E9

A case against the lognormal distribution

✦ **Abstract.** The lognormal distribution is, in some respects, of great simplicity. This is one reason why, next to the Gaussian, it is widely viewed as the practical statistician's best friend. From the viewpoint described in Chapter E5, it is short-run concentrated and long-run even. This makes it the prototype of the state of slow randomness, the difficult middle ground between the wild and mild state of randomness. Metaphorically, every lognormal resembles a liquid, and a very skew lognormal resembles a glass, which physicists view as a very viscous liquid.

A hard look at the lognormal reveals a new phenomenon of delocalized moments. This feature implies several drawbacks, each of which suffices to make the lognormal dangerous to use in scientific research. Population moments depend overly on *exact* lognormality. Small sample sequential moments oscillate to excess as the sample size increases. A non-negligible concentration rate can only represent a transient that vanishes for large samples. ✦

AFTER LÉVY, ZIPF AND PARETO were described as providing inspiration to scaling and fractal geometry, Chapter E4 also listed a widely-followed nemesis. Robert Gibrat, the author of *Les inégalités économiques* (Gibrat 1932), remains foremost among the many who claim that economic inequalities (presumably all of them) can be described and explained by the lognormal. As is well-known, Λ is called lognormal when $G = \log\Lambda$ is Gaussian. Section 1 recalls the basic facts about the lognormal, and describes in parallel several reasons why it is liked, and counterbalancing reasons why its assets are misleading. In a word: this distribution should be avoided. A major reason, elaborated in Section 2, is that a near-lognormal's *population* moments are overly sensitive to departures from exact lognormalities. A second major reason, elaborated in Section 3, is that the *sample moments* are not to be trusted, because the sequential sample moments oscillate with sample size in erratic and unmanagable manner.

Once again, the preceding paragraph and the rest of this book avoid endless and tiresome repetition of the terms "density," "distribution," "random variable," and the like. It is better to deal with such words as "Gaussian," "lognormal," "Bernoulli," "Poisson," and "scaling" as common names. For example, if there is no loss of intelligibility and the context allows, "lognormal" will be a synonym either of "lognormal distribution," or of "lognormal random variable." Only a slip of the pen can make me use the word "normal" as synonym of "Gaussian." The reason is that in this book the norm is randomness that used to be called "anomalous" and that Chapter E5 describes as "wild." Since the word "lognormal" will not change, I try not to think about its undesirable root.

Some statisticians tell practicing scientists that there is no need to deal with many different random variables, because every variable can be transformed into a Gaussian ... or even a uniform variable. This transformation is discussed and dismissed in Chapter E5.

The lognormal claims to represent both the bell and the tails in distribution of personal income, though only roughly. The scaling is concerned with the tail only, but claims to represent that part in more precise, more enlightening and more useful fashion. The L-stable is claimed in Chapter E10 to represent the tails well and the bell, reasonably. More generally, the lognormal, the scaling and other narrower-purpose distributions continually compete in the many fields of science where skew long-tailed histograms are a fact of life and concentration ratios are not small. My research life began by facing the conflict between the lognormal and the scaling in the study of word frequencies.

The endless conflict between the lognormal and the scaling is illustrated on Figure 1. It is annoying and boring, and its very existence is irritating and implies that the two distributions differ less than their vastly different analytic forms would suggest. Section 3 will show that such is indeed the case: many lognormals can be approximated over wide spans of values of the variable by judiciously chosen scaling, and conversely.

This assertion does *not* endorse the claim by statisticians who despise log-log plots, that "everyone knows that *every* log-log plot is straight; therefore, a straight log-log plot cannot mean anything." If this were true, the scaling distribution could not be conceivably proved wrong ("falsified" in Popper's terminology.) But it would not be a candidate for serious scientific discourse. Be that as it may, all log-log plots *are not* straight.

The lognormal's properties helped Chapter E5 draw a deep difference between mild, slow, and wild "states of randomness." The Gaussian is mildly random. The scaling thrives on its own wildness: it faces the many difficulties due to skewness and long-tailedness, and this is why it is usable and realistic. The lognormal lies between the mild and the wild, in the state of "slow randomness;" it even provides an excellent illustration of this intermediate state and its pitfalls. It is beloved because it passes as mild: moments are easy to calculate and it is easy to take for granted that

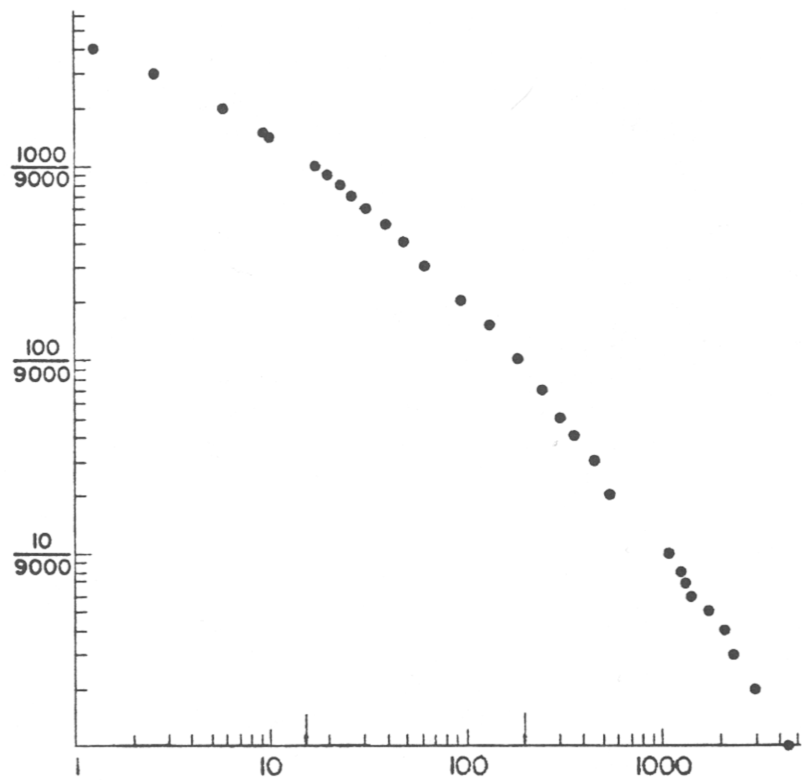

FIGURE C9-1. Illustration of how a sample of a very skew lognormal random variable can "pass" as being from a scaling. The abscissa is $\log x$, the ordinate is $\log \mathrm{Fr}\,(X > x)$, with Fr the frequency in a sample. This is the plot of the distribution (cumulated from the tail) of a sample of 9 000 lognormal variables X, where $\log X$ has zero mean and a standard deviation equal to $\log_e 10$. The graph "passes" as straight. The arrow near $x = 12$ marks the mean, and the arrow near $x = 150$, the mean plus one standard deviation.

they play the same role as for the Gaussian. But they do not. They hide the difficulties due to skewness and long-tailedness behind limits that are overly sensitive and overly slowly attained.

In the metaphor of "states of randomness," the contrast between liquid and solid leaves room for glasses. These hard objects used to be viewed as solids, but their properties are *not* explained by the theory of solids (as a matter of fact, they remain poorly explained). In time, strong physical reasons arose for viewing glasses as being very viscous fluids. The glassy state is a convenient metaphor to characterize the lognormal, but also a challenge that will be taken up in this chapter. Therefore, the lognormal's wondrous properties are irrelevant and thoroughly misleading; it *is not* the statisticians' best friend, perhaps even their worst one. For those reasons, and because of the importance of the topic, this chapter was added to bring together some points also made in other chapters.

Given the serious flaws of the lognormal, there are strong *practical* reasons to prefer the scaling. But scientists learn to live with practical difficulties, when there are solid *theoretical* reasons for doing so. The scaling has diverse strong theoretical points in its favor, while Chapter E8 shows that the usual theoretical argument in favor of lognormality is weak, incomplete and unconvincing. Unfortunately, the fields where the lognormal and the scaling compete lack convincing explanations.

Helpful metaphors. There are many issues that the scaling distribution faces straight on, but the lognormal distribution disguises under a veneer. The lognormal distribution is a wolf in sheep's skin, while the scaling density is a wolf in its own skin; when living among wolves, one must face them on their own terms.

References. The literature on the theory and occurrences of the lognormal is immense and I do not follow it systematically. Aitchison & Brown 1957 was up-to-date when I took up this topic, and I marvelled even then at the length of the mathematical developments built on foundations I viewed as flimsy. See also Johnson, Kotz & Balakrishnan 1994.

A warning against a confusion between "lognormal" and "logBrownian." To my continuing surprise, "lognormal" is also applied here and there to the "logBrownian" model according to which log (price) performs a Brownian motion, à la Bachelier 1900. The only feature common to those two models (not counting the evidence against both) is slight: the logBrownian model asserts that where a price is known at time $t = 0$, its value at time t is a lognormal random variable.

1. INTRODUCTION

1.1 The lognormal's density and its population moments

Let $V = \log\Lambda$ be Gaussian, that is, of probability density

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{(v-\mu)^2}{2\sigma^2} \right\}.$$

Then the probability density of the lognormal $\Lambda = \exp V$ is

$$p(\lambda) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{\lambda} \exp\left\{ -\frac{(\log\lambda-\mu)^2}{2\sigma^2} \right\}.$$

The simplest distinction between states of randomness (Chapter E5) involves the convexity of $\log p(\lambda)$ and the finiteness of the variance.

The cup-convexity of the tail of $\mathbf{\log p(\lambda)}$**.** For the lognormal, there is a "bell" where $\log p(\lambda)$ is cap-convex, and a tail where $\log p(\lambda)$ is cup-convex. Most of the probability is in the bell when σ^2 small, and in the tail when σ^2 is large. If generalized to other distributions, this definition sensibly states that the Gaussian has no tail. Because of the cup-convexity of $\log p(\lambda)$ in the tail, Chapter E5 calls the lognormal "long-tailed."

Finiteness of the moments. An easy classical calculation of $E\Lambda^q$ needed in the sequel yields

$$\begin{aligned}
E\Lambda^q &= \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty \lambda^{q-1} \exp\left\{ -\frac{(\log\lambda-\mu)^2}{2\sigma^2} \right\} d\lambda \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty \exp\left\{ -\frac{(\log\lambda-\mu)^2}{2\sigma^2} + (q-1)\log\lambda \right\} d\lambda \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^\infty \exp\left\{ -\frac{(v-\mu)^2}{2\sigma^2} + qv \right\} dv. \\
&= \frac{1}{\sigma\sqrt{2\sigma}} \exp\left\{ \mu q + \frac{\sigma^2 q^2}{2} \right\} \int_{-\infty}^\infty \exp\left\{ -\frac{[v-(\mu+\sigma^2 q)]^2}{2\sigma^2} \right\} dv.
\end{aligned}$$

Hence the following result, valid for all $q(-\infty < q < \infty)$

$$E\Lambda^q = \exp\left\{\mu q + \frac{\sigma^2 q^2}{2}\right\} < \infty \text{ and } [E\Lambda^q]^{1/q} = \exp\left\{\mu + \frac{\sigma^2 q}{2}\right\} < \infty.$$

Ways of normalizing Λ. One can set $\mu = 0$ by choosing the unit in which λ is measured. To achieve $E\Lambda = 1$, it suffices to set $\mu = -\sigma^2/2$. Using the notations $\mu = -m$ and $\sigma^2 = 2m$.

$$E\Lambda^q = \exp[-qm + q^2 m] = \exp[mq(q-1)].$$

In particular, $E\Lambda^2 = \exp(2m) = \exp\sigma^2$, and the variance is $\nu^2 = \exp(2m) - 1$.

Skewness and long-tailedness. The lognormal's skewness and kurtosis confirm that, as $m \rightarrow \infty$, the distribution becomes increasingly skew and long-tailed. But skewness and kurtosis are less telling than the above-mentioned notions of "bell" and "tail".

The reader is encouraged to draw several lognormal densities, normalized to $E\Lambda = 1$ and parametrized by the standard deviation ν. On both sides of the point of coordinates 1 and $p(1)$, include an interval of length 2ν. As soon as $m > (\log 2)/2 \sim 0.35$, this interval extends to the left of the ordinate axis. This fact underlines the unrepresentative nature of the standard deviation, even in cases of moderate skewness.

This fact also brings to mind one of the deep differences that exist between physics and economics. In physics, moments of low order have a clear theoretical interpretation. For example, the population variance is often an energy that must be finite. In economics, to the contrary, the population variance is nothing but a tool of statistical analysis. Therefore, the only real interest is restricted to the insights that population moments can yield, concerning phenomena ruled by *sample moments*. This chapter will show that the predictions drawn from the lognormal are too confused to be useful while those drawn from the scaling are clear-cut.

1.2 Three main reasons why the lognormal is liked, and more-than-counterbalancing reasons why it should be avoided

An asset: the lognormal density and the formulas for its moments are very simple analytically. So are *products* of lognormals.

A more-than-counterbalancing drawback: the distributions of sums are unmanageably complicated. Dollars and firm sizes do not multiply; they add and subtract. But sums of lognormals are not lognormal and their analytic expressions are unmanageable. That is, the lognormal has invariance properties, but not useful ones.

This is a severe handicap from the viewpoint of the philosophy of invariances described in Chapter E1 and throughout this book. Once again, each scientific or engineering problem involves many fluctuating quantities, linked by a host of necessary relations. A pure curve-fitting doctrine proposes for each quantity the best-fitting theoretical expression, chosen in a long list of all-purpose candidates. But there is no guarantee at all that the separately best fitting expressions are linked by the relations the data must satisfy. For example, take the best fit to one-day and two-day price changes. The distribution of the sum of one day fit need not be analytically manageable, and, even if it is, need not be identical to the distribution of a two-day fit.

Major further drawback: Section 2 shows that the population moments of the lognormal are not at all *robust with respect to small deviations from* absolutely precise *lognormality*. Because of this lack of robustness, X being approximately Gaussian is not good enough from the viewpoint of the population moments of $\exp X$. The known simple values of $E\Lambda^q$ are destroyed by seemingly insignificant deviations. The technical reason behind this feature will be described and called "localization of the moments." Hence, unless lognormality is verified with absolute precision, the moments' values are effectively arbitrary.

The deep differences between the lognormal as an exact or an approximate distribution were unexpected and led to confusions even under the pen of eminent scientists. Few are the flaws in the *Collected Works* of Andrei N. Kolmogorov (1903-1987), but his influential papers on lognormality (especially in the context of turbulence) are deeply flawed. Hard work to correct those flaws led M 1972j{N14} and M 1974f{N15} to results on multifractals that overlap several fields of inquiry and greatly contributed to fractal geometry and the present discussion.

Another major drawback: Section 3 shows that the sequential sample moments of the lognormal behave very erratically. This additional drawback tends to prevent the first one from actually manifesting itself. The population moments of a lognormal or approximate lognormal will eventually be approached, but how rapidly? The answer is: "slowly."

When the lognormal Λ is very skew, sample size increases, the answer is that the sequential sample average undergoes very rough fluctuations, and does not reach the expectation until an irrelevant long-run (corresponding to asymptotically vanishing concentration). In the middle-run, the sample and population averages are largely unrelated and the formulas that give the scatter of the sequential sample moments of the lognormal are impossibly complicated and effectively useless. This behavior is best explained graphically, the Figure captions being an integral part of the text. Figure 2 uses simulated lognormal random variables. while Figure 3 uses data.

Powers of the lognormal being themselves lognormal, all sample moments are averages of lognormals. Their small, and medium sample variability is exteeme and *not* represented by simple rules deduced from lognormality. By contrast, the scaling interpolations of the same data yields simple rules for the very erratic sample variability. Erratically behaving sample moments and diverse other difficulties that the scaling distribution faces straight on, are characteristic of wild randomness.

A widely assumed asset: it is believed that the lognormal is "explained" by a random "proportional effect" argument. Aside from its formal simplicity, the greatest single asset of the Gaussian is that it is the limit in the most important central limit theorem. That theorem's limit is not affected by small changes in the assumptions, more precisely, limit Gaussianity defines a "domain of "universality," within which details do not count. Similarly, the lognormal is ordinarily viewed as being justified via so-called "proportionate effect" models. They represent log X as the sum of independent proportionate effects, then invoke the central limit theorem to conclude that log Z must be approximately Gaussian.

A more-than-counterbalancing drawback: the random proportional effect models yield the Gaussian character of log Λ as an approximation and the conclusions concerning Λ cannot be trusted. In most scientific problems, the lack of exactitude of central limit approximations makes little difference. The number of conceivable multiplicative terms of proportionate effect is not only finite (as always in science) but small. Therefore, the Gaussian involved in the limit theorem is *at best a distant* asymptotic approximation to a preasymptotic reality. When John Maynard Keynes observed that in the long-run we shall be all dead, he implied that asymptotics is fine, but economists should be concerned with what will happen in some middle run. Unfortunately, we deal with one of those cases where, because of the already-mentioned sensitivity, approximations are not sufficient.

Under the lognormal assumption, the basic phenomenon of industrial concentration must be interpreted as a transient that can occur in a small sample, but vanishes asymptotically. In an industry including N firms of lognormally distributed size, how does the relative size of the largest depend on N? This topic is discussed in Chapter E7 and E8.

In the long-run regime $N \to \infty$, the relative size of the largest of N lognormal addends decreases and soon becomes negligible. Hence, a sizeable relative size of the largest, could only be a transient and could only be observed when there are few firms. Furthermore, the formulas that deduce the degree of concentration in this transient are complicated, evade intuition, and must be obtained without any assistance from probability limit theorems.

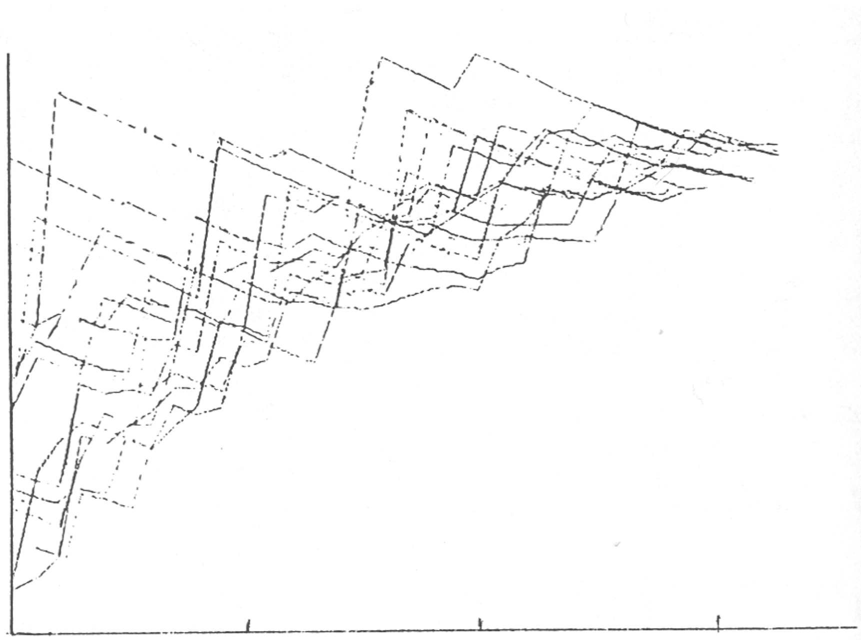

FIGURE C9-2. My oldest illustration of the erratic behavior of the sample averages of very skew approximately lognormal random variables. Several samples were generated, each containing over 10,000 values. Then the sample average $N^{-1}\sum_{\lambda=1}^{N} X_\lambda$ was computed for each sample, and plotted as a line.

Both coordinates are logarithmic. In an initial "transient" zone, the averages scatter over several orders of magnitude. The largest average is often so far removed from the others, that one is tempted to call it an outlier and to disregard it. The approximate limit behavior guaranteed by the law of large numbers is far from being approached. The expectation EX is far larger than the bulk of sample values X_n, which is why huge sample sizes are required for the law of large numbers to apply.

In addition, the limit depends markedly on the Gaussian generator. In this instance, $\log X_n = \sum I_n - 6$, where the I_n are 12 independent pseudorandom variables with uniform distribution. With a different approximation, the limit would be different, but the convergence, equally slow and erratic.

To the contrary, in an industry in which firm size is scaling, the relative size of the largest firm will depend little on the number of firms. Furthermore, the asymptotic result relative to a large number of firms remains a workable first-order approximation where the number of firms is not very large.

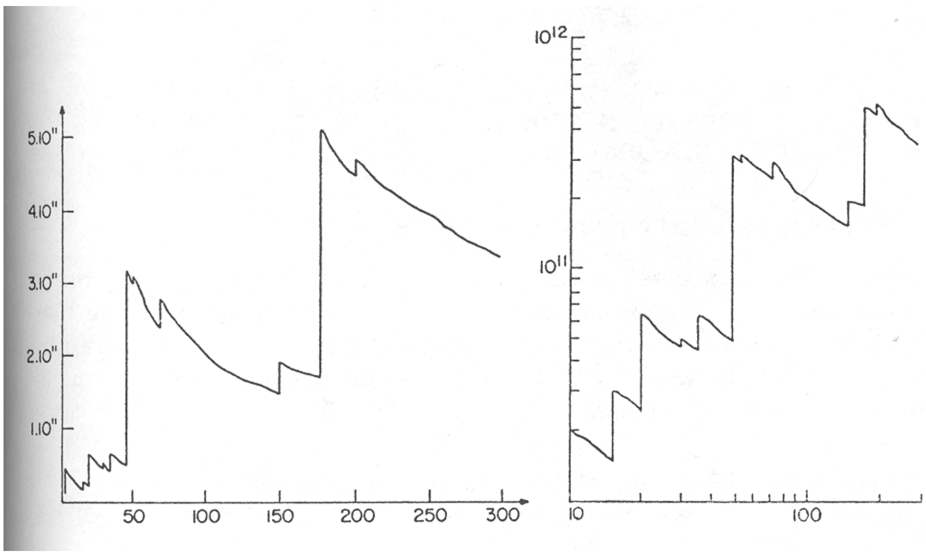

FIGURE C9-3. Illustration of the erratic behavior of the sample mean square of a set of very skew natural data, namely the populations of the 300 largest cities in the USA. This old graph was hand-drawn in 1986. The alphabetical order was picked as approximately random and $N^{-1}\sum X_n^2$ was computed for every value of N. The curve to the left uses linear coordinates in units of 10^{11}; the curve to the right uses log-log coordinates.

There is not even a hint of convergence.

In light of this Figure, examine two conflicting claims. Gibrat 1932 claims that the distribution of city populations is lognormal, and Auerbach 1913 that this distribution is scaling. It may well be that both expressions fit the histograms. But it is clear that the fitted lognormal only describes the asymptotic behavior of the sample mean square and gives no information until the sample size enters the asymptotic range. However, the sample of city sizes is exhaustive, and cannot be increased any further, hence the notion of asymptotic behavior is a figment of the imagination. To the contrary, the fitted scaling distribution does predict the general shape of this Figure.

Conclusion. Even in the study of the transients, it is better to work with the scaling approximation to the lognormal than with the lognormal itself. This scaling approximation makes one expect a range of sizes in which the concentration depends little on N.

The 3 and 4-parameter generalized lognormals. They will not be discussed here. To all the defects of the 2-parameter original, the generalizations add defects of their own. Simplicity is destroyed, the moments are equally meaningless and Gibrat's purported justifications, already shaky for the lognormal, lose all credibility when parameters are added.

2. THE POPULATION MOMENTS OF A NEAR-LOGNORMAL ARE LOCALIZED, THEREFORE OVERLY SENSITIVE TO DEPARTURES FROM EXACT LOGNORMALITY

2.1 Summary of a first argument against the lognormal

The expressions obtained in the Section 1.1 prove to be of little consequence unless the lognormal holds with exactitude beyond anything that any scientist or engineer can reasonably postulate for a statistical distribution. Otherwise, the classical and easily evaluated population moments are *devoid of practical relevance*.

2.2 Even when G is an acceptable Gaussian approximation of Z, the moments of e^G may drastically differ from the moments of e^Z

This sensitivity is a very serious failing. When a theoretical probability distribution is characterized by only a few parameters, a host of properties are intimately tuned to each other. It suffices to verify a few to predict the other. Moving on from a theoretical distribution to one obtained by fitting, one hopes that "small" errors of fitting yield small prediction errors. Such is, indeed, the case for the Gaussian G, but not for the lognormal $\Lambda = e^G$. The trouble is that the practical use of the lognormal consists of predictions that are very sensitive to departure of G from *exact* Gaussianity.

The sensitivity of the lognormal will not be proved theoretically, but will instead be illustrated by comparing a) the Gaussian, and the following near-Gaussian examples: b) a Bernoulli random variable B obtained as sum of K binomial variables bounded by $\max B$, c) a Poisson random variable, P, and d) a gamma random variable Γ obtained as the sum of γ exponentials. Textbooks prove that B, P and Γ can be made "nearly identical" to a normal G. The underlying concept of "near identity" is critical; for sound reasons, it is called "weak" or "vague." Let us show that it allows the moments of the "approximations" e^P, e^B and e^Γ to depend on q in ways that vary with the approximation, and do not match the patterns that is characteristic of e^G.

a) *The lognormal*. To match the Poisson's property that $EP = EP^2 = \rho$, we set $EG = \mu = \rho$ and $\sigma^2 = \rho$. It follows that $[E(e^G)^q]^{1/q} = \exp[\rho(1 + q/2)]$. Thus, $[E(e^G)^q]^{1/q}$ is finite for all q, and increases exponentially.

b) *The logBernoulli*. Here, $[E(e^B)^q]^{1/q} \leq \exp(K \ \max \ B)$. Thus, $[E(e^B)^q]^{1/q}$ is bounded; in the vocabulary of states of randomness expounded in Chapters E5, e^B is mildly random, irrespective of K, but this property is especially devoid of contents from the viewpoint of the small- and the middle-run.

c) *The logPoisson*. $[E(e^P)^q]^{1/q} = \exp[\rho(e^q - 1)/q]$. Thus, $[E(e^P)^q]^{1/q}$ is finite but increases more rapidly than any exponential. Like U_e, the lognormal e^P belongs to the state of slow randomness

d) *The log-gamma*. $E(e^{\Gamma/\alpha})^q = \infty$ when $q > \alpha$. Thus, $e^{\Gamma/\alpha}$ is of the third level of slow randomness when $\alpha > 2$, and is wildly random when $\alpha < 2$.

Expectations. By design, the bells of G and P are very close when ρ is large, but $E(e^G) = \exp(1.5\rho)$ and $E(e^P) = \exp(1.7\rho)$ are very different; this shows that the expectation is not only affected by the bell, which is roughly the same for G and P, but also by their tails, which prove to be very different.

The coefficients of variation. They are

$$\frac{E[(e^G)^2]}{[E(e^G)]^2} - 1 = e^\rho - 1 \text{ and } \frac{E[(e^P)^2]}{[E(e^P)]^2} - 1 = \exp[(e-1)^2\rho] - 1 \sim e^{3\rho} - 1.$$

The dependence on the tails is even greater for e^P than it is for $E\Lambda$.

Higher order moments differ even more strikingly. In short, it does not matter that a large ρ insures that B and P are nearly normal from the usual viewpoint of the "weak-vague" topology. The "predictive error" $E(e^P)^k - E(e^G)^k$ *is not* small. Less good approximations Z yield values of the moments $E(e^Z)^k$ that differ even more from $E(e^G)^k$.

Illustration of the appropriateness of the term "weak topology." In a case beyond wild randomness that is (thankfully) without application but serves as warning, consider $Z = \exp \tilde{V}_N$ where $\tilde{V}_N$ is a normalized sum of scaling addends with $\alpha \geq 2$. By choosing N large enough, the bells of $\tilde{V}_N$ and G are made to coincide as closely as desired. Moreover, $EV^2 < \infty$, hence the central limit theorem tells us that $\tilde{V}_N$ converges to a Gaussian G, that is, comes "close" to G in the "weak", "vague" sense. The underlying topology is powerful enough for the central limit theorem, but for $q > \alpha$ moments *cannot* be matched, since $E\tilde{V}_N^q = \infty$ while $EG^q < \infty$, and *all* positive moments $E\exp(q\tilde{V}_N)$ are infinite, due to the extraordinarily large values of some events that are so extraordinarily rare that they do not matter.

2.3 The moments of the lognormal are sensitive because they are localized, while those of the Gaussian are delocalized

The formula $E\Lambda^q = \exp(\mu q + \sigma^2 q^2/2)$ reduces all the moments of the lognormal to two parameters that describe the middle bell. However, let us consider a general U and take a close look at the integral

$$EU^q = \int u^q p(u) du.$$

For many cases, including the lognormal and the Gaussian, the integrand $u^q p(u)$ has a maximum for $u = \tilde{u}_q$, and one can approximate $q \log u + \log p(u)$ near its maximum by a parabola of the form $-(u - \tilde{\mu}_q)/2\tilde{\sigma}_q^2$, and the integral is little changed if integration is restricted to a "leading interval" of the form $[-\tilde{\sigma}_q + \tilde{\mu}_q, \tilde{\mu}_q + \tilde{\sigma}_q]$, where $\tilde{\sigma}_q$ is the width of the maximum of $u^q p(u)$. When q' is allowed to vary continuously instead of being integer-valued and close to q, the corresponding leading intervals always overlap. We shall now examine what happens as q' moves away from q. There is continuing overlap in the Gaussian, but not in the lognormal case. It follows that different moments of the lognormal are determined by different portions of the density $p(u)$; therefore, it is natural to describe them as *localized*. By small changes in the tail of $p(u)$, one can strongly modify the moments, not independently of each other, to be sure, but fairly independently. This fact will help explain the observations in Section 2.1.

2.3.1 The Gaussian's moments are thoroughly delocalized. Here,

$$\log\,[u^q p(u)] = \text{ a constant } + q \log u - \frac{u^2}{2\sigma^2}.$$

At its maximum, which is $\tilde{\mu}_q = \sigma\sqrt{q}$, the second derivative is $2/\sigma^2$, hence $\tilde{\sigma}_q = \sigma/\sqrt{2}$. Successive leading intervals overlap increasingly as q increases. Numerically, the second percentile of $|G|$ is given in the tables as roughly equal to 2.33. Values around the second percentile greatly affect moments of order 5 or 6. The value $|G| = 3$ is encountered with probability 0.0026, and its greatest effect is on the moment of order $q = 9$. Therefore, samples of only a few thousand are expected to yield nice estimates of moments up to a fairly high order.

2.3.2 The lognormal's moments are localized. For the lognormal, a good choice variable of integration, $v = \log u$ yields the formula in Section 1.1

$$E\Lambda^q = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\mu q + \frac{\sigma^2 q^2}{2}\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{[v-(\mu+\sigma^2 q)]^2}{2\sigma^2}\right\} dv.$$

When $\mu = -m$ and $\sigma^2 = 2m$, so that $E\Lambda = 1$,

$$\tilde{\mu}_q = \mu + \sigma^2 q = m(2q-1), \quad \text{while} \quad \tilde{\sigma}_q = \sigma = \sqrt{2m} \text{ is independent of } q.$$

Consequences of the dependence of $\tilde{\mu}_q$ *and* $\tilde{\sigma}_q$ on q, *concerning the localized character of the moments of the lognormal*. The midpoints of the leading intervals corresponding to q and $q + \Delta q$ differ by $\sigma^2 \Delta q$. When σ is small, the leading intervals overlap only with neighbors. When $\sigma^2 > 2\sigma$, integer qs yield non-overlapping leading intervals.

Consequences of the values of $\tilde{\mu}_q$ *and* $\tilde{\sigma}_q$ *concerning the direct estimation of the moments* $E\Lambda^q$ *from the data on a lognormal* Λ . One can evaluate $E\Lambda^q$ from the mean and variance as estimated from the distribution of $\log \Lambda$, or from Λ itself. The latter method shows that the population moments of the lognormal are delocalized and overly dependent on separate intervals of rare values.

The moment $E\Lambda$. As soon as $m = 2.33$, $\log \tilde{\lambda}_1$ lies on the distribution's first percentile to the right. That is, the estimation of $E\Lambda$ from the λ data is dominated by one percent of the data. As soon as $m = 3.10$, $\log \tilde{\lambda}_1$ corresponds to the first per-mil to the right. That is, the estimation of $E\Lambda$ from the λ data is dominated by one-thousandth of the data.

The moment $E\Lambda^2$. Its estimation is dominated by $\log \tilde{\lambda}_2 \sim \mu + 2\sigma^2 = 3m$. The percent and per-mil thresholds now occur, respectively, for $m = 0.77$ and $m = 1.03$. Therefore, the empirical variance makes no sense, except for very small m and/or a very large sample size.

The moment $E\Lambda^3$. Its estimation, hence the value of the empirical skewness, is dominated by $\log \tilde{\lambda}_3 \sim \mu + 3\sigma^2 = 5m$.

The rest of the argument is obvious and the practical meaninglessness of its estimate is increasingly accentuated as q increases.

Implications of the sensitivity of the population moments to the confidence a scientist may place in them. For both the Gaussian and the lognormal, a standard formula extrapolates all the $E\Lambda^q$ and the tail's shape from two characteristics of the bell, namely, μ and σ^2. For the Gaussian, the extrapolation is safe. For the lognormal, the extrapolated high moments cannot be trusted, unless the underlying distribution is known in full mathematical precision, allowing no approximation. But *absolute exactitude for all* λ is not of this world. For example, Section 1.2 mentioned that the statisticians' attachment to the lognormal is rationalized via the Central Limit Theorem, but this theorem says *nothing* of the tails. Moreover, due to the localization of the lognormal, high order moments largely depend on a tail that is effectively unrelated to the bell.

Prediction in economics demands such as extrapolation from the fitted distribution to larger samples and corresponding larger values. From this viewpoint, data for which one may hesitate between the lognormal and the scaling distributions are genuinely difficult to handle. By fitting the scaling distribution the difficulties are made apparent and can be faced. By contrast, lognormal fitting hides them and prevents them from being recognized, because it fails to be sensitive in the regions in which sensitivity matters. The decision between lognormal or the scaling cannot be helped by the development of better statistical techniques. When data are such that the scaling and lognormal representations are equally defensible, and the limited goal is compression of data for the purpose of filing them away, one may just as well flip a coin. But we must move beyond that limited goal.

3. THE POPULATION MOMENTS OF THE LOGNORMAL BEING LOCALIZED, THE FINITE SAMPLE MOMENTS OSCILLATE IN ERRATIC AND UNMANAGEABLE MANNER

3.1 Summary of the second argument against the lognormal

Population moments can be evaluated in two ways: by *theory*, starting from a known distribution function, or by *statistics*, starting from sample moments in a sufficiently large sample. For the lognormal, Section 2 took up the first method. We now propose to show that the sensitivity of the population moments to rare events has another unfortunate consequence:

the second method to estimate the population moments is no better than the first.

3.2 From exhaustive to sequential sample moments

Every form of science used to depend heavily on the possibility of reducing long lists of data to short lists of "index numbers," such as the moments. But Section 5.1 of Chapter E5 argues that computer graphics decreases this dependence. Moreover, the heavy reliance on moments seems, perhaps unconsciously, related to the notion of statistical sufficiency. As is well-known, the sample average is sufficient for the expectation of a Gaussian, meaning that added knowledge about the individual values in the sample brings no additional information. This is true for estimating the expectations of the Gaussian but not in general. I always believed, in fact, that sample moments pushed concision to excess. This is why my old papers, beginning with M 1963b{E14}, did not simply evaluate a q-th moment, but made sure to record a whole distribution.

3.3 The lognormal's sequential sample moment

Given a set of $N = \max n$ data and an integer q, the sequential qth sample moment is defined by

$$S_q(n) = \frac{1}{n} \sum_{m=1}^{n} U_m^q.$$

The question is how $S_q(n)$ varies as n increases from 1 to $N = \max n$.

For the lognormal Λ and near-lognormals with $EU^q < \infty$, we know that $S_q(n)$ does converge to a limit as $n \to \infty$. But Section 3.4 will show that the sample sizes needed for reliable estimation of the population moments may be colossal, hence impractical. For reasonable sample sizes, convergence is erratic. With a significant or even high probability, the sample moments will seem to vary aimlessly, except that, overall, they appear to increase.

The key fact is that, for large enough q, the event that $U_m^q < EU^q$ has a very high probability, hence also the event that $S_q(n) < EU^q$. Colossal sample sizes are needed to allow $S_q(n)$ to reach up to EU^q.

The nature and intensity of those difficulties depends on skewness. In the limit $\sigma \ll 1$ and $E\Lambda = 1$, one has $\mu = \sigma^2/2$, hence $|\mu| \ll 1$ and $\Lambda = \exp[\sigma(G - \mu)] \sim 1 + \sigma(G - \mu)$. That is, Λ is near Gaussian, and one anticipates sample moments converging quickly. Low-order moments confirm this anticipation. However, Λ^q being also lognormal, the qth moment of one lognormal is the sample average of a less skewed one. Since a large enough q makes the parameters $\sigma_q = q\sigma$ and $\mu = -q^2\sigma^2/2$ as large as desired, Λ^q become arbitrarily far from being Gaussian.

That is, every lognormal's sufficiently high moments eventually misbehave irrespective of the value of σ. Since the moments' behavior does not depend on q and σ separately, but through their product $q\sigma$, we set $q = 1$, and study averages for a lognormal having the single parameter σ.

3.4 The growing sequence of variable "effective scaling exponents" that controls the behavior of the sequential moments of the lognormal

The scaling and lognormal distributions are best compared on log-log plots of the tail densities, but those plots are complicated. To the contrary, the log-log plot of the density are very simple and give roughly the same result. An effective α exponent $\tilde{\alpha}(\lambda)$ is defined by writing

$$\frac{d}{d\lambda} \log p(\lambda) = \frac{d}{d\lambda}\left\{ -\log(\sigma\sqrt{2}) - \log\lambda - \frac{(\log\lambda - \mu)}{2\sigma^2} \right\}$$
$$= -1 - \frac{\log\lambda - \mu}{\sigma^2} = -\tilde{\alpha}(\lambda) - 1.$$

After reduction to $E\Lambda = 1$,

$$\tilde{\alpha}(\lambda) = \frac{\log\lambda - \mu}{\sigma^2} = \frac{1}{2} + \frac{\log\lambda}{\sigma^2}.$$

From Section 2.2.2, the values of λ that contribute most to $E\Lambda^q$ satisfy $\log\lambda \sim \tilde{\mu}_q = \sigma^2(q - 1/2)$, hence yield an effective $\tilde{\alpha}(q) \sim q$. For example, the range corresponding to $q = 1$ yields an effective $\tilde{\alpha}(q) \sim 1$. Within a sample of finite size $N = \max n$, one can say that the behavior of the sequential $S_q(n)$ is not affected by the tail of the density, only by a finite portion, and for the lognormal that finite portion corresponds to an effective $\tilde{\alpha}$ that grows, but slowly.

The existence of an effective $\tilde{\alpha}$ follows from the localization of moments. An effective $\tilde{\alpha}$ is not defined for the Gaussian, or can be said to increase so rapidly that small samples suffice to make it effectively infinite. By contrast, the scaling distribution has a constant $\tilde{\alpha}$, which is the true α.

We know that a scaling X makes specific predictions concerning the distribution of sequential sample moments, and those predictions are simple and identical in the middle and the long-run. When the q-th population moment diverges for $q > \alpha$, the sequential moment $S_q(n)$ has no limit for $n \rightarrow \infty$, but the renormalized form $N^{-q/\alpha}\Sigma X_n^q$ tends in distribution to a L-stable variable of exponent α/q and maximal skewness. In particular, median $[N^{-1}\Sigma X_n^q]$ is finite and proportional to $N^{-1+q/\alpha}$, and the scatter of the sample q-th moment, as represented by the ratio of X_n^q to its median, also tends in distribution to a L-stable random variable.

Why inject the wildness of infinite population moments into a discussion in which all moments are actually safe and finite? Because the very same behavior that some authors used to describe as "improper" is needed to predict about how the sequential moment of the lognormal varies with sample size. While this behavior is practically impossible to obtain from direct analytic derivations, it is readily described from a representative "effective" sequence of scaling distributions.

For small N, the sample $S_q(n)$ will behave as if the lognormal "pretended" to be scaling with a very low α, that is, to be wild with an infinite $E\Lambda$, suggesting that it will *never* converge to a limit value. For larger samples, the lognormal mimics a scaling distribution with $1 < \tilde{\alpha} < 2$, which has a finite $E\Lambda$, but an infinite $E\Lambda^2$. As the sample increases, so does the effective $\tilde{\alpha}(\lambda)$ and the sample variability of the average decreases. It is only as $\lambda \rightarrow \infty$, therefore $\tilde{\alpha}(\lambda) \rightarrow \infty$, that the lognormal distribution eventually acknowledges the truth: it has finite moments of all orders, and $S_q(n)$ ultimately converges. Those successive ranges of values of λ are narrow and overlap when σq is small, but are arbitrarily wide and non-overlapping when σq is large.

But where will the convergence lead? Suppose that Λ is not exactly, only nearly lognormal. The qualitative argument will be the same, but the function $\tilde{\alpha}(\lambda)$ will be different and the ultimate convergence will end up with different asymptotics.

Sequential sample moments that behave erratically throughout a sample are often observed in data analysis, and must be considered a fact of life.