Pure appl. geophys. 162 (2005) 1187–1228
0033 – 4553/05/071187 – 42
DOI 10.1007/s00024-004-2666-3

Pure and Applied Geophysics

Approximating the Distribution of Pareto Sums

I. V. Zaliapin,[1] Y. Y. Kagan,[2] and F. P. Schoenberg[3]

Abstract—Heavy tailed random variables (rvs) have proven to be an essential element in modeling a wide variety of natural and human-induced processes, and the sums of heavy tailed rvs represent a particularly important construction in such models. Oriented toward both geophysical and statistical audiences, this paper discusses the appearance of the Pareto law in seismology and addresses the problem of the statistical approximation for the sums of independent rvs with common Pareto distribution $F(x) = 1 - x^{-\alpha}$ for $1/2 < \alpha < 2$. Such variables have infinite second moment which prevents one from using the Central Limit Theorem to solve the problem. This paper presents five approximation techniques for the Pareto sums and discusses their respective accuracy. The main focus is on the median and the upper and lower quantiles of the sum's distribution. Two of the proposed approximations are based on the Generalized Central Limit Theorem, which establishes the general limit for the sums of independent identically distributed rvs in terms of stable distributions; these approximations work well for large numbers of summands. Another approximation, which replaces the sum with its maximal summand, has less than 10% relative error for the upper quantiles when $\alpha < 1$. A more elaborate approach considers the two largest observations separately from the rest of the observations, and yields a relative error under 1% for the upper quantiles and less than 5% for the median. The last approximation is specially tailored for the lower quantiles, and involves reducing the non-Gaussian problem to its Gaussian equivalent; it too yields errors less than 1%. Approximation of the observed cumulative seismic moment in California illustrates developed methods.

Key words: Pareto distribution, Pareto truncated distribution, seismic moment distribution, stable distributions, approximation of Pareto sums.

1. Introduction

Statistical data analysis, a significant part of modern Earth Sciences research, is led by the intuition of researchers traditionally trained to think in terms of "averages," "means," and "standard deviations." Curiously, an essential part of relevant natural processes does not allow such an interpretation, and appropriate statistical models do not have finite values of these characteristics. Seismology

[1] Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California, 90095-1567, USA (e-mail: zal@ess.ucla.edu)
[2] Department of Earth and Space Sciences, University of California, Los Angeles, California, 90095-1567, USA (e-mail: ykagan@ucla.edu)
[3] Department of Statistics, University of California, Los Angeles, California, 90095-1554, USA (e-mail: frederic@stat.ucla.edu)

presents a superb example of such a situation with one of its fundamental laws, describing the distribution of the seismic moment (or energy) released in earthquakes as a power law, having both infinite mean and standard deviation. This paper addresses the problem of the statistical approximation for sums of power-law variables, which are used to describe the total seismic moment or energy released by multiple earthquakes, for example.

In describing natural physical systems, most variables should in principle have an upper limit, hence the study of tapered or truncated power-law distributions is important. However, in many cases one observes geophysical variables over relatively small scales, and in such cases distributions such as the Pareto may provide adequate representations and are considered crucial elementary tools for discussion. Moreover, in order properly to reject the fit of the pure Pareto distribution to seismic data sets, bounds on the sums of Pareto random variables are an important tool and hence worthy of investigation. In addition, sums of Pareto variables are of general concern in many other applications including finance, where the use of the pure Pareto distribution has not generally been invalidated. Further, whereas for truncated or tapered Pareto random variables classical statistical approaches provide reasonable estimates, for pure (untruncated, untapered) Pareto variables such evaluations are unavailable in most cases.

These are the reasons that in this work we mainly address the problem of evaluating sums of pure Pareto variables, and only in some limited cases do we consider the sums of tapered or truncated Pareto variables. A more complete and rigorous investigation of the transitional cases involving sums of tapered Pareto variables is a task for future work.

1.1 Pareto Distribution

Many natural and human-induced phenomena exhibit power-law behavior: for instance the power law is observed to approximate the distribution of sizes of earthquakes and volcanic eruptions, solar-flares, lightning strikes, river networks, forest fires, extinctions of biological species, war casualties, internet traffic, stock returns, insurance pay-offs, and cities (see *e.g*., MANDELBROT, 1983; RICHARDSON, 1960; TURCOTTE, 1997; BARTON and LA POINTE, 1995; SORNETTE, 2003; NEWMAN *et al*., 1994); this list can easily be extended. The power-law size distribution implies that the number $N(x)$ of objects of size larger than x decreases as a power of x:

$$N(x) \propto x^{-\alpha}, \quad \alpha > 0. \tag{1}$$

The power law is scale-invariant since the scale change $y = ax$ affects only the normalization constant in (1). Properly normalized, the law (1) is known as the *Pareto distribution*, its cumulative distribution function (cdf) $F(x)$ and probability distribution function (pdf) $f(x)$ are given by

$$F(x) = 1 - x^{-\alpha},$$
$$f(x) = \alpha x^{-1-\alpha}, \quad x > 1, \alpha > 0. \tag{2}$$

In many cases, the distribution (2) appears to fit well to the largest observations, $x > x_0$, but not for the entire sample. We describe such processes, where the survivor function $1 - F(x) \propto x^{-\alpha}$, as having power-law *tails*. This description is especially useful when describing and modeling processes with large deviations, a situation where one is primarily interested in the largest possible observations, and the specific distribution of the smaller observations may be neglected. A distribution that assigns a nonignorable probability to extremely large observations is called *heavy tailed*. One refers to a random variable (rv) as heavy tailed if it has an infinite second moment (infinite variation). For Pareto random variables, this corresponds to the case $0 < \alpha < 2$; for $\alpha \leq 1$ a Pareto rv has also infinite first moment (expectation).

1.2 Earthquake Size Distribution

The distribution of earthquake sizes is described by the well-known Gutenberg-Richter (GR) magnitude-frequency relation (GUTENBERG and RICHTER, 1941, 1944; SCHOLZ, 2002):

$$\log_{10}[N(m)] = a - bm, \quad b \approx 1, \tag{3}$$

where $N(m)$ is the annual number of earthquakes with magnitude equal or larger than m. Recalling that earthquake moment magnitude is related to the scalar seismic moment M (in NM) via (KANAMORI, 1977; BEN-ZION, 2003)

$$m = \frac{2}{3}\log_{10} M - c, \tag{4}$$

we see that the GR relation is a power law for the number $N(M)$ of earthquakes with seismic moment above M:

$$N(M) \propto M^{-\alpha}, \quad \alpha = \frac{2}{3}b. \tag{5}$$

In this paper we consider scalar seismic moment as the only quantitative measure of earthquake size. However, for historical and other reasons we sometimes convert the moment into moment magnitude, using $c = 6$ in Eq. (4). Thus in these cases the moment magnitude is used as a proxy for seismic moment.

Introducing the appropriate normalization, one obtains the Pareto pdf for seismic moments:

$$f(M) = \alpha M_t^{\alpha} M^{-1-\alpha}, \quad M_t \leq M, \tag{6}$$

where M_t is a catalog completeness threshold (or observational cutoff) and $\alpha \approx \frac{2}{3}$ is the index parameter of the distribution. In order for the total seismic moment to be finite, the distribution density must decay faster than M^{-2}. Thus simple considerations of the finiteness of the seismic moment flux or deformational energy available for earthquake generation (KNOPOFF and KAGAN, 1977) require the Pareto relation (6) to be modified for large seismic moments. This can be done, for example, by applying an exponential taper to the survivor function $1 - F(M)$, so that it takes the form

$$1 - F(M) = (M_t/M)^{\alpha} \exp\left(\frac{M_t - M}{M_c}\right), \quad M_t \leq M < \infty. \tag{7}$$

Here M_c is the *corner* moment, a parameter that primarily controls the distribution in the upper ranges of M (VERE-JONES *et al.*, 2001; KAGAN and SCHOENBERG, 2001; KAGAN, 2002a). To illustrate the above expressions, Figure 1 displays cumulative

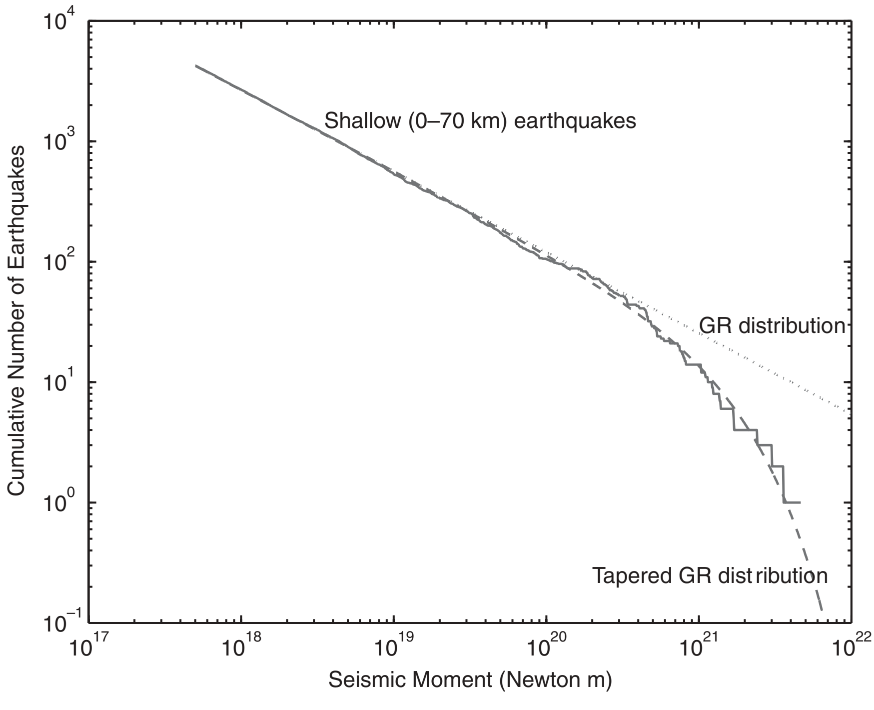


Figure 1

Number of earthquakes with seismic moment larger than or equal to M as a function of M for the shallow earthquakes in the Harvard catalog during 1977/1/1–2002/12/31. Power-law approximation (Gutenberg-Richter law, Eq. 5) is shown by dotted line. Tapered Gutenberg-Richter distribution (Eq. 7) which is the GR law restricted at large seismic moments by an exponential taper is shown by dashed line. The slope of the linear part of the curve corresponds to α-value 0.673 ± 0.011 and the corner moment $M_c = 1.6 \times 10^{21}$ Nm.

histograms for the scalar seismic moment of shallow earthquakes in the Harvard catalog (EKSTRÖM *et al.*, 2003) during 1977–2002. The curves display a scale-invariant (Pareto) segment (linear in the log-log plot) for small and moderate values of the seismic moment M. But for large M, the curve clearly bends downward.

Although the index parameter for the seismic moment distribution has been argued to have a universal value $\alpha \approx \frac{2}{3}$ at least for large ($m > 5.5$) earthquakes (Eqs. (3), (5); KAGAN, 2002a, b; BIRD and KAGAN, 2004), some geometrical variables which depend on seismic moments, such as slips during earthquakes, or widths and lengths of earthquake ruptures, appear to have power-law distributions with varying index values (KAGAN, 1994; ANDERSON and LUCO, 1983; WELLS and COPPERSMITH, 1994). Thus it is relevant to consider Pareto random variables with a broad range of indices, and in this work we consider the range $1/2 \leq \alpha < 2$.

For the reasons explained above, for index $\alpha \leq 1$ one is often interested in the case of the Pareto distribution with an upper bound or taper. MCCAFFREY (1997), SHEN-TU *et al.* (1998) and HOLT *et al.* (2000) considered the statistical bounds for the sums of upper-truncated Pareto rvs, for the case when several earthquakes approaching the maximum size are observed. In these cases the distribution of the sum can be approximated using the Central Limit Theorem (see more in Section 5). However, in actual earthquake catalogs these largest events are very rarely observed, so that bounds on the sums must often be estimates even when no earthquake of size approximately M_c is registered. Similar problems are encountered in the estimation of the parameters of the seismic moment-frequency law (7). KAGAN and SCHOENBERG (2001) and KAGAN (2002a) discuss the difficulty of estimating the corner moment with insufficient data.

1.3 Heavy-Tailed Sums

In many situations one is interested in *sums* of power-law distributed variables. For example, the total deformation of rock materials or of the Earth's surface due to earthquakes may be modeled as the sum of seismic moment tensors (KOSTROV, 1974; KAGAN, 2002b), the latter obeying the Pareto distribution; cumulative economic losses or casualties due to natural catastrophes are modeled as sums of power-law distributed variables (KAGAN, 1997; PISARENKO, 1998; RODKIN and PISARENKO, 2000); the total pay-off of an insurance company is modeled as the sum of individual pay-offs, each of which is distributed according to a power law, etc. The Pareto distribution is the simplest of the heavy-tailed distributions, thus the properties of Pareto sums are easier to study.

For $0 < \alpha < 2$ the second statistical moment of a Pareto rv is infinite which renders useless many of the conventional statistical techniques commonly used since the early 19th century. The most prominent of these is the Central Limit Theorem (FELLER, 1971, **2**, VIII, 4), which justifies approximating the sum of independent rvs by a Gaussian (normal) distribution. The sums of heavy tailed rvs cannot be approximated this way. However, such sums, when suitably normalized, typically approach a well-defined limiting distribution which depends on α. The entire family of such limits is known as *stable probability distributions with parameter* α (SAMORODNITSKY and TAQQU, 1994; UCHAIKIN and ZOLOTAREV, 1999; see also Section 2.1 below.) The Gaussian distribution is a special case corresponding to $\alpha = 2$.

Stable distributions, which (except for the Gaussian case) have power-law tails, recently became an object of intense mathematical and practical development (MANDELBROT, 1983; ZOLOTAREV, 1986; UCHAIKIN and ZOLOTAREV, 1999; NOLAN, 2005; RACHEV and MITTNIK, 2000; RACHEV, 2003). Their use is widespread in physics, finance, and other disciplines. The sums of large numbers of heavy-tailed rvs are investigated in many publications, especially in finance, and are well described in terms of stable distributions (MITTNIK *et al.*, 1998; RACHEV and MITTNIK, 2000; RACHEV, 2003; EMBRECHTS *et al.*, 1997). However, little is known about the distribution of the sum of an arbitrary (intermediate to small) number of Pareto rvs.

1.4 Approximation Problem

The aim of this study is to approximate the distribution of Pareto sums with arbitrary numbers of summands. Formally, let X_i, $i = 1, \dots, n$ be independent identically distributed (iid) rvs with a common Pareto distribution (2), and let S_n denote their sum

$$S_n = \sum_{i=1}^{n} X_i. \tag{8}$$

We seek to approximate certain quantiles $z_q(n)$ of S_n:

$$\text{Prob}\ \{S_n < z_q(n)\} = q, \quad 0 \leq q \leq 1,\, n > 1, \tag{9}$$

where Prob $\{A\}$ means the probability of event A. The *median* $z_{1/2}$ as well as *lower and upper bounds* z_q, z_{1-q}, for $q \ll 1$, are of special interest. As mentioned above, the problem becomes non-trivial for $0 < \alpha < 2$ when X_i have infinite second moments; for $0 < \alpha \leq 1$ these rvs also have infinite expectation.

The remainder of this paper is organized as follows. In Section 2 we construct two approximations for the sum (8) by using stable distributions. Stable distributions are discussed in more detail in Section 2.1; this discussion should help to provide a basic understanding of these complex distributions which sometimes display quite counter-intuitive behavior. Section 3 describes a different approach, involving the use of the distribution of the largest observation to approximate that of the entire sum (8). It is shown that for $\alpha < 1$ this seemingly crude replacement may result in fairly good approximations for the upper $(1 - q \geq 0.95)$ quantile. The results of Sections 2–3 are asymptotic; they give good approximations for the sums of a large number n of summands. However, unlike the Gaussian case, when "large" typically means $n > 30$, the sum of heavy-tailed variables sometimes converges very slowly, so that a sufficient number of summands allowing one to apply the asymptotics may in some cases be greater than $n = 10^4$. To address this problem, Section 4 describes two techniques specially tailored for approximating the sum S_n of an arbitrary number n of summands. Section 4.1 introduces an approach based on the analysis of order statistics to approximate the median and upper quantiles of S_n. Section 4.2 shows how to transform our non-Gaussian problem to a Gaussian equivalent and use the Central Limit Theorem to approximate the lower quantile. Approximation for the sum of truncated Pareto rvs is considered briefly in Section 5. The study's results are discussed in Section 6. We keep in the main text only the essential formulae; the detailed mathematical derivations are placed in Appendices.

2. *Approximating by Stable Distributions*

In this section we approximate Pareto sums with *stable distributions*, which is advocated by the Generalized Central Limit Theorem discussed in Section 2 below. We start with a brief discussion of stable distributions and their basic properties.

2.1 Univariate Stable Distributions

Stable distributions are a rich class of probability distributions that allow skewness (asymmetry), heavy tails and have many intriguing mathematical properties (ZOLOTAREV, 1986; SAMORODNITSKY and TAQQU, 1994; UCHAIKIN and ZOLOTAREV, 1999). A random continuous variable X is said to have a stable distribution if for any $n \geq 2$, there is a positive number C_n and a real number D_n such that

$$X_1 + X_2 + \cdots + X_n \stackrel{d}{=} C_n X + D_n, \tag{10}$$

where $X_1, X_2, \ldots, X_n$ are independent realizations of X, and $\stackrel{d}{=}$ means that the variables have the same distribution. This property makes stable distributions of special importance when working with sums of rvs. The lack of closed formulas for densities and distribution functions for all but a few stable distributions (Gaussian, Cauchy and Lévy) has been a major drawback to the use of stable distributions by practitioners. There are now reliable computer programs to compute stable densities, distribution functions and quantiles (NOLAN, 1997; MCCULLOCH and PANTON, 1997, 1998). With these programs, it is possible to use stable models in a variety of practical problems.

Different authors have provided several quite distinct representations for stable distributions (see Introduction in ZOLOTAREV, 1986; pp. 7–9 in SAMORODNITSKY and TAQQU, 1994; Section 3.6 in UCHAIKIN and ZOLOTAREV, 1999), which complicates the practical use of these distributions and often leads to confusion (HALL, 1981). Moreover, there are many misprints and other errors in published formulae for stable distributions (see, for instance, remark 2.12 in ZOLOTAREV, 1986). Below we use only the expressions which have been tested by comparing with tabulated values (McCULLOCH and PANTON, 1997, 1998).

Stable distributions have been proposed as a model for many types of physical and economic systems. There are several reasons for using a stable distribution to describe a system. The first is the presence of solid theoretical reasons for expecting non-Gaussian stable behavior, e.g. reflection of a rotating mirror yielding the Cauchy distribution, hitting times for a 1-D Brownian motion yielding the Lévy distribution, the gravitational field of stars yielding the Holtsmark 3-D distribution, or the stress distribution in a solid with defects following the Cauchy law (see FELLER, 1971 and UCHAIKIN and ZOLOTAREV, 1999 for other examples). The second reason is the Generalized Central Limit Theorem (see Section 2.2, Eq. (19)) which states that the only possible non-trivial limit of normalized sums of continuous iid terms is a stable law. Some observed quantities are posited to be the sum of many individual terms — the price movements of a stock, the noise in a communication system, etc., and hence a stable model may be used to describe such systems. A third argument for the use of stable distributions in modeling physical systems is empirical: many large data sets exhibit heavy tails and skewness. The strong empirical evidence for these features combined with the Generalized Central Limit Theorem is used by many to justify the use of stable models.

The univariate stable distribution is generally characterized by four parameters: its index α, the parameter β characterizing the degree of skewness, a scale parameter, and a shift parameter (SAMORODNITSKY and TAQQU, 1994). In this work we will consider only normalized stable distributions with the scale parameter equal to 1.0 and the shift parameter equal to zero. Moreover, since the Pareto variables we consider here are positive, the sums of these variables converge to a maximally asymmetrical (maximally-skewed) stable distribution, corresponding to $\beta = 1$. Thus, the stable variables considered below depend only on one parameter α, and their pdf and cdf are denoted as f_α and F_α, respectively.

In Figure 2 we show an example of the pdf for the stable distribution with index $\alpha = 2/3$ (15) and a corresponding Pareto distribution. The stable distributions with $\alpha < 1$ and $\beta = 1$ are concentrated on the positive x-axis, whereas maximally-skewed distributions with $\alpha \geq 1$ and $\beta = 1$ have support over the whole x-axis.

Recall that the standard Gaussian (normal) cdf with expectation μ and standard deviation σ is given by

$$\Phi(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) dy. \tag{11}$$

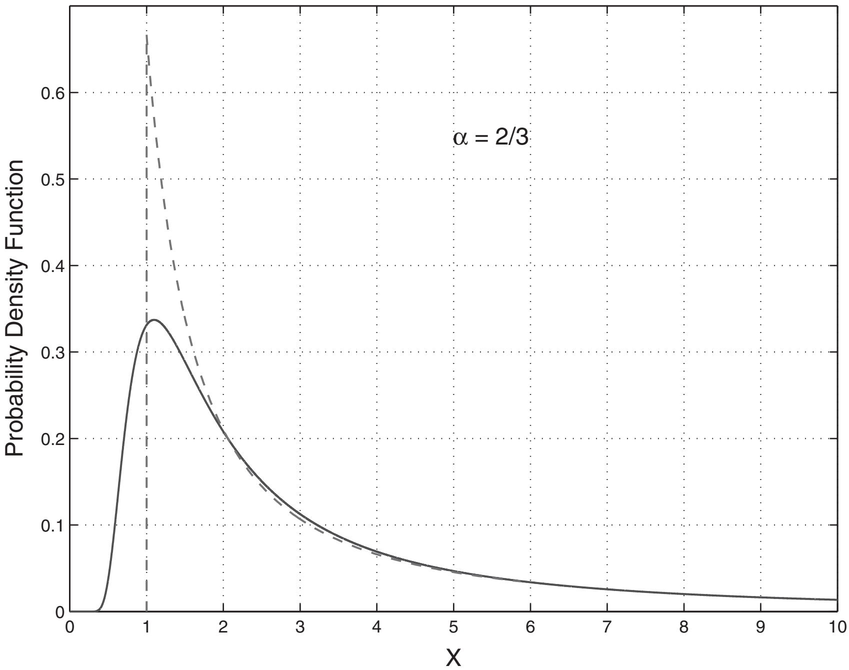


Figure 2
Probability density functions for the stable (Eq. 15, solid line) and Pareto (Eq. 2, dashed line) distributions, both with $\alpha = 2/3$

The Gaussian *stable* distribution, $\alpha = 2$, for which the skewness index β is not defined, is usually written in a slightly different form: $F_2(x) = \Phi(x; 0, 2)$ so its pdf becomes

$$f_2(x) = \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{x^2}{4}\right). \tag{12}$$

Only one of the stable maximally-skewed distributions, the Lévy, with $\alpha = 1/2$, can be expressed through elementary functions. This distribution has density

$$f_{1/2}(x) = \frac{1}{x\sqrt{2\pi x}} \exp\left(-\frac{1}{2x}\right), \tag{13}$$

and cdf

$$F_{1/2}(x) = 2\left[1 - \Phi\left(\frac{1}{\sqrt{x}}; 0, 1\right)\right], \tag{14}$$

where Φ is defined in (11).

Two more maximally-skewed distributions, for $\alpha = 2/3$ and $\alpha = 3/2$, can be expressed through special functions (Zolotarev, 1954):

$$f_{2/3}(x) = \frac{\sqrt{3}}{x\sqrt{\pi}} \exp\left(-\frac{16}{27x^2}\right) W_{1/2,1/6}\left(\frac{32}{27x^2}\right) \quad \text{for } x > 0, \tag{15}$$

where $W_{1/2,1/6}$ is a Whittaker function (GRADSHTEYN and RYZHIK, 1980, p. 1059). The Whittaker function $W_{k,\mu}(z)$ can be calculated using the confluent hypergeometric function $U(a, b, z)$ (WOLFRAM, 1999, pp. 770–771)

$$W_{k,\mu}(z) = \frac{z^{0.5+\mu} \mathrm{U}(0.5 - k + \mu, 1 + 2\mu, z)}{e^{0.5z}}. \tag{16}$$

Similar expressions for $f_{3/2}(x)$ are (ZOLOTAREV, 1954)

$$f_{3/2}(x) = -\frac{\sqrt{3}}{x\sqrt{\pi}} \exp\left(\frac{x^3}{27}\right) W_{1/2,1/6}\left(-\frac{2x^3}{27}\right) \quad \text{for } x < 0, \tag{17}$$

and

$$f_{3/2}(x) = \frac{1}{2x\sqrt{3\pi}} \exp\left(\frac{x^3}{27}\right) W_{-1/2,1/6}\left(\frac{2x^3}{27}\right) \quad \text{for } x > 0. \tag{18}$$

The cumulative functions for distributions (15), and (17, 18) can be obtained by numerical integration (WOLFRAM, 1999).

Though stable distributions usually cannot be summarized via elementary functions, their tails often allow fairly simple approximations given in Appendix A. These approximations may be useful for q close to 0 or 1.

2.2 Generalized Central Limit Theorem Approximation

Statistical inference relevant to the sum (8) can be made using the **Generalized Central Limit Theorem** (GCLT) (SAMORODNITSKY and TAQQU, 1994, p. 50; UCHAIKIN and ZOLOTAREV, 1999, p. 62), which states that the properly normalized sum S_n (8) of a large number n of iid Pareto rvs with common distribution (2) may be approximated by a stable distribution:

$$\lim_{n\to\infty} \text{Prob}\left\{\frac{S_n - b_n}{n^{1/\alpha} C_\alpha} < x\right\} = F_\alpha(x), \tag{19}$$

where $F_\alpha(x)$ is a stable cdf with index α. The normalization and shift coefficients are given by (see Appendix B)

$$C_\alpha = \begin{cases} [\Gamma(1-\alpha)\cos(\pi\alpha/2)]^{1/\alpha}, & \alpha \neq 1, \\ \pi/2, & \alpha = 1, \end{cases} \tag{20}$$

$$b_n = \begin{cases} 0, & 0 < \alpha < 1, \\ \frac{\pi n^2}{2} \int\limits_1^\infty \sin\left(\frac{2x}{\pi n}\right) dF(x), & \alpha = 1, \\ n\alpha/(\alpha - 1), & 1 < \alpha < 2, \end{cases} \tag{21}$$

where $\Gamma(x)$ is the gamma function. The integral in (21) for $\alpha = 1$ can be expanded as (GRADSHTEYN and RYZHIK, 1980, Eqs. 3.761.3, 8.230.2)

$$\begin{aligned} b_n &= n \log n + n \left[\frac{\pi n}{2} \sin\left(\frac{2}{\pi n}\right) - C - \log\frac{2}{\pi} - \int_0^{2/(\pi n)} \frac{\cos t - 1}{t} dt \right] \\ &\cong n \log n + n \left[1 - C - \log\frac{2}{\pi} \right], \end{aligned} \tag{22}$$

where

$$C \approx 0.5772\ldots \tag{23}$$

is the Euler constant.

It follows from the GCLT (19) that an arbitrary quantile z_q (9) of the sum S_n can be approximated as

$$z_q \approx z_q^{(1)} \equiv n^{1/\alpha} x_q C_\alpha + b_n, \tag{24}$$

where x_q solves the equation

$$F_\alpha(x_q) = q. \tag{25}$$

In some cases, the applicability of the approximation $z_q^{(1)}$ is seriously affected by the low convergence rate in the GCLT (19); we investigate the quality of this approximation in detail in Section 2.3 below.

An approximation for upper quantiles z_{1-q}, $q < 0.05$, can be obtained by noticing that the tail $1 - F_\alpha(x)$ of a stable distribution has a simple asymptotic (SAMORODNITSKY and TAQQU, 1994; UCHAIKIN and ZOLOTAREV, 1999):

$$\lim_{x\to\infty} x^\alpha (1 - F_\alpha) = C_\alpha^{-\alpha}, \tag{26}$$

where C_α is defined by (20). Applying the GCLT for the tail

$$\lim_{n\to\infty} \text{Prob}\left\{ \frac{S_n - b_n}{n^{1/\alpha} C_\alpha} > x \right\} = 1 - F_\alpha(x), \tag{27}$$

and equating the probability in (27) to q, one obtains

$$z_q^{(2)} \approx n^{1/\alpha} q^{-1/\alpha} + b_n, \tag{28}$$

where b_n is defined in (21), (22).

The approximation $z_q^{(2)}$ is easier to use than $z_q^{(1)}$ since the former is written in a closed form and does not involve solving the equation (25). This convenience is achieved at a price: $z_q^{(2)}$ only provides a satisfactory approximation for the upper quantiles, while $z_q^{(1)}$ is applicable for any value of q.

2.3 Evaluating Approximation Quality

The quality of the approximation $z_q^{(1)}$ (Eq. 24) may be evaluated by simulating the sums of Pareto rvs with different α-values. Note that the result depends essentially on the quality of the approximation (Eq. 19), so in assessing this approximation, we are simultaneously evaluating the convergence rate in the GCLT.

Simulation of Pareto rvs is especially easy since a synthetic realization may be constructed via

$$X_i = R^{-1/\alpha}, \tag{29}$$

where R is a rv uniformly distributed in the interval $(0, 1]$. For each fixed value of n and α we simulated $\geq 10^6$ realizations of the sum S_n to compute its quantiles z_q (9). We define an approximation's relative error as

$$\Delta_q^{(1)} = z_q^{(1)}/z_q - 1. \tag{30}$$

To abbreviate the notation, we do not explicitly note the dependence on n and α of the quantiles and their approximations and errors.

As mentioned previously, the approximation of S_n using the GCLT requires a very large number of summands. For instance, Figure 3 displays the quality of the approximation $z_q^{(1)}$ for $\alpha = 1/2$ and $n = 1000$. Even for 1000 summands, disagreements between the distribution of the sum and its approximation are easily discerned. This disagreement is even larger for smaller values of x (not shown in the figure). Indeed, the relative errors $\Delta_q^{(1)}$ for the upper quantiles approach the limit of zero rather quickly as the number n of summands increases, whereas for lower quantiles the convergence is quite slow. This is demonstrated in Figure 4, which shows the dependence of the relative error $\Delta_q^{(1)}$ for $q = 0.02$ and 0.98 on the number n of summands for $\alpha = 2/3$ and $\alpha = 1.5$. We follow McCulloch and Panton (1997) in evaluating the quantiles 0.02 and 0.98 rather than the more commonly used 0.025 and 0.975.

Such behavior is easy to understand in light of Figure 2. The upper tail of the Pareto distribution decays as $x^{-\alpha}$, similar to the decay of the upper tail of the stable distribution. On the other hand, the lower tail of the Pareto distribution has an abrupt truncation at $x = 1$ (see Eq. 2), whereas the stable distributions are smooth everywhere. Thus, a large number of variables must be summed in order for the lower quantiles of S_n to be reasonably approximated by those of F_α.

Table 1 collects the simulation results for selected values of α. Synthetic values of the sum are compared to stable quantiles as tabulated in McCulloch and Panton (1997), except for $\alpha = 2/3$ where we calculate F_α by integrating (15). The stable distribution approximates the Pareto sum upper 0.98 quantile quite well: only for α approaching 1.0 and 2.0 from below does the number of summands necessary to achieve 10% relative accuracy exceed two. As mentioned above, the approximation

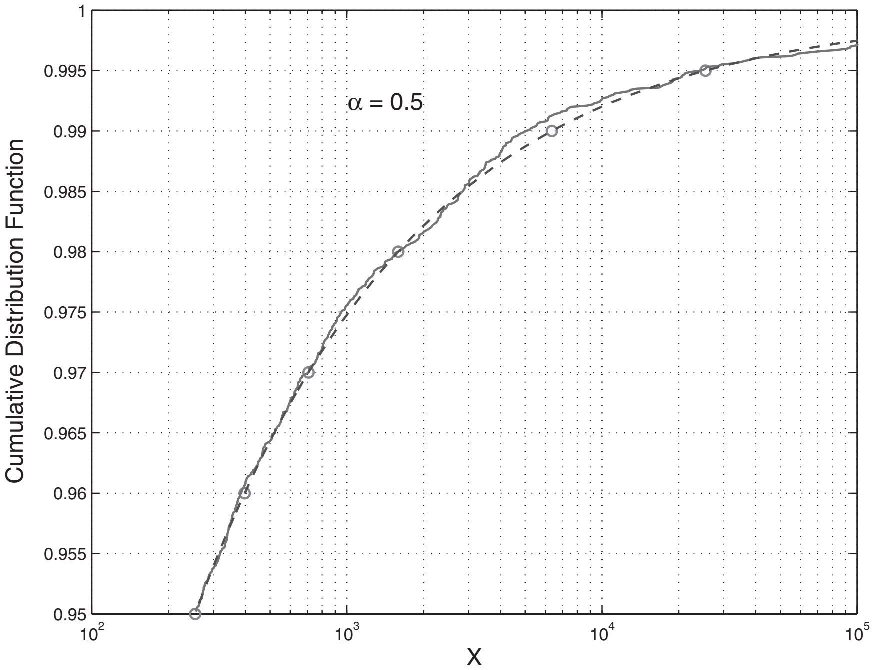


Figure 3
Quality of approximation $z_q^{(1)}$, Eq. (24) (stable distribution approximation). Solid line — cdf for simulated Pareto sum, $n = 1000$, dashed line — calculation by (14), circles — from McCulloch and Panton (1997) tables.

deteriorates substantially for the lower quantile: even in the best cases several dozens of summands are needed to yield $< 10\%$ relative error.

Table 2 (columns 6, 8, and 10) shows the relative error $\Delta_q^{(1)}$ of the approximation $z_q^{(1)}$ for selected values of α, q, and relatively small values of n. In each case S_n is calculated by using 10^7 simulated Pareto summands (thus $10^7 n$ values are drawn from the Pareto distribution). The error $\Delta_q^{(2)}$ of the approximation $z_q^{(2)}$ is given in the eleventh column.

3. Replacing the Sum with the Maximum

Approximation of the sums of Pareto rvs can be obtained by noticing that in many important cases the largest of n such observations, M_n, has the same order of magnitude as the entire sum S_n.

To gather some intuition, let us define

$$r_n = S_n / M_n. \tag{31}$$

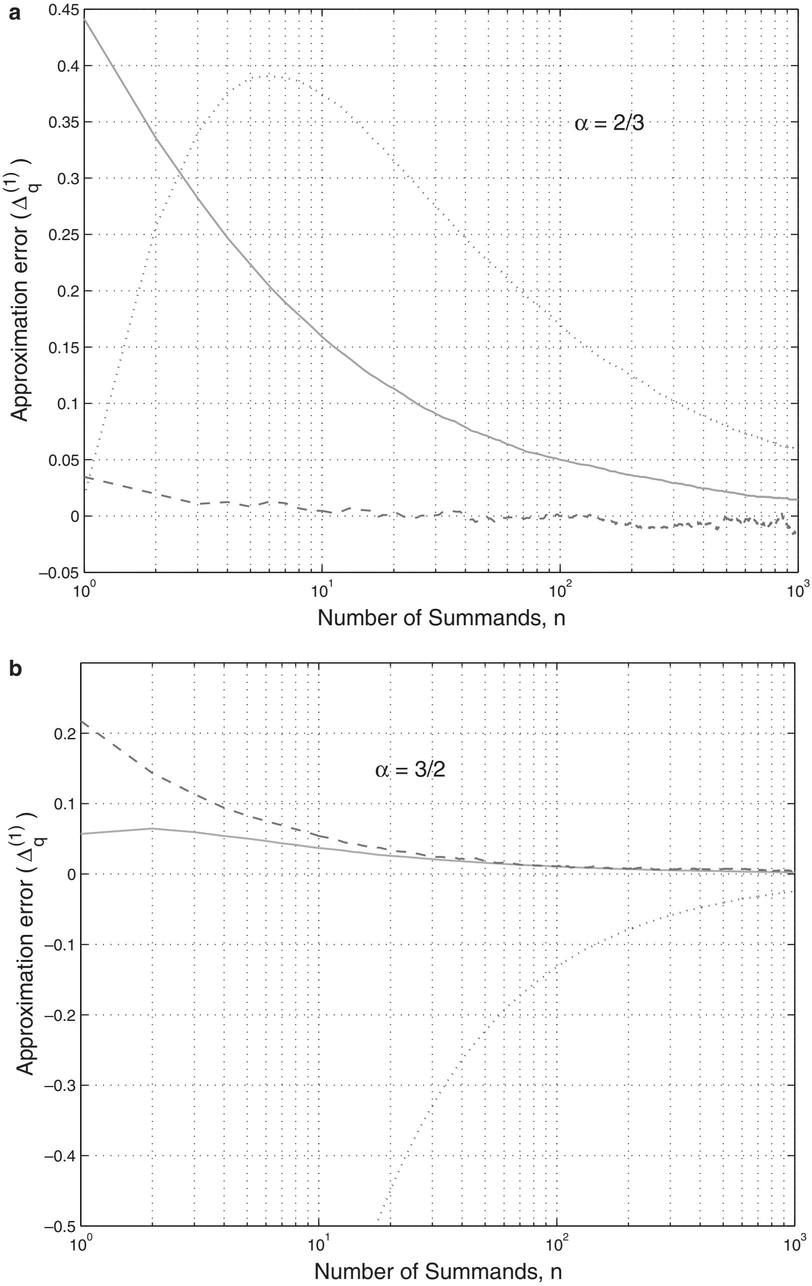


Figure 4

Relative error $\Delta_q^{(1)}$ of approximation $z_q^{(1)}$ (Eq. (24), stable distribution approximation). Dotted curve is for lower 0.02 quantile, solid curve for the median, and dashed curve for the 0.98 quantile (a) $\alpha = 2/3$, number of realizations $n = 3 \times 10^6$, F_α is obtained by integration of expression (15). (b) $\alpha = 3/2$, $n = 10^6$, F_α is taken from McCulloch and Panton (1997) tables.

Table 1

Number n of summands necessary to approximate the sum S_n by stable distributions ($z_q^{(1)}$) with error *$\Delta_q^{(1)}$ 0.1*

	Quantile		
α	0.02	0.50	0.98
0.50	20 ↑	2 ↑	2 ↓
0.60	25 ↓	3 ↓	2 ↓
0.66	270 ↓	18 ↓	2 ↓
2/3	320 ↓	25 ↓	2 ↓
0.80	> 10000 ↓	3000 ↓	2 ↓
0.90	>> 10000 ↓	> 10000 ↓	8 ↓
0.94	>> 10000 ↓	>> 10000 ↓	2000 ↓
0.98	>> 10000 ↓	>> 10000 ↓	> > 10000 ↓
1.00	55 ↑	2 ↓	2 ↓
1.02	65 ↑	2 ↓	2 ↓
1.10	75 ↑	2 ↓	2 ↓
1.20	85 ↑	2 ↓	2 ↓
1.30	100 ↑	2 ↓	2 ↓
1.50	140 ↑	2 ↓	3 ↓
1.66	230 ↑	2 ↓	8 ↓
1.80	300 ↑	3 ↓	30 ↓
1.90	1000 ↑	6 ↓	35 ↓
1.94	1800 ↑	7 ↓	500 ↓
1.98	6500 ↑	9 ↓	4000 ↓

↑ means that the value is approached from below; ↓ means that the value is approached from above.

In the case of iid Pareto summands, the expectation of the ratio r_n takes the form (see Appendix C)

$$E(r_n) = \begin{cases} \left[1 - nB\left(n, \alpha^{-1}\right)\right]/(1-\alpha), & \alpha \neq 1 \\ \sum_{k=1}^{n} 1/k, & \alpha = 1, \end{cases} \tag{32}$$

where $B(\cdot,\cdot)$ is the beta function.

For $\alpha = 2/3$ this expression can be simplified (ABRAMOWITZ and STEGUN, 1972, Eq. 6.1.12) as

$$E(r_n) = 3\left[1 - \frac{2^n n!}{(2n+1)!!}\right], \tag{33}$$

where we use the standard notation $(2n+1)!! = 1 \cdot 3 \cdot \ldots (2n+1)$. It is easy to derive the asymptotics of $E(r_n)$ as $n \to \infty$:

$$E(r_n) \cong \begin{cases} 1/(1-\alpha), & \alpha < 1 \\ n^{1-1/\alpha}\Gamma(1/\alpha)/(\alpha-1), & \alpha > 1 \\ C + \log(n), & \alpha = 1, \end{cases} \tag{34}$$

Table 2

Relative errors of approximations considered in the paper for selected values of a, n, and q

α	n	Simulated quantiles[1]			Approximation relative errors						
					$q = 0.02$		$q = 0.5$		$q = 0.98$		
		$q_{0.2}$	$q_{0.5}$	$q_{0.98}$	$\Delta^{(1)}_{0.02}$ (Eq. 24)	$\Delta^{(5)}_{0.02}$ (Eq. 64)	$\Delta^{(1)}_{0.5}$ (Eq. 24)	$\Delta^{(4)}_{0.5}$ (Eq. 55)	$\Delta^{(1)}_{0.98}$ (Eq. 24)	$\Delta^{(2)}_{0.98}$ (Eq. 28)	$\Delta^{(3)}_{0.98}$ (Eq. 37)	$\Delta^{(4)}_{0.98}$ (Eq. 56)
1/2	2	2.50	14.94	10000.17	−0.535	−0.002	−0.076	−0.001	0.000	0.000	−0.020	0.000
1/2	5	10.32	89.24	62436.65	−0.297	0.003	−0.033	0.037	0.001	0.001	−0.019	0.000
1/2	10	34.92	351.03	249338.20	−0.169	0.004	−0.016	0.057	0.002	0.003	−0.017	0.002
1/2	20	127.78	1392.25	1004949.00	−0.091	−0.002	−0.008	0.068	−0.005	−0.005	−0.025	−0.006
1/2	50	753.85	8654.80	6237558.00	−0.037	−0.008	−0.003	0.075	0.002	0.002	−0.018	0.001
1/2	100	2960.02	34628.25	25040967.00	−0.019	−0.002	−0.003	0.075	−0.002	−0.002	−0.022	−0.003
2/3	2	2.36	8.63	1012.34	0.257	−0.001	0.336	0.000	0.013	−0.012	−0.027	−0.001
2/3	5	8.44	37.29	4029.96	0.389	−0.003	0.221	0.017	0.006	−0.019	−0.034	−0.002
2/3	10	24.14	111.27	11406.04	0.373	0.005	0.158	0.023	0.006	−0.020	−0.035	−0.001
2/3	20	71.32	327.36	32489.57	0.314	0.003	0.113	0.027	−0.002	−0.027	−0.041	−0.007
2/3	50	302.14	1345.80	128040.70	0.226	0.001	0.070	0.026	0.001	−0.024	−0.038	−0.003
2/3	100	896.63	3882.27	363796.40	0.168	0.000	0.049	0.026	−0.003	−0.028	−0.043	−0.007
1	2	2.23	5.11	104.72	−1.658	0.001	−0.020	−0.001	0.043	−0.015	−0.055	−0.003
1	5	6.99	16.90	271.33	−0.870	0.000	0.012	0.007	0.023	−0.033	−0.088	0.003
1	10	17.28	40.49	555.47	−0.493	0.002	0.016	0.006	0.012	−0.043	−0.109	0.000
1	20	42.92	94.75	1127.16	−0.269	0.001	0.015	0.004	0.009	−0.044	−0.122	0.003
1	50	140.46	283.08	2888.23	−0.115	−0.002	0.011	0.001	0.001	−0.052	−0.143	−0.001
1	100	337.76	636.17	5851.76	−0.059	−0.001	0.008	−0.001	0.000	−0.052	−0.154	−0.001
3/2	2	2.15	3.67	24.01	−2.453	0.004	0.064	0.000	0.141	0.147	−0.109	0.000
3/2	5	6.21	10.60	50.39	−1.291	0.003	0.050	0.004	0.080	0.085	−0.218	0.001
3/2	10	14.13	23.00	88.03	−0.765	−0.004	0.037	0.002	0.051	0.056	−0.289	−0.001
3/2	20	32.03	48.98	153.80	−0.449	−0.002	0.026	−0.001	0.036	0.040	−0.354	0.001
3/2	50	92.64	130.02	326.42	−0.223	0.005	0.016	−0.003	0.020	0.024	−0.439	0.001
3/2	100	203.00	268.74	583.28	−0.132	0.005	0.010	−0.004	0.012	0.016	−0.502	0.002

[1]using 10^7 numerical simulations for each fixed α, n.

where C is given by (23). In Figure 5 we display dependence of $E(r_n)$ on the number n of summands using $\alpha = 2/3$. The asymptotic value $E(r_n) = 3$ is reached relatively slowly: for $n = 100$, $E(r_n) = 2.73$; and even for $n = 1000$, $E(r_n) = 2.92$.

For $\alpha < 1$ the coefficient r_n is on the order of unity which means that the major contribution to the sum S_n is made by the maximal observation M_n. It is important to note, however, that the variation of r_n may be significant. To illustrate the relation between S_n and M_n we show their scatterplot in Figure 6 for $\alpha = 2/3$, $n = 1000$. One can see that departures from the relation $S_n \sim M_n/(1-\alpha)$ suggested by (31, 34) are significant, especially at the lower and upper tails of S_n and M_n. Moreover, for any fixed value of S_n or M_n the ratio r_n varies substantially.

Figure 6 suggests the possibility of approximating the upper quantiles of S_n with those of M_n, since one clearly sees that the largest possible values of the sum lie close to the line $S_n = M_n$. Using the well-known distribution of the maximum (FELLER, 1971):

$$\text{Prob}\,\{M_n < x\} = F^n(x) = (1 - x^{-\alpha})^n, \tag{35}$$

one can approximate the distribution of the normalized sum as

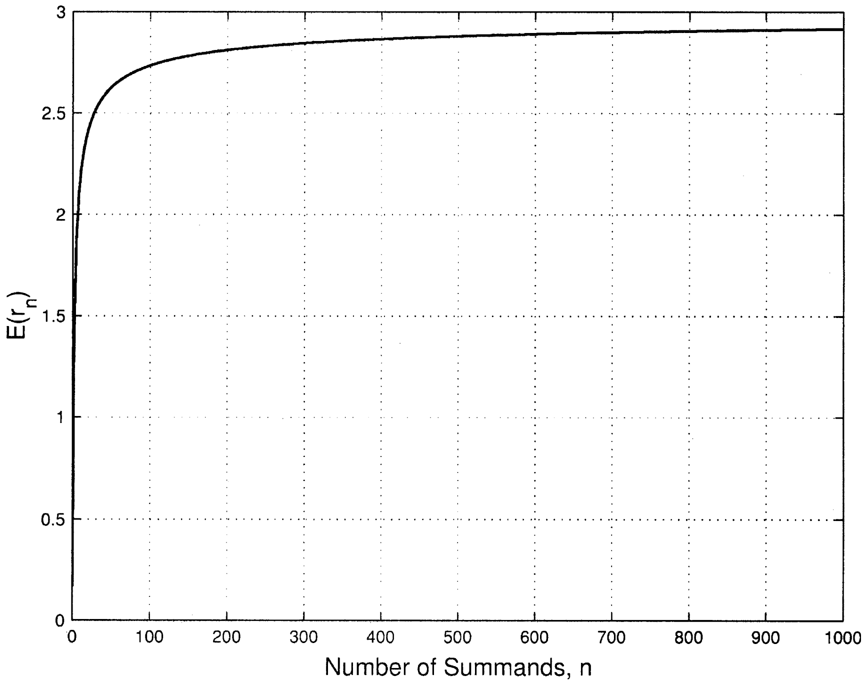


Figure 5
Dependence of $E(r_n)$ on the number n of Pareto summands with $\alpha = 2/3$.

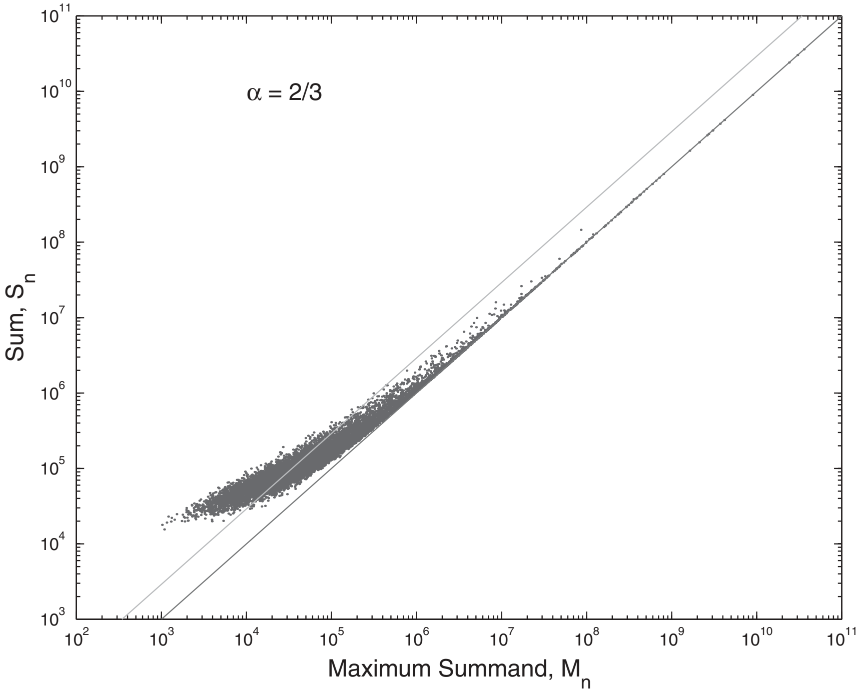


Figure 6
Value of the sum S_n vs. the maximal summand M_n for the Pareto distribution with $\alpha = 2/3$, the number of summands $n = 1000$. Lower line is $S_n = M_n$, the upper line is $S_n = M_n/(1-\alpha)$

$$\text{Prob}\left\{\frac{S_n}{C_\alpha n^{1/\alpha}} < z\right\} \approx \text{Prob}\left\{\frac{M_n}{C_\alpha n^{1/\alpha}} < z\right\} \approx \exp\{-(zC_\alpha)^{-\alpha}\}. \tag{36}$$

Equating the probability (36) to q and adding the shift b_n (21), which is important for $\alpha \geq 1$, we obtain

$$z_q^{(3)} = n^{1/\alpha}[\log(1/q)]^{-1/\alpha} + b_n. \tag{37}$$

The relative error $\Delta_q^{(3)}$ of this approximation is shown in column 12 of Table 2. One sees that for $\alpha < 1$ the approximation is quite close, whereas for $\alpha > 1$ the relative errors are large.

4. More Precise Techniques

This Section develops techniques that allow one to approximate the sums S_n with a higher degree of accuracy (relative error $< 1\%$) for an arbitrary number n of summands.

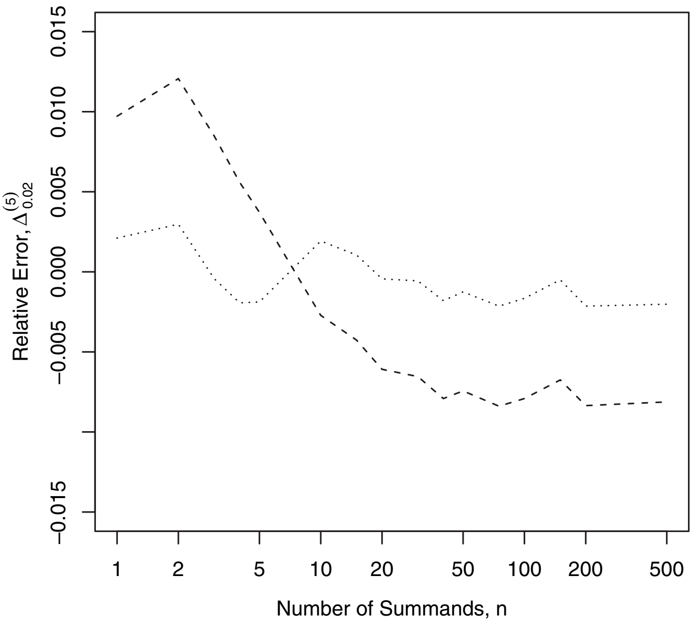


Figure 7
Relative errors $\Delta_q^{(5)}$ for the approximation $z_q^{(5)}$ with $p^* = q^{1/2}$ (dashed line) and p^* given by (64) (dotted line) as functions of the number of summands n, for $\alpha = 2/3$ and $q = 0.02$.

4.1 Approximating the Median and Upper Bound

Here we expand on the idea of using the largest observations to approximate the sum (8) of Pareto rvs. Our approach is based on considering the *variational series*

$$X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n} \tag{38}$$

formed by the *order statistics* $X_{1,n} = \min\{X_1, \ldots, X_n\}$, …, $X_{n,n} = \max\{X_1, \ldots, X_n\}$. Introducing the notations

$$S_{k,n} = \sum_{i=1}^{k} X_{i,n}, \;\; T_{k,n} = \sum_{i=k+1}^{n} X_{i,n}, \quad k \leq n \tag{39}$$

we may rewrite (8) as

$$S_n = \sum_{i=1}^{p} X_{i,n} + \sum_{i=p+1}^{n} X_{i,n} = S_{p,n} + T_{p,n}. \tag{40}$$

Essential properties of this representation follow from the following result.

Theorem 1. *The m-th moment $E(X_{k,n})^m$ of the k-th order statistic from Pareto distribution (2) is finite for $m < \alpha(n-k+1)$ and is given by*

$$EX_{k,n}^m = \frac{n!\ \Gamma(n-k+1-m/\alpha)}{(n-k)!\ \Gamma(n+1-m/\alpha)}. \tag{41}$$

The joint second moment $E(X_{r,n}X_{s,n})$, $r > s$ is finite if

$$\min\{n-r+1, (n-s+1)/2\} > 1/\alpha,$$

and is given by

$$E(X_{r,n}X_{s,n}) = \frac{n!\ \Gamma(n-r+1-1/\alpha)\Gamma(n-s+1-2/\alpha)}{(n-r)!\ \Gamma(n-s+1-1/\alpha)\Gamma(n+1-2/\alpha)}. \tag{42}$$

Proof. See Nevzorov (2001, Assignment 6.2.)

An important consequence of Theorem 1 is the following

Corollary 1. *If $p = n - \lfloor m/\alpha \rfloor$, where $\lfloor x \rfloor$ is the integer nearest to x from below, then all the summands in $S_{p,n}$ have finite m-th moment, and all the summands in $T_{p,n}$ have infinite m-th moment.*

As follows from this corollary, the number $n_m = \lfloor m/\alpha \rfloor$ of order statistics with infinite m-th moments does not depend on the sample size n. Seemingly counter-intuitive, this fact is well known in statistics (see Sen, 1959); its direct proof is given in Appendix D. Importantly, for $1/3 < \alpha < 1$ there are no more than two (upper) order statistics with infinite mathematical expectations, and for $2/3 < \alpha < 1$ there are only two (upper) order statistics with infinite second moment.

Below we use the representation (40) with $p = n - 2$ which gives $T_{n-2,n} = X_{n-1,n} + X_{n,n}$. Recalling that order statistics obey the Markov property (Nevzorov, 2001, Remark 4.3.)

$$\text{Prob}\ \{X_{k+1,n} < x | X_{1,n}, \ldots, X_{k,n}\} = \text{Prob}\ \{X_{k+1,n} < x | X_{k,n}\}, \tag{43}$$

and for the common distribution $F(x)$ of X_i

$$\text{Prob}\ \{X_{k+1,n} > x | X_{k,n} = u\} = \left(\frac{1-F(x)}{1-F(u)}\right)^{n-k}, \tag{44}$$

one can prove the following result.

Theorem 2. *The sum $T_{n-2,n}$ of the upper two order statistics of n Pareto rvs has the following distribution:*

$$\begin{aligned} T(x,\alpha,n) &= \text{Prob}\ \{X_{n-1,n} + X_{n,n} < x\} \\ &= n(n-1)\alpha^2 \int_2^x \int_1^{z/2} y^{-\alpha-1}(z-y)^{-\alpha-1}(1-y^{-\alpha})^{n-2}\, dy dz \end{aligned} \tag{45}$$

$$= n(n-1)\alpha^2 \sum_{k=0}^{n-2} (-1)^k C_{n-2}^k \int_2^x z^{-\alpha(k+2)-1} B(1/z, 1/2, -\alpha(k+1), -\alpha)\, dz, \qquad (46)$$

where $B(x_0, x_1, a, b)$ *is the generalized incomplete beta function* (WOLFRAM, 1999)

$$B(x_0, x_1, a, b) = \int_{x_0}^{x_1} t^{a-1}(1-t)^{b-1} dt. \qquad (47)$$

Proof. See Appendix E.

The Markov property (43) can be used as well to derive the distribution of a larger number of the upper order statistics. We restrict ourselves to the upper two order statistics and now proceed with the sum $S_{n-2,n}$ of the lower $(n-2)$ order statistics. Using (41), (42) we find that for $\alpha \geq 1/2$ the sum $S_{n-2,n}$ has finite expectation

$$m_1(\alpha, n) = \sum_{k=1}^{n-2} \frac{n!\,\Gamma(n-k+1-1/\alpha)}{(n-k)!\,\Gamma(n+1-1/\alpha)}. \qquad (48)$$

For $\alpha > 2/3$ it also has finite second moment

$$\begin{aligned} m_2(\alpha, n) = {} & \sum_{k=1}^{n-2} \frac{n!\,\Gamma(n-k+1-2/\alpha)}{(n-k)!\,\Gamma(n+1-2/\alpha)} \\ & + 2\sum_{r=2}^{n-2} \sum_{s=1}^{r-1} \frac{n!\,\Gamma(n-r+1-1/\alpha)\Gamma(n-s+1-2/\alpha)}{(n-r)!\,\Gamma(n-s+1-1/\alpha)\Gamma(n+1-2/\alpha)}. \qquad (49) \end{aligned}$$

For $\alpha = 1/2$ and 1 the expressions (48), (49) can be simplified:

$$m_1(1, n) = n\sum_{k=2}^{n-1} \frac{1}{k} = n(C - 1 + \log(n-1) + o(1)); \qquad (50)$$

$$m_1(1/2, n) = n(n-2); \qquad (51)$$

$$\begin{aligned} m_2(1, n) = {} & n(n-2) + 2n(n-1) \\ & \times \left[\log(n-2)(C - 1 + \log(n-2)) - \sum_{j=2}^{n-2} \frac{\log(j-1)}{j} + o(\log(n)) \right]. \qquad (52) \end{aligned}$$

In particular, it follows that for $\alpha = 1$, as $n \to \infty$,

$$\begin{aligned} m_1(1, n) &= n\log(n) + o(n\log(n)), \\ m_2(1, n) - m_1^2(1, n) &= \mathrm{Var}\,\left(S_{n-2,n}\right) = n^2[\log(n)]^2 + o\left\{n^2[\log(n)]^2\right\}. \end{aligned} \qquad (53)$$

To obtain bounds on the distribution of $S_{n-2,n}$, we note that its summands can be considered iid rvs with the common distribution

$$\text{Prob}\ \{Y < x\} = \ \text{Prob}\ \{X_i < x | X_i < X_{n-1,n}\}. \tag{54}$$

From Theorem 1 it follows that for $\alpha > 2/3$ they have finite first and second moments, and one may apply the Central Limit Theorem (FELLER, 1971, **2**, VIII, 4) to obtain

Corollary 2. *For $\alpha > 2/3$ the sum $S_{n-2,n}$ of the $(n-2)$ lower order statistics of n Pareto rvs converges in probability, as $n \to \infty$, to a normally distributed rv with first two moments $m_1(\alpha, n)$, $m_2(\alpha, n)$ given by (48), (49).*

Thus, to approximate the sum S_n we propose the representation (40) with $p = n - 2$. Then the distribution of $T_{n-2,n}$ is given by Theorem 2, while $S_{n-2,n}$ allows the normal approximation of Corollary 2. The median $z_{1/2}$ and the upper quantile z_q, $q > 0.95$ of S_n are approximated as

$$z_{1/2}^{(4)} = m_1(\alpha, n) + T^{-1}(1/2, \alpha, n); \tag{55}$$

$$z_q^{(4)} = m_1(\alpha, n) + \kappa(\alpha, n) + T^{-1}(q, \alpha, n), \tag{56}$$

where

$$\kappa(\alpha, n) = \begin{cases} \sqrt{m_2(\alpha, n) - m_1^2(\alpha, n)}, & \alpha > 2/3, \\ 0, & \alpha \leq 2/3. \end{cases} \tag{57}$$

Both of these approximations involve inversion of the distribution $T(x, \alpha, n)$, which must be done numerically. Numerical integration is more efficient using the representation (45) of Theorem 2. The relative errors $\Delta_q^{(4)}$ of these approximations are shown in columns 9 and 13 of Table 2. The approximations appear to be very close for all values of $\alpha < 2$, even for small numbers of summands. For $n = 2$, m_1 and m_2 become zero and approximations $z^{(4)}$ coincide with theoretical quantiles of the distribution T.

In constructing the approximations $z^{(4)}$ we obtain the exact quantiles for the sum of the upper two statistics $T_{n-2,n}$ by inverting its distribution T, and assume that, due to the normal asymptotic of Corollary 2, the quantiles of $S_{n-2,n}$, can be expressed via its first two moments. For instance, in (55) the median of S_{n-2} is naturally approximated by the mathematical expectation $m_1(\alpha, n)$. Indeed, such a construction is seriously affected by the fact that $S_{n-2,n}$ and $T_{n-2,n}$ are statistically dependent rvs, so the direct summation of their quantiles is a purely heuristic device. We notice though that for $\alpha < 1$ the major contribution to the sum S_n is made by the two upper statistics, $T_{n-2,n}$, while the contribution from the rest of the summands $(S_{n-2,n})$ becomes negligible, and so does their statistical variation. On the other hand, for $\alpha > 1$, the contribution from $T_{n-2,n}$ becomes less important compared to that of the large number of summands in $S_{p,n}$. In both cases, the statistical variation of one of the terms in (40) is negligible compared to the second one, consequently the direct summation of their quantiles results in a reasonable approximation.

4.2 Approximating the Lower Bound

An approximation for the q-quantile z_q (9) of the sum S_n, for small q, may be constructed based on the following idea. For any value y_n,

$$\begin{aligned}\text{Prob }\{S_n < z_q\} &= \text{Prob }\{S_n < z_q | M_n \le y_n\} \text{ Prob }\{M_n \le y_n\} \\ &+ \text{Prob }\{S_n < z_q | M_n > y_n\} \text{ Prob }\{M_n > y_n\}.\end{aligned} \tag{58}$$

The term Prob $\{S_n < z_q | M_n \le y_n\}$in (58) can be approximated using the Central Limit Theorem (FELLER, 1971, **2**, VIII, 4), since the variables being summed are now truncated and hence have finite moments. The resulting approximation will be close for sufficiently small values of y_n and sufficiently large n. At the same time, for any reasonably large value of y_n, and for small q, the quantity Prob $\{S_n < z_q | M_n > y_n\}$ in (58) will be infinitesimal; hence the entire final term in (58) may be considered negligible. Thus, in approximating z_q, we suggest choosing an appropriate value of y_n, and considering the approximation

$$\begin{aligned}\text{Prob }\{S_n < z_q\} &\approx \text{Prob}\{S_n < z_q | M_n \le y_n\} \text{ Prob }\{M_n \le y_n\} \\ &\approx \Phi\left(\frac{z_q - n\mu_{y_n}}{\sigma_{y_n}\sqrt{n}}; 0, 1\right) \text{ Prob}\{M_n \le y_n\},\end{aligned} \tag{59}$$

where Φ is the standard normal distribution function (11), and

$$\mu_y = E[X_1 | X_1 \le y] = \begin{cases} \frac{\alpha}{1-\alpha}\left(y^{1-\alpha} - 1\right)/(1 - y^{-\alpha}), & \alpha \ne 1, \\ \log(y)/\left(1 - y^{-1}\right), & \alpha = 1, \end{cases} \tag{60}$$

$$\sigma_y^2 = V[X_1 | X_1 \le y] = \begin{cases} \frac{\alpha}{2-\alpha}\left(y^{2-\alpha} - 1\right)/(1 - y^{-\alpha}) - \mu_y^2, & \alpha \ne 2, \\ 2\log(y)/\left(1 - y^{-2}\right) - \mu_y^2, & \alpha = 2 \end{cases} \tag{61}$$

are the conditional mean and variance of each summand, given the restriction on the maximum.

Note that there is a tradeoff in choosing y_n in (59): if one selects too small a value of y_n, then the term Prob $\{S_n < z_q | M_n > y_n\}$ in (58) is not negligible, thus the resulting approximation may not be satisfactory. On the other hand, if y_n is too large, then the approximation of Prob $\{S_n < z_q | M_n < y_n\}$ using the Central Limit Theorem may be unsatisfactory; this is particularly true for small n.

One option is to choose some value p^* to represent the probability Prob $\{S_n < z_q | M_n \le y_n\}$ in (59). Then for sufficiently small values of q as defined in (9), one has $q/p^* =$ Prob $\{M_n \le y_n\} = \left\{1 - (\frac{1}{y_n})^\alpha\right\}^n$, and solving this for y_n one obtains

$$y_n = \left[1 - (q/p^*)^{1/n}\right]^{-1/\alpha}. \tag{62}$$

The Central Limit Theorem approximation in (59) then yields

$$z_q^{(5)} = \sigma_{y_n}\sqrt{n}\Phi^{-1}(p^*) + n\mu_{y_n}, \tag{63}$$

where μ_{y_n} and σ_{y_n} are given by (60–61), and y_n is given by (62).

A naive choice for p^* is $\sqrt{q}$; this seems to balance the aforementioned tradeoff, since then Prob $\left\{S_n < z_q | M_n \le y_n\right\} =$ Prob $\{M_n \le y_n\} = \sqrt{q}$. Using simulations, one may obtain approximations of various quantiles and determine, for each such simulated quantile, the optimal choice of p^*. We found that inspection of plots of optimal choices of p^* versus q, n, and α (not shown) revealed nearly quadratic variation with q, and approximately threshold relationships with n and α, suggesting the following formula as a possible choice of p^*:

$$p^* = 0.136 + 0.235q + q^2 + 0.0066\min(n, 10) - 0.05\max(\alpha, 1). \tag{64}$$

Equation (64) was derived essentially by trial-and-error and is not meant to reflect an optimal choice of p^*, but rather a simple summary of the relationships observed between p^*, q, n, and α which may be useful in practice.

The dotted curve in Figure 7 shows the relative errors, as a function of n, for the resulting approximation $z_q^{(5)}$ with p^* given by (64). For comparison, the dashed curve in Figure 7 shows relative errors for the solution to (63) with $p^* = \sqrt{q}$. Here $\alpha = 2/3$ and $q = 0.02$; the results for other values of α and other small values of q are similar.

The values reported in column 7 of Table 2 reflect the approximation $z_q^{(5)}$ with p^* as in (64). One sees that the approximation matches the true quantile quite closely, with relative errors consistently less than 1%, even for small n.

5. *Cumulative Pareto Sums: Linear vs. Nonlinear Regimes*

As mentioned in Section 1.2, it is important to consider the case of the Pareto distribution with an upper bound or taper, especially for $\alpha \le 1.0$. An example of such a distribution is shown in Figure 1, which demonstrates that an exponential taper applied to the cdf of the Pareto distribution results in a close approximation to the empirical distribution of seismic moments. A commonly employed alternative is to use a Pareto distribution that is simply truncated at some value y; the corresponding cdf is

$$F(x) = \frac{1 - x^{-\alpha}}{1 - y^{-\alpha}}, \quad 1 < x \le y. \tag{65}$$

For the truncated Pareto distribution, simulated values may be constructed via

$$X_i = \{R[1 - y^{-\alpha}] + y^{-\alpha}\}^{-1/\alpha}, \tag{66}$$

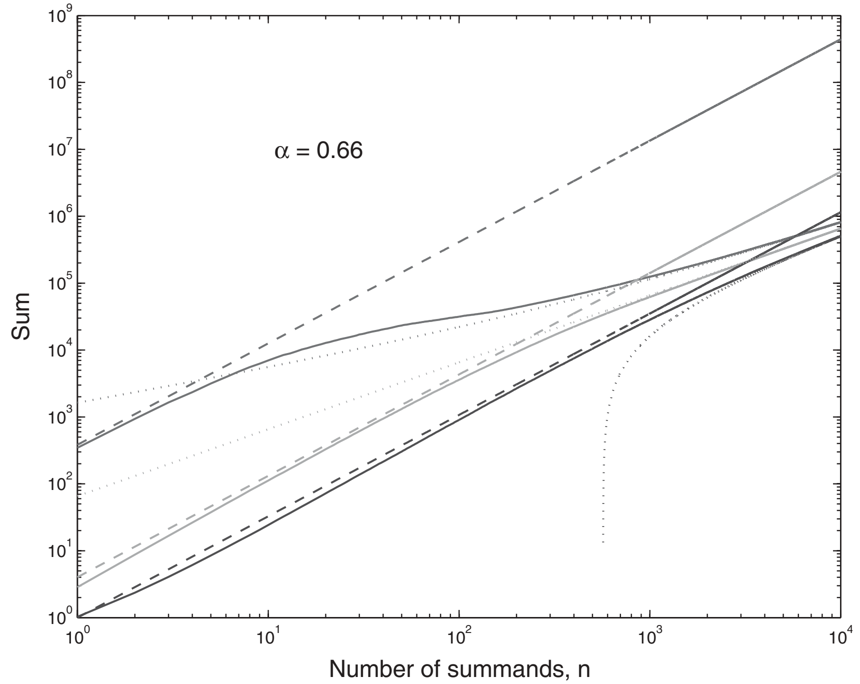


Figure 8
Quantiles for the sum S_n of truncated Pareto variables (upper limit $y = 3.4 \times 10^4$, Eq. 65) and their approximations as functions of the number of summands, n. Two approximations are considered: via the stable distribution, Eq. (24) (dashed lines) and Gaussian, Eq. (67) (dotted lines). Solid lines represent quantiles of simulated Pareto sums. The upper three curves are for the 0.98 quantile, middle three curves are for the median, and lower three curves are for 0.02 quantile.

where R is a uniform rv as in (29). The ratio $E(r_n) = E(S_n/M_n)$ for the truncated Pareto distribution and $\alpha < 1$ can be evaluated as in (32), and is given in Appendix C, Eq. (94).

Figure 8 displays an example of simulated sums (S_n) for the Pareto distribution with $\alpha = 0.66$ truncated at $y = 3.4 \times 10^4$ compared to the stable distribution quantiles. The upper quantiles depart from the theoretical curve for the stable distribution starting with $n = 2$, whereas the behavior of the lower quantile is essentially unaffected by the truncation until n exceeds 10^3.

When the number of summands is large, the truncation point y dominates the behavior of the quantiles. The sum is then distributed asymptotically according to the Gaussian law:

$$\lim_{n\to\infty} F_{S_n}(x) = \Phi\left(\frac{x - n\mu_y}{\sigma_y\sqrt{n}}; 0, 1\right), \tag{67}$$

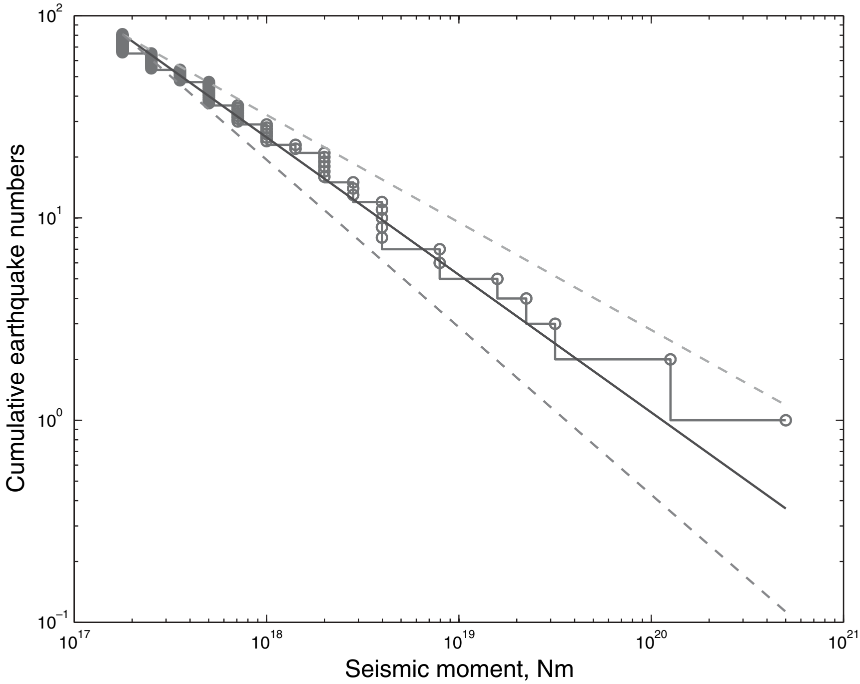


Figure 9

Number of earthquakes (shown by circles) with seismic moment larger than or equal to M as a function of M for the earthquakes in the TOPPOZADA *et al.* (2000) catalog. Latitude limits 37.0° N – 39.0°N, magnitude threshold 5.5, the total number of events 81. Solid line - - - approximation by the Pareto distribution, $\alpha = 0.680 \pm 0.076$, dashed lines 95% confidence limits (AKI, 1965), conditioned by the total number of earthquakes observed. These limits only reflect the uncertainty in the estimation of α (not in the number of observed earthquakes).

where Φ is the normal cdf (11), and the parameters μ_y and σ_y are given by (60), (61). In Figure 8 the Gaussian approximation for the truncated Pareto sums is shown to provide a satisfactory approximation to the sum of truncated Pareto rvs, when the number n of terms being summed exceeds 500.

The median of the distribution of S_n for truncated Pareto rvs increases non-linearly with n: for $n < 100$ the median increases at a rate proportional to $n^{1/\alpha}$ (PISARENKO, 1998; RODKIN and PISARENKO, 2000; HUILLET and RAYNAUD, 2001). This behavior differs sharply from the linear increase of all quantiles of the distributions of sums of rvs with finite first statistical moment (i.e., for distributions with $\alpha \geq 1$). For larger numbers of summands, quantiles of sums of truncated Pareto distribution increase linearly.

RODKIN and PISARENKO (2000, their Eqs. (19) and (21)) introduce two particular values of n:

$$n_1 = \frac{1-\alpha}{\alpha} \times y^a \log 2, \tag{68}$$

and

$$n_2 = \frac{9(1-\alpha)^2 y^a}{\alpha(2-\alpha)}. \tag{69}$$

They argue that if $n < n_1$ the sums behave statistically as if the distribution has not been truncated, whereas for $n > n_2$ the Gaussian approximation for the sums is appropriate. For the example displayed in Figure 8, $n_1 = 350$ and $n_2 = 1152$. The curves in Figure 8 appear to confirm the validity of the conclusions of RODKIN and PISARENKO (2000) for the lower quantiles.

6. Discussion

We presented five approximations for the quantiles of the distribution of the sums of Pareto variables. The approximations $z_q^{(1,2)}$ (Eqs. 24, 28) are based on stable asymptotics for Pareto sums; $z_q^{(3)}$ (Eq. 37) uses replaces the sum S_n with its maximal summand; $z_q^{(4)}$ (Eqs. 55–56) elaborates the latter idea considering separately the sum of the largest two observations and the rest of the lower terms to approximate the median and the upper quantiles; and $z_q^{(5)}$ (Eq. 63) approximates the lower quantiles by introducing an upper truncation point on the maximum observed value. We consider the relative error $\Delta_q^{(k)}$ (30) for all our approximations: Table 2 collects the errors for selected values of α, n, and q.

The accuracy of the estimates given in Table 2 is three digits. This accuracy is comparable with that of our best approximations and therefore is sufficient for this study. We notice that Theorem 2 provides the theoretical distribution for the sum of the two upper order statistics, therefore for $n = 2$ the values of $z^{(4)}$ are theoretical values of the corresponding quantiles, not approximations. Accordingly, the errors $\Delta^{(4)}$ for $n = 2$ shown in Table 2 are statistical fluctuations of empirical quantiles (in a sample of 10^7) around their theoretical values. These errors are small: for $\alpha \neq 1$ they are less than or equal to 0.001. For $\alpha = 1$ the error increases to 0.003.

The lower quantile approximation $z_q^{(5)}$, $q \ll 1$, indeed provides a closer approximation than $z_q^{(1)}$. The former uniformly has a relative error less than 1% for any n, while the latter starts producing reliable approximations only for $n > 50$. This is due to that fact that the normal approximation used in $z_q^{(5)}$ is much better suited for the small values at the lower tail of the S_n distribution than the stable approximation used in $z_q^{(1)}$. For large n ($n > 10^3$) the quality of $z_q^{(1)}$ becomes acceptable, as seen in Figure 4.

The quality of **the median** approximation $z^{(1)}_{1/2}$ increases rapidly with n, providing acceptable results for $n \geq 20$. For $n < 100$ an order of magnitude better approximation is provided by $z^{(4)}_{1/2}$, whose relative error is less than 1% for $\alpha \geq 1$, and less than 10% for $1/2 \leq \alpha < 1$. As discussed in Section 4.1, the quality of this approximation is affected by the existence of statistical dependence between the highest and lower summands within S_n, and as a result its accuracy decreases with increasing n for $\alpha < 1$.

The upper quantile approximations are all quite accurate, unlike the case of medians and lower quantiles where the choice of approximation appears to be more important. The most rough approximation $z^{(3)}_q$ which simply replaces the sum S_n with its maximal summand provides less than 5% error for $\alpha \leq 1$. Indeed, for $\alpha > 1$ its performance deteriorates, since the largest observation no longer makes an important contribution to the sum S_n. The stable-law approximations $z^{(1,2)}_q$ demonstrate similar performance for $\alpha > 3/2$ and $\alpha < 2/3$, while for $2/3 \leq \alpha \leq 3/2$, $z^{(1)}_q$ appears to be preferable. The most accurate approximation, with less than 1% relative error for all the cases considered, is demonstrated by the approximation $z^{(4)}_q$. By comparison of $z^{(4)}_q$ with $z^{(3)}_q$, one observes how important it is to consider more than just the single largest observation, as well as the importance of the contribution from the smaller summands, which becomes crucial for $\alpha > 1$.

In searching for a best approximation among the five proposed above, we observe a tradeoff between the quality and simplicity. This is especially prevalent for the upper quantiles, where the directly calculated approximations $z^{(2,3)}_q$ are clearly inferior to the more elaborate approximation $z^{(4)}_q$, which involves solving an integral equation. For large samples with $n \geq 10^3$ we recommend using $z^{(1)}_q$. When n is intermediate to small, $n < 10^3$, the upper quantiles can be well approximated by $z^{(2)}_q$; for the lower quantiles $z^{(5)}_q$ appears to perform well. The upper quantile is best approximated by $z^{(4)}_q$, which can be replaced in favor of the simpler $z^{(1)}_q$ when $\alpha < 1$. Generally, the choice of approximation should be dictated by the range of α, n, and q considered, as well as consideration of the relative error rate and computational simplicity desired.

To illustrate the application of our results, Figure 9 shows the distribution of seismic moment for Californian seismicity, $m \geq 5.5$, during the last two centuries. We use the Toppozada *et al.* (2000) earthquake catalog and convert its magnitudes into seismic moments using Eq. (4). Unlike Figure 1, with such a data set one does not observe fewer earthquakes of large seismic moment than expected according to the Pareto law. Indeed, the curvature for large M in Figure 9 may even suggest that the Pareto distribution underestimates the frequency of earthquakes in this seismic moment range. However, as the 95% confidence limits indicate, this departure from the Pareto distribution is not statistically significant (see also discussion on the magnitude accuracy in the catalog below).

In these illustrative displays we do not try to apply the truncated or tapered Pareto distribution (as in Eq. (7) or in Section 5) for evaluating the statistical bounds. As mentioned in Section 1, the proper application of these distributions would require a significant increase in the scope of our investigations.

Continuing with the application of our results to the TOPPOZADA *et al.* catalog, Figure 10 shows the cumulative seismic moment for Californian seismicity, during the last two centuries. The largest earthquake in the plot is the San Francisco 1906 event (m7.8). Similar displays are often used in geophysics to study accumulation of seismic moment (PETERSON and SENO, 1984; JAUMÉ and SYKES, 1996; TRIEP and SYKES, 1997; RODKIN and PISARENKO, 2000; TOPPOZADA *et al.,* 2002). Figure 11 presents the same data and approximations in semi-logarithmic scale with the number of earthquakes shown at the x-axis.

Using the Pareto distribution with $\alpha = 2/3$ as a model for the seismic moment distribution, the observed cumulative moment would be well described by the statistical bounds for Pareto sums given in columns $3 - 5$ of Table 2 and shown (dashed lines) in Figues 10 and 11. For comparison, our approximation by the stable distribution, $z_q^{(1)}$ (dotted lines), is also displayed. As one can see from Table 2, other approximations work much better than $z_q^{(1)}$ (especially for the lower quantile and the median) therefore their respective deviations from the simulated Pareto quantiles in Figures 10 and 11 would be even smaller.

The observed values of the cumulative seismic moment appears to be well within the bounds given by both approximations. However, these bounds indicate that seismic moment release for such data can be estimated only with considerable uncertainty: the lower 2% bound is an order of magnitude smaller than the median, whereas the upper 98% bound exceeds the median by nearly a factor 100. Note that the seismic data particularly prior to 1840 are thought to be of inferior quality, with significant numbers of events missing and their magnitudes possibly substantially underestimated; this may explain why the observed cumulative line starts out of the prescribed statistical bounds.

Although as mentioned in Section 1, the earthquake size distribution should ideally have an upper bound, here we show statistical limits for the unlimited Pareto distribution to illustrate the enormous range of these bounds. The corner moment estimate for California and similar tectonic regions appears to be in $10^{21} < M < 10^{21.6}$ Nm ($m\,8.0 - m\,8.5$) range (BIRD and KAGAN, 2004), thus we have only one earthquake in the sample which approaches the upper limit. Hence, the bounds shown are representative of possible uncertainties in estimating the moment rate, using naive summation of earthquake moments.

Several methods can be used to improve the estimate of the total seismic moment release rate. One can increase the size of the available earthquake dataset and thus include more large earthquakes, either by using older data or by increasing the spatial size of the region under consideration. Both of these methods have clear disadvantages: in the former case the additional data may be

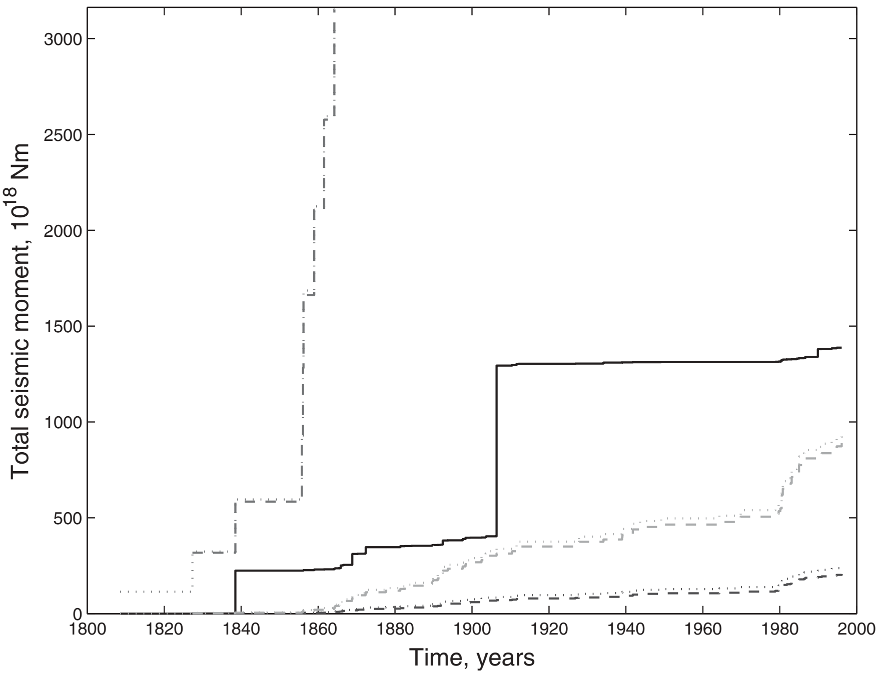


Figure 10

Cumulative seismic moment (solid line) for California as a function of time. Latitude limits 37.0° N – 39.0°N, magnitude threshold 5.5. TOPPOZADA *et al.* (2000) catalog is used. We display quantiles of simulated Pareto ($\alpha = 2/3$) sums with $q = 0.02, 0.5$, and 0.98 (dashed lower, middle, and upper curves, respectively). Bounds with $q = 0.02, 0.5$, and 0.98 obtained using the stable distribution approximation ($z_q^{(1)}$, Eq. 24) are shown by dotted curves.

of inferior quality; in the latter case the resolution of the study deteriorates. As another alternative, one can use tectonic and geodetic data to constrain the moment release rate (KAGAN, 2002b; BIRD and KAGAN, 2004) by integrating the seismic moment-frequency relation (7). If the corner or maximum seismic moment is known or can be estimated (KAGAN and SCHOENBERG, 2001; KAGAN, 2002a), we can use the methods developed in Section 5 to evaluate the bounds.

Figures 10 and 11 suggest that for seismological applications the quality of the approximation $z_q^{(1)}$ may be sufficient for an adequate description of the bounds on the distribution of the total seismic moment release. Indeed, when the quality of the available data subject to approximation are of such low accuracy and when the variability of the sum S_n is so extreme, even a relative error in the approximation as large as 10% – 20% may be acceptable.

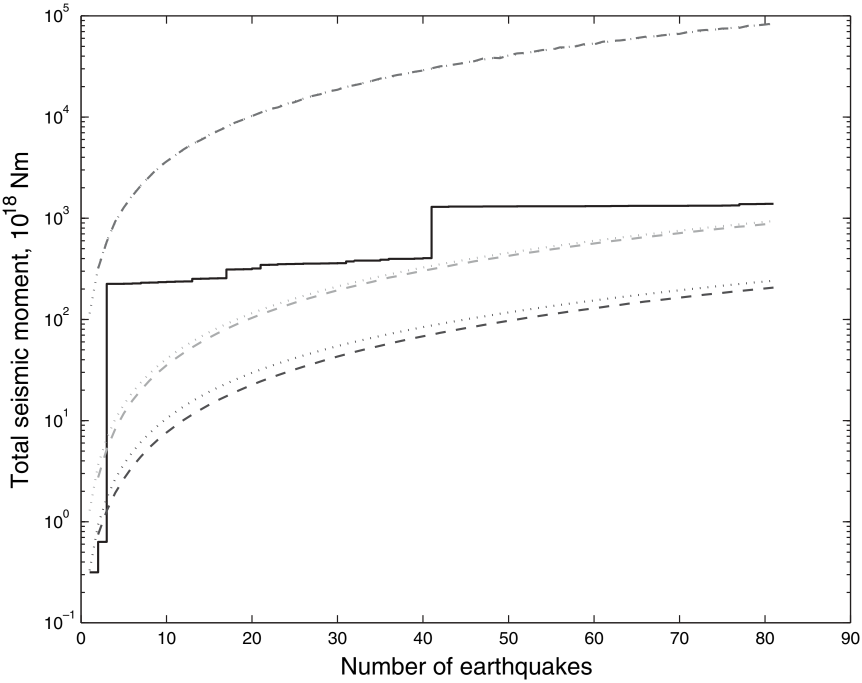


Figure 11

Logarithmic cumulative seismic moment (solid line) for California as a function of the number of events, n. Latitude limits 37.0°N – 39.0°N, magnitude threshold 5.5. Toppozada *et al.* (2000) catalog is used. We display quantiles of simulated Pareto ($\alpha = 2/3$) sums with $q = 0.02, 0.5$, and 0.98 (dashed lower, middle, and upper curves, respectively). Bounds obtained using the stable distribution approximation ($z_q^{(1)}$, Eq. 24) are shown by dotted curves.

Acknowledgements

We thank V. F. Pisarenko for stimulating and valuable discussions, and for providing us with his unpublished notes. We are grateful to D. Vere-Jones and D. Sornette for their discussion and support. Reviews by Renata Rotondi, by an anonymous reviewer and suggestions by Associate editor Yehuda Ben-Zion have enhanced revision of the manuscript. Ilya Zaliapin was partly supported by independent International Association (INTAS), grant 0748 and by grant NSF-ATM 0327558. Yan Kagan appreciates partial support from the Southern California Earthquake Center (SCEC). SCEC is funded by NSF Cooperative Agreement EAR-0106924 and USGS Cooperative Agreement 02HQAG0008. This material is based upon work supported in part by the National Science Foundation under Grant No. DMS-0306526. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Publication 770, SCEC.

Appendices

A. Asymptotic Formulas for Stable Distributions

In comparing the Pareto sums to the appropriately scaled stable distributions, the upper and lower tails of these distributions are of special relevance to us. For maximally-skewed distributions, the lower tail is 'light', i.e., its density decays faster than any power of x (ZOLOTAREV, 1986; SAMORODNITSKY and TAQQU, 1994; UCHAIKIN and ZOLOTAREV, 1999). On the other hand, the upper tail is governed by a power-law decay with the exponent $\alpha < 2$, such tails are called 'heavy'.

There exist series expansions for stable distributions that are known to converge relatively quickly in the tails. One may take the leading term $F_\alpha^{(0)}(x)$ of these series as an approximation of the distribution at the tails. The form of the leading terms depends on the index α and on the tail (upper or lower) to be approximated. Below we present three approximations for the lower quantiles (Eqs. 70–73) and two approximations for the upper quantiles (Eqs. 74–75).

For $\alpha < 1$ by integrating the pdf from LINNIK (1954, Eq. 1) or SKOROHOD (1954, Eq. IV) we obtain

$$F_\alpha^{(0)}(x) = \frac{1}{\sqrt{2\alpha}}\left\{1 - \mathrm{erf}\left[\sqrt{\frac{1-\alpha}{\alpha}}\left(\frac{\alpha}{\cos(\pi\alpha/2)}\right)^{\frac{1}{2(1-\alpha)}} x^{\frac{\alpha}{2(\alpha-1)}}\right]\right\}, \quad x > 0, \tag{70}$$

where the error function erf (x) is given by

$$\mathrm{erf}\ (x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)\,dt. \tag{71}$$

For $\alpha = 1$ integrating the pdf from IBRAGIMOV and LINNIK (1971, Theorem 2.4.4) yields

$$F_\alpha^{(0)}(x) = \frac{1}{\sqrt{2}}\left\{1 - \mathrm{erf}\left[\sqrt{\frac{2}{\pi}}\exp\left(-\frac{1}{2} - \frac{\pi x}{4}\right)\right]\right\}, \quad -\infty < x < \infty \tag{72}$$

and for $\alpha > 1$ integrating the pdf from SKOROHOD (1954, Eq. VI) yields

$$F_\alpha^{(0)}(x) = \frac{1}{\sqrt{2\alpha}}\left\{1 - \mathrm{erf}\ \left[\sqrt{\frac{\alpha-1}{\alpha}}\left(\frac{-\alpha}{\cos(\pi\alpha/2)}\right)^{\frac{1}{2(1-\alpha)}}(-x)^{\frac{\alpha}{2(\alpha-1)}}\right]\right\}, \quad -\infty < x < \infty. \tag{73}$$

As UCHAIKIN and ZOLOTAREV (1999, p. 127) note, equation (70) is exact for $\alpha = 0.5$ (see Eq. 14), and (73) corresponds to the Gaussian distribution $\Phi(x; 0, 2)$ for $\alpha = 2.0$.

For the upper quantile, $x \to \infty$, one may use the approximation (SAMORODNITSKY and TAQQU, 1994, Eqs. 1.2.8, 1.2.9)

$$F_\alpha^{(0)}(x) = 1 - [x^\alpha \Gamma(1-\alpha)\cos(\pi\alpha/2)]^{-1}, \tag{74}$$

if $\alpha \neq 1$, and

$$F_\alpha^{(0)}(x) = 1 - \frac{2}{\pi x}, \tag{75}$$

for $\alpha = 1$.

In Figure 12a we display an example of approximating the lower quantiles using Eq. (70). In general, the approximations (Eqs. 70–73) work reasonably well, with relative errors of just a few percent. For the upper quantiles (Eqs. 74–75) the fit is good only for certain ranges of α. Figure 12b demonstrates, for instance, that for the 0.98 quantile the expression (74) is within 10% of the theoretical values for $\alpha < 0.85$ and for $1.21 < \alpha < 1.73$, whereas for values of α closer to 2 the fit of the approximation is only satisfactory for very high quantiles such as 0.999.

B. Normalization Coefficients in Generalized Central Limit Theorem

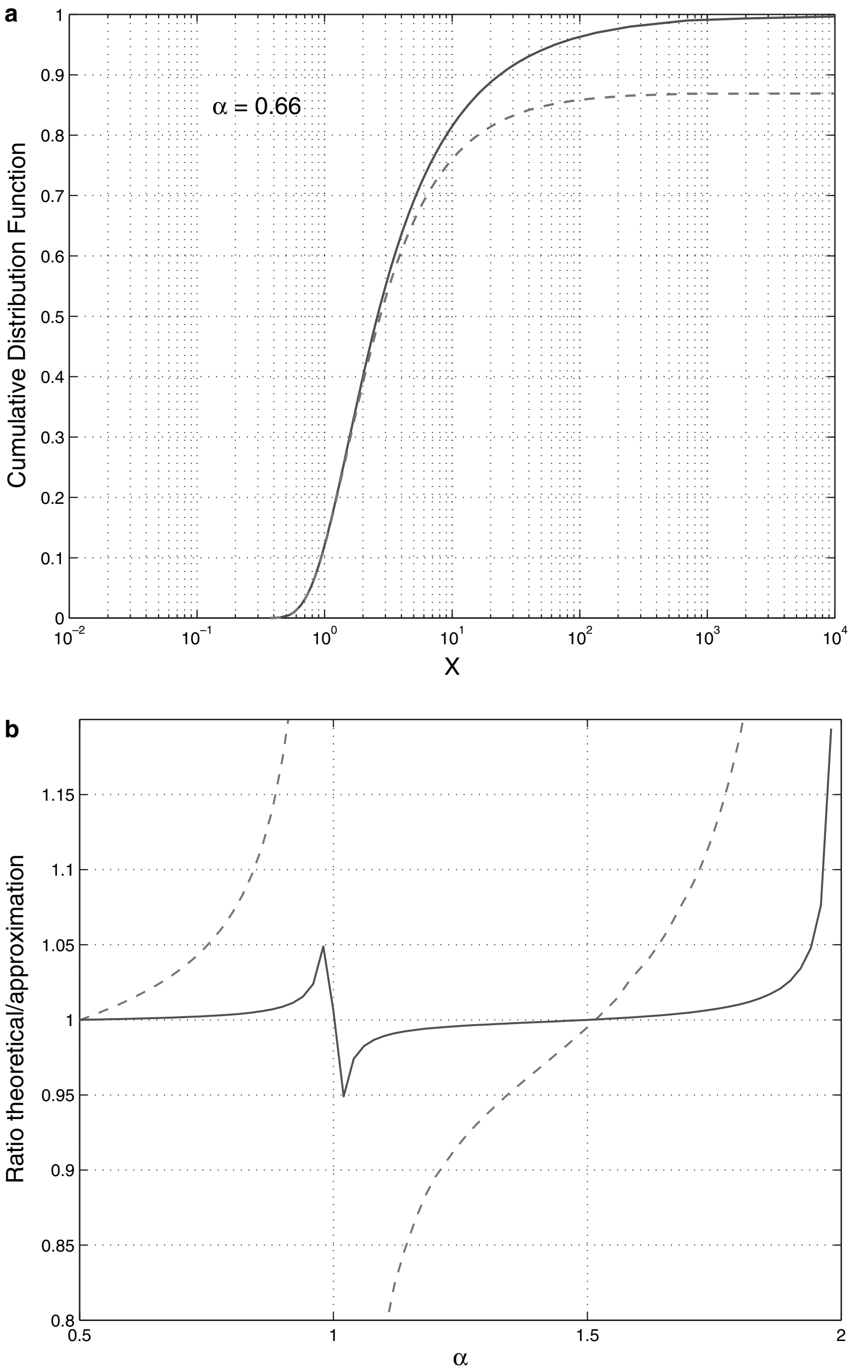
Here we follow IBRAGIMOV and LINNIK (1971) (referred to below as IL71) to derive the normalization coefficient C_α in the Generalized Central Limit Theorem. Although this section is a mere extraction from (IL71); we believe it is worth presenting the complete derivation here. Since the computational aspects of stable laws were negligible in the 1960s, IBRAGIMOV and LINNIK (1971) do not give explicit formulae for C_α. Such formulas appeared in MIJNHEER (1975, see his Eq. 2.2.2 and 2.2.4, p. 17), but were unfortunately misprinted. These misprinted formulas were borrowed by SAMORODNITSKY and TAQQU (1994, Theorem 1.8.1, p. 50). As a result, coefficients that are important for all practical purposes appear misprinted in the most popular references on the subject. The correct formula is given in UCHAIKIN and ZOLOTAREV (1999) though in an alternative form, thus below (Eqs. 85–87) we also demonstrate the equivalence of their Table 2.1, p. 62, and our Eq. (20).

Recall (IL71, Theorem 2.2.1, p. 39) that the characteristic function f_α of a stable law F_α with $0 < \alpha < 2$ can be represented as

$$\begin{aligned}\log f(t) = i\gamma t &+ \int_{-\infty}^{0} \left(\mathrm{e}^{itu} - 1 - \frac{itu}{1+u^2}\right) dM(u)\\ &+ \int_{0}^{\infty} \left(\mathrm{e}^{itu} - 1 - \frac{itu}{1+u^2}\right) dN(u),\end{aligned} \tag{76}$$

with

$$M(u) = c_1(-u)^{-\alpha}, \quad N(u) = -c_2 u^{\alpha},$$
$$c_1 \geq 0, \quad c_2 \geq 0, \quad c_1 + c_2 > 0.$$


a
α = 0.66
Cumulative Distribution Function
X
b
Ratio theoretical/approximation
α

Theorem 2.2.2 (IL71, p. 43) evaluates explicitly the integrals on the right-hand side of (76) to show that

$$\log f(t) = i\gamma t - c|t|^\alpha \left(1 - i\beta \frac{t}{|t|} w(t, \alpha)\right), \tag{77}$$

where α, β, γ, c are constants ($c \geq 0$, $|\beta| \leq 1$) and

$$w(t, a) = \begin{cases} \tan(\pi\alpha/2), & \alpha \neq 1, \\ (2/\pi) \log|t|, & \alpha = 1. \end{cases}$$

In the proof of Theorem 2.2.2 one finds the following (unnumbered) expressions for c (pages from IL71):

$$c = \begin{cases} -\alpha L(\alpha)(c_1 + c_2)\cos(\pi\alpha/2), & 0 < \alpha < 1, & \text{p.44} \\ -\alpha M(\alpha)(c_1 + c_2)\cos(\pi\alpha/2), & 1 < \alpha < 2, & \text{p.45} \\ \pi(c_1 + c_2)/2, & \alpha = 1, & \text{p.45} \end{cases} \tag{78}$$

with

$$L(\alpha) = \int_0^\infty (\mathrm{e}^{-u} - 1)\frac{du}{u^{1+\alpha}} = -\frac{\Gamma(1-\alpha)}{\alpha} < 0.$$

$$M(\alpha) = \int_0^\infty (\mathrm{e}^{-u} - 1 + u)\frac{du}{u^{1+\alpha}} = -\frac{\Gamma(2-\alpha)}{\alpha(\alpha-1)} > 0.$$

Substituting $L(\alpha)$, $M(\alpha)$ into (78) yields

$$c = \begin{cases} \Gamma(1-\alpha)(c_1 + c_2)\cos(\pi\alpha/2), & \alpha \neq 1, \\ \pi(c_1 + c_2)/2, & \alpha = 1. \end{cases} \tag{79}$$

If we restrict ourselves to the case $c = 1$, which can always be achieved by proper rescaling of variables, then

$$c_1 + c_2 = \begin{cases} [\Gamma(1-\alpha)\cos(\pi\alpha/2)]^{-1}, & \alpha \neq 1, \\ 2/\pi, & \alpha = 1. \end{cases} \tag{80}$$

We seek the coefficients B_n such that the normalized sum

$$\frac{X_1 + X_2 \ldots + X_n - b_n}{B_n}$$

◀

Figure 12
Approximations for the tails of stable distributions. (a) Stable cdf (McCulloch and Panton, 1987) with $\alpha = 0.66$ (solid line) and its approximation by Eq. (70) (dashed line). (b) Ratio between theoretical and approximated values of the upper quantiles 0.98 (dashed) and 0.999 (solid) as functions of the α-index.

of independent random variables X_i with the same distribution function $F(x)$ converges to a stable distribution $S(x)$ with exponent α $(0 < \alpha < 2)$. Theorem 2.6.1 (IL71, p. 76) states that is this case

$$1 - F(x) + F(-x) = (c_1 + c_2)\frac{h(x)}{x^\alpha}(1 + o(1)), \tag{81}$$

where $h(x)$ is a slowly varying function. Furthermore, from Theorems 1.7.3 (IL71, p. 35) and 2.2.1 (IL71, p. 39) (see also Eqs. (2.6.3) and (2.6.4) in IL71) it follows that

$$n[1 - F(B_n x) + F(-B_n x)] \to (c_1 + c_2)x^{-\alpha}, \tag{82}$$

for the appropriate choice of the constants B_n. Combining (81) and (82) one finds

$$\frac{n\, h(B_n x)}{B_n^\alpha} \to 1. \tag{83}$$

Condition (83) is the most general condition describing the normalization coefficients B_n in the GCLT.

In the case of the Pareto distribution we may find an explicit expression for B_n. Since

$$1 - F(x) + F(-x) = x^{-\alpha}$$

then

$$h(x) = (c_1 + c_2)^{-1}$$

and we finally obtain

$$C_\alpha := \frac{B_n}{n^{1/\alpha}} = (c_1 + c_2)^{-1/\alpha} = \begin{cases} [\Gamma(1-\alpha)\cos(\pi\alpha/2)]^{1/\alpha}, & \alpha \neq 1, \\ \pi/2, & \alpha = 1. \end{cases} \tag{84}$$

As we mentioned above, Uchaikin and Zolotarev (1999, their Table 2.1) give the formulae for the B_n coefficient in a different form (for $\alpha \neq 1$):

$$B_n = \left[\frac{\pi n}{2\Gamma(\alpha)\sin(\pi\alpha/2)}\right]^{1/\alpha}, \tag{85}$$

The case $\alpha = 1$ is obtained as a limit of Eq. (85)

$$\lim_{\alpha \to 1} B_n = (\pi n)/2, \tag{86}$$

(see the second Eq. 20). The expression (85) can be easily transformed into the first Eq. (20), using Eq. (6.1.17) in Abramowitz and Stegun (1972):

$$\Gamma(\alpha)\Gamma(1-\alpha) = \frac{\pi}{2\sin(\pi\alpha/2)\cos(\pi\alpha/2)}. \tag{87}$$

C. Proof of (34)

Consider the ratio X_1/M_n. Its cdf $G(z)$ can be expressed in terms of the conditional distribution $F(x|y)$ of X_1 conditioned on the fixed maximum $M_n = y$ (PISARENKO, 2003, private communication):

$$G(z) = \text{Prob}\left\{\frac{X_1}{M_n} < z\right\} = \int_0^\infty \text{Prob}\left\{\frac{X_1}{y} < z|y\right\} dF^n(y)$$
$$= \int_0^\infty \text{Prob}\{X_1 < yz|y\} dF^n(y)$$
$$= \int_0^\infty F(yz|y)\, dF^n(y),$$

and the density $g(z)$ is obtained by taking the derivative with respect to z:

$$g(z) = \int_0^\infty y f(yz|y)\, dF^n(y). \tag{88}$$

The conditional density $f(x|y)$ corresponding to the distribution $F(x|y)$ is given by

$$f(x|y) = \frac{\delta(y-x)}{n} + \left(1 - \frac{1}{n}\right)\frac{f(x)H(y-x)}{F(y)}. \tag{89}$$

Here $H(x) = 1$ for $x > 0$ and 0 for $x \leq 0$. The first term on the rhs of (89) corresponds to the case $M_n = X_1$, whose probability is $1/n$ by symmetry; the complementary event $X_1 < M_n$ occurs with probability $(1 - 1/n)$. Substituting the pdf (89) into (88) we obtain

$$g(z) = \frac{\delta(z-1)}{n} + (n-1)\int_0^\infty y f(yz) f(y) F^{n-2}(y)\, dy, \quad 0 < z \leq 1. \tag{90}$$

This expression can be used to calculate the expectation of the ratio $r_n = S_n/M_n$:

$$
\begin{aligned}
E(r_n) &= n\int_0^1 zg(z)dz \\
&= 1 + n(n-1)\int_0^\infty yf(y)F^{n-2}(y)\int_0^1 zf(yz)dydz \\
&= 1 + n(n-1)\int_0^\infty yf(y)F^{n-2}(y)\left(\frac{F(y)}{y} - \frac{1}{y^2}\int_0^y F(x)dx\right)dy \\
&= 1 + n(n-1)\left(\int_0^\infty f(y)F^{n-1}(y)dy - \int_0^\infty T(y)f(y)F^{n-2}(y)dy\right) \\
&= 1 + (n-1)\left(\int_0^\infty dF^n(y)\right) - n\left(\int_0^\infty T(y)dF^{n-1}(y)\right) \\
&= n\left(1 - \int_0^\infty T(y)dF^{n-1}(y)\right),
\end{aligned} \tag{91}
$$

where $T(y) = 1/y\int_0^y F(u)du$. For the Pareto distribution (2) this becomes

$$
T(y) = \frac{1}{y}\int_1^y (1 - x^{-\alpha})dx = 1 - \frac{1}{1-\alpha}\left(y^{-\alpha} - \alpha y^{-1}\right), \tag{92}
$$

and plugging (92) into (91) one obtains

$$
E(r_n) = E\left(\frac{S_n}{M_n}\right) = \begin{cases} \frac{1-nB\left(n,\alpha^{-1}\right)}{1-\alpha}, & \alpha \neq 1 \\ \sum_{k=1}^n \frac{1}{k}, & \alpha = 1. \end{cases} \tag{93}
$$

For the truncated Pareto distribution (65) on $[1, X_p]$ we have (*cf.* PISARENKO, 1998, Eq.12)

$$
\begin{aligned}
E(r_n) = n\Big\{1 - (n-1)\left(1 - X_p^{-\alpha}\right)^{-n}[B(z,1,1,n-1) \\
- \frac{1}{1-\alpha}B(z,1,2,n-1) + \frac{\alpha}{1-\alpha}B(z,1,1+1/\alpha,n-1)\Big]\Big\},
\end{aligned} \tag{94}
$$

where $z = X_p^{-\alpha}$ and $B(\cdot,\cdot,\cdot,\cdot)$ is defined by (47).

D. Existence of Moments for Ordered Statistics

The general distribution of the order statistics $X_{k,n}$ from a sample with a common cdf $F(x)$ is given by (NEVZOROV, 2001, Eq. 2.1):

$$F_{n-k,n}(x) = P\{X_{n-k,n} < x\} = \sum_{m=n-k}^{n} C_m^n (F(x))^m (1 - F(x))^{n-m}. \tag{95}$$

For the Pareto distribution (2) this gives

$$F_{n-k,n}(x) = 1 - C_{n-k-1}^{n} x^{-(k+1)\alpha} + o\left(x^{-(k+1)\alpha}\right), \tag{96}$$

so the decay rate of the tail

$$\left[1 - F_{n-k,n}(x)\right] \sim x^{-(k+1)\alpha}$$

does not depend on the sample volume n, hence the existence of moments for $X_{n-k,n}$ is determined by k only.

In general, the existence of finite moments for order statistics is described by SEN (1959):

Theorem 3. *If $E(|X|^{\alpha}) < \infty$ for some α, then the moment $\mu_{k,n}^{(r)} = E\left(X_{k,n}\right)^r$ exists for all k such that*

$$\frac{r}{\alpha} \leq k \leq n + 1 - \frac{r}{\alpha}.$$

E. Proof of Theorem 2

Denote by $F_{n-1,n}$ the distribution of the order statistic $X_{n-1,n}$ and by $\bar{F}_{n,n}(x|u)$ the conditional distribution of the upper order statistic $X_{n,n}$:

$$1 - \bar{F}_{n,n}(x|u) = \text{Prob } \{X_{n,n} > x | X_{n-1,n} = u\}. \tag{97}$$

Let $\bar{f}_{n,n}(x|u) = \frac{d}{dx}\bar{F}_{n,n}(x|u)$ and $f_{n-1,n}(x) = \frac{d}{dx}F_{n-1,n}(x)$. Then

$$\begin{aligned} f_{n-1,n}(x) &= \frac{d}{dx}\left[n(1 - F(x))F^{n-1}(x) + F^n(x)\right] \\ &= n(n-1)f(x)F^{n-2}(x)(1 - F(x)); \end{aligned} \tag{98}$$

$$\bar{f}_{n,n}(x|u) = \frac{d}{dx}\left[1 - \frac{1 - F(x)}{1 - F(u)}\right] = \frac{f(x)}{1 - F(u)}. \tag{99}$$

The distribution of $T_{n-2,n} = X_{n-1,n} + X_{n,n}$ is calculated using the Markov property (43):

$$\begin{aligned}\mathrm{Prob}\{X_{n-1,n}+X_{n,n}<x\} &= \int_2^x\int_1^{z/2} f_{n-1,n}(y)\bar{f}_{n,n}(z-y)\,dydz \\ &= n(n-1)\int_2^x\int_1^{z/2} f(y)f(z-y)F^{n-2}(y)\,dydz. \qquad (100)\end{aligned}$$

The internal upper integration limit is $z/2$ due to the inequality $X_{n-1,n} \le X_{n,n}$. For the Pareto distribution this yields

$$\begin{aligned}\mathrm{Prob}\{X_{n-1,n}+X_{n,n}<x\} &= n(n-1)\alpha^2\int_2^x\int_1^{z/2} y^{-\alpha-1}(z-y)^{-\alpha-1}(1-y^{-\alpha})^{n-2}\,dy\,dz \\ &= n(n-1)\alpha^2\sum_{k=0}^{n-2}(-1)^k C_{n-2}^k\int_2^x\int_1^{z/2} y^{-\alpha(k+1)-1}(z-y)^{-\alpha-1}\,dy\,dz.\end{aligned}$$

(101)

References

ABRAMOWITZ, M. and STEGUN, I.A., *Handbook of Mathematical Functions* (Dover, NY 1972), 1046 pp.

AKI, K. (1965), *Maximum Likelihood Estimate of b in the Formula* $\log N = a - bM$ *and its Confidence Limits*, Bull. Earthquake Res. Inst. Tokyo Univ. *43*, 237–239.

ANDERSON, J.G. and LUCO, J.E. (1983), *Consequences of Slip Rate Constraints on Earthquake Occurrence Relation*, Bull. Seismol. Soc. Am. *73*, 471–496.

BARTON, C.C. and LA POINTE, P.R., eds., *Fractals in the Earth Sciences* (Plenum Press, New York 1995) 265 pp.

BEN-ZION, Y. Appendix 2, *Key Formulas in Earthquake Seismology*, In *International Handbook of Earthquake and Engineering Seismology*, Part B, 1857–1875 (Academic Press 2003).

BIRD, P. and KAGAN, Y.Y. (2004), *Plate-tectonic Analysis of Shallow Seismicity: Apparent Boundary Width, Beta, Corner Magnitude, Coupled Lithosphere Thickness, and Coupling in Seven Tectonic Settings*, Bull. Seismol. Soc. Am. 94 (6), 2380–2399.

EKSTRÖM, G., DZIEWONSKI, A.M., MATERNOVSKAYA, N.N., and NETTLES, M. (2003), *Global Seismicity of 2001: Centroid-moment Tensor Solutions for 961 Earthquakes*, Phys. Earth Planet. Inter. *136*(3–4), 165–185.

EMBRECHTS, P., KLÜPPELBERG, C., and MIKOSCH, T., *Modelling Extremal Events for Insurance and Finance* (Springer-Verlag, Berlin 1997) 645 pp.

FELLER, W., *An Introduction to Probability Theory and its Applications*, *Vol. 2*, 2nd ed. (J. Wiley, New York 1971), 509 pp.

GRADSHTEYN, I.S. and RYZHIK, I.M., *Table of Integrals, Series, and Products* (Acad. Press, NY 1980) 1160 pp.

GUTENBERG, B. and RICHTER, C.F. (1941), *Seismicity of the Earth*, Geol. Soc. Am. Special papers, *34*, 1–131.

GUTENBERG, B. and RICHTER, C.F. (1944), *Frequency of Earthquakes in California*, Bull. Seism. Soc. Am. *34*, 185–188.

HALL, P. (1981), *A Comedy of Errors: The Canonical Form for a Stable Characteristic Function*, Bull. London Math. Soc. *13*(1), 23–27.

HOLT, W.E., CHAMOTROOKE, N., LE PICHON, X., HAINES, A.J., SHEN-TU, B., and REN, J. (2000), *Velocity Field in Asia Inferred from Quaternary Fault Slip Rates and Global Positioning System Observations*, J. Geophys. Res. *105*, 19,185–19,209.

HUILLET, T. and RAYNAUD, H.F. (2001), *On Rare and Extreme Events*, Chaos, Solitons and Fractals, **12**, 823–844.

IBRAGIMOV, I.A. and YU. V. LINNIK, *Independent and Stationary Sequences of Random Variables* (Groningen, Wolters-Noordhoff 1971) 443 pp. Russian original 1965.

JAUMÉ, S.C. and SYKES, L.R. (1996), *Evolution of Moderate Seismicity in the San Francisco Bay region, 1850 to 1993: Seismicity Changes Related to the Occurrence of Large and Great Earthquakes*, J. Geophys. Res. *101*, 765–89.

KAGAN, Y.Y. (2002a), *Seismic Moment Distribution Revisited: I. Statistical Results*, Geophys. J. Int. *148*, 520–541.

KAGAN, Y.Y. (2002b), *Seismic Moment Distribution Revisited: II. Moment Conservation Principle*, Geophys. J. Int. *149*, 731–754.

KAGAN, Y.Y. (1997), *Earthquake Size Distribution and Earthquake Insurance*, Communications in Statistics: Stochastic Models *13*(4), 775–797.

KAGAN, Y.Y. (1994), *Observational Evidence for Earthquakes as a Nonlinear Dynamic Process*, Physica D *77*, 160–192.

KAGAN, Y.Y. and SCHOENBERG, F. (2001), *Estimation of the upper cutoff parameter for the tapered Pareto distribution*, J. Appl. Probab. *38A*, 158–175.

KANAMORI, H. (1977), *The Energy Release in Great Earthquakes*, J. Geophys. Res. *82*, 2981–2987.

KNOPOFF, L. and KAGAN, Y.Y. (1977), *Analysis of the Theory of Extremes as Applied to Earthquake Problems*, J. Geophys. Res. *82*, 5647–5657.

KOSTROV, B.V. (1974), *Seismic Moment and Energy of Earthquakes, and Seismic Flow of Rock*, Izv. Acad. Sci. USSR, Phys. Solid Earth, January 13–21.

LINNIK, YU. V. (1954), *On Stable Probability Laws with Exponent less than One (in Russian)*, Doklady AN USSR *94*, 619–621.

MANDELBROT, B.B., *The Fractal Geometry of Nature* (W. H. Freeman, San Francisco, Calif. 1983) 2nd edition, 468 pp.

MCCAFFREY, R. (1997), *Statistical Significance of the Seismic Coupling Coefficient*, Bull. Seismol. Soc. Am. *87*, 1069–1073.

MCCULLOCH, J.H. and PANTON, D.B. (1997), *Precise Tabulation of the Maximally-skewed Stable Distributions and Densities*, Comput. Statist. Data Anal. *23*, 307–320; Erratum, *26*, 101. http://www.econ.ohio-state.edu/jhm/fracden

MCCULLOCH, J.H. and PANTON, D.B. *Table of the Maximally-skewed Stable Distributions*, In (ADLER, R.J., FELDMAN, R.E., and TAQQU, M.S. eds.), *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, (Birkhäuser, Boston 1998) pp. 501–507.

MIJNHEER, J.L., *Sample Path Properties of Stable Processes*, (Math. Centrum, Amsterdam 1975) 124 pp.

MITTNIK, S., RACHEV, S.T., and KIM, J.R. (1998), *Chi-square-type Distributions for Heavy-tailed Variates*, Economet. Theor. *14*(3), 339–354.

NEVZOROV, V.B., *Records: Mathematical Theory* (Translations of Mathematical Monographs, v. 194) (American Mathematical, Society 2001) pp. 164; Russian original 2000.

NEWMAN, W.I., GABRIELOV, A., and TURCOTTE, D.L. eds., *Nonlinear Dynamics and Predictability of Geophysical Phenomena*, Geoph. Monogr. *83* (Washington, American Geophysical Union 1994) 107 pp.

NOLAN, J.P. (1997), *Numerical Calculation of Stable Densities and Distribution Functions*, Communications in Statistics: Stochastic Models *13*, 759–774. http://academic2.american.edu/~jpnolan/stable/quantile.dat

NOLAN, J.P. *Stable Distributions – Models for Heavy-Tailed Data* (Boston, Birkhäuser 2005) in press.

PETERSON, E.T. and SENO, T. (1984), *Factors Affecting Seismic Moment Release Rates in Subduction Zones*, J. Geophys. Res. *89*, 10,233–10,248.

PISARENKO, V.F. (1998), *Non-linear Growth of Cumulative Flood Losses with Time*, Hydrological Processes *12*, 461–470.

RACHEV, S. and MITTNIK, S., *Stable Paretian Models in Finance* (J. Wiley, NY 2000) 855 pp.

RACHEV, S.T., ed., *Handbook of Heavy Tailed Distributions in Finance* (Elsevier, Amsterdam 2003) 680 pp.

RICHARDSON, L.F., *Statistics of Deadly Quarrels* (Wright, Q. and Lienau, C.C. eds., Pittsburgh, Boxwood Press 1960).

RODKIN, M.V. and PISARENKO, V.F. (2000), *Damage and Casualties in Earthquakes: A Statistical Analysis*, Vychislitel'naya Seismologiya (Comput. Seismol.) *31*, Geos, Moscow, 242–272 (in Russian).

SAMORODNITSKY, G. and TAQQU, M.S., *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance* (New York, Chapman and Hall 1994) 632 pp.

SCHOLZ, C.H., *The Mechanics of Earthquakes and Faulting* (Cambridge University Press, Cambridge 2002) 2nd ed., 471 pp.

SEN, P.K. (1959), *On the Moments of Sample Quantiles*, Bull. Calcutta Stat. Assoc. *9*, 1–19.

SHEN-TU, B.M., HOLT, W.E., and HAINES, A.J. (1998), *Contemporary Kinematics of the Western United States Determined from Earthquake Moment Tensors, Very Long Baseline Interferometry, and GPS Observations*, J. Geophys. Res. *103*, 18,087–18,117.

SKOROHOD, A.V. (1954), *Asymptotic Formulas for Stable Distribution Laws (in Russian)*, Doklady AN USSR *98*, 731–734.

SORNETTE, D., *Critical Phenomena in Natural Sciences (Chaos, Fractals, Self-organization, and Disorder: Concepts and Tools)*, (New York, Springer 2003) 2nd ed., 528 pp.

TOPPOZADA, T., BRANUM, D., PETERSEN, M., HALLSTROM, C., CRAMER, C., and REICHLE, M. (2000), *Epicenters of and Areas Damaged by $M \geq 5$ California Earthquakes, 1800–1999*, Map sheet 49, Div. Mines Geology, California.

TOPPOZADA, T.R., BRANUM, D.M., REICHLE, M.S., and HALLSTROM, C.L. (2002), *San Andreas Fault Zone, California: $M \geq 5.5$ Earthquake History*, Bull. Seismol. Soc. Am. *92*, 2555–2601.

TRIEP, E.G. and SYKES, L.R. (1997), *Frequency of Occurrence of Moderate to Great Earthquakes in Intracontinental Regions: Implications for Changes in Stress, Earthquake Prediction, and Hazards Assessments*, J. Geophys. Res. *102*, 9923–9948.

TURCOTTE, D.L., *Fractals and Chaos in Geology and Geophysics*, 2nd ed. (Cambridge University Press, Cambridge 1997) 398 pp.

UCHAIKIN, V.V. and ZOLOTAREV, V.M., *Chance and Stability: Stable Distributions and Their Applications* (VSP International Science Publishers, Utrecht 1999) 596 pp.

VERE-JONES, D., ROBINSON, R., and YANG, W.Z. (2001), *Remarks on the Accelerated Moment Release Model: Problems of Model Formulation, Simulation and Estimation*, Geophys. J. Int. *144*, 517–531.

WELLS, D.L. and COPPERSMITH, K.J. (1994), *New Empirical Relationships Among Magnitude, Rupture Length, Rupture Width, Rupture Area, and Surface Displacement*, Bull. Seismol. Soc. Am. *84*, 974–1002.

WOLFRAM, S., *The Mathematica Book*, 4th ed. (Cambridge University Press 1999) 1470 pp.

ZOLOTAREV, V.M. (1954), *Expression of the Density of a Stable Distribution with Exponent α Greater than One by Means of a Density with Exponent* $1/\alpha$, (in Russian), Dokl. Akad. Nauk SSSR *98*, 735–738.

ZOLOTAREV, V.M., *One-Dimensional Stable Distributions*, (Amer. Math. Soc. Providence, R.I. 1986) pp. 284; Russian original 1983.

(Received November 1, 2003, accepted June 7, 2004)