$$

where α determines the range of the initial log-normal behavior and α and the first term on the right-hand side are chosen to ensure the proper normalization of $G(x)$. While one can solve for $G(x)$ in terms of $g(x)$ by using the method of Mellin transform, here we just demonstrate that $G(x)$ has a nonanalytic part leading to the Pareto–Lévy tail. This is accomplished by noting that $G(x)$ satisfies the recursion relation

$$G(x) = \frac{\lambda}{N}G(x/N) + \left[\alpha + \frac{\lambda}{1-\lambda}\right]^{-1}\left[\alpha g(x) + (1-\alpha)g\left(\frac{x}{N}\right)\right].$$

Since $g(x)$ is assumed to be analytic, any singular behavior satisfies the equation

$$G_s(x) = \frac{\lambda}{N}G_s(x/N),$$

which has the solution

$$G_s(x) = x^{-1-\mu}A(x),$$

where $\mu = \ln(1/\lambda)/\ln N$ and $A(x) = A(x/N)$—i.e., $A(x)$ is oscillatory, periodic in $\log x$ with period $\log N$. We note that μ appears here naturally in the form of a fractal dimension (11, 12) and is $\approx 3/2$ for the income distribution in Fig. 1. If $\lambda N \geq 1$, then $0 < \mu < 1$ and $G(x)$ represents the distribution function of a divergent branching (bifurcating) process in which the mean value of $G(x)$ is infinite. If $\lambda N < 1$ but $\lambda N^2 > 1$, then $1 < \mu < 2$ and the mean of $G(x)$ will be finite. However, the fluctuations about the mean will be finite. The connection to the tail of the Lévy distribution can now be made.

Lévy (13) investigated the statistics of a sum of independent random variables

$$X_n = x_1 + x_2 + \ldots + x_n \qquad [11]$$

for cases inapplicable for the central limit theorem. For simplicity, we postulate the average value of each x_j to vanish, $\langle x_j \rangle = 0$. In the case that each x_j has a gaussian distribution with $\sigma_i^2 = \langle x_i^2 \rangle$, the distribution function of the sum X_n is also gaussian with the dispersion function

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2. \quad [12]$$

Lévy considered the case of the x_js being devoid of second and higher moments and whether there exist non-gaussian distributions common to each x_j such that X_n will have the same type of distribution as the individual x_j. This is the case for the Cauchy distribution

$$F_j(x) = (a_j/\pi)(a_j^2 + x^2) \text{ for the } j\text{th } x. \quad [13]$$

Then the distribution for X_n is

$$F_j(x) = (A/\pi)(A^2 + x^2) \qquad (A = a_1 + \ldots + a_n). \quad [14]$$

In general, Lévy showed that, when the distribution of the individual x_js is given by the Fourier integral

$$F_j(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx} \exp(-|k|^\alpha a_j dk) \qquad (0 < \alpha \leq 2), \quad [15]$$

the distribution of X_n is

$$F(x) = \frac{1}{2\pi}\int_{\infty}^{\infty} e^{ikx} \exp(-|k|^\alpha A dk) \qquad (0 < \alpha \leq 2), \quad [16]$$

with

$$A^\alpha = a_1^\alpha + a_2^\alpha + \ldots + a_n^\alpha. \quad [17]$$

The Cauchy case (Eq. 13) corresponds to $\alpha = 1$ and the Gauss case corresponds to $\alpha = 2$. The Gauss distribution is singular. It is the only case that enjoys the existence of moments for x_j and for X_n. When $\alpha < 2$, it is not difficult to show (3) that, for large x,

$$F(x) \sim \frac{\alpha a}{\pi x^{\alpha+1}} \Gamma(\alpha) \sin\frac{1}{2}\pi\alpha \qquad (0 < \alpha < 2), \quad [18]$$

whose inverse power character shows $F(x)$ to have the Pareto form.

This research was partially supported by the Advanced Research Projects Agency of the U.S. Department of Defense.

1. Meijer, P. H. E., Mountain, R. D. & Soulen, R. J., Jr., eds. (1981) *Sixth International Conference on Noise in Physical Systems* (National Bureau of Standards, Washington, D.C.), special publ. no. 614.
2. Shockley, W. (1957) *Proc. IRE* **45**, 279–290.
3. Feller, W. (1966) *An Introduction to Probability Theory and Its Applications* (Wiley, New York), Vol. 2, pp. 491, 548.
4. van der Ziel, A. (1950) *Physica* **16**, 359–372.
5. Bell, D. A. (1960) *Electrical Noise* (Van Nostrand, London), p. 227.
6. Machlup, S. (1981) in *Sixth International Conference on Noise in Physical Systems*, eds. Meijer, P. H. E., Mountain, R. D. & Soulen, R. J., Jr. (National Bureau of Standards, Washington, D.C.), special publ. no. 614, pp. 157–160.
7. Pareto, V. (1897) *Cours d'Economie Politique* (Lausanne).
8. Davis, H. T. (1941) *Theory of Econometrics* (Bloomington, Indiana).
9. U.S. National Resources Committee (1939) *Consumer Incomes in the U.S. 1935–1936* (Government Printing Office, Washington, D.C.).
10. Mandelbrot, B. B. (1977) *Fractal: Form, Chance, and Dimension* (Freeman, San Francisco).
11. Hughes, B. D., Shlesinger, M. F. & Montroll, E. W. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3287–3291.
12. Shlesinger, M. F. & Hughes, B. D. (1981) *Physica* **A109**, 597–608.
13. Lévy, P. (1937) *Théorie de l'Addition des Variables Aleatoires* (Gauthiers-Villars, Paris).