Global Fluctuations and Gumbel Statistics

Eric Bertin

Department of Theoretical Physics, University of Geneva, CH-1211 Geneva 4, Switzerland

(Received 7 June 2005; published 18 October 2005)

We explain how the statistics of global observables in correlated systems can be related to extreme value problems and to Gumbel statistics. This relationship then naturally leads to the emergence of the generalized Gumbel distribution $G_a(x)$, with a real index a, in the study of global fluctuations. To illustrate these findings, we introduce an exactly solvable nonequilibrium model describing an energy flux on a lattice, with local dissipation, in which the fluctuations of the global energy are precisely described by the generalized Gumbel distribution.

DOI: 10.1103/PhysRevLett.95.170601 PACS numbers: 05.40.−a, 02.50.−r, 05.70.−a

The ubiquitous appearance of asymmetric distributions in the study of fluctuations of global quantities in correlated systems has raised a lot of interest in recent years. Such non-Gaussian distributions, characterized by an exponential tail on one side and a rapid falloff on the other side, have been observed in many models or experimental systems, in the context of turbulence [1–5], equilibrium critical systems [1,6–9], nonequilibrium models exhibiting self-organized criticality [6,10], interface models [11], $1/f$ noise [12], Langevin equations [13], granular gas models [3,14], or even the statistics of the level of the Danube river [15]. Quite strikingly, this analogy is not only qualitative, but also many of the distributions observed in these very different systems actually fall [1,6,9,14,15], once suitably rescaled, close to the so-called Bramwell-Holdsworth-Pinton (BHP) distribution describing the magnetization of the XY model in the low temperature limit, as well as the roughness of the two-dimensional Edwards-Wilkinson surface model [7,16]. Yet, as not all data collapse onto the BHP curve [4,8], a more general distribution has been proposed to describe the data, namely, the generalized Gumbel distribution $G_a(x)$, which includes a continuous shape parameter a [4,6–8,14,15,17,18] (see below for a definition). Interestingly, this distribution, with $a = 1$, turns out to be the exact one for periodic Gaussian $1/f$ noise [12]; it is also very close to the BHP one for $a \approx \pi/2$ [6,14].

The distribution $G_a(x)$ originates, for integer values of a, from the study of extreme value statistics [19,20], and describes the fluctuations of the ath largest value in a large set of identically distributed (independent) random variables z_i [21]. Accordingly, there is no obvious theoretical motivation for the use of the distribution $G_a(x)$ in the study of fluctuations of global quantities. Rather, it is usually considered as a convenient fitting function, and a theoretical understanding of its relevance is still lacking. Indeed, the question of the underlying role of extreme values in correlated systems has been repeatedly asked in the literature [9,10,12,17,22]. Still, attempts to identify an extremal process dominating the dynamics of such systems have failed up to now [9,17]. All the above body of results thus leads to the following questions: First, what is the precise relationship (if any) between global fluctuations in correlated systems and extreme value statistics? Second, could one find a simple physical model for which the fluctuations of a global quantity would be exactly described by a generalized Gumbel distribution?

Global fluctuations in complex correlated systems are often hard to tackle analytically precisely due to strong correlations between local microscopic variables. Yet, in some cases, statistically independent collective variables—such as Fourier modes [7,12,23,24]—can be defined, so that a problem of correlated random variables may be converted into a problem of independent random variables, with nonidentical distributions—otherwise the central limit theorem would hold.

In this Letter, we explain how the statistics of global quantities, expressed as sums of nonidentically distributed random variables, is related to extreme value problems, and how the generalized Gumbel distribution $G_a(x)$, with a real index a, emerges in the study of global fluctuations. Interestingly, it turns out that such a relationship does not rely on an extremal process hidden in the dynamics of global variables, contrary to usual conjectures. These results are illustrated on a nonequilibrium cascade model in which the fluctuations of the total energy are exactly described by the generalized Gumbel distribution $G_a(x)$, where a depends continuously on the microscopic parameters of the model.

Our starting point is the observation [12] that the integrated power spectrum w of periodic Gaussian $1/f$ noise is distributed, after a suitable rescaling $x = (w - \langle w \rangle)/\sigma_w$ (where σ_w^2 is the variance of w) according to the Gumbel distribution $G_1(x)$. The model for $1/f$ noise used in [12] consists in a large number N of statistically independent Gaussian Fourier modes with complex amplitudes $c_n = c_{-n}^*$. Introducing $y_n \equiv |c_n|^2 + |c_{-n}|^2$, one has by definition $w = \sum_{n=1}^{N} y_n$. The distribution of y_n reads

$$p_n(y_n) = n\kappa e^{-n\kappa y_n}, \tag{1}$$

so that w is simply the sum of N independent random variables y_n, with nonidentical exponential distributions.

Note that the appearance of a non-Gaussian distribution is not surprising in itself, since the sum of the variances of the y_n's converges when $N \to \infty$ [as $\mathrm{var}(y_n) = 1/(n^2\kappa^2)$], so that the central limit theorem is not expected to hold (for a detailed discussion on this point, see [25]). Still, the fact that a Gumbel distribution precisely emerges may suggest the existence of some "hidden" extremal processes dominating the fluctuations of w, but no clear evidence for such processes has been found yet [9,17].

Actually, a different perspective may be necessary to understand the relationship between the two problems. Indeed, instead of looking for extremal processes hidden in sums of random variables, one may look for sums of random variables with decreasing amplitudes when studying the statistics of extreme values. To this aim, we introduce the following procedure, illustrated in Fig. 1. Considering a set of N random variables $z_n > 0$ ($1 \le n \le N$), all drawn from the same distribution $P(z)$, we introduce the variables z'_n defined by ordering the original variables z_n: $z'_n = z_{\sigma(n)}$, where $\sigma(n)$ is a permutation over the interval $[1, N]$ such that $z'_1 \ge z'_2 \ge \cdots \ge z'_N$. Thus z'_1 is simply the maximum value of the set $\{z_n\}$. We also define the interval y_n between z'_{n+1} and z'_n:

$$y_n = z'_n - z'_{n+1} \qquad (1 \le n \le N-1); \qquad y_N = z'_N. \tag{2}$$

With these notations, one can write

$$\max_{1\le n\le N}(z_n) \equiv z'_1 = \sum_{n=1}^{N} y_n. \tag{3}$$

As a result, a problem of extreme value can be mapped onto a problem of the sum of random variables. Still, it should be noticed that although the original variables z_n are independent, the y_n's are *a priori* correlated.

In the following, we show that in the specific case where $P(z)$ is an exponential distribution $P(z) = \kappa e^{-\kappa z}$, the y_n's prove independent and distributed according to Eq. (1). The distribution $P_N(\{y_n\})$ reads

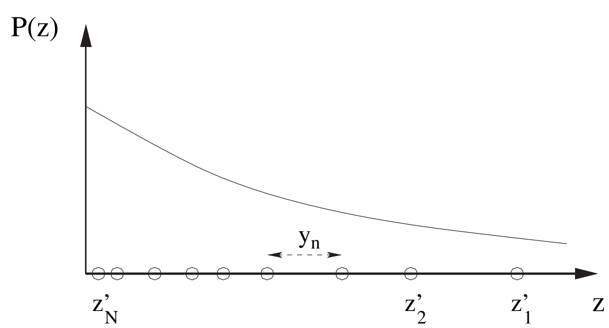

FIG. 1. Sketch of the notations used in the text. N random values of z are drawn according to the probability density $P(z)$. These values are relabeled into $z'_1 \ge \cdots \ge z'_N$, and the interval between two successive z'_n is denoted by y_n.

$$P_N(\{y_n\}) = \kappa^N N! \int_0^\infty dz_N e^{-\kappa z_N} \cdots \int_{z_2}^\infty dz_1 e^{-\kappa z_1} \delta(y_N - z_N) \times \prod_{n=1}^{N-1} \delta[y_n - (z_n - z_{n+1})], \tag{4}$$

where the integral over z_n is from z_{n+1} to ∞, for $1 \le n \le N-1$. This expression can be understood as follows: either the variables $\{z_n\}$ are already ordered, which straightforwardly gives the above integrals, or they are not, and then can be ordered through a permutation, which leads to the $N!$ factor in front. Making the change of variables $v_n = z_n - z_{n+1}$ ($1 \le n \le N-1$) and $v_N = z_N$ in Eq. (4), the different integrals factorize and one finds

$$P_N(\{y_n\}) = \prod_{n=1}^{N} n\kappa e^{-n\kappa y_n}. \tag{5}$$

Thus it turns out that in the specific case $P(z) = \kappa e^{-\kappa z}$ the y_n's are independent variables, distributed as the squared Fourier amplitudes in the $1/f$ noise model, i.e., according to Eq. (1). But as the sum of the y_n's is precisely the maximum value of a set of exponentially distributed variables $\{z_n\}$, we know that this sum has to be distributed (after a suitable rescaling) according to $G_1(x)$, so that one recovers immediately the results of [12]. Accordingly, a clear relationship appears between the statistics of extreme values and that of the sums of variables with decreasing amplitudes. This relationship can actually be understood at two different levels. On the one hand, starting from a set of (possibly correlated) variables $\{z_n\}$, one can always define the interval y_n between two successive variables z'_n obtained by ordering the set $\{z_n\}$ [see Eq. (3)]. Thus, on general grounds, the maximum value of correlated variables z_n can be formally written as a sum of correlated variables y_n, but the corresponding extreme value distribution is usually unknown. On the other hand, it seems that the maximum value of a set $\{z_n\}$ of *independent* variables is related to a sum of *independent* variables $\{y_n\}$ only in the case where the z_n's are drawn from an exponential distribution, leading to the Gumbel distribution $G_1(x)$. Indeed, the factorization property of the exponential is essential to derive Eq. (5) from Eq. (4).

The above result leads to some rather unexpected consequences. From the very definition of the variables $\{z'_n\}$, z'_k is precisely the kth largest value of the original set $\{z_n\}$. So we know that z'_k follows, once rescaled as $x = (z'_k - \langle z'_k \rangle)/\sigma_k$ with $\sigma_k^2 = \mathrm{var}(z'_k)$, the generalized Gumbel distribution $G_k(x)$ [19]. The distribution $G_a(x)$ is defined for any positive real value a by

$$G_a(x) = \frac{\theta_a a^a}{\Gamma(a)} \exp\{-a[\theta_a(x + \nu_a) + e^{-\theta_a(x+\nu_a)}]\}, \tag{6}$$

with

$$\theta_a^2 = \frac{d^2 \ln\Gamma}{da^2}, \qquad \nu_a = \frac{1}{\theta_a}\left(\ln a - \frac{d \ln\Gamma}{da}\right), \tag{7}$$

where $\Gamma(a)$ is the Euler Gamma function. Besides, z'_k may also be expressed as a sum:

$$z'_k = \sum_{n=k}^{N} y_n = \sum_{n=1}^{N-k+1} \tilde{y}_n, \tag{8}$$

with $\tilde{y}_n \equiv y_{n+k-1}$ distributed according to

$$p_{n,k}(\tilde{y}_n) = (n+k-1)\kappa e^{-(n+k-1)\kappa\tilde{y}_n}. \tag{9}$$

Thus the sum of independent random variables drawn from (9) is distributed, after a suitable rescaling, according to $G_k(x)$ in the limit $N \to \infty$. But then, one can forget the original extreme value problem, and consider only the statistics of the sum, so that there is no more reason to restrict k to be integer. Since the generalized Gumbel distribution is obtained for integer k, it seems plausible that it also holds for real values $k = a > 0$. To be more specific, considering independent variables u_n with distribution

$$p_{n,a}(u_n) = (n+a-1)\kappa e^{-(n+a-1)\kappa u_n}, \quad 1 \le n \le N, \tag{10}$$

the sum $X = \sum_{n=1}^{N} u_n$ is precisely distributed, in the limit $N \to \infty$, according to the generalized Gumbel distribution $G_a(x)$, where $x = (X - \langle X \rangle)/\sigma_X$. This result, suggested by the above argument, can be derived exactly without reference to the extreme value problem [26].

We now illustrate the above result on a simple nonequilibrium stochastic model, which is defined by the following rules [27]. On each site $n = 1, \ldots, N$ of a one-dimensional lattice, a positive continuous variable ρ_n—to be thought of as an energy—is introduced. The (asynchronous) dynamics is defined through three different physical mechanisms involving energy, namely, injection on, say, the left boundary, transport from one site to its right neighbor, and local dissipation. More precisely, an amount of energy between μ and $\mu + d\mu$ can be either injected on the leftmost site $n = 1$ with a rate (probability per unit time) $J(\mu)d\mu$, transferred from site n to site $n+1$ with rate $\phi(\mu)d\mu$, or removed (i.e., dissipated) from site n with rate $\Delta(\mu)d\mu$—see Fig. 2. On the rightmost site $n = N$, the transferred energy is actually dissipated. Note that the above rates do not depend on the values of the local energies ρ_n, apart from the obvious constraint that one cannot withdraw from site n (either for transport or dissipation) an energy μ greater than ρ_n.

The master equation governing the evolution of the probability distribution $P(\{\rho_n\}, t)$ reads

$$\frac{\partial P}{\partial t} = \int_0^{\rho_1} d\mu J(\mu) P(\{\rho_1 - \mu, \rho_j\}, t) - \int_0^{\infty} d\mu J(\mu) P(\{\rho_j\}, t) + \sum_{n=1}^{N-1} \int_0^{\rho_{n+1}} d\mu \phi(\mu) P(\{\rho_n + \mu, \rho_{n+1} - \mu, \rho_j\}, t)$$
$$+ \sum_{n=1}^{N} \int_0^{\infty} d\mu [\Delta(\mu) + \phi(\mu)\delta_{n,N}] P(\{\rho_n + \mu, \rho_j\}, t) - \sum_{n=1}^{N} \int_0^{\rho_n} d\mu [\phi(\mu) + \Delta(\mu)] P(\{\rho_j\}, t), \tag{11}$$

where ρ_j generically stands for all the variables that are not affected by μ. In the following, we focus on the specific case where $J(\mu) = e^{-\beta\mu}\phi(\mu)$ and $\Delta(\mu) = (e^{\lambda\mu} - 1)\phi(\mu)$, introducing two positive parameters β and λ. With these assumptions, the steady-state distribution $P(\{\rho_n\})$ proves factorized and can be computed exactly [28]; it turns out to be precisely the same as Eq. (10):

$$P(\{\rho_n\}) = \prod_{n=1}^{N} (\lambda n + \beta) e^{-(\lambda n + \beta)\rho_n}, \tag{12}$$

with the identification $\lambda = \kappa$ and $\beta = (a-1)\kappa$—note that $P(\{\rho_n\})$ does not depend on the specific form of $\phi(\mu)$. As a result, the fluctuations of the total energy $E = \sum_{n=1}^{N} \rho_n$ are described in the infinite N limit—after rescaling E to ensure zero mean and unit variance—by the generalized Gumbel distribution $G_a(x)$, with $a = 1 + \beta/\lambda$ (see Fig. 3). Interestingly, in the limit of low dissipation $\lambda \to 0$, one recovers a Gaussian distribution, since $G_a(x)$ converges to a Gaussian for $a \to \infty$. Qualitatively, the parameter a may be thought of as the number of sites having roughly the same energy, of the order of $1/\beta$.

Finally, we note that the "cascade" mechanism illustrated by the above stochastic model should be considered as one possible mechanism, but perhaps not as the unique one. Indeed, in some systems such as freely evolving granular gases [14], Gumbel distributions are indeed observed even though the global quantity of interest cannot be

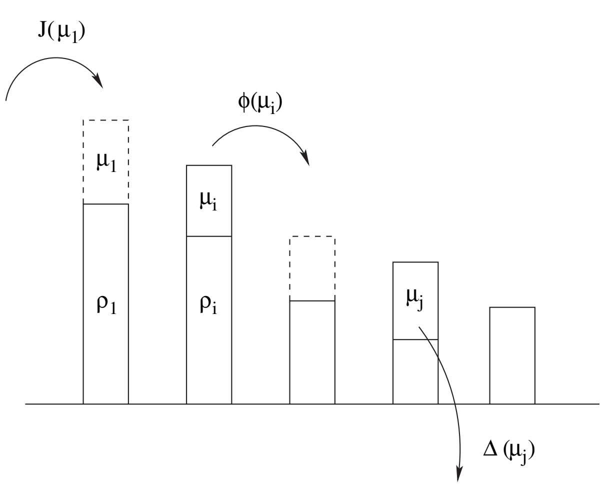

FIG. 2. Schematic view of the model, with the three different mechanisms: injection on the leftmost site with rate $J(\mu_1)$, transport from site i to $i+1$ with rate $\phi(\mu_i)$, and dissipation on site j with rate $\Delta(\mu_j)$.

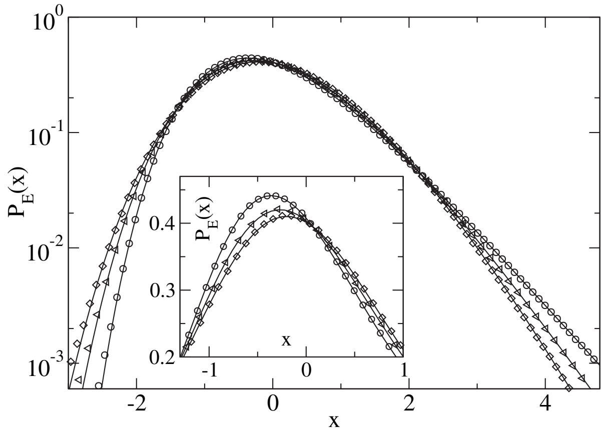

FIG. 3. Distribution $P_E(x)$ of the rescaled (global) energy of the model, $x = (E - \langle E \rangle)/\sigma_E$. The analytical result $G_a(x)$ (solid lines) is compared with numerical simulations with $\phi(\mu) = 1$, for $a \equiv 1 + \beta/\lambda = 1.7$ (○), 3.3 (◁), and 6 (◇), showing excellent agreement. Inset: zoom on the top of the curves, on a linear scale.

written in an obvious way as a sum of independent collective variables. Yet, it must be noticed that such a granular system does not reach a steady state since no energy is injected; fluctuations are then measured in a scaling regime where the average kinetic energy continuously decreases. Accordingly, one might expect another physical mechanism to be at play in this case.

In summary, we have shown that the generalized Gumbel distribution $G_a(x)$ appearing in numerous experimental and numerical studies should not be interpreted as a signature of some hidden extremal process, but on the contrary, as the distribution associated to an infinite sum of independent and exponentially distributed random variables u_n ($n \geq 1$), with mean value $[(n + a - 1)\kappa]^{-1}$. If a is the integer, the variables u_n can be interpreted as the intervals y_n between two successive (ordered) random values drawn from an exponential distribution $P(z) = \kappa e^{-\kappa z}$, so that the ath largest value among the z_n's can be written as the sum of the y_n's for $n \geq a$. Thus a clear connection between global fluctuations and extreme value statistics has been established. Besides, we have proposed a simple nonequilibrium model, defined through microscopic stochastic rules, for which the natural global quantity is exactly described by the generalized Gumbel distribution $G_a(x)$, with $a > 1$ a real value related to the parameters of the model. Such a simple model might be considered as a kind of ''ideal'' model that may be extended in several directions to describe in a more precise way some realistic systems. In particular, one expects that changing slightly the dynamical rules should yield a global energy distribution which is still close to a Gumbel distribution. In addition, the present model may be useful to study other issues of nonequilibrium statistical physics, as there are very few known solvable models including dissipation.

The author is grateful to I. Bena, M. Clusel, O. Dauchot, M. Droz, P. Holdsworth, C. Mazza, F. van Wijland, and Z. Rácz for fruitful discussions and interesting comments on the manuscript. This work has been supported in part by the Swiss National Science Foundation.

[1] S. T. Bramwell, P. C. W. Holdsworth, and J.-F. Pinton, Nature (London) **396**, 552 (1998).

[2] J.-F. Pinton, P. C. W. Holdsworth, and R. Labbé, Phys. Rev. E **60**, R2452 (1999).

[3] S. Aumaître, S. Fauve, S. McNamara, and P. Poggi, Eur. Phys. J. B **19**, 449 (2001).

[4] A. Noullez and J.-F. Pinton, Eur. Phys. J. B **28**, 231 (2002).

[5] B. Portelli, P. C. W. Holdsworth, and J.-F. Pinton, Phys. Rev. Lett. **90**, 104501 (2003).

[6] S. T. Bramwell *et al.*, Phys. Rev. Lett. **84**, 3744 (2000).

[7] S. T. Bramwell *et al.*, Phys. Rev. E **63**, 041106 (2001).

[8] B. Portelli, P. C. W. Holdsworth, M. Sellitto, and S. T. Bramwell, Phys. Rev. E **64**, 036111 (2001).

[9] M. Clusel, J.-Y. Fortin, and P. C. W. Holdsworth, Phys. Rev. E **70**, 046112 (2004).

[10] K. Dahlstedt and H. J. Jensen, J. Phys. A **34**, 11 193 (2001).

[11] M. Plischke, Z. Rácz, and R. K. P. Zia, Phys. Rev. E **50**, 3589 (1994).

[12] T. Antal, M. Droz, G. Györgyi, and Z. Rácz, Phys. Rev. Lett. **87**, 240601 (2001); Phys. Rev. E **65**, 046140 (2002).

[13] J. Farago, J. Stat. Phys. **107**, 781 (2002).

[14] J. J. Brey, M. I. Garcia de Soria, P. Maynar, and M. J. Ruiz-Montero, Phys. Rev. Lett. **94**, 098001 (2005).

[15] S. T. Bramwell, T. Fennell, P. C. W. Holdsworth, and B. Portelli, Europhys. Lett. **57**, 310 (2002).

[16] V. Aji and N. Goldenfeld, Phys. Rev. Lett. **86**, 1007 (2001).

[17] B. Portelli and P. C. W. Holdsworth, J. Phys. A **35**, 1231 (2002).

[18] S. C. Chapman, G. Rowlands, and N. W. Watkins, J. Phys. A **38**, 2289 (2005).

[19] E. J. Gumbel, *Statistics of Extremes* (Columbia University Press, New York, 1958).

[20] J.-P. Bouchaud and M. Mézard, J. Phys. A **30**, 7997 (1997).

[21] This requires that z has no upper bound, and that $P(z)$ decays faster than any power law at large z.

[22] G. Györgyi, P. C. W. Holdsworth, B. Portelli, and Z. Rácz, Phys. Rev. E **68**, 056116 (2003).

[23] J. Villain, J. Phys. (Paris) **36**, 581 (1975).

[24] M. Clusel, J.-Y. Fortin, and P. C. W. Holdsworth (to be published).

[25] W. Feller, *An Introduction to Probability Theory and its Applications* (Wiley, New York, 1966), Vol. II.

[26] E. Bertin (to be published).

[27] The present model is inspired by, but still quite different from, cascade models for turbulence (see, e.g., [5]). It should rather be thought of as a generic model with boundary injection and bulk dissipation.

[28] Technical details, as well as results for more general $J(\mu)$ and $\Delta(\mu)$, will be reported elsewhere [26].