Random Multiplicative Processes: An Elementary Tutorial

S. Redner
Center for Polymer Studies and Department of Physics
Boston University, Boston, MA, 02215, USA

Abstract

An elementary discussion of the statistical properties of the *product* of N independent random variables is given. Our motivation is to emphasize the essential differences between the asymptotic $N \to \infty$ behavior of a random product and the asymptotic behavior of a *sum* of random variables – a random additive process. For this latter process, it is widely appreciated that the asymptotic behavior of the sum and its distribution is provided by the central limit theorem. However, no such universal principle exists for a random multiplicative process. In this case, the ratio between the average value of the product, $\langle P \rangle$, and the most probable value, $P_{\rm mp}$, diverges exponentially in N as $N \to \infty$. Within a continuum approximation, the classical log-normal form is often invoked to describe the distribution of the product. We show, however, that the log-normal provides a poor approximation for the asymptotic behavior of the average value, and also for the higher moments of the product. A procedure for computing the correct leading asymptotic behavior of the moments is outlined. The implications of these results for simulations of random multiplicative processes are also discussed. For a such a simulation, the numerically observed "average" value of the product is of the order of $P_{\rm mp}$, and it is only when the simulation is large enough to sample a finite fraction of all the states in the system that a monotonic crossover to the true average value $\langle P \rangle$ occurs. We provide an idealized, but quantitative account for this crossover.

P.A.C.S. Numbers: 02.50.+s, 05.40.+j.

1. Introduction

An important component of an elementary statistical mechanics course is a discussion of the theory of random walks.[1] Usually, an initial treatment is based on a one-dimensional lattice random walk, which is a sequence of equal length displacements whose direction is chosen randomly at each step. One of the basic goals in the study of random walks is to find the average displacement of the probability distribution after a large number of steps N. This example is a realization of a *random additive process*, as the displacement, r, is the *sum* of random steps. For the one-dimensional random walk, the probability distribution for the displacement is the binomial function. In the limit as $N \to \infty$, the central limit theorem[2–4] guarantees that this distribution approaches a Gaussian function, with the $2k^{\text{th}}$ moment of the displacement, $\langle r^{2k} \rangle$, varying as N^k. This limiting Gaussian distribution is a *universal* property of a wide class of stochastic processes, in that details of the underlying random process at the single step level are irrelevant in determining the asymptotic properties of the distribution of displacements.[3] Thus the existence of the central limit theorem provides crucial information for understanding the asymptotic behavior of a random additive process. This universality principle has stimulated extensive study of a variety of physical realizations of random walks, in order to elucidate the limits of applicability of the central limit theorem.[5–7]

In contrast to the well-studied and (relatively) well-understood situation of random additive processes, there is a lesser degree of appreciation of the statistical properties of a *product* of random variables in the physics research literature. This apparent dearth belies the ubiquity of random multiplicative processes. They underlie a diverse range of natural phenomena such as the distribution of incomes, body weights, rainfall, fragment sizes in rock crushing processes, *etc.*[5,8] Random multiplicative processes have also been found to underlie a range of physical processes which fall under the rubric of multifractal phenomena.[9,10] Quite recently, the notions of random multiplicative processes have been applied to diffusive transport in random media.[11] Given the ubiquity and recent interest in random multiplicative processes, it should prove to be fruitful to provide a relatively complete, yet elementary treatment of their statistical properties.

As a specific example, consider a successive process of rock fragmentation in which the size of a fragment evolves according to $x_0 \to x_1 \to x_2 \to \cdots \to x_N$, with a reduction factor at each stage of breakup, $r_n = x_n/x_{n-1} < 1$ that has some well-behaved distribution. The size of a fragment at the N^{th} level is thus given by the *product* of the relative reduction factors, $x_N = (\prod_{k=1}^{N} r_k)x_0$. The primary foal of this paper is to elucidate some of the basic statistical properties of such a *product* of N random variables. It is hoped that the ensuing discussion represents a useful self-contained presentation of the basic features of random multiplicative processes which will fill the apparent gap in the literature.

We shall argue that the behavior of such a product is considerably richer than that of a sum of N random variables. A crucial feature of such a process is that extreme events, although exponentially rare in N, are exponentially different from the typical, or most probable value of the product. Thus it turns out to be necessary to properly account for the extremes in the distribution of the product in order to compute averages correctly.

In the limit of large N, a time-honored approximation for describing the distribution of products is based on noting that the logarithm of the product, $\ln P$, is merely the sum of N random variables, so that $\ln P$ obeys a Gaussian distribution. This leads to the classic *log-normal* form for the distribution of the product.[8] By this construction, however, information about the tail of the distribution has been lost, and these details are crucial in determining the higher moments of the product. We shall show explicitly how the log-normal form fails in providing an accurate description of the statistical properties of the product. Correspondingly, one of our basic results will be to derive an accurate continuum limit for a random multiplicative process from which the correct asymptotic behavior of the higher moments can be obtained. In the context of asymptotic expansions, this continuum limit is a Gaussian function, but one in which the location of the peak depends on the order of the moment being computed.

The fact that the average is dominated by rare events has fundamental implications for numerical studies of systems governed by a random multiplicative process. If one samples only an infinitesimal fraction of the total number of states of the system, as is the case in most realistic situations, then by definition, one will detect the typical value of an observable. As the scale of the simulation is increased, progressively more extreme events become accessible, and the observed average also increases. However, it is only when one has the resources to sample a finite fraction of all the states of the system that the measurement will converge to the true average value of the observable. The quantitative description of this crossover between the most probable value of a random product and its "true" average value is another major goal of this paper.

While many of our basic results are straightforward to derive, it appears that they are not as widely-known as one might expect. We have therefore endeavored to give a pedagogical discussion in what follows.

2. A Binomial Multiplicative Process

To be concrete, consider a binary sequence in which the real, positive numbers z_1 and z_2 appear independently and with probabilities p and q, respectively. Without loss of generality, we take $z_1 > z_2$. If there are N elements in the sequence, we ask what is the average value of this N-fold product, $\langle P \rangle$? In order to compute $\langle P \rangle$, define $p(n)$ to be the probability that the binary product of N independent factors assumes the value $z_1^n z_2^{N-n}$.

This probability is simply the binomial

$$p(n) = \binom{N}{n} p^n q^{N-n}, \tag{1}$$

where $\binom{N}{n} = \frac{N!}{n!\,(N-n)!}$. By averaging over all possible outcomes of the product, one finds the *average* value

$$\langle P \rangle = \sum_{n=0}^{N} \binom{N}{n} p^n q^{N-n} \cdot z_1^n z_2^{N-n} = (pz_1 + qz_2)^N. \tag{2}$$

On the other hand, the most probable event is one in which the product contains Np factors of z_1 and Nq factors of z_2. This is obtained by maximizing the probabilistic factor $\binom{N}{n} p^n q^{N-n}$ with respect to n in Eq. (2). Consequently, the *most probable* value of the product, $P_{\rm mp}$, is simply

$$P_{\rm mp} = (z_1^p z_2^q)^N. \tag{3}$$

While the most probable event yields a good approximation for the average value of the sum in a random additive process, we see that it is inadequate for determining the average value of the product in a random multiplicative process. In fact, the ratio $\langle P \rangle / P_{\rm mp}$ diverges exponentially in N as $N \to \infty$.

Another way to compute $P_{\rm mp}$ is to consider the logarithm of the product. This is a random additive process for which the average value and the most probable value diverge at the same rate as $N \to \infty$. Therefore by computing $\langle \ln P \rangle$, one also obtains the value of $\ln P_{\rm mp}$. Then by re-exponentiating, one has

$$P_{\rm mp} = e^{\langle \ln P \rangle}, \tag{4a}$$

while by definition

$$\langle P \rangle = \langle e^{\ln P} \rangle. \tag{4b}$$

Mathematically, the interchange of exponentiation and averaging is not generally justified. In fact, by expanding Eqs. (4) in power series and using the basic fact[2–4] that $\langle x^k \rangle > \langle x \rangle^k$, it is clear that $\langle P \rangle > P_{\rm mp}$. In general, there is no reason why $P_{\rm mp}$ and $\langle P \rangle$ should have similar values, or even be of the same order of magnitude for a sufficiently broad distribution.

The disparity between $\langle P \rangle$ and $P_{\rm mp}$ can be nicely illustrated by considering several special cases with $z_1 = 2$ and $z_2 = 1/2$: (i) For $p = q = 1/2$, $\langle P \rangle = (5/4)^N = e^{N \ln(5/4)}$, while $P_{\rm mp} = 1$; (ii) For $p = \frac{1}{3}$ and $q = \frac{2}{3}$, the 1/2's are twice as likely to occur as 2's in the sequence of numbers comprising the product. In this case $\langle P \rangle = 1$, while $P_{\rm mp} = ((1/2) \cdot (1/2) \cdot 2)^{N/3} = 2^{-N/3} = e^{-\frac{1}{3} N \ln 2}$; (iii) For p strictly within the range $\frac{1}{3}$ to

$\frac{1}{2}$, one has the curious situation where $\langle P\rangle \sim e^{+\alpha N} \to \infty$ and $P_{\rm mp} \sim e^{-\beta N} \to 0$, where α and β are positive constants, as $N \to \infty$.

The essential reason for the large discrepancy between $\langle P\rangle$ and $P_{\rm mp}$ is the relatively important role played by rare events. For example, a sequence consisting only of N factors of z_1 occurs with an exponentially small probability, but the value of this product is exponentially large compared to the typical value. Consequently, this extreme event makes a finite contribution to $\langle P\rangle$, and a dominant contribution to the higher moments of the product, $\langle P^k\rangle$. From Eq. (2), we see that $\langle P^k\rangle$ reduces to $(pz_1^k)^N$ as $k \to \infty$, *i.e.*, the $k^{\rm th}$ moment is determined solely by the most extreme event. A closely related feature is that the moments obey the inequalities $\langle P^k\rangle (pz_1^k + qz_2^k)^N \gg \langle P\rangle^k = (pz_1 + qz_2)^{Nk}$, and more generally, $\langle P^k\rangle \gg \langle P^{k-1}\rangle^{k/(k-1)}$. These relations also show that there does not exist a unique scale which governs the scaling of all the moments of the product. That is, $\langle P^k\rangle$ cannot be written in the form $\langle P^k\rangle \sim a_k\langle P\rangle^k$, with a_k a non-singular function of k and N. This loss of scaling stems from the long tail in the underlying distribution of products. However, as the order of the moment goes to ∞, the contribution of the single event where the product has the value z_1^N dominates in the value of the moment. In this circumstance, a conventional scaling picture is restored since the value of the moment is essentially determined by a single event.

An additional intriguing feature of the binomial multiplicative process is the sensitivity of $\langle P\rangle$ to short-range correlations in the sequence of variables that are being multiplied. As a simple example, consider the case where $z_1 = 2$, $z_2 = 1/2$, with these two factors occurring with equal probability. Further suppose that there are "no immediate reversals" in the sequence of z_1's and z_2's that comprise the product. That is, when a 2 first appears in the sequence, the next element must also be a 2. Only after the second appearance of a 2 does the sequence become uncorrelated again. For a random walk process with this type of correlation (persistent random walk), it is well-known that the asymptotic properties of the mean displacement are unaffected.[6] However, for a random multiplicative process, this nearest-neighbor correlation is equivalent to replacing the sequence of N correlated variables, which may be either 2 or 1/2, by a sequence of $N/2$ independent variables, which may be either 4 or 1/4. For this new sequence, $\langle P\rangle = (\sqrt{17/8})^N \gg (5/4)^N$ as $N \to \infty$. The increase in $\langle P\rangle$ compared to the original binomial process becomes much more pronounced as the range of correlation in the sequence becomes longer. The origin of this increase stems from the relatively larger role played by "rare" events, *i.e.*, a sequence containing only 2's becomes relatively more likely as the correlation range increases. This simple but remarkable result shows that there is no analog of a central limit theorem for a random multiplicative process, as the mean value depends on correlations among the factors comprising the product.

3. The continuum limit

An important aspect of treating any stochastic process is determining the distribution of relevant observables in the continuum limit. For a product of random variables, the exact distribution function for the discrete system is simply the binomial function given in Eq. (1). From this form, one is naturally led to first apply Stirling's approximation, and then expand the resulting distribution about its maximum to arrive at the classical log-normal form. While this is the time-honored approach for deriving the continuum limit in virtually all stochastic processes, it is an ill-founded approximation for a random multiplicative process. Let us follow this general prescription, however, in order to illustrate the pitfalls associated with the Gaussian approximation. We then present an appropriate expansion procedure that leads to the correct continuum limit.

The $k^{\rm th}$ moment of the product can be approximated by

$$\langle P^k \rangle \simeq \int_{-N}^{+N} e^{\ln p(n)}\, z_1^{nk} z_2^{(N-n)k}\, dn. \tag{5}$$

The Gaussian approximation is based on first applying Stirling's formula to the binomial distribution of Eq. (1) to yield[1]

$$-\ln p(n) \simeq n \ln \frac{n}{N} + (N-n)\ln(1-\frac{n}{N}) - n\ln p - (N-n)\ln q + \frac{1}{2N}\ln(2\pi n(1-\frac{n}{N})). \tag{6}$$

We now expand this approximation for $\ln p(n)$ about its maximum which is located at $n = Np$, with corrections to this condition that vanish in the limit $N \to \infty$. Then by re-exponentiating this expansion, one obtains the Gaussian form

$$p_{\rm Gaussian}(n) \simeq \frac{1}{\sqrt{2\pi Npq}} \exp\left\{ -\frac{(n-Np)^2}{2Npq} \right\}. \tag{7}$$

By writing a generic value of the product as $P = (\frac{z_1}{z_2})^n z_2^N$ and $P_{\rm mp} = (\frac{z_1}{z_2})^{Np} z_2^N$, Eq. (7) can be recast as

$$p_{\rm Gaussian}(n) \simeq \frac{1}{\sqrt{2\pi Npq}} \exp\left\{ -\frac{(\ln P - \ln P_{\rm mp})^2}{2Npq(\ln \frac{z_1}{z_2})^2} \right\}. \tag{8}$$

This is the well-known *log-normal* distribution,[8] whose name reflects the fact that the logarithm of the product is normally distributed.

The $k^{\rm th}$ moment of the product within the Gaussian approximation is, from Eq. (7),

$$\langle P^k \rangle_{\rm Gaussian} \simeq \frac{1}{\sqrt{2\pi Npq}} \int_{-\infty}^{+\infty} e^{-\frac{(n-Np)^2}{2Npq}}\, z_1^{nk} z_2^{(N-n)k}\, dn, \tag{9}$$

and we evaluate this integral by completing the square in the exponential to yield

$$\langle P^k\rangle_{\mathrm{Gaussian}} = \exp\left\{\frac{Npq}{2}k^2(\ln\frac{z_1}{z_2})^2 + N(p-q)k\ln\frac{z_1}{z_2} + \frac{Nk}{2}\ln z_1 z_2\right\}. \qquad (10)$$

It is instructive to compare the values for $\langle P^k\rangle_{\mathrm{Gaussian}}$ with the exact result for $\langle P^k\rangle$ given in Eq. (2). For $z_1 \approx z_2$ and for $p, q \approx \frac{1}{2}$ the agreement between the exact result and the Gaussian approximation is reasonable because the role of extreme events is lessened by having z_1 close to z_2 and $p \approx q$. However, when either of these relations is not satisfied, then the agreement becomes poor (see Table 1). This is especially true for large k because the Gaussian approximation predicts that the k^{th} moment of an N-fold product increases as $e^{\gamma N k^2}$, where $\gamma = \frac{pq}{2}(\ln\frac{z_1}{z_2})^2$, while the exact value increases only as $e^{Nk\ln z_1}$.

What is wrong with the Gaussian approximation? The essential flaw can be seen in comparing Eqs. (5) and (9). In Eq. (5), the k^{th} moment equals the probability of an event, times the k^{th} power of the value of that event, averaged over all events. The k^{th} power of the product is exponentially large in N, so that both factors in the integral are of the same order of magnitude. However in writing the Gaussian approximation of Eq. (9), the binomial distribution has first been expanded about the point where it achieves a maximum. The process of completing the square in the exponent in Eq. (9) is then tantamount to finding the maximum of the function in the exponent in which one portion of this function has already been expanded to second order. This represents an incorrect application of Laplace's method[12] to provide an asymptotic expansion of the integral in Eq. (9).

To see explicitly where such an approach leads, we change variables from n to $x = n/N$ and rewrite Eq. (6) as $\ln p(x) \simeq -Nf(x)$, where

$$f(x) = -N(x\ln x + (1-x)\ln(1-x) - x\ln p - (1-x)\ln q + \frac{1}{2N}\ln(2\pi N x(1-x)). \qquad (11)$$

Eq. (5) then becomes

$$\langle P^k\rangle \simeq \int_{-1}^{+1} e^{-N[f(x)+g(x)]}\, N\, dx, \qquad (12)$$

where $g(x) = kx\ln z_1 + k(1-x)\ln z_2$. In the Gaussian approximation, $f(x)$ is first replaced by $f_{\mathrm{Gaussian}}(x) = f(x_0) + \frac{1}{2}(x-x_0)^2 f''(x_0)$, where $x_0 = p$. The completion of the square in the exponential in Eq. (9) is equivalent to expanding about the maximum of $f_{\mathrm{Gaussian}}(x) + g(x)$. The interested reader can readily verify that this generally is a poor approximation for the location of the peak of the exact exponent function, $h(x) \equiv f(x) + g(x)$.

To perform the integral correctly by the Laplace method, we expand $h(x)$ about its true maximum. Writing

$$\begin{aligned} h(x) = x\ln x + (1-x)\ln(1-x) - x\ln p - (1-x)\ln q + \frac{1}{2N}\ln(2\pi N x(1-x)) \\ - kx\ln(\frac{z_1}{z_2}) - k\ln z_2, \end{aligned} \qquad (13)$$

differentiating with respect to x, and setting this expression to zero, one finds that $h(x)$ has a maximum at

$$x^* \simeq \frac{\xi}{1+\xi}, \qquad \xi = \frac{pz_1^k}{qz_2^k} \tag{14}$$

in the limit $N \to \infty$. Notice that this maximum is always greater than the value $x_{\rm mp} = p$ that corresponds to the most probable value of the product. This shows that the dominant contribution to $\langle P^k \rangle$ generally comes from product values which are larger than $P_{\rm mp}$. At $x = x^*$, straightforward algebra gives

$$h(x^*) = \ln \frac{\xi}{(1+\xi)pz_1^k} + \frac{1}{2N} \ln 2\pi N \frac{\xi}{(1+\xi)^2}, \tag{15a}$$

and

$$h''(x^*) = \frac{(1+\xi)^2}{\xi}. \tag{15b}$$

Therefore the $k^{\rm th}$ moment becomes

$$\begin{aligned} \langle P^k \rangle &\simeq \int_{-1}^{+1} e^{-Nh(x)} N dx \\ &\sim \int_{-1}^{+1} e^{-N[h(x^*)+\frac{1}{2}(x-x^*)^2 h''(x^*)+\ \ldots]} N dx \\ &\sim e^{-Nh(x^*)} \sqrt{\frac{2\pi N}{h''(x^*)}}. \end{aligned} \tag{16}$$

Employing the expressions for $h(x^*)$ and $h''(x^*)$ given in Eqs. (15) yields

$$\langle P^k \rangle \sim (pz_1^k + qz_2^k)^N. \tag{17}$$

Thus the correct value of $\langle P^k \rangle$ is obtained if the continuum limit is formulated appropriately. This formulation only requires that Laplace's method is applied to the *exact* form of the exponent function in Eq. (12). In this circumstance, Laplace's method is guaranteed to give the correct leading asymptotic behavior of the integral.[12] The shortcoming of the log-normal approximation stems from the fact that an approximation to one of the two terms in the exponent function has been made *before* Laplace's method is applied. It is also noteworthy that the correct continuum distribution function cannot be cast as a single function, such as a simple log-normal. From Eqs. (12) and (16) the product of the distribution function times the value of the $k^{\rm th}$ moment can itself be represented by a Gaussian, but one in which the location of the peak depends on k. This emphasizes that there does not exist a unique scale which accounts for all the moments of the product, but rather, each moment is governed by distinct portions of the underlying distribution. It is important to keep these basic features in mind when writing the continuum distribution functions for stochastic processes in which extreme events play a major role.

4. The dependence of the "observed" average value on the sample size

While the true average value of an N-fold product of random variables is governed by extreme events which are at the tail of the distribution, most probable events will tend to dominate in a typical numerical simulation of such a process. That is, most simulations are performed in situations for which one can sample only an infinitesimal fraction of all the states of the system. Therefore the "observed" average value of the product in such a simulation will actually be close to $P_{\rm mp}$. As the number of realizations R in the simulation ensemble is increased, however, there is a corresponding increase in the accessibility to the tails of the distribution. In a given simulation, this increased access will manifest itself in the sporadic appearance of exceptional realizations which will cause the observed averages to fluctuate wildly as a function of R. However for a large enough ensemble, all events are sampled with the correct weight and the fluctuating observations as a function of R will cross over and converge to the true average. We now give a quantitative analysis of this "crossover" within an idealized picture in which the increased access to the tails occurs systematically as a function of R. This approximate treatment predicts a smooth, but sharp increase in the observed average as the number of realizations in the simulation reaches a crossover value $R^* \sim e^{\alpha N}$, where α is a constant. While our analysis is rather crude, to the best of our knowledge, an explicit description of this crossover has not appeared in the literature previously.

A crucial step in our analysis is determining the expected value of the largest product in an ensemble of R realizations of an N-fold product. For this largest product, the number of appearances of the factor z_1 (the larger factor in the product) attains a maximal value n_+ which we will show increases only logarithmically with the number of realizations R. Consequently, R needs to be of the order of e^N, *i.e.*, of the order of all the states in the system, before the extreme events that lead to the correct averages are effectively sampled. To estimate n_+, we employ the Gaussian approximation for the probability distribution of the product, $p(n)$. This will provide an accurate estimate for n_+ as long as R is not too large so that n_+ falls within several standard deviations of the Gaussian peak. This is very likely to be true in any realistic simulation. The case where R approaches 2^N can also be treated within the present framework, but at the expense of introducing unenlightening technical complications.

To determine n_+, we require the probability that in an ensemble of R independent realizations, there exists a single realization for which the expected maximal number of z_1's in an N-fold product is greater than or equal to n_+. In other words, we wish to specify the lower limit for the integral over $p(n)$ such that the probability of having $n \geq n_+$ is equal to $1/R$. That is,

$$\int_{n_+}^{N} p(n)\, dn = \frac{1}{R}. \tag{18}$$

Employing the Gaussian approximation for $p(n)$, Eq. (18) becomes

$$\frac{1}{\sqrt{2\pi Npq}} \int_{n_+}^{\infty} e^{-\frac{(n-Np)^2}{2Npq}}\, dn = \frac{1}{R}. \tag{19}$$

This can be recast as $\frac{1}{R} = \frac{1}{2}\mathrm{erfc}(u_+)$, where $u_+ = \frac{(n-Np)}{\sqrt{2Npq}}$ and $\mathrm{erfc}(x)$ is the error function complement. Finally, using the asymptotic expansion for the error function[13] yields the fundamental result

$$n_+ \simeq Np + \sqrt{2Npq \cdot \ln R}. \tag{20}$$

Correspondingly, there is a minimal number $n_- \simeq Np - \sqrt{2Npq \cdot \ln R}$ which specifies the smallest value of the product in an ensemble of R realizations. Thus n_+ is greater than the typical value of $n = Np$ by an amount that is of the order of $\sqrt{N}$ when the scale of the simulation is small, *i.e.*, typical random walk fluctuations, and n_+ approaches the maximum possible value of N only when the simulation is large enough to sample a finite fraction of all events, *i.e.*, when $R \gtrsim R^* \sim \mathcal{O}(2^N)$. When R reaches R^*, the square root term in Eq. (20) becomes of the order to N, and the most extreme events in the ensemble are now accessible. We emphasize that although this approach predicts that n_+ is a smoothly increasing function of R, n_+ will actually be a sporadically increasing function in a given simulation, and it is only the expected value of n_+ which will conform to Eq. (20).

Due to the predicted systematic dependence of the magnitude of the extreme event on the number of realizations R in a simulation, there will be a corresponding smooth dependence of observed averages on R. Within this formulation, the k^{th} moment of the product in a simulation containing R realizations becomes

$$\begin{aligned} \langle P^k \rangle_R &\simeq \int_{x_-}^{x_+} e^{-Nh(x)}\, N dx \\ &\sim \int_{x_-}^{x_+} e^{-N[h(x^*)+\frac{1}{2}(x-x^*)^2 h''(x^*)+\dots]}\, N dx, \end{aligned} \tag{21}$$

where $x_+ = n_+/N$. The qualitative behavior of $\langle P \rangle_R$ is determined by the relative position of the peak in the integrand compared to the limits of the integral, as illustrated in Fig. 1. For relatively small R, both limits of the integral lie far to the left of the peak, so that $\langle P \rangle_R$ is vanishingly small. As the number of realizations R increases, the upper limit of the integral, x_+, slowly increases and eventually reaches the location of the peak at x^*.

From Eqs. (14) and (20), the number of realizations, R^*, required for x_+ to reach x^* is given by

$$\ln R^* = \frac{N}{2pq}\left(p - \frac{\xi}{1+\xi}\right)^2 = \frac{N}{2pq}\left(p - \frac{pq(\frac{z_1}{z_2})^k}{q + p(\frac{z_1}{z_2})^k}\right)^2. \tag{22}$$

When $R = R^* \simeq e^{\alpha N}$, with α dependent on the details of the multiplicative process, we have integrated over one-half of the contribution to $\langle P^k\rangle_R$ in Eq. (21), and $\langle P\rangle_R \simeq \frac{1}{2}\langle P^k\rangle_\infty$ (see Figs. 1 and 2). For $R > R^*$, x_+ has passed the far side of the peak and $\langle P^k\rangle_R$ very quickly approaches $\langle P^k\rangle_\infty$. Since the width of the peak is of order $1/\sqrt{N}$, $\langle P\rangle_R$ is sharply increasing in the narrow range $x_+ \simeq x^* \pm \mathcal{O}(\frac{1}{\sqrt{N}})$, and this corresponds to $R \simeq R^* \pm \mathcal{O}(e^{\sqrt{N}})$ (Fig. 2). Furthermore, since there is a distinct peak location for each value of k, which is an increasing function of k, progressively more realizations are needed in a simulation to estimate the higher moments to the same degree of accuracy as the low-order moments.

The value of the k^{th} moment of the product in an ensemble containing R realizations can now be obtained by substituting in the expressions for $h(x^*)$ and $h''(x^*)$ given in Eqs. (15) into Eq. (21). This yields,

$$\langle P^k\rangle_R \simeq \frac{1}{\sqrt{\pi}}(pz_1^k + qz_2^k)^N \int_{u_-}^{u_+} e^{-u^2}\, du, \tag{23}$$

where

$$u_\pm = \sqrt{\tfrac{1}{2}Nh''(x^*)}(x_\pm - x^*). \tag{24}$$

We expect that this expression is quantitatively correct when $u_+ > 0$, where we are essentially performing a Laplace expansion of an exponential integral. For $u_+ < 0$, we are integrating a rapidly varying exponential function over a range where there is no extremum in the integrand. In this case the Gaussian approximation is not justified, but it does provide a simple and correct qualitative picture of the behavior of the moments on the number of realizations.

Eq. (23) can be rewritten in terms of error functions, and there are two cases depending on the relative position of the limits with respect to the peak (cf. Fig. 1). We thereby find

$$\langle P^k\rangle_R \simeq \begin{cases} \frac{1}{2}(pz_1^k + qz_2^k)^N[\mathrm{erf}(|u_-|) - \mathrm{erf}(|u_+|)] & \text{if } u_+ \le 0 \ (R < R^*); \\ \\ \frac{1}{2}(pz_1^k + qz_2^k)^N[\mathrm{erf}(u_+) + \mathrm{erf}(|u_-|)] & \text{if } u_+ > 0 \ (R > R^*). \end{cases} \tag{25}$$

Thus the observed value of the k^{th} moment suddenly crosses over from a relatively small number to the true value of the k^{th} moment when R passes through R^*. While a real simulation will not actually exhibit such a smooth behavior, Eq. (25) provides an appealing and quantitative account of the expected nature of the crossover.

5. Conclusions

The statistical properties of the product of N random variables has been outlined. We have shown that the distribution of the product and the behavior of the moments are crucially sensitive to extreme events. Consequently, there is no analog of a central limit theorem, as in the case of random additive processes, in which typical events are sufficient to determine the statistical properties of the sum of a large number of random variables. We have shown explicitly why the log-normal approximation fails to adequately represent the statistical properties of the product in the continuum limit. It is worthwhile to be cognizant of these shortcomings, given the wide range of phenomena for which the log-normal is invoked. We have also provided the correct continuum limit which can be viewed as a log-normal function, but one whose precise form depends on the order of the moment being considered. This emphasizes that there does not exist a unique scale which accounts for all the moments of the product.

We have also discussed how these features would appear in numerical simulations of random multiplicative processes. Numerically observed "averages" are determined by the extreme events that appear in a finite number of realizations of a random product, and we have derived the condition which specifies the nature of these extreme events. When the size of the simulation ensemble is of the order of the total number of possible realizations of the product, then the most extreme events will appear. It is only when this finally happens that a simulation can provide accurate numerical estimates. The logarithmic dependence of the magnitude of the extreme event on the size of the ensemble provides the basis for an idealized, but quantitative account of the crossover to asymptotic behavior. The basic message of this analysis is that numerical estimates from realistic-scale simulations of a random multiplicative process have no relation to true average values. A nice graphics demonstration of these general features has been given by R. L. Blumberg.[14]

Acknowledgements

I wish to thank Profs. B. Shapiro and H. E. Stanley for very useful discussions. I am also grateful to H. E. Stanley for a careful reading of the manuscript and for many helpful suggestions. The Center for Polymer Studies is supported in part by grants from the ARO, NSF, and ONR. This finanical support is greatfully acknowledged.

References

1. See *e.g.*, F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, New York, 1965).
2. J. L. Doob, *Stochastic Processes* (Wiley, New York, 1953).
3. B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables* (Addison-Wesley, Cambridge, MA, 1954).

4. W. Feller, *An Introduction to Probability Theory and its Applications* (Wiley, New York, 1968).
5. E. W. Montroll and M. F. Shlesinger, *The Wonderful World of Random Walks*, *in Nonequilibrium Phenomena II* ed. by J. L. Lebowitz and E. W. Montroll (North-Holland, Amsterdam, 1984)
6. G. H. Weiss and R. J. Rubin, *Random Walks: Theory and Selected Applications*, *Adv. Chem. Phys.* **52**, 363-505 (1983).
7. S. Havlin and D. ben-Avraham, *Diffusion in Disordered Media*, *Adv. Phys.* **36**, 695-798 (1987).
8. See *e.g.,* J. Aitcheson and J. A. C. Brown, *The Lognormal Distribution* (Cambridge University Press, London, England, 1957), and references therein.
9. S. Redner, *Random Multiplicative Processes and Multifractals*, *in Universality in Condensed-Matter Physics* Proceedings of the Les Houches Winter Conference, March 1988 (Springer-Verlag).
10. For a recent review, see *e.g.*, H. E. Stanley and P. Meakin, *Multifractal Phenomena in Physics and Chemistry*, *Nature* **335**, 405-409 (1988).
11. S. Havlin, R. L. Blumberg, M. Schwartz, H. E. Stanley, and A. Bunde, *Random Multiplicative Processes and Transport in Structures with Correlated Spatial Disorder*, *Phys. Rev. Lett.* **61**, 1428-1431 (1988).
12. C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978).
13. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).
14. R. L. Blumberg, private communication; see also Ref. 10.

Table 1. Comparison between the exact value and the Gaussian estimate of the average value of the product of N factors for the case where $p = q = \frac{1}{2}$ and $z_2 = z_1^{-1}$. Since $\langle P \rangle$ and $\langle P \rangle_{\rm Gaussian}$ both grow as μ^N, we write the value of μ to provide a comparison between two numbers which are of the same order of magnitude. The agreement is reasonable for small z_1, but becomes poor for larger values of z_1.

z_1	$\mu_{\rm exact}$	$\mu_{\rm Gaussian}$
2	1.2500	1.2715
3	1.6667	1.8285
4	2.1250	2.6141
5	2.6000	3.6515
10	5.0500	14.167
20	10.025	88.873

Figure 1. Qualitative behavior of the integral in Eq. (21). For small R, the limits of the integral are close together and on the tail of the distribution. As R increases, more of the peak is included within the integration limits, and the value of the integral increases sharply as $R \to R^*$ as shown in Fig. 2.

Figure 2. A sketch of the "observed" value of the k^{th} moment in a simulation, $\langle P^k \rangle_R$, as a function of the logarithm of the number of realizations, R. The sharp increase in $\langle P^k \rangle_R$ moves progressively to the right for larger values of k. As an illustrative example, we have chosen $p = q = \frac{1}{2}$, $z_1 = z_2^{-1} = 4$, $k = 1$, and $N = 100$. The value of $\ln R^* \simeq 38.93$, corresponding to $R^* \sim 8 \times 10^{16}$. Thus to find the moments of a product of only 100 random numbers numerically, a simulation beyond the scope of present-day computation is required!