J. Appl. Probab. **38A**, 901–918 (2001)
Printed in England

ESTIMATION OF THE UPPER CUTOFF PARAMETER FOR THE TAPERED PARETO DISTRIBUTION

Y. Y. KAGAN[1] AND
F. SCHOENBERG,[2] *University of California*

Abstract

The tapered (or generalized) Pareto distribution, also called the modified Gutenberg–Richter law, has been used to model the sizes of earthquakes. Unfortunately, maximum likelihood estimates of the cutoff parameter are substantially biased. Alternative estimates for the cutoff parameter are presented, and their properties discussed.

Keywords: Parameter estimation; tapered Pareto distribution; Gutenberg–Richter relation; maximum likelihood estimation; method of moments; earthquakes

AMS 2000 Subject Classification: Primary 62F10
Secondary 62F12, 62F25, 62F15

1. Introduction

In this paper, we investigate the problem of estimating the parameters used in models for the sizes of earthquakes. Such models are important for the investigation and quantification of seismic hazard, i.e. the risk of an earthquake exceeding a given size within a given time period in the future, which is a subject of concern not only in seismology but also in civil engineering, urban planning, and insurance.

Our work was chiefly inspired by David Vere-Jones, who has been one of the principal architects of the general area of statistical seismology, in which the subject of our paper lies. In addition to his numerous significant advances on closely related themes, Vere-Jones has been a main proponent of the use of the tapered Pareto law in modeling earthquake sizes.

This paper is addressed to both statisticians and seismologists, and in what follows we provide more explanatory material than is customary. General properties of the earthquake process and its stochastic modelling, as well as some principles of statistical analysis, have been discussed by Kagan [11] and Kagan and Vere-Jones [14]. Earthquake occurrences are most commonly modeled as point processes; Daley and Vere-Jones [2] provide a thorough introduction to such models.

Extensive global catalogs of seismic moment tensor solutions with thousands of events have become available since 1977 (see [5], [6] and [24], and references therein).

Received September 2000
[1] Postal address: Dept of Earth and Space Sciences, University of California, Los Angeles CA 90095–1567, USA. Email: ykagan@ucla.edu
[2] Postal address: Dept of Statistics, University of California, Los Angeles, CA 90095–1554, USA. Email: frederic@stat.ucla.edu

These catalogs are thought to be of significantly superior quality (in terms of precision, reliability, completeness, and accuracy) compared to instrumental earthquake catalogs available since the end of the 19th century. An example is the Harvard catalog [5], which presently contains information on 16706 earthquakes for the period 1977/1/1–1999/12/31. Many of these earthquakes occur at shallow depths (less than 70 km), and represent a major hazard to humans.

Modern earthquake catalogs such as the Harvard dataset contain the following information about moderate and large earthquakes: origin times; locations; sizes; and orientations of earthquake focal mechanisms ([11], [16]). An earthquake's location may be specified either via hypocentral coordinates (the spatial location where the earthquake rupture begins) or via earthquake centroid coordinates (the location of the center of gravity for the seismic moment release). Historically, earthquake sizes were recorded in terms of empirical magnitude, a measure which was first proposed in the 1930s by Charles Richter for characterizing earthquakes in California. Presently, an earthquake's size is typically measured in terms of either scalar seismic moment, M, or on a logarithmic scale via moment magnitude, m. Seismic moment is measured in *Newton-meters* (Nm) or in *dyne cm* ($1\,\mathrm{Nm} = 10^7\,\mathrm{dyne\,cm}$) ([5]) and may be converted into moment magnitude via the function ϕ:

$$m \simeq \phi(M) = \tfrac{2}{3}\log_{10} M - 6, \tag{1}$$

with M expressed in units of Nm.

The distribution of earthquake sizes has been the subject of considerable research; see Kagan [11], Vere-Jones [29] and especially Utsu [26] for a thorough review. Earthquake magnitudes have traditionally been modelled via the Gutenberg–Richter (G–R) law, an exponential law in terms of earthquake magnitudes which transforms into a Pareto distribution in terms of scalar seismic moments.

Due to limitations on the sensitivity of seismographic networks, small earthquakes are generally missing from earthquake catalogs. Hence one typically fits a Pareto distribution that is truncated on the left at some positive value a, called the *completeness threshold*. For instance, the Harvard catalog is thought to be incomplete below the moment threshold of about $10^{17.7}$ Nm, corresponding to a magnitude of approximately 5.8 [12]. The Harvard catalog, when restricted to shallow earthquakes above this magnitude, contains 3765 events, and we shall refer only to this portion of the catalog in the remainder.

Additionally, the upper tail of the Pareto distribution has been modified in various ways, based on physical and statistical principles. Considerations of finiteness of seismic moment flux or of deformational energy suggest that the upper tail of the seismic moment distribution decays to zero more rapidly than that of the Pareto ([15], [25], [32]); this agrees with empirical observations that the G–R law appears to overpredict the frequencies of large seismic moments (see Figure 1). A simplistic modification proposed by several researchers is to use a Pareto distribution that is truncated at some upper threshold θ (e.g. [10], [26]).

Kagan [10] has argued that a sharp upper threshold does not agree with the known behavior of dissipative physical dynamic systems and fundamental seismological scaling principles. Furthermore, the fact that observations of seismic moments are recorded

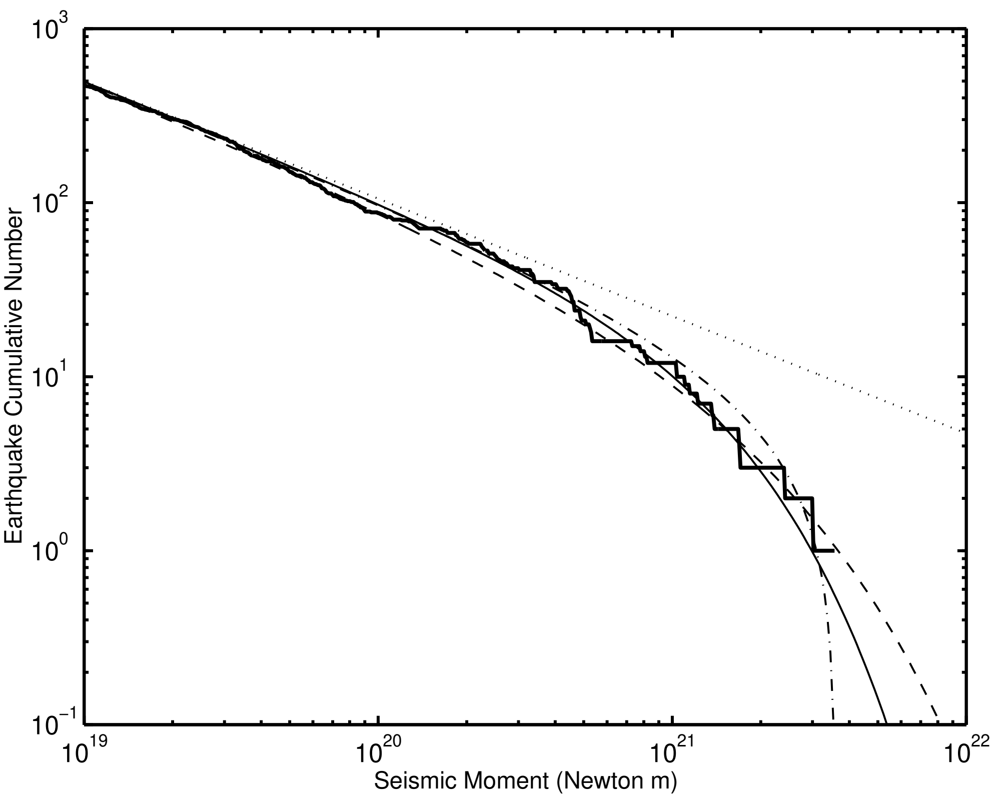

FIGURE 1: Earthquake cumulative number versus log seismic moment for the global shallow earthquakes in the 1977/1/1–1999/12/31 Harvard catalog. The curves show the numbers of events with seismic moment $\geq M$. Four approximations to the empirical distribution, for which the parameters β and θ are fitted by the maximum likelihood estimator (MLE), are shown: (i) the original G–R law without upper cutoff (the classical Pareto distribution) $(\cdots\cdots)$; (ii) the two-sided truncated Pareto distribution $(-\cdot-\cdot-)$; (iii) the left-truncated Gamma distribution with negative shape parameter (– – – –); and (iv) the tapered Pareto distribution (———).

with error contradicts the notion of a fixed point θ such that earthquakes with moment $\theta + \varepsilon$ are observed with very different frequencies compared to those with moment $\theta - \varepsilon$ (see [10]).

As an alternative, a Pareto distribution which tapers to zero via an exponential function, instead of sudden truncation, has been proposed by Vere-Jones *et al.* [30]. The tapered Pareto distribution, sometimes called the *modified G–R* law by seismologists [1] is characterized by its gradual rather than steep decrease in frequency at the upper cutoff moment, and provides better agreement with both fundamental seismological principles and earthquake catalogs ([7], [10]).

For example, Figure 1 shows the fit of various distributions to the Harvard catalog. One sees that the tapered Pareto distribution fits the data much better than either the simple Pareto distribution or the Pareto distribution truncated on the right at a maximum seismic moment. There is little perceptible difference between the four theoretical curves and the empirical distribution for seismic moments in the range 5×10^{17} Nm to 10^{19} Nm (not shown).

Several alternative models for the earthquake size distribution have been proposed. One example is the left-truncated Gamma distribution with negative shape parameter ([11], [13], [17]), which is similar to the tapered Pareto law: the former involves applying an exponential taper to the Pareto *density* function, while the latter involves applying an exponential taper to the *survivor* function. In addition, different parametric forms for the tapered upper moment of the Pareto distribution have been proposed ([18], [25], [26]). The tapered Pareto distribution of Vere-Jones *et al.* [30] is preferred largely because of its simplicity and paucity of free parameters; available data from catalogs of seismic moments do not seem to warrant more than two or three degrees of freedom (see Table 2 of [26]).

Note also that the arguments used to motivate the modified G–R law may suggest tapering rather than truncation at the lower end of the distribution as well, as discussed e.g. in [19] and [30]. Such modifications appear to be of limited seismological concern, however, and will not be treated here. Instead, our focus is on estimation of the one-sided tapered Pareto distribution described by Vere-Jones *et al.* [30] and Jackson and Kagan [7].

2. The tapered Pareto distribution

2.1. Characterization

The Pareto distribution and its variants have been used in numerous applications. See chapter 20 of [9] for a review, including a survey of historical developments and properties of the Pareto distribution and its various estimates.

The tapered Pareto (or *modified G–R*) law has cumulative distribution function

$$F(x) = 1 - \left(\frac{a}{x}\right)^{\beta} \exp\left(\frac{a-x}{\theta}\right) \qquad (a \le x < \infty), \tag{2}$$

and density

$$f(x) = \left(\frac{\beta}{x} + \frac{1}{\theta}\right)\left(\frac{a}{x}\right)^{\beta} \exp\left(\frac{a-x}{\theta}\right) \qquad (a \le x < \infty), \tag{3}$$

where a is the observational completeness threshold, β is a shape parameter governing the power-law decrease in frequency with seismic moment, and θ is an *upper cutoff* parameter governing the location of the exponential taper to zero in the frequency of large events. The distribution is a special case of what has been called a *generalized* Pareto distribution; see e.g. equation (20.152) of [9]. In fact, the distribution appears to have been first proposed by Vilfredo Pareto himself in 1897 ([20], pp. 305-306, Eqs. 2 and 5).

Typically the lower threshold moment a in (1) is presumed known, and only the parameters θ and/or β must be estimated. [Note that the notation here differs slightly from the usual seismological notation (cf. [7], [12]), which uses the notation M_t and either $M_{\max}$ or M_c in place of a and θ, respectively, and the symbol m_c to denote $\phi(\theta)$, the *magnitude* of the upper cutoff, where ϕ is defined in (1). Estimable parameters are denoted here by Greek letters, in agreement with statistical convention.]

The characteristic function associated with F defined in (2) is

$$\begin{aligned}\phi(t) = \mathrm{E}[\exp(itX)] &= \int_a^\infty \exp\left(\frac{a-(1-\theta it)x}{\theta}\right)\left(\frac{\beta}{x}+\frac{1}{\theta}\right)\left(\frac{a}{x}\right)^\beta \mathrm{d}x \\ &= \exp(a/\theta)\beta a^\beta \frac{\theta it}{\theta it-1}\int_a^\infty x^{-\beta-1}\exp[(\theta it-1)x/\theta]\,\mathrm{d}x + \frac{\exp(ait)}{1-\theta it} \\ &= -\beta it\exp(a/\theta)\left[\frac{a(1-\theta it)}{\theta}\right]^\beta \Gamma(-\beta, a(1-\theta it)/\theta) + \frac{\exp(ait)}{1-\theta it}, \qquad (4)\end{aligned}$$

where $\Gamma(y,z) = \int_z^\infty \mathrm{e}^{-t}t^{y-1}\,\mathrm{d}t$ denotes the incomplete gamma function; $\Gamma(\cdot,\cdot)$ satisfies the relation

$$\Gamma(y+1,z) = y\Gamma(y,z) + z^y\mathrm{e}^{-z} \qquad (0<z<\infty,\ y\neq 0,-1,\ldots). \qquad (5)$$

2.2. Moments

The moments of F may be obtained directly from (3). For instance, the first moment $\mu = \mathrm{E}(X)$ is given by

$$\begin{aligned}\mu &= \int_a^\infty x\left(\frac{\beta}{x}\right)\left(\frac{a}{x}\right)^\beta \exp\left(\frac{a-x}{\theta}\right)\mathrm{d}x + \int_a^\infty x\left(\frac{1}{\theta}\right)\left(\frac{a}{x}\right)^\beta\exp\left(\frac{a-x}{\theta}\right)\mathrm{d}x \\ &= \beta a^\beta\exp(a/\theta)\int_a^\infty x^{-\beta}\exp(-x/\theta)\,\mathrm{d}x + a^\beta\exp(a/\theta)/\theta\int_a^\infty x^{1-\beta}\exp(-x/\theta)\,\mathrm{d}x. \qquad (6)\end{aligned}$$

By a simple change of variable one obtains

$$\int_a^\infty x^{-\beta}\exp(-x/\theta)\,\mathrm{d}x = \theta^{1-\beta}\Gamma(1-\beta, a/\theta), \qquad (7)$$

and integration by parts yields

$$\begin{aligned}\int_a^\infty x^{1-\beta}\exp(-x/\theta)\mathrm{d}x &= \theta a^{1-\beta}\exp(-a/\theta) + \theta(1-\beta)\int_a^\infty x^{-\beta}\exp(-x/\theta)\,\mathrm{d}x \\ &= \theta a^{1-\beta}\exp(-a/\theta) + \theta(1-\beta)\theta^{1-\beta}\Gamma(1-\beta,a/\theta). \qquad (8)\end{aligned}$$

Substituting (7) and (8) into (6) yields, after some cancellation,

$$\mu = a + a^\beta\theta^{1-\beta}\exp(a/\theta)\Gamma(1-\beta, a/\theta). \qquad (9)$$

Continuing in this fashion one readily obtains

$$\mathrm{E}(X^2) = a^2 + 2a^\beta\theta^{2-\beta}\exp(a/\theta)\Gamma(2-\beta,a/\theta), \qquad (10)$$

and a general formula for the moments of higher order,

$$\mathrm{E}(X^k) = a^k + ka^\beta\theta^{k-\beta}\exp(a/\theta)\Gamma(k-\beta,a/\theta). \qquad (11)$$

2.3. Simulation

Vere-Jones *et al.* [30] note that the product form of the distribution (2) admits an interpretation in terms of *competing risks*, namely, the survivor function $S(x) = 1 - F(x)$ is the product of the survivor functions of a Pareto random variable and an exponential random variable. This observation shows that a very simple method for simulating from F is to take the minimum of independently simulated Pareto and exponential random variables (see [30] for details).

3. Estimation of θ and β

The parameters of the tapered Pareto distribution are most commonly fitted by maximum likelihood. When independent, identically distributed (i.i.d.) observations $x_1, \ldots,$ x_n come from the distribution (2), the log-likelihood function takes the form

$$\log L(\beta, \theta) = \sum_{i=1}^{n} \log\left(\frac{\beta}{x_i} + \frac{1}{\theta}\right) + \beta n \log a - \beta \sum_{i=1}^{n} \log x_i + \frac{an}{\theta} - \frac{1}{\theta} \sum_{i=1}^{n} x_i. \qquad (12)$$

Setting the derivative of $\log L$ with respect to the parameters θ and β to zero yields the relations

$$\frac{\theta}{n} \sum_{i=1}^{n} \frac{x_i}{\beta\theta + x_i} = \bar{x} - a \qquad (13)$$

and

$$\theta \sum_{i=1}^{n} \frac{1}{\beta\theta + x_i} = \sum_{i=1}^{n} \log x_i - n \log a, \qquad (14)$$

where $\bar{x}$ is the sample mean $(x_1 + \ldots + x_n)/n$. Approximate simultaneous solutions to these equations may be obtained via a numerical fitting routine such as Newton–Raphson optimization (see [30]).

It is convenient to write

$$\eta = \frac{1}{\theta}, \qquad A = \frac{1}{n} \sum_{i=1}^{n} \log \frac{x_i}{a}, \qquad B = \bar{x} - a. \qquad (15)$$

Then, as noted in section A4 of [30], the maximum likelihood estimates $\hat{\beta}$ and $\hat{\eta}$ satisfy

$$\hat{\beta} A + \hat{\eta} B = 1, \qquad (16)$$

so, writing $v_i = x_i/a$,

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{1 - \hat{\eta}(B - Av_i)} = 1\,. \qquad (17)$$

The solution $\hat{\eta}$ to this last equation must satisfy $0 \leq \hat{\eta} < 1/(B - A \min_i\{v_i\})$, and consequently, assuming that the maximum likelihood estimates satisfy

$$\hat{\eta} > 0, \qquad \hat{\beta} > 0, \qquad (18)$$

a Newton–Raphson iteration starting from $\hat{\eta}_0 = 1/B$ converges to $\hat{\eta}$.

It is generally important to discriminate between the case where two parameters are estimated and the case where only one is estimated, since in the one-parameter case only one of the equations based on the derivative of the log-likelihood is justified. An exception is the case where either β or η is known to be zero. When $\beta = 0$, for instance, the distribution (2) reduces to a purely exponential form, and the maximum likelihood estimate $\hat{\eta} = 1/\hat{\theta} = 1/B$, in agreement with (16). Similarly, when η is known to be 0, the distribution (2) is a pure Pareto, and $\hat{\beta} = 1/A$, as is consistent with (16).

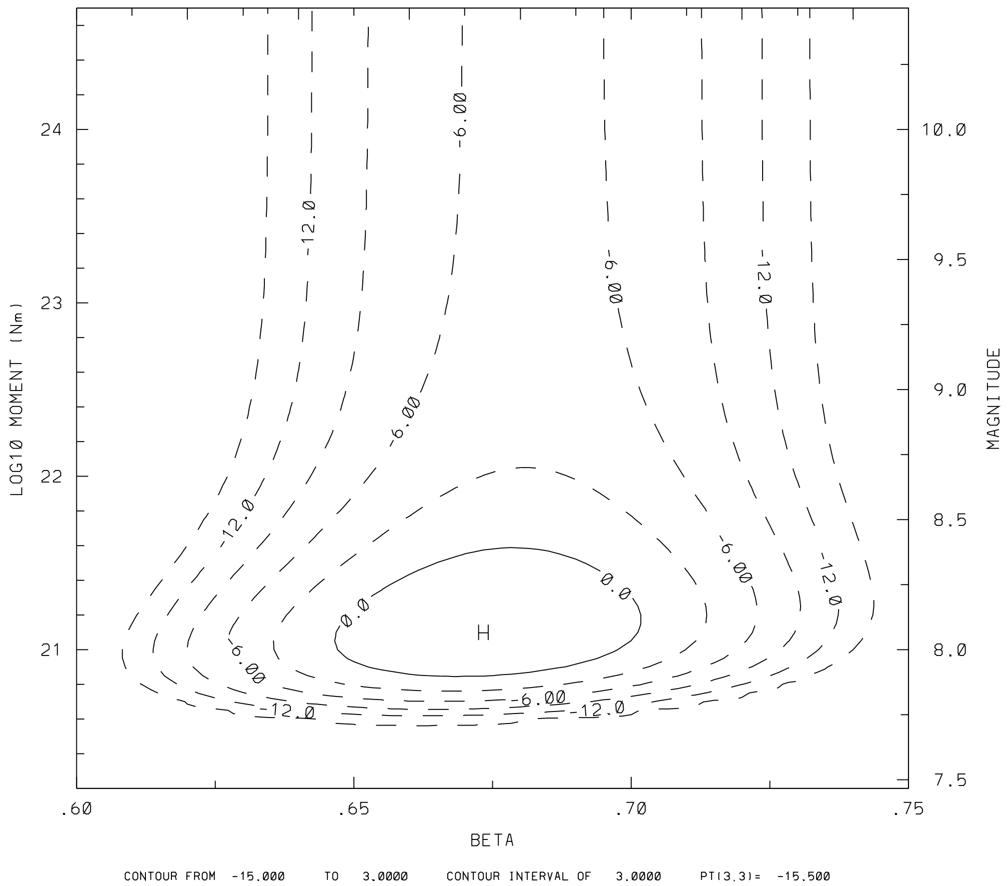

FIGURE 2: Log-likelihood map for the distribution of scalar seismic moments: The Harvard catalogue time span is January 1, 1977 to December 31, 1999; the completeness threshold a is $10^{17.7}$ Nm; the number of shallow events is 3765. Approximation by the tapered Pareto distribution. The solid contour line denotes a 95% confidence bound based on asymptotic theory [32].

In practice, the problem of estimating β is nowhere near as difficult as that of estimating θ. To illustrate this, Figure 2 shows a contour plot of $\log L(\theta)$ as a function of β and θ, using the data from the Harvard catalog. The likelihood function has been normalized so that its maximum is 3.0. The log-likelihood, as a function of β, appears nearly symmetric around a well-defined maximum, and approximate 95%-confidence intervals can be constructed in a straight-forward manner, based on the convergence of the negative reciprocal of elements of the diagonal of the Hessian of the log-likelihood function to χ^2 (chi-squared) random variables (see e.g. [31], Chapter 13.8). By contrast, the log-likelihood as a function of θ decays very slowly from its maximum and is highly non-symmetric. An approximate 95%-confidence interval, based on asymptotic relations, corresponds to the contour labeled 0.0. Although the interval is bounded in Figure 2, for smaller subcatalogs this is not the case. Even for relatively large datasets of the size of the full Harvard catalog, simulations show that confidence intervals based on asymptotic theory have inappropriate coverages.

The source of this problem is the fact that estimation of θ is essentially dominated by only the largest events. Hence in the Harvard catalog of 3765 shallow earthquakes, it is just a small number of them that heavily influence the determination of the upper cutoff parameter θ [21]: indeed, only 12 of these earthquakes have seismic moment in excess of 10^{21} Nm (8.0 in magnitude), the value corresponding to the maximum likelihood estimator (MLE) of θ (see Figures 1 and 2). Since it is these events which dominate the estimation of θ, the relevance of asymptotic results for maximum likelihood estimates, such as asymptotic normality, consistency, and unbiasedness, is questionable in the absence of numerous events *above the upper cutoff itself.*

In addition to being easily estimable, there is some evidence that the parameter β may be constant globally, or at least for certain classes of shallow earthquakes, and hence need not be estimated simultaneously with θ using individual local earthquake catalogs. Indeed, Kagan [12] has shown that there is no statistically significant variation in β for all subduction and continental seismic regions; estimated values of β for all such earthquakes range from 0.60 to 0.70. Further, recent analysis of oceanic earthquakes [1] suggests that the distributions of these events have β-values similar to subduction and continental earthquakes. By contrast, whereas $m_c = \phi(\theta)$ is typically about 8.0 for subduction and continental regions (see Figure 2), estimates of $\phi(\theta)$ vary from 5.8 to 7.2 for oceanic earthquakes [1].

Further, there exists theoretical justification for the global value of β: Vere-Jones ([27], [28]) derived the value $\beta = 0.5$ for a critical branching process which was proposed as a general model for the earthquake size distribution (see also further discussion by Kagan [11] and Kagan and Vere-Jones [14]). If the process is slightly subcritical, the resulting distribution is similar to the tapered Pareto distribution (2). Observational estimates of β are slightly higher than the theoretical value mentioned above; the cause of this discrepancy is not yet clear ([11], [12]).

Thus, given a particular earthquake catalog, one may wish to assume a global value of β and concentrate exclusively on the estimation of θ. Since seismologists usually use the G–R distribution with b-value corresponding to $\beta = \frac{2}{3}$, we use this value in our considerations below; the results do not significantly change if a slightly different value for β is assumed, e.g. β in the range 0.60 to 0.63 as suggested in [12]. How our results below are affected by variations in β is discussed further at the end of Section 5. However the focus of the remainder of this paper is on estimation of θ alone.

4. Estimation of θ only

We now consider the case where one wishes to estimate only the cutoff parameter θ, the other parameters a and β being known.

4.1. Maximum likelihood estimates

In the case where only θ is being estimated, the log-likelihood function (12) is a function of only one variable. Thus one need merely obtain a solution to the single nonlinear equation (13). An approximate solution may be obtained from any of a variety of standard numerical procedures.

The maximum likelihood estimator $\hat{\theta}$ is justified primarily because of its desirable asymptotic properties. For small samples, however, maximum likelihood estimates of the cutoff parameter θ can be heavily biased, and conventional formulae for standard errors and quantiles based on asymptotic relations may be highly misleading. As mentioned in Section 2, even for relatively large catalogs such as the Harvard dataset, asymptotic properties are less relevant than small-sample properties when it comes to estimating θ, since the determination of $\hat{\theta}$ depends essentially on the very few largest events.

Consider i.i.d. observations $x_1, \ldots, x_n$ from any distribution truncated from above at an unknown value θ. Pisarenko and others ([21], [22]) derived an expression for the bias in the MLE of the truncation point for a general class of truncated random variables. The origin of the bias in the maximum likelihood estimator $\hat{\theta}$ comes from the fact that the likelihood function $L(\theta)$ is at its peak when $\theta = \max\{x_i\}$. Hence the estimator $\hat{\theta}$ is *never* greater than the true value of θ, for any realizations $x_1, \ldots, x_n$, so inflation of this estimator is unilaterally sensible.

In the case where $x_1, ..., x_n$ are uniformly distributed random variables on $[0, \theta]$, for instance, it is well known that the maximum likelihood estimator $\hat{\theta} = \max\{x_i\}$ is biased by a factor of $(n-1)/n$ (see e.g. p. 289 of [30]). In the case where x_i are distributed according to a truncated Pareto law, the bias in the MLE of θ is much greater than in the uniform case, the relative infrequency of large values contributing to the expected discrepancy between $\max\{x_i\}$ and θ.

The case is essentially similar for estimating θ in the tapered Pareto distribution (2) using a small sample. Unfortunately, due to the lack of an expression in closed form for the maximum likelihood estimator for θ, no simple formula for its bias is available. However, the bias may be approximated via simulation; see Section 5 below.

Note that relation (16) suggests the estimator $\check{\theta}$ given by

$$\check{\theta} = \frac{B}{1 - \beta A}. \tag{19}$$

Since it is based on likelihood equations, one may expect the behavior of $\check{\theta}$ to be similar to that of the MLE $\hat{\theta}$. However, since β is not being estimated, equation (14), which is obtained by setting the partial derivative of the log-likelihood function with respect to β to zero, is meaningless. In practice the denominator in (19) tends to be very nearly zero, so the estimator $\check{\theta}$ is extremely unstable; this is confirmed by simulations as described briefly in Section 5 below.

4.2. Estimation of θ based on moments

As an alternative to the maximum likelihood estimator $\hat{\theta}$, an estimator of θ based on the first two moments of X may be constructed.

Using (5), one may rewrite (10) as

$$\mathrm{E}(X^2) = a^2 + 2a\theta + 2(1-\beta)a^{\beta}\theta^{2-\beta}\exp(a/\theta)\Gamma(1-\beta, a/\theta). \tag{20}$$

Combining (9) and (20) yields

$$\mathrm{E}(X^2) = a^2 + 2a\theta + 2\theta(1-\beta)(\mu - a), \tag{21}$$

and rearranging terms gives

$$\theta = \frac{\mathrm{E}(X^2) - a^2}{2[a\beta + (1-\beta)\mu]}. \tag{22}$$

Given i.i.d. observations $x_1, \ldots, x_n$, one may substitute the uncentered first and second sample moments $\bar{x} = \sum_{i=1}^n x_i/n$ and $\sum_{i=1}^n x_i^2/n$ in equation (22), leading to the estimator

$$\tilde{\theta} = \frac{\sum x_i^2/n - a^2}{2[a\beta + (1-\beta)\bar{x}]}. \tag{23}$$

The simplicity and closed analytic form of the equation for $\tilde{\theta}$ facilitates the derivation of properties of the estimator, including small-sample properties. For the MLE $\hat{\theta}$, by contrast, estimates of small-sample bias and rates of convergence are unavailable.

Theorem 1. *Suppose $x_1, \ldots, x_n$ are independently sampled from the distribution* (2), *with $a > 0$. Then the method of moments estimator $\tilde{\theta}$ is asymptotically unbiased and consistent; its mean is given by*

$$\mathrm{E}(\tilde{\theta}) = \theta + \frac{(\beta-1)[2a^3 + 3a^2\theta\beta + (\sigma^2+\mu^2)(6\theta - 3\theta\beta - 2\mu)]}{4n[a\beta + (1-\beta)\mu]^2} + O\left(n^{-2}\right). \tag{24}$$

Proof. The moments of $\tilde{\theta}$ may readily be approximated using the technique of *linearization* (or more appropriately *polynomialization*), described e.g. in [3], [4] and [23].

Let $g_n(u) = \frac{1}{2}\left(\sum_i x_i^2/n - a^2\right)/[a\beta + (1-\beta)u]$, and let $y_n = \bar{x} - \mu$. Then we may write

$$\begin{aligned}\tilde{\theta} = g_n(\mu + y_n) &= g_n(\mu) + y_n g_n'(\mu) + y_n^2 g_n''(\mu)/2! + \cdots \\ &= \frac{\sum x_i^2/n - a^2}{2[a\beta + (1-\beta)\mu]} \sum_{k=0}^{\infty} \left(\frac{y_n(\beta-1)}{a\beta + (1-\beta)\mu}\right)^k,\end{aligned} \tag{25}$$

provided the Taylor series converges. Note the convergence property

$$\Pr\left\{\left|\frac{y_n(\beta-1)}{a\beta + (1-\beta)\mu}\right| < 1\right\} \to 1 \qquad (n \to \infty). \tag{26}$$

Indeed, for any positive ε, δ, $\Pr\{|y_n| > \varepsilon\} \le \sigma^2/(n\varepsilon^2) < \delta$ for $n > \sigma^2/(\delta\varepsilon^2)$ by Chebyshev's inequality, since y_n has mean 0 and variance σ^2/n.

In order to show convergence of the mean of $\tilde{\theta}$ via the expected value of the Taylor expansion (25), it is not enough to show that the Taylor series converges on a set with probability going to one; one must also ensure that the expected value decreases to zero on the set S where the Taylor series fails to converge. But this fact is ensured by noting that the expected value of $\tilde{\theta}$ is bounded even on S: the denominator in (23) must be no smaller than $2a$ since each $x_i \ge a$; hence $\mathrm{E}(\tilde{\theta}; S) \le P(S) \times \frac{1}{2}[\mathrm{E}(X^2) + a^2]/a$, which converges to zero with $n \to \infty$ since $P(S) \to 0$.

Since $\tilde{\theta}$ is expressed in (25) as a polynomial function of the first two sample moments, computation of the approximate moments of $\tilde{\theta}$ is straightforward. For instance, the expected value of $\tilde{\theta}$ is given by

$$\begin{aligned}\mathrm{E}(\tilde{\theta}) &= \mathrm{E}\left[\frac{\sum x_i^2/n - a^2}{2[a\beta + (1-\beta)\mu]}\right] + \mathrm{E}\left[\frac{(\sum x_i^2/n - a^2)(\bar{x}-\mu)(\beta-1)}{2[a\beta + (1-\beta)\mu]^2}\right] + O\left(n^{-2}\right) \\ &= \theta + \tfrac{1}{2}(\beta-1)\left[a\beta + (1-\beta)\mu\right]^{-2}\left[\mathrm{E}\left(\bar{x}\sum x_i^2/n\right) - \mu\mathrm{E}(X^2)\right] + O\left(n^{-2}\right),\end{aligned} \tag{27}$$

since the expectation of the higher-order terms in (25) are of the order of n^{-2}.

Note that

$$\mathrm{E}\left(\bar{x}\sum x_i^2/n\right) = \mathrm{E}(X^3)/n + \mu\mathrm{E}(X^2) - \mu\mathrm{E}(X^2)/n, \tag{28}$$

and that, using (5) and (11),

$$\mathrm{E}(X^3) = a^3 + (3\theta/2)\left[(2-\beta)E(X^2) + \beta a^2\right]. \tag{29}$$

Combining (27), (28) and (29) yields (24). Hence $\tilde{\theta}$ is asymptotically unbiased. A calculation similar to (27) shows that the variance of $\tilde{\theta}$ is of the order of $1/n$, which together with the asymptotic unbiasedness of $\tilde{\theta}$ implies that $\tilde{\theta}$ is consistent. □

Relation (24) suggests adjusting the estimator $\tilde{\theta}$ by subtracting its bias. Note however that the approximate bias in $\tilde{\theta}$ depends on the quantities μ, σ^2, and θ, which are generally unknown. One may plug in estimates of these quantities to obtain an *adjusted* moment estimator

$$\tilde{\theta}_a \ := \ \tilde{\theta} \ - \ \frac{(\beta-1)[2a^3 + 3a^2\tilde{\theta}\beta + (\hat{\sigma}^2 + \bar{x}^2)(6\tilde{\theta} - 3\tilde{\theta}\beta - 2\bar{x})]}{4n[a\beta + (1-\beta)\bar{x}]^2}. \tag{31}$$

Although the estimator $\tilde{\theta}_a$ may be very nearly unbiased, the adjustment factor can introduce substantial variance in some cases. These features are discussed further in Section 5.

4.3. Average likelihood estimation of θ

A third estimator, $\breve{\theta}$, may be obtained by computing the mean of θ with respect to the likelihood measure $\mathrm{d}L(\theta)$, namely,

$$\breve{\theta} := \frac{\int \theta L(\theta)\,\mathrm{d}\theta}{\int L(\theta)\,\mathrm{d}\theta}. \tag{32}$$

This Average Likelihood Estimator (ALE) is equivalent to the posterior mean of θ obtained in Bayesian estimation, starting with a non-informative prior on θ. (Note that some authors call this estimator the *mean* likelihood estimator, abbreviated *mele*; see e.g. equation 4.4.15 of Jenkins and Watts [8].)

Unfortunately, in many cases the integral in the numerator of (31) does not converge. When the sample size n is relatively small, the likelihood function L decays very slowly as $\theta \to \infty$, as seen in Figure 2. In such cases, one may choose to modify the ALE $\breve{\theta}$ in some fashion. For instance, instead of a non-informative prior, one may adopt a prior that gives very little weight to large values of θ. However, clearly in this framework the resulting estimate will depend heavily on the choice of prior.

Alternatively, one may set $\eta = \theta^{-1}$ and estimate η using the ALE $\breve{\eta}$. One may then wish to employ the *inverse* ALE estimate $\breve{\theta}_i := 1/\breve{\eta}$ as an estimator of θ.

Figure 3 shows the log-likelihood as a function of η, for a simulated dataset. While the log-likelihood decays very slowly as a function of θ in Figure 2, one sees from Figure 3 that the log-likelihood decays very rapidly with η and hence the integrals in (31) generally converge when estimating η.

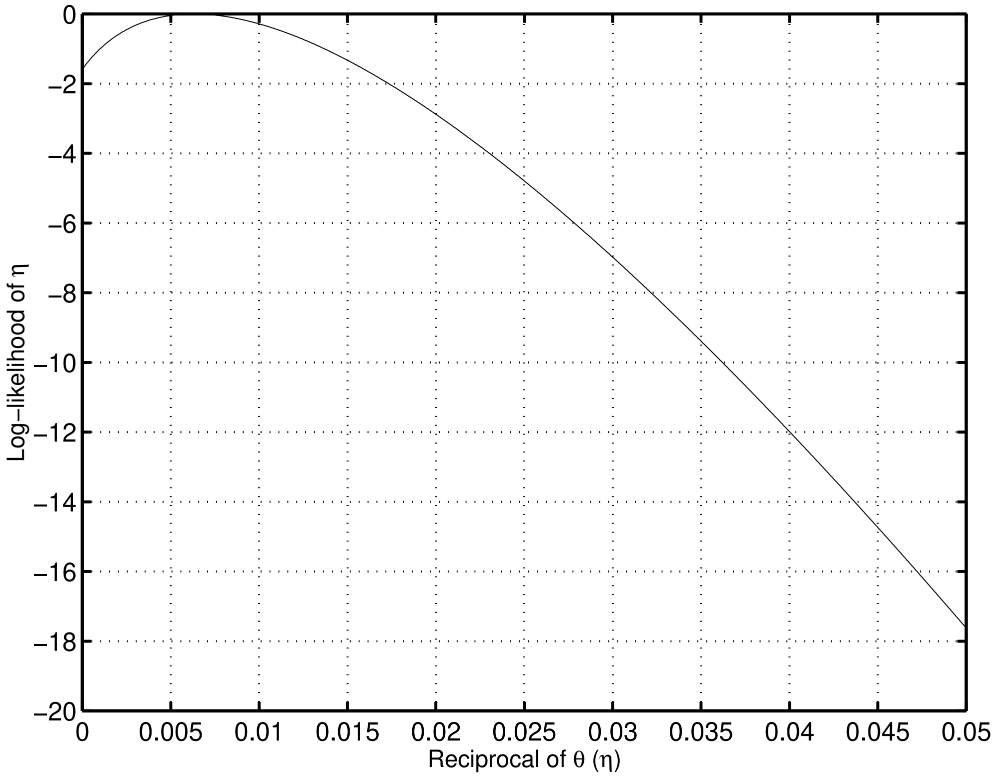

FIGURE 3: $\log L$ versus $\eta = 1/\theta$, for a simulated dataset consisting of $n = 100$ independent draws from equation (2), with $a = 1$, $\beta = \frac{2}{3}$, $\theta = 1000$.

In practice, computation of $\breve{\theta}$ and $\breve{\theta}_i$ is nontrivial. Typically, the integrals in (31) must be evaluated numerically, with each computation requiring an evaluation of the likelihood L, which itself requires a pass through the n observations. Further, in the numerical evaluation of the integrals one must experiment with various upper limits and step-sizes in order to ensure convergence and accuracy, respectively, of the numerical integration.

5. Comparison of estimators

The estimators described previously are compared, first on a linear scale (i.e. estimates of θ are compared to θ) and subsequently on a logarithmic scale, where estimates of θ are converted to units of magnitude via the function ϕ defined in (1), and compared to $\phi(\theta)$.

5.1. Comparison on linear scale

The method of moments estimator $\tilde{\theta}$ has several attractive features. First, it is extremely easy to compute: no iteration is required, nor use of numerical approximation routines. Second, as noted in Section 4.2, expressions for the approximate small-sample moments of $\tilde{\theta}$ can be obtained. Third, the variance of $\tilde{\theta}$ appears in many instances to

TABLE 1: Simulated estimates of θ

	$\hat{\theta}$			$\tilde{\theta}$			$\tilde{\theta}_a$			$\breve{\theta}_i$		
n	bias	sd	rmse	bias	sd	rmse	bias	sd	rmse	bias	sd	rmse
25	–335	1257	1301	–612	674	910	–30	2139	2139	–763	371	848
50	–140	1330	1337	–459	752	881	128	2081	2085	–642	429	772
100	–6	1240	1240	–311	765	826	167	1738	1746	–489	470	678
250	48	914	915	–160	675	694	108	1117	1122	–270	487	557
500	36	638	639	–88	555	562	58	740	742	–139	456	477
1000	20	435	435	–47	428	431	27	496	497	–65	378	384
2500	9	267	267	–19	287	311	11	304	304	–25	261	262
5000	4	187	187	–10	207	207	5	213	213	–12	187	187

be a bit smaller than that of $\hat{\theta}$. One may expect the adjusted method of moments estimator, $\tilde{\theta}_a$, to have considerably smaller bias than $\hat{\theta}$; however, $\tilde{\theta}_a$ has larger variance due to the high variance of the adjustment factor. In fact, when the sample size is small (i.e. only a few hundred earthquakes, and hence only a handful of *large* events, are observed), the bias in the unadjusted estimators $\hat{\theta}$, $\tilde{\theta}$, and $\breve{\theta}_i$ is very substantial, and use of the adjustment factor may be sensible despite the resulting increase in variance.

The above results regarding the relative performance of estimators may be demonstrated experimentally via simulation. For instance, Table 1 summarizes the performance of various estimators of θ for the case where $a = 1$, $\beta = \frac{2}{3}$, $\theta = 1000$. Each row of Table 1 shows the result of thousands of simulated catalogs; for each such catalog, an estimate of θ is obtained from n independent draws from the distribution (2), as described in Section 2.3. The bias (expressed as the estimated value minus the true value), the standard deviation (sd) and the root-mean-squared error (rmse) of the estimates are reported in Table 1. Whereas as a general principle of estimation there tends to be a tradeoff between bias and variance, the rmse, which can be computed as the square root of the sum of the squared bias and the variance, provides a useful measure indicating the typical size of an estimator's error (see e.g. [9], p. 128). With the exception of $\breve{\theta}_i$, for each estimator and each row of Table 1, the total number of simulated events (i.e. n times the number of simulated catalogs) is approximately 2.5×10^8. Because of the computational burden in computing $\breve{\theta}_i$, for this estimator the total number of simulated events in each row is 5×10^7.

The entries in Table 1 show that the bias in the unadjusted estimators $\hat{\theta}$, $\tilde{\theta}$, and $\breve{\theta}_i$ is very large when the sample size is small, and diminishes quite rapidly as the sample size increases. For the maximum likelihood estimator $\hat{\theta}$, somewhat surprisingly, the bias becomes positive for samples of size 250 before beginning to approach zero. The *adjusted* moment estimator $\tilde{\theta}_a$ is very nearly unbiased but has such large variance that its root-mean-squared error is larger than that of the other estimators. Note however that the observed bias in $\tilde{\theta}_a$ may not reasonably be attributed to chance variation. Indeed, since only the first two terms in the Taylor expansion (25) were used in constructing the adjustment factor in $\tilde{\theta}_a$, and since the unknown θ appears in the adjustment factor and is estimated using the biased estimator $\tilde{\theta}$, it is to be expected that the bias in $\tilde{\theta}_a$ is nonzero.

Though the bias in the ALE for small samples is very large, the root-mean-squared error is in each case smaller for the ALE than for the other estimators. Note also that in almost every case the rmse is smaller for the unadjusted method of moments estimator than for the MLE. However, the differences between the estimators appear to dissipate quickly as n increases. For very large sample sizes of many thousands of events, all the estimators considered here perform well. Even for $n = 1000$, the difference between $\tilde{\theta}$ and $\breve{\theta}_i$ is hardly substantial and may arguably be insufficient to justify the large increase in computational burden.

Note that the MLE $\hat{\theta}$ in Table 1 is obtained by finding an approximate numerical solution to the single equation (13), treating β as known. One may alternatively investigate the maximum likelihood estimate of θ based on the estimation of two parameters, i.e. the simultaneous solution to equations (13) and (14) or the inverse of the estimate $\hat{\eta}$ in (17). The estimate based on the two-parameter case has uniformly substantially higher rmse than the one-parameter estimate $\hat{\theta}$. In addition, the estimate $\breve{\theta}$ defined via (19) has substantially higher bias and variance compared to $\hat{\theta}$ for each set of simulations; its rmse is typically a factor of two or more greater than that of $\hat{\theta}$. As mentioned previously, the source of this problem seems to be the fact that the denominator in (19) is typically very nearly zero. In fact, in many cases the denominator (and hence the resulting estimate) is negative. This problem appears to be particularly severe for relatively small values of the ratio a/θ. In cases where the two terms $\hat{\beta}A$ and $\hat{\eta}B$ in the left-hand-side of (16) are of roughly equal size, the estimate $\breve{\theta}$ may perform reasonably well, but in the typical seismological case where η is usually very small, the estimator $\breve{\theta}$ is prohibitively unstable.

A word should be said about the choice of parameters in the simulations of Table 1. The results of the different estimators relative to one another were qualitatively similar when different choices of these parameters were used. In addition, the bias and variance of estimates of θ appear to depend critically on the number of events in the sample greater than θ, which in turn depends on the ratio $\rho = a/\theta$, and this fact can be used so as to shed light on the bias and variance in estimators of θ for catalogs of values of ρ different than $1/1000$. For instance, the global earthquake size distribution, shown in Figures 1 and 2, is dominated by shallow earthquakes in subduction and continental areas ([12], [16]), and parameter estimates suggest that the ratio $\rho = a/\theta$ is approximately $1/2000$ for these earthquakes in the Harvard catalog (see Figure 2). Catalogs of oceanic earthquakes tend to have much smaller empirical values of the ratio [1]. Fortunately our simulation results can be easily applied to these earthquakes and other earthquake catalogs, in the following manner. Suppose one obtains n' independent observations from a distribution with cutoff parameter θ' and completeness threshold a', and let $\rho' = a'/\theta'$. Then the appropriate value of n signifying which row in Table 1 corresponds to catalogs with a similar number of large events, and hence similar values of bias (relative to θ) and variance, satisfies:

$$n = n' \, (\rho'/\rho)^{\beta} \exp(\rho' - \rho). \tag{33}$$

For example, an earthquake catalog with $n' = 794$ events taken from the Harvard data with $\rho' = 1/2000$ has approximately the same number of events exceeding θ as a simulated catalog of $n = 500$ events from a distribution with $\rho = 1/1000$, and hence is approximately equivalent in terms of statistical properties of estimates of θ. Hence,

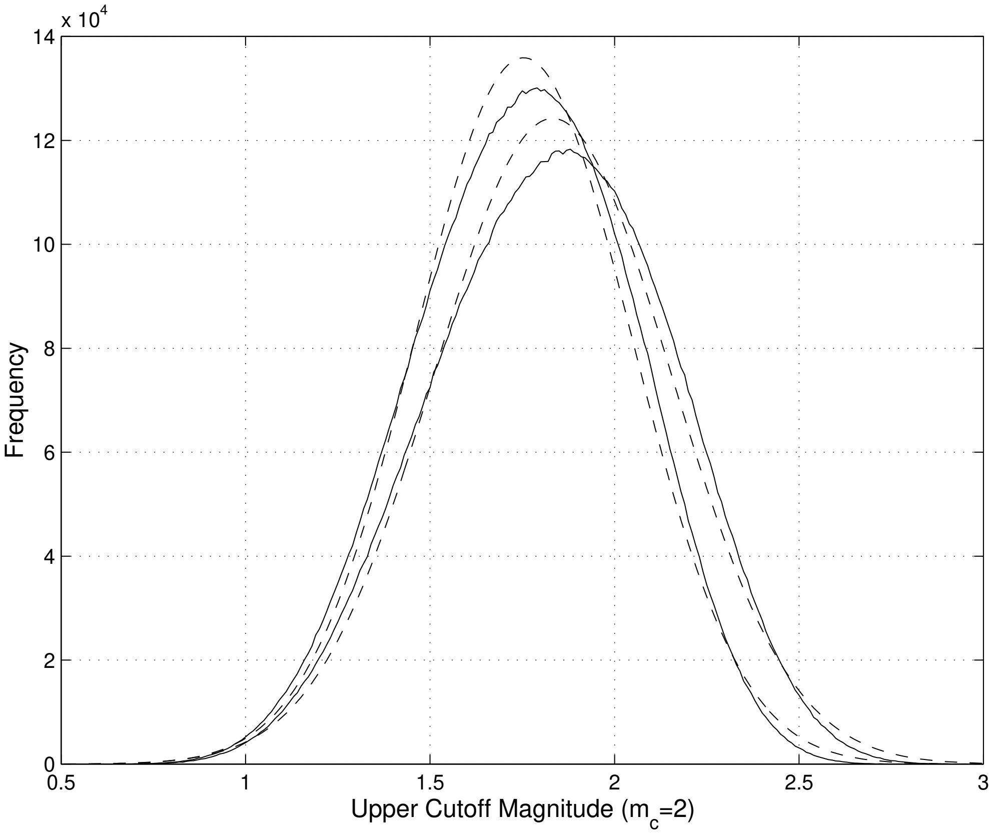

FIGURE 4: Histograms of $\phi(\hat{\theta})$ and $\phi(\tilde{\theta})$ (solid lines) for simulations each consisting of $n = 100$ independent draws from (2), with $a = 1$, $\beta = \frac{2}{3}$, $\theta = 1000$. The total number of simulated catalogs is 10^7. The Gaussian curves (dashed lines) are fitted to have the same mean and standard deviation as the simulated data. The proximity of the histograms of $\phi(\tilde{\theta}_a)$ and $\phi(\breve{\theta}_i)$ (not shown to avoid excessive overlap) to Gaussian curves was similar. For $\tilde{\theta}$ (left-most curves), the mean and sd of m_c are 1.753 and 0.293, respectively. For $\hat{\theta}$, the mean and sd of m_c are 1.832 and 0.320, respectively.

one need merely multiply the entries of the fifth row of Table 1 by ρ/ρ' to obtain approximations of the bias, sd, and rmse of the various estimators as applied to this sample subset of the Harvard catalog. This similarity is confirmed by our simulations.

5.2. Comparison on logarithmic scale.

It is important to assess estimates of θ on a logarithmic scale, for three reasons. First, the *moment magnitude* of an earthquake, expressed as the logarithm of the seismic moment via (1), is a quantity of great significance in seismology. Second, seismic moment data are obtained by an inversion process applied to seismographic recordings; the errors in the recordings and the inversion process appear to be approximately *multiplicative* [6]. Thus on a logarithmic scale, the errors in the seismic moment are approximately additive. Third, although the exact distributions of the estimators of θ have not yet been obtained, simulations indicate that the estimators $\hat{\theta}$, $\tilde{\theta}$, and $\breve{\theta}_i$

TABLE 2: Simulated estimates of θ on magnitude scale

	$\phi(\hat{\theta})$			$\phi(\tilde{\theta})$			$\phi(\tilde{\theta}_a)$			$\phi(\breve{\theta}_i)$		
n	bias	sd	rmse	bias	sd	rmse	bias	sd	rmse	bias	sd	rmse
25	−.463	.471	.660	−.568	.430	.712	−.423	.511	.663	−.657	.383	.760
50	−.291	.398	.493	−.386	.362	.529	−.262	.428	.502	−.462	.321	.563
100	−.168	.320	.361	−.247	.293	.383	−.151	.340	.372	−.302	.260	.399
250	−.072	.225	.236	−.126	.211	.246	−.068	.236	.246	−.151	.191	.243
500	−.037	.165	.169	−.072	.161	.176	−.037	.174	.178	−.081	.150	.170
1000	−.019	.119	.121	−.040	.121	.127	−.021	.127	.129	−.042	.114	.121
2500	−.007	.076	.076	−.017	.081	.083	−.009	.083	.083	−.017	.075	.077
5000	−.004	.053	.053	−.008	.059	.060	−.005	.059	.059	−.009	.054	.055

are very nearly log-normally distributed: for example, Figure 4 shows how close to Gaussian are the histograms of the logarithm of these estimates for the simulations summarized in Table 1. It follows that distributions of estimators of θ, expressed in terms of *magnitudes* rather than moments, are well-summarized by their means and standard deviations.

Table 2 presents the results of estimating $m_c = \phi(\theta)$, with ϕ defined as in (1), via the logarithms of the four estimators in Table 1. That is, for each of the simulations in Table 1, estimates $\hat{\theta}$, $\tilde{\theta}$, $\tilde{\theta}_a$, and $\breve{\theta}_i$ are obtained as before, converted to magnitude scale via the conversion (1), and compared to $\phi(\theta)$.

The results of Table 2 are somewhat different from those of Table 1. On the logarithmic scale, the root-mean-squared errors are very comparable for all four estimators regardless of sample size, and in contrast to the seismic moment scale of Table 1, the MLE $\hat{\theta}$ typically has the smallest root-mean-squared error of the four. Unlike on the seismic moment scale, the bias on the magnitude scale is negative for every estimator even for larger samples. Another important difference is that on the magnitude scale, the adjusted method of moments estimator $\tilde{\theta}_a$ offers little improvement in terms of bias compared to $\hat{\theta}$, and indeed in many cases actually has a bit *greater* bias than the MLE. Since the adjustment factor is based on the bias of $\tilde{\theta}$ in terms of seismic moment rather than magnitude, it is not surprising that the adjustment factor provides relatively little improvement in bias on the magnitude scale. On the other hand, it should be noted that as with the seismic moment scale, adjustment of the method of moments estimator on the logarithmic scale does result in smaller bias and greater variance for each sample size, when compared to the unadjusted method of moments estimator $\tilde{\theta}$. Further, as in Table 1, the differences between the estimators in Table 2 diminish rapidly as the sample size is increased. The results in Table 2 can be used directly for catalogs with different values of the ratio ρ' if the catalog size n is adjusted according to (32).

Additional simulations were performed to evaluate the influence of variations in β on estimates of θ. In these simulations we created synthetic catalogs using varying β-values, but in the estimation of θ, β was assumed to be $\frac{2}{3}$. We found that a decrease in the true value of β generally caused an increase in the average of the estimate $\phi(\tilde{\theta})$, with the amount of change depending on n and ρ. Conversely, an increase in β resulted in decreased average values of $\phi(\tilde{\theta})$. For example, for $n = 100$ and $\rho = \frac{1}{1000}$, a 3% decrease in the true value of β caused $\phi(\tilde{\theta})$ to increase by about 0.03 on average, thus decreasing the bias in $\phi(\tilde{\theta})$ from -0.247 to -0.215. Meanwhile, a 3% increase β caused $\phi(\tilde{\theta})$ to increase by about 0.03 on average, thus increasing the bias in $\phi(\tilde{\theta})$ from -0.247 to -0.279. For $n = 1000$, 3% variations in β resulted in corresponding changes of 0.02 in the bias of $\phi(\tilde{\theta})$. The effects on $\phi(\hat{\theta})$ of variations in β were very similar to those on $\phi(\tilde{\theta})$. The fact that the use of an improper value of β has an impact on the bias in the estimation of θ underscores the importance of accuracy in the determination of β. Fortunately, as discussed in Section 3, β typically appears to fall within a narrow range and may readily be estimated quite accurately. Since the bias in estimates of $\phi(\theta)$ induced by assuming an incorrect value of β appears to be comparable to the size of the error in β, this bias is not overwhelming when compared to typical variations in estimates of $\phi(\theta)$.

6. Summary

In estimating the upper cutoff parameter θ of the tapered Pareto distribution, the MLE $\hat{\theta}$ has substantial bias for small samples. The integrated likelihood required in the computation of the ALE $\breve{\theta}$, in addition to being quite cumbersome computationally, typically diverges. The inverse ALE $\breve{\theta}_i$ is certainly preferable to $\breve{\theta}$, though it is similarly difficult to compute and offers little improvement in performance when compared to $\hat{\theta}$, particularly when performance is evaluated on a logarithmic scale.

The method of moments estimator $\tilde{\theta}$ yields similar root-mean-squared error to the MLE and ALE, on both logarithmic and seismic moment scales, and differences in performance may be offset by the fact that the estimator $\tilde{\theta}$ is vastly simpler and faster to compute than the MLE and ALE. Though the adjusted method of moments estimator $\tilde{\theta}_a$ may appear to be preferable to the unadjusted estimators for small and moderate sample sizes due to its reduction in bias, it has increased root-mean-squared error due to the high variance of the adjustment factor. Here even samples of 1000 earthquakes may be considered relatively small, since estimation of θ is dominated by only the largest few observations. For very large sample sizes of hundreds of thousands of events, all the estimators considered here perform adequately. The choice between $\tilde{\theta}$ and $\tilde{\theta}_a$ appears to depend largely upon whether one is more interested in obtaining an estimate of low bias or of low variance; the adjusted estimator $\tilde{\theta}_a$ appears to offer substantially decreased bias at the expense of much additional variation.

Acknowledgements

We would like to thank D. Vere-Jones, Y. Ogata, V. F. Pisarenko, and D. D. Jackson for their helpful discussions. The editor, D. Daley, also provided numerous useful comments and suggested improvements. Yan Kagan appreciates partial support from the National Science Foundation through grant EAR 00-01128, as well as partial support from the Southern California Earthquake Center (SCEC). SCEC is funded by NSF Cooperative Agreement EAR-8920136 and USGS Cooperative Agreements 14-08-0001-A0899 and 1434-HQ-97AG01718. The SCEC contribution number is 557.

References

[1] Bird, P., Kagan, Y. Y. and Jackson, D. D. (2000). Frequency-magnitude distribution, effective lithosphere thickness, and seismic efficiency of oceanic transforms and spreading ridges. *Eos Trans. AGU* **81**(22), (abstract), p. WP147.

[2] DALEY, D. J. AND VERE-JONES, D. (1988). *An Introduction to the Theory of Point Processes.* Springer-Verlag, New York.
[3] DAVID, F. N. AND JOHNSON, N. L. (1951). The effect of non-normality on the power function of the F-test in the analysis of variance. *Biometrika* **38**, 43–57.
[4] DUSENBERRY, W. E. AND BOWMAN, K. O. (1977). The moment estimator for the shape parameter of the Gamma distribution. *Comm. Stat. B* **1**, 1–19.
[5] DZIEWONSKI, A. M., EKSTRÖM, G. AND MATERNOVSKAYA, N. N. (2000). Centroid-moment tensor solutions for July-September 1999. *Phys. Earth Planet. Inter.* **119**, 311–319.
[6] FROHLICH, C. AND DAVIS, S. D. (1999). How well constrained are well-constrained T, B, and P axes in moment tensor catalogs? *J. Geophys. Res.* **104**, 4901–4910.
[7] JACKSON, D. D. AND KAGAN, Y. Y. (1999). Testable earthquake forecasts for 1999. *Seism. Res. Lett.* **70**(4), 393–403.
[8] JENKINS, G. M. AND WATTS, D. G. (1968). *Spectral Analysis and its Applications.* Holden-Day, San Francisco.
[9] JOHNSON, N. L., KOTZ, S. AND BALAKRISHNAN, N. (1995). *Continuous Univariate Distributions.* 2nd edn. Wiley, New York.
[10] KAGAN, Y. Y. (1993). Statistics of characteristic earthquakes. *Bull. Seismol. Soc. Amer.* **83**, 7–24.
[11] KAGAN, Y. Y. (1994). Observational evidence for earthquakes as a nonlinear dynamic process. *Physica D* **77**, 160–192.
[12] KAGAN, Y. Y. (1999). Universality of the seismic moment-frequency relation. *Pure Appl. Geoph.* **155**, 537–573.
[13] KAGAN, Y. Y. AND KNOPOFF, L. (1984). A stochastic model of earthquake occurrence. *Proc. 8th Int. Conf. Earthq. Eng., San Francisco, Calif.* **1**, 295–302.
[14] KAGAN, Y. Y. AND VERE-JONES, D. (1996). Problems in the modelling and statistical analysis of earthquakes. In *Lecture Notes in Statistics (Athens Conference on Applied Probability and Time Series Analysis)* **114**, C. C. Heyde, Yu. V. Prohorov, R. Pyke, and S. T. Rachev, eds. New York, Springer, 398–425.
[15] KNOPOFF, L. AND KAGAN, Y. Y. (1977). Analysis of the theory of extremes as applied to earthquake problems. *J. Geophys. Res.* **82**, 5647–5657.
[16] LAY, T. AND WALLACE, T. C. (1995). *Modern Global Seismology*. Academic Press, San Diego.
[17] MAIN, I. G. (1996). Statistical physics, seismogenesis, and seismic hazard. *Rev. Geophys.* **34**, 433–462.
[18] MOLCHAN, G. M. AND PODGAETSKAYA, V. M. (1973). Parameters of global seismicity, I. In *Computational Seismology* **6**, Keilis-Borok, V. I., ed., Nauka, Moscow, 44–66 (in Russian).
[19] OGATA, Y. AND KATSURA, K. (1993). Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues. *Geophys. J. Int.* **113**, 727–738.
[20] PARETO, V. (1897). *Cours d'Économie Politique*, Tome Second, Lausanne, F. Rouge, quoted by Pareto, V. (1964), *Œuvres Complètes*, Publ. by de Giovanni Busino, Genève, Droz, v. II.
[21] PISARENKO, V. F. (1991). Statistical evaluation of maximum possible earthquakes. *Phys. Solid Earth* **27**(9), 757–763, (English translation).
[22] PISARENKO, V. F., LYUBUSHIN, A. A., LYSENKO, V. B. AND GOLUBEVA, T. V. (1996). Statistical estimation of seismic hazard parameters: maximum possible magnitude and related parameters. *Bull. Seis. Soc. Amer.* **86**, 691–700.
[23] SHENTON, L. R., BOWMAN, K. O. AND SHEEHAN, D. (1971). Sampling moments of moments associated with univariate distributions. *J. Roy. Statist. Soc. Ser. B* **33**, 444–457.
[24] SIPKIN, S. A., BUFE, C. G. AND ZIRBES, M. D. (2000). Moment-tensor solutions estimated using optimal filter theory: global seismicity, 1998. *Phys. Earth Planet. Inter.* **118**, 169–179.
[25] SORNETTE, D. AND SORNETTE, A. (1999). General theory of the modified Gutenberg–Richter law for large seismic moments. *Bull. Seismol. Soc. Amer.* **89**, 1121–1030.
[26] UTSU, T. (1999). Representation and analysis of the earthquake size distribution: a historical review and some new approaches. *Pure Appl. Geophys.* **155**, 509–535.
[27] VERE-JONES, D. (1976). A branching model for crack propagation. *Pure Appl. Geophys.* **114**, 711–725.
[28] VERE-JONES, D. (1977). Statistical theories of crack propagation. *Math. Geol.* **9**, 455–481.
[29] VERE-JONES, D. (1992). Statistical methods for the description and display of earthquake catalogues. In *Statistics in the Environmental and Earth Sciences*, A. T. Walden and P. Guttorp, eds., E. Arnold, London, 220–244.
[30] VERE-JONES, D., ROBINSON, R. AND YANG, W. Z. (2000). Remarks on the accelerated moment release model. Submitted to *Geophys. J. Int.*
[31] WILKS, S. S. (1962). *Mathematical Statistics.* Wiley, New York.
[32] WYSS, M. (1973). Towards a physical understanding of the earthquake frequency distribution. *Geophys. J. Roy. Astronom. Soc.* **31**, 341–359.