Proportional growth with or without diffusion, and other explanations of scaling

✦ **Abstract.** However useful and "creative" scaling may be, it is not accepted as an irreducible scientific principle. Several isolated instances of scaling are both unquestioned and easy to reduce to more fundamental principles, as will be seen in Section 1. There also exists a broad class of would-be universal explanations, many of them variants of proportional growth of U, with or without diffusion of log U. This chapter shows why, countering widely accepted opinion, I view those explanations as unconvincing and unacceptable.

The models to be surveyed and criticized in this expository text were scattered in esoteric and repetitive references. Those I quote are the earliest I know. Many were rephrased in terms of the distribution of the sizes of firms. They are easily translated into terms of other scaling random variables that are positive. The two-tailed scaling variables that represent change of speculative prices (M 1963b{E14}) pose a different problem, since the logarithm of a negative change has no meaning. ✦

THIS CHAPTER MEETS HEAD-ON the legitimate and widespread wish to explain the prevalence of scaling in finance and other fields.

Section 1 describes scattered instances in which scaling is explained fully by a brief and mathematically straightforward argument: by eliminating an intrinsic variable between two intrinsically meaningful exponentials. Because there are so few of them, those instances acquire an otherwise undeserved standing.

Section 2 provides a careful analysis and critique of a typical attempt to explain scaling by multiplicative diffusion of U, that is, ordinary ("Fickian") diffusion of $\log U$. Deep reservations about this approach motivated a preference for viewing scaling as a postulate that brings economies of thought, good fit and a useful basis for practical work. The scaling distribution must be thoroughly understood, and consequences explored, without waiting for explanation.

Section 3 describes a classic first-order model, in which randomness is limited to a transient, and models involving permanent randomness.

Reasons why this text came to be. In the late 1950s and early 1960s, much of my research concerned scaling in economics. But questions I found fascinating kept being described by others as not worth detailed study. Stated in today's vocabulary, the reason was that scaling was viewed as haging been explained by very simple arguments involving a "principle of proportional effect." Therefore, scaling "should be expected," and there was not much to it. M 1982F{*FGN*} alludes to those events in Chapter 42, titled "The Path to Fractals."

My constantly restated response involved several separate points.

A) A successful explanation could be described as proceeding "upstream" from scaling, while the consequences scaling proceed "downstream." In the case of scaling, upstream considerations happen not to affect my downstream investigations and the latter prove to be surprising or even shocking, therefore, extremely worthwhile.

This first response rarely convinced my prospective audience, making additional responses necessary.

B) A careful look shows that, with a few exceptions, the existing upstream explanations are oversold in one or more of several ways.

B1) Nearly all lean heavily on probability limit theorems that concern the state of a system in a "long-run when we shall all be dead" (to quote J. M. Keynes once again). Arbitrarily set initial conditions do not affect the limit, but the pre-asymptotic behavior may be poorly approximated by the theoretical asymptotics. Chapter E9 shows that this "defect" is especially nefarious when the limit is lognormal.

B2) In physics, the long-run is attainable, and "universality" often prevails. This grand word means that the same result is obtained from seemingly different assumptions, therefore details do not matter much. In particular, the interaction between particles leads to the same equilibrium distribution whether the total energy of a system is fixed, or allowed to fluctuate. To the contrary, would-be models of scaling outside of physics are overly sensitive to the choice between these last two assumptions, as will be argued in Section 2.

B3) In addition to clearly stated assumptions, many explanations of scaling use additional hypotheses that are rarely stated and by no means compelling. Seemingly imperceptible changes of unstated assumptions often yield a final outcome that is *completely different* and *usually non-scaling*. For example, they generate the lognormal distribution, instead of the scaling distribution that one wishes to explain.

B4) Some would-be "explanations" are circular and/or mathematically incorrect. It will be seen that diffusion demands *second-order* differential equations, yet some authors inplicitly believe that the same effects can be obtained by using a *first-order* difference equation. To achieve a scaling output, such models must begin with a scaling input.

The above-listed features were either not noticed or not appreciated. Reluctantly, I wrote M 1959s, which overwhelmed many readers, then M 1963g, which took a broader view. But at that point, the sudden wide interest in M 1963b{E14} seemed to eliminate the need for M 1963g, and that text was left unpublished. Unfortunately, the hopes that led to questionable models prove to be durable. This motivates the present text, loosely based on M 1963g, with two short papers added in appendices. This survey leaves aside a very novel generation of scaling via multifractals (M 1972j{N14}); Chapters E1 and E6 mention and use them to model price variation.

Aside from Section 1.2.4, and the criticism of diffusion models, this chapter surveys the work of others. It does not pretend to be either exhaustive, or entirely accurate in terms of historical credit. But it hopes to discourage the tedious process of piecemeal and independent discovery of models that are essentially equivalent and equally unconvincing. This text should provide the reader with background to tackle other questions of interest.

1. EXCEPTIONAL SCALING DISTRIBUTIONS THAT ARE COMPLETELY EXPLAINED IN A FEW LINES

Some models sketched in this Section are properly random, others obtain the scaling distribution by straightforward elimination of a common variable between two exponential relations.

1.1 Properly random models

Section 1.1.1. and 1.1.2. describe a few examples that take only a moment and yield scaling with $\alpha = 1$. Section 1.1.3 refers to a classical example.

1.1.1. Ratios of independent random variables and the effects of small denominators. *Ratios of Gaussian variables*. The ratio $R = Y/X$ of two independent Gaussian variables X and Y with $EX = EY = 0$ and $EX^2 = EY^2 = 1$ is well-known to be a Cauchy variable, with the density $1/[\pi(1+z^2)]$. Here is a perspicuous geometric proof. It follows from the assumptions on X and Y that $Z = X + iY$ is an isotropic random variable. Hence, its curves of equal probability density are circles, and $\theta = \tan^{-1}(Y/X)$ is uniformly distributed from 0 to 2π. The density of $R = Y/X$ follows immediately. The Cauchy density is only asymptotically scaling of exponent $\alpha = 1$ (it is L-stable).

Ratios of exponential variables. When X and Y are both exponential with $EX = EY = 1$, the curves of equal probability density of $Z = X + iY$ are no longer circles (Z is not isotropic), but lines with an equation of the form $X + Y =$ constant. For every given $X + Y$, hence also unconditionally, it follows that the ratio $(Y - X)/(Y + X)$ is uniformly distributed from –1 to 1. That is, U being defined as uniform from 0 to 1, R satisfies

$$\frac{Y-X}{Y+X} = \frac{R-1}{R+1} = 1 - \frac{2}{R+1} = 1 - 2U, \text{ hence } R = \frac{1}{U} - 1.$$

Conclusion, $R + 1$ follows *exactly* the scaling distribution of exponent $\alpha = 1$. No reference was found for this unbeatably simple result, but it would be surprising if it were new.

The scaling character of ratios whose denominators are often enough very small. In the 1960s, interest was aroused by econometric techniques that conclude with ratios having a small denominator and infinite moments. Those moments' divergence was viewed as an anomaly to be avoided, but anyone interested in explaining scaling should hold the precisely opposite view. Indeed, "it would suffice" to reexpress scaling quantities as ratios in which the denominator can be small. Students of mechanics know, at least since Poincaré, that small denominators lead to chaotic behavior.

1.1.2. Car queues on a one-lane road. Starting with many cars trying to maintain a constant randomly selected speed, the system will evolve to one in which cars queue behind a slow car. What will be the steady-state distribution of the length of this queue? In the simplest case, the intended speeds are independent and identically distributed random variables U_m with $\Pr\{U < u\} = F(u)$. The length of the queue is the first value of m such that $U_m < U_1$. When u_1 is known, $\Pr\{M = m \mid U = u_1\} = [1 - F(u)]^{m-1} F(u_1)$. Assume that there is a zero probability of anyone trying to drive at some minimum speed. Thus,

$$\Pr\{M=m\} = \int dF(u)[1-F(u)]^{m-1}F(u) = \frac{1}{m} - \frac{1}{m-1}\ ; \Pr\{M \geq m\} = \frac{1}{m}\ .$$

The queue length follows *exactly* the scaling distribution of exponent $\alpha = 1$.

1.1.3. Number of tosses of a coin before a prescribed high gain level is first reached. It is well-known that this number's tail probability is $\sim u^{-1/2}$, which is asymptotically scaling with $\alpha = 1/2$. A concrete application of this distribution is inserted at the end of this Chapter as Appendix I.

1.2 Straightforward elimination of an intrinsic variable between two intrinsically meaningful exponentials

This section proceeds in more or less chronologic order, tackling mutations, gravitation cosmic rays and words in discourse.

1.2.1. Attraction from within a very thin cone. In a Euclidean space of dimension E, create a cloud of unit point masses called stars, pick one star as the origin Ω and draw a very thin one-sided cone with apex at Ω and height R. If fluctuations are neglected, the proportion of stars found in this cone and also in a sphere of radius ρ is $(\rho/R)^E$. Assume that attraction follows the generalized Newton's law $u = \rho^{-N}$. Eliminating ρ, we find

$$\text{Prob }\{\text{attraction} > u\} = R^{-E}(u^{-1/N})^E = R^{-E}u^{-E/N}.$$

This is a scaling distribution. Three arbitrary features are the cutoff at a finite R, the restriction to a very thin cone, and the fact that we only consider the attraction from one star other than Ω.

Holtsmark's problem. Appendix III of M 1960i{E10} carries out a full argument for the physical case $E = 3$ and $N = 2$, but the same method generalizes without problem to all cases where $N > E/2$.

1.2.2. Bacterial mutations (Luria & Delbrück 1943). Disregarding the fact that numbers of bacteria are random and are integers, choose the time unit so that bacteria multiply and mutate deterministically and exponentially as follows: In the absence of mutation, a culture that contains b_0 bacteria at time $t = 0$ contains $b_0 e^t$ bacteria at time t. In the presence of mutations at the rate m, the size of the culture becomes $b_0 \exp[t(1-m)]$. Also suppose that a clone that descends from a mutation attains the size $\exp(gx)$ at age x. The number of clones will increase without end, the biggest clone being the oldest one.

If a mutant clone is picked at random among those present after a long time t, $\Pr\{\text{age} \geq x\} = \exp[-x(1-m)]$, hence $\Pr\{U \geq u\} = \exp(-gx)$. Eliminating x, we find in the limit the scaling distribution

$$\Pr\{U \geq u\} = u^{-D} \text{ with } D = (1-m)/g.$$

This argument helped molecular biology take off. A full and explicit treatment taking account of fluctuations was first provided in M 1974d, which is incorporated in this chapter as Pre-Publication Appendix II.

1.2.3. The energy of incoming cosmic rays (Fermi 1949.) The energy of primary cosmic rays is found to follow a scaling distribution with $\alpha \sim 1.7$ over 11 units of $\log_{10}$ (energy). The breadth of this range lies beyond economists wildest dreams; even in physics it is rarely encountered.

Fermi's assumptions easily translate into terms of firm growth, disappearance and replacement: all firms that exceed the size $\tilde{u}$ grow exponentially, until they die, but the population is replenished at a uniform rate to insure a steady-state distribution of sizes. We choose the unit of time so that T units of time after the size $\tilde{u}$ was exceeded

$$\text{firm size } = u(T) = \tilde{u}e^{T}.$$

Independently of its size $u \geq \tilde{u}$, a firm is given the probability αdT of disappearing during the time increment dT. It follows that the average "lifetime" of a firm is $1/\alpha$, and that

$$\Pr\{\text{ a firm survives for the time } > T\} = \exp(-\alpha T).$$

Under these assumptions,

$$\Pr\{U > u\} = \Pr\{T > \log(u/\tilde{u})\} = \exp(-\log(u/\tilde{u}) = (u/\tilde{u})^{-\alpha}.$$

This means that $\Pr\{\text{size} > u\}$ is scaling with the exponent α. Q. E. D.

The expected change of a firm's size. The probability of dying out in the next unit of time is αdt, and the firms that do not die out will grow by the factor $(1 + dt)$. Hence, neglecting second-degree terms

$$E[U(t+dt) \mid u(t)] = u(t)(1-\alpha dt)(1+dt) = [1 + (1-\alpha)dt - (dt)^2]u(t),$$

hence

$$E[U(t+dt)\,|\,u(t)] - u(t) = (1-\alpha)dt.$$

Thus, an overall steady state will be established. But a given firm will, on the average, increase in size if $\alpha < 1$, and decrease in size if $\alpha > 1$. If $\alpha = 1$, the expected change of $U(t)$ vanishes; $U(t)$ is a martingale (see Chapter E1 or M 1966b{E19}).

A Fermi-like model with variable immigration rate $\varphi(t)$. If $\varphi(t) = \exp(\beta t)$, the expected number of firms of size u is

$$\int_0^\infty \exp(-\alpha T)\exp[\beta(t-T)]dt = \lambda \exp(at)\left(\frac{u}{\tilde{u}}\right)^{-(\alpha+\beta)} u^{-1}du.$$

If $\alpha + \beta > 0$, the number of firms surviving to time t will have a finite expected value equal to

$$\lambda \int_0^\infty \exp(-\alpha T)\exp[(t-T)]dt = \frac{\lambda \exp(\beta t)}{(\alpha+\beta)}.$$

Contrary to Fermi's original process, the number of firms now increases without bound, and there is no steady-state. But the expected *relative* number of firms of size u is scaling with exponent $\alpha + \beta$.

More generally, the distribution of firm sizes is the Laplace transform of the immigration rate $\varphi(t)$. This model may yield *any distribution of firm sizes*, as long as the inverse Laplace transform is positive.

1.2.4. A different example of straightforward elimination between two exponentials: word frequencies, lexical trees, and "Zipf's" law (M 1951, M 1982F{FGN}, Chapter 38.) Take an alphabet of $M+1$ letters L_m, with L_0 denoting the improper letter "space". Let "typing monkeys" use this alphabet to produce a random text in which L_0 is used with the probability p_0, and each of the other letters, with the probability $(1-p_0)/M$. There will be M^k distinct words made of k proper letters followed by space, each with the probability

$$p = p_0[(1-p_0)/M]^k = p_0 e^{-k \log B}, \text{ by definition of } B = (1-p_0)/M.$$

The most probable word, corresponding to $k=0$, is the shortest one. Now rank the other words by decreasing probability. In this ordering, a k -letter word will have a rank satisfying

$$r \propto M^k; \text{ that is } k = \frac{\log r}{\log M}.$$

Eliminating k between r and p yields the scaling distribution

$$p \propto p_0 \exp(- \log r \frac{\log B}{\log M}) = P_0 r^{-1/\alpha},$$

with

$$\frac{1}{\alpha} = \frac{\log B}{\log M} = \frac{-\log(1-p_0) + \log M}{\log M} = 1 + |\log_M (1-p_0)| > 1.$$

This scaling expression was found to be obeyed by words in large but homogenous samples of homogenous natural discourse, such as some long books. In the (unattainable!) limit $\alpha = 1$, this expression is called *Zipf's law for word frequencies*.

There is *nothing more* to Zipf's law with $\alpha > 1$. The derivation merely relies on compensation between two exponentials.

Markovian discourse and other generalizations yield the same result for $r \to \infty$ (M 1954b) But the probability distribution of "m-grams" formed by m letters *is not expected to be scaling*. Scaling does not take over until after the m-grams for all values of m have been sorted out in order of decreasing probability.

For word frequencies, the compensation between two exponentials can be rephrased in several ways. There is a "thermodynamical" or "information-theoretical" restatement (M 1982F{*FGN*}, Chapter 38); it looks learned and is enlightening to the specialist, but brings nothing new for most readers.

(In an amusing tongue-in-cheek etymology, Lee Sallows described the number of letters in a word as being its logorithm, from *logos*=word and *arithmos*=number. Thus, the simplest model described above assigns to a word a *logorithm* that is proportioned to the *logarithm* of its inverse probability.)

2. "RANDOM PROPORTIONATE EFFECT" AND ITS FLAWS

Improvements that inject randomness in the models of Luria-Delbruck and Fermi will be examined in Sections 3.2 and 3.3. But we begin by considering a widely popular approach that recreates the conditions of Section 1.2, by introducing an artificial variable that enters into two exponentials, and can be eliminated to yield the scaling.

2.1 Introduction to models in which the logarithm of the firm size performs a discrete random walk, a Brownian motion, or a diffusion

The overall scheme is familiar: when U follows the scaling distribution $\Pr\{U > u\} \sim (u/\tilde{u})^{-\alpha}$, the auxiliary variable $V = \log_e(U/\tilde{u})$ satisfies the exponential distribution $\Pr\{V \geq v\} = \exp(-\alpha v)$. To explain U by explaining V, one must a) motivate the transformation from U to V, and b) explain why V should be exponential. Task a) is difficult, but task b) seems formally very easy, because the exponential distribution plays a central and well-understood role in physics. It is not a surprise that a number of models of scaling are more or less obvious and/or conscious economic translations of various classical models of statistical thermodynamics.

Section 3.1 will classify those models as being of either the first or the second order and, therefore, as leading to a flow or a diffusion. But certain issues must first be faced, and this is done best by focusing on a bare bones diffusion model.

2.1 Tempting dynamic explanation of the exponential distribution for V by diffusion contained by a reflecting barrier

We begin with a random walk that is the simplest example of second order and diffusion. The physics background is a collection of particles that a) are subjected to a uniform downward gravity force, b) form a gas at a uniform temperature and density, and c) are constrained to remain in a semi-infinite vertical tube with a closed bottom and an open top. These particles' final equilibrium distribution will be a compromise: a) *gravity* alone would pull them down, b) *heat motion* alone would diffuse them to infinity, and c) the tube's *bottom* prevents downward diffusion. The result is classical in physics: acting together, these three tendencies create at the height z an exponential density distribution of the form $\exp(-\alpha z)$, where $1/\alpha$ increases with the temperature. Therefore, scaling could be explained by any model which (consciously or not) will re-interpret the above three physical forces in terms of economic variables.

2.2. A surrogate for diffusion of log *U*, based on a biased random walk

Let time t be an integer and $\log_e U$ be of the form kc, where k is an integer and $c > 0$. Between times t and $t+1$, allow the following possibilities:

a) $\log_e U$ can increase by c, the probability being p;

b) $\log_e U$ can decrease by c, the probability being $1-p$.

A necessary condition for equilibrium. For a distribution of $\log U$ to be invariant by the above transformations, it is *necessary* that the expected number of firms growing from size e^{kc} to size $e^{(k+1)c}$ be equal to the expected number of firms declining from size $e^{(k+1)c}$ to size e^{kc}. This can be written

$$\frac{\Pr\{\log_e U > (k+1)c\} - \Pr\{\log_e U > kc\}}{\Pr\{\log_e U > kc\} - \Pr\{\log_e U > (k-1)c\}} = \frac{p}{1-p} .$$

A steady-state solution in which $P(\log U > \nu) = \exp(-\alpha v)$ requires $\exp(-\alpha c) = p/(1-p)$. The condition $\alpha > 0$ requires $p < 1-p$ or $p < 1/2$ (this is an economic counterpart of the "force of gravity" referred to in section 2.1). Taken by themselves, the conditions a) and b) make firm sizes *decrease* on the average. Hence, to insure that the number of firms remains time-invariant above a lowest value $\tilde{u}$, one needs an additional factor, the counterpart of the closed bottom in Section 2.1.

c) A reflecting barrier can indifferently be interpreted in either of two principal ways. The firms that go below $\log_U = \log_e \tilde{u} - (1/2)c$ are given a new chance to start in life at the level $\tilde{u}$, or become lost but replaced by a steady influx of new firms starting at the level $\tilde{u}$.

Altogether, $P(k, t) = \Pr\{\log_e U = kc$ at time $t\}$ satisfies the following system of equations

$$P(k, t+1) = pP(k-1, t) + (1-p)P(k+1, t) \quad \text{if } k > k_o \log_e \tilde{u}/c,$$
$$P(\log_e(\tilde{u}/c), t+1) = (1-p)P(\log \tilde{u}/c, t) + (1-p)P(1 + \log_e(\tilde{u}/c), t).$$

Due to the "diffusive" character of these equations, there is a steady-state limit function $P(k, t)$, independent of the initial conditions imposed at a preassigned starting time $\tilde{t}$. That limit is the scaling distribution. (*Proof:* Under the steady-state condition $P(k, t+1) = P(k, t)$, the second equation yields $pP(k_o, t) = (1-p)P(1 + k_o, t)$, then the first equation gives the same identity by induction on $k_o + 1$, $k_o + 2$, and so on.) Therefore, conditions a), b), and c) provide a possible generating model of the exponential distribution for log U.

Champernowne's formally generalized random walk of log U. Champernowne 1953 offers a more general model. It seems sophisticated, but I came to view it as violating a cardinal rule of model-making, namely, that the assumptions must not be heavier and less transparent than the outcome. Details are found in M 1961e{E11}.

2.3 Aside on diffusion *without* a reflecting barrier: as argued by Gibrat, it yields the lognormal distribution, but no steady-state

As mentioned in Chapter E, the scaling and lognormal are continually suggested as alternative representations of the same phenomena, and the lognormal is believed by its proponents (Gibrat 1932) to result from a proportional effect diffusion. This is correct, but only up to a point. It is true that *in the absence of a reflecting barrier*, diffusion of log U does not yield an exponential, but a Gaussian, hence the lognormal for U. Indeed, after T tosses of a coin, $\log_e U(t+T) - \log_e U(t)$ is the sum of T independent variables taking the values c or $-c$ with the respective probabilities p and $1-p$. By the central limit theorem,

$$\frac{\log U(t+T) - \log U(t) - T\,c(2p-1)}{[2\,Tp\,(1-p)]^{1/2}}$$

will tend towards a reduced Gaussian variable. Hence, if at $T=0$ all firms have equal sizes, $\log_e U(t+T)$ will become lognormally distributed as $T \rightarrow \infty$. The same holds for other diffusion models without reflecting barrier.

While the above argument is widely accepted, it has a lethal drawback: the lognormal describes an instantaneous state, not a steady state distribution; for example, in time, its variance increases without bound.

2.4 Misgivings concerning the relevance to economics of the model of scaling based on the diffusion of log $U = V$

In the models for V that lead to a proper steady-state (Sections 2.1 and 2.2), the transformation $U = \exp V$ seems to work a miracle of alchemy: the metamorphosis of a mild variable V into the wild variable U, in the sense described in Chapter E5. But I propose to argue that no metamorphosis took place, because the conclusions reached ex-post destroy the intuition that justified ex-ante the diffusion of V. This contradiction between the ex-post and the ex-ante comes on top of the limitations stressed under B1) in the second page of this chapter.

Reminder of why a diffusion of V is a reasonable idea in physics. Physics is fortunate to have a good and simple reason why diffusion models are good at handling the exchanges of energy between gas molecules. In gases, the energy of even the most energetic molecule is negligible with respect to the total energy of a reservoir. Hence, all kinds of things can happen to one molecule, while hardly anything changes elsewhere. This enormous source of simplicity is a major historical reason why the theory of gases proved relatively easy to develop. For example, the wonderful property of universality (described under B2 at the beginning of this chapter) implies that it makes no difference whether the total "energy" $\sum V$ in a gas is fixed (as it is in a "microcanonical" system) or allowed to fluctuate (as it is in a "canonical" system.) This is why the diffusion model's conclusions do not contradict its premises.

Reasons why diffusion of $\log U$ *seems* not to be *a reasonable idea in economics*. The transformation $V = \log U$ may seem innocuous, but it introduces a major change when $\alpha < 2$. The original U (firm size, income and the like) is additive, but *not* the logarithm $V = \log U$, and U is "wildly random." The concentration characteristic of wild randomness has very concrete consequences.

Firm and city sizes are arguably scaling with $\alpha \sim 1$ (Chapter E13 and Auerbach 1913), and it is typical for a country's largest city to include 15% of the total population. It surely will matter, both intuitively and technically, whether $\sum U$ is fixed or allowed to fluctuate. When the total employment or population are kept fixed, it seems far-fetched to assume that the largest firm or city could grow or wane without influencing a whole industry or country. I cannot even imagine which type of observation could confirm that such is the case. Ex-post, this would be an interesting prediction about economics. Ex-ante, however, this presumed property is surely no more obvious than scaling itself, hence, cannot be safely inserted in an explanatory model of scaling.

Conclusion: a baffling embarrassment of apparent riches. The economic predictions yielded by the diffusion of $\log U$ are baffling. The model does yield a scaling distribution for U, but the model conclusion makes its premises highly questionable. Additionally, the argument ceases to be grounded in thermodynamics, because the latter does not handle situations where canonical and microcanonical models do not coincide. It is true that neither the success nor the failure of a model in physics can guarantee its success or failure in economics.

All told, the models of U based on the diffusion of $\log U$ leaves an embarrassment of riches. The user's response is a matter of psychology.

The *utter pessimist* will say that flawed models are not worth further discussion.

For the *moderate pessimist*, the diffusion of $\log U$ is questionable but worth playing with. I favor this attitude and M 1961e{E11} is an example of its active application.

The *moderate optimist* will not want to look too closely at a gift horse; even in physics, it is common that stochastic models yield results that contradict their assumptions. For example, the search for the "most likely state" applies Stirling's approximation to the factorial $N!$; but this argument eventually yields Ns equal 0 or 1, so that Stirling's approximation is unjustified even though it is successful. This is true, but physics also knows how to read the same results without assumption that contradict the conclusions; economics does not know how to do it. The feature that saves many stochastic models is that errors somehow seem to "cancel out." But in the presence of concentration the counterparts of "the contributing errors" are neither absolutely nor relatively small. For this reason, I view the moderately optimistic position as very hard to adopt.

The widespread *very optimistic* and unquestioning view of diffusion as explaining scaling is altogether indefensible. Models based on the diffusion of $\log U$ do not make me change my policy of viewing scaling as a postulate that brings economies of thought, good fit and a useful basis for practical work. It remains my strongly held belief that scaling must be thoroughly understood, and its consequences explored, without waiting to explain them. Nevertheless, after a short digression, we shall devote Section 3 to an examination of variant models.

2.4. A static characterization of the exponential: it is the "most-likely," and the "expected" steady-state of Brownian motion with a barrier

Among the other models proposed to account for an exponential V in finance and economics, most are also dynamic, if only because "static" is a derogatory word and "dynamic" a compliment.

It is very well-known in thermodynamics that the state which a system attains as the final result of random intermolecular shocks can also be characterized as being "the most likely." The exponential distribution $P(v) = \exp[-\alpha(v-\tilde{v})]$ is indeed obtained by maximizing the expression $-\sum P(v) \log P(v)$ under the two constraints $v \geq \tilde{v}$ and $\sum vP(v) = \text{constant}$. A longer proof shows that the exponential is also "the average state."

As applied to the distribution of income, this approach goes back at least to Cantelli 1921. I see no good reason for retaining it: in physics, the expressions $\sum vP(v)$ and $\sum -P(v)\log P(v)$ have independent roles, the former as an energy and the latter as an entropy (communication theory calls it quantity of information.) But those roles do not give those expressions any status in the present context.

Many variants seem to avoid the transformation $V = \log_e U$. Castellani 1951 assumes without motivation the (equivalent) postulate that u is submitted to a "downward force" proportional to $1/u$.

3. PROPORTIONATE GROWTH AND SCALING WITH EITHER TRANSIENT OR PERMANENTLY DIFFUSIVE RANDOMNESS

Let us start from scratch. The basic idea of *random proportionate effect* is that when t increases by dt, the increment du is proportional to the present value of u, hence the change of $v = \log u$ is independent of v. However, those words are far from completely specifying the variation of V in time. Many distinct behaviors are obtained by varying the boundary conditions and the underlying differential equation itself. To end on a warning: due to time pressure, the algebra in this section has not been checked through and misprints may have evaded attention.

3.1 First and second-order random proportionate effects

Most important is the fact that the underlying equation can be of either first or second order.

Flow models ruled by a first-order equation. Solution of first-order differential equations largely preserve their initial conditions and essentially correspond to a nonrandom proportionate growth of U, in which the value of u only depends upon "age." This assumption must be recognized and may or may not be realistic. The model found in Yule 1924, corresponds to a first-order finite-difference equation. It predicts a random transient in youth, followed by effectively nonrandom, mature growth. This feature makes it inapplicable to economics, even though it is defensible in the original context of the theory of evolution, and its restatement in terms of bacterial mutation (see Appendix II) helped provide the first full solution of a classic problem in Luria & Delbrück 1943. The analysis in Section 3.2 applies with insignificant changes to a slight variant of Yule's process advanced in Simon 1954. After a short transient, the growth postulated in that paper becomes non-random.

Diffusion models ruled by a second-order equation. An actual "diffusion" embodies a permanently random proportionate effect. The initial conditions may become thoroughly mixed up, but lateral conditions (such as the nature of a side-barrier, the kind of immigration, etc.) have a permanent influence, even on the qualitative structure of the solution.

In the model in Aström 1962, the boundary is "natural," in the terminology of Feller, and no lateral conditions need or can be imposed. However, the equation possesses certain special features that amount to additional hypotheses in the model that the equations embody.

In any event, many assumptions are necessary if the diffusion model is to lead to the desired scaling solution and/or to a steady-state solution. One must not artificially stress some assumptions, while not acknowledging others of special importance in the existence of the reflecting barrier that entered in Section 2.1, but not Section 2.3.

3.2 The first-order finite difference model in Yule 1924

Except for an initial genuinely stochastic transient, the model in Yule 1924 is undistinguishable from the model in Fermi 1949 changed by the addition of exponential immigration, and restriction to a vanishing decay constant $\alpha = 0$. The calculations are much more involved than Fermi's, but it is useful to analyze Yule's model fully and concisely because there are occasions where it is applicable.

The fundamental assumption is that firm size is quantized and that during each time increment dt, the size of a firm has a known probability dt of increasing by unity. It is well-known that a firm growth from size 1 to size u during the elapsed time T has the geometric distribution for the probability that a firm grew from size 1 to size u during the elapsed time T. To this assumption, Yule adds the further hypothesis that new firms immigrate into the system with a probability dt during the time increment from t to $t + dt$. Then, at the calendar time t, the probability that one of these firms is of size u is given by

$$e^{-T}(1 - e^{-T})^{u-1}.$$

To this, Yule adds another hypothesis: new firms immigrate into the system with a probability $\varphi(t)dt$ during the time-increment from t to $t + dt$. Then, at the calendar time t, the probability that one of these firms be of size u is given by

$$f(u, t) = \int_{-\infty}^{t} e^{-(t-\tilde{t})}[1 - e^{-(t-\tilde{t})}]^{u-1}\varphi(\tilde{t})d\tilde{t}.$$

As to the number of firms at time t, its expected value is given by

$$n(t) = \int_{-\infty}^{t} \varphi(\tilde{t})d\tilde{t},$$

and the expected value of the total size of all firms is

$$k(t) = \int_{-\infty}^{t} e^{t-\tilde{t}}\varphi(\tilde{t})d\tilde{t}.$$

In Yule's case,

$$\varphi(\tilde{t}) = \alpha e^{\alpha\tilde{t}} \text{ for } \tilde{t} > 0.$$

Writing $\exp(\tilde{t} - t) = y$, one obtains

$$f(u, t) = \alpha\int_{0}^{t} e^{-(t-\tilde{t})}[1 - e^{-(t-\tilde{t})}]^{u-1}e^{\alpha\tilde{t}}d\tilde{t} = \alpha e^{\alpha t}\int_{0}^{e^{-t}} y^{\alpha}(1-y)^{u-1}dy.$$

This is Euler's incomplete Beta integral. For e^{-t} close to 1,

$$f(u, t) \sim \alpha e^{\alpha t}\,\frac{\Gamma(\alpha+1)\Gamma(u)}{\Gamma(\alpha+u+1)}\,.$$

For large u,

$$f(u, t) \sim \alpha e^{\alpha t}\Gamma(\alpha+1)u^{-(\alpha+1)}.$$

Yule's model predicts the birth rate

$$k(t) = \alpha\int_{0}^{t} d^{t-\tilde{t}}e^{\alpha\tilde{t}}d\tilde{t} = \frac{\alpha}{\alpha-1}\,(e^{\alpha t} - e^{t}).$$

If $\alpha < 1$, $k(t) \sim \frac{\alpha}{1-\alpha} e^t$, and $f(u, k) \sim (1-\alpha)^{\alpha} \alpha^{1-\alpha} \Gamma(\alpha+1) k^{\alpha} u^{-(\alpha+1)}$.

If $\alpha > 1$, $k(t) \sim \frac{\alpha}{\alpha-1} e^{\alpha t}$, and $f(u, k) \sim \frac{\Gamma(\alpha+1)}{\alpha-1} k u^{-(\alpha+1)}$.

The relative rate of addition of new firms equals $(1-\alpha)^{-1}$. In the case of firm sizes, the mechanism of proportional growth cannot be presumed to start until the size has exceeded some sizable quantity $\tilde{u}$. Therefore, the Yule model is *not applicable*, since the correspondingly modified model will be based upon the probability that a population grows from size $\tilde{u}$ to size u in the time T, which is well-known to be given by the negative binomial distribution (Feller 1950). Therefore,

$$f(u, t) = \int_0^t \frac{\Gamma(u)\varphi(\tilde{t})d\tilde{t}}{\Gamma(u-\tilde{u}+1)\Gamma(\tilde{u})} e^{-\tilde{u}(t-\tilde{t})}[1-e^{-(t-\tilde{t})}]^{u-\tilde{u}}.$$

Using Yule's rate of addition of new population,

$$f(u, t) = \alpha e^{\alpha t} \frac{\Gamma(u)}{\Gamma(u-\tilde{u}+1)\Gamma(\tilde{u})} \int_0^t e^{-\tilde{u}T}[l-e^{-(t-\tilde{t})}]^{u-\tilde{u}}$$

$$= \alpha e^{\alpha t} \frac{\Gamma(u)}{\Gamma(u-\tilde{u}+1)\Gamma(\tilde{u})} \int_0^{e^{-t}} y^{\tilde{u}-1+\alpha}(1-y)^{u-\tilde{u}} dy \,.$$

If t is large,

$$f(u, t) = \alpha e^{\alpha t} \frac{\Gamma(u)}{\Gamma(u-\tilde{u}+1)\Gamma(\tilde{u})} \frac{\Gamma(\tilde{u}+\alpha)\Gamma(u-\tilde{u}+1)}{\Gamma(\alpha+u+1)}$$

$$= \alpha e^{\alpha t} \frac{\Gamma(\tilde{u}+\alpha)}{\Gamma(\tilde{u})} \frac{\Gamma(u)}{\Gamma(\alpha-u+1)} \,.$$

As expected, $f(u,t)$ is the Yule distribution truncated to $u \geq \tilde{u}$.

Now examine non-exponential rates $\varphi(\tilde{t})$. If $\tilde{u}$ is large (e.g, if it exceeds 100), the kernel function $e^{-ux}(1-e^{-x})^{u-\tilde{u}}$ becomes *extremely* peaked near its maximum for $e^x = u/\tilde{u}$. To approximate this kernel, expand $\log\,[e^{-\tilde{u}T}(1-e^{-T})^{u-\tilde{u}}]$ in Taylor series up to terms of second-order. The exponential of the result is

$$e^{-\tilde{u}T}(1-2^{-T})^{u-\tilde{u}} \sim \left\{\left(\frac{\tilde{u}}{u}\right)^{(\tilde{u}/u)}\left(1-\frac{\tilde{u}}{u}\right)^{(1-\tilde{u}/u)}\right\} u \ \exp\left\{-\frac{[T-\log(u/\tilde{u})]^2}{2(1/\tilde{u}-1/u)}\right\}.$$

From this representation, it follows that the full integral that gives $f(u,t)$ can be replaced by the integral carried over a "four-standard-deviation" interval around $\log(u/\tilde{u})$, namely

$$\log\left(\frac{u}{\tilde{u}}\right)-2\sqrt{\frac{1}{\tilde{u}}-\frac{1}{u}} \le t \le \log\left(\frac{u}{\tilde{u}}\right)+2\sqrt{\frac{1}{\tilde{u}}-\frac{1}{u}}\ .$$

That is, for all practical purposes, $f(u,t)$, will not depend upon the firms that started at any other times.

As a consequence, suppose that the size u has exceeded a sizeable threshold, such as $\tilde{u}=10{,}000$. Yule's model predicts that, from then on, $u(t)/10{,}000$ will differ from $\exp(t-\tilde{t})$ by, at most, a quantity of the order of $10{,}000^{-1/2}=1\%$. That is, the growth of sizes *will for all practical purposes proceed with little additional randomness.*

To summarize, except for insignificant "noise," Yule's model makes the prediction that the size of the larger firms increases exponentially. Among the firms larger than "Smith and Co.," almost all will have reached the threshold size 10,000 before Smith and Co., and all the ranking by size of most larger firms is identical to their ranking by date of achievement of the threshold size 10,000.

It is easy to ascertain that the same conclusion will be reached if the definition of Yule's process is modified, as long as the modification does not affect the fundamental differential equations of that process. In all cases, U exhibits a stochastic transient during take-off, lasting as long as a firm is small. However, if and when its size becomes considerable, all sources of new randomness will have disappeared. This makes Yule's model indistinguishable in practice from Fermi's with zero death-rate, and there is little gain from Yule's far heavier mathematics. In Simon 1955, the independent variable is the actual total population k at time t, but – after a short transient – growth is non-random.

3.3 A "Fermi-like" model involving permanent diffusion

To eliminate the principal drawback of the model in Fermi 1949, as applied to economics, exponential lifetimes can be combined with growth that allows *random* increments of $\log_e U$. When this argument was presented to my IBM colleague M. S. Watanabe, he noted it could be applied to the cosmic-ray problem, but my model turned out to be simply a variant of Fermi's approach, as defined by the following assumptions.

Let dt and $\sigma\sqrt{dt}$ be the mean value and the standard deviation of the change of V per time increment dt. Let t be the time elapsed since the moment when a firm exceeds the size $\tilde{u}$, i.e., when V first exceeds the value 0. The distribution of V after the time t will be Gaussian, i. e., the probability that $v \le V \le v + dv$ will be

$$\frac{1}{\sigma\sqrt{2\pi t}} \exp[-\frac{(v-t)^2}{2t\sigma^2}]dv.$$

Now, the probability that $v < V \le v + dv$, regardless of "age," will be

$$\frac{\lambda}{\sigma\sqrt{2}\ \pi} \left\{ \int_0^\infty t^{-1/2} \exp\left[-bt - \frac{(v-t)^2}{2t\sigma^2}\right] dt \right\} dv.$$

A result in p. 146 of Bateman 1954, reduces this probability to

$$\frac{\lambda}{\sqrt{2\sigma}} (b + \frac{1}{2\sigma^2})^{-1/2} \exp\left\{ \frac{v}{\sigma^2} - \frac{v\sqrt{2}}{\sigma} \sqrt{b + \frac{1}{2\sigma^2}} \right\},$$

which is exponential distribution with

$$\alpha = -\frac{1}{\sigma^2} + \frac{\sqrt{2}}{\sigma} \sqrt{b + \frac{1}{2\sigma^2}}.$$

To check that $\alpha > 0$, multiply the positive expression by

$$\tilde{\alpha} = \frac{1}{\sigma^2} + \frac{\sqrt{2}}{2^\sigma} \sqrt{b + \frac{1}{2\sigma^2}},$$

The product is $\alpha\tilde{\alpha} = 2b/\sigma^2 > 0$, therefore $\alpha > 0$, as it should.

The value $\alpha = 1$ corresponds to

$$\sqrt{2b+1}\ /\sigma^2 = \sigma(1 + \frac{1}{\sigma^2}) = \sigma + \frac{1}{\sigma} \text{ or } \sigma^2 = 2(b-1).$$

Again $\alpha = 1$ corresponds to the martingale relation for U. Indeed,

$$\exp\,[dt + (1/2)\sigma^2 dt] \sim 1 + [1 + \sigma^2/2]dt,$$

when Z is the expected value of a variable whose logarithm is Gaussian with mean dt and standard deviation $\sigma\sqrt{dt}$. Hence, whichever b,

$$E[U(t+1) \mid u(t)] \sim u(t)(1 - bd)[\,+ dt(1 + \sigma^2/2)]$$
$$\sim u(t)[1 + (\sigma^{-2}/2 + 1 - b)dt],$$

and the martingale relation indeed requires $\sigma^2 = 2(b-1)$.

3.4 Aström's diffusion model

While studying a problem of control engineering, Aström 1962 introduced the continuous time variant of the difference equation

$$U(t+1) = (1 - m' + G)U(t) + m''.$$

Here $m', m'' > 0$ and $\sigma > 0$ are constants, and G is a Gaussian variable of mean zero and unit standard deviation. For large u, this equation is another variant of random proportionate effect. However, it presents the originality that the reflecting barrier is replaced by the correction expressed by the positive "drift" m'', which is independent of u and negligible for large u. This gives interest to the equation. Since we expect the density to be asymptotically scaling, it is reasonable to try as a solution a density of the form

$$f(u) = \varphi(u)u^{-(\alpha+1)},$$

where $\varphi(u)$ rapidly tends to a limit as u increases.

To avoid advanced mathematics, let us first determine α by requiring $u(t+1)$ and $(1 - m'G\sigma)u(t)$ to have the same distribution for large u. This can be written as

$$u^{-(\alpha+1)} = \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty w^{-1} \exp\left\{ -\frac{[u - (1-m')w]^2}{2\sigma^2 w^2} \right\} w^{-(\alpha+1)} dw;$$

defining y as u/w, this requirement becomes

$$u^{-(\alpha+1)} = u^{-(\alpha+1)} \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty y^\alpha \exp\left\{ - \frac{[v-(1-m')]^2}{2\sigma^2} \right\} dy.$$

In the limit where the basic equation changes from difference to differential, m' and σ^2 are small. Letting $\psi(y) = \log_2 v - [v-(1-m')]^2(2\sigma^2)^{-1}$, we see that $\exp[\psi(y)]$ is non-negligible only in the neighborhood of its maximum, which occurs approximately for $y = \tilde{y} = 1 - m' + \alpha\sigma^2$. Near $\tilde{y}$, $\psi(y) \sim \psi(\tilde{y}) + (y-\tilde{y})\psi'(\tilde{v}) + (1/2)(v-\tilde{v})\psi'(\tilde{v})$, hence the steady-state condition:

$$1 = \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty \exp[\psi(v)]dy = \exp[\psi(\tilde{v})]\sigma^{-1}[\psi''(\tilde{v})]^{-1/2}.$$

Easy algebra yields

$$\alpha = 1 + \frac{2m'}{\sigma^2}.$$

We again encounter the increasingly familiar relation between the sign of $\alpha - 1$ and that of the regression of $u(t+1)$ on $u(t)$. In particular, $\alpha = 1$ corresponds to the "martingale" case $m' = 0$.

In the present case, however, $\varphi(u)$ is determined by the equation itself, therefore, the model does not need special boundary conditions at $u = 0$. Suffices to say that the Fokker-Planck equation yields

$$\varphi(u) = \exp(-2m''/\sigma^2) = \exp(-\beta/u),$$

so that, normalizing the density $f(u)$ to add to one, we obtain

$$f(u) = \frac{\beta^\alpha}{\Gamma(\alpha)} u^{-(\alpha+1)} \exp(-\beta/u).$$

a) U_A is the inverse of a Gamma-variable of exponent α. Hence, $\Sigma 1/u_n$ is a "sufficient statistic" for the estimation of α from a set of sample values u_n. Unfortunately, this statistic is unusable, being overly dependent upon values that are small and fall in the range where data do not exactly follow the inverse of a gamma distribution. This exemplifies the difficulties which are encountered in the statistical theory of the estimation of the scaling exponent.

b) For $\alpha = 1$, U_A reduces to the Fréchet distribution that rules the largest of N identical scaling variables of exponent 1.

c) For the special value $\alpha = 1/2$, U_A happens to be identical to the L-stable distribution of exponent $\alpha = 1/2$ and of maximum skewness, in other words identical to the limit of the distribution of $(1/N^2)$-th of the sum of N identical scaling variables of exponent 1/2.

&&&&& PRE-PUBLICATION APPENDICES &&&&&

APPENDIX I (M 1964o): RANDOM WALKS, FIRE DAMAGE, & RISK

Actuarial science accumulated a substantial store of knowledge. This paper discusses the risk due to fire. Very similar mechanisms apply in many other problems, hence a more general goal is to illustrate why it is said that the scaling distribution constitutes a "source of anxiety for the risk theory of insurance."

I.1 Introduction

In somewhat simplified form, the following statement summarizes an empirical law established by Benckert & Sternberg 1957.

"The damage a fire causes to a house follows the scaling distribution."

That is, suppose that the damage is greater than a minimum threshold m = \$20 and smaller than M, defined as the maximum destroyable amount of the building. Then

$$\text{for } m < x < M, \qquad \Pr\{\text{damage } \geq x\} = x^{-\alpha} m^{\alpha};$$
$$\text{but} \qquad \Pr\{\text{damage } = M\} = M^{-\alpha} m^{\alpha}.$$

This law applies to all classes of Swedish houses outside of Stockholm. The reported values of α range between 0.45 and 0.56; we shall take $\alpha = 1/2$ and investigate the consequences.

I.2 A model of fire damage amount

The scaling distribution of exponent 0.5 plays a central role in probability theory: it is the distribution of the returns to equilibrium in the game of tossing a fair coin. This theory is developed in most textbooks of probability (such as Feller 1950), and can be translated into insurance terms.

Our first assumption is that the intensity of a fire is characterized by a single number, designated by U, which can take only integer values: there is no fire when $U=0$; a fire starts when U becomes equal to 1, and it ends when either U becomes equal to 0 again, or when all that can possibly burn has already burned out.

We also assume that, at any instant of time, there is a probability $p=1/2$ that the fire encounters new material so that its intensity increases by 1, and there is a probability $q=1/2$ that the absence of new materials or the action of fire-fighters decreases the intensity by 1. In the preceding statement, 'time' is to be measured by the extent of damage. If there is no finite maximum extent of damage, and no lower bound to recorded damages, the duration of a fire will be an even number given by a classical result concerning the return to equilibrium in coin-tossing:

$$\Pr\{\text{duration of a fire} = x\} = 2\binom{1/2}{x/2}(-1)^{x/2-1}.$$

Except for the first few values of x, this expression is proportional to $x^{-3/2}$. Damages smaller than the minimum threshold m are not properly recorded. Hence, the duration of a fire (i.e., the extent of damage) will be given by the scaling distribution with exponent $1/2$:

$$\Pr\{\text{duration of a fire } > x > m\} = (x/m)^{-1/2}.$$

Finally, take account of the fact that the fire *must* end if and when the whole house has burnt out. We see that the prediction of the above argument coincides precisely with the Benckert-Sternberg finding.

I.3 Relations involving the size of the property and the expected amount of the damage due to fire

It is easy to compute the expected value of the random variable considered in Section 1. One finds

$$\text{Expected fire damage } = \int_m^M x(1/2)x^{-(1/2+1)}m^{1/2}dx + M(M/m)^{-1/2}$$
$$= 2\sqrt{Mm} - m.$$

This value tends to infinity as $M \rightarrow \infty$.

On the other hand, according to von Savitsch & Benktander 1953, the expected number of fires per house in the course of a year is a linear function of M. If this is indeed so, it would imply that, for large values of M, the rate of insurance should be proportional to $\sqrt{M}$.

When the distribution of property sizes M is known, a simple argument yields

$$\Pr\{\text{damage per fire} > x\} = \Pr\ \{M > m\}(x/m)^{-1/2}.$$

Moreover, if von Savitsch & Benktander 1953 is correct, one has

$$d\ \Pr\{\text{fire damage per year } > x\} = Cxd\ \Pr\ \{\text{ damage } > x\}$$
$$= Cxd\ \left[\Pr\ \{M > m\}(x/m)^{-1/2}\right].$$

Let the distribution of M be itself scaling with the exponent α^*. This is the case in all kinds of liability amounts. The distribution of damage in a single fire will then be scaling with exponent $\alpha^* + 1/2$; if $\alpha^* > 1/2$, the distribution of damage per year will be scaling with exponent $\alpha^* - 1/2$. This demonstrates that the mathematical manipulations based on the scaling distribution are especially convenient.

I.4 Generalization

The random walk with $p = q = 1/2$ represents a kind of equilibrium state between the fire and the fire-fighters. If the quantity of combustible property were unbounded, such a random walk will surely die out, although its expected duration would be infinite.

To the contrary, if $p > q$, the fire-fighting efforts would be inadequate, and there would be a nonvanishing probability that the fire continue forever.

If $p < q$, the probability of the fire running forever would again be zero, and the expected duration of the fire would be finite. The law giving the duration of the fire would then take the form:

$$d\,\Pr\{\text{duration of a fire } \geq x > m\} \sim c^* \exp(-cx)x^{-3/2}dx,$$

where c^* and c are two constants depending upon m, p, and q. If x is actually bounded and $q-p$ (and hence c) is small, the above formula will be indistinguishable in practice from a scaling distribution with an exponent slightly greater than 1/2; this is perhaps an explanation of the more precise experimental results of Benckert and Sternberg.

Generalizations of the random walk provided by many classical models of diffusion may be rephrased in terms of fire risks. A random walk model in which the intensity of the fire goes up, down, or remains unchanged would only change the time scale.

I.5 Another application

The results of Benckert & Sternberg 1957 strongly recall those of Richardson 1960 and Weiss 1963, and the model that has been sketched above is easy to translate in terms of Richardson's problem. It would be fascinating to ponder how relevant that translation may be.

Annotation to Appendix I: The importance of the scaling distribution in the theory of risk. The topic of Appendix I may seem narrowly specialized, but many risks against which one seeks insurance or reinsurance do follow scaling distributions. The sole reason for focussing on fire damage was the abundance and quality of the data on Swedish wooden houses.

Until recently, however, no one faced such risks squarely, quite to the contrary. Once, an insurance representative visited to report that his industry was bound one day to face risks following the scaling distributions, but he never called again. More significantly, a (short-lived) manager I once had at IBM pointedly described the present paper as childish and insignificant in comparison with the bulk of my work. Events suggest, to the contrary, that this paper was a quiet long-term investment. Important occurrences of very long-tailed risk distribution that are being investigated include Zajdenweber 1995ab.

APPENDIX II (M 1974d): THE NUMBER OF MUTANTS IN AN OLD CULTURE OF BACTERIA

Luria & Delbrück 1943 observed that, in old cultures of bacteria that have mutated at random, the distribution of the number of mutants is extremely long-tailed. Here, this distribution is derived (for the first time) exactly and explicitly. The rates of mutation will be allowed to be either positive or infinitesimal, and the rate of growth for mutants will be allowed to be either equal, greater or smaller than for nonmutants. Under the realistic limit condition of a very low mutation rate, the number of mutants is shown to be a L-stable random variable, of maximum skewness, β, whose exponent α is essentially the ratio of the growth rates of nonmutants and of mutants. Thus, the probability of the number of mutants exceeding the very large value m is proportional to $m^{-\alpha-1}$; it is "asymptotically scaling." The unequal growth rate cases $\alpha \neq 1$ are solved for the first time. In the $\alpha = 1$ case, a result of Lea & Coulson is rederived, interpreted, and generalized. Various paradoxes involving divergent moments that were encountered in earlier approaches are either absent or fully explainable. The mathematical techniques used are standard and will not he described in detail; this paper is primarily a collection of results.

II.1 Introduction

Let the bacteria in a culture grow, and sometimes mutate, at random, for a long time. In an occasional culture, the number of mutants will be enormous, which means that "typical values," such as the moments or the most probable value, give a very incomplete description of the overall distribution. Also, when the same mutation experiment is replicated many times, the number of mutants in the most active replica may exceed by orders of magnitude the sum of the numbers of mutants in all other replicas taken together. Luria & Delbrück 1943 first observed the above fact, and also outlined an explanation that played a critical role in the birth of molecular biology: The advantage of primogeniture is so great that the clone to which it gives rise has time to grow to a very much larger size than other clones in the same replica, or than the largest clone grown in a more typical replica in which no early mutation happened to be included.

Interest in expressing this explanation quantitatively, by describing the full distribution of the numbers of mutants, first peaked with Lea & Coulson 1949, Kendall 1952, Armitage 1952, 1953 and Bartlett 1966, but the solutions advanced were not definitive. Several investigators only calculated moments. Also, rates of growth were always assumed to be the same for mutants and non-mutants, and the rate of mutation to be very small. Kendall 1950 was very general; it may include in principle the results to be described, but not explicitly.

This short paper describes the whole distribution (for the first time) under assumptions that seem both sufficiently general to be realistic and sufficiently special for the solution to be exact and near explicit – in the sense that the Laplace transform of the distribution is given in closed analytic form.

The extreme statistical variability characteristic of the Luria & Delbrück experiment is also found in other biological experiments in progress. One may therefore hope that the present careful study, which settles the earliest and simplest such problem, would provide guidance in the future. It may help deal with some new cases when very erratic behavior is unavoidable, and in other instances, it may help avoid very erratic random behavior and thus achieve better estimates of such quantities as rates of mutation.

II.2 Background material: assumptions and some known distributions

Assumptions. (A) At time $t = 0$ the culture includes no mutant, but includes a large number b_0 of nonmutants.

(B) Between times t and $t + dt$, a bacterium has the probability mdt of mutating.

(C) Back mutation is possible.

(D) Neither the mutants nor the nonmutants die.

(E) The rate of mutation m is so small that one can view each mutation as statistically independent of all others.

(F) Mutants and nonmutants multiply at rates that may be different. The scale of time is so selected that, between the instants t and $t + dt$, the probability of division is gdt for a mutant and dt for a nonmutant.

Non-mutants. A bacterium that mutates may be considered by its non-mutant brethren as having died. Therefore, $N(t, m)$, defined as the number of non-mutant bacteria at the instant t, follows the well-known "simple birth and death process" (see, e.g., Feller 1950, Vol. I, 3rd ed., p. 454). When $b_0 \gg 1$, the variation of $N(t, m)$ is to a good approximation deterministic

$$N(t, m) \sim EN(t, m) \sim b_0 e^{t(1-m)}.$$

Non-random clones. A clone being all the progeny of one mutation, denote by $K(t, m)$ the number of clones at the instant t. From Assumption (E), $K(t, m)$ is so small relative to $N(t, m)$ that different mutations can be considered statistically independent. It follows that $K(t, m)$ is a Poisson random variable of expectation

$$m\int_0^t fN(s, m)ds = b_0 m(l-m)^{-1}[e^{t(1-m)} - 1].$$

Random clones. Denote by $Y(t, m, g)$ the number of mutants in a clone selected at random (each possibility having the same probability) among the clones that have developed from mutations that occurred between the instants 0 and t. The distribution of $Y(t, m, g)$ will be seen to depend on its parameters through the combinations e^{gt} and $\alpha = (1-m)/g$. Since eventually we shall let $m \to 0$, α nearly reduces to the ratio of growth rates, $1/g$. One can prove that, after a finite t,

$$\Pr\{Y(t,m,g) \geq y\} = \alpha[1 - e^{-t(1-m)}]^{-1}\int_1^{\exp(gt)} fv^{\alpha - y}(v-1)^{y-1}dv.$$

In the case $\alpha = 1$, this yields explicitly

$$\Pr\{Y(t, m, g) \geq y\} = y^{-1}[1 - e^{-gt}]^{y-1}.$$

The generating function (g.f.) of Y, denoted $\tilde{Y}$, equals

$$\tilde{Y}(b, t, m, g) = \alpha[1 - e^{-t(1-m)}]\int_1^{\exp(gt)} fv^{-\alpha-1}\{[v(e^b - 1) + 1]^{-1}dv\}.$$

As $t \to \infty$, while m and g are kept constant, Y tends to a limit random variable $Y(\alpha)$ that depends only on α. When $\alpha = 1$,

$$\Pr\{Y(\alpha) = y\} = \int_0^1 fv(1-v)^{y-1}dv = \frac{1}{y(y+1)},$$

a result known to Lea & Coulson 1949. For all α,

$$\Pr\{Y(\alpha) \geq y\} = \alpha\frac{\Gamma(\alpha)\Gamma(y)}{\Gamma(\alpha + y)}.$$

For large y,

$$\Pr\{Y(\alpha) \geq y\} \sim \Gamma(\alpha+1)y^{-l-\alpha}.$$

The $Y(\alpha)$ thus constitutes a form of asymptotically "hyperbolic" or "Pareto" random variable of exponent α. The population moment $EY^h(\alpha)$ is finite if $h < \alpha$ but infinite if $h \geq \alpha$. For example, the expectation of $Y(\alpha)$ is finite if, and only if, $\alpha > 1$ and the variance is finite if, and only if, $\alpha > 2$. Infinite moments are a vital part of the present problem.

II.3 The total number of mutants

$M(t, m, g)$ will denote the number of mutant bacteria at the instant t. Thus, $M(0, m, g) = 0$, and

$$M(t, m, g) = \sum_{k=1}^{K(t,m,g)} Y_k(t, m, g).$$

Denote its g.f. by $\tilde{M}(b, t, m, g)$. Since K is a Poisson random variable of expectation EK, $\log \tilde{M}(b, t, m, g) = EK[\tilde{Y}(b, t, m, g) - 1]$.

The distributions of K and Y both depend on t (and are therefore interrelated). For this reason, the standard theorems concerning the limit behavior of sums (Gnedenko & Kolmogorov 1954, Feller 1950, Vol. II) are not applicable here. Fortunately, the special analysis that is required is straightforward. An approximate formal application of the standard theorems – letting the Y converge to the $Y(\alpha)$ and then adding K of them – would be unjustified, but some of its results nevertheless remain applicable. (Some of the paradoxes encountered in the analyses circa 1950 are related to cases where inversion of limit procedures is unjustified.)

One correct formal inference concerns the choices of a scale factor $S(K)$ and location factor $L(K)$, so as to ensure that the probability distribution of $R = S(K)[M - L(K)]$ tends to a nondegenerate limit as $K \to \infty$. Setting

$$\delta = [b_0 m(l-m)^{-1}]^{-1/\alpha},$$

the scale factors are as follows:

$$\begin{aligned}
\alpha > 2 &: L(K) = EY &;\quad S(K) = (EK)^{-1/2} \\
1 < \alpha < 2 &: L(K) = EY &;\quad S(K) = (EK)^{-1/\alpha} \\
\alpha = 1 &: L(K) = \log EK &;\quad S(K) = (EK)^{-1/\alpha} = \delta e^{-gt} \\
\alpha < 1 &: L(K) = 0 &;\quad S(K) = (EK)^{-1/\alpha}
\end{aligned}$$

With these scale factors, the limits depend on α, as follows.

The case $\alpha > 2$**.** Here, $\lim_{t \to 0} (EK)^{-1/2}(M - EM)$ can be shown to be Gaussian. Nothing original!

The case $\alpha < 1$**.** Here, $\lim_{t \to 0} (EK)^{-1/\alpha}\sum_{k=1}^{K} Y_k$ can be shown to have a g.f. equal to

$$\tilde{R}(b, \infty, m, g) = \exp\left\{\alpha \int_0^\delta fbw^{-\alpha}(bw + 1)^{-1} dw\right\}.$$

The corresponding limit r.v., call it $R(\infty, \alpha, \Delta)$, appears for the first time (to the best of my knowledge) in the present context. The fact that it is nondegenerate (not reduced to either 0 or ∞) confirms that the above standardization was well chosen. Moreover, near $b = 0$, $\tilde{R}(b, \infty, m, g)$ has a good expansion in Taylor series, so all moments of $R(\infty, \alpha, \delta)$ converge. However, this convergence has limited significance because, in actual practice, m is extremely small and Δ is extremely large, so the moments of $R(\infty, \alpha, \delta)$ are themselves enormous and tell us very little about the distribution of $R(\infty, \alpha, \delta)$. On the other hand, as was shown by Luria & Delbrück, the birth and mutation process is illuminated by a sort of "diagonal" procedure. In this procedure, while t increases, m and/or b_0 change in such a way that $\Delta \to \infty$ while $g > 1$, while α remains between 0 and 1. As a result, the function $\tilde{R}$ tends towards

$$\exp\left\{-\alpha b^\alpha \int_0^\infty fz^{-\alpha}(1 + z)^{-1} dz\right\} = \exp\left[-b^\alpha \alpha\pi / \sin(\alpha\pi)\right].$$

This is an unfamiliar expression for a well-known function, namely the g.f. a positive Lévy stable random variable of maximal skewness, $\beta = 1$ (Gnedenko & Kolmogorov 1954, Feller 1950, Volume II). It is also the limit one would have obtained for $K \to \infty$ if we let $Y \to Y(\alpha)$ and then consider the similarly standardized sum of K independent random variables of the form $Y(\alpha)$. In the limit, all the moments of order $h > \alpha$ (including all integer moments) diverge. As a practical consequence, the statistical estimation of m and g from values of M is both complicated and unreliable. Traditionally, statistics has relied heavily on sample averages, but when the population averages are infinite, the behavior of the sample averages is extremely erratic, and any method that involves them must be avoided.

***The case* $1 < \alpha < 2$.** Here, $\lim_{t \to \infty} (EK)^{-1/\alpha} \sum_{k=1}^{K} [Y_k - EY_k]$ can be shown to have the g.f.

$$\exp \left\{ \alpha \int_0^{\delta} f b^2 w^{-\alpha+1} (bw+1)^{-1} dw \right\}.$$

As $\Delta \to \infty$, this function tends towards

$$\exp \left[-b^{\alpha} \alpha \pi / \sin (\alpha \pi) \right],$$

which is again the g.f. of a stable random variable of exponent α and maximal skewness, i.e., of the limit of a similarly standardized sum of K independent random variables of the form $Y(\alpha)$. The theory of these limits is well known, but their shape is not; see M 1960i{E10}, M & Zarnfaller 1959.

***The case* $\alpha = 1$.** Here, $\lim_{t \to \infty} (EK)^{-1} \sum_{k=1}^{K} [Y_k - \log EK]$ can be shown to have the g.f.

$$\exp \left[b \log b + b \log (1 + 1/b\Delta) \right].$$

As $\Delta \to \infty$, this function tends towards $\exp [b \log b]$, corresponding to the stable density of exponent $\alpha = 1$ and maximal skewness, $\beta = 1$, sometimes called the "asymmetric Cauchy" density. It was derived (but not identified) in Lea & Coulson 1949, which concerns the case when the mutation rate m is small, and the growth rates for the mutants and the nonmutants are equal, so that $\alpha \sim 1$.

II.4 The total number of bacteria and the degree of concentration

Designate by $B(t, m, g) = N(t, m) + M(t, m, g)$ the number of bacteria of either kind at the instant t. In the straightforward special case $g = 1$, the function $B(t, m, g)$ follows a "simple birth process" or "Yule process"; see Feller

1968. When $b_0 \gg 1$, the growth of B is for all practical purposes deterministic and exponential, meaning that $B(t) \sim b_0 e^t$.

In the cases $g \neq 1$, things are much more complex, but much of the story is told by the orders of magnitude for large t: $M(t, m, g) \sim e^{gt}$ and $N(t, m, g) \sim e^{(l-m)t}$.

When $\alpha < 1$, $B(t, m, g) \sim M(t, m, g)$, meaning that the mutants – which we know are subject to very large fluctuations – become predominant.

When $\alpha > 1$, $B(t, m, g) \sim e^{(1-m)t}$ with little relative fluctuation, since the random factor that multiplies t is nearly the same as it would be if there had been no mutation. Thus, the dependence of B upon g is asymptotically eliminated.

Now examine the "degree of concentration" of the mutants, namely the ratio ρ of the number of mutants in the largest of the K clones in a replication, divided by the total number of mutants in the other clones of this replication.

Luria & Delbrück discovered that an alternative ratio can be quite large. This ratio is the number of mutants in the largest among H replications, divided by the sum of the number of mutants in the other of the replications. It can be shown that the above two ratios follow the same distribution, so it will suffice to study the first, beginning with two extreme cases.

Consider the case where a mutation brings enough competitive disadvantage and enough decrease in the growth rate to result in $\alpha \gg 2$. Then, the number of young and small clones increases much faster than the size of the single oldest clone in an experiment. Therefore, it is conceivable that a negligible proportion of mutants will be descended from either this oldest clone or any other single clone. This expectation is indeed confirmed. We know that if $\alpha > 2$ the quantity $M(t, m, g)$ tends towards a Gaussian limit, so the contribution of any individual addend Y_k to their sum is indeed negligible.

Now consider the opposite extreme case, where a mutation brings enough competitive advantage and enough increase in the growth rate to result in $\alpha \ll 1$. Then, the size of the oldest clone in an experiment (corresponding to the earliest mutation) grows much faster than the number of fresh clones. It is conceivable, therefore, that the largest clone in an experiment is comparable in size to the sum of all the other clones. An appreciable proportion of the mutants could descend from the single largest clone. This expectation is indeed confirmed in two different ways. First, it has been shown by Darling 1952 (see also Feller 1950, Volume II, p. 439, problem 20), that if $\alpha < 1$, the ratio ρ does not tend to zero as $K \to \infty$. Rather, its distribution tends to a nondegenerate limit, and $E(1/\rho)$ has the nondegenerate limit $\alpha/(1-\alpha)$. As α varies from 0 to 1, this limit varies from 0 to ∞. That is, when mutation causes an enormous increase in growth rate so that the value of α is very small, $1/\rho$ is nearly 0 on the average, and the limit value of ρ for large K is often very large. When, on the contrary, mutation brings very slight advantage, so that α is very nearly 1, $1/\rho$ is very large on the average and ρ tends to be small. But its values can be seen to be widely scattered, and large values are not unlikely.

The limits described by the preceding theorem are approached rapidly when α is small, but very slowly when α is near 1. Thus, in the Lea & Coulson case corresponding to $\alpha = 1$, the value of K must be very large for ρ to become negligible. For ordinary values of K, the typical value of ρ is non-negligible, and the dispersion of ρ around this typical value is very wide, so that the original argument of Luria & Delbrück is justified.

Note: The formulae on random clones described in Section 2 restate some results obtained in Yule 1924. Yule's paper is known to have introduced the birth process, but has otherwise been neglected. It treated a nominally different problem: our "growth" was his "increase in the number of species in one genus," our "mutation" was his "starting of a new genus." Simon 1954 attempted to modify Yule's argument to obtain diffusion from less strong first-order assumptions. This attempt unavoidably failed.

Editorial comment. This reprint corrects a horrendous typographical error. In the original, both in the abstract and at the end of Section 2, the exponent $-\alpha$ was printed as $-\alpha-1$.