INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **39** (2006) 7607–7619 doi:10.1088/0305-4470/39/24/001

Generalized extreme value statistics and sum of correlated variables

Eric Bertin[1] **and Maxime Clusel**[2]

[1] Department of Theoretical Physics, University of Geneva, CH-1211 Geneva 4, Switzerland
[2] Institut Laue-Langevin, 6 rue Horowitz BP156 X, 38042 Grenoble cedex, France

E-mail: eric.bertin@physics.unige.ch and clusel@ill.fr

Received 9 January 2006
Published 31 May 2006
Online at stacks.iop.org/JPhysA/39/7607

Abstract
We show that generalized extreme value statistics—the statistics of the kth largest value among a large set of random variables—can be mapped onto a problem of random sums. This allows us to identify classes of non-identical and (generally) correlated random variables with a sum distributed according to one of the three (k-dependent) asymptotic distributions of extreme value statistics, namely the Gumbel, Fréchet and Weibull distributions. These classes, as well as the limit distributions, are naturally extended to real values of k, thus providing a clear interpretation to the onset of Gumbel distributions with non-integer index k in the statistics of global observables. This is one of the very few known generalizations of the central limit theorem to non-independent random variables. Finally, in the context of a simple physical model, we relate the index k to the ratio of the correlation length to the system size, which remains finite in strongly correlated systems.

PACS numbers: 02.50.−r, 05.40.−a

1. Introduction

One of the cornerstones of probability theory is the celebrated central limit theorem, stating that under general assumptions, the distribution of the sum of independent random variables converges, once suitably rescaled, to a Gaussian distribution. This theorem provides one of the foundations of statistical thermodynamics and, from a more practical point of view, legitimates the use of Gaussian distributions to describe fluctuations appearing in experimental or numerical data.

However, this general theorem breaks down in several situations of physical interest, when the assumptions made to derive it are no longer fulfilled. For instance, if the random variables contributing to the sum have an infinite variance, the distribution of the sum is no longer Gaussian, but becomes a Lévy distribution [1]. This has deep physical consequences,

for instance, in the context of laser cooling [2] or in that of glasses [3, 4]. Though the central limit theorem applies to a class of correlated variables, the martingale differences [5, 6], it is generally inapplicable to sums of strongly correlated, or alternatively, strongly non-identical random variables. In this case, no general mathematical theory is available, but experimental as well as numerical distributions of the fluctuations of global observables in correlated systems often present an asymmetric shape, with an exponential tail on one side and a rapid fall-off on the other. Such distributions are usually well described by the so-called generalized Gumbel distribution with a real index a. The latter is one of the limit distributions appearing, for integer values $a = k$, as the distribution of the kth extremal value of a set of independent and identically distributed random variables [7, 8]. For instance, Bramwell and co-workers argued that the generalized Gumbel distribution with $a \approx 1.5$ reasonably describes the large scale fluctuations in many correlated systems [9]. Since then, the generalized Gumbel distribution has been reported in various contexts [10–14]. Yet, the interpretation of such a distribution in the context of extreme value statistics (EVS) is far from obvious. This issue led to a debate around the possible existence of a (hidden) extreme process which might dominate the statistics of sums of correlated variables in physically relevant situations [15–18]. Still, no evidence for such a process has been found yet. And even if one accepts that some extremal process is at play, a conceptual difficulty remains, as speaking of the kth largest value for non-integer k simply does not make sense.

In this paper, we propose another interpretation of the usual EVS in terms of statistics of sums of correlated variables. Within this framework the extension to real positive values of k is straightforward and avoids any logical problem. Usual EVS then appears as a particular case of a more general problem of limit theorem for sums of correlated variables belonging to a given class.

The paper is organized as follows. In section 2, we present the link between usual EVS and sums of correlated variables, and propose a more general problem of random sums, which is studied in the following sections. In section 3, we present in detail the simple problem of sum—already presented briefly by one of us in a previous publication [19]—associated with the EVS of exponential random variables. Section 4 deals with the extension to the more general problem of random sums defined in section 2, and it is shown that the asymptotic distribution of the sum is either a (generalized) Gumbel, Fréchet or Weibull distribution. Finally, we discuss in section 5 some physical interpretations of our results.

2. Equivalence between extreme value statistics and sums of correlated variables

2.1. Standard extreme value statistics

The asymptotic theory of EVS, which found applications for instance in physics [20], hydrology [21], seismology [22] and finance [23], is also a part of the important results of probability theory. Consider a random variable X, distributed according to the probability density P, and let $\mathfrak{X}_N$ be a set of N realizations of this random variable. Now instead of considering the sum of these N realizations as in the central limit theorem, one is interested in the study of the distribution of the kth maximal value z_k of $\mathfrak{X}_N$. The major result of EVS is that, in the limit $N \to \infty$, the probability density of z_k does not depend on the details of the original distribution P: there are only three different types of distributions, depending on the asymptotic behaviour of P [7, 8]. In particular, in the case where P decreases faster than any power laws at large x, the variable z_k is distributed according to the Gumbel distribution

$$G_k(z_k) = \frac{k^k\theta_k}{\Gamma(k)}\exp[-k\theta_k(z_k+\nu_k) - k\,\mathrm{e}^{-\theta_k(z_k+\nu_k)}],$$
$$\theta_k^2 = \Psi'(k), \qquad \nu_k = \frac{1}{\theta_k}[\ln k - \Psi(k)], \tag{1}$$

where $\Psi(x) = \frac{\mathrm{d}}{\mathrm{d}x}\ln\Gamma(x)$ is the digamma function. Note that from the point of view of EVS, $k \in \mathbb{N}^*$ by construction. Yet, if one forgets about EVS, expression (1) is also formally valid for $k = a$ real and positive.

2.2. Reformulation as a problem of sum

Let us consider again the set $\mathfrak{X}_N$ defined here above. We then introduce the ordered set of random variables, $z_k = x_{\sigma(k)}$, where σ is the ordering permutation such that $z_1 \geqslant z_2 \cdots \geqslant z_N$. One can now define the increments of z with the following relations:

$$u_k \equiv z_k - z_{k+1}, \qquad \forall\, 1 \leqslant k \leqslant N-1,$$
$$u_N \equiv z_N.$$

Then by definition, the following identity holds:

$$z_n \equiv \mathrm{Max}^{(n)}[\mathfrak{X}_N] = \sum_{k=n}^{N} u_k, \tag{2}$$

where $\mathrm{Max}^{(n)}$ is the nth largest value of the set $\mathfrak{X}_N$, showing that an extreme value problem may be reformulated as a problem of sum of random variables U_k.

Besides, as these new random variables have been obtained through an ordering process, they are *a priori* non-independent and non-identically distributed. This is confirmed by the computation of the joint probability $\tilde{J}_{k,N}(u_k,\ldots,u_N)$ which is given by

$$\begin{aligned}\tilde{J}_{k,N}(u_k,\ldots,u_N) &= N!\int_0^\infty \mathrm{d}z_N\,P(z_N)\int_{z_N}^\infty \mathrm{d}z_{N-1}\,P(z_{N-1})\ldots\int_{z_2}^\infty \mathrm{d}z_1\,P(z_1)\\ &\quad\times\ \delta(u_N - z_N)\prod_{n=k}^{N-1}\delta(u_n - z_n + z_{n+1}),\\ &= N!\left[\prod_{n=k}^{N} P\left(\sum_{i=n}^{N} u_i\right)\right]\int_{u_k+\cdots+u_N}^\infty \mathrm{d}z_{k-1}\,P(z_{k-1})\cdots\int_{z_2}^\infty \mathrm{d}z_1\,P(z_1).\end{aligned}$$

By recurrence it is easy to demonstrate the following relation:

$$\int_u^\infty \mathrm{d}z_{k-1}\,P(z_{k-1})\ldots\int_{z_2}^\infty \mathrm{d}z_1\,P(z_1) = \frac{1}{(k-1)!}F(u)^{k-1},$$

with $F(u) \equiv \int_u^\infty \mathrm{d}z\,P(z)$, leading to

$$\tilde{J}_{k,N}(u_k,\ldots,u_N) = \frac{N!}{(k-1)!}F\left(\sum_{i=k}^{N} u_i\right)^{k-1}\prod_{n=k}^{N} P\left(\sum_{i=n}^{N} u_i\right).$$

The probability $\tilde{J}_{k,N}$ is actually a probability at $N' = N+1-k$ points, and a shift of indices allows one to write it in the final form:

$$J_{N'}(u_1,\ldots,u_{N'}) = \frac{(N'+k-1)!}{(k-1)!}F\left(\sum_{i=1}^{N'} u_i\right)^{k-1}\prod_{n=1}^{N'} P\left(\sum_{i=n}^{N'} u_i\right). \tag{3}$$

In the general case, this expression does not factorize, i.e. one cannot find N' functions π_n such that

$$J_{N'}(u_1, \ldots, u_{N'}) = \prod_{n=1}^{N'} \pi_n(u_n),$$

and the random variables U_n are thus non-independent.

2.3. Extension of the joint probability

Up to now, we only reformulated the usual extreme value statistics as a problem of sum, by some trivial manipulations. Basically, we have two equivalent problems with the same asymptotic distribution, e.g., the standard Gumbel distribution if P is in the Gumbel class. Let us now point out that equation (3) is a particular case of the following joint distribution:

$$J_N(u_1, \ldots, u_N) = \frac{\Gamma(N)}{Z_N} \Omega\left[F\left(\sum_{n=1}^{N} u_n\right)\right] \prod_{n=1}^{N} P\left(\sum_{i=n}^{N} u_i\right), \tag{4}$$

where $\Omega(F)$ is an (arbitrary) positive function of F, and where Z_N is given by

$$Z_N = \int_0^1 \mathrm{d}v\, \Omega(v)(1-v)^{N-1}. \tag{5}$$

Equation (3) is recovered by choosing $\Omega(F) = F^{k-1}$. Note that by extending the definition of the joint probability, we lost the equivalence between the two original problems. In other words, we are generalizing the problem of statistics of sums and *not* the extreme value one. In the rest of this paper, we study the limit distribution of a sum of correlated random variables satisfying the generalized joint probability (4).

Let us note finally that one can easily generate from the joint probability (4) other joint probabilities that lead to the same asymptotic distribution for the sum, by summing over a set $\mathcal{S}$ of permutations over $[1, \ldots, N]$:

$$J_N^{\mathcal{S}}(u_1, \ldots, u_N) = \frac{1}{\mathcal{N}(\mathcal{S})} \sum_{\sigma \in \mathcal{S}} J_N(u_{\sigma(1)}, \ldots, u_{\sigma(N)}),$$

with $\mathcal{N}(\mathcal{S})$ being the cardinal of S. This may allow in particular some symmetry properties between the variables to be restored. Indeed, starting for instance from independent and non-identically distributed random variables, this procedure leads to a set of non-independent and identically distributed random variables with the same statistics for the sum.

3. The exponential case

Though in general the joint probability does not factorize, there is a particular case where it does. For pedagogical purposes, we discuss this case separately, before dealing with the general case in section 4. To allow the joint probability J_N defined by (4) to factorize, the function P has to satisfy the following property:

$$\forall (x, y), \qquad P(x+y) = P(x)P(y).$$

That is to say that P must be exponential: $P(x) = \kappa\, \mathrm{e}^{-\kappa x}$, with $\kappa > 0$. As a result, one has $F(x) = \mathrm{e}^{-\kappa x}$, and the factorization criterion on J_N translates into the following condition on Ω:

$$\forall (F_1, F_2), \qquad \Omega(F_1 F_2) = \Omega(F_1)\Omega(F_2),$$

so that Ω must be a power law: $\Omega(F) = F^{a-1}$ (the prefactor may be set to 1 without loss of generality), with $a > 0$ to ensure the convergence of the integral defining Z_N given in equation (5). In such a case, one finds

$$Z_N = \frac{\Gamma(N)\Gamma(a)}{\Gamma(N+a)}, \tag{6}$$

and the joint probability could be written in the factorized form (4), with

$$\pi_n(u_n) = (n+a-1)\kappa\, \mathrm{e}^{-(n+a-1)\kappa u_n}. \tag{7}$$

Therefore, for $a = k$ integer, if one sums independent and non-identically distributed random variables U_n, $n = 1, \ldots, N$, obeying (7), the distribution of the sum converges to a Gumbel distribution G_k in the limit $N \to \infty$. Note that the particular case $k = 1$ has been previously studied by Antal *et al* in the context of $1/f$ noise [24]. In the following, we shall establish that for a real, the distribution of the sum of U_ns converges towards the generalized Gumbel distribution G_a for $N \to \infty$, as announced in the previous publication [19].

To this aim, let us define the random sum $S_N = \sum_{n=1}^{N} U_n$, where U_n is distributed according to (7), with a real and positive. The distribution of S_N is denoted by Υ_N. As U_ns are independent, the Fourier transform of Υ_N is simply given by

$$\mathcal{F}[\Upsilon_N](\omega) = \prod_{n=1}^{N} \mathcal{F}[\pi_n](\omega) = \prod_{n=1}^{N} \left(1 + \frac{\mathrm{i}\omega}{\kappa(n+a-1)}\right)^{-1}. \tag{8}$$

The first two cumulants of the distribution read

$$\langle S_N \rangle = \sum_{n=1}^{N} \langle U_n \rangle = \frac{1}{\kappa} \sum_{n=1}^{N} \frac{1}{n+a-1}, \tag{9}$$

$$\sigma_N^2 = \sum_{n=1}^{N} \mathrm{Var}(U_n) = \frac{1}{\kappa^2} \sum_{n=1}^{N} \frac{1}{(n+a-1)^2}. \tag{10}$$

In order to get a well-defined distribution in the limit $N \to \infty$, let us introduce the reduced variable μ by

$$\mu = \frac{s - \langle S_N \rangle}{\sigma}, \qquad \text{with} \quad \sigma = \lim_{N\to\infty} \sigma_N.$$

Note that the fact $\sigma < \infty$ breaks Lindeberg's condition, allowing an eventual breakdown of the central limit theorem [25]. The distribution of μ, Φ_N, is then given by $\Phi_N(\mu) = \sigma \Upsilon_N(\sigma\mu + \langle S_N \rangle)$. The Fourier transform of $\Phi = \lim_{N\to\infty} \Phi_N$ reads, using (8) and (9)

$$\mathcal{F}[\Phi_\infty](\omega) = \lim_{N\to\infty} \mathcal{F}[\Phi_N](\omega) = \prod_{n=1}^{\infty} \left(1 + \frac{\mathrm{i}\omega}{\sigma\kappa(n+a-1)}\right)^{-1} \exp\left(\frac{\mathrm{i}\omega}{\sigma\kappa(n+a-1)}\right).$$

The last part of the previous expression has to be compared with the Fourier transform given in the appendix, leading to

$$\Phi_\infty(\mu) = G_a(\mu).$$

This result shows that it is possible to obtain quite directly the generalized Gumbel distribution with a real index a, not from an extremal process, but from a sum of independent non-identically distributed random variables. As already pointed out, the fact that the random variables U_n are uncorrelated is a specificity of the exponential distribution. A natural question is then to know whether this result survives for a more general class of distributions P and functions Ω. We address this question in the next section.

4. Generalized extreme value distributions

In the general case, the random variables U_n are non-independent, so we have to deal with the joint probability (4). Accordingly, the route to the asymptotic limit distribution will be quite different from the exponential case.

4.1. Distribution of the sum for finite N

Consider a set of realizations $\{u_n\}$ of N (correlated) random variables U_n, with the joint probability (4). We then define as above the random variable $S_N = \sum_{n=1}^{N} U_n$, and let Υ_N be the probability density of S_N. Then Υ_N is given by

$$\Upsilon_N(s) = \int_0^\infty \mathrm{d}u_N \ldots \mathrm{d}u_1\, J_N(u_1, \ldots, u_N)\delta\left(s - \sum_{n=1}^{N} u_n\right).$$

Inserting (4), one obtains

$$\Upsilon_N(s) = \frac{\Gamma(N)}{Z_N} P(s)\Omega(F(s))I_N(s),$$

with

$$I_N(s) = \int_0^\infty \mathrm{d}u_N\, P(u_N) \int_0^\infty \mathrm{d}u_{N-1}\, P(u_N + u_{N-1}) \ldots \int_0^\infty \mathrm{d}u_1 \delta\left(s - \sum_{n=1}^{N} u_n\right).$$

To evaluate I_N, let us start by integrating over u_1, using

$$\int_0^\infty \mathrm{d}u_1\, \delta\left(s - \sum_{n=1}^{N} u_n\right) = \Theta\left(s - \sum_{n=2}^{N} u_n\right),$$

where Θ is the Heaviside distribution. This changes the upper bound of the integral over u_2 by $u_2^{\max} = \max\left(0, s - \sum_{n=2}^{N} u_n\right)$. Then the integration over u_2 leads to

$$\int_0^{u_2^{\max}} \mathrm{d}u_2\, P(u_2) = \left[F\left(\sum_{n=3}^{N} u_n\right) - F(s)\right]\Theta\left(s - \sum_{n=3}^{N} u_n\right).$$

By recurrence it is then possible to show that

$$I_N(s) = \frac{1}{\Gamma(N)}(1 - F(s))^{N-1},$$

finally yielding the following expression for Υ_N:

$$\Upsilon_N(s) = \frac{1}{Z_N} P(s)\Omega(F(s))(1 - F(s))^{N-1}. \tag{11}$$

In the following sections, we assume that $\Omega(F)$ behaves asymptotically as a power law $\Omega(F) \sim \Omega_0 F^{a-1}$ when $F \to 0$ $(a > 0)$. Under this assumption, we deduce from equation (11) the different limit distributions associated with the different classes of asymptotic behaviour of P at large x.

4.2. The Gumbel class

In this section we focus on the case where P is in the Gumbel class, that is, $P(x)$ decays faster than any power law at large x. Note that the exponential case studied above precisely belongs to this class. Our aim is to show that, after a suitable rescaling of the variable,

the limit distribution obtained from (11) is the generalized Gumbel distribution G_a, where a characterizes the asymptotic behaviour of $\Omega(F)$ for $F \to 0$.

To that purpose, we define s_N^* by $F(s_N^*) = a/N$. If a is an integer, this is nothing but the typical value of the ath largest value of s in a sample of size N. As P is unbounded we have

$$\lim_{N\to\infty} s_N^* = +\infty.$$

Let us introduce $g(s) = -\ln F(s)$ and, assuming $g'(s_N) \neq 0$, define the rescaled variable v by

$$s = s_N^* + \frac{v}{g'(s_N^*)}. \tag{12}$$

For large N, one can perform a series expansion of g around s_N^*:

$$g(s) = g(s_N^*) + v + \sum_{n=2}^{\infty} \frac{1}{n!} \frac{g^{(n)}(s_N^*)}{g'(s_N^*)^n} v^n.$$

For P in the Gumbel class, $g^{(n)}(s_N^*)/g'(s_N^*)^n$ is bounded as a function of n so that the series converges. In addition, one has

$$\lim_{N\to\infty} \frac{g^{(n)}(s_N^*)}{g'(s_N^*)^n} = 0, \quad \forall n \geqslant 2,$$

so that $g(s)$ may be written as

$$g(s) = g(s_N^*) + v + \varepsilon_N(v), \qquad \text{with} \quad \lim_{N\to\infty} \varepsilon_N(v) = 0. \tag{13}$$

Given that $P(s) = g'(s)F(s)$, one gets using equations (11) and (13)

$$\begin{aligned}\Phi_N(v) &= \frac{1}{g'(s_N^*)} \Upsilon_N(s) \\ &= \frac{1}{Z_N} \frac{g'(s)}{g'(s_N^*)} F(s)\Omega(F(s))(1 - F(s))^{N-1},\end{aligned}$$

where s is given by equation (12). For P in the Gumbel class, it can be checked that, for fixed v

$$\lim_{N\to\infty} \frac{g'(s_N^* + v/g'(s_N^*))}{g'(s_N^*)} = 1.$$

Besides, $F(s_N^* + v/g'(s_N^*)) \to 0$ when $N \to \infty$, so that one can use the small F expansion of $\Omega(F)$. Altogether, one finds

$$\Phi_N(v) \underset{N\to\infty}{\sim} \frac{\Omega_0}{Z_N} \left(\frac{a}{N}\right)^a \mathrm{e}^{-av - a\varepsilon_N(v)} \left[1 - \frac{a}{N} \mathrm{e}^{-v-\varepsilon_N(v)}\right]^{N-1}.$$

Using a simple change of variable in equation (5), one can show that

$$\lim_{N\to\infty} \frac{N^a Z_N}{\Omega_0} = \Gamma(a).$$

It is then straightforward to take the asymptotic limit $N \to \infty$, leading to

$$\Phi_\infty(v) = \frac{a^a}{\Gamma(a)} \exp[-av - a\,\mathrm{e}^{-v}].$$

In order to recover the usual expression for the generalized Gumbel distribution, one simply needs to introduce the reduced variable

$$\mu = \frac{v - \langle v \rangle}{\sigma_v},$$

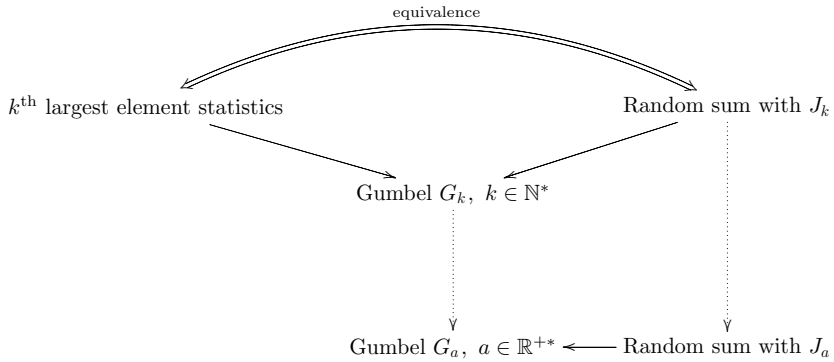

Figure 1. Generalization of EVS *via* sums of non-independent variables.

with, Ψ being the digamma function,

$$\langle v \rangle = \ln a - \Psi(a), \qquad \sigma_v^2 = \Psi'(a).$$

The variable μ is then distributed according to a generalized Gumbel distribution (1).

To sum up, if one considers the sum S_N of $N \gg 1$ random variables linked by the joint probability (4), then the asymptotic distribution of the reduced variable μ defined by

$$\mu = \frac{s_N - \langle S_N \rangle}{\sigma_N},$$

with

$$\langle S_N \rangle = s_N^* + \frac{\ln a - \Psi(a)}{g'(s_N^*)}, \qquad \sigma_N = \frac{\sqrt{\Psi'(a)}}{g'(s_N^*)}.$$

is the generalized Gumbel distribution (1). More generally, the approach developed in this paper to relate EVS to sums of non-independent random variables, and then generalize the problem of sum, may be summarized as shown in figure 1 on the example of the Gumbel class.

4.3. Extension to the Fréchet and Weibull classes

Let us now consider the cases where P belongs either to the Fréchet class, that is, P has a power law tail

$$P(x) \underset{x\to\infty}{\sim} \frac{P_0}{x^{1+\eta}}, \qquad \eta > 0,$$

with $P_0 > 0$ being a constant, or to the Weibull class, that is, P has an upper bound A and behaves as a power law in the vicinity of A:

$$\begin{aligned} &\forall x > A, \qquad P(x) = 0 \\ &P(x) \underset{x\to A^-}{\sim} \tilde{P}_0 (A - x)^{\beta-1}, \qquad \beta > 0, \end{aligned} \tag{14}$$

with $\tilde{P}_0$ being an arbitrary positive prefactor. A calculation similar to that used in the Gumbel case allows the asymptotic distributions to be determined. Considering first the Fréchet class, we define s_N^* through $F(s_N^*) = a/N$ as above, and introduce the scaling variable $v = s/s_N^*$. This gives for the distribution $\Phi_N^{\mathrm{f}}(v)$,

$$\Phi_N^{\mathrm{f}}(v) \underset{N\to\infty}{\sim} \frac{a^a \Omega_0}{N^a Z_N} \frac{\eta}{v^{1+a\eta}} \left(1 - \frac{a}{N v^\eta}\right)^{N-1}.$$

Taking the limit $N \to \infty$, the asymptotic distribution reads

$$\Phi^{\mathrm{f}}_{\infty}(v) = \frac{a^a}{\Gamma(a)} \frac{\eta}{v^{1+a\eta}} \mathrm{e}^{-a/v^{\eta}}.$$

One can rescale the variable v with respect either to the mean value or to the standard deviation σ. In this latter case, one finds

$$\sigma = \frac{a^{1/\eta}}{\Gamma(a)} \left[\Gamma(a)\Gamma\left(a - \frac{2}{\eta}\right) - \Gamma\left(a - \frac{1}{\eta}\right)^2 \right]^{1/2}.$$

Introducing $\mu = v/\sigma$ we get

$$F_a(\mu) = \frac{\eta \Lambda^a}{\Gamma(a)} \frac{1}{\mu^{1+a\eta}} \exp(-\Lambda \mu^{-\eta}),$$

$$\Lambda = \Gamma(a)^{\eta} \left[\Gamma(a)\Gamma\left(a - \frac{2}{\eta}\right) - \Gamma\left(a - \frac{1}{\eta}\right)^2 \right]^{-\eta/2}.$$

Calculations for the Weibull class are very similar, given that the scaling variable is now $v = (A - s)/(A - s^*_N)$. Skipping the details, the asymptotic distribution reads

$$\Phi^{\mathrm{w}}_{\infty}(v) = \frac{\beta a^a}{\Gamma(a)} v^{a\beta - 1} \mathrm{e}^{-a v^{\beta}}.$$

After rescaling to normalize the second moment

$$\sigma = \frac{a^{-1/\beta}}{\Gamma(a)} \left[\Gamma(a)\Gamma\left(a + \frac{2}{\beta}\right) - \Gamma\left(a + \frac{1}{\beta}\right)^2 \right]^{1/2},$$

one obtains

$$W_a(\mu) = \frac{\beta}{\Gamma(a)} \Lambda^a \mu^{\beta a - 1} \exp(-\Lambda \mu^{\beta}),$$

$$\Lambda = \Gamma(a)^{-\beta} \left[\Gamma(a)\Gamma\left(a + \frac{2}{\beta}\right) - \Gamma\left(a + \frac{1}{\beta}\right)^2 \right]^{\beta/2}.$$

Thus here again, the distributions appearing in EVS are generalized into distributions of sums, including a real parameter a coming from the asymptotic behaviour of the function Ω.

Note that the generalized Fréchet and Weibull distributions found above may also arise from a different statistical problem. As already noted by Gumbel [7] in the context of EVS, the Fréchet and Weibull distributions may be related to the Gumbel distribution through a simple change of variable. If a random variable X is distributed according to a Gumbel distribution G_a, then the distribution of the variable $Y = \mathrm{e}^{\lambda X}$ ($\lambda > 0$) is a generalized Fréchet distribution. As X is defined as a sum of correlated random variables associated with a joint probability (4) with P in the Gumbel class, Y appears as the product of variables with P in the Fréchet class. Similarly, the generalized Weibull distribution may also be obtained from the Gumbel distribution by the change of variable $Y = \mathrm{e}^{-\lambda X}$.

In summary, one sees that the reformulation of standard EVS by means of the joint probability (4) allows a natural and straightforward definition of all the generalized extreme value statistics. As already stressed, this generalization breaks the equivalence with the extreme value problem: there is no associated extreme process leading to the generalized extreme value distributions G_k, F_k or W_k with k being real. At the level of the joint probability (4), the equivalence with EVS only holds if $\Omega(F)$ is a pure power-law of F, with an *integer* exponent k. Yet, concerning the asymptotic distribution, extreme value distributions are recovered as soon as $\Omega(F) \sim \Omega_0 F^{k-1}$ for $F \to 0$ (with k integer), that is, when Ω is regular in $F = 0$.

5. Physical interpretation

Up to now we presented a mathematical result about sums of random variables, linked by the joint probability (4). This joint probability leads to non-Gaussian distributions that may be interpreted as the result of correlations between those variables. All the information about the correlation is included in J_N. From a physicist point of view, this is not completely satisfying: the degree of correlation is indeed usually quantified by the correlation length. At first sight it is disappointing that we cannot extract such a quantity from (4). One has to realize however that up to now we only dealt with *numbers*, without giving any physical meaning to any of those numbers. To get some physical information from our result, we have to put some physics in it first[3], giving an interpretation of u_n, explaining also how those quantities are arranged in time or space, introducing then a notion of distance or time between u_n and the dimensionality of space. The results presented in the previous sections could therefore describe various physical situations. In this section, we illustrate this idea by studying a simple model.

Let us consider a one-dimensional lattice model with a continuous variable ϕ_x on each site $x = 1, \ldots, L$. Although we do not specify the dynamics explicitly, we have in mind that ϕ_xs are strongly correlated. One can define the Fourier modes $\hat{\phi}_q$ associated with ϕ_x through

$$\hat{\phi}_q = \frac{1}{\sqrt{L}} \sum_{x=1}^{L} \phi_x \, \mathrm{e}^{\mathrm{i}qx}.$$

A natural global observable is the integrated power spectrum ('energy') $E = \sum_x \phi_x^2$. From the Parseval theorem, E may also be expressed as a function of $\hat{\phi}_q$ as $E = \sum_q |\hat{\phi}_q|^2$. We now assume that the squared amplitudes $u_n = |\hat{\phi}_q|^2$, with $q = 2\pi n/L$, follow the statistics defined by equation (4). For simplicity, we consider the simplest case where $P(z) = \kappa \, \mathrm{e}^{-\kappa z}$ and $\Omega(F) = F^{a-1}$, although more complicated situations may be dealt with. This actually generalizes the study of the $1/f$ noise by Antal *et al* [24]. The power spectrum is given by

$$\begin{aligned} \langle |\hat{\phi}_q|^2 \rangle &= \left[\kappa \left(\frac{|q|L}{2\pi} + a - 1 \right) \right]^{-1} \propto \frac{1}{|q| + m}, \\ m &= \left(\frac{2\pi(a-1)}{L} \right)^{-1}. \end{aligned} \tag{15}$$

The correlation function $C(r)$, given by the inverse Fourier transform of $\langle |\hat{\phi}_q|^2 \rangle$,

$$C(r) = \mathcal{F}^{-1}[\langle |\hat{\phi}_q|^2 \rangle](r), \tag{16}$$

can be computed using (15), leading to

$$C(r) \propto -\sin(mr)\,\mathrm{si}(mr) - \cos(mr)\,\mathrm{ci}(mr),$$

where si and ci are respectively the sine and cosine integral functions [27]. The correlation length is therefore defined as the typical length scale of $C(r)$:

$$\xi = m^{-1} = \frac{L}{2\pi(a-1)}.$$

Thus, ξ appears to be proportional to the system size, which was expected from the breaking of the central limit theorem[4]. The particular case of a $1/f$ noise ($a = 1$) then corresponds to

[3] 'Mathematician prepares abstract reasoning ready to be used, if you have a set of axioms about the real world. But the physicist has meaning to all his phrases.' Feynman in [26].

[4] Dividing a system of linear size L into (essentially) independent subsystems with a size proportional to ξ, the number of subsystems remains finite when $L \to \infty$ if $\xi \sim L$, so that the central limit theorem should not hold.

a highly correlated system, with $\xi/L \to \infty$. For $a \approx 1.5$, a value often reported in physical systems [9], one gets $\xi/L \simeq 0.15$. The correlation is weaker but ξ still diverges with L [18].

Altogether, this simple model may be thought of as a minimal model, which allows some generic properties of more complex correlated physical systems to be understood.

Obviously not all correlated systems will exhibit generalized EVS. A well-known counter-example is the 2D Ising model at its critical temperature [28]. Strictly speaking the generalized EVS can only be obtained if there exists P such that the joint probability is given by (4). The number of observations of distributions close to a generalized Gumbel distribution suggests however that expression (4) is general enough to reasonably approximate the real joint probability in many situations. This would explain the ubiquity of the generalized Gumbel in correlated systems.

More practically, the generalized extreme value distributions also appear as natural fitting functions for global fluctuations. If one is measuring fluctuations of some global quantities, it seems quite reasonable to fit them with an asymptotic distribution which could possibly take into account a violation of the hypothesis of the central limit theorem. The generalized EVS are one of the few such distributions. Using the reduced variable μ, there is only one free parameter in the Gumbel class, a, which quantifies the deviation from the CLT.

6. Conclusion

In this paper, we established that the so-called generalized extreme value distributions are the asymptotic distributions of random sums, for particular classes of random variables defined by equation (4), which do not satisfy the hypothesis underlying the central limit theorem. Interestingly, this is one of the very few known generalizations of the central limit theorem to non-independent random variables. In this framework, it becomes clear that it is vain to look for a hidden extreme process when one of the generalized extreme value distributions is observed in a problem of global fluctuations. Therefore, qualifying such distributions of generalized extreme value statistics is somehow misleading. The parameter a quantifies the dependence of the random variables, although further physical inputs are needed to give a physical interpretation to this parameter. Within a simple model of independent Fourier modes, a is related in a simple way to the ratio of the correlation length to the system size, a ratio that remains constant in the thermodynamic limit for strongly correlated systems. Besides, we believe that the classes of random variables defined by equation (4) may be regarded as reference classes in the context of random sums breaking the central limit theorem, due to their mathematically simple form. Along this line of thought, it is not so surprising that distributions close to the generalized Gumbel distribution are so often observed in the large scale fluctuations of correlated systems, as for instance in the case of the XY model at low temperature.

Acknowledgments

It is a pleasure to thank Steven T Bramwell, Michel Droz, Jean-Yves Fortin and Peter C W Holdsworth for useful discussions and comments. MC thanks University College London for hospitality during completion of this work.

Appendix. Fourier transform of the generalized Gumbel distribution

In this appendix, we give an expression for the Fourier transform of generalized Gumbel distribution G_a defined by

$$G_a(x) = \frac{\theta a^a}{\Gamma(a)} \exp[-a\theta(x+\nu) - a\,\mathrm{e}^{\theta(x+\nu)}],$$

with $\theta^2 = \Psi'(a)$ and $\theta\nu = \ln a - \Psi(a)$. The Fourier transform of G_a is defined by

$$\mathcal{F}[G_a](\omega) = \int_{-\infty}^{+\infty} \mathrm{d}x\, \mathrm{e}^{-\mathrm{i}\omega x} G_a(x).$$

Letting $u = a\exp(-\theta(x+\nu))$, we obtain

$$\begin{aligned}\mathcal{F}[G_a](\omega) &= \frac{1}{\Gamma(a)} a^{-\mathrm{i}\omega/\theta}\, \mathrm{e}^{\mathrm{i}\nu\omega} \int_0^{\infty} \mathrm{d}u\, u^{\mathrm{i}\omega/\theta + a - 1}\, \mathrm{e}^{-u},\\ &= \frac{\mathrm{e}^{\mathrm{i}\nu\omega}}{a^{\mathrm{i}\omega/\theta}} \frac{\Gamma\left(\frac{\mathrm{i}\omega}{\theta} + a\right)}{\Gamma(a)}.\end{aligned}$$

Using the identity [27]

$$\Gamma(1+z) = \mathrm{e}^{-\gamma z} \prod_{n=1}^{\infty} \frac{\mathrm{e}^{z/n}}{1+\frac{z}{n}}, \tag{A.1}$$

one finally obtains, with $\alpha = a - 1$,

$$\begin{aligned}\mathcal{F}[G_a](\omega) &= \frac{\mathrm{e}^{\mathrm{i}(\nu - \frac{\gamma}{\theta})\omega}}{(1+\alpha)^{\mathrm{i}\omega/\theta}} \prod_{n=1}^{\infty} \frac{\exp\left(\frac{\mathrm{i}\omega}{n\theta}\right)}{1+\frac{\mathrm{i}\omega}{\theta(n+\alpha)}}\\ &= \left\{\frac{\mathrm{e}^{\mathrm{i}(\nu - \frac{\gamma}{\theta})\omega}}{(1+\alpha)^{\mathrm{i}\omega/\theta}} \exp\left[\frac{\mathrm{i}\omega}{\theta} \sum_{n=1}^{\infty}\left(\frac{1}{n} - \frac{1}{n+\alpha}\right)\right]\right\} \prod_{n=1}^{\infty} \frac{\exp\left(\frac{\mathrm{i}\omega}{n(\theta+\alpha)}\right)}{1+\frac{\mathrm{i}\omega}{\theta(n+\alpha)}}.\end{aligned}$$

The identity (A.1) leads to

$$\frac{\mathrm{d}\ln\Gamma(1+\alpha)}{\mathrm{d}\alpha} = -\gamma + \sum_{n=1}^{\infty}\left(\frac{1}{n} - \frac{1}{n+\alpha}\right), \tag{A.2}$$

and therefore we obtain, using (10),

$$\theta^2 = \frac{\mathrm{d}^2\ln\Gamma(1+\alpha)}{\mathrm{d}\alpha^2} = \sum_{n=1}^{\infty} \frac{1}{(n+\alpha)^2} = \sigma^2.$$

Furthermore, relation (A.2) leads to

$$\begin{aligned}(\nu\theta - \gamma) - \ln(1+\alpha) + \sum_{n=1}^{\infty}\left(\frac{1}{n} - \frac{1}{n+\alpha}\right) &= (\nu\theta - \gamma) - \ln a + \left(\gamma + \frac{\mathrm{d}\ln\Gamma}{\mathrm{d}a}\right),\\ &= 0.\end{aligned}$$

So finally we have

$$\mathcal{F}[G_a](\omega) = \prod_{n=1}^{\infty} \frac{\exp\left(\frac{\mathrm{i}\omega}{\theta(n-1+a)}\right)}{1+\frac{\mathrm{i}\omega}{\theta(n-1+a)}}.$$

References

[1] Lévy P 1954 *Théorie de l'addition de variables aléatoires* (Paris: Gauthier-Villard) (édition J. Gabay, 2003)

[2] Bardou F, Bouchaud J-P, Aspect A and Cohen-Tannouji C 2002 *Lévy Statistics and Laser Cooling* (Cambridge: Cambridge University Press)

[3] Bouchaud J-P and Georges A 1990 *Phys. Rep.* **195** 127

[4] Bouchaud J-P 1992 *J. Phys. I (France)* **2** 1705

[5] Billingsley P 1961 *Proc. Am. Math. Soc.* **12** 788
[6] Ibragimov I A 1963 *Theor. Probab. Appl.* **8** 83
[7] Gumbel E J 1958 *Statistics of Extremes* (New York: Columbia University Press) (Dover publication, 2004)
[8] Galambos J 1987 *The Asymptotic Theory of Extreme Order Statistics* (New York: Wiley)
[9] Bramwell S T *et al* 2000 *Phys. Rev. Lett.* **84** 3744
[10] Noullez A and Pinton J-F 2002 *Eur. Phys. J.* B **28** 231
[11] Chamon C *et al* 2004 *J. Chem. Phys.* **121** 10120
[12] Pennetta C, Alfinito E, Reggiani L and Ruffo S 2004 *Semicond. Sci. Technol.* **19** S164
[13] Duri A, Bissig H, Trappe V and Cipelletti L 2005 *Phys. Rev.* E **72** 051401
[14] Varotsos P A, Sarlis N V, Tanaka H K and Skordas E S 2005 *Phys. Rev.* E **72** 041103
[15] Dahlstedt K and Jensen H J 2001 *J. Phys. A: Math. Gen.* **34** 11193
[16] Watkins N W, Chapman S C and Rowland G 2002 *Phys. Rev. Lett.* **89** 208901
[17] Bramwell S T *et al* 2002 *Phys. Rev. Lett.* **89** 208902
[18] Clusel M, Fortin J-Y and Holdsworth P C W 2004 *Phys. Rev.* E **70** 046112
[19] Bertin E 2005 *Phys. Rev. Lett.* **95** 170601
[20] Bouchaud J-P and Mézard M 1997 *J. Phys. A: Math. Gen.* **30** 7997
[21] Katz R W, Parlangi M B and Naveau P 2002 *Advances in Water Resources* **25** 1287
[22] Sornette D, Knopoff L, Kagan Y and Vannest C 1996 *J. Geophys. Res.* **101** 13883
[23] Longin F 2000 *J. Bank. Finance* **24** 1097
[24] Antal T, Droz M, Györgyi G and Rácz Z 2001 *Phys. Rev. Lett.* **87** 240601
[25] Feller W 1968 *An Introduction to Probability Theory and its Applications* vol 1, 3rd edn (New York: Wiley)
[26] Feynman R P 1967 *The Character of Physical Laws* (Cambridge, MA: MIT Press)
[27] Gradshteyn I S and Ryzhik I M 1994 *Table of Integrals, Series, and Products* 5th edn (New York: Academic)
[28] Bruce A D 1981 *J. Phys. C: Solid State Phys.* **14** 3667