Log-normal distribution from a process that is not multiplicative but is additive

Hideaki Mouri

Meteorological Research Institute, Nagamine, Tsukuba 305-0052, Japan
(Received 12 February 2013; published 14 October 2013)

The central limit theorem ensures that a sum of random variables tends to a Gaussian distribution as their total number tends to infinity. However, for a class of positive random variables, we find that the sum tends faster to a log-normal distribution. Although the sum tends eventually to a Gaussian distribution, the distribution of the sum is always close to a log-normal distribution rather than to any Gaussian distribution if the summands are numerous enough. This is in contrast to the current consensus that any log-normal distribution is due to a product of random variables, i.e., a multiplicative process, or equivalently to nonlinearity of the system. In fact, the log-normal distribution is also observable for a sum, i.e., an additive process that is typical of linear systems. We show conditions for such a sum, an analytical example, and an application to random scalar fields such as those of turbulence.

I. INTRODUCTION

Consider a random variable that takes positive values, $z > 0$. If its logarithm $\ln z$ obeys a Gaussian (normal) distribution, z itself is said to obey a log-normal distribution [1],

$$f(z) = \frac{1}{z\sqrt{2\pi\kappa_2}} \exp\left[-\frac{(\ln z - \kappa_1)^2}{2\kappa_2}\right]. \qquad (1)$$

Here, $\kappa_1 = \langle \ln z \rangle$ and $\kappa_2 = \langle(\ln z - \langle \ln z \rangle)^2\rangle$ are the average and variance of $\ln z$. This distribution is positively skewed with a long tail on the side of $z > \langle z \rangle$, except for the case of $\kappa_2 \ll 1$ where the distribution is rather close to a nonskewed Gaussian distribution.

The log-normal distribution is important because it has been observed very often at least as a good approximation in the natural and the social sciences [2]. In physics, examples include fragment size [3], crystal size [4], wave transmittance in random media [5], turbulence energy [6], dissipation rate of turbulence energy [7], durations of transient events [8,9], stellar mass [10], and cosmological density fluctuations [11]. Also for temporal fluctuations of some nonlinear systems [12,13], a fundamental parameter is considered to vary slowly in a log-normal distribution.

The current consensus is that any log-normal distribution is due to a multiplicative process, i.e., a product of random variables [14], $\prod_{n=1}^{N} z_n$. Its logarithm is a sum of random variables, $\sum_{n=1}^{N} \ln z_n$. This tends to a Gaussian distribution as $N \to \infty$, according to the central limit theorem [1], if those summands have finite variances, if none of the summands dominates the others, and also if the summands are not highly dependent on one another. It follows that $\prod_{n=1}^{N} z_n$ tends to a log-normal distribution. The multiplicative process is in turn attributed to nonlinearity of the system.

However, for a class of positive random variables, we observe that their sum becomes log-normal before it becomes Gaussian. That is, so long as the total number N of the summands is finite and is large enough, the distribution of the sum is always close to a log-normal distribution rather than to any Gaussian distribution. Since this log-normal distribution tends to a Gaussian distribution as $N \to \infty$, our observation is still in accordance with the central limit theorem.

Thus, log-normality could also be observed for a sum of random variables, i.e., an additive process typical of linear systems that are divisible into independent subsystems. This possibility has to be considered in each observation of the log-normal distribution. We describe the general theory in Sec. II and an analytical example in Sec. III. They are applied to random scalar fields such as those of turbulence in Sec. IV. We conclude with remarks in Sec. V.

II. GENERAL THEORY

The additive process is studied by using an average of positive random variables $z_n > 0$,

$$\bar{z}_N = \frac{1}{N} \sum_{n=1}^{N} z_n. \qquad (2a)$$

Instead of the usual moments $\langle z^m \rangle$, we use the cumulants $\langle z^m \rangle_c = d^m \ln\langle\exp(i\xi z)\rangle/d(i\xi)^m|_{\xi=0}$. They are $\langle z^2 \rangle_c = \langle(z - \langle z \rangle)^2\rangle$, $\langle z^3 \rangle_c = \langle(z - \langle z \rangle)^3\rangle$, $\langle z^4 \rangle_c = \langle(z - \langle z \rangle)^4\rangle - 3\langle(z - \langle z \rangle)^2\rangle^2$, and so on [1]. The cumulant for $m = 2$ is just the variance.

For simplicity (see Sec. I), we assume that the random variables z_n are independent of one another and are identically distributed with the distribution of some random variable z_*. Their cumulants are all assumed to be finite. From these assumptions, since any cumulant of a sum of independent random variables is the sum of cumulants of the variables [1], we have the cumulants of $\bar{z}_N$ as

$$\langle\bar{z}_N^m\rangle_c = \frac{1}{N^m} \sum_{n=1}^{N} \langle z_n^m \rangle_c = \frac{\langle z_*^m \rangle_c}{N^{m-1}}. \qquad (2b)$$

Here, $\langle z_*^m \rangle_c$ has been substituted into each of $\langle z_n^m \rangle_c$. The skewness of $\bar{z}_N$ is

$$\frac{\langle\bar{z}_N^3\rangle_c}{\langle\bar{z}_N^2\rangle_c^{1.5}} = \frac{1}{N^{0.5}} \frac{\langle z_*^3 \rangle_c}{\langle z_*^2 \rangle_c^{1.5}}. \qquad (2c)$$

The kurtosis of $\bar{z}_N$ is

$$\frac{\langle\bar{z}_N^4\rangle_c}{\langle\bar{z}_N^2\rangle_c^{2}} = \frac{1}{N} \frac{\langle z_*^4 \rangle_c}{\langle z_*^2 \rangle_c^{2}}. \qquad (2d)$$

With an increase in N, these skewness and kurtosis decay to the Gaussian value of 0. A faster decay to the Gaussian value of 0 is obtained for $\langle \bar{z}_N^m \rangle_c / \langle \bar{z}_N^2 \rangle_c^{m/2} \propto 1/N^{m/2-1}$ at $m \geqslant 5$. Thus, $\bar{z}_N$ tends to a Gaussian distribution. This is the central limit theorem [1].

However, the above discussion does not determine the shapes of the far tails of the distribution at deviations of $\bar{z}_N$ from $\langle \bar{z}_N \rangle$ that are larger than several times $\langle \bar{z}_N^2 \rangle_c^{0.5}$. They are ignored in the central limit theorem [15,16]. In fact, $\bar{z}_N > 0$ is by definition not exactly Gaussian. The theorem is nevertheless of very practical use because the shapes of those tails are not reliably determined with any finite number of data obtained from any actual observation. Also, in any actual system, N is finite. We define that $\bar{z}_N$ is observed to be Gaussian, regardless of its exact distribution, if its skewness and kurtosis are close enough to 0.

The additive process of Eq. (2a) is regarded as if it were multiplicative. For $z_n = \langle z_n \rangle (1 + \varepsilon_n) > 0$, we consider N such that $N \gg \langle \varepsilon_n^2 \rangle_c^{0.5} = \langle z_*^2 \rangle_c^{0.5} / \langle z_* \rangle$. Then, large deviations of ε_n from $\langle \varepsilon_n \rangle = 0$ are ignored to focus on ε_n within the range of $-N \ll \varepsilon_n \ll +N$,

$$\frac{1}{N} \sum_{n=1}^{N} (1 + \varepsilon_n) = 1 + \frac{\varepsilon_1}{N} + \cdots + \frac{\varepsilon_N}{N} \simeq \left[\prod_{n=1}^{N} (1 + \varepsilon_n) \right]^{1/N}.$$
(3a)

The arithmetic average of z_n is thereby approximated as their geometric average,

$$\frac{1}{N} \sum_{n=1}^{N} z_n \simeq \left(\prod_{n=1}^{N} z_n \right)^{1/N}.$$
(3b)

From Eqs. (2a) and (3b), we obtain an approximation as a multiplicative process,

$$\bar{z}_N \simeq \left(\prod_{n=1}^{N} z_n \right)^{1/N} \quad \text{or} \quad \ln \bar{z}_N \simeq \frac{1}{N} \sum_{n=1}^{N} \ln z_n.$$
(4a)

This is to be used as an exact relation. Indeed, if z_n were distributed only within a finite range, Eq. (4a) would be exact in the limit $N \to \infty$. The cumulants of $\ln \bar{z}_N$ are

$$\langle (\ln \bar{z}_N)^m \rangle_c = \frac{1}{N^m} \sum_{n=1}^{N} \langle (\ln z_n)^m \rangle_c = \frac{\langle (\ln z_*)^m \rangle_c}{N^{m-1}}.$$
(4b)

The skewness and kurtosis of $\ln \bar{z}_N$ are

$$\frac{\langle (\ln \bar{z}_N)^3 \rangle_c}{\langle (\ln \bar{z}_N)^2 \rangle_c^{1.5}} = \frac{1}{N^{0.5}} \frac{\langle (\ln z_*)^3 \rangle_c}{\langle (\ln z_*)^2 \rangle_c^{1.5}}$$
(4c)

and

$$\frac{\langle (\ln \bar{z}_N)^4 \rangle_c}{\langle (\ln \bar{z}_N)^2 \rangle_c^2} = \frac{1}{N} \frac{\langle (\ln z_*)^4 \rangle_c}{\langle (\ln z_*)^2 \rangle_c^2}.$$
(4d)

With an increase in N, they decay to the Gaussian value of 0. Thus, although $\bar{z}_N$ is observed to become Gaussian, $\bar{z}_N$ is also observed to become log-normal.

The above observations are approximate in that they have ignored the large deviations of $\bar{z}_N$ from $\langle \bar{z}_N \rangle$. Those ignored to observe the log-normal distribution are not necessarily the same as those ignored to observe the Gaussian distribution.

Since the log-normal distribution has to obey asymptotes such as $\langle z^3 \rangle_c / \langle z^2 \rangle_c^{1.5} = 3 \langle (\ln z)^2 \rangle_c^{0.5}$ and $\langle z^4 \rangle_c / \langle z^2 \rangle_c^2 = 16 \langle (\ln z)^2 \rangle_c$ at $\langle (\ln z)^2 \rangle_c \ll 1$ [1], it tends eventually to the Gaussian distribution as $N \to \infty$ and hence as $\langle (\ln \bar{z}_N)^2 \rangle_c \to 0$, in accordance with the central limit theorem.

The large deviations have also been ignored in Eqs. (3) and (4), but the resultant log-normality is not necessarily less accurate than the Gaussianity, which results from a similar approximation in the central limit theorem. We also note that Eqs. (3) and (4) could become much more accurate if the same constant were added to each of z_n so that $\langle \varepsilon_n^2 \rangle_c^{0.5}$ were smaller (see Sec. IV).

For $\bar{z}_N$ to tend to a log-normal distribution faster than to any Gaussian distribution, necessary conditions are obtained by comparing Eqs. (4c) and (4d) with Eqs. (2c) and (2d),

$$\left| \frac{\langle (\ln z_*)^3 \rangle_c}{\langle (\ln z_*)^2 \rangle_c^{1.5}} \right| < \left| \frac{\langle z_*^3 \rangle_c}{\langle z_*^2 \rangle_c^{1.5}} \right|$$
(5a)

and

$$\left| \frac{\langle (\ln z_*)^4 \rangle_c}{\langle (\ln z_*)^2 \rangle_c^2} \right| < \left| \frac{\langle z_*^4 \rangle_c}{\langle z_*^2 \rangle_c^2} \right|.$$
(5b)

These are also practically sufficient conditions because it follows that $\bar{z}_N$ at large N is always observed to be log-normal rather than to be Gaussian so long as the observation is based on the skewness and kurtosis as in our definition. If the distributions of the individual summands z_n were not identical, their typical values could be used for Eq. (5).

The conditions of Eq. (5) hold for an extensive class of positively skewed distributions. It is actually known that the skewness of such a distribution, $\langle z^3 \rangle_c / \langle z^2 \rangle_c^{1.5} \gg 0$, is much reduced by transforming the variable z into its logarithm $\ln z$ and thereby by reducing the positive tail [1,17], i.e., $\langle (\ln z)^3 \rangle_c / \langle (\ln z)^2 \rangle_c^{1.5} \ll \langle z^3 \rangle_c / \langle z^2 \rangle_c^{1.5}$ but still $\langle (\ln z)^3 \rangle_c / \langle (\ln z)^2 \rangle_c^{1.5} > - \langle z^3 \rangle_c / \langle z^2 \rangle_c^{1.5}$. The log-normal distribution is hence observable for some extensive class of additive processes. Examples are shown in the sections below.

III. ANALYTICAL EXAMPLE

As an example to observe the log-normal distribution for an additive process, we study the gamma distribution of a positive random variable $z > 0$ [1],

$$f(z) = \frac{1}{\Gamma(\gamma)} z^{\gamma-1} \exp(-z) \quad \text{with} \quad \gamma > 0.$$
(6)

Here, Γ is the gamma function. The parameter γ defines the shape of the distribution, which is positively skewed especially when γ is close to 0. We have the cumulants of z as

$$\langle z^m \rangle_c = \gamma (m-1)!.$$
(7a)

Then, the skewness and kurtosis of z are

$$\frac{\langle z^3 \rangle_c}{\langle z^2 \rangle_c^{1.5}} = \frac{2}{\sqrt{\gamma}} \quad \text{and} \quad \frac{\langle z^4 \rangle_c}{\langle z^2 \rangle_c^2} = \frac{6}{\gamma}.$$
(7b)

For the cumulants of $\ln z$ of the gamma distribution, the definition $\langle (\ln z)^m \rangle_c = d^m \ln \langle \exp(i\xi \ln z) \rangle / d(i\xi)^m |_{\xi=0}$ is known to yield an analytical formula [1,17],

$$\langle (\ln z)^m \rangle_c = \frac{d^m}{d\gamma^m} \ln \Gamma(\gamma).$$
(8a)

The right-hand side is a polygamma function. By using its asymptote for $\gamma \to \infty$ [18], we obtain

$$\langle (\ln z)^m \rangle_c \to (-1)^m \frac{(m-2)!}{\gamma^{m-1}} \quad \text{for} \quad m \geqslant 2. \quad (8b)$$

The skewness and kurtosis of $\ln z$ tend as

$$\frac{\langle (\ln z)^3 \rangle_c}{\langle (\ln z)^2 \rangle_c^{1.5}} \to -\frac{1}{\sqrt{\gamma}} \quad \text{and} \quad \frac{\langle (\ln z)^4 \rangle_c}{\langle (\ln z)^2 \rangle_c^2} \to \frac{2}{\gamma}. \quad (8c)$$

These asymptotes are good approximations of the skewness and kurtosis for $\gamma \gtrsim 10^0$, as shown in Fig. 1 (dotted and solid curves).

The gamma distribution is reproductive [1]. Recall the additive process of Eq. (2a). If its independent random variables z_n are identically distributed with the gamma distribution for $\gamma = \gamma_*$, the functional form of Eq. (7a) ensures that $\sum_{n=1}^{N} z_n$ obeys the gamma distribution for $\gamma = N\gamma_*$. The shape is the same for the distribution of $\bar{z}_N = \sum_{n=1}^{N} z_n/N$.
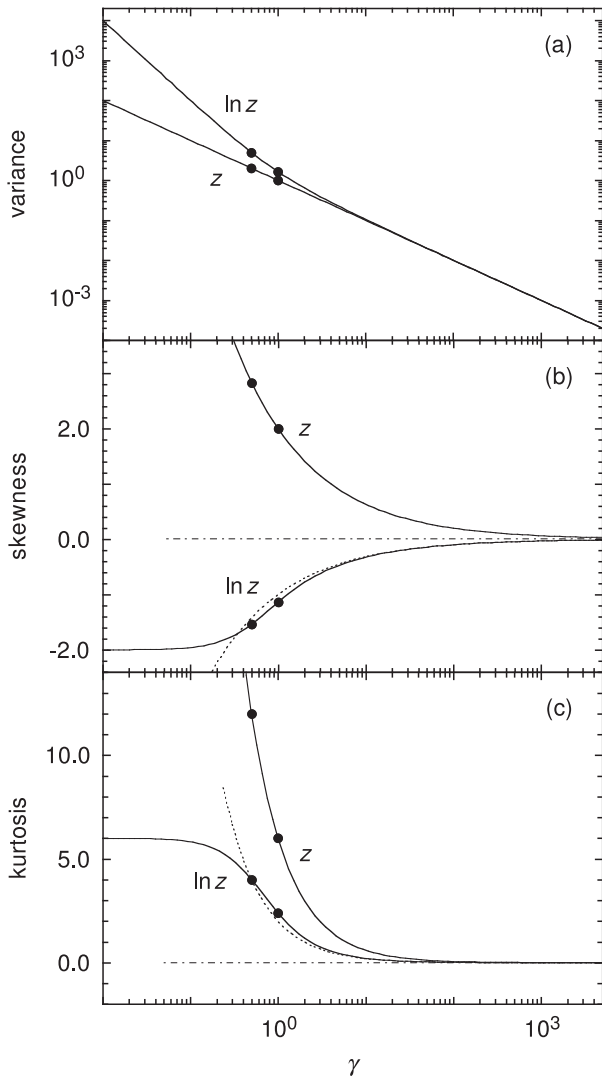

With an increase in $N = \gamma/\gamma_*$, it follows from Eqs. (7) and (8) for $m \geqslant 3$ that $\langle (\ln \bar{z}_N)^m \rangle_c / \langle (\ln \bar{z}_N)^2 \rangle_c^{m/2}$ decays to the Gaussian value of 0 faster than $\langle \bar{z}_N^m \rangle_c / \langle \bar{z}_N^2 \rangle_c^{m/2}$ by a factor of $m-1$ in units of $N^{m/2-1}$. In particular, the skewness and kurtosis of $\ln \bar{z}_N$ decay faster than those of $\bar{z}_N$ by factors of 2 and 3 (see also Fig. 1). Thus, $\bar{z}_N$ tends to a log-normal distribution faster than to any Gaussian distribution. If N is large enough, $\bar{z}_N$ is always observed to be log-normal, according to our definition of the observation in Sec. II.

The gamma distribution arises from various processes in physics. Of importance are those for $\gamma_* = 1/2$ and 1 (filled circles in Fig. 1). While $\gamma_* = 1/2$ corresponds to the distribution of the square of a zero-mean Gaussian random variable, $\gamma_* = 1$ corresponds to an exponential distribution. The gamma distribution is hence expected to explain some of observations of the log-normal distribution in terms of the additive process. For example, such an observation for event duration τ [2,8,9] could be explained by a series of subevents $\sum_{n=1}^{N} \tau_n$, if each of them is exponential, $f(\tau_n) \propto \exp(-\tau_n/\tau_*)$. Another example is the energy E of a canonical ensemble in statistical mechanics [19], which is to be used in the Appendix.

The above features of the gamma distribution confirm our theory in Sec. II. Being consistent with Eq. (5), which is satisfied for any of $\gamma_* > 0$, the distribution observed at any of $N = \gamma/\gamma_* \gg 1/\gamma_*$ is log-normal rather than Gaussian (see Fig. 1). For $\gamma_* = 1/2$, we use Eq. (8a) to obtain the right-hand sides of Eqs. (4c) and (4d) as $-1.54/N^{0.5}$ and $+4/N$. They are close to $-1.41/N^{0.5}$ and $+4/N$, the right-hand sides of Eq. (8c) for $\gamma = N\gamma_* = N/2$. A similar result is found for $\gamma_* = 1$, where those of Eqs. (4c) and (4d) are $-1.14/N^{0.5}$ and $+2.4/N$. Thus, the relations of Eq. (4) are reliable, despite the fact that we have ignored the large deviations of $\bar{z}_N$ from $\langle \bar{z}_N \rangle$.

IV. APPLICATION TO RANDOM FIELD

The log-normal distribution is to be observed for large-scale fluctuations of a random scalar field that is homogeneous along a one-dimensional position x. Although we are interested in the energy $v^2(x)$ of a velocity component v of turbulence as in our past works [6,20], the distribution also applies to, e.g., the cosmological density fluctuations [21]. If time t is considered instead of position x, the scalar is just the energy of some stochastic process [22]. The particular field studied here corresponds to the relaxed stationary state of the Ornstein-Uhlenbeck process that represents Brownian motion and Johnson noise [23].

To focus on the process for the log-normal distribution, we study an idealized Gaussian random field, where v at each position x obeys a zero-mean Gaussian distribution. Any statistical feature of such a field is determined by the two-point correlation of v alone. Its functional form is set to be exponential,

$$\langle v(x+r)v(x) \rangle \propto \exp\left(-\frac{r}{L_v}\right). \quad (9a)$$

Here, the average $\langle \cdot \rangle$ is taken over the position x. The two-point correlation of v^2 is

$$\langle [v^2(x+r) - \langle v^2 \rangle][v^2(x) - \langle v^2 \rangle] \rangle \propto \exp\left(-\frac{r}{L_{v^2}}\right), \quad (9b)$$
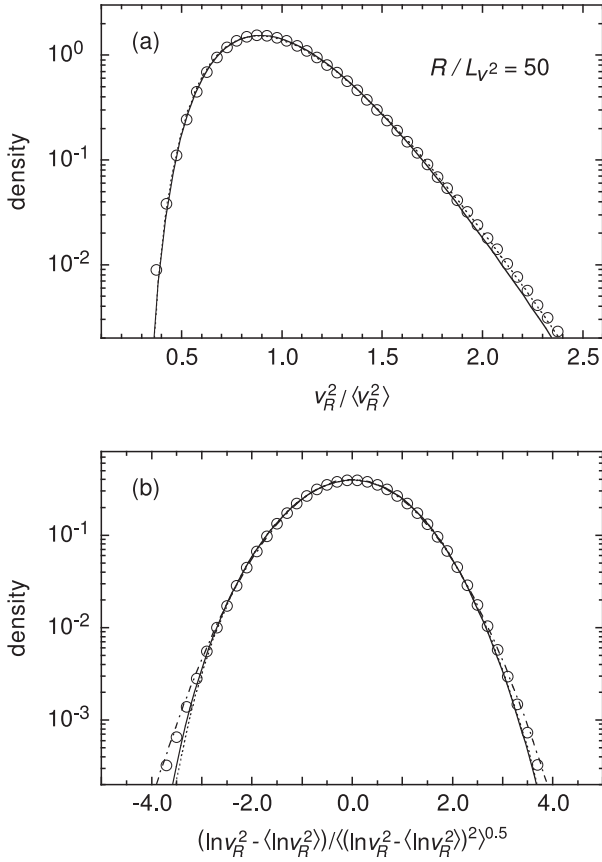
FIG. 1. Variance (a), skewness (b), and kurtosis (c) of z and of $\ln z$ as functions of γ for the gamma distribution of Eq. (6). The variance of z is normalized with the square of $\langle z \rangle$. We indicate the values at $\gamma = 1/2$ and 1 with filled circles. The dotted curves are asymptotes of Eq. (8c). The dot-dashed lines denote the Gaussian value of 0.

FIG. 2. Probability density distributions of $\bar{v}_R^2/\langle\bar{v}_R^2\rangle$ (a) and of $(\ln\bar{v}_R^2 - \langle\ln\bar{v}_R^2\rangle)/\langle(\ln\bar{v}_R^2 - \langle\ln\bar{v}_R^2\rangle)^2\rangle^{0.5}$ (b) at $R = 50L_{v^2}$ for the random field of Eq. (9). The solid and the dotted curves are model predictions described in Sec. IV and in the Appendix. The dot-dashed curve denotes the Gaussian distribution.

with the correlation length

$$L_{v^2} = \frac{L_v}{2}. \tag{9c}$$

This field is numerically calculated by making use of a simple and exact algorithm for the Ornstein-Uhlenbeck process [23, 24]. Since its fluctuations exist at any small scale [23], the calculations are made with different resolutions if necessary.

The one-dimensional field is divided into segments with length R [6,20]. Over each of the segments, the center of which is tentatively defined as x_*, we average the energy v^2 as

$$\bar{v}_R^2(x_*) = \frac{1}{R}\int_{-R/2}^{+R/2} v^2(x_* + x)\,dx. \tag{10}$$

This corresponds to the additive process of Eq. (2a) in the case of $R \gg L_{v^2}$, where the correlation of v^2 is negligible. The segment serves as a linear system because it is divisible into independent subsegments [20]. Within the individual subsegments, we could regard v^2 as constants. Since the overall distribution of v^2 corresponds to Eq. (6) for $\gamma = 1/2$ and thereby satisfies Eq. (5) (see Sec. III), a log-normal distribution could be observable for $\bar{v}_R^2$ among the segments with length $R \gg L_{v^2}$.

Figure 2 compares the distribution of $\bar{v}_R^2$ with that of $\ln\bar{v}_R^2$ at $R = 50L_{v^2}$ (open circles). While the distribution of $\bar{v}_R^2$ is
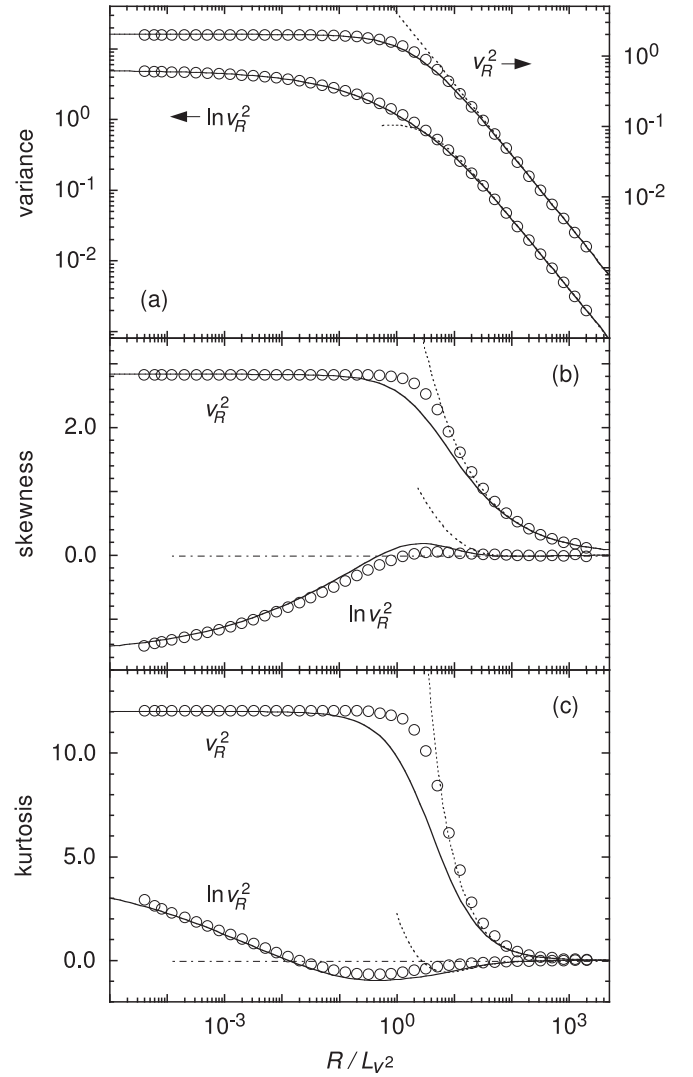

FIG. 3. Variance (a), skewness (b), and kurtosis (c) of $\bar{v}_R^2$ and of $\ln\bar{v}_R^2$ as functions of R/L_{v^2} for the random field of Eq. (9). The variance of $\bar{v}_R^2$ is normalized with the square of $\langle\bar{v}_R^2\rangle$. The solid and the dotted curves are model predictions described in Sec. IV and in the Appendix. The dot-dashed lines denote the Gaussian value of 0.

positively skewed, that of $\ln\bar{v}_R^2$ is symmetric and is close to a Gaussian distribution (dot-dashed curve). Thus, $\bar{v}_R^2$ is log-normal at least as a good approximation, which is also known in terms of the Ornstein-Uhlenbeck process [25].

Figure 3 compares the skewness and kurtosis of $\bar{v}_R^2$ with those of $\ln\bar{v}_R^2$ as functions of R/L_{v^2} (open circles). With an increase in R/L_{v^2}, they decay from the values of v^2 or of $\ln v^2$ to the Gaussian value of 0 (dot-dashed lines). The decay is much faster for $\ln\bar{v}_R^2$. Its skewness and kurtosis are already close to 0 at $R/L_{v^2} \simeq 10^0$, where the skewness and kurtosis of $\bar{v}_R^2$ are still close to the values of v^2. Thus, $\bar{v}_R^2$ tends to a log-normal distribution much faster than to any Gaussian distribution.

These features and others, i.e., features of the variances of $\bar{v}_R^2$ and of $\ln\bar{v}_R^2$, are reproduced by a model described in the Appendix (solid curves in Figs. 2 and 3). The model assumes that $\bar{v}_R^2$ at $R \gg L_{v^2}$ is determined by an additive process of an averaging over independent subsegments with length $4L_{v^2}$

[20]. Their number per segment with length R is

$$N = \frac{R}{4L_{v^2}}, \qquad (11a)$$

which is the single parameter of the model. The gamma distribution of Eq. (6) is used for z defined as

$$z = \frac{N + 1/\sqrt{2}}{\sqrt{2}} \left(\frac{\bar{v}_R^2}{\langle \bar{v}_R^2 \rangle} - \frac{\sqrt{2} - 1}{\sqrt{2}} \frac{N}{N + 1/2} \right) > 0, \quad (11b)$$

with

$$\gamma = \frac{(N + 1/\sqrt{2})^2}{2N + 1}. \qquad (11c)$$

The distributions of $\bar{v}_R^2/\langle \bar{v}_R^2 \rangle$ and of z at each R have the same shape. However, those of $\ln(\bar{v}_R^2/\langle \bar{v}_R^2 \rangle)$ and of $\ln z$ do not because a constant $(\sqrt{2} - 1)/\sqrt{2} \times N/(N + 1/2)$ has been added to $\bar{v}_R^2/\langle \bar{v}_R^2 \rangle$. Actually among the values of $\bar{v}_R^2/\langle \bar{v}_R^2 \rangle$ calculated for Fig. 2, only 0.0001% of them are below the constant. This constant leads to a higher accuracy for $\bar{v}_R^2$ than for z in an approximation as some multiplicative process (see Sec. II). As a result, $\bar{v}_R^2$ in Fig. 3 is more closely log-normal than z in Fig. 1.

We could observe a large-scale log-normal distribution for any homogeneous field of positive random variable, if Eq. (5) holds at each position, and also if any of the n-point correlations decays fast enough for the corresponding nonlinearity to become negligible at the large scales. The functional form of the correlation affects only details of the approach to the log-normal distribution. To normalize the scale R, we generalize the definition of the correlation length L_{v^2} [20],

$$L_{v^2} = \frac{\int_0^\infty \langle [v^2(x + r) - \langle v^2 \rangle][v^2(x) - \langle v^2 \rangle] \rangle dr}{2\langle v^2 \rangle^2}. \quad (12)$$

Figure 4 shows that $\bar{v}_R^2$ at $R \gg L_{v^2}$ in a variety of turbulent flows is log-normal [6] and is consistent with our model [20], regardless of the correlation functions that depend on the configuration of the flow. Not so consistent with our model is $\bar{v}_R^2$ at $R \ll L_{v^2}$ (not shown here). This is because v at each position x is not exactly Gaussian. Also, no fluctuations exist below the scale of the Kolmogorov length [26], which is not 0 so long as the Reynolds number is finite.

V. CONCLUDING REMARKS

The central limit theorem ensures that a sum of random variables is observed to become Gaussian with an increase in the total number of the variables [1], where the observation is just an approximation to ignore too large deviations of the sum from its average [15,16]. However, if the distribution of each of the variables is positively skewed and satisfies the practical conditions of Eq. (5), the sum is observed to become log-normal before it is observed to become Gaussian. This is in contrast to the current consensus that any log-normal distribution is due to a multiplicative process, i.e., a product of random variables [14], or equivalently to nonlinearity of the system. In fact, some are due to additive processes in linear systems that are divisible into independent subsystems. An analytical example is the gamma distribution of Eq. (6).
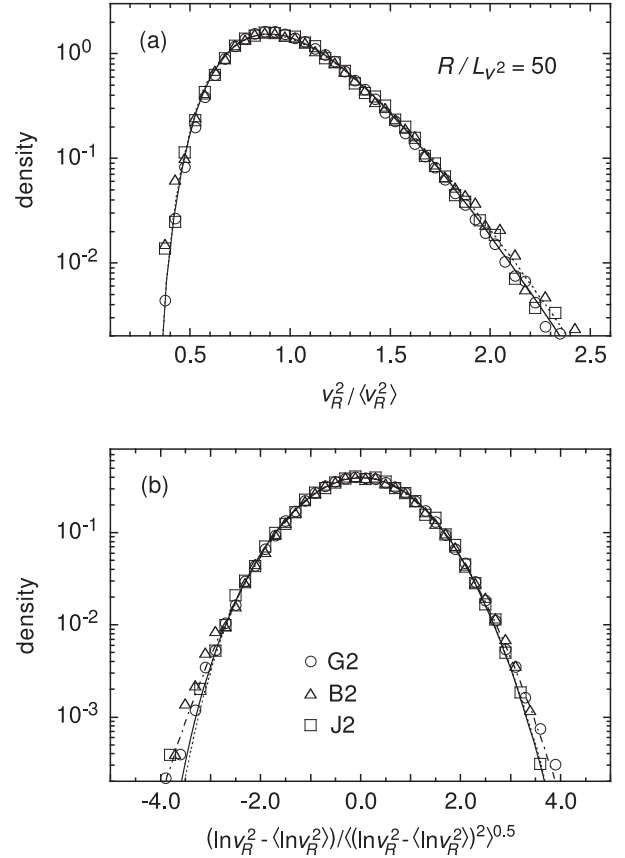

FIG. 4. As in Fig. 2 but for the data of a turbulent grid flow G2 (circles), a turbulent boundary layer B2 (triangles), and a turbulent jet J2 (squares) obtained in our past experiments using a wind tunnel [20].

Through Eq. (11), it reproduces the log-normal distribution observed for large-scale fluctuations of random scalar fields such as those of turbulence [6].

There have been studies similar to ours, i.e., studies for a sum of random variables that are exactly log-normal. No exact explicit formulas are known, but numerical calculations have shown that the sum is approximately log-normal [27,28]. This serves as a result for a special class of additive processes and is consistent with our result.

While we have discussed that a log-normal distribution is observable for a sum of positively skewed variables, the same discussion implies that a Gaussian distribution is observable for a product of negatively skewed variables. That is, Gaussianity is not necessarily due to an additive process in a linear system.

Thus far, we have focused on the cases where the central limit theorem holds because the variances of the summands are finite. If their distributions have, say, power-law tails of $z^{-1-\alpha}$ with $0 < \alpha < 2$ in the limit $z \to \infty$, the variances are not finite and the theorem does not hold. Then, the sum tends to a non-Gaussian stable distribution [1,29]. Its tail remains in the form of $z^{-1-\alpha}$. There is no notion of large deviations because they have to be much larger than the square root of the variance.

The log-normal distribution has been often related to such a power-law distribution [30], by assuming that both of them arise from nonlinearity of the system. Indeed, the standard

process for the log-normal distribution, i.e., multiplication of random variables, leads instead to a power-law distribution if a minimum value is set for the variables [31] or if random noise is added at each step of the multiplication [32]. These two distributions are nevertheless distinct. Although the log-normal distribution could have a long tail, all of its cumulants are finite. The power-law distribution has some infinite cumulants as shown above for the case of the variance.

Having observed that the log-normal distribution also arises from additive processes in linear systems, we are rather interested in its relation to the Gaussian distribution. It is of particular interest to reconsider the existing observations of the log-normal distribution [2–13]. The observed skewness is often small so that the distribution is confusingly similar to a Gaussian distribution [2]. Such a case might turn out to be an example for a log-normal distribution from a sum, which tends to a Gaussian distribution with a further increase in the total number of the summands. A promising approach is to study the distribution beyond the range of the central limit theorem, i.e., large deviations from the average, or to study the process itself, e.g., whether it is for a linear or for a nonlinear system.

APPENDIX: THEORETICAL MODEL OF EQ. (11)

The model of Eq. (11) is an extension of our past model [20]. We begin by obtaining the variance of $\bar{v}_R^2$ among the segments with length R [22],

$$
\langle(\bar{v}_R^2 - \langle\bar{v}_R^2\rangle)^2\rangle
$$
$$
= \frac{2}{R^2}\int_0^R (R-r)\langle[v^2(x+r) - \langle v^2\rangle][v^2(x) - \langle v^2\rangle]\rangle dr.
$$
$$(A1a)$$

If the two-point correlation of v^2 is negligible at large r, Eqs. (12) and (A1a) yield

$$
\langle(\bar{v}_R^2 - \langle\bar{v}_R^2\rangle)^2\rangle \to \frac{4L_{v^2}}{R}\langle v^2\rangle^2 \quad \text{for} \quad R \to \infty. \quad (A1b)
$$

Thus, L_{v^2} is naturally incorporated into our model. On the other hand, if the distribution of v is Gaussian,

$$
\langle(\bar{v}_R^2 - \langle\bar{v}_R^2\rangle)^2\rangle \to 2\langle v^2\rangle^2 \quad \text{for} \quad R \to 0. \quad (A1c)
$$

These asymptotes are used to approximate Eq. (A1a) at any R as

$$
\langle(\bar{v}_R^2 - \langle\bar{v}_R^2\rangle)^2\rangle \simeq \frac{4L_{v^2}}{R + 2L_{v^2}}\langle v^2\rangle^2. \quad (A1d)
$$

If the fluctuations of $\bar{v}_R^2$ can be regarded as in an equilibrium state, Eq. (A1d) has an analog in statistical mechanics, i.e., a formula for thermal fluctuations of energy E in a canonical ensemble at temperature T [19],

$$
\langle(E - \langle E\rangle)^2\rangle = CT^2 \quad \text{with} \quad C = \frac{d\langle E\rangle}{dT}. \quad (A2)
$$

This is the case even in turbulence [20]. We have defined the number of subsegments with length $4L_{v^2}$ as $N = R/4L_{v^2}$ in Eq. (11a), which is used to relate Eq. (A1d) with Eq. (A2),

$$
E = N[\bar{v}_R^2 - (1 - \sqrt{\zeta})\langle v^2\rangle] + \frac{\bar{v}_R^2}{2} \quad (A3a)
$$

and

$$
T = \frac{2N+1}{2\sqrt{\zeta}N+1}\langle v^2\rangle. \quad (A3b)
$$

Here, $\zeta > 0$ is a constant. Then,

$$
\langle E\rangle = CT \quad \text{with} \quad C = \frac{(\sqrt{2\zeta}N + 1/\sqrt{2})^2}{2N+1}. \quad (A3c)
$$

The first term in the right-hand side of Eq. (A3a) represents an additive process. This term is dominant over the second term in the thermodynamic limit of $N \gg 1$, where Eq. (A3) reproduces formulas of thermodynamics [20].

The energy distribution $f(E)$ in any canonical ensemble is determined by the heat capacity C through a series of relations in statistical mechanics [19]. From C in Eq. (A3c), we obtain $f(E)$ in the form of the gamma distribution of Eq. (6),

$$
f(E) = \frac{E^{C-1}\exp(-E/T)}{\Gamma(C)T^C}. \quad (A4a)
$$

Especially in the thermodynamic limit, the distribution is reduced to our past model [20],

$$
f(E) = \frac{E^{\zeta N-1}\exp(-E/T)}{\Gamma(\zeta N)T^{\zeta N}} \quad \text{for} \quad N \gg 1. \quad (A4b)
$$

This model corresponds to the exact Eq. (A1b) rather than to the approximate Eq. (A1d). By also using Eqs. (7a) and (A3a), we have $\langle(\bar{v}_R^2)^m\rangle_c \propto 1/N^{m-1}$ and hence $\propto 1/R^{m-1}$ in accordance with Eq. (2b).

To set the value of ζ, we assume universality of $f(E)$ at $N \gg 1$. Since any interaction occurs only between adjacent members of the subsegments, the energy distribution $f(E)$ at $N = R/4L_{v^2} \gg 1$ is determined by many random steps of energy transfer from the subsegment length scale $4L_{v^2}$ to the segment length scale R. These should have randomized any effect of the individual subsegments [26], as implied by our results in Figs. 2 and 4. Consider first a special case where the energies of the subsegments are independent of one another and are identically distributed with the distribution of the square of a zero-mean Gaussian random variable. The resultant $f(E)$ at $N \gg 1$ is the gamma distribution for $\gamma = N/2$ (see Sec. III), which is described by Eq. (A4b) with

$$
\zeta = \tfrac{1}{2}. \quad (A5)
$$

Then, also in other general cases, the universality leads to $\zeta = 1/2$. This value reproduces the law of equipartition of energy $\langle E\rangle = NT/2$ [19] at $N \gg 1$ in Eq. (A3c). From Eqs. (A3), (A4a), and (A5), we obtain Eq. (11).

Lastly, we note that Eq. (A4a) does not serve as a very good model at the scales of $N = R/4L_{v^2} \simeq 1$ because we have not considered their details (see Fig. 3). The model could be refined by refining its approximation for $\langle(\bar{v}_R^2 - \langle\bar{v}_R^2\rangle)^2\rangle$ in Eq. (A1d). However, such refinement depends on the functional form of the two-point correlation, and hence it is

not universal. The exact and universal model of Eq. (A4b) for $N \gg 1$ is rather reliable at least over a part of those scales (dotted curves in Figs. 2–4), where $\bar{v}_R^2$ has become almost log-normal.

[1] M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, 4th ed. (Griffin, London, 1977), Vol. 1.

[2] See E. Limpert, W. A. Stahel, and M. Abbt, BioScience **51**, 341 (2001) for a review of observations of the log-normal distribution in the natural and the social sciences.

[3] A. N. Kolmogorov, Dokl. Akad. Nauk SSSR **31**, 99 (1941).

[4] P. Feltham, Acta Metall. **5**, 97 (1957).

[5] V. I. Tatarskii, *Wave Propagation in a Turbulent Medium* (McGraw-Hill, New York, 1961).

[6] H. Mouri, A. Hori, and M. Takaoka, Phys. Fluids **21**, 065107 (2009).

[7] A. N. Kolmogorov, J. Fluid Mech. **13**, 82 (1962).

[8] N. Cianos and E. T. Pierce, Technical Report, Stanford Research Institute, Menlo Park, 1972.

[9] B. McBreen, K. J. Hurley, R. Long, and L. Metcalfe, Mon. Not. R. Astron. Soc. **271**, 662 (1994).

[10] G. E. Miller and J. M. Scalo, Astrophys. J. Suppl. Ser. **41**, 513 (1979).

[11] E. Hubble, Astrophys. J. **79**, 8 (1934).

[12] A. M. Oboukhov, J. Fluid Mech. **13**, 77 (1962).

[13] C. Beck and E. G. D. Cohen, Physica A **322**, 267 (2003).

[14] R. Gibrat, Bull. Statist. Gen., Fr. **19**, 469 (1930).

[15] H. Cramér, Actualités Sci. Indust. **736**, 5 (1938).

[16] See H. Touchette, Phys. Rep. **478**, 1 (2009) for a review of the theory of the large deviations, including a demonstration that the central limit theorem serves as an approximation to ignore them.

[17] N. L. Johnson, Biometrika **36**, 149 (1949).

[18] *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1965).

[19] H. B. Callen, *Thermodynamics and an Introduction to Thermostatics*, 2nd ed. (Wiley, New York, 1985).

[20] H. Mouri, A. Hori, Y. Kawashima, and K. Hashimoto, Phys. Fluids **23**, 125110 (2011).

[21] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, Astrophys. J. **304**, 15 (1986).

[22] S. O. Rice, Bell Syst. Tech. J. **23**, 282 (1944); **24**, 46 (1945); reprinted in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York, 1954). This selection also includes papers on the Ornstein-Uhlenbeck process.

[23] See D. T. Gillespie, Am. J. Phys. **64**, 225 (1996) for a review of the mathematics of the Ornstein-Uhlenbeck process.

[24] D. T. Gillespie, Phys. Rev. E **54**, 2084 (1996).

[25] T. Matsumoto and M. Takaoka, Phys. Rev. E **87**, 013008 (2013).

[26] A. N. Kolmogorov, Dokl. Akad. Nauk SSSR **30**, 301 (1941).

[27] L. F. Fenton, IRE Trans. Commun. Syst. **8**, 57 (1960).

[28] S. C. Schwartz and Y. S. Yeh, Bell Syst. Tech. J. **61**, 1441 (1982).

[29] B. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).

[30] See M. Mitzenmacher, Internet Math. **1**, 226 (2004) for a review comparing the log-normal and the power-law distributions in terms of the processes that yield them.

[31] D. G. Champernowne, Econ. J. **63**, 318 (1953).

[32] H. Kesten, Acta Math. **131**, 207 (1973).