HAL
open science

Extremum statistics: a framework for data analysis

S. C. Chapman, G. Rowlands, N. W. Watkins

To cite this version:

S. C. Chapman, G. Rowlands, N. W. Watkins. Extremum statistics: a framework for data analysis. Nonlinear Processes in Geophysics, 2002, 9 (5/6), pp.409-418. hal-00302141

HAL Id: hal-00302141

https://hal.science/hal-00302141v1

Submitted on 18 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nonlinear Processes in Geophysics (2002) 9: 409–418

Nonlinear Processes in Geophysics
© European Geosciences Union 2002

Extremum statistics: a framework for data analysis

S. C. Chapman[1], **G. Rowlands**[1], **and N. W. Watkins**[2]

[1]Physics Department, University of Warwick, Coventry CV4 7AL, UK
[2]British Antarctic Survey, High Cross, Madingley Rd., Cambridge CB3 0ET, UK

Received: 24 October 2001 – Revised: 21 January 2002 – Accepted: 8 February 2002

Abstract. Recent work has suggested that in highly correlated systems, such as sandpiles, turbulent fluids, ignited trees in forest fires and magnetization in a ferromagnet close to a critical point, the probability distribution of a global quantity (i.e. total energy dissipation, magnetization and so forth) that has been normalized to the first two moments follows a specific non-Gaussian curve. This curve follows a form suggested by extremum statistics, which is specified by a single parameter a ($a = 1$ corresponds to the Fisher-Tippett Type I ("Gumbel") distribution).

Here we present a framework for testing for extremal statistics in a global observable. In any given system, we wish to obtain a, in order to distinguish between the different Fisher-Tippett asymptotes, and to compare with the above work. The normalizations of the extremal curves are obtained as a function of a. We find that for realistic ranges of data, the various extremal distributions, when normalized to the first two moments, are difficult to distinguish. In addition, the convergence to the limiting extremal distributions for finite data sets is both slow and varies with the asymptote. However, when the third moment is expressed as a function of a, this is found to be a more sensitive method.

1 Introduction

The study of systems exhibiting non-Gaussian statistics is of considerable current interest (see, e.g. Sornette, 2000, and references therein). These statistics are often observed to arise in finite sized, multi-body systems, exhibiting correlation over a broad range of scales, leading to emergent phenomenology, such as self-similarity and in some cases fractional dimension (Bohr et al., 1998). The apparent ubiquitous nature of this behavior has led to interest in self-organized criticality (Bak, 1997; Jensen, 1998) as a paradigm; other highly correlated systems include those exhibiting fully developed turbulence. In solar terrestrial physics in particular, problems of interest include MHD turbulence in the solar wind and in the Earth's magnetotail. Irregular or bursty transport and energy release in the latter has recently led to complex system approaches such as SOC (see the review by Chapman and Watkins, 2001). These complex systems are often characterized by a lack of scale, and in particular, by the exponents of the power law probability distributions (PDF) of patches of activity in the system. Examples of these patches of activity include energy dissipated by avalanches in sandpiles, vortices in turbulent fluids, ignited trees in forest fires and magnetization in a ferromagnet close to the critical point. In the Earth's magnetotail, patches of activity in the aurora, as seen by POLAR UVI have been used as a proxy for the energy released in bursty magnetotail transport, in order to infer its scaling properties (Lui et al., 2000; Uritsky et al., 2001). The challenge is to distinguish the system from an uncorrelated Gaussian process, by demonstrating self-similarity and to determine the power law exponents. To do this directly is nontrivial, requiring measurements of the individual patches or activity events over many decades. Here we consider what may be a more readily accessible measure: the statistics of a global average quantity, such as the total energy dissipation, magnetization and so forth.

An important hypothesis that is the subject of this paper is that the data arise from an extremum process, i.e. that some unknown selection process operates such that the observed global quantity is dominated by the largest events selected from ensembles of individual "patches" of activity. This is a real possibility for two reasons. First, measurements of physical systems, and in particular, observations of natural systems, inevitably incorporate instrumental thresholds and this may affect the statistics of a global quantity comprising activity summed over patches. Second, there has recently been considerable interest in a series of intriguing results from turbulence experiments (Labbe et al., 1996; Pinton et al., 1999; Bramwell et al., 1998), and numerical models exhibiting correlations (Bramwell et al. (2000), (see also Aji and Goldenfeld, 2001; Zheng and Trimper, 2001; Bramwell et al., 2001).

Correspondence to: S. C. Chapman
(sandrac@astro.warwick.ac.uk)

These studies reveal statistics of a global quantity (i.e. E) that follow curves that are of the form of one of the limiting extremal distribution (Gumbel, 1958; Fisher and Tippett, 1928):

$$P(E) = K(e^{y-e^y})^a \quad y = b(E-s), \tag{1}$$

where K, b and s are obtained by normalizing to the first two moments ($M_0 = 1$, $M_1 = 0$, $M_2 = 1$), and the single parameter a appears to be close to the value $\pi/2$.

For an infinitely large ensemble, there are two limiting distributions that we consider here. The Fisher-Tippett type I (or "Gumbel") extremal distribution is of the form (1), but with $a = 1$ and arises from selecting the largest events from ensembles with distributions that fall off exponentially or faster. Since we wish to construct a framework that could encompass all highly correlated systems, we also treat the case where the distribution of "patches" is a power law. An example is the Potts model (Cardy, 1996) for magnetization, where connected bonds form clusters, the size of which is power law distributed at the critical point. In this case, the relevant extremal distribution is Fisher-Tippett type II (or "Frechet").

Here we provide a framework for comparing data with Fisher-Tippett type I and II extremal curves. This essentially requires obtaining the normalizations of these curves in terms of the moments of the data and ultimately as functions of the single parameter a.

We find that the curves of form (1), which are obtained by normalizing to the first two moments, are difficult to distinguish if a is in the range $(1, 2)$ or from Frechet curves given a realistic range of data. Furthermore, we demonstrate that slow convergence with respect to the size of the data set, to the limiting $a = 1$ extremal distribution has the consequence that, for a large but finite ensemble, the extremal distribution of an uncorrelated Gaussian process is indistinguishable from the $a = \pi/2$ curve. To overcome these limitations we suggest two much more sensitive methods for determining whether or not the curve is of the form (1), and, if so, the corresponding value of a. These methods are based on the third moment, and the peak of the distribution, both of which we obtain here as a function of a.

2 Extremum statistics: general results

To facilitate the work here we first develop some results from extremum statistics (for further background reading, see Sornette (2000); Gumbel (1958); Bouchaud and Potters (2000)). If the maximum Q^* drawn from an ensemble of M patches of activity Q with distribution $N(Q)$ is $Q^* = max\{Q_1, ..Q_M\}$, then the probability distribution (PDF) for Q^* is given by

$$P_m(Q^*) = MN(Q^*)(1 - N_>(Q^*))^{M-1}, \tag{2}$$

where M is the number of patches in the ensemble and

$$N_>(Q^*) = \int_{Q^*}^{\infty} N(Q)dQ. \tag{3}$$

We now obtain P_m for large M, Q. For general PDF $N(Q)$ we can write (for appropriate choice of the function $g(Q^*)$):

$$(1 - N_>)^M = e^{-Mg(Q^*)} \tag{4}$$

and for small $N_>(Q^*)$ we have

$$g(Q^*) = -\ln(1 - N_>(Q^*)) \sim N_> + \frac{N_>^2}{2}. \tag{5}$$

We now consider a characteristic value of Q^*, namely $\tilde{Q}^*$, such that by definition

$$Mg(\tilde{Q}^*) = q, \tag{6}$$

so that

$$q = Mg(\tilde{Q}^*) \approx MN_>(\tilde{Q}^*) + M\frac{N_>^2(\tilde{Q}^*)}{2} + \cdots. \tag{7}$$

We now expand $g(Q^*)$ about $\tilde{Q}^*$ to obtain

$$g(Q^*) = g(\tilde{Q}^*) + g'(\tilde{Q}^*)\Delta Q^* + \frac{g''(\tilde{Q}^*)}{2}(\Delta Q^*)^2 + \cdots \tag{8}$$

and from (5) we have

$$g'(Q^*) = -N(Q^*) - N(Q^*)N_> + \cdots \tag{9}$$

$$g''(Q^*) = -N'(Q^*) - N'(Q^*)N_> + N^2(Q^*) + \cdots, \tag{10}$$

where g', g'' denote differentiation with respect to Q^*, $\Delta Q^* = Q^* - \tilde{Q}^*$, and we have used $N'_> = dN_>/dQ^* = -N$. Inverting expansion (7) gives

$$MN_>(\tilde{Q}^*) = q\left[1 - \frac{q}{2M} + \cdots\right] \approx M\left(1 - e^{-\frac{q}{M}}\right). \tag{11}$$

We obtain from (5) and its derivatives with respect to Q^*:

$$g(\tilde{Q}^*) = \frac{q}{M}\left(1 - \frac{q}{2M}\right) + \frac{1}{2}\left(\frac{q}{M}\right)^2 + \cdots$$
$$= \frac{q}{M} + 0\left(\frac{q}{M}\right)^3, \tag{12}$$

which to relevant order is consistent with (6), and

$$g'(\tilde{Q}^*) = -N(\tilde{Q}^*)\left[1 + \frac{q}{M} + \cdots\right]. \tag{13}$$

For q finite as $M \to \infty$ this gives $g'(\tilde{Q}^*) = -N(\tilde{Q}^*)$ and $MN_>(\tilde{Q}^*) = q$.

We can now consider the extremal statistics of specific PDF $N(Q)$, and more importantly show that $P_m(Q^*)$ can be written in the universal form (1).

2.1 Gaussian and Exponential $N(Q)$

If $N(Q)$ falls off sufficiently fast in Q, i.e. is Gaussian or exponential, it is sufficient to consider lowest order only in (5) giving $g(Q^*) \sim N_>$ (Gumbel, 1958; Bouchaud and Mezard, 1997) and $q = MN_>(\tilde{Q}^*)$. Expanding (3) in Q^* near $\tilde{Q}^*$ gives to this order:

$$MN_>(Q^*) = M\int_{\tilde{Q}^*}^{\infty} N(Q)dQ - MN(\tilde{Q}^*)\Delta Q^*$$
$$= q\left[1 - \frac{MN(\tilde{Q}^*)}{q}\Delta Q^* + \cdots\right] \approx qe^{-M\frac{N(\tilde{Q}^*)}{q}\Delta Q^*} \tag{14}$$

Expanding $N(Q)$ about Q^* yields

$$N(Q^*) = N(\tilde{Q}^*)\left[1 + \frac{N'(\tilde{Q}^*)}{N(\tilde{Q}^*)}\Delta Q^* + \cdots\right] \approx N(\tilde{Q}^*)e^{\frac{N'(\tilde{Q}^*)}{N(\tilde{Q}^*)}\Delta Q^*}. \quad (15)$$

As to this order $(1 - N_>)^{M-1} \approx e^{-MN_>}$ we then have from (2)

$$P_m(Q^*) = MN(Q^*)(1 - N_>(Q^*))^{M-1} \approx MN(Q^*)e^{-MN_>} \sim (e^{u-e^u})^a, \quad (16)$$

with

$$a = -\frac{N'(\tilde{Q}^*)N_>(\tilde{Q}^*)}{N^2(\tilde{Q}^*)} \quad (17)$$

and

$$u = \ln\left(\frac{MN_>(\tilde{Q}^*)}{a}\right) - \frac{N(\tilde{Q}^*)}{N_>(\tilde{Q}^*)}\Delta Q^*. \quad (18)$$

Since throughout we are considering $\tilde{Q}^*$ to be large ($M \to \infty$, q finite), we have the effective value of a as that given by (17) in the limit $\tilde{Q}^* \to \infty$. For $N(Q)$ exponential the above gives $a = 1$. In the particular case of the exponential, all the summations, which in the above we have truncated, can be resummed exactly and give $a \equiv 1$, recovering the result of Bouchaud and Mezard (1997).

For $N(Q)$ Gaussian we cannot obtain a exactly in this way, but as we shall see it is instructive to make an estimate. Given $N(Q) = N_0 \exp(-\lambda Q^2)$ and expanding Eqs. (14), (15) and (16) to next order we obtain

$$P_m = \bar{P}_m e^{R(u)}$$
$$R = -\frac{\ln^2(q)}{4\lambda\tilde{Q}^{*2}} + \bar{u}\left(1 + \frac{2\ln(q)}{4\lambda\tilde{Q}^{*2}}\right) - \frac{\bar{u}^2}{4\lambda\tilde{Q}^{*2}} - e^{\bar{u}}, \quad (19)$$

where we have used $u = -2\lambda\tilde{Q}^*\Delta Q^*$ and $\bar{u} = u + \ln(q)$. To lowest order in $\Delta Q^*/\tilde{Q}^*$ (i.e. $\tilde{Q}^* \to \infty$) we have a universal PDF with $a = 1$, but to next order, i.e., neglecting only the term in $\bar{u}^2$ in (19), we have a universal distribution of form (1,16) with

$$a \equiv \left(1 + \frac{2\ln(q)}{4\lambda\tilde{Q}^{*2}}\right) \neq 1. \quad (20)$$

2.2 Power law $N(Q)$

The PDF of patches $N(Q)$ may, however, be a power law and in this case it will fall off sufficiently slowly with Q so that we need to go to next order, as in (7). If we consider a normalizable source PDF

$$N(Q) = \frac{N_0}{(1 + Q^2)^k}, \quad (21)$$

then for large Q ($Q \gg 1$) we have $N(Q) \sim N_0/Q^{2k}$ and then using (3) and (7)

$$\tilde{Q}^*N(\tilde{Q}^*) = (2k-1)N_>(\tilde{Q}^*) = (2k-1)\frac{q}{M}(1 - \frac{q}{2M}), \quad (22)$$

which with the above general expressions for $g(\tilde{Q}^*)$ and its derivatives substituted into (8) gives an expression for $g(Q^*)$

$$g(Q^*) = \frac{q}{M}\left[1 - (2k-1)\frac{\Delta Q^*}{\tilde{Q}^*} + k(2k-1)(\frac{\Delta Q^*}{\tilde{Q}^*})^2 \cdots\right]. \quad (23)$$

We also require an expression for $N(Q^*)$, again expanding about $\tilde{Q}^*$ and obtaining the derivatives of $N(\tilde{Q}^*)$ from those of $g(\tilde{Q}^*)$ and via (11) gives

$$N(Q^*) = N(\tilde{Q}^*)\left[1 - 2k\frac{\Delta Q^*}{\tilde{Q}^*} + k(2k+1)(\frac{\Delta Q^*}{\tilde{Q}^*})^2\right], \quad (24)$$

which can be rearranged as

$$N(Q^*) = N(\tilde{Q}^*)e^{\left[-2k\frac{\Delta Q^*}{\tilde{Q}^*} + k(\frac{\Delta Q^*}{\tilde{Q}^*})^2\right]}. \quad (25)$$

After some algebra (23) can be rearranged to give

$$Mg(Q^*) = qe^{\left[-(2k-1)\frac{\Delta Q^*}{\tilde{Q}^*} + \frac{2k-1}{2}(\frac{\Delta Q^*}{\tilde{Q}^*})^2\right]}. \quad (26)$$

These two expressions combine to finally give

$$P_m(Q) \equiv P_m(Q^*) \sim (e^{\bar{u} - e^{\bar{u}}})^a \quad (27)$$

with

$$\bar{u} = -\ln(a) - \ln(q) - (2k-1)\frac{\Delta Q^*}{\tilde{Q}^*}(1 - \frac{\Delta Q^*}{2\tilde{Q}^*}) \quad (28)$$

and

$$a = \frac{2k}{2k-1}. \quad (29)$$

To lowest order, neglecting the $(\Delta Q^*/\tilde{Q}^*)^2$ term (28) reduces to (18).

Hence, a power law PDF has maximal statistics $P_m(Q)$ which, when evaluated to next order, can be written in the form of a universal curve (i.e. of form (1,16)) with a correction that is non-negligible at the asymptotes. This can be seen (Jenkinson, 1955; Bouchaud and Potters, 2000) to be consistent with the well-known result due to Frechet, where (following the notation of Bouchaud and Potters, 2000) if we have PDF

$$N(x) \sim \frac{1}{|x|^{1+\mu}}, \quad (30)$$

then

$$N_> \sim \frac{1}{x^\mu} \quad (31)$$
$$P_m(x^*) = \frac{\mu}{(x^*)^{1+\mu}}e^{-\frac{1}{(x^*)^\mu}},$$

which we can write in the form

$$P_m(x^*) = \mu e^{\frac{\mu+1}{\mu}\ln(\frac{\mu+1}{\mu})}(e^{u-e^u})^a \tag{32}$$

$$u = -\mu\ln(x^*) - \ln\left(\frac{\mu+1}{\mu}\right), \tag{33}$$

which is of universal form (1,16) in u. Noting that here $\mu = 2k-1$ and $a = (\mu+1)/\mu$ and that to second order

$$\frac{\Delta Q^*}{\tilde{Q}^*}(1 - \frac{\Delta Q^*}{2\tilde{Q}^*}) = \ln\left(1 + \frac{\Delta Q^*}{\tilde{Q}^*}\right), \tag{34}$$

we simply identify $1 + \Delta Q^*/\tilde{Q}^*$ with $\tilde{x}^*$ to obtain (28). To next order in $\Delta Q^*/\tilde{Q}^*$ the analogue of (28) still yields the right-hand side of (34).

2.3 Convergence to the limiting distributions

The above results should be contrasted with the derivation of Fisher and Tippett (Fisher and Tippett, 1928). Central to Fisher and Tippett (1928) and later derivations is that a single ensemble of NM patches has the same statistics as the N ensembles (of M patches), of which it is comprised. The fixed point of the resulting functional equation (Bhavsar and Barrow, 1985) for arbitrarily large N and M is $a = 1$ for the exponential and Gaussian PDF, and the Frechet result for power law PDF. Here we consider a finite sized system so that although the number of realizable ensembles of the system can be taken to be arbitrarily large, the number of patches M per ensemble is always large but finite. More importantly, the rate of convergence with M depends on the PDF $N(Q)$. For an exponential or power law PDF we are able to resum the above expansion exactly to obtain a and convergence then just depends on terms $O(1/M)$ and above. This procedure is not possible for $N(Q)$ Gaussian, instead we consider the characteristic Q^*, that is $\tilde{Q}^*$ which for M to be arbitrarily large, Q^* should be large as well. Rearranging (7) to lowest order for $N(Q) = N_0\exp(-\lambda Q^2)$ yields $\sqrt{\lambda}\tilde{Q}^* \sim \sqrt{\ln(M)}$, implying significantly slower convergence. This is further discussed in Sornette (2000).

The extremal distributions are thus essentially a family of curves that are approximately of universal form (1,16) and are asymmetric with a handedness that just depends on the sign of Q; we have assumed Q to be positive, whereas one could choose Q to be negative, in which case $N(Q) \to N(|Q|)$. This would correspond to, say, power absorbed, rather than emitted, from a system. The single parameter a that distinguishes the extremal PDF then just depends on the PDF of the individual events. For $N(Q)$ exponential we then recover exactly the well-known result (Gumbel, 1958; Bouchaud and Mezard, 1997) $a = 1$. For a power law PDF a is determined by k via (29). We have also demonstrated that for a Gaussian PDF with finite but large M and N, that $a \neq 1$ and we will explore the significance of this in Sect. 3.1.

3 Normalization to the first two moments

To compare these curves with data we need $P(\bar{Q}) \equiv P_m(Q^*)$ in normalized form. This has moments

$$M_n = \int_{-\infty}^{\infty} y^n \bar{P}(y)dy, \tag{35}$$

which we will obtain as a function of a and then insist that $M_0 = 1$, $M_1 = 0$ and $M_2 = 1$.

Setting $M_1 = 0$ (and $M_0 = 1$, $M_2 = 1$) in our analysis of extremal distributions does not require any assumptions about the form of the PDF except that the moments exist. It will allow us to write the analytically obtained extremal distributions as functions of single parameter a.

3.1 Extremal distributions arising from Gaussian and exponential $N(Q)$

For Gaussian and exponential PDF we have

$$\bar{P}(y) = K(e^{u-e^u})^a \tag{36}$$

$$u = b(y-s). \tag{37}$$

This has moments which converge for all n. From Appendix A we have that the nth moment:

$$\begin{aligned} M_n &= \frac{1}{b}\int_{-\infty}^{\infty} \bar{P}(y)d\eta\frac{[\ln(a) + bs - \eta]^n}{b^n} \\ &= Ke^{-a\ln(a)}\frac{d^n}{da^n}\Gamma(a), \end{aligned} \tag{38}$$

where $\eta = \ln(a) - u$.

To normalize we insist that $M_0 = 1$, $M_1 = 0$ and $M_2 = 1$. The necessary integrals can be expressed in terms of derivatives of the Gamma function $\Gamma(a)$ (Gradshteyn and Ryzhik, 1980) and we obtain in Appendix A:

$$\begin{aligned} b^2 &= \Psi'(a) \\ K &= \frac{b}{\Gamma(a)}e^{a\ln(a)} \\ s &= -\frac{(\Psi(a) - \ln(a))}{b}, \end{aligned} \tag{39}$$

where

$$\Psi(a) = \frac{1}{\Gamma(a)}\frac{d\Gamma(a)}{da}$$

$$\Psi'(a) = \frac{d\Psi}{da}.$$

The ambiguity in the sign of b (and hence s) corresponds to the two solutions for $P(\bar{Q})$ for positive and negative Q.

We can now plot the curves, i.e., normalized to the first two moments and these are shown in Fig. 1. Experimental measurements of a global PDF $P(E)$ normalized to M_0 would be plotted M_2P versus $(E - M_1)/M_2$. In the main plot we show normalized distributions of the form (1,16) for $a = 1, \pi/2$ and 2. It is immediately apparent that the curves are difficult to distinguish over several decades in $\bar{P}(y)$ and thus in order to obtain a good estimate for a, the numerical or real experiments would require good statistics over a

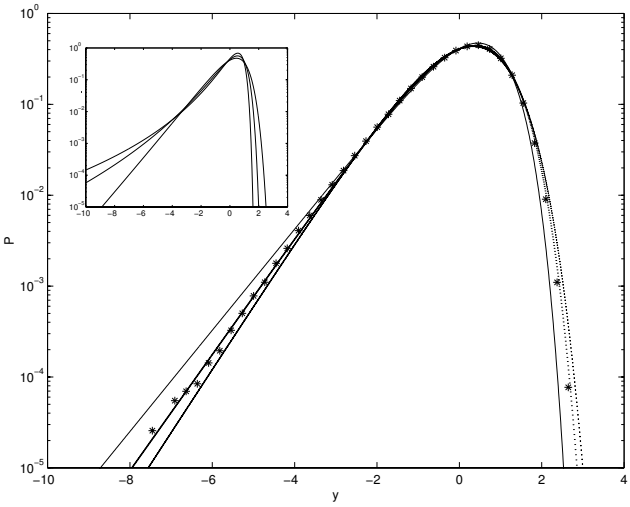


Fig. 1. Curves of the form (1) for $a = 1, \pi/2, 2$. Overlaid (*) are the numerically calculated extremal statistics of an uncorrelated Gaussian process (see text), and inset for comparison are Frechet curves plotted on the same scale (see Fig. 2).

dynamic range of about 4 decades, something which is not readily achievable.

In Fig. 1 we have also over plotted (*) the extremal PDF of ensembles of uncorrelated numbers that are Gaussian distributed, calculated numerically. We randomly select M uncorrelated variables Q_j, $j = 1, M$ and to specify the handedness of the extremum distribution, the Q_j are defined negative and $N(|\ Q\ |)$ is normally distributed. This would physically correspond to a system where the global quantity $\tilde{Q}$ is negative, i.e. power consumption in a turbulent fluid, as opposed to power generation. To construct the global PDF we generate T ensembles, that is select T samples of the largest negative number $Q_i^* = min\{Q_1..Q_M\}_i, i = 1, T$. For the data shown in the figure $M = 10^5$ and $T = 10^6$, this gives $\sqrt{\lambda}\tilde{Q}^* \sim \sqrt{\ln(M)} \simeq 3$ so that for the Gaussian we are far from the $a = 1$ limit (Fisher and Tippett, 1928). The numerically calculated PDF lies close to $a = \pi/2$. Such a value of a on these curves thus does not give direct evidence of a *correlated* process; in addition, it is necessary to establish that the data considered do not arise as the result of an extremal process.

Generally, plotting data in this way is an insensitive method for determining a and thus distinguishing the statistics of the underlying physical process. The question of interest is whether we can determine the form of the curve, and the value of a from data with a reasonable dynamic range; we address this question in Sect. 4.

3.2 Frechet distributions arising from power law $N(Q)$

For power law PDF (21) we use the Frechet distribution which we first write as:

$$P(Q^*) = K(e^{u-e^u})^a \tag{40}$$

$$u = \alpha + \beta \ln(1 + \frac{Q^*}{\tilde{Q}^*}), \tag{41}$$

which reduces to the form of (37) for $\Delta Q^*/\tilde{Q}^* \ll 1$. From (28), (21) and (33) we identify

$$\beta = -\mu = -(2k - 1). \tag{42}$$

The procedure of normalizing to the moments is only valid provided that they exist. For the power law PDF (21) we have (see also Bury, 1999):

$$M_n = \int_0^\infty \frac{Q^n H(Q) dQ}{(1 + Q^2)^k},$$

which converges for $Q \to 0$ and for $Q \to \infty$

$$M_n \sim \int^\infty \frac{Q^n H(Q) dQ}{Q^{2k}},$$

which if $H(Q) \to H_0$ as $Q \to \infty$

$$M_n \to \int^\infty \frac{dQ}{Q^{2k-n}} \simeq \frac{1}{Q^{2k-n-1}}\ |_{Q\to\infty},$$

which converges if $2k > n + 1$.

We now evaluate the moments. Again, we insist that $M_0 = 1$, $M_1 = 0$ and $M_2 = 1$ and in Appendix B obtain:

$$\alpha = -\beta \ln\left(\frac{a^{\frac{1}{\beta}}}{\Gamma(1 + 1/\beta)}\right)$$

$$K = \pm\beta a^a \left[\Gamma(1 + \frac{2}{\beta}) - \Gamma^2(1 + \frac{1}{\beta})\right]^{\frac{1}{2}} \tag{43}$$

$$\tilde{Q}^* = \frac{\Gamma(1 + \frac{1}{\beta})}{\left[\Gamma(1 + \frac{2}{\beta}) - \Gamma^2(1 + \frac{1}{\beta})\right]^{\frac{1}{2}}},$$

where $\beta = -(2k - 1)$. The normalization constants are thus also expressible as functions of $a = 2k/(2k - 1)$.

For convergence, these curves exist for power law of index $\infty > 2k > 3$ i.e. $1 < a < 3/2$. This is significant since processes exhibiting intermittency as a consequence of long-range correlations typically have k lower than this (Jensen, 1998), and we will consider alternative methods in Sect. 5.

In Fig. 2, we plot the normalized Fisher-Tippett type II or Frechet PDF for $k = 2, 5, 100$ and for comparison, the Fisher-Tippett type I ("Gumbel") PDF with $a = 1$. From (29) $a = 1$ corresponds to $k \to \infty$ and it is straightforward to demonstrate from the algebra that in this limit, the normalized Frechet PDF tends to Gumbel's asymptote $a = 1$. Hence, on this plot we see that for $k = 100$ these are indistinguishable, and differences between the Frechet and Gumbel PDF only appear on such a plot around the mean for $k < 3$ approximately. This demonstrates that these extremal curves arising from an uncorrelated Gaussian, exponential or power law $N(Q)$ will all be difficult to distinguish from the curve (1,16) with $a \neq 1$. We now consider more sensitive methods to determine a.

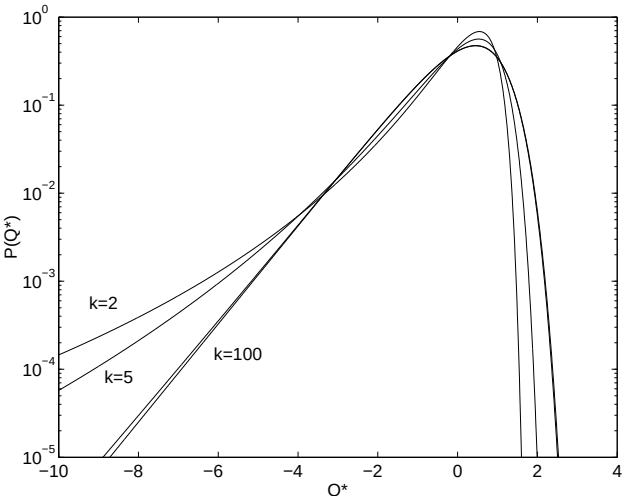


Fig. 2. Frechet PDF normalized to the first two moments for PDF $N(Q) = 1/(1+Q^2)^k$, $k = 2, 5, 100$.

4 Sensitive indicators of a; the mean and the third moment

The question of interest is whether we can determine a with sufficient accuracy from data with a reasonable dynamic range. We consider two possibilities here. First, a uniformly sampled process will have the most statistically significant values on the extremal curve near the peak, and in particular, from the figures we see that the Frechet distributions for small k will be most easily distinguished in this way. For the Frechet PDF the peak is at $u = 0$, i.e., it has coordinates

$$\bar{P}_m = \frac{K}{e^a} \quad \bar{y} = \tilde{Q}\left[e^{-\frac{\alpha}{\beta}} - 1\right] \tag{44}$$

on the normalized curve with K, $\tilde{Q}$, α, β known as functions of a from Appendix B. The coordinates of the peak of the PDF from the data plotted with $M_0 = 1$, $M_1 = 0$ and $M_2 = 1$ can thus be graphically inverted to give an estimate of a.

For PDF that represent a power law with large k, either exponential or Gaussian, we consider the normalized extremal PDF; then the coordinates of the maximum of $\bar{P}(y)$ is at $u = 0$, $y = s$, i.e.:

$$\bar{P}_m = \frac{K}{e^a} = \frac{\sqrt{\Psi'(a)}e^{-a(1-\ln(a))}}{\Gamma(a)} \tag{45}$$

with K, s from (A14). These can again be graphically inverted to obtain a; Fig. 3 shows $\bar{P}$ and $\bar{y}$ versus k for the Frechet PDF.

A more sensitive indicator may be the third moment of $\bar{P}$ of the curve (1,16), which, after some algebra (Appendix A), can be written as

$$M_3 = -\frac{\Psi''(a)}{(\Psi'(a))^{\frac{3}{2}}} \tag{46}$$

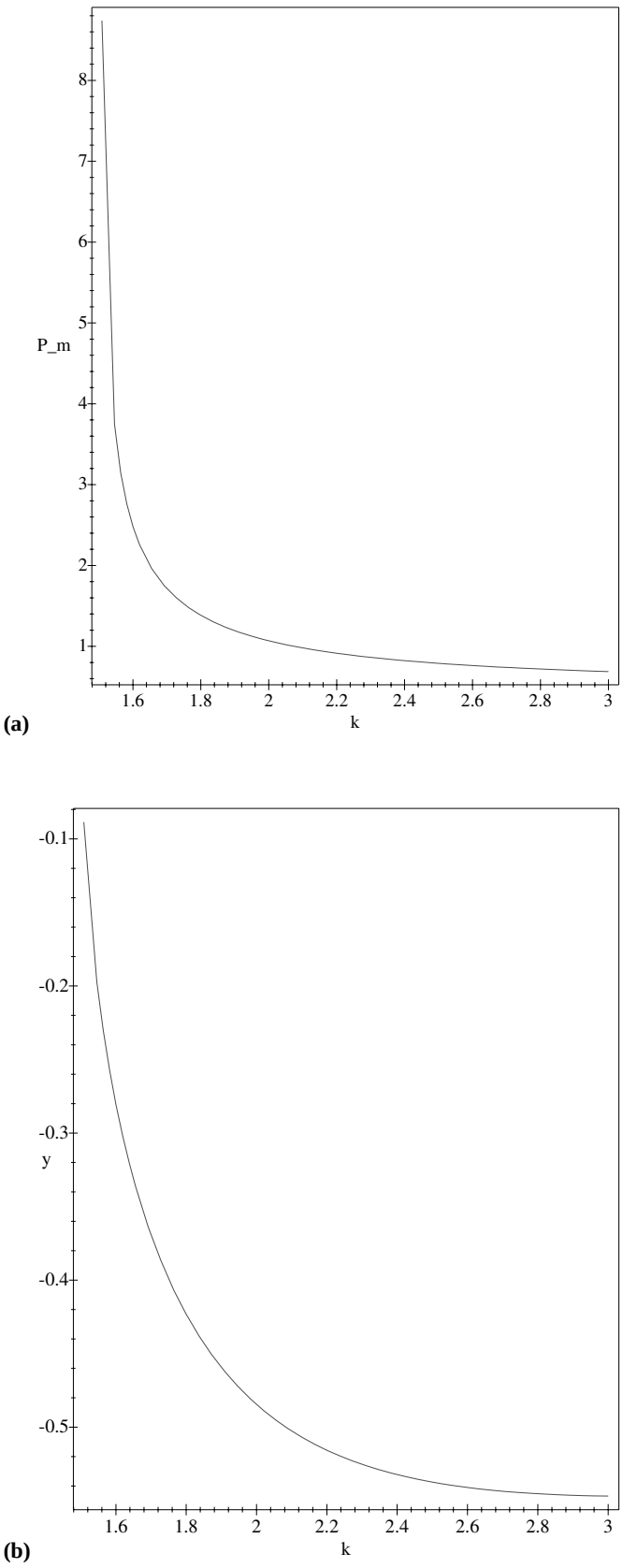


Fig. 3. The peak **(a)** and its location **(b)** as a function of k for Frechet curves.

for a Gaussian or exponential PDF, i.e. with (37) and

$$M_3 = \frac{\left[\Gamma(1+\frac{3}{\beta}) - 3\Gamma(1+\frac{2}{\beta})\Gamma(1+\frac{1}{\beta}) + 2\Gamma^3(1+\frac{1}{\beta})\right]}{\left[\Gamma(1+\frac{2}{\beta}) - \Gamma^2(1+\frac{1}{\beta})\right]^{\frac{3}{2}}} \quad (47)$$

for a power law PDF (Appendix B), i.e. with (41); the latter then converging for $k > 2$. Again, these refer to one of the two possible solutions for $P(\bar{Q})$; the other solution corresponding to $y \to -y$ ($Q^* \to -Q^*$) in Eqs. (37) and (reffrechu) which in turn gives $M_3 \to -M_3$.

The third moment is plotted versus a and k, respectively, in Fig. 4 for the Gumbel and Frechet curves. Inspection of Fig. 4 shows that over most of the range, M_3 is more sensitive than $\bar{P}$. For Frechet curves, M_3 only has convergence for relatively large k ($k > 2, a < 4/3$); for smaller k, $\bar{P}$ can distinguish the Frechet distributions ($k > 3/2, a < 3/2$ for convergence).

5 A method for small k

For $N(Q)$ power law, we can only use the properties of the normalized Frechet PDF above for $k > 3/2$. If k is smaller than this the second moment will not exist. We can, however, obtain a useful result for $k > 1$ by using the first moment only, i.e. by insisting $M_0 = 1, M_1 = 0$. We need another condition and can arbitrarily insist $P(u = 0) = 1$ (insisting that all the maxima of the Frechet PDF have the same height) which gives the condition

$$K e^{-a} = 1. \quad (48)$$

From B6 and B5

$$\frac{K\tilde{Q}^*}{\beta g^{1/\beta} a^a} = 1, \quad (49)$$

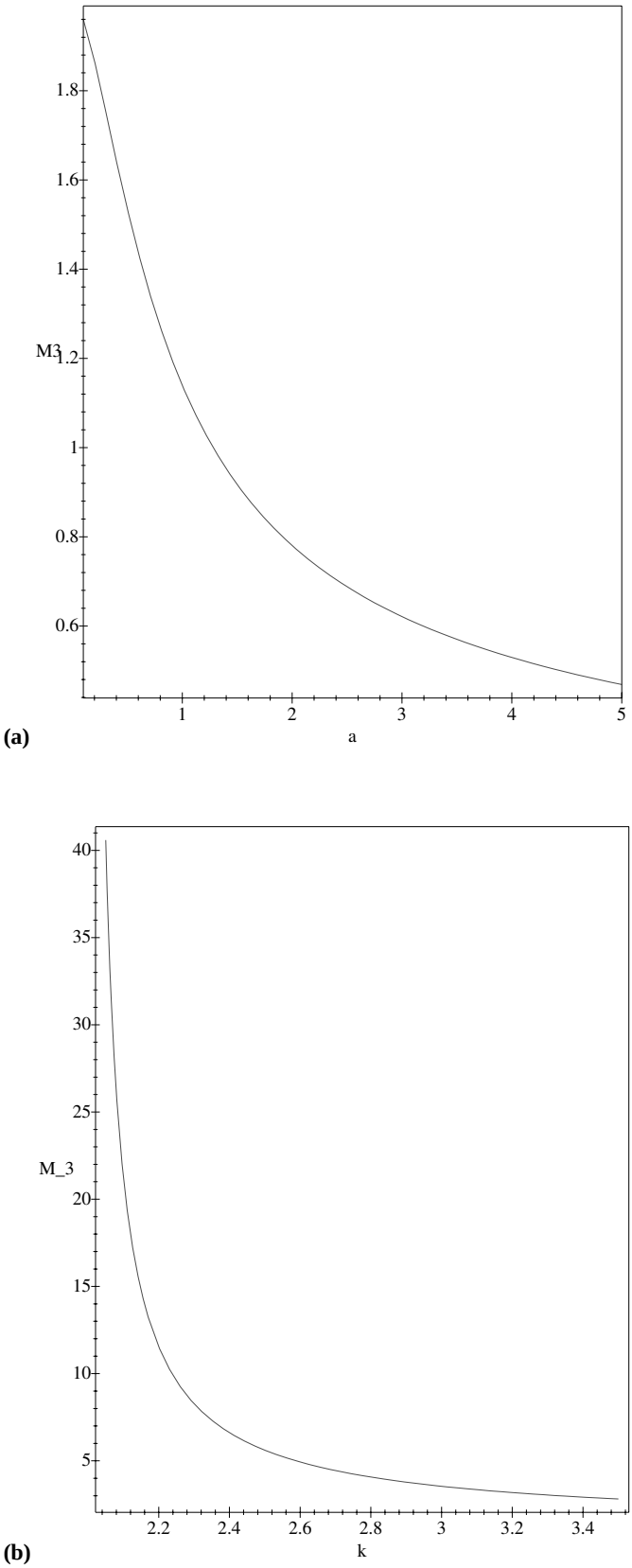
which, with $g^{1/\beta} = \Gamma(1 + 1/\beta)$ from Appendix B, gives $\tilde{Q}^*$ in terms of a and β (or k). Similarly, we use (B5); $g = ae^{\alpha}$ to obtain α in terms of a and β.

This then gives

$$P_m(Q^*) = K\left(e^{u-e^u}\right)^a,$$
$$u = \alpha + \beta\ln(1 + \frac{\Delta Q^*}{\tilde{Q}^*}),$$
$$\alpha = \beta\ln\left(\Gamma(1+\frac{1}{\beta})\right) - \ln(a),$$
$$\tilde{Q}^* = \beta e^{a(\ln(a)-1)},$$
$$K = e^a.$$

6 Conclusions

Recent work has suggested that the probability distribution of some global quantity, such as total power needed to drive rotors at a constant velocity in a turbulent fluid, or total magnetization in a ferromagnet slightly off the critical point, when



(a) **(b)**

Fig. 4. The third moment as a function of a for **(a)** curves of form (1) and **(b)** Frechet curves.

normalized to the first two moments, follows a non-Gaussian, universal curve. This curve is of the same form as that found from the extremal statistics of a process that falls off exponentially or faster at large values (i.e. Fisher-Tippett type I or "Gumbel"); whereas for an extremal process, the parameter specifying the curve $a = 1$, for the correlated processes $a > 1$.

In this paper, a framework has been developed to compare data with Fisher-Tippett type I ("Gumbel") and type II ("Frechet") asymptotes by obtaining the curves, and their normalizations, as a function of a single parameter a. We find:

1. The Fisher-Tippett type I and type II curves and their corresponding values of a are most easily distinguished by considering either the third moment, or the position of the peak, as functions of a, the functional forms for which are given here.

   For realistic ranges of data, simply comparing curves normalized to the first two moments, for example, in Bramwell et al. (1998, 2000), is insufficient to adequately distinguish either curves of the form of type I ("Gumbel") but with a values in the range $[1, 2]$, or most type II ("Frechet") curves.

2. Convergence to the limiting form of the extremal curve $a = 1$ (Gumbel's asymptote Fisher and Tippett (1928)) is sufficiently slow for an uncorrelated Gaussian such that for a large but realistic size of data set one obtains $a \approx \pi/2$. Data that falls on this curve is thus not sufficient to unambiguously distinguish a global observable of a system that has correlations (Bramwell et al., 1998, 2000), from that of an uncorrelated, extremal process.

Comparison with data is then facilitated in the following way. First, the data distribution is normalized to M_0 (to obtain the PDF $N(Q)$). Second, the data is plotted on semilog axes under the following normalization: $N(Q) \times M_2$ versus $(Q - M_1)/M_2$. Any Gaussian PDF on such a plot will fall on a single inverted parabola; similarly, any Gumbel (Fisher-Tippett I) process will fall on a single curve. Finally, M_3 is calculated for the data; we then can compare the data with an extremal process by inverting $M_3(a)$ obtained here for a Fisher-Tippett type I or II distribution. Overlaying these curves (augmented by other quantitative comparisons) then essentially constitutes a fitting procedure; but more importantly, in addition, the value of a is related to the underlying distribution, as we have discussed.

This and related techniques will have relevance, in particular, for regions where transport is dominated by turbulence, in the solar wind and magnetosphere in circumstances where multi-point and long time interval in situ measurements are difficult to obtain.

Acknowledgements. The authors would like to thank G. King, M. P. Freeman, D. Sornette and J. D. Barrow for illuminating discussions. SCC was supported by PPARC.

Appendix A Moments of the Gumbel distribution and the normalization b, K and s as a function of a

We consider a family of curves of the form

$$P(y) = Ke^{-au-ae^{-u}} \tag{A1}$$

with $u = b(y - s)$ where K, b, s, are constants to be derived as functions of a. We write

$$\eta = \ln a - b(y - s) = \ln a - u, \tag{A2}$$

then $ae^{-u} = e^{\eta}$ and $d\eta = -bdy$, and the nth moment is given by

$$M_n = \int_{-\infty}^{\infty} y^n P(y)dy = \frac{1}{b}\int_{-\infty}^{\infty} P(y)d\eta \frac{[\ln(a) + bs - \eta]^n}{b^n}. \tag{A3}$$

Then, using A2, we write $P(y)$ (A1) as

$$P(y) = K\, e^{-a(\ln(a)-\eta)-e^{\eta}} = \bar{K}e^{a\eta-e^{\eta}}, \tag{A4}$$

where $\bar{K} = Ke^{-a\ln(a)}$.

Now to within a constant we can write M_n as:

$$\tilde{M}_n = \int_{-\infty}^{\infty} \eta^n P(y)d\eta = \bar{K}\int_{-\infty}^{\infty} \eta^n e^{a\eta-e^{\eta}} d\eta, \tag{A5}$$

so that $M_0 = \tilde{M}_0/b$. Using the substitution $\tau = e^{\eta}$ A5 becomes

$$\tilde{M}_n = \bar{K}\int_0^{\infty} (\ln\tau)^n \tau^{a-1} e^{-\tau} d\tau = \bar{K}\frac{d^n}{da^n}\Gamma(a), \tag{A6}$$

where $\Gamma(a)$ is the Gamma function. Thus

$$\begin{aligned}
\tilde{M}_0 &= \bar{K}\Gamma(a) \\
\tilde{M}_1 &= \bar{K}\Gamma(a)\Psi(a) = \tilde{M}_0\Psi(a) \\
\tilde{M}_2 &= \bar{K}\Gamma(a)[\Psi^2(a) + \Psi'(a)] \\
&= \tilde{M}_0(\Psi^2(a) + \Psi'(a)),
\end{aligned} \tag{A7}$$

where

$$\Psi(a) = \frac{d\Gamma(a)}{da}\frac{1}{\Gamma(a)}.$$

We now insist that $M_0 = 1$, $M_1 = 0$ and $M_2 = 1$. Thus

$$M_0 = \frac{\tilde{M}_0}{b} = \frac{\bar{K}\Gamma(a)}{b} = 1 \tag{A8}$$

and

$$M_1 = 0 = \frac{1}{b^2}\int_{-\infty}^{\infty} P(y)d\eta[\ln(a) + bs - \eta]$$
$$= \frac{1}{b^2}\left[(\ln(a) + bs)\tilde{M}_0 - \tilde{M}_1\right], \tag{A9}$$

so

$$\frac{\tilde{M}_1}{\tilde{M}_0} = \ln(a) + bs = \Psi(a) \tag{A10}$$

from A7. Thus

$$bs = \Psi(a) - \ln(a). \tag{A11}$$

Also,

$$M_2 = 1 = \frac{1}{b^3}\int_{-\infty}^{\infty} P(y)d\eta[\ln(a) + bs - \eta]^2$$
$$= \frac{1}{b^3}\left[(\ln(a) + bs)^2\tilde{M}_0 - 2(\ln a + bs)\tilde{M}_1 + \tilde{M}_2\right], \tag{A12}$$

which, using A7 and A10, rearranges to give

$$M_2 = 1 = \frac{\tilde{M}_0}{b^3}\Psi'(a). \tag{A13}$$

This finally gives the normalisation of the universal curve

$$b^2 = \Psi'(a)$$
$$\bar{K} = \frac{b}{\Gamma(a)}$$
$$\text{that is } K = \frac{b}{\Gamma(a)}e^{a\ln(a)} \tag{A14}$$
$$s = \frac{(\Psi(a) - \ln(a))}{b}.$$

The above results will also yield an expression for the third moment in terms of a. Following A3 and A5 we have

$$M_3 = \frac{1}{b^4}\int_{-\infty}^{\infty} P(y)d\eta[\ln(a) + bs - \eta]^3$$
$$= \frac{1}{b^4}\Big[(\ln(a) + bs)^3\tilde{M}_0 - 3(\ln a + bs)^2$$
$$\tilde{M}_1 + 3(\ln(a) + bs)\tilde{M}_2 - \tilde{M}_3\Big]. \tag{A15}$$

Then A6 gives

$$\tilde{M}_3 =$$
$$\tilde{M}_0\left[\Psi(a)(\Psi^2(a) + \Psi'(a)) + 2\Psi(a)\Psi'(a) + \Psi''(a)\right] \tag{A16}$$

which, with A7 and A10, rearranges to give

$$M_3 = -\frac{\Psi''(a)}{(\Psi'(a))^{3/2}}. \tag{A17}$$

Appendix B Moments of the Frechet distribution and normalization as a function of a.

The moments of a Frechet distribution are obtained from Bury (1999). Here we wish to consider PDF of the form (19) which has extremum statistics

$$P_m(Q) = K(e^{u-e^u})^a, \tag{B1}$$

where, following (25–32), we write:

$$u = \alpha + \beta\ln(1 + \frac{Q}{\tilde{Q}}), \tag{B2}$$

where here we use the notations $Q \equiv \Delta Q^*$, $\tilde{Q} \equiv \tilde{Q}^*$, i.e. Q refers to extremal values. From (26), α and $\beta = (2k-1)$ are constants. We can then define the moments of $P_m(Q)$:

$$M_n = \int_{-\tilde{Q}}^{\infty} Q^n\, dQ\, P_m(Q) \tag{B3}$$

since from B2 $u \to \infty$ as $Q \to \infty$ and $u \to -\infty$ as $Q \to -\tilde{Q}$. Using the substitution $ae^u = \zeta$ we obtain after some algebra

$$M_n = \bar{K}\tilde{Q}^n\int_0^{\infty}((\frac{\zeta}{g})^{1/\beta} - 1)^n\, \zeta^{a-1+1/\beta}\, e^{-\zeta}\, d\zeta, \tag{B4}$$

where the constants

$$g = ae^{\alpha} \quad \text{and} \quad \bar{K} = \frac{K\tilde{Q}}{\beta g^{\frac{1}{\beta}}a^a}. \tag{B5}$$

By taking the expansion $u = \alpha + \beta Q/\tilde{Q}$ it is straightforward to verify that B4 yields the results from Appendix A. We now insist that $M_0 = 1$, $M_1 = 0$ and $M_2 = 1$. B4 then gives

$$M_0 = 1 = \bar{K}\Gamma(\bar{a}), \quad \text{where} \quad \bar{a} = a + 1/\beta \tag{B6}$$

and

$$M_1 = 0 = \bar{K}\tilde{Q}[\frac{\Gamma(\bar{a} + 1/\beta)}{g^{1/\beta}} - \Gamma(\bar{a})],$$

that is

$$\Gamma(\bar{a} + \frac{1}{\beta}) = g^{1/\beta}\Gamma(\bar{a}) \tag{B7}$$

and using B7 we have from B4:

$$M_2 = 1 = \bar{K}\tilde{Q}^2[\frac{\Gamma^2(\bar{a})\Gamma(\bar{a} + 2/\beta)}{\Gamma^2(\bar{a} + 1/\beta)} - \Gamma(\bar{a})],$$

that is

$$1 = \tilde{Q}^2[\frac{\Gamma(\bar{a})\Gamma(\bar{a} + 2/\beta)}{\Gamma^2(\bar{a} + 1/\beta)} - 1] \tag{B8}$$

using B6.

Now from the main text (27) $a = \frac{2k}{2k-1}$ and since

$$\beta = -(2k-1)$$
$$\bar{a} = a + 1/\beta = 1 \tag{B9}$$

and $\Gamma(\bar{a}) = \Gamma(1) = 1$.

B7 then gives $g^{1/\beta} = \Gamma(1 + 1/\beta)$. B8 then gives $\tilde{Q}$:

$$\tilde{Q} = \pm\frac{\Gamma(1 + \frac{1}{\beta})}{\left[\Gamma(1 + \frac{2}{\beta}) - \Gamma^2(1 + \frac{1}{\beta})\right]^{\frac{1}{2}}} \tag{B10}$$

then B7 gives K as

$$K = \pm\frac{\beta a^a\Gamma(1 + 1/\beta)}{\tilde{Q}} \tag{B11}$$

and since $g = ae^{\alpha}$, B6 gives an expression for α:

$$(ae^{\alpha})^{\frac{1}{\beta}} = \frac{K\tilde{Q}}{\beta a^a} \tag{B12}$$

that is:

$$\alpha = -\beta \ln \left(\frac{a^{\frac{1}{\beta}}}{\Gamma(1+1/\beta)} \right) \tag{B13}$$

which completes the normalization of B1,B2 as functions of k or a.

Using B7 we have from B4 an expression for the third moment:

$$M_3 = \bar{K}\tilde{Q}^3 \Big[\frac{\Gamma(\bar{a}+\frac{3}{\beta})\Gamma^3(\bar{a})}{\Gamma^3(\bar{a}+\frac{1}{\beta})} - \frac{3\Gamma(\bar{a}+\frac{2}{\beta})\Gamma^2(\bar{a})}{\Gamma^2(\bar{a}+\frac{1}{\beta})} + \frac{3\Gamma(\bar{a}+\frac{1}{\beta})\Gamma(\bar{a})}{\Gamma(\bar{a}+\frac{1}{\beta})} - \Gamma(\bar{a}) \Big]. \tag{B14}$$

Expansion in $1/\beta$ readily shows that to lowest order result A17 is recovered.

Then, using B9, B10 and B11, B13 can be rearranged to give $M_3(\beta)$, and hence, M_3 as a function of k or a:

$$M_3 = \frac{\left[\Gamma(1+\frac{3}{\beta}) - 3\Gamma(1+\frac{2}{\beta})\Gamma(1+\frac{1}{\beta}) + 2\Gamma^3(1+\frac{1}{\beta})\right]}{\left[\Gamma(1+\frac{2}{\beta}) - \Gamma^2(1+\frac{1}{\beta})\right]^{\frac{3}{2}}}. \tag{B15}$$

References

Aji, V., and Goldenfeld, N.: Fluctuations in finite critical and turbulent systems, Phys. Rev. Lett., 84, 1007–1010, 2001.

Bak, P.: How nature works: the science of self-organised criticality, Oxford University Press, Oxford, 1997.

Bhavsar, S. P. and Barrow, J. D.: First ranked galaxies in groups and clusters, Mon. Not. Roy. Astron. Soc., 213, 857–869, 1985.

Bohr, T., Jensen, M. H., Paladin, G., and Vulpiani, A.: Dynamical systems approach to turbulence, Cambridge University Press, Cambridge, 1998.

Bouchaud, J. P. and Mezard, M.: Universality classes for extreme value statistics, J. Phys. A, 30, 7997, 1997.

Bouchaud, J. P. and Potters, M.: Theory of financial risks: from statistical physics to risk management, Cambridge University Press, Cambridge, 2000.

Bramwell, S. T., Holdsworth, P. C. W., and Pinton, J. F.: Universality of rare fluctuations in turbulence and critical phenomena, Nature, 396, 552–554, 1998.

Bramwell, S. T., Christensen, K., Fortin, J. Y., et al.: Universal fluctuations in correlated systems, Phys. Rev. Lett., 84, 3744–3747, 2000.

Bramwell, S. T. et al.: Reply to comment on "universal fluctuations in correlated systems", Phys. Rev. Lett., 87, 18 8902, 2001.

Bury, K.: Statistical distributions in engineering, Cambridge University Press, Cambridge, 1999.

Cardy, J. E.: Scaling and renormalization in statistical physics, Cambridge University Press, Cambridge, 1996.

Chapman, S. C. and Watkins, N. W.: Avalanching and self-organised criticality: a paradigm for geomagnetic activity?, Space. Sci. Rev., 95, 293–307, 2001.

Fisher, R. A. and Tippett, L. H. C.: Limiting forms of the frequency distribution of the largest and smallest members of a sample, Proc. Camb. Phil. Soc., 24, 180–190, 1928.

Gradshteyn, I. S. and Ryzhik, I. M.: Table of integrals, series and products, Academic Press, 1980.

Gumbel, E. J.: Statistics of extremes, Columbia university press, New York, 1958.

Jenkinson, A. F.: The frequency distribution of the annual maximum (or minimum) values of meteorological elements, Q. J. Roy. Meteorological Soc., 81, 158–171, 1955.

Jensen, H. J.: Self-organised criticality: emergent complex behaviour in physical and biological systems, Cambridge University Press, Cambridge, 1998.

Labbe, R., Pinton, J. F., and Fauve, S.: Power fluctuations in turbulent swirling flows, J. Phys. II, France, 6, 1099–1110, 1996.

Lui, A. T. Y., Chapman, S. C., Liou, K., Newell, P. T., Meng, C. I., Brittnacher, M., and Parks, G. K.: Is the dynamic magnetosphere an avalanching system?, Geophys. Res. Lett., 27, 911–915, 2000.

Pinton, J. F., Holdsworth, P., and Labbe, R.: Power fluctuations in a closed turbulent shear flow, Phys. Rev. E, 60, R2452–R2455, 1999.

Sornette, D.: Critical phenomena in the natural sciences – chaos, fractals, selforganization and disorder: concepts and tools, Springer, Berlin, 2000.

Uritsky, V. M., Klimas, A. J., and Vassiliadis, D.: J. Geophys. Res., submitted, 2001.

Zheng, B., and Trimper, S.: Comment on "universal fluctuations in correlated systems", Phys. Rev. Lett., 87, 18 8901, 2001.