Global fluctuations and Gumbel statistics

Eric Bertin
Department of Theoretical Physics, University of Geneva, CH-1211 Geneva 4, Switzerland
(Dated: September 17, 2018)

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.

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 fall-off on the other side, have been observed in many models or experimental systems, in the context of turbulence, equilibrium critical systems, nonequilibrium models exhibiting self-organized criticality, interface models, $1/f$ noise, Langevin equations, granular gas models, or even the statistics of the level of the Danube river. Quite strikingly, this analogy is not only qualitative, but many of the distributions observed in these very different systems actually fall, once suitably rescaled, close to the so-called Bramwell-Holdsworth-Pinton (BHP) distribution.

In this Letter, we explain how the statistics of global quantities, expressed as sums of non-identically 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 that the integrated power spectrum $w$ of periodic Gaussian $1/f$ noise is distributed, after a suitable rescaling, as a central Gaussian $G_1(x)$ for $a = 1$, but out to be the exact one for periodic Gaussian $1/f$-noise; it is also very close to the BHP one for $a \approx \pi/2$. The distribution $G_a(x)$ originates, for integer values of $a$, from the study of extreme value statistics, and describes the fluctuations of the $a$th largest value in a large set of identically distributed (independent) random variables $z_i$. 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. Still, attempts to identify an extremal process dominating the dynamics of such systems have failed up to now. All the above body of results thus leads to the following questions. First, what is (if any) the precise relationship 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 —like Fourier modes— can be defined, so that a problem of correlated random variables may be converted into a problem of independent random variables, with non-identical distributions —otherwise the central limit theorem would hold.

The distribution $G_a(x)$ is simply the sum of statistically independent Gaussian Fourier modes with complex amplitudes $\varepsilon_n$, distributed, after a suitable rescaling $x = (w - \langle w \rangle)/\sigma_w$, according to the Gumbel distribution $G_1(x)$. The model for $1/f$ noise used in consists in a large number $N$ of statistically independent Gaussian Fourier modes with complex amplitudes $\varepsilon_n$. Introducing $y_n \equiv |\varepsilon_n|^2 + |\varepsilon_{-n}|^2$, one has by definition $w = \sum_{n=1}^N y_n$. The distribution of $y_n$ reads

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

so that $w$ is simply the sum of $N$ independent random variables $y_n$, with non-identical exponential distributions. Note that the appearance of a non-Gaussian distribution is not surprising in itself, since the sum of
where the integral over $z_n$ is from $z_{n+1}$ to $\infty$, for $1 \leq n \leq N - 1$. 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 $v_n = z_n - z_{n+1}$ ($1 \leq n \leq N - 1$) and $v_N = z_N$ in Eq. (3), 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. (4). 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 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 $k$th 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 $\sigma_k^2 = \text{var}(z'_k)$, the generalized Gumbel distribution $G_k(x)$. The distribution $G_a(x)$ is defined for any positive real value $a$ by:

$$G_a(x) = \frac{\theta_a^a a^{a-1} e^{-\theta_a a^a}}{\Gamma(a)} \exp \left\{-a \left[\theta_a (x + \nu_a) + e^{-\theta_a (x + \nu_a)}\right]\right\} \tag{6}$$

with

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

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

\[ z^*_k = \sum_{n=k}^{N} y_n = \sum_{n=1}^{N-k+1} \tilde{y}_n \]  \hspace{1cm} (8)

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

\[ p_{n,k}(\tilde{y}_n) = (n + k - 1)\kappa e^{-(n+k-1)\kappa \tilde{y}_n} \]  \hspace{1cm} (9)

Thus the sum of independent random variables drawn from \( G_k(x) \) 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

\[ p_{n,a}(u_n) = (n+a-1)\kappa e^{-(n+a-1)\kappa u_n}, \quad 1 \leq n \leq N \]  \hspace{1cm} (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 \( 25 \).

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 \( \rho_n \) 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 \( \mu \) 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 \( \rho_n \); apart from the obvious constraint that one cannot withdraw from site \( n \) (either for transport or dissipation) an energy \( \mu \) greater than \( \rho_n \).

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

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

where \( \rho_j \) generically stands for all the variables that are not affected by \( \mu \). In the following, we focus on the specific case where \( J(\mu) = e^{-\beta \mu} \phi(\mu) \) and \( \Delta(\mu) = (e^{\lambda n} - 1) \phi(\mu) \), introducing two positive parameters \( \beta \) and \( \lambda \). 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):

\[
P(\{\rho_n\}) = \prod_{n=1}^N (\lambda n + \beta) e^{-(\lambda n + \beta)\rho_n} \]  \hspace{1cm} (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 like freely evolving granular gases \([14]\), Gumbel distributions are indeed observed even though the global quantity of interest cannot be 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 \(G_\alpha(x)\) 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 \((n+a-1)\). If \(a\) is integer, the variables \(u_n\) can be interpreted as the intervals \(y_n\) between two successive (ordered) random values drawn from an exponential distribution \(P(z) = ke^{-\alpha z}\), so that the \(\alpha\)th 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 \(G_\alpha(x)\), with \(\alpha > 1\) 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 partially supported by the Swiss National Science Foundation.

---

[1] S. T. Bramwell, P. C. W. Holdsworth, and J.-F. Pinton, Nature (London) \textbf{396}, 552 (1998).
[2] J.-F. Pinton, P. C. W. Holdsworth, and R. Labbé, Phys. Rev. E \textbf{60}, R2452 (1999).
[3] S. Aumaitre, S. Fauve, S. McNamara, and P. Poggi, Eur. Phys. J. B \textbf{19}, 449 (2001).
[4] A. Noullez and J.-F. Pinton, Eur. Phys. J. B \textbf{28}, 231 (2002).
[5] B. Portelli, P. C. W. Holdsworth, and J.-F. Pinton, Phys. Rev. Lett. \textbf{90}, 104501 (2003).
[6] S. T. Bramwell et al., Phys. Rev. Lett. \textbf{84}, 3744 (2000).
[7] S. T. Bramwell et al., Phys. Rev. E \textbf{63}, 041106 (2001).
[8] B. Portelli, P. C. W. Holdsworth, M. Sellitto, and S. T. Bramwell, Phys. Rev. E \textbf{64}, 036111 (2001).
[9] M. Clusel, J.-Y. Fortin, and P. C. W. Holdsworth, Phys. Rev. E \textbf{70}, 046112 (2004).
[10] K. Dahlstedt and H. J. Jensen, J. Phys. A: Math. Gen. \textbf{34}, 11193 (2001).
[11] M. Plischke, Z. Rácz, and R. K. P. Zia, Phys. Rev. E \textbf{50}, 3589 (1994).
[12] T. Antal, M. Droz, G. Györgyi, and Z. Rácz, Phys. Rev. Lett. \textbf{87}, 240601 (2001); Phys. Rev. E \textbf{65}, 046140 (2002).
[13] J. Farago, J. Stat. Phys. \textbf{107}, 781 (2002).
[14] J. J. Brey, M. I. Garcia de Soria, P. Maynar, and M. J. Ruiz-Montero, Phys. Rev. Lett. \textbf{94}, 098001 (2005).
[15] S. T. Bramwell, T. Fennell, P. C. W. Holdsworth, and B. Portelli, Europhys. Lett. \textbf{57}, 310 (2002).
[16] V. Aji and N. Goldenfeld, Phys. Rev. Lett. \textbf{86}, 1007 (2001).
[17] B. Portelli and P. C. W. Holdsworth, J. Phys. A: Math. Gen. \textbf{35}, 1231 (2002).
[18] S. C. Chapman, G. Rowlands and N. W. Watkins, J. Phys. A: Math. Gen. \textbf{38}, 2289 (2005).
[19] E. J. Gumbel, \textit{Statistics of Extremes} (Columbia University Press, New York, 1958).
[20] J.-P. Bouchaud and M. Mézard, J. Phys. A: Math. Gen. \textbf{30}, 7997 (1997).
[21] J. Villain, J. Phys. (Paris) \textbf{36}, 581 (1975).
[22] M. Clusel, J.-Y. Fortin, and P. C. W. Holdsworth, to be published.
[23] G. Györgyi, P. C. W. Holdsworth, B. Portelli, and Z. Ráczi, Phys. Rev. E \textbf{68}, 056116 (2003).
[24] W. Feller, \textit{An Introduction to Probability Theory and its Applications}, Vol. II. (Wiley, New-York, 1966).
[25] E. Bertin, to be published.
[26] This requires that \(z\) has no upper bound, and that \(P(z)\) decays faster than any power law at large \(z\).
[27] The present model is inspired by, but still quite differ-
ent 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.