

# Log-normal Distributions across the Sciences: Keys and Clues

ECKHARD LIMPert, WERNER A. STAHEL, AND MARKUS ABBT

**A**s the need grows for conceptualization, formalization, and abstraction in biology, so too does mathematics' relevance to the field (Fagerström et al. 1996). Mathematics is particularly important for analyzing and characterizing random variation of, for example, size and weight of individuals in populations, their sensitivity to chemicals, and time-to-event cases, such as the amount of time an individual needs to recover from illness. The frequency distribution of such data is a major factor determining the type of statistical analysis that can be validly carried out on any data set. Many widely used statistical methods, such as ANOVA (analysis of variance) and regression analysis, require that the data be normally distributed, but only rarely is the frequency distribution of data tested when these techniques are used.

The Gaussian (normal) distribution is most often assumed to describe the random variation that occurs in the data from many scientific disciplines; the well-known bell-shaped curve can easily be characterized and described by two values: the arithmetic mean  $\bar{x}$  and the standard deviation  $s$ , so that data sets are commonly described by the expression  $\bar{x} \pm s$ . A historical example of a normal distribution is that of chest measurements of Scottish soldiers made by Quetelet, Belgian founder of modern social statistics (Swoboda 1974). In addition, such disparate phenomena as milk production by cows and random deviations from target values in industrial processes fit a normal distribution.

However, many measurements show a more or less skewed distribution. Skewed distributions are particularly common when mean values are low, variances large, and values cannot be negative, as is the case, for example, with species abundance, lengths of latent periods of infectious diseases, and distribution of mineral resources in the Earth's crust. Such skewed distributions often closely fit the log-normal distribution (Aitchison and Brown 1957, Crow and Shimizu 1988, Lee 1992, Johnson et al. 1994, Sachs 1997). Examples fitting the normal distribution, which is symmetrical, and the log-normal distribution, which is skewed, are given in Figure 1. Note that body height fits both distributions.

Often, biological mechanisms induce log-normal distributions (Koch 1966), as when, for instance, exponential growth

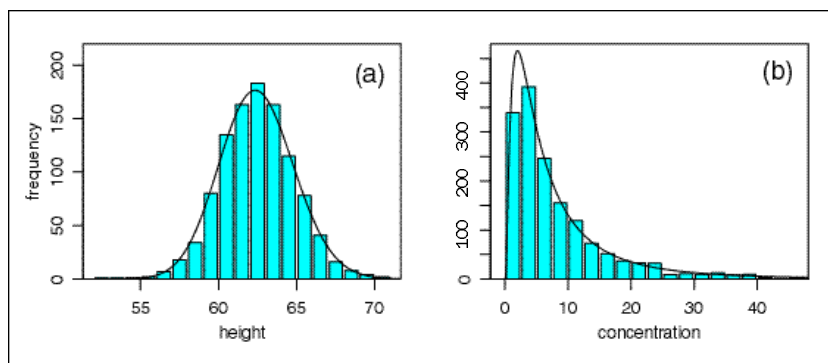
ON THE CHARMS OF STATISTICS, AND HOW MECHANICAL MODELS RESEMBLING GAMBLING MACHINES OFFER A LINK TO A HANDY WAY TO CHARACTERIZE LOG-NORMAL DISTRIBUTIONS, WHICH CAN PROVIDE DEEPER INSIGHT INTO VARIABILITY AND PROBABILITY—NORMAL OR LOG-NORMAL: THAT IS THE QUESTION

is combined with further symmetrical variation: With a mean concentration of, say,  $10^6$  bacteria, one cell division more—or less—will lead to  $2 \times 10^6$ —or  $5 \times 10^5$ —cells. Thus, the range will be asymmetrical—to be precise, multiplied or divided by 2 around the mean. The skewed size distribution may be why “exceptionally” big fruit are reported in journals year after year in autumn. Such exceptions, however, may well be the rule: Inheritance of fruit and flower size has long been known to fit the log-normal distribution (Groth 1914, Powers 1936, Sinnot 1937).

What is the difference between normal and log-normal variability? Both forms of variability are based on a variety of forces acting independently of one another. A major difference, however, is that the effects can be additive or multiplicative, thus leading to normal or log-normal distributions, respectively.

---

Eckhard Limpert (e-mail: Eckhard.Limpert@ipw.agr.ethz.ch) is a biologist and senior scientist in the Phytopathology Group of the Institute of Plant Sciences in Zurich, Switzerland. Werner A. Stahel (e-mail: stahel@stat.math.ethz.ch) is a mathematician and head of the Consulting Service at the Statistics Group, Swiss Federal Institute of Technology (ETH), CH-8092 Zürich, Switzerland. Markus Abbt is a mathematician and consultant at FJA Feilmeier & Junker AG, CH-8008 Zürich, Switzerland. © 2001 American Institute of Biological Sciences.



**Figure 1. Examples of normal and log-normal distributions. While the distribution of the heights of 1052 women (a, in inches; Snedecor and Cochran 1989) fits the normal distribution, with a goodness of fit  $p$  value of 0.75, that of the content of hydroxymethylfurfural (HMF,  $\text{mg}\cdot\text{kg}^{-1}$ ) in 1573 honey samples (b; Renner 1970) fits the log-normal ( $p = 0.41$ ) but not the normal ( $p = 0.0000$ ). Interestingly, the distribution of the heights of women fits the log-normal distribution equally well ( $p = 0.74$ ).**

Some basic principles of additive and multiplicative effects can easily be demonstrated with the help of two ordinary dice with sides numbered from 1 to 6. Adding the two numbers, which is the principle of most games, leads to values from 2 to 12, with a mean of 7, and a symmetrical frequency distribution. The total range can be described as 7 plus or minus 5 (that is,  $7 \pm 5$ ) where, in this case, 5 is not the standard deviation. Multiplying the two numbers, however, leads to values between 1 and 36 with a highly skewed distribution. The total variability can be described as 6 multiplied or divided by 6 (or  $6 \times/6$ ). In this case, the symmetry has moved to the multiplicative level.

Although these examples are neither normal nor log-normal distributions, they do clearly indicate that additive and multiplicative effects give rise to different distributions. Thus, we cannot describe both types of distribution in the same way. Unfortunately, however, common belief has it that quantitative variability is generally bell shaped and symmetrical. The current practice in science is to use symmetrical bars in graphs to indicate standard deviations or errors, and the sign  $\pm$  to summarize data, even though the data or the underlying principles may suggest skewed distributions (Factor et al. 2000, Keesing 2000, Le Naour et al. 2000, Rhew et al. 2000). In a number of cases the variability is clearly asymmetrical because subtracting three standard deviations from the mean produces negative values, as in the example  $100 \pm 50$ . Moreover, the example of the dice shows that the established way to characterize symmetrical, additive variability with the sign  $\pm$  (plus or minus) has its equivalent in the handy sign  $\times/$  (times or divided by), which will be discussed further below.

Log-normal distributions are usually characterized in terms of the log-transformed variable, using as parameters the expected value, or mean, of its distribution, and the standard deviation. This characterization can be advanta-

geous as, by definition, log-normal distributions are symmetrical again at the log level.

Unfortunately, the widespread aversion to statistics becomes even more pronounced as soon as logarithms are involved. This may be the major reason that log-normal distributions are so little understood in general, which leads to frequent misunderstandings and errors. Plotting the data can help, but graphs are difficult to communicate orally. In short, current ways of handling log-normal distributions are often unwieldy.

To get an idea of a sample, most people prefer to think in terms of the original rather than the log-transformed data. This conception is indeed feasible and advisable for log-normal data, too, because the familiar properties of the normal distribution have their analogies in the log-normal distribution. To improve comprehension of log-normal distributions, to encourage

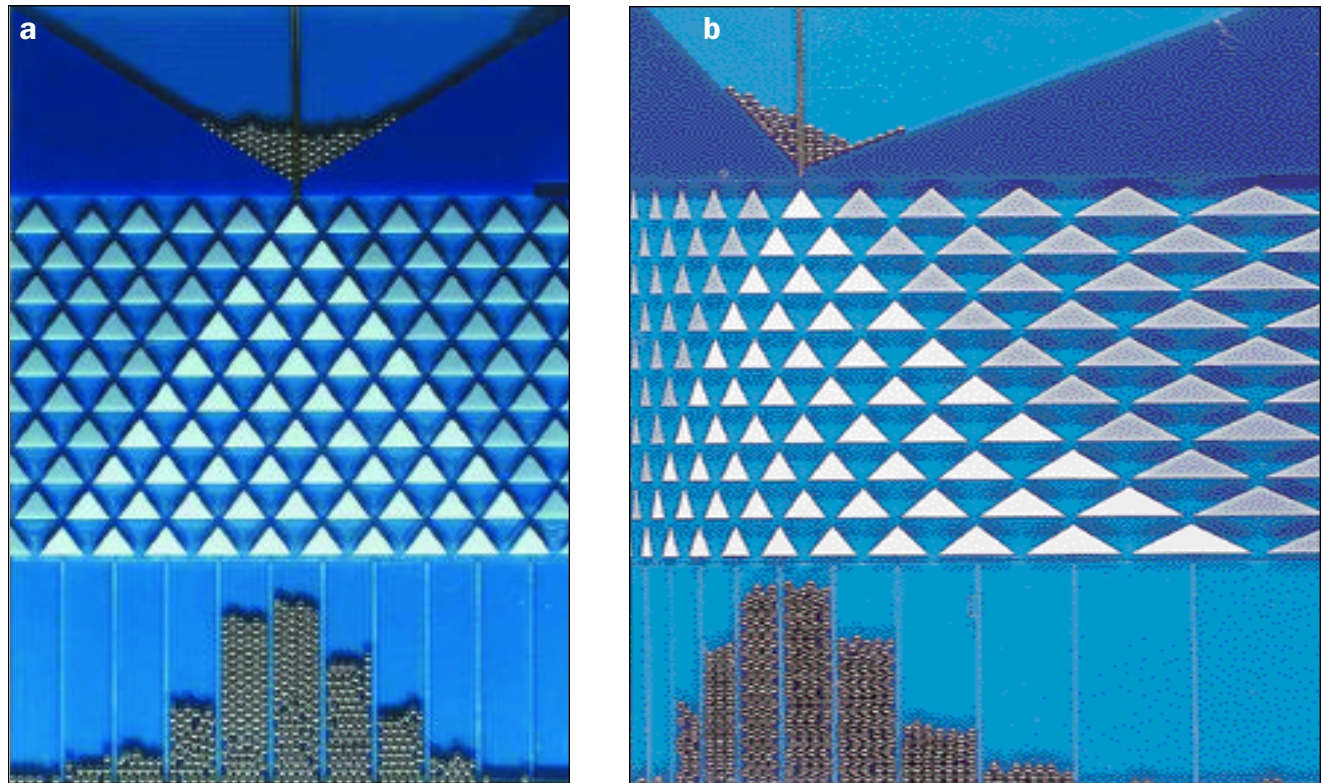
their proper use, and to show their importance in life, we present a novel physical model for generating log-normal distributions, thus filling a 100-year-old gap. We also demonstrate the evolution and use of parameters allowing characterization of the data at the original scale. Moreover, we compare log-normal distributions from a variety of branches of science to elucidate patterns of variability, thereby reemphasizing the importance of log-normal distributions in life.

### ***A physical model demonstrating the genesis of log-normal distributions***

There was reason for Galton (1889) to complain about colleagues who were interested only in averages and ignored random variability. In his thinking, variability was even part of the “charms of statistics.” Consequently, he presented a simple physical model to give a clear visualization of binomial and, finally, normal variability and its derivation.

Figure 2a shows a further development of this “Galton board,” in which particles fall down a board and are deviated at decision points (the tips of the triangular obstacles) either left or right with equal probability. (Galton used simple nails instead of the isosceles triangles shown here, so his invention resembles a pinball machine or the Japanese game Pachinko.) The normal distribution created by the board reflects the cumulative additive effects of the sequence of decision points.

A particle leaving the funnel at the top meets the tip of the first obstacle and is deviated to the left or right by a distance  $c$  with equal probability. It then meets the corresponding triangle in the second row, and is again deviated in the same manner, and so forth. The deviation of the particle from one row to the next is a realization of a random variable with possible values  $+c$  and  $-c$ , and with equal probability for both of them. Finally, after passing  $r$  rows of triangles, the particle falls into



**Figure 2.** Physical models demonstrating the genesis of normal and log-normal distributions. Particles fall from a funnel onto tips of triangles, where they are deviated to the left or to the right with equal probability (0.5) and finally fall into receptacles. The medians of the distributions remain below the entry points of the particles. If the tip of a triangle is at distance  $x$  from the left edge of the board, triangle tips to the right and to the left below it are placed at  $x + c$  and  $x - c$  for the normal distribution (panel a), and  $x \cdot c$  and  $x / c$  for the log-normal (panel b, patent pending),  $c$  and  $c$  being constants. The distributions are generated by many small random effects (according to the central limit theorem) that are additive for the normal distribution and multiplicative for the log-normal. We therefore suggest the alternative name *multiplicative normal distribution* for the latter.

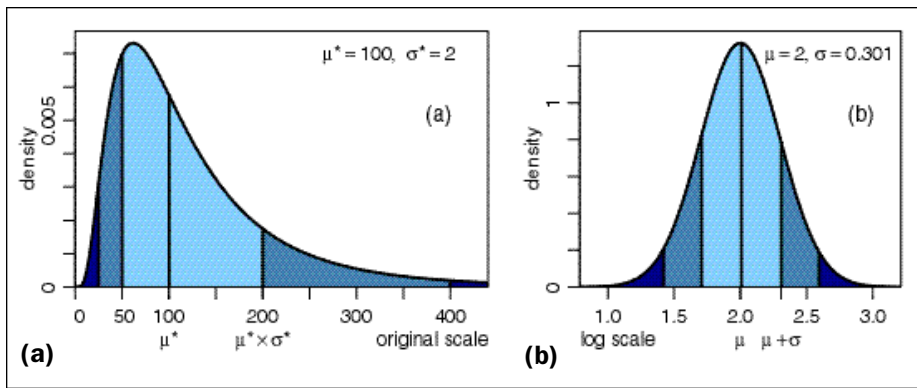
one of the  $r + 1$  receptacles at the bottom. The probabilities of ending up in these receptacles, numbered  $0, 1, \dots, r$ , follow a binomial law with parameters  $r$  and  $p = 0.5$ . When many particles have made their way through the obstacles, the height of the particles piled up in the several receptacles will be approximately proportional to the binomial probabilities.

For a large number of rows, the probabilities approach a normal density function according to the central limit theorem. In its simplest form, this mathematical law states that the sum of many ( $r$ ) independent, identically distributed random variables is, in the limit as  $r \rightarrow \infty$ , normally distributed. Therefore, a Galton board with many rows of obstacles shows normal density as the expected height of particle piles in the receptacles, and its mechanism captures the idea of a sum of  $r$  independent random variables.

Figure 2b shows how Galton's construction was modified to describe the distribution of a product of such variables, which ultimately leads to a log-normal distribution. To this aim, scalene triangles are needed (although they appear to be isosceles in the figure), with the longer side to the right. Let the distance from the left edge of the board to the tip of the first obstacle below the funnel be  $x_m$ . The lower corners of the

first triangle are at  $x_m \cdot c$  and  $x_m / c$  (ignoring the gap necessary to allow the particles to pass between the obstacles). Therefore, the particle meets the tip of a triangle in the next row at  $X = x_m \cdot c$ , or  $X = x_m / c$ , with equal probabilities for both values. In the second and following rows, the triangles with the tip at distance  $x$  from the left edge have lower corners at  $x \cdot c$  and  $x / c$  (up to the gap width). Thus, the horizontal position of a particle is multiplied in each row by a random variable with equal probabilities for its two possible values  $c$  and  $1/c$ .

Once again, the probabilities of particles falling into any receptacle follow the same binomial law as in Galton's device, but because the receptacles on the right are wider than those on the left, the height of accumulated particles is a "histogram" skewed to the left. For a large number of rows, the heights approach a log-normal distribution. This follows from the multiplicative version of the central limit theorem, which proves that the product of many independent, identically distributed, positive random variables has approximately a log-normal distribution. Computer implementations of the models shown in Figure 2 also are available at the Web site <http://stat.ethz.ch/vis/log-normal> (Gut et al. 2000).



**Figure 3.** A log-normal distribution with original scale (a) and with log arithmetic scale (b). Areas under the curve, from the median to both sides, correspond to one and two standard deviation ranges of the normal distribution.

J. C. Kapteyn designed the direct predecessor of the log-normal machine (Kapteyn 1903, Aitchison and Brown 1957). For that machine, isosceles triangles were used instead of the skewed shape described here. Because the triangles' width is proportional to their horizontal position, this model also leads to a log-normal distribution. However, the isosceles triangles with increasingly wide sides to the right of the entry point have a hidden logical disadvantage: The median of the particle flow shifts to the left. In contrast, there is no such shift and the median remains below the entry point of the particles in the log-normal board presented here (which was designed by author E. L.). Moreover, the isosceles triangles in

the Kapteyn board create additive effects at each decision point, in contrast to the multiplicative, log-normal effects apparent in Figure 2b.

Consequently, the log-normal board presented here is a physical representation of the multiplicative central limit theorem in probability theory.

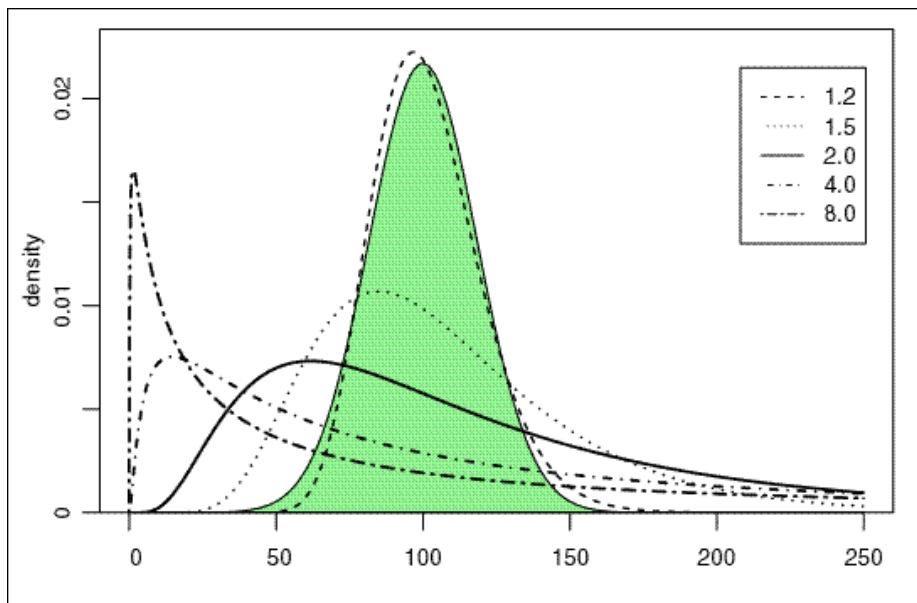
### Basic properties of log-normal distributions

The basic properties of log-normal distribution were established long ago (Weber 1834, Fechner 1860, 1897, Galton 1879, McAlister 1879, Gibrat 1931, Gaddum 1945), and it is not difficult to characterize log-normal distributions

mathematically. A random variable  $X$  is said to be log-normally distributed if  $\log(X)$  is normally distributed (see the box on the facing page). Only positive values are possible for the variable, and the distribution is skewed to the left (Figure 3a). Two parameters are needed to specify a log-normal distribution. Traditionally, the mean  $\mu$  and the standard deviation (or the variance  $\sigma^2$ ) of  $\log(X)$  are used (Figure 3b). However, there are clear advantages to using "back-transformed" values (the values are in terms of  $x$ , the measured data):

(1)  $\mu = e^\mu$ ,  $\sigma = e^\sigma$ . We then use  $X(\mu, \sigma)$  as a mathematical expression meaning that  $X$  is distributed according to the log-normal law with median  $\mu$  and multiplicative standard deviation  $\sigma$ .

The median of this log-normal distribution is  $\text{med}(X) = \mu e^{-\sigma^2}$ , since  $\mu$  is the median of  $\log(X)$ . Thus, the probability that the value of  $X$  is greater than  $\mu$  is 0.5, as is the probability that the value is less than  $\mu$ . The parameter  $\sigma$ , which we call multiplicative standard deviation, determines the shape of the distribution. Figure 4 shows density curves for some selected values of  $\sigma$ . Note that  $\mu$  is a scale parameter; hence, if  $X$  is expressed in different units (or multiplied by a constant for other reasons), then  $\mu$  changes accordingly but  $\sigma$  remains the same.



**Figure 4.** Density functions of selected log-normal distributions compared with a normal distribution. Log-normal distributions  $\Lambda(\mu^*, \sigma^*)$  shown for five values of multiplicative standard deviation,  $\sigma^*$ , are compared with the normal distribution ( $100 \pm 20$ , shaded). The  $\sigma^*$  values cover most of the range evident in Table 2. While the median  $\mu^*$  is the same for all densities, the modes approach zero with increasing shape parameters  $\sigma^*$ . A change in  $\mu^*$  affects the scaling in horizontal and vertical directions, but the essential shape  $\sigma^*$  remains the same.

Distributions are commonly characterized by their expected value  $\mu$  and standard deviation  $\sigma$ . In applications for which the log-normal distribution adequately describes the data, these parameters are usually less easy to interpret than the median  $\mu$  (McAlister 1879) and the shape parameter  $\sigma$ . It is worth noting that  $\sigma$  is related to the

coefficient of variation by a monotonic, increasing transformation (see the box below, eq.2).

For normally distributed data, the interval  $\mu \pm \sigma$  covers a probability of 68.3%, while  $\mu \pm 2\sigma$  covers 95.5% (Table 1). The corresponding statements for log-normal quantities are

$$\left[ \mu / \sigma, \mu + \sigma \right] = \mu^{x/\sigma} \quad (\text{contains } 68.3\%) \text{ and}$$

$$\left[ \mu / (\sigma)^2, \mu + (\sigma)^2 \right] = \mu^{x/(\sigma)^2} \quad (\text{contains } 95.5\%).$$

This characterization shows that the operations of multiplying and dividing, which we denote with the sign  $x/$  (times/divide), help to determine useful intervals for log-normal distributions (Figure 3), in the same way that the operations of adding and subtracting ( $\pm$ , or plus/minus) do for normal distributions. Table 1 summarizes and compares some properties of normal and log-normal distributions.

The sum of several independent normal variables is itself a normal random variable. For quantities with a log-normal distribution, however, multiplication is the relevant operation for combining them in most applications; for example, the product of concentrations determines the speed of a simple

chemical reaction. The product of independent log-normal quantities also follows a log-normal distribution. The median of this product is the product of the medians of its factors. The formula for  $\bar{x}^*$  of the product is given in the box below (eq.3).

For a log-normal distribution, the most precise (i.e., asymptotically most efficient) method for estimating the parameters  $\mu^*$  and  $\sigma^*$  relies on log transformation. The mean and empirical standard deviation of the logarithms of the data are calculated and then back-transformed, as in equation 1. These estimators are called  $\bar{x}^*$  and  $s^*$ , where  $\bar{x}^*$  is the geometric mean of the data (McAlister 1879; eq. 4 in the box below). More robust but less efficient estimates can be obtained from the median and the quartiles of the data, as described in the box below.

As noted previously, it is not uncommon for data with a log-normal distribution to be characterized in the literature by the arithmetic mean  $\bar{x}$  and the standard deviation  $s$  of a sample, but it is still possible to obtain estimates for  $\mu^*$  and

### Definition and properties of the log-normal distribution

A random variable  $X$  is log-normally distributed if  $\log(X)$  has a normal distribution. Usually, natural logarithms are used, but other bases would lead to the same family of distributions, with rescaled parameters. The probability density function of such a random variable has the form

$$f(x) = \frac{1}{x \cdot \sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} (\log(x) - \mu)^2\right)$$

A shift parameter can be included to define a three-parameter family. This may be adequate if the data cannot be smaller than a certain bound different from zero (cf. Aitchison and Brown 1957, page 14). The mean and variance are  $\exp(\mu + \sigma^2/2)$  and  $(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)$ , respectively, and therefore, the coefficient of variation is

$$(2) \quad cv = (\exp(\sigma^2) - 1)^{1/2}$$

which is a function in  $\sigma$  only.

The product of two independent log-normally distributed random variables has the shape parameter

$$(3) \quad \sigma^*(X_1 \cdot X_2) = \exp\left(\left([\log \sigma^*(X_1)]^2 + [\log \sigma^*(X_2)]^2\right)^{1/2}\right)$$

since the variances at the log-transformed variables add.

Estimation: The asymptotically most efficient (maximum likelihood) estimators are

$$(4) \quad \bar{x}^* = \exp\left(\frac{1}{n} \sum_{i=1}^n \log(x_i)\right) = \left(\prod_{i=1}^n x_i\right)^{1/n}$$

$$s^* = \exp\left(\left[\frac{1}{n-1} \sum_{i=1}^n \left[\log\left(\frac{x_i}{\bar{x}^*}\right)\right]^2\right]^{1/2}\right).$$

The quartiles  $q_1$  and  $q_2$  lead to a more robust estimate  $(q_1/q_2)^c$  for  $s^*$ , where  $1/c = 1.349 = 2 \cdot \Phi^{-1}(0.75)$ ,  $\Phi^{-1}$  denoting the inverse standard normal distribution function. If the mean  $\bar{x}$  and the standard deviation  $s$  of a sample are available, i.e. the data is summarized in the form  $\bar{x} \pm s$ , the parameters  $\mu^*$  and  $s^*$  can be estimated from them by using  $\bar{x} / \omega$  and  $\exp[\sqrt{c} \log(\omega)]$ , respectively, with  $\omega = 1 + (s/\bar{x})^2 = 1 + cv^2$ ,  $cv$  = coefficient of variation. Thus, this estimate of  $s^*$  is determined only by the  $cv$  (eq. 2).

**Table 1. A bridge between normal and log-normal distributions.**

Property	Normal distribution (Gaussian, or additive normal,distribution)	Log-normal distribution (Multiplicative normal distribution)
Effects (central limit theorem)	Additive	Multiplicative
Shape of distribution	Symmetrical	Skewed
Models		
Triangle shape	Isosceles	Scalene
Effects at each decision point x	$x \pm c$	$x^x / c$
Characterization		
Mean	$\bar{x}$ , Arithmetic	$\bar{x}^*$ , Geometric
Standard deviation	s, Additive	$s^*$ , Multiplicative
Measure of dispersion	$cv = s/\bar{x}$	$s^*$
Interval of confidence		
68.3%	$\bar{x} \pm s$	$\bar{x}^* \times / s^*$
95.5%	$\bar{x} \pm 2s$	$\bar{x}^* \times / (s^*)^2$
99.7%	$\bar{x} \pm 3s$	$\bar{x}^* \times / (s^*)^3$

Notes: cv = coefficient of variation;  $\times /$  = times/divide, corresponding to plus/minus for the established sign  $\pm$ .

\* (see the box on page 345). For example, Stehmann and De Waard (1996) describe their data as log-normal, with the arithmetic mean  $\bar{x}$  and standard deviation s as  $4.1 \pm 3.7$ . Taking the log-normal nature of the distribution into account, the probability of the corresponding  $\bar{x} \pm s$  interval (0.4 to 7.8) turns out to be 88.4% instead of 68.3%. Moreover, 65% of the population are below the mean and almost exclusively within only one standard deviation. In contrast, the proposed characterization, which uses the geometric mean  $\bar{x}^*$  and the multiplicative standard deviation  $s^*$ , reads  $3.0 \times / 2.2$  (1.36 to 6.6). This interval covers approximately 68% of the data and thus is more appropriate than the other interval for the skewed data.

### Comparing log-normal distributions across the sciences

Examples of log-normal distributions from various branches of science reveal interesting patterns (Table 2). In general, values of  $s^*$  vary between 1.1 and 33, with most in the range of approximately 1.4 to 3. The shapes of such distributions are apparent by comparison with selected instances shown in Figure 4.

**Geology and mining.** In the Earth's crust, the concentration of elements and their radioactivity usually follow a log-normal distribution. In geology, values of  $s^*$  in 27 examples varied from 1.17 to 5.6 (Razumovsky 1940, Ahrens 1954, Malanca et al. 1996); nine other examples are given in Table 2. A closer look at extensive data from different reefs (Krige 1966) indicates that values of  $s^*$  for gold and uranium increase in concert with the size of the region considered.

**Human medicine.** A variety of examples from medicine fit the log-normal distribution. Latent periods (time from infection to first symptoms) of infectious diseases have often

been shown to be log-normally distributed (Sartwell 1950, 1952, 1966, Kondo 1977); approximately 70% of 86 examples reviewed by Kondo (1977) appear to be log-normal. Sartwell (1950, 1952, 1966) documents 37 cases fitting the log-normal distribution. A particularly impressive one is that of 5914 soldiers inoculated on the same day with the same batch of faulty vaccine, 1005 of whom developed serum hepatitis.

Interestingly, despite considerable differences in the median  $\bar{x}^*$  of latency periods of various diseases (ranging from 2.3 hours to several months; Table 2), the majority of  $s^*$  values were close to 1.5. It might be worth trying to account for the similarities and dissimilarities in  $s^*$ . For instance, the small  $s^*$  value of 1.24 in the example of the

Scottish soldiers may be due to limited variability within this rather homogeneous group of people. Survival time after diagnosis of four types of cancer is, compared with latent periods of infectious diseases, much more variable, with  $s^*$  values between 2.5 and 3.2 (Boag 1949, Feinleib and McMahon 1960). It would be interesting to see whether  $\bar{x}^*$  and  $s^*$  values have changed in accord with the changes in diagnosis and treatment of cancer in the last half century. The age of onset of Alzheimer's disease can be characterized with the geometric mean  $\bar{x}^*$  of 60 years and  $s^*$  of 1.16 (Horner 1987).

**Environment.** The distribution of particles, chemicals, and organisms in the environment is often log-normal. For example, the amounts of rain falling from seeded and unseeded clouds differed significantly (Biondini 1976), and again  $s^*$  values were similar (seeding itself accounts for the greater variation with seeded clouds). The parameters for the content of hydroxymethylfurfural in honey (see Figure 1b) show that the distribution of the chemical in 1573 samples can be described adequately with just the two values. Ott (1978) presented data on the Pollutant Standard Index, a measure of air quality. Data were collected for eight US cities; the extremes of  $\bar{x}^*$  and  $s^*$  were found in Los Angeles, Houston, and Seattle, allowing interesting comparisons.

**Atmospheric sciences and aerobiology.** Another component of air quality is its content of microorganisms, which was—not surprisingly—much higher and less variable in the air of Marseille than in that of an island (Di Giorgio et al. 1996). The atmosphere is a major part of life support systems, and many atmospheric physical and chemical properties follow a log-normal distribution law. Among other examples are size distributions of aerosols and clouds and parameters of turbulent processes. In the context of turbulence, the thermal flux from the Earth's surface is formed by eddies, the

**Table 2. Comparing log-normal distributions across the sciences in terms of the original data.  $\bar{x}$  is an estimator of the median of the distribution, usually the geometric mean of the observed data, and  $s^*$  estimates the multiplicative standard deviation, the shape parameter of the distribution; 68% of the data are within the range of  $\bar{x} \pm s^*$ , and 95% within  $\bar{x} \pm (s^*)^2$ . In general, values of  $s^*$  and some of  $\bar{x}$  were obtained by transformation from the parameters given in the literature (cf. Table 3). The goodness of fit was tested either by the original authors or by us.**

Discipline and type of measurement	Example	n	$\bar{x}^*$	$s^*$	Reference
<b>Geology and mining</b>					
Concentration of elements	Ga in diabase	56	17 mg · kg <sup>-1</sup>	1.17	Ahrens 1954
	Co in diabase	57	35 mg · kg <sup>-1</sup>	1.48	Ahrens 1954
	Cu	688	0.37%	2.67	Razumovsky 1940
	Cr in diabase	53	93 mg · kg <sup>-1</sup>	5.60	Ahrens 1954
	<sup>226</sup> Ra	52	25.4 Bq · kg <sup>-1</sup>	1.70	Malanca et al. 1996
	Au: small sections	100	(20 inch-dwt.) <sup>a</sup>	1.39	Krige 1966
	large sections	75,000	n.a.	2.42	Krige 1966
	U: small sections	100	(2.5 inch-lb.) <sup>a</sup>	1.35	Krige 1966
large sections	75,000	n.a.	2.35	Krige 1966	
<b>Human medicine</b>					
Latency periods of diseases	Chicken pox	127	14 days	1.14	Sartwell 1950
	Serum hepatitis	1005	100 days	1.24	Sartwell 1950
	Bacterial food poisoning	144	2.3 hours	1.48	Sartwell 1950
	Salmonellosis	227	2.4 days	1.47	Sartwell 1950
	Poliomyelitis, 8 studies	258	12.6 days	1.50	Sartwell 1952
	Amoebic dysentery	215	21.4 days	2.11	Sartwell 1950
Survival times after cancer diagnosis	Mouth and throat cancer	338	9.6 months	2.50	Boag 1949
	Leukemia myelocytic (female)	128	15.9 months	2.80	Feinleib and McMahon 1960
	Leukemia lymphocytic (female)	125	17.2 months	3.21	Feinleib and McMahon 1960
	Cervix uteri	939	14.5 months	3.02	Boag 1949
Age of onset of a disease	Alzheimer	90	60 years	1.16	Horner 1987
<b>Environment</b>					
Rainfall	Seeded	26	211,600 m <sup>3</sup>	4.90	Biondini 1976
	Unseeded	25	78,470 m <sup>3</sup>	4.29	Biondini 1976
HMF in honey	Content of hydroxymethylfurfural	1573	5.56 g kg <sup>-1</sup>	2.77	Renner 1970
Air pollution (PSI)	Los Angeles, CA	364	109.9 PSI	1.50	Ott 1978
	Houston, TX	363	49.1 PSI	1.85	Ott 1978
	Seattle, WA	357	39.6 PSI	1.58	Ott 1978
<b>Aerobiology</b>					
Airborne contamination by bacteria and fungi	Bacteria in Marseilles	n.a.	630 cfu m <sup>-3</sup>	1.96	Di Giorgio et al. 1996
	Fungi in Marseilles	n.a.	65 cfu m <sup>-3</sup>	2.30	Di Giorgio et al. 1996
	Bacteria on Porquerolles Island	n.a.	22 cfu m <sup>-3</sup>	3.17	Di Giorgio et al. 1996
	Fungi on Porquerolles Island	n.a.	30 cfu m <sup>-3</sup>	2.57	Di Giorgio et al. 1996
<b>Phytomedicine</b>					
Fungicide sensitivity, EC <sub>50</sub> Banana leaf spot	Untreated area	100	0.0078 µg · ml <sup>-1</sup> a.i.	1.85	Romero and Sutton 1997
	Treated area	100	0.063 µg · ml <sup>-1</sup> a.i.	2.42	Romero and Sutton 1997
	After additional treatment	94	0.27 µg · ml <sup>-1</sup> a.i.	>3.58	Romero and Sutton 1997
Powdery mildew on barley	Spain (untreated area)	20	0.0153 µg · ml <sup>-1</sup> a.i.	1.29	Limpert and Koller 1990
	England (treated area)	21	6.85 µg · ml <sup>-1</sup> a.i.	1.68	Limpert and Koller 1990
<b>Plant physiology</b>					
Permeability and solute mobility (rate of constant desorption)	<i>Citrus aurantium</i> /H <sub>2</sub> O/Leaf	73	1.58 10 <sup>-10</sup> m s <sup>-1</sup>	1.18	Baur 1997
	<i>Capsicum annuum</i> /H <sub>2</sub> O/CM	149	26.9 10 <sup>-10</sup> m s <sup>-1</sup>	1.30	Baur 1997
	<i>Citrus aurantium</i> /2,4-D/CM	750	7.41 10 <sup>-7</sup> 1 s <sup>-1</sup>	1.40	Baur 1997
	<i>Citrus aurantium</i> /WL110547/CM	46	2.6310 <sup>-7</sup> 1 s <sup>-1</sup>	1.64	Baur 1997
	<i>Citrus aurantium</i> /2,4-D/CM	16	n.a.	1.38	Baur 1997
	<i>Citrus aurantium</i> /2,4-D/CM + acc.1	16	n.a.	1.17	Baur 1997
	<i>Citrus aurantium</i> /2,4-D/CM	19	n.a.	1.56	Baur 1997
<i>Citrus aurantium</i> /2,4-D/CM + acc.2	19	n.a.	1.03	Baur 1997	
<b>Ecology</b>					
Species abundance	Diatoms (150 species)	n.a.	12.1 i/sp	5.68	May 1981
	Plants (coverage per species)	n.a.	0.4%	7.39	Magurran 1988
	Fish (87 species)	n.a.	2.93%	11.82	Magurran 1988
	Birds (142 species)	n.a.	n.a.	33.15	Preston 1962
	Moths in England (223 species)	15,609	17.5 i/sp	8.66	Preston 1948
	Moths in Maine (330 species)	56,131	19.5 i/sp	10.67	Preston 1948
	Moths in Saskatchewan (277 species)	87,110	n.a.	25.14	Preston 1948
<b>Food technology</b>					
Size of unit (mean diameter)	Crystals in ice cream	n.a.	15 µm	1.5	Limpert et al. 2000b
	Oil drops in mayonnaise	n.a.	20 µm	2	Limpert et al. 2000b
	Pores in cocoa press cake	n.a.	10 µm	1.5-2	Limpert et al. 2000b

Table 2. (continued from previous page)

Discipline and type of measurement	Example	<i>n</i>	$\bar{x}^*$	<i>s</i> <sup>*</sup>	Reference
<b>Linguistics</b>					
Length of spoken words in phone conversation	Different words	738	5.05 letters	1.47	Herdan 1958
	Total occurrence of words	76,054	3.12 letters	1.65	Herdan 1958
Length of sentences	G. K. Chesterton	600	23.5 words	1.58	Williams 1940
	G. B. Shaw	600	24.5 words	1.95	Williams 1940
<b>Social sciences and economics</b>					
Age of marriage	Women in Denmark, 1970s	30,200	(12.4) <sup>a</sup> 10.7 years	1.69	Preston 1981
Farm size in England and Wales	1939	n.a.	27.2 hectares	2.55	Allanson 1992
	1989	n.a.	37.7 hectares	2.90	Allanson 1992
Income	Households of employees in Switzerland, 1990	1.7x10 <sup>6</sup>	sFr. 6,726	1.54	Statistisches Jahrbuch der Schweiz 1997

<sup>a</sup> The shift parameter of a three-parameter log-normal distribution.

Notes: n.a. = not available; PSI = Pollutant Standard Index; acc. = accelerator; i/sp = individuals per species; a.i. = active ingredient; and cfu = colony forming units.

size of which is distributed log-normally (Limpert et al. 2000b).

**Phytomedicine and microbiology.** Examples from microbiology and phytomedicine include the distribution of sensitivity to fungicides in populations and distribution of population size. Romero and Sutton (1997) analyzed the sensitivity of the banana leaf spot fungus (*Mycosphaerella fijiensis*) to the fungicide propiconazole in samples from untreated and treated areas in Costa Rica. The differences in  $\bar{x}^*$  and *s*<sup>\*</sup> among the areas can be explained by treatment history. The *s*<sup>\*</sup> in untreated areas reflects mostly environmental conditions and stabilizing selection. The increase in *s*<sup>\*</sup> after treatment reflects the widened spectrum of sensitivity, which results from the additional selection caused by use of the chemical.

Similar results were obtained for the barley mildew pathogen, *Blumeria (Erysiphe) graminis* f.sp. *hordei*, and the fungicide triadimenol (Limpert and Koller 1990) where, again, *s*<sup>\*</sup> was higher in the treated region. Mildew in Spain, where triadimenol had not been used, represented the original sensitivity. In contrast, in England the pathogen was often treated and was consequently highly resistant, differing by a resistance factor of close to 450 ( $\bar{x}^*$  England /  $\bar{x}^*$  Spain). To obtain the same control of the resistant population, then, the concentration of the chemical would have to be increased by this factor.

The abundance of bacteria on plants varies among plant species, type of bacteria, and environment and has been found to be log-normally distributed (Hirano et al. 1982, Loper et al. 1984). In the case of bacterial populations on the leaves of corn (*Zea mays*), the median population size ( $\bar{x}^*$ ) increased from July and August to October, but the relative variability expressed (*s*<sup>\*</sup>) remained nearly constant (Hi-

rano et al. 1982). Interestingly, whereas *s*<sup>\*</sup> for the total number of bacteria varied little (from 1.26 to 2.0), that for the subgroup of ice nucleation bacteria varied considerably (from 3.75 to 8.04).

**Plant physiology.** Recently, convincing evidence was presented from plant physiology indicating that the log-normal distribution fits well to permeability and to solute mobility in plant cuticles (Baur 1997). For the number of combinations of species, plant parts, and chemical compounds studied, the median *s*<sup>\*</sup> for water permeability of leaves was 1.18. The corresponding *s*<sup>\*</sup> of isolated cuticles, 1.30, appears to be considerably higher, presumably because of the preparation of cuticles. Again, *s*<sup>\*</sup> was considerably higher for mobility of the herbicides Dichlorophenoxyacetic acid (2,4-D) and WL110547 (1-(3-fluoromethylphenyl)-5-U-<sup>14</sup>C-phenoxy-1,2,3,4-tetrazole). One explanation for the differences in *s*<sup>\*</sup> for water and for the other chemicals may be extrapolated from results from food technology, where, for transport through filters, *s*<sup>\*</sup> is smaller for simple (e.g., spherical) particles than for more complex particles such as rods (E. J. Windhab [Eidgenössische Technische Hochschule, Zurich, Switzerland], personal communication, 2000).

Chemicals called accelerators can reduce the variability of mobility. For the combination of *Citrus aurantium* cuticles and 2,4-D, diethyladipate (accelerator 1) caused *s*<sup>\*</sup> to fall from 1.38 to 1.17. For the same combination, tributylphosphate (accelerator 2) caused an even greater decrease, from 1.56 to 1.03. Statistical reasoning suggests that these data, with *s*<sup>\*</sup> values of 1.17 and 1.03, are normally distributed (Baur 1997). However, because the underlying principles of permeability remain the same, we think these cases represent log-normal distributions. Thus, considering only statistical reasons may lead to misclassification, which may handicap further analysis. One



**Table 3. Established methods for describing log-normal distributions.**

Characterization	Disadvantages
<b>Graphical methods</b> Density plots, histograms, box plots	Spacey, difficult to describe and compare
<b>Indication of parameters</b> Logarithm of X, mean, median, standard deviation, variance	Unclear which base of the logarithm should be chosen; parameters are not on the scale of the original data and values are difficult to interpret and use for mental calculations
Skewness and curtosis of X	Difficult to estimate and interpret

question remains: What are the underlying principles of permeability that cause log-normal variability?

**Ecology.** In the majority of plant and animal communities, the abundance of species follows a (truncated) log-normal distribution (Sugihara 1980, Magurran 1988). Interestingly, the range of  $s^*$  for birds, fish, moths, plants, or diatoms was very close to that found within one group or another. Based on the data and conclusions of Preston (1948), we determined the most typical value of  $s^*$  to be 11.6.<sup>1</sup>

**Food technology.** Various applications of the log-normal distribution are related to the characterization of structures in food technology and food process engineering. Such disperse structures may be the size and frequency of particles, droplets, and bubbles that are generated in dispersing processes, or they may be the pores in filtering membranes. The latter are typically formed by particles that are also log-normally distributed in diameter. Such particles can also be generated in dry or wet milling processes, in which log-normal distribution is a powerful approximation. The examples of ice cream and mayonnaise given in Table 2 also point to the presence of log-normal distributions in everyday life.

**Linguistics.** In linguistics, the number of letters per word and the number of words per sentence fit the log-normal distribution. In English telephone conversations, the variability  $s^*$  of the length of all words used—as well as of different words—was similar (Herdan 1958). Likewise, the number of words per sentence varied little between writers (Williams 1940).

**Social sciences and economics.** Examples of log-normal distributions in the social sciences and economics include age of marriage, farm size, and income. The age of first marriage in Western civilization follows a three-parameter log-normal distribution; the third parameter corresponds to age at puberty (Preston 1981). For farm size in England and

Wales, both  $\bar{x}^*$  and  $s^*$  increased over 50 years, the former by 38.6% (Allanson 1992). For income distributions,  $\bar{x}^*$  and  $s^*$  may facilitate comparisons among societies and generations (Aitchison and Brown 1957, Limpert et al. 2000a).

### Typical $s^*$ values

One question arises from the comparison of log-normal distributions across the sciences: To what extent are  $s^*$  values typical for a certain attribute? In some cases, values of  $s^*$  appear to be fairly restricted, as is the case for

the range of  $s^*$  for latent periods of diseases—a fact that Sartwell recognized (1950, 1952, 1966) and Lawrence reemphasized (1988a). Describing patterns of typical skewness at the established log level, Lawrence (1988a, 1988b) can be regarded as the direct predecessor of our comparison of  $s^*$  values across the sciences. Aitchison and Brown (1957), using graphical methods such as quantile–quantile plots and Lorenz curves, demonstrated that log-normal distributions describing, for example, national income across countries, or income for groups of occupations within a country, show typical shapes.

A restricted range of variation for a specific research question makes sense. For infectious diseases of humans, for example, the infection processes of the pathogens are similar, as is the genetic variability of the human population. The same appears to hold for survival time after diagnosis of cancer, although the value of  $s^*$  is higher; this can be attributed to the additional variation caused by cancer recognition and treatment. Other examples with typical ranges of  $s^*$  come from linguistics. For bacteria on plant surfaces, the range of variation of total bacteria is smaller than that of a group of bacteria, which can be expected because of competition. Thus, the ranges of variation in these cases appear to be typical and meaningful, a fact that might well stimulate future research.

### Future challenges

A number of scientific areas—and everyday life—will present opportunities for new comparisons and for more far-reaching analyses of  $s^*$  values for the applications considered to date. Moreover, our concept can be extended also to descriptions based on sigmoid curves, such as, for example, dose–response relationships.

**Further comparisons of  $s^*$  values.** Permeability and mobility are important not only for plant physiology (Baur 1997) but also for many other fields such as soil sciences and human medicine, as well as for some industrial processes. With the help of  $\bar{x}^*$  and  $s^*$ , the mobility of different chemicals through a variety of natural membranes could easily be assessed, allowing for comparisons of the membranes with one another as well as with those for filter actions of soils or with technical membranes and filters. Such comparisons will undoubtedly yield valuable insights.

<sup>1</sup>Species abundance may be described by a log-normal law (Preston 1948), usually written in the form  $S(R) = S_0 \cdot \exp(-a^2 R^2)$ , where  $S_0$  is the number of species at the mode of the distribution. The shape parameter  $a$  amounts to approximately 0.2 for all species, which corresponds to  $s^* = 11.6$ .

**Farther-reaching analyses.** An adequate description of variability is a prerequisite for studying its patterns and estimating variance components. One component that deserves more attention is the variability arising from unknown reasons and chance, commonly called *error variation*, or in this case,  $s^*_E$ . Such variability can be estimated if other conditions accounting for variability—the environment and genetics, for example—are kept constant. The field of population genetics and fungicide sensitivity, as well as that of permeability and mobility, can demonstrate the benefits of analyses of variance.

An important parameter of population genetics is migration. Migration among regions leads to population mixing, thus widening the spectrum of fungicide sensitivity encountered in any one of the regions mentioned in the discussion of phytomedicine and microbiology. Population mixing among regions will increase  $s^*$  but decrease the difference in  $\bar{x}^*$ . Migration of spores appears to be greater, for instance, in regions of Costa Rica than in those of Europe (Limpert et al. 1996, Romero and Sutton 1997, Limpert 1999).

Another important aim of pesticide research is to estimate resistance risk. As a first approximation, resistance risk is assumed to correlate with  $s^*$ . However,  $s^*$  depends on genetic and other causes of variability. Thus, determining its genetic part,  $s^*_G$ , is a task worth undertaking, because genetic variation has a major impact on future evolution (Limpert 1999). Several aspects of various branches of science are expected to benefit from improved identification of components of  $s^*$ . In the study of plant physiology and permeability noted above (Baur 1997), for example, determining the effects of accelerators and their share of variability would be elucidative.

### **Sigmoid curves based on log-normal distributions.**

Dose–response relations are essential for understanding the control of pests and pathogens (Horsfall 1956). Equally important are dose–response curves that demonstrate the effects of other chemicals, such as hormones or minerals. Typically, such curves are sigmoid and show the cumulative action of the chemical. If plotted against the logarithm of the chemical dose, the sigmoid is symmetrical and corresponds to the cumulative curve of the log-normal distribution at logarithmic scale (Figure 3b). The steepness of the sigmoid curve is inversely proportional to  $s^*$ , and the geometric mean value  $\bar{x}^*$  equals the “ED<sub>50</sub>,” the chemical dose creating 50% of the maximal effect. Considering the general importance of chemical sensitivity, a wide field of further applications opens up in which progress can be expected and in which researchers may find the proposed characterization  $\bar{x}^*/s^*$  advantageous.

### **Normal or log-normal?**

Considering the patterns of normal and log-normal distributions further, as well as the connections and distinctions between them (Tables 1, 3), is useful for describing and explaining phenomena relating to frequency distributions in life. Some important aspects are discussed below.

**The range of log-normal variability.** How far can  $s^*$  values extend beyond the range described, from 1.1 to 33? Toward the high end of the scale of possible  $s^*$  values, we found one  $s^*$  larger than 150 for hail energy of clouds (Federer et al. 1986, calculations by W. A. S.). Values below 1.2 may even be common, and therefore of great interest in science. However, such log-normal distributions are difficult to distinguish from normal ones—see Figures 1 and 3—and thus until now have usually been taken to be normal.

Because of the general preference for the normal distribution, we were asked to find examples of data that followed a normal distribution but did not match a log-normal distribution. Interestingly, original measurements did not yield any such examples. As noted earlier, even the classic example of the height of women (Figure 1a; Snedecor and Cochran 1989) fits both distributions equally well. The distribution can be characterized with 62.54 inches  $\pm$  2.38 and 62.48 inches  $\times$  1.039, respectively. The examples that we found of normally—but not log-normally—distributed data consisted of differences, sums, means, or other functions of original measurements. These findings raise questions about the role of symmetry in quantitative variation in nature.

### **Why the normal distribution is so popular.**

Regardless of statistical considerations, there are a number of reasons why the normal distribution is much better known than the log-normal. A major one appears to be symmetry, one of the basic principles realized in nature as well as in our culture and thinking. Thus, probability distributions based on symmetry may have more inherent appeal than skewed ones. Two other reasons relate to simplicity. First, as Aitchison and Brown (1957, p. 2) stated, “Man has found addition an easier operation than multiplication, and so it is not surprising that an additive law of errors was the first to be formulated.” Second, the established, concise description of a normal sample— $\bar{x} \pm s$ —is handy, well-known, and sufficient to represent the underlying distribution, which made it easier, until now, to handle normal distributions than to work with log-normal distributions. Another reason relates to the history of the distributions: The normal distribution has been known and applied more than twice as long as its log-normal sister distribution. Finally, the very notion of “normal” conjures more positive associations for nonstatisticians than does “log-normal.” For all of these reasons, the normal or Gaussian distribution is far more familiar than the log-normal distribution is to most people.

This preference leads to two practical ways to make data look normal even if they are skewed. First, skewed distributions produce large values that may appear to be outliers. It is common practice to reject such observations and conduct the analysis without them, thereby reducing the skewness but introducing bias. Second, skewed data are often grouped together, and their means—which are more normally distributed—are used for further analyses. Of course, following that procedure means that important features of the data may remain undiscovered.

### **Why the log-normal distribution is usually the better model for original data.**

As discussed above, the connection between additive effects and the normal distribution parallels that of multiplicative effects and the log-normal distribution. Kapteyn (1903) noted long ago that if data from one-dimensional measurements in nature fit the normal distribution, two- and three-dimensional results such as surfaces and volumes cannot be symmetric. A number of effects that point to the log-normal distribution as an appropriate model have been described in various papers (e.g., Aitchison and Brown 1957, Koch 1966, 1969, Crow and Shimizu 1988). Interestingly, even in biological systematics, which is the science of classification, the number of, say, species per family was expected to fit log-normality (Koch 1966).

The most basic indicator of the importance of the log-normal distribution may be even more general, however. Clearly, chemistry and physics are fundamental in life, and the prevailing operation in the laws of these disciplines is multiplication. In chemistry, for instance, the velocity of a simple reaction depends on the product of the concentrations of the molecules involved. Equilibrium conditions likewise are governed by factors that act in a multiplicative way. From this, a major contrast becomes obvious: The reasons governing frequency distributions in nature usually favor the log-normal, whereas people are in favor of the normal.

For small coefficients of variation, normal and log-normal distributions both fit well. For these cases, it is natural to choose the distribution found appropriate for related cases exhibiting increased variability, which corresponds to the law governing the reasons of variability. This will most often be the log-normal.

### **Conclusion**

This article shows, in a nutshell, the fundamental role of the log-normal distribution and provides insights for gaining a deeper comprehension of that role. Compared with established methods for describing log-normal distributions (Table 3), the proposed characterization by  $\bar{x}^*$  and  $s^*$  offers several advantages, some of which have been described before (Sartwell 1950, Ahrens 1954, Limpert 1993). Both  $\bar{x}^*$  and  $s^*$  describe the data directly at their original scale, they are easy to calculate and imagine, and they even allow mental calculation and estimation. The proposed characterization does not appear to have any major disadvantage.

On the first page of their book, Aitchison and Brown (1957) stated that, compared with its sister distributions, the normal and the binomial, the log-normal distribution “has remained the Cinderella of distributions, the interest of writers in the learned journals being curiously sporadic and that of the authors of statistical textbooks but faintly aroused.” This is indeed true: Despite abundant, increasing evidence that log-normal distributions are widespread in the physical, biological, and social sciences, and in economics, log-normal knowledge has remained dispersed. The question now is this: Can we begin to bring the wealth of knowledge we have on normal and log-normal distributions to the public? We feel that

doing so would lead to a general preference for the log-normal, or multiplicative normal, distribution over the Gaussian distribution when describing original data.

### **Acknowledgments**

This work was supported by a grant from the Swiss Federal Institute of Technology (Zurich) and by COST (Coordination of Science and Technology in Europe) at Brussels and Bern. Patrick Flüttsch, Swiss Federal Institute of Technology, constructed the physical models. We are grateful to Roy Snaydon, professor emeritus at the University of Reading, United Kingdom, to Rebecca Chasan, and to four anonymous reviewers for valuable comments on the manuscript. We thank Donna Verdier and Herman Marshall for getting the paper into good shape for publication, and E. L. also thanks Gerhard Wenzel, professor of agronomy and plant breeding at Technical University, Munich, for helpful discussions.

Because of his fundamental and comprehensive contribution to the understanding of skewed distributions close to 100 years ago, our paper is dedicated to the Dutch astronomer Jacobus Cornelius Kapteyn (van der Heijden 2000).

### **References cited**

- Ahrens LH. 1954. The log-normal distribution of the elements (A fundamental law of geochemistry and its subsidiary). *Geochimica et Cosmochimica Acta* 5: 49–73.
- Aitchison J, Brown JAC. 1957. *The Log-normal Distribution*. Cambridge (UK): Cambridge University Press.
- Allanson P. 1992. Farm size structure in England and Wales, 1939–89. *Journal of Agricultural Economics* 43: 137–148.
- Baur P. 1997. Log-normal distribution of water permeability and organic solute mobility in plant cuticles. *Plant, Cell and Environment* 20: 167–177.
- Biondini R. 1976. Cloud motion and rainfall statistics. *Journal of Applied Meteorology* 15: 205–224.
- Boag JW. 1949. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society B*, 11: 15–53.
- Crow EL, Shimizu K, eds. 1988. *Log-normal Distributions: Theory and Application*. New York: Dekker.
- Di Giorgio C, Krempff A, Guiraud H, Binder P, Tiret C, Dumenil G. 1996. Atmospheric pollution by airborne microorganisms in the City of Marseilles. *Atmospheric Environment* 30: 155–160.
- Factor VM, Laskowska D, Jensen MR, Woitach JT, Popescu NC, Thorgeirsson SS. 2000. Vitamin E reduces chromosomal damage and inhibits hepatic tumor formation in a transgenic mouse model. *Proceedings of the National Academy of Sciences* 97: 2196–2201.
- Fagerström T, Jagers P, Schuster P, Szathmari E. 1996. Biologists put on mathematical glasses. *Science* 274: 2039–2041.
- Fechner GT. 1860. *Elemente der Psychophysik*. Leipzig (Germany): Breitkopf und Härtel.
- . 1897. *Kollektivmasslehre*. Leipzig (Germany): Engelmann.
- Federer B, et al. 1986. Main results of Grossversuch IV. *Journal of Climate and Applied Meteorology* 25: 917–957.
- Feinleib M, McMahon B. 1960. Variation in the duration of survival of patients with the chronic leukemias. *Blood* 15–16: 332–349.
- Gaddum JH. 1945. Log normal distributions. *Nature* 156: 463, 747.
- Galton F. 1879. The geometric mean, in vital and social statistics. *Proceedings of the Royal Society* 29: 365–367.
- . 1889. *Natural Inheritance*. London: Macmillan.
- Gibrat R. 1931. *Les Inégalités Economiques*. Paris: Recueil Sirey.
- Groth BHA. 1914. The golden mean in the inheritance of size. *Science* 39: 581–584.

- Gut C, Limpert E, Hinterberger H. 2000. A computer simulation on the web to visualize the genesis of normal and log-normal distributions. <http://stat.ethz.ch/vis/log-normal>
- Herdan G. 1958. The relation between the dictionary distribution and the occurrence distribution of word length and its importance for the study of quantitative linguistics. *Biometrika* 45: 222–228.
- Hirano SS, Nordheim EV, Army DC, Upper CD. 1982. Log-normal distribution of epiphytic bacterial populations on leaf surfaces. *Applied and Environmental Microbiology* 44: 695–700.
- Horner RD. 1987. Age at onset of Alzheimer's disease: Clue to the relative importance of etiologic factors? *American Journal of Epidemiology* 126: 409–414.
- Horsfall JG. 1956. Principle of fungicidal actions. *Chronica Botanica* 30.
- Johnson NL, Kotz S, Balakrishnan N. 1994. *Continuous Univariate Distributions*. New York: Wiley.
- Kapteyn JC. 1903. *Skew Frequency Curves in Biology and Statistics*. Astronomical Laboratory, Groningen (The Netherlands): Noordhoff.
- Keesing F. 2000. Cryptic consumers and the ecology of an African Savanna. *BioScience* 50: 205–215.
- Koch AL. 1966. The logarithm in biology. I. Mechanisms generating the log-normal distribution exactly. *Journal of Theoretical Biology* 23: 276–290.
- . 1969. The logarithm in biology. II. Distributions simulating the log-normal. *Journal of Theoretical Biology* 23: 251–268.
- Kondo K. 1977. The log-normal distribution of the incubation time of exogenous diseases. *Japanese Journal of Human Genetics* 21: 217–237.
- Krige DG. 1966. A study of gold and uranium distribution patterns in the Klerksdorp Gold Field. *Geoexploration* 4: 43–53.
- Lawrence RJ. 1988a. The log-normal as event-time distribution. Pages 211–228 in Crow EL, Shimizu K, eds. *Log-normal Distributions: Theory and Application*. New York: Dekker.
- . 1988b. Applications in economics and business. Pages 229–266 in Crow EL, Shimizu K, eds. *Log-normal Distributions: Theory and Application*. New York: Dekker.
- Lee ET. 1992. *Statistical Methods for Survival Data Analysis*. New York: Wiley.
- Le Naour F, Rubinstein E, Jasmin C, Prenant M, Boucheix C. 2000. Severely reduced female fertility in CD9-deficient mice. *Science* 287: 319–321.
- Limpert E. 1993. Log-normal distributions in phytomedicine: A handy way for their characterization and application. *Proceedings of the 6th International Congress of Plant Pathology*; 28 July–6 August, 1993; Montreal, National Research Council Canada.
- . 1999. Fungicide sensitivity: Towards improved understanding of genetic variability. Pages 188–193 in *Modern Fungicides and Antifungal Compounds II*. Andover (UK): Intercept.
- Limpert E, Koller B. 1990. Sensitivity of the Barley Mildew Pathogen to Triadimenol in Selected European Areas. Zurich (Switzerland): Institute of Plant Sciences.
- Limpert E, Finckh MR, Wolfe MS, eds. 1996. *Integrated Control of Cereal Mildews and Rusts: Towards Coordination of Research Across Europe*. Brussels (Belgium): European Commission. EUR 16884 EN.
- Limpert E, Fuchs JG, Stahel WA, 2000a. Life is log normal: On the charms of statistics for society. Pages 518–522 in Häberli R, Scholz RW, Bill A, Welti M, eds. *Transdisciplinarity: Joint Problem-Solving among Science, Technology and Society*. Zurich (Switzerland): Haffmans.
- Limpert E, Abbt M, Asper R, Graber WK, Godet F, Stahel WA, Windhab EJ, 2000b. Life is log normal: Keys and clues to understand patterns of multiplicative interactions from the disciplinary to the transdisciplinary level. Pages 20–24 in Häberli R, Scholz RW, Bill A, Welti M, eds. *Transdisciplinarity: Joint Problem-Solving among Science, Technology and Society*. Zurich (Switzerland): Haffmans.
- Loper JE, Suslow TV, Schroth MN. 1984. Log-normal distribution of bacterial populations in the rhizosphere. *Phytopathology* 74: 1454–1460.
- Magurran AE. 1988. *Ecological Diversity and its Measurement*. London: Croom Helm.
- Malanca A, Gaidolfi L, Pessina V, Dallara G. 1996. Distribution of <sup>226</sup>Ra, <sup>232</sup>Th, and <sup>40</sup>K in soils of Rio Grande do Norte (Brazil). *Journal of Environmental Radioactivity* 30: 55–67.
- May RM. 1981. Patterns in multi-species communities. Pages 197–227 in May RM, ed. *Theoretical Ecology: Principles and Applications*. Oxford: Blackwell.
- McAlister D. 1879. The law of the geometric mean. *Proceedings of the Royal Society* 29: 367–376.
- Ott WR. 1978. *Environmental Indices*. Ann Arbor (MI): Ann Arbor Science.
- Powers L. 1936. The nature of the interaction of genes affecting four quantitative characters in a cross between *Hordeum deficiens* and *H. vulgare*. *Genetics* 21: 398–420.
- Preston FW. 1948. The commonness and rarity of species. *Ecology* 29: 254–283.
- . 1962. The canonical distribution of commonness and rarity. *Ecology* 43: 185–215, 410–432.
- . 1981. Pseudo-log-normal distributions. *Ecology* 62: 355–364.
- Razumovsky NK. 1940. Distribution of metal values in ore deposits. *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS* 9: 814–816.
- Renner E. 1970. *Mathematisch-statistische Methoden in der praktischen Anwendung*. Hamburg (Germany): Parey.
- Rhew CR, Miller RB, Weiss RF. 2000. Natural methyl bromide and methylene chloride emissions from coastal salt marshes. *Nature* 403: 292–295.
- Romero RA, Sutton TB. 1997. Sensitivity of *Mycosphaerella fijiensis*, causal agent of black sigatoka of banana, to propiconazole. *Phytopathology* 87: 96–100.
- Sachs L. 1997. *Angewandte Statistik. Anwendung statistischer Methoden*. Heidelberg (Germany): Springer.
- Sartwell PE. 1950. The distribution of incubation periods of infectious disease. *American Journal of Hygiene* 51: 310–318.
- . 1952. The incubation period of poliomyelitis. *American Journal of Public Health and the Nation's Health* 42: 1403–1408.
- . 1966. The incubation period and the dynamics of infectious disease. *American Journal of Epidemiology* 83: 204–216.
- Sinnot EW. 1937. The relation of gene to character in quantitative inheritance. *Proceedings of the National Academy of Sciences* 23: 224–227.
- Snedecor GW, Cochran WG. 1989. *Statistical Methods*. Ames (IA): Iowa University Press.
- Statistisches Jahrbuch der Schweiz. 1997. Zurich (Switzerland): Verlag Neue Züricher Zeitung.
- Stehmann C, De Waard MA. 1996. Sensitivity of populations of *Botrytis cinerea* to triazoles, benomyl and vinclozolin. *European Journal of Plant Pathology* 102: 171–180.
- Sugihara G. 1980. Minimal community structure: An explanation of species abundance patterns. *American Naturalist* 116: 770–786.
- Swoboda H. 1974. *Knauer's Buch der Modernen Statistik*. München (Germany): Droemer Knauer.
- van der Heijden P. 2000. Jacob Cornelius Kapteyn (1851–1922): A Short Biography. (16 May 2001; [www.strw.leidenuniv.nl/~heijden/kapteynbio.html](http://www.strw.leidenuniv.nl/~heijden/kapteynbio.html))
- Weber H. 1834. *De pulsa resorptione auditu et tactu. Annotationes anatomicae et physiologicae*. Leipzig (Germany): Koehler.
- Williams CB. 1940. A note on the statistical analysis of sentence length as a criterion of literary style. *Biometrika* 31: 356–361.