Scale Invariance and Lack of Self-Averaging in Fragmentation

P. L. Krapivsky*, I. Grosse*†, and E. Ben-Naim‡
*Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215
†Institute for Molecular Biology and Biochemistry, Free University Berlin, Arnimallee 22, 14195 Berlin, Germany
‡Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545

We derive exact statistical properties of a recursive fragmentation process. We show that introducing a fragmentation probability $0 < p < 1$ leads to a purely algebraic size distribution, $P(x) \propto x^{-2p}$, in one dimension. In d dimensions, the volume distribution diverges algebraically in the small fragment limit, $P(V) \sim V^{-\gamma}$, with $\gamma = 2p^{1/d}$. Hence, the entire range of exponents allowed by mass conservation is realized. We demonstrate that this fragmentation process is non-self-averaging as the moments $Y_\alpha = \sum_i x_i^\alpha$ exhibit significant sample to sample fluctuations.

PACS numbers: 05.40.+j, 64.60.Ak, 62.20.Mk

Numerous physical phenomena are characterized by a set of variables, say $\{x_j\}$, which evolve according to a random process, and are subject to the conservation law $\sum_j x_j = \text{const}$. An important example is fragmentation, with applications ranging from geology [1] and fracture [2] to the breakup of liquid droplets [3] and atomic nuclei [4,5]. Other examples include spin glasses [6], where x_j represents the equilibrium probability of finding the system in the $j^{\rm th}$ valley, genetic populations, where x_j is the frequency of the $j^{\rm th}$ allele [7,8], and random Boolean networks [9,10].

In most cases, stochasticity governs both the way the fragments are produced and the number of fragmentation events they experience. For example, in fragmentation of solid objects due to impact with a hard surface fragments may bounce several times before coming to a rest [11]. The typical number of fragmentation events may vary greatly depending on the initial kinetic energy. Another seemingly unrelated example is provided by DNA segmentation algorithms [12], where homogeneous subsequences are produced recursively from an inhomogeneous sequence until a predefined homogeneity level is reached. Here, the number of segmentation events is determined by the degree of homogeneity of the original sequence.

In this study, we examine a fragmentation process with two types of fragments: stable fragments which do not break anymore and unstable fragments. We show that the size distribution is algebraic, and that the entire range of power-laws allowed by the underlying conservation law can be realized by tuning the fragmentation probability. Additionally, these fragmentation processes are characterized by large sample to sample fluctuations, as seen from analysis of the moments of the fragment size distribution.

Specifically, we consider the following *recursive fragmentation process*. We start with the unit interval and choose a break point l in $[0,1]$ with a uniform probability density. Then, with probability p, the interval is divided into two fragments of lengths l and $1-l$, while with probability $q = 1-p$, the interval becomes "frozen" and never fragmented again. If the interval is fragmented, we recursively apply the above fragmentation procedure to both of the resulting fragments.

First, let us examine the average total number of fragments, N. With probability q a single fragment is produced, and with probability p the process is repeated with two fragments. Hence $N = q + 2pN$, yielding

$$N = \begin{cases} q/(1-2p), & \text{if } p < 1/2; \\ \infty, & \text{if } p \geq 1/2. \end{cases} \tag{1}$$

The average total number of fragments becomes infinite at the critical point $p_c = 1/2$, reflecting the critical nature of the underlying branching process [13].

Next, we study $P(x)$, the density of fragments of length x. The recursive nature of the process can be used to obtain the fragment length density

$$P(x) = q\delta(x-1) + 2p\int_x^1 \frac{dy}{y}\, P\left(\frac{x}{y}\right). \tag{2}$$

The second term indicates that a fragment can be created only from a larger fragment, and the y^{-1} kernel reflects the uniform fragmentation density. Eq. (2) can be solved by introducing the Mellin transform

$$M(s) = \int dx\, x^{s-1} P(x). \tag{3}$$

Eqs. (2) and (3) yield $M(s) = q + 2ps^{-1}M(s)$ and as a result

$$M(s) = q\left[1 + \frac{2p}{s-2p}\right]. \tag{4}$$

The average total number of fragments $M(1) = N$ is consistent with Eq. (1), and the total fragment length $M(2) = 1$ is conserved in accord with $1 = \int dx\, xP(x)$. (Here and in the following the integration is carried over the unit interval, i.e., $0 < x < 1$.) The inverse Mellin transform of Eq. (4) gives

$$P(x) = q\left[\delta(x-1) + 2p\, x^{-2p}\right]. \tag{5}$$

Apart from the obvious δ-function, the length density is a purely algebraic function. In particular, the fragment distribution diverges algebraically in the limit of small fragments. Given such an algebraic divergence near the origin, $P(x) \sim x^{-\gamma}$, length conservation restricts the exponent range to $\gamma < 2$. In our case $\gamma = 2p$, and since $0 < p < 1$, the entire range of acceptable exponents emerges by tuning the only control parameter p.

Interestingly, there is no analytic change in the fragment length distribution as the critical point $p_c = \frac{1}{2}$ is passed. Yet, the fragment length distribution becomes independent of the initial interval length at the critical point. Starting from an interval of length L, Eq. (5) can be generalized to yield

$$P(x) = q\delta(x - L) + 2pqL^{2p-1}x^{-2p}. \qquad (6)$$

Thus, the critical point may be detected by observing the point at which the segment distribution becomes independent of the original interval length.

The recursive fragmentation process can be generalized to d dimensions. For instance, in two dimensions we start with the unit square, choose a point (x_1, x_2) with a uniform probability density, and divide, with probability p, the original square into four rectangles of sizes $x_1 \times x_2$, $x_1 \times (1 - x_2)$, $(1 - x_1) \times x_2$, and $(1 - x_1) \times (1 - x_2)$. With probability q, the square becomes frozen and we never again attempt to fragment it. The process is repeated recursively whenever a new fragment is produced.

Let $P(\mathbf{x})$, $\mathbf{x} \equiv (x_1, \ldots, x_d)$, be the probability density of fragments of size $x_1 \times \cdots \times x_d$. This quantity satisfies

$$P(\mathbf{x}) = q\delta(\mathbf{x} - \mathbf{1}) + 2^d p \int \frac{d\mathbf{y}}{y_1 \cdots y_d} P\left(\frac{x_1}{y_1}, \ldots, \frac{x_d}{y_d}\right), \quad (7)$$

with $\int d\mathbf{y} = \int dy_1 \cdots \int dy_d$. Following the steps leading to Eq. (4), we find that the d-dimensional Mellin transform, defined by $M(\mathbf{s}) = \int d\mathbf{x}\, x_1^{s_1-1} \cdots x_d^{s_d-1} P(\mathbf{x})$ with the shorthand notation $\mathbf{s} \equiv (s_1, \ldots, s_d)$ obeys

$$M(\mathbf{s}) = q\left[1 + \frac{\gamma^d}{s_1 \cdots s_d - \gamma^d}\right], \quad \text{with} \quad \gamma = 2p^{1/d}. \quad (8)$$

Eq. (8) gives the total average number of fragments, $N = M(\mathbf{1}) = q/(1 - 2^d p)$ if $p < 2^{-d}$ and $N = \infty$ if $p \geq 2^{-d}$. One can also verify that the total volume $M(\mathbf{2}) = 1$ is conserved. Interestingly, there is an additional infinite set of conserved quantities: all moments whose indices belong to the hyper-surface $s_1^* \cdots s_d^* = 2^d$ satisfy $M(\mathbf{s}^*) = 1$. In a continuous time formulation of this process the same moments were found to be integrals of motion [14–16]. The existence of an infinite number of conservation laws is surprising, because only volume conservation has a clear physical justification.

Next, we study the volume density $P(V)$ defined by

$$P(V) = \int d\mathbf{x}\, P(\mathbf{x})\, \delta\left(V - x_1 \cdots x_d\right). \qquad (9)$$

The Mellin transform, $M(s) = \int dV V^{s-1} P(V)$, can be obtained from Eq. (8) by setting $s_i = s$:

$$M(s) = q\left[1 + \frac{\gamma^d}{s^d - \gamma^d}\right]. \qquad (10)$$

Using the $d^{\rm th}$ root of unity, $\zeta = e^{2\pi i/d}$, and the identity $(s^d - 1)^{-1} = d^{-1} \sum_{0 \leq k \leq d-1} \zeta^k/(s - \zeta^k)$, $M(s)$ can be expressed as a sum over simple poles at $\gamma\zeta^k$. Consequently, the inverse Mellin transform is given by a linear combination of d power laws:

$$P(V) = q\left[\delta(V - 1) + \frac{\gamma}{d}\sum_{k=0}^{d-1} \zeta^k V^{-\gamma\zeta^k}\right]. \qquad (11)$$

One can verify that this expression equals its complex conjugate, and hence, it is real. Additionally, the one-dimensional case (5) is recovered by setting $d = 1$.

The small-volume tail of the distribution can be obtained by noting that the sum in Eq. (11) is dominated by the first term in the series, which leads to

$$P(V) \simeq \frac{\gamma q}{d} V^{-\gamma} \qquad \text{as} \qquad V \to 0. \qquad (12)$$

Although the value of the exponent $\gamma = 2p^{1/d}$ changes with the dimension d, the possible range of exponents for this process remains the same since $0 < 2p^{1/d} < 2$ when $0 < p < 1$. In the infinite dimension limit, the volume density becomes universal: $P(V) \sim V^{-2}$.

The leading behavior of $P(V)$ in the large size limit can be derived by using the Taylor expansion and the identity $\sum_{k=0}^{d-1} \zeta^{kn} = \delta_{n,0}$ for $n = 0, \ldots, d - 1$. One finds that in higher dimensions the volume distribution vanishes algebraically near its maximum value,

$$P(V) \simeq \frac{\gamma^d}{(d-1)!}(1 - V)^{d-1} \qquad \text{as} \qquad V \to 1. \quad (13)$$

In fact, the entire multivariate fragment length density can be also obtained explicitly. This can be achieved by expanding the geometric series,

$$\frac{\gamma^d}{s_1 \cdots s_d - \gamma^d} = \sum_{n \geq 0} \prod_{i=1}^{d} \left(\frac{\gamma}{s_i}\right)^{n+1},$$

and performing the inverse Mellin transform for each variable separately. Using the identity $\int dx\, x^{s-1}\left(\ln \frac{1}{x}\right)^n = n! s^{-n-1}$ gives

$$P(\mathbf{x}) = q\left[\delta(\mathbf{x} - \mathbf{1}) + \gamma^d F_d(z)\right], \qquad (14)$$

with the shorthand notations

$$F_d(z) = \sum_{n=0}^{\infty} \left(\frac{z^n}{n!}\right)^d \qquad \text{and} \qquad z = \gamma\left(\prod_{i=1}^{d} \ln \frac{1}{x_i}\right)^{1/d}. \quad (15)$$

In two dimensions, $F_2(z) = I_0(2z)$ where I_0 is the modified Bessel function.

The small size behavior of $P(\mathbf{x})$ can be obtained by using the steepest decent method. The leading tail behavior, $F_d(z) \simeq (2\pi z)^{\frac{1-d}{2}} e^{zd}$ for $z \gg 1$, corresponds to the case when at least one of the lengths is small, i. e. $x_i \ll 1$. Returning to the original variables we see that the fragment distribution exhibits a "log-stretched-exponential" behavior

$$P(\mathbf{x}) \sim \left[\prod_{i=1}^{d} \ln \frac{1}{x_i}\right]^{-\frac{d-1}{2d}} \exp\left[d\gamma \left(\prod_{i=1}^{d} \ln \frac{1}{x_i}\right)^{1/d}\right]. \quad (16)$$

The fragment size distribution represents an average over infinitely many realizations of the fragmentation process, and hence, it does not capture sample to sample fluctuations. These fluctuations are important in non-self-averaging systems, where they do not vanish in the thermodynamic limit. Useful quantities for characterizing such fluctuations are the moments [17,18]

$$Y_\alpha = \sum_i x_i^\alpha, \quad (17)$$

where the sum runs over all fragments.

We are interested in the average values $\langle Y_\alpha \rangle$ and $\langle Y_\alpha Y_\beta \rangle$. For integer α, $\langle Y_\alpha \rangle$ is the probability that α points randomly chosen in the unit interval belong to the same fragment. The expected value of Y_α satisfies

$$\langle Y_\alpha \rangle = q + p\langle Y_\alpha \rangle \int dy\, [y^\alpha + (1-y)^\alpha]. \quad (18)$$

The first term corresponds to the case where the unit interval is not fragmented, and the second term describes the situation when at least one fragmentation event has occurred. Eq. (18) gives

$$\langle Y_\alpha \rangle = q\left[1 + \frac{2p}{\alpha + 1 - 2p}\right] \quad (19)$$

if $\alpha > 2p-1$, and $\langle Y_\alpha \rangle = \infty$ if $\alpha \le 2p-1$. As expected, Eq. (19) agrees with the moments of $P(x)$ obtained by integrating Eq. (5), $\langle Y_\alpha \rangle = \int dx\, x^\alpha P(x)$.

However, higher order averages such as $\langle Y_\alpha Y_\beta \rangle$ do not follow directly from the fragment size density. For integer α and β, $\langle Y_\alpha Y_\beta \rangle$ is the probability that, if $\alpha+\beta$ points are chosen at random, the first α points all lie on the same fragment, and the last β points all lie on another (possibly the same) fragment. This quantity satisfies

$$\begin{aligned}\langle Y_\alpha Y_\beta \rangle &= q + p\langle Y_\alpha Y_\beta \rangle \int dy\, \left[y^{\alpha+\beta} + (1-y)^{\alpha+\beta}\right] \quad (20)\\ &+ p\langle Y_\alpha \rangle \langle Y_\beta \rangle \int dy\, \left[y^\alpha (1-y)^\beta + (1-y)^\alpha y^\beta\right],\end{aligned}$$

which yields

$$\begin{aligned}\langle Y_\alpha Y_\beta \rangle &= q + \frac{2pq}{\alpha+\beta+1-2p} \quad (21)\\ &+ 2p\,\frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+1)}\,\frac{\langle Y_\alpha \rangle \langle Y_\beta \rangle}{\alpha+\beta+1-2p}\end{aligned}$$

if $\alpha, \beta, \alpha+\beta > 2p-1$, and $\langle Y_\alpha Y_\beta \rangle = \infty$ otherwise.

Eq. (21) shows that $\langle Y_\alpha Y_\beta \rangle \neq \langle Y_\alpha \rangle \langle Y_\beta \rangle$, and in particular, $\langle Y_\alpha^2 \rangle \neq \langle Y_\alpha \rangle^2$. Therefore, fluctuations in Y_α are significant and the recursive fragmentation process is non-self-averaging. Hence, statistical properties obtained by averaging over all realizations are insufficient to probe sample to sample fluctuations. This behavior is reminiscent of the lack of self-averaging found in fragmentation processes that exhibit a shattering transition [19].

In principle, higher order averages such as $\langle Y_\alpha^n \rangle$ can be calculated recursively by the procedure outlined above. The resulting expressions are cumbersome and not terribly illuminating. Instead, one may study the distribution $Q_\alpha(Y)$ which obeys

$$\begin{aligned}Q_\alpha(Y_\alpha) &= q\delta(Y_\alpha - 1) \quad (22)\\ &+p \int dl \int_0^{Y_\alpha} dZ\, \frac{1}{l^\alpha}\, Q_\alpha\left(\frac{Z}{l^\alpha}\right) \frac{1}{(1-l)^\alpha}\, Q_\alpha\left(\frac{Y_\alpha - Z}{(1-l)^\alpha}\right).\end{aligned}$$

In addition to the recursive nature of the process, we have employed extensivity, i.e., $\langle Y_\alpha \rangle \propto L^\alpha$ in an interval of length L.

Clearly, $Y_0 = N$ and $Y_1 = 1$, and therefore, $Q_1(Y_1) = \delta(Y_1 - 1)$ and $Q_0(N)$ can also be determined analytically as well. Generally, different behaviors emerge for $\alpha > 1$ and $\alpha < 1$. We concentrate on the former case where the support of the distribution $Q_\alpha(Y)$ is the interval [0,1]. The Laplace transform, $R_\alpha(\lambda) = \int_0^1 dY_\alpha\, e^{-\lambda Y_\alpha} Q_\alpha(Y_\alpha)$, obeys

$$R_\alpha(\lambda) = q\, e^{-\lambda} + p\int_0^1 dl\, R_\alpha\left[\lambda l^\alpha\right] R_\alpha\left[\lambda (1-l)^\alpha\right]. \quad (23)$$

The behavior of $Q_\alpha(Y_\alpha)$ in the limit $Y_\alpha \to 0$ is reflected by the asymptotics of $R_\alpha(\lambda)$ as $\lambda \to \infty$. Substituting $R_\alpha(\lambda) \sim \exp(-A\lambda^\beta)$ into both sides of Eq. (23), evaluating the integral using the method of steepest decent, and equating the left and right hand sides gives $\beta = 1/\alpha$. Consequently, we find that the distribution has an essential singularity near the origin,

$$Q_\alpha(Y_\alpha) \sim \exp\left(-B Y_\alpha^{-\frac{1}{\alpha-1}}\right) \qquad \text{as} \qquad Y_\alpha \to 0. \quad (24)$$

Extremal properties can be viewed as an additional probe of sample to sample fluctuations. Thus, we consider the length density of the largest fragment, $\mathcal{L}(x)$. Clearly, $\mathcal{L}(x) = P(x)$ for $x \ge 1/2$, i.e.,

$$\mathcal{L}(x) = q\delta(x-1) + 2pq\, x^{-2p} \qquad \text{for} \quad x \ge 1/2. \quad (25)$$

In the complementary case of $x < 1/2$, $\mathcal{L}(x)$ satisfies

$$\begin{aligned}\mathcal{L}(x) &= 2qp^2 \mathcal{L}_-\left(\frac{x}{1-x}\right) + 2p^2 \int_{1-x}^{1} \frac{dy}{y}\, \mathcal{L}\left(\frac{x}{y}\right) \quad (26)\\ &+ 2p^3 \int_x^{1-x} \frac{dy}{y}\, \mathcal{L}\left(\frac{x}{y}\right) \mathcal{L}_-\left(\frac{x}{1-y}\right),\end{aligned}$$

where $\mathcal{L}_-(u) = \int_0^u dv\, \mathcal{L}(v)$. The first term on the right-hand side of Eq. (26) describes the situation when the unit interval was fragmented into two intervals of lengths x and $1-x$, with stable smaller fragment and unstable longer fragment (hence the factor qp^2). The latter $\mathcal{L}_-$ factor guarantees that subsequent fragmentation of the unstable interval does not lead to a fragment longer than x. If one of the 1st generation fragments is shorter that x, then only the longest 1st generation fragment contributes; this leads to the second term on the right-hand side of Eq. (26). The next term describes the situation when both 1st generation fragments are longer than x, so the longest fragment can arise out of breaking any of the two fragments. The factor $\mathcal{L}_-$ guarantees that the longest fragment of length x comes from the corresponding 1st generation fragment, and the factor p^3 guarantees that both 1st generation fragments remain unstable. Since $\mathcal{L}(x)$ obeys different equations in different regions, it looses analyticity on the boundaries. Namely, $\mathcal{L}(x)$ possesses an infinite set of singularities at $x = 1/k$ which become weaker as k increases. Similar singularities underly extremal properties of a number of random processes, including random walks, spin glasses, random maps, and random trees [7,8,17,18,20].

For completeness, we note that the above results extend to a complementary class of fragmentation processes where first fragmentation occurs, and then fragments remain unstable with probability p. In this case, the δ function drops, and the the size distribution becomes purely algebraic, $P(x) = 2qx^{-2p}$.

In summary, we have found that recursive fragmentation is scale free, i.e., the fragment length distribution is purely algebraic. In higher dimensions, the volume distribution is a linear combination of d power laws, and consequently, an algebraic divergence characterizes the small-fragment tail of the distribution. A number of recent impact fragmentation experiments reported algebraic mass distributions with the corresponding exponents ranging from 1 to 2 [11]. It will be interesting to further examine whether our simplified model is suitable for describing fragmentation of solid objects.

We have also found that the recursive fragmentation process exhibits a number of features that arise in other complex and disordered systems, such as non-self-averaging behavior and the existence of an infinite number of singularities in the distribution of the largest fragment. These features indicate that large sample to sample fluctuations exist, and that knowledge of first order averages may not be sufficient for characterizing the system. Our 1D model is equivalent to applying the aforementioned DNA segmentation algorithm to a random sequence. It will be interesting to study self-averaging and extremal properties of genetic sequences, which are known to have commonalities with disordered systems. Indeed, if these subtle features are found for genetic sequences as well, this would suggest that much caution should be exercised in statistical analysis of DNA.

We thank S. Redner and O. Weiss for fruitful discussions, and DOE, NSF, ARO, NIH, and DFG for support.

[1] D. L. Turcotte, J. Geophys. Res. **91**, 1921 (1986).
[2] B. R. Lawn and T. R. Wilshaw, *Fracture of Brittle Solids* (Cambridge University Press, Cambridge, 1975).
[3] R. Shinnar, J. Fluid Mech. **10**, 259 (1961).
[4] K. C. Chase, P. Bhattacharyya, and A. Z. Mekjian, Phys. Rev. C **57**, 822 (1998).
[5] For a general review of fragmentation, see e.g. S. Redner, in: *Statistical Models for the Fracture of Disordered Media*, ed. H. J. Herrmann and S. Roux (Elsevier Science, New York, 1990).
[6] M. Mézard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
[7] P. G. Higgs, Phys. Rev. E **51**, 95 (1995).
[8] B. Derrida and B. Jung-Muller, J. Stat. Phys. **94**, 277 (1999).
[9] S. A. Kauffman, *The Origin of Order: Self-Organization and Selection in Evolution* (Oxford University Press, London, 1993).
[10] H. Flyvbjerg and N. J. Kjaer, J. Phys. A **21**, 1695 (1988).
[11] T. Ishii and M. Matsushita, J. Phys. Soc. Jap. **61**, 3474 (1992); L. Oddershede, P. Dimon, and J. Bohr, Phys. Rev. Lett. **71**, 3107 (1993); T. Kadono, Phys. Rev. Lett. **78**, 1444 (1997).
[12] P. Bernaola-Galván, R. Román-Roldán, and J. L. Oliver, Phys. Rev. E **53**, 5181 (1996); Phys. Rev. Lett. **80**, 1344 (1998); W. Li *et al.*, Genome Research **8**, 916 (1998).
[13] T. E. Harris, *The theory of branching processes* (Dover, New York, 1989).
[14] P. L. Krapivsky and E. Ben-Naim, Phys. Rev. E **50**, 3502 (1994); Phys. Lett. A **196**, 168 (1994); E. Ben-Naim and P. L. Krapivsky, Phys. Rev. Lett. **76**, 3234 (1996).
[15] G. J. Rodgers and M. K. Hassan, Phys. Rev. E **50**, 3459 (1994).
[16] D. Boyer, G. Tarjus, and P. Viot, Phys. Rev. E **51**, 1043 (1995).
[17] For a review of non-self-averaging phenomena, see B. Derrida, Physica D **107**, 186 (1997).
[18] B. Derrida and H. Flyvbjerg, J. Phys. A **20**, 5273 (1987); J. Physique **48**, 971 (1987).
[19] D. L. Maslov, Phys. Rev. Lett. **71** 1268 (1993).
[20] L. Frachebourg, I. Ispolatov, and P. L. Krapivsky, Phys. Rev. E **52**, R5727 (1995).