

Comparison of Pareto and tapered Pareto distributions for environmental phenomena.

Frederic P. Schoenberg and Rakhee D. Patel

Abstract. The Pareto distribution is often used to describe environmental phenomena such as the sizes of earthquakes or wildfires, or the interevent times or distances between such environmental disturbances. Because it is heavy-tailed, the Pareto distribution, or power-law distribution as it is occasionally called, suggests that a higher frequency of extremely large values occur compared to other, more familiar distributions such as the normal, exponential, or uniform. However, an alternative distribution called the tapered Pareto has been shown in some cases to fit as well or better to data than the Pareto distribution, and the tapered Pareto distribution is not heavy-tailed, suggesting a far lower frequency of extreme events. Even with rather large datasets, it is often quite difficult to distinguish which of these distributions is preferable, as they only differ markedly in the extreme upper tail where few, if any, observations are recorded. This article reviews the evidence and arguments related to these two competing distributions, especially in the context of earthquakes and wildfires.

1. Introduction.

Many environmental disturbances, such as earthquakes, wildfires, and population abundances for various species, are often modeled using the Pareto distribution, which is characterized by extremely heavy tails. These heavy tails are consistent with the relatively frequent occurrence of extremely large events, compared with more conventional distributions like the normal, exponential, uniform, etc., and thus lead to the contention that extremely huge events, or *dragon-kings*, may occur with surprisingly high frequency.

On the other hand, some researchers have proposed that some of these observed phenomena, such as earthquake sizes (Jackson and Kagan 1999, Vere-Jones et al. 2001, Kagan and Schoenberg 2001) or wildfire sizes (Cumming 2001, Schoenberg 2003), may instead be described by a *tapered* Pareto distribution which has much smaller exponential tails but is otherwise similar to the Pareto distribution. Indeed, the tapered Pareto distribution has been shown to offer superior fit not only to the distribution of wildfire sizes and earthquake sizes, but also to earthquake interevent times, earthquake interevent distances, and the areas and perimeters of Voronoi cells of earthquake epicenters, for example (Schoenberg et al. 2008). The aim of this paper is to survey evidence and arguments for the Pareto versus the tapered Pareto distribution, as applied to environmental data.

The issue of discrimination between competing models whose differences manifest mainly in the upper tail is intimately linked to the notion of *dragon-kings*, which refers to the case where the density in the extreme upper tail is substantially higher than what one would predict based on simple extrapolation of the densities of smaller events.

In the case of the tapered Pareto distribution, the hypothesis is that events in the extreme upper tail are substantially *less* frequent than one would predict using the simple Pareto distribution; hence the idea behind the tapered Pareto distribution is essentially the opposite of the *dragon-kings* hypothesis. The main issues underlying the comparison of such distributions and of inference about extreme upper tails given limited observations of these tail events appear to be central to discussion of the *dragon-kings* hypothesis.

2. The Pareto distribution and its variants.

For $x \geq \alpha$, the Pareto distribution has cumulative distribution function (CDF)

$$(1) \quad F(x) = 1 - (\alpha/x)^\beta,$$

and density

$$f(x) = \beta/\alpha (\alpha/x)^{\beta+1},$$

where $\alpha > 0$ is a lower truncation point which is typically known a priori, and β is a parameter to be estimated using the data. The Pareto distribution and its variants have been used in numerous applications. See Chapter 20 of Johnson et al. (1995) for a review of applications, properties and estimates of parameters of the Pareto distribution.

The Pareto distribution has a special self-similarity or fractal property, which lies in the fact that the survival function $[S(x) = 1 - F(x)]$ of the Pareto distribution is linear when plotted on log-log scale. As a result, the shape of the density function (on log-log scale) looks essentially the same at all scales. In other words, the density at any value x , relative to the density at any other value y , only depends on the ratio of x to y and not the values themselves. The slope of the log survival function is the exponent β , which is often called the fractal dimension of the data.

The Pareto distribution arises in theory in certain situations as a result of the fact that the distribution of extreme events conditioned to be larger than some threshold can under certain quite general conditions be characterized by a generalized Pareto distribution (Gnedenko 1943, Pickands 1975).

The *tapered* Pareto distribution has cumulative distribution function, for $x \geq \alpha$,

$$(2) \quad F(x) = 1 - (\alpha/x)^\beta \exp\{(\alpha - x)/\theta\},$$

and density

$$f(x) = (\beta/x + 1/\theta) (\alpha/x)^\beta \exp\{(\alpha - x)/\theta\},$$

where again α is a lower truncation point which is typically known a priori, β is a shape parameter governing the power-law decrease in frequency as x increases, and θ is a parameter governing the location of the exponential taper to zero in the frequency of large events.

There are, of course, numerous alternative distributions that have been proposed for environmental data such as earthquake and wildfire sizes as well as many models that describe tail behavior in general. For environmental phenomena,

perhaps most relevant among these alternatives is the *truncated* Pareto distribution, where the density is proportional to the Pareto density up to some truncation point τ , beyond which the density is identically zero (see e.g. Anderson and Luco 1983, Kagan 1993, Utsu 1999). In finance, important alternatives include the stretched exponential and log-Weibull distributions. For example, Malevergne et al. (2005) test distributions of returns against various parametric models and find that the sample tails decay faster than the Pareto distribution, but slower than the stretched exponential distribution, ultimately concluding that the log-Weibull distribution, which provides a smooth interpolation between the Pareto and stretched exponential, might be more appropriate, though none of the standard distributions seem to provide excellent fit to the entire range of the distribution.

3. Physical arguments.

Inconsistencies between the truncated Pareto distribution and basic physical principles have been well noted. For instance, the fact that observations of environmental phenomena are typically recorded with error does not seem to agree with the notion of a fixed point τ such that events just slightly greater in size than τ are observed with very different frequencies compared to those of size τ (Kagan 1993, Kagan and Schoenberg 2001). Similarly, fundamental geophysical scaling principles in many environmental fields and our understanding of dissipative physical dynamic systems seem to contradict the idea of a sudden truncation point in the density of environmental events. The truncated Pareto distribution will thus not be discussed in what follows, and we will focus on the Pareto and tapered Pareto distributions.

In contrast to the truncated Pareto law, the Pareto distribution is quite consistent with much of the geophysical literature on fractals and scaling, and the simplicity of the distribution and universality of its self-similarity is physically attractive. However, the Pareto distribution is also not without its physical inconsistencies. Indeed, in some circumstances the size of the support of the distribution is limited by physical constraints (a wildfire cannot burn an area larger than the Earth, for instance), so a fitted Pareto distribution that may attribute substantial density to values above these constraints can be troublesome. In addition, the fitted Pareto model for some phenomena is so heavy-tailed that the model implies infinite mean and/or variance, which can stand in contradiction with basic physical principles. In seismology, for instance, the finiteness of deformational energy available for earthquake generation prohibits a seismic moment distribution with infinite mean or variance (Wyss 1973, Knopoff and Kagan 1977, Sornette and Sornette 1999, Bird and Kagan 2004). Similarly, Molchan (2005) showed that a unified scaling law for earthquake interevent times cannot extend perpetually. An excellent survey of these and other physical arguments associated with the fractal, Pareto-based approach to seismology is given in Ben-Zion (2008).

An undesirable feature of the tapered Pareto distribution, from a physical perspective, is that it lacks the exact self-similarity or fractality feature of the pure Pareto law. For tapered Pareto phenomena, one observes near fractality close to the minimum observable size α , and this approximate fractality gradually dissipates as one extends toward the upper right tail of the distribution. Note that unlike the truncated Pareto distribution, the tapered Pareto distribution is not altogether inconsistent with the dragon-kings phenomenon, as the distribution permits extremely large events, though at a reduced frequency compared to the pure Pareto distribution.

4. Similarity of the two distributions.

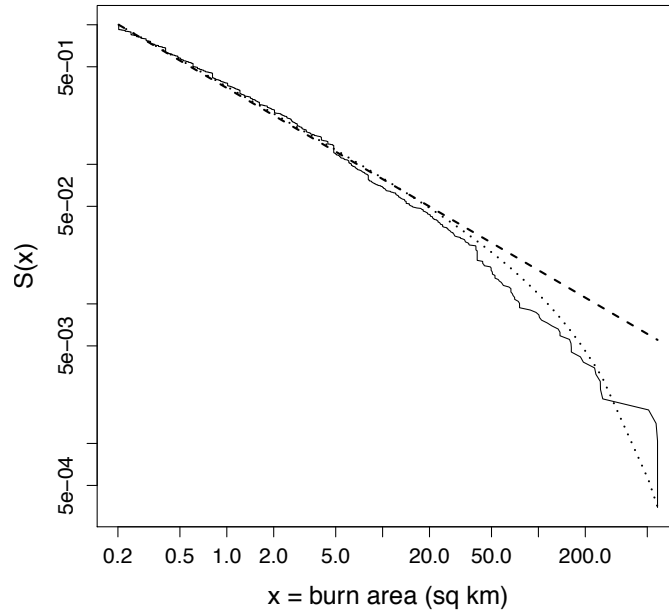


Figure 1. Survival function $S(x) = 1 - F(x)$ of wildfires of size at least $\alpha = 0.20235$ sq. km (50 acres) on CDF protected lands from January 1999 to November 2006. The solid curve is the empirical survival function, the dashed curve is that of the Pareto distribution fit by MLE, with an estimate of β of 0.65286, and the dotted curve is $S(x)$ for the tapered Pareto distribution fit by MLE, with estimates of β and θ of 0.63639 and 201.351, respectively.

The Pareto and tapered Pareto distributions are so similar, other than in the extreme upper tail, and occurrences in the upper tail are often so rare, that the discrimination between the two models is difficult (Molchan et al. 1997, Pisarenko and Sornette 2003, 2004).

For example, consider the survival plot of California wildfire sizes in Figure 1, plotted on a log-log scale along with Pareto and tapered Pareto distributions, fit by

maximum likelihood. The data consist of all wildfires on protected areas of the California Department of Forestry and Fire Protection (CDFFP) between January, 1990 and November, 2006, catalogued by the U.S. Geological Survey, Western Ecological Research Center. For further information about the data, see Keeley (2006). The plot illustrates the difficulty of model selection, as both the Pareto and tapered Pareto distributions appear to provide reasonable fit to the bulk of the data and are nearly indistinguishable except in the upper tail, where the data are extremely sparse.

Even when one has access to thousands of recorded events, the problem of discerning which model is more apt can essentially be considered a small sample problem, since the results of goodness-of-fit tests comparing the Pareto and tapered Pareto distributions can be substantially influenced by just one or two extremely large events (Kagan and Schoenberg 2001).

5. Differences between the distributions.

The implications of the Pareto and tapered Pareto distributions can be substantially different in the upper tail of the distribution, however. For instance, consider the application of the Pareto and tapered Pareto distributions, estimated by maximum likelihood estimation (MLE), to the distribution of scalar seismic moments using the 3765 shallow worldwide Harvard catalog events above magnitude 5.8, from 1/1/1977 to 12/31/1999, analyzed by Kagan and Schoenberg (2001). Using the tapered Pareto distribution, the MLEs of (β, θ) are $(0.67, 10^{21} \text{ Newton m})$. With these parameters, one would expect an event of magnitude at least 10.0 (scalar seismic moment $\geq 10^{24}$) every 10^{436} years according to the tapered Pareto distribution. Using the pure Pareto distribution with the same estimate of $\beta = 0.67$, one would expect an earthquake of magnitude at least 10.0 every 102 years. The fitted tapered Pareto distribution predicts a magnitude 12.0 or higher earthquake more rarely than once every 10^{434300} years, whereas the pure Pareto predicts such an event every 10500 years. The two different distributions obviously have radically distinct implications for policy making, urban planning, and engineering.

6. Estimation.

With either the Pareto or tapered Pareto distribution, one may readily estimate the parameters by maximum likelihood given at least a few hundred observations. One argument against the tapered Pareto distribution is that estimates of the parameter θ are typically dependent on estimates of the lower truncation point α , and may be somewhat unstable in cases where α is unknown and must be estimated using data (Pisarenko and Sornette, 2003). Such cases appear to be quite unusual, however; in most applications, a lower truncation point, e.g. the completeness threshold of a catalog of earthquakes or wildfires, is imposed by researchers and is not estimated using the same data as those used to estimate the parameters in the Pareto or tapered Pareto distribution.

Mixing distributions can be problematic, as a mixture of two distinct Pareto distributions, for instance, is not necessarily Pareto (Pisarenko and Sornette, 2003). This concern is raised in Ben-Zion (2008), who considers the possibility of different distributions for different seismic regimes. Kilpatrick and Ives (2003) show that population abundances for individual species and for collections of species seem consistent with the Pareto distribution, but discuss how interactions between species can affect estimates of β for collections of species.

7. Occam's razor.

Because it is always possible to find a new, more complicated model to fit better to a particular dataset, one may argue that preference should typically be given to simpler and perhaps to older models, unless the improvement of the newer, more complex model is really substantial. This argument is similar to the principle known as Occam's razor (see e.g. Ariew 1976). The distribution function associated with the tapered Pareto law is doubtless a bit more complex than that of the Pareto, and there is an additional parameter, θ , to be estimated. However, as far as age is concerned, the Pareto and tapered Pareto distributions are actually twins, as both appear to have been introduced in the same paper by Vilfredo Pareto in 1897 (see Pareto 1897, pp. 305-306, eqs. 2 and 5).

On the one hand, the simplicity of the Pareto distribution is attractive, and there is a sense in which, if the pure Pareto distribution is correct, the upper tail is bound to deviate a bit from the Pareto simply by chance when one observes a finite sample. On the other hand, statistical tests tend to favor the tapered Pareto by a statistically significant margin for some data sets, offering evidence against the hypothesis that the deviation is due to chance alone or the existence of too small a sample size to discriminate. Some of these tests and their results are discussed in the subsequent Section.

8. Statistical tests.

A commonly used method for assessing local goodness-of-fit is to compare the empirical survival function with the proposed (e.g. Pareto or tapered Pareto) distribution. Confidence bounds can be constructed simply using the binomial distribution, since the empirical cumulative distribution function can be represented by $A(x)/n$, where $A(x)$ is the number of observations less than or equal to x , n is the sample size, and A is binomial[n , $p = F(x)$].

Another standard method for comparing two models is the likelihood ratio test, which is convenient in the case where the Pareto distribution is nested within a proposed alternative such as the tapered Pareto. The likelihood ratio evaluates asymptotically to a Chi-square distribution with degrees of freedom equal to the difference in dimensionality between the hypothesized models -- which is 1, for a

test of the Pareto against the tapered Pareto -- and is asymptotically efficient (Wilks, 1938).

Utsu (1999) compares the Pareto distribution and modified versions of it, including an exponentially tapered version similar to (2), for various Japanese and global datasets of earthquake sizes, using the likelihood-based Akaike Information Criterion (AIC) as a model selection criterion. The results are rather mixed, as in some cases the pure Pareto law provides the best fit using AIC, while in several others, one of the modified models offers slightly superior fit.

Cumming (2001) discusses the distribution of forest fire sizes in Alberta, Canada and uses the Anderson-Darling test as well as quantile plots to show that the Pareto distribution overpredicts large fires, but that the truncated Pareto distribution fits quite well.

Pisarenko and Sornette (2003) derive a statistical test based on a bootstrapping approach involving normalized deviations from the survivor function, with the emphasis upon finding a *change point* where the distribution significantly bends away from the Pareto's log-linear survivor function. The test is used in Pisarenko and Sornette (2003, 2004) to identify statistically significant departures from the Pareto distribution of scalar seismic moments for subduction and mid-ocean range zones, analyzed separately, whereas the survivor function does not show these deviations in the mid-ranges of the distribution. Pisarenko et al. (2004) also propose a statistic they call *TP*, which is based on the first two normalized log-moments of the Pareto CDF, and which converges to zero as the sample size grows to infinity if the data are consistent with the Pareto distribution. The *TP* statistic is cumulative, rather than a measure that can examine local self-similarity at each point. Pisarenko et al. (2004) show using various earthquake datasets that though the Pareto law seems to apply to the middle part of the scalar seismic moment distribution when looking at the survival function, there are in fact significant deviations from the pure Pareto for most of the scaling region as the *TP* statistic shows regular oscillations in the data as the lower threshold is moved.

Song et al. (2005) propose a box-covering method and a renormalizing procedure to assess the extent of a power law relationship and use it to investigate the self-similarity of various complex networks, such as linked webpages from a part of the world wide web, a social network of actors and whether or not they were linked in at least one film, biological networks of protein interactions in *E. coli* and *Homo sapiens* and cellular networks using graphical representations of biochemical pathways for various species of Archaea, Bacteria and Eukarya. Song et al. (2005) find that the relationship between the number of boxes needed to cover the network and the size of the box closely follows the Pareto law.

An alternative is the statistic $g_k(x) = (1/k) [f(x)/f(kx)]^{1/(\beta+1)}$, introduced in Patel and Schoenberg (2011), which specifically examines the local self-similarity of the distribution in question using the ratio of estimated densities at two specified values,

x and kx , for any constant $k > 0$. This statistic is approximately unity for a pure Pareto-distributed random variable and can be estimated using non-parametric density estimation for f , for example, and maximum likelihood estimation for β . The statistic can then be compared to unity to test data against a Pareto null hypothesis, for various values of k that determine how locally to examine the self-similarity. The survival function has an advantage over this local self-similarity test in that accurate density estimation is not necessary for the estimation of the survival function. However, because the survival function is an aggregate measure, it has comparatively low power at discerning local departures from self-similarity. On the other hand, the test statistic $g^*_k(x)$ may be more powerful at detecting lack of fit locally, i.e. for values of x in a particular range. Patel and Schoenberg (2011) use this test to identify statistically significant departures from the pure Pareto distribution for Los Angeles County wildfires between 30 and 50 km² and interevent times of Southern California earthquakes greater than 5 days.

9. Human induced tapering.

Note that it may be possible to reconcile both the Pareto and tapered Pareto distributions by using the Pareto distribution to model an underlying phenomenon and the tapered Pareto to describe our observations of that phenomenon. That is, an environmental phenomenon in question, in its purest form, may follow a pure Pareto law closely, but perhaps human interaction and technology interfere to force a kind of tapering. For instance, while perhaps wildfires might naturally spread to sizes that might be consistent with a Pareto distribution, it is possible that at some point (possibly around 75 km², as speculated in Patel and Schoenberg 2011), intervention by firefighters might become significant, causing the distribution of observed wildfire sizes to be more like the tapered Pareto distribution.

10. Conclusion.

Speculation about the distribution of events in the extreme upper tail, where there are few or no observations, lies perhaps more in the realm of philosophy of science rather than statistics. Proponents of the Pareto distribution may appeal to Occam's razor, contending that when in doubt, the simpler model is preferable. Indeed, the argument that the Pareto model fits the thousands of points in the bulk of the distribution and therefore can be expected to fit to the upper tail as well is similar to a fundamental scientific principle of empiricism: just as one can be confident that the sun will rise tomorrow based on the observation that the sun has risen on each of thousands of days in the past, one may trust the fit of the Pareto distribution to the murky upper tail because of the good fit to the thousands of points in the center of the distribution. The tapered Pareto distribution seems simple enough to warrant serious consideration, however, and when sufficiently many observations of heavy-tailed environmental variables are available, they tend to support the tapered Pareto distribution in the upper tail in many cases. When thousands of observations are present, modeling the distribution using a 2 or 3-parameter model cannot reasonably be considered overfitting. Furthermore, physical arguments often favor

the tapered Pareto distribution; in seismology, for instance, considerations of seismic flux suggest that the moments of the distribution of earthquake size are finite. Similarly, wildfire sizes are ultimately bounded by the size of the Earth, so some sort of tapering may yield a better approximation to reality in the extreme upper tail, just as, incidentally, evidence of the ultimate mortality of the sun (Zeilik and Gregory, 1998) may suggest tapering our confidence in its future existence as well.

Anderson, J.G. and Luco, J.E. (1983). Consequences of slip rate constraints on earthquake occurrence relation. *Bull. Seis. Soc. Amer.* **73**, 471-496.

Ariew, Roger (1976). *Ockham's Razor: A Historical and Philosophical Analysis of Ockham's Principle of Parsimony*. Champaign-Urbana, University of Illinois.

Ben-Zion, Y. (2008). Collective behavior of earthquakes and faults: continuum-discrete transitions, progressive evolutionary changes, and different dynamic regimes. *Reviews of Geophysics* **46**, 1-70.

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. Seis. Soc. Amer.* **94**(6), 2380-2399 (plus electronic supplement).

Cumming, S.G. (2001). A parametric model of the fire-size distribution. *Canadian Journal of Forest Research* **31**, 1297-1303.

Gnedenko, B.V. (1943), Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, **44**, 423-453.

Jackson, D.D., and Kagan, Y.Y. (1999). Testable earthquake forecasts for 1999. *Seismological Research Letters* **70**(4), 393-403.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, 2nd ed., Wiley, NY.

Kagan, Y.Y. (1993). Statistics of characteristic earthquakes. *Bull. Seis. Soc. Amer.* **83**, 7-24.

Kagan, Y.Y., and Schoenberg, F. (2001). Estimation of the upper cutoff parameter for the tapered Pareto distribution. *J. Appl. Prob.* **38A**, Supplement: Festschrift for David Vere-Jones, D. Daley, editor, 158-175.

Keeley, J.E. (2006). South coast bioregion. In *Fire in California's Ecosystems*, N.G. Sugihari, J.W. van Wangtendonk, K.E. Shaffer, J. Fites-Kaufman, and A.E. Thoede, eds. University of California Press, pp. 350-390.

- Kilpatrick, A.M., and Ives, A.R. (2003). Species interactions can explain Taylor's power law for ecological time series. *Nature* **422**, 65-68.
- Knopoff, L. and Kagan, Y.Y. (1977). Analysis of the theory of extremes as applied to earthquake problems. *J. Geophys. Res.* **82**, 5647-5657.
- Malevergne, Y., Pisarenko, V., and Sornette, D. (2005). Empirical distributions of log-returns: between Stretched Exponential and the Pareto Law? *Qualitative Finance* **5**(4), 379-401.
- Molchan, G. (2005). Interevent time distribution in seismicity: a theoretical approach. *Pure and Applied Geophysics* **162**, 1135-1150.
- Molchan, G., Kronrod, T., Panza, G.F. (1997). Multi-scale seismicity model for seismic risk. *Bulletin of the Seismological Society of America* **87** (5), 1220-1229.
- Pareto, V. (1897). *Cours d'Économie Politique*, Tome Second, Lausanne, F. Rouge, quoted by Pareto, V. (1964), *Oeuvres Complètes*, published by de Giovanni Busino, Genève, Droz, vol II.
- Patel, R.D., and Schoenberg, F.P. (2011). A graphical test for local self-similarity in univariate data. *Journal of Applied Statistics*, to appear.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, **3**, 119-131.
- Pisarenko, V.F., and Sornette, D. (2003). Characterization of the frequency of extreme events by the Generalized Pareto Distribution, *Pure and Applied Geophysics* **160** (12), 2343-2364.
- Pisarenko, V.F., and Sornette, D. (2004). Statistical detection and characterization of a deviation from the Gutenberg-Richter distribution above magnitude 8. *Pure and Applied Geophysics* **161** (4), 839-864.
- Pisarenko, V.F., Sornette, D., and Rodkin, M. (2004). A new approach to characterize deviations in the seismic energy distribution from the Gutenberg-Richter law. *Computational Seismology* **35**, 138-159.
- Schoenberg, F.P., Barr, C., and Seo, J. (2008). The distribution of Voronoi cells generated by Southern California earthquake epicenters. *Environmetrics*, **20**(2), 159-171.
- Song, C., Havlin, S., and Makse, H. (2005). Self-similarity of complex networks. *Nature* **433**, 392-395.

- 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.
- Utsu, T. (1999). Representation and analysis of the earthquake size distribution: a historical review and some new approaches. *Pure Appl. Geophys.* **155**, 509-535.
- Vere-Jones, D., Robinson, R., and Yang, W.Z., (2001). Remarks on the accelerated moment release model: problems of model formulation, simulation and estimation, *Geophysical Journal International* **144**, 517-531.
- Wilks, S. S. (1938). The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics* **9**, 60-62.
- Wyss, M. (1973). Towards a physical understanding of the earthquake frequency distribution. *Geophys. J. R. Astr. Soc.* **31**, 341-359.
- Zeilik, M.A.; Gregory, S.A. (1998). *Introductory Astronomy & Astrophysics (4th ed.)*. Saunders College Publishing, Philadelphia, PA.