Eur. Phys. J. B (2016) 89: 16
DOI: 10.1140/epjb/e2015-60452-3

THE EUROPEAN PHYSICAL JOURNAL B

Regular Article

Numerical tools for obtaining power-law representations of heavy-tailed datasets*

Marc L. Mansfield[a]

Bingham Research Center, Utah State University, Vernal, 84078 Utah, USA

Received 5 June 2015 / Received in final form 24 September 2015
Published online 20 January 2016 –

Abstract. Many empirical datasets have highly skewed, non-Gaussian, heavy-tailed distributions, dominated by a relatively small number of data points at the high end of the distribution. Consistent with their role as stable distributions, power laws have frequently been proposed to model such datasets. However there are physical situations that require distributions with finite means. Such situations may call for power laws with high-end cutoffs. Here, I present a maximum-likelihood technique for determining an optimal cut-off power law to represent a given dataset. I also develop a new statistical test of the quality of fit. Results are demonstrated for a number of benchmark datasets. Non-power-law datasets can frequently be represented by power laws, but this is a trivial result unless the dataset spans a broad domain. Nevertheless, I demonstrate that there are non-power-law distributions, including broad log-normal distributions, whose tails can be fit to power laws over many orders of magnitude. Therefore, caution is called for whenever power laws are invoked to represent empirical data.

1 Introduction

Heavy-tailed distributions are skewed and asymmetric, usually with means much larger than medians or modes, and with large standard deviations. Power-law distributions in which a variable x obeys a distribution function proportional to $x^{-\lambda}$ for λ a positive real constant, are prototypes of heavy-tailed distributions [1–5]. They have been used, to cite a few examples, as models of the distribution of personal wealth or income, initial stellar masses, the distribution of biological species among genera, city sizes, the diameters of lunar craters, the size of computer files in internet traffic, citation frequency of scientific papers, waiting-time distributions in models of non-Fickian diffusion, and the occurrence frequency of words in many languages [2,4–13]. It is believed that the prevalence of power-law phenomena is related to the role power laws play as stable distributions [5,6]. The Gaussian distribution is certainly the most common stable distribution in nature: according to the Central Limit Theorem, any process or phenomenon resulting from the combination of a large number of stochastic variables follows a Gaussian distribution as long as each individual variable has finite mean and variance. If we relax the finite-mean-and-variance condition, then generalizations of the Central Limit Theorem call for stable distributions with power-law tails [6]. Under this interpretation, power laws are to heavy-tailed phenomena as Gaussian distributions are to more mundane phenomena.

Elsewhere, I will propose the modeling of heavy-tailed measurement sets of environmental pollutants with power laws [14]. Such measurement sets contain "hot spots" or "super-emitters", i.e., individual measurements that are significantly larger (sometimes by orders of magnitude) than the mean or median. Sampling errors tend to be large and to have negative bias. In part because of this, there are growing concerns that estimates of anthropogenic methane emissions are too low [15–19]. Therefore, a better understanding of sampling error and other statistical properties of heavy-tailed datasets is needed.

Often, power laws have infinite means, which in some contexts is unacceptable. For example, Earth and its inhabitants are incapable of generating an infinite amount of methane. An obvious remedy is to employ power laws with high-end cutoffs.

A central problem in statistics is the "finite-sampling problem", i.e., we wish to infer properties of a distribution function when only a finite sample of the distribution is available. The Central Limit Theorem provides important guidance in this regard, but it can be inconclusive when applied to a dataset that is obviously heavy-tailed. The primary purpose of this paper is to present numerical tools for representing heavy-tailed datasets in terms of power laws, including first, obtaining a power-law representation of an arbitrary dataset, and second, assessing the quality of fit. As I explain below, this power-law representation is only a first step in resolving the finite-sampling problem. It does provide an estimate of the mean of the distribution,

* Supplementary material in the form of one pdf file available from the Journal web page at:
http://dx.doi.org/10.1140/epjb/e2015-60452-3

[a] e-mail: marc.mansfield@usu.edu

but it is unable to place error bars on the result. This problem of error estimation is an active research area and is beyond the scope of this paper.

Clauset et al. [3], and Stumpf and Porter [5] have warned against a tendency to see empirical power laws where none actually exist. A third purpose of this paper is to present many examples of such behavior. Because all differentiable functions are locally linear, and because the analysis presented here is based on a linearizing transformation of the power law, it seems capable almost always of finding a power law if the dataset is restricted to a sufficiently narrow domain. The Stumpf-Porter rule of thumb is that empirical power laws are suspicious unless they extend over at least two orders of magnitude. This is supported by many of the examples considered here. However, I will also present examples of non-power-law datasets (including the tails of broad log-normal distributions) that conform to power laws over two orders or more. This finding provides another reason to be cautious before assigning power-law behavior to empirical datasets.

A number of authors advise determining the maximum-likelihood distribution (MLD) to obtain an optimal power law for a given dataset [2,3,20]. I also follow that approach. However, as shown here, the recommended formula for the MLD is not always accurate for power laws with cutoffs. Therefore, the maximum-likelihood computation has been generalized as explained below.

2 Power-law distributions

A general approach would allow for a power law over a finite interval $x \in [a, b]$ and other behavior outside:

$$P(x) = \begin{cases} g_1(x) & \text{if } x \in (-\infty, a) \\ \frac{\beta}{x^\lambda} & \text{if } x \in [a, b] \\ g_2(x) & \text{if } x \in (b, +\infty) \end{cases} \tag{1}$$

β is a normalization constant. There are at least three good reasons for introducing the lower and upper cutoffs a and b and the functions g_1 and g_2. First, in many empirical examples, the power law holds only for the tail of the distribution [2]. Indeed, the stable distributions of the Generalized Central Limit Theorem generally show power-law behavior only in their tails [6]. Furthermore some examples from the emissions literature include $x < 0$, behavior to which power laws cannot apply. In all such cases, g_1 represents the dependence at smaller x. Second depending on λ, $a \to 0$ or $b \to \infty$ may introduce singularities, and a cutoff at a or b may be required in order for the distribution to be well behaved. Third, cutoffs may be imposed by the physical situation. The value b might represent a physical boundary, limitation, or restriction. For example methane in soil gas can never be greater than complete saturation, 10^6 ppm. We program cutoffs into any model and avoid singularities through appropriate choices of a, b, g_1 and g_2. The main purpose of the g_2 function is to allow for a gradual, as opposed to an abrupt, cutoff at the high end of the data. However, since most datasets do not contain enough information to characterize g_2, I opt for an abrupt high-end cutoff and usually assume g_2 is zero.

I now introduce properties of these functions when g_1 and g_2 are both zero:

$$P(x) = \begin{cases} 0 & \text{if } x \in (-\infty, a) \\ \frac{\beta}{x^\lambda} & \text{if } x \in [a, b] \\ 0 & \text{if } x \in (b, +\infty). \end{cases} \tag{2}$$

Equation (2) is completely determined by the three parameters a, b, and λ, so I use the notation (a, b, λ) to label the power law defined by equation (2). Interesting power-law behavior occurs when a and b are separated by an order of magnitude or more. Therefore, we usually assume that $b \gg a$. The normalization constant and the mean can be written succinctly as

$$\beta = \frac{1}{\Omega(a, b, q)}, \tag{3}$$

$$\mu = \frac{\Omega(a, b, 1+q)}{\Omega(a, b, q)}, \tag{4}$$

where

$$q = 1 - \lambda \tag{5}$$

and where Ω represents the definite integral:

$$\Omega(a, b, z) = \int_a^b x^{z-1} dx = \frac{b^z - a^z}{z}. \tag{6}$$

If $z < 0$, then Ω diverges as $a \to 0$, while if $z > 0$, it diverges as $b \to \infty$. The moments obey

$$\langle x^m \rangle = \int_a^b x^m \beta x^{-\lambda} dx = \frac{\Omega(a, b, m+q)}{\Omega(a, b, q)}. \tag{7}$$

Equation (6) has a removable singularity at $z = 0$. Then, we have

$$\Omega(a, b, 0) = \ln b - \ln a. \tag{8}$$

For numerical work, the following series (with infinite convergence interval) can be used arbitrarily closely to the singularity:

$$\Omega(a, b, z) = \sum_{j=1}^{\infty} \frac{z^{j-1}}{j!} \left[(\ln b)^j - (\ln a)^j \right]. \tag{9}$$

I occasionally consider the complete real domain for λ: $\lambda \in (-\infty, +\infty)$. $\lambda = 0$ corresponds to a distribution that is uniform over $[a, b]$, and $\lambda < 0$ corresponds to a monotone-increasing power law.

When $\lambda \in (-\infty, 1)$, the 0th and all higher moments of the power law diverge in the limit $b \to \infty$, which means that b must be finite to guarantee normalizability of the distribution. When $\lambda \in (1, 2)$, the 0th moment does not diverge, but the first and all higher ones do. Then, (a, ∞, λ) is normalizable, but it has an infinite mean. Some situations might allow this [4] but otherwise when $\lambda < 2$, $b < \infty$ is required to enforce a finite mean (as mentioned above,

power-law behavior emerges for complex processes with internal variables that do not have finite means or variances. Power laws with cutoffs seem reasonable when means or variances are large but finite). Furthermore, when $\lambda < 2$, the mean is very sensitive to the value of b. If $\lambda > 2$ it is permissible to build a physical model that lets b be large and unspecified (for example, in the case of stellar masses, λ is usually cited as about 2.35, and an upper cutoff is generally not given [11]). For similar reasons, because Ω diverges in the limit $a \to 0$ whenever $z < 0$, it is often necessary to maintain $a > 0$.

λ controls the thinning-out rate of the high-end members of the distribution. When $\lambda = 0$, there is no thinning out, the distribution is uniform. $\lambda \in (0, 2)$ produces a distribution that thins out only slowly: equation (4) includes a term in $b^{1+q} = b^{2-\lambda}$, indicating that the mean is sensitive to the high-end members of the distribution. When $\lambda > 2$, the term in $b^{2-\lambda}$ vanishes at large b, indicating that the high-end members occur so rarely that they barely affect the mean. However, the standard deviation carries a term in $b^{3-\lambda}$, indicating that the high-end members influence it unless $\lambda > 3$. As λ increases, $(a,\ b,\ \lambda)$ becomes sharper, and $\lambda \to \infty$ brings us progressively closer to a Dirac-δ function.

The geometric mean of equation (2)

$$\mu_g = e^{\langle \ln x \rangle} \tag{10}$$

obeys

$$\ln \mu_g = \frac{B \ln B - A \ln A}{q\,(B - A)} - \frac{1}{q}, \tag{11}$$

where $A = a^q$, and $B = b^q$. For later reference, observe that the above can also be written:

$$\ln \mu_g = \frac{\partial \ln \beta}{\partial \lambda}. \tag{12}$$

Equation (11) has the following dimensionless form

$$\ln G = \frac{M^q \ln M}{M^q - 1} - \frac{1}{q} \tag{13}$$

where $M = b/a$ and $G = \mu_g/a$.

We let $\Phi(x)$ represent the accumulated distribution of $P(x)$:

$$\Phi\,(x) = \int_{-\infty}^{x} dx' P\,(x') = \beta \int_{a}^{x} x'^{-\lambda} dx' = \frac{\Omega\,(a, x, 1-\lambda)}{\Omega\,(a, b, 1-\lambda)}, \tag{14}$$

$\Phi(x)$ is obviously a linear function of x^q. Therefore, a power-law distribution is a good representation of a dataset over an interval $x \in [a, b]$ if there exists a real exponent q such that Φ is linear in x^q over the interval. Below, we make use of this linearizing transformation to test the quality of fit of a power law to a dataset.

3 The maximum-likelihood distribution

This section describes the maximum-likelihood technique. It provides a straightforward way to obtain a power law that represents an observational dataset [2,3,20]. Let $\{x_1, x_2, \ldots, x_n\}$ represent a dataset. The maximum-likelihood distribution (MLD) is the power law $(a^*\ b^*\ \lambda^*)$ that maximizes the logarithm of the probability:

$$S = \ln \left[\prod_{j=1}^{n} P\,(x_j; a, b, \lambda) \right] = n\,(\ln \beta - \lambda \ln \mu_g)\,, \tag{15}$$

where asterisks denote quantities corresponding to the maximum in S, where

$$\mu_g = \left(\prod x_j \right)^{1/n} \tag{16}$$

is the geometric mean of the dataset, and where β is given by equation (3).

Let $x_{\min}$ and $x_{\max}$ represent the smallest and largest entries, respectively, of the set. Then it must be true that $a \leqslant x_{\min}$ and $b \geqslant x_{\max}$. The derivatives of S with respect to a and b are

$$\frac{\partial S}{\partial a} = n\beta a^{-\lambda} \quad \text{and} \quad \frac{\partial S}{\partial b} = -n\beta b^{-\lambda} \tag{17}$$

and are respectively always positive and negative. Therefore, S is maximized with respect to a or b at these values:

$$a^* = x_{\min} \quad \text{and} \quad b^* = x_{\max}. \tag{18}$$

The condition for λ^* is $\partial S/\partial \lambda = 0$, or

$$\frac{\partial \ln \beta}{\partial \lambda} = \ln \mu_g, \tag{19}$$

which is identical to equation (12). However, equation (12) refers to the geometric mean of a continuous power law, as defined in equation (10), whereas equation (19) refers to the geometric mean of a discrete dataset, as defined in equation (16). This indicates that the MLD is the power law that has the same geometric mean as the dataset. Equation (13) can be solved parametrically for q, which then yields $\lambda^* = 1 - q$.

In the limit $M^q \longrightarrow 0$, equation (13) becomes

$$\lambda^* = 1 + (\ln G)^{-1}\,. \tag{20}$$

Equation (20) is the form published elsewhere [2,3], but its use is not advisable unless M^q is indeed negligible in equation (13). Because these authors usually consider distributions with $\lambda > 2$ and with b effectively infinite, equation (20) is adequate and their work does not require the generalization given by equation (13). On the other hand, the pollution data I examine [14] often displays $\lambda^* \in (1, 2)$ requiring finite b, and equation (13) must be used.

It is important to realize that the MLD usually does not solve the finite sampling problem, and it is a common statistical fallacy to assume that it does. This is true even when the MLD passes quality-of-fit tests such as those discussed below. First, although the MLD maximizes equation (15) over the space of all (a, b, λ)-functions, any other (a, b, λ)-function with comparably large S is also a candidate for the "true" distribution, i.e., the distribution from

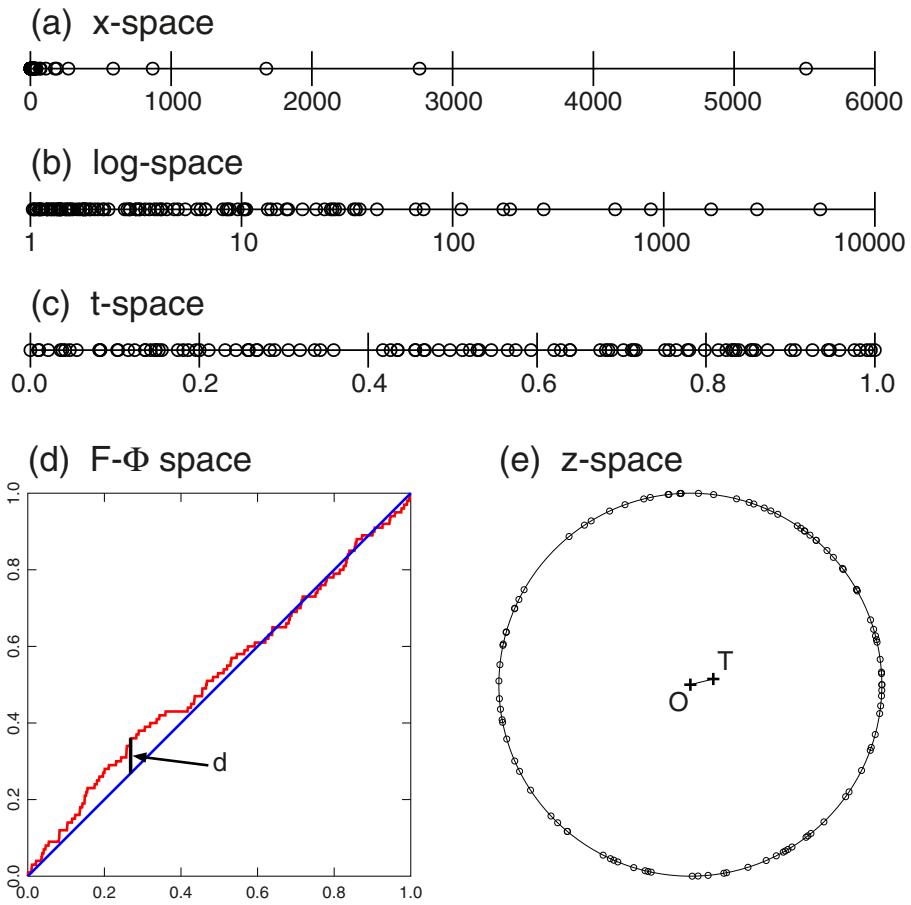

Fig. 1. Schematic representation of the transformations used to test the quality of fit of the maximum-likelihood distribution (MLD).

which the sample was taken. This can be seen, for example, in equation (17). Depending on the values of a, b, and λ, $n\beta b^{-\lambda}$ can be small, meaning that S can be a weak function of b, and that $b^* = x_{\max}$ can be a poor estimate of the "true" b. Second, as demonstrated by many of the examples below, the "true" distribution may not even be an (a, b, λ)-function.

4 Verifying the quality of the MLD

Any arbitrary dataset yields an MLD power law: some (a, b, λ)-function maximizes S. Therefore, it is still necessary to test the quality of fit. The tests considered here are related to the transformation

$$t_j = \Phi(x_j) \tag{21}$$

which maps the dataset $\{x_1, x_2, \ldots, x_n\}$ to a new dataset $\{t_1, t_2, \ldots, t_n\}$ and where Φ is the accumulated distribution belonging to the MLD, defined in equation (14). See Figure 1 for a graphical representation of this transformation. It displays a heavy-tailed dataset of 100 elements first in x-space (Fig. 1a), in which a large fraction of the points congregate near zero, and also on a logarithmic scale (Fig. 1b), in which the dataset still does not appear to be uniform. The set $\{t_1, t_2, \ldots, t_n\}$ obtained from the appropriate MLD is also shown (Fig. 1c). If the MLD is an accurate representation of the dataset, then the set $\{t_j\}$ is uniform over the interval [0, 1]. Let $F(x)$ equal the fraction of points in the dataset with values less than or equal to x, and plot $F(x)$ along the ordinate and $\Phi(x)$ along the abscissa. When the MLD is a good representation of the dataset, we expect $F(x) \cong \Phi(x)$ and the plot should be approximately linear, i.e., we expect the red trace in Figure 1d to lie close to the blue diagonal (this is essentially equivalent to the Q-Q formalism). The F-Φ plot has the following graphical interpretation: as we move horizontally through t-space from 0 to 1, each time we encounter one of the t_j's, we displace vertically by an amount $1/n$. (Another graphical test would be to plot histograms of the dataset in log-log representation. However this can be inconclusive for smaller datasets that give sparse histograms).

F-Φ plots have a discrete structure, rising through steps of $1/n$, and are subject to n-dependent statistical fluctuations. Therefore, an objective test of quality of fit that takes into account the expected fluctuation size is also useful. I have developed the "ring test" to accomplish this. The ring test is a statistical p-test, based on a test statistic to be defined below and its p-value will be denoted p_r. It can be calculated from the dataset $\{x_j\}$ and

from the three power-law parameters (a, b, λ). Indeed, it can be calculated from any arbitrary (a, b, λ) triple. When calculated from the (a^*, b^*, λ^*) triple, I denote it as p_r^*.

The ring test is also based on the uniformity of the power law in t-space. If the dataset $\{t_j = \Phi(x_j)\}$ is uniform in the interval [0, 1], then the 2-vectors

$$\mathbf{z}_j = (\cos 2\pi t_j, \sin 2\pi t_j) \tag{22}$$

are uniform around the circumference of the unit circle, as in Figure 1e, and the center of mass of this collection of 2-vectors is close to the origin:

$$\mathbf{T} = \frac{1}{n}\sum z_j \cong (0, 0)\,. \tag{23}$$

The ring-value is defined as

$$p_r = e^{-nT^2}, \tag{24}$$

where $T = (T_x^2 + T_y^2)^{1/2}$ is the distance of $\mathbf{T}$ from the origin, i.e., the length of the line segment $\overline{\mathbf{OT}}$ in Figure 1e. p_r represents the probability that the $\mathbf{T}$-vector belonging to the dataset in question lies closer to the origin than the $\mathbf{T}$-vectors of all arbitrarily chosen datasets drawn from the power law (a, b, λ). In other words, when p_r is less than some small number α we would say that the assumption that the dataset follows the MLD power law is not statistically significant at significance level α. (In fact, small p_r indicates one of two possibilities, either the dataset does not follow the distribution, or it does, but we have encountered an exceptional event). This property of the ring-value is derived in the Appendix.

The ring test is not limited to power-law fits, but can be applied generally. The Φ-function appearing in equation (21) can be the accumulation of any distribution function to which a dataset has been fit.

The ring test is not a perfect indicator. T can also be small for distributions that have enough symmetry to balance the ring non-uniformly. Therefore, it is advisable always to combine the ring test with a visual inspection of the F-Φ plot.

The ring test is similar to the Kolmogorov-Smirnov test [21] in that it supplies a p-value by which we may test the null hypothesis that an empirical dataset was drawn from some distribution function. Although discussed here within the context of a power-law MLD, both tests can be applied to any distribution function, and both actually test the null hypothesis that the transformation $t = \Phi(x)$ generates a uniform distribution. They differ in that the test statistic for the ring test is the length, T, of the line segment $\overline{\mathbf{OT}}$ in Figure 1e, while that of the Kolmogorov-Smirnov test is the maximum vertical distance, d, between the F-Φ trace and the diagonal, as in Figure 1d. The main advantage of the ring test is that the length T has simpler statistics than the length d, so that the associated p-value is obtained directly from equation (24), while the p-value of the Kolmogorov-Smirnov test must be obtained through look-up tables [21] or by Monte Carlo techniques [3]. The primary disadvantage of the ring test is, as mentioned above, that a non-uniform t-distribution might have enough symmetry to give $T \cong 0$.

5 Benchmark datasets

Thirty-eight benchmark datasets are used to provide examples of the analysis given here. Some of their properties are summarized in Table 1, and more complete characterizations, including procedures for generating each dataset, are given in the supplementary material. Some of these have been chosen because they fit the subjective definition of a heavy-tailed distribution but they do not follow the simple power law formula, thereby testing the capability of the scheme to distinguish such situations. Others have been chosen because they obviously do not have heavy tails, with the goal to see how the analysis performs under those conditions. The selection of datasets for inclusion here was arbitrary. In particular, none were ever discarded because they did not give good power-law fits.

The datasets have been assigned to several different classes. Class A consists of datasets resulting from an analytic distribution function. Class P contains datasets designed to have a power-law tail. For all such datasets, the design value of λ appears in Table 2. Class AP designates the intersection of classes A and P. Class O contains datasets generated by other mathematical rules, and class E contains empirical datasets. Some of the class-O datasets are quasi-power-law sets, in which case a design value of λ is also given in Table 2.

Specific procedures for generating each dataset are given in the supplementary material. For class A or AP, a dataset $\{x_j\}$ of n elements can be generated from an analytic $P(x)$, if we begin with a set of random numbers $\{R_j\}$ distributed uniformly in (0, 1), and then take each $x_j$as the solution of

$$R_j = \Phi(x_j)\,. \tag{25}$$

$R = \Phi(x)$ is solved directly or parametrically for x.

Figure 2 displays the F-Φ plots and Table 1 the p_r^* values obtained for the 38 benchmark datasets. For comparison, the Kolmogorov-Smirnov p-value [21] has also been computed, and is labeled p_{KS}^*. Obviously, not all the datasets pass these tests for conformity to the power law. However, we can still look for power-law behavior in the tail of the distribution. After all, one reason proposed for the occurrence of power laws in nature is that the physical situation calls for a stable distribution, but stable distributions generally obey power laws only in their tails [6]. Furthermore, only the tails of many empirical datasets follow power laws [2]. We partition the dataset into a "body" and a "tail" by designating a parameter D All points $x_j < D$ are assigned to the body, and all points $x_j \geqslant D$ are assigned to the tail. We let n_b and n_t denote, respectively, the number of points in the body and the tail. In the event that all points in the original dataset conform acceptably to a power law, then we assign them all to the tail and define the body to be empty.

The separation into a body and a tail is somewhat arbitrary. There is no one "right" answer, because we are essentially asking at what point the asymptote of a function becomes an adequate representation of the function.

Table 1. Properties of the benchmark datasets, before partitioning into bodies and tails.

Name	n	min	median	mean	max $= b^*$	p_r^*	P_{KS}^*
Class AP							
Exponential Wing	1000	1.25e–04	8.69e–03	0.105	0.984	0.89	0.96
Cauchy	2001	4	1032	4142	4.45 e+05	0.00	0.00
PLi_2.5	3000	1.00	1.59	3.01	314	0.36	0.60
PLi_3	3000	1.00	1.43	2.01	45.9	0.24	0.06
PL_0	3000	2.94	4964	5037	1.00 e+04	0.43	0.50
PL_0.5	3000	1.08	2635	3442	1.00 e+04	0.86	0.94
PL_1.5	3000	1.00	3.92	112	8.93 e+03	0.97	0.96
PL_2	3000	1.00	2.05	13.5	9.15 e+03	0.85	0.93
PL_2.5	3000	1.00	1.58	2.90	235	0.52	0.49
PL_3	3000	1.00	1.42	1.97	93.7	0.84	0.94
Sum	2500	1.00	3.04	70.6	7.48 e+03	0.03	0.05
Class P							
Drift	2048	1.51e+05	4.57e+06	5.33 e+08	6.30 e+11	0.00	0.00
Stable10	2000	1.74	16.9	1277	8.90 e+05	0.00	0.00
Stable100	2000	6.44	73.8	5.89 e+05	1.12 e+09	0.00	0.00
Class A							
Primelike	1000	2.80	4824	4833	9996	0.81	0.80
Exponential	1000	2.41	669	967	6766	0.00	0.00
Gaussian Wing	1000	6.13e–05	2.57e–02	0.114	0.603	0.05	0.20
Parabola A	2000	0.0927	203	254	952	0.00	0.00
Circle A	2000	5.68e–04	0.456	0.476	1.13	0.00	0.00
Ramp	2000	1.79	300	333	975	0.00	0.00
Crossover	1000	1.005	30.3	345	9421	0.00	0.00
Parabola B	2000	0.583	517	558	1.49 e+03	0.00	0.00
Circle B	2000	1.51e–05	0.367	0.480	1.98	0.00	0.00
Cosine	2000	0.0771	271	300	947	0.00	0.00
Log-normal A	2500	1.22e–05	115	3.05 e+06	6.20 e+09	0.00	0.00
Log-normal B	2500	8.26e–04	4.81	69.9	2.83 e+04	0.00	0.00
Log-normal C	2500	1.80e–02	1.10	1.79	85.7	0.00	0.00
Log-normal D	2500	0.374	4.88	6.27	45.2	0.00	0.00
Log-normal E	2500	7.48e–04	9.85e–03	1.28 e–02	1.78e–01	0.00	0.00
Log-normal F	2500	3.27e–06	13.2	6.32 e+04	5.85e+07	0.00	0.00
Log-normal G	2500	8.62e–07	3.16	3.17 e+03	2.18 e+06	0.00	0.00
Log-normal H	2500	8.77e–05	7.70	464	7.20 e+04	0.00	0.00
Class E							
Ouray Ozone (ppb)	699	13.1	52.4	56.5	141.6	0.00	0.00
Gas Production (10^3 scf)	11069	1	1560	3633	1.94 e+05	0.00	0.00
Movies (10^6 US$)	200	286	391	454	1608	0.18	0.54
Class O							
Piecewise Uniform 1	1000	1.00	3.95	101.4	8138	0.25	0.28
Piecewise Uniform 2	1004	1.003	3.97	187	15234	0.00	0.00
Primes	10 000	2	48 615	49 617	1.05e+05	0.70	0.98

Table 2. Properties of the benchmark datasets, after partitioning into bodies and tails.

Name (Design λ)	Characteristics of Body		Characteristics of Tail					
	n_{b}	$x_{\mathrm{b,max}}$	n_{t}	$x_{\mathrm{t,min}} = a^*$	$\lambda*$	p_r^*	p_{KS}^*	b^*/a^*
Class AP								
Exponential Wing (1)	0	N/A	1000	1.25 e–4	1.02	0.89	0.96	8000
Cauchy (2)	1568	2993	433	3002	1.97	0.46	0.60	150
PLi_2.5 (2.5)	0	N/A	3000	1.00	2.49	0.36	0.60	300
PLi_3 (3)	0	N/A	3000	1.00	2.94	0.24	0.06	50
PL_0 (0)	0	N/A	3000	2.94	–0.01	0.43	0.50	3000
PL_0.5 (0.5)	0	N/A	3000	1.08	0.49	0.87	0.94	9000
PL_1.5 (1.5)	0	N/A	3000	1.00	1.49	0.97	0.96	9000
PL_2 (2)	0	N/A	3000	1.00	1.97	0.85	0.93	9000
PL_2.5 (2.5)	0	N/A	3000	1.00	2.49	0.52	0.49	200
PL_3 (3)	0	N/A	3000	1.00	2.99	0.84	0.94	90
Sum (1.5)	1538	4.42	962	4.43	1.54	0.75	0.80	1700
Class P								
Drift (1.5)	968	4.00e+06	1080	4.01e+06	1.50	0.75	0.29	1.6e+05
Stable10 (1.6)	1403	39.3	597	39.3	1.58	0.50	0.61	20 000
Stable100 (1.6)	1137	98.9	863	99.8	1.65	0.58	0.67	1.1e+07
Class A								
Primelike	0	N/A	1000	2.803	0.08	0.81	0.80	4000
Exponential	877	2014	123	2025	3.25	0.34	0.60	3
Gaussian Wing	524	0.0299	476	0.0301	0.72	0.26	0.66	20
Parabola A	1802	557	198	559	5.25	0.28	0.70	1.7
Circle A	1640	0.800	360	0.801	4.17	0.34	0.70	1.4
Ramp	1852	725	148	729	7.14	0.26	0.53	1.3
Crossover	801	319	199	334	1.64	0.71	0.96	30
Parabola B	1791	1089	209	1.09 e+03	7.51	0.31	0.44	1.4
Circle B	1646	0.879	354	0.880	3.33	0.25	0.21	2
Cosine	1789	587	211	588	5.70	0.30	0.64	1.6
Log-normal A	2225	3.53 e+04	275	3.61 e+04	1.47	0.48	0.48	1.7 e+05
Log-normal B	2259	113	241	114	1.96	0.57	0.39	250
Log-normal C	2255	4.17	245	4.21	3.15	0.34	0.36	20
Log-normal D	2209	11.3	291	11.3	3.62	0.54	0.32	4
Log-normal E	2340	2.92 e–02	160	2.95 e–02	4.33	0.39	0.73	6
Log-normal F	2189	2.87 e+03	311	2.97 e+03	1.43	0.52	0.75	20 000
Log-normal G	2278	638	222	679	1.52	0.57	0.89	3000
Log-normal H	1968	83.0	532	83.9	1.52	0.61	0.83	900
Class E								
Ouray Ozone	572	68.9	127	69.1	2.47	0.20	0.20	2
Gas Production	9868	8112	1201	8114	2.62	0.27	0.03	20
Movies	0	N/A	200	286	3.33	0.18	0.54	6
Class O								
Piecewise Uniform 1 (1.5)	0	N/A	1000	1.000	1.48	0.25	0.28	8000
Piecewise Uniform 2 (1.5)	772	19.76	232	20.9	1.44	0.65	0.09	700
Primes	0	N/A	10000	2	0.11	0.70	0.98	50 000

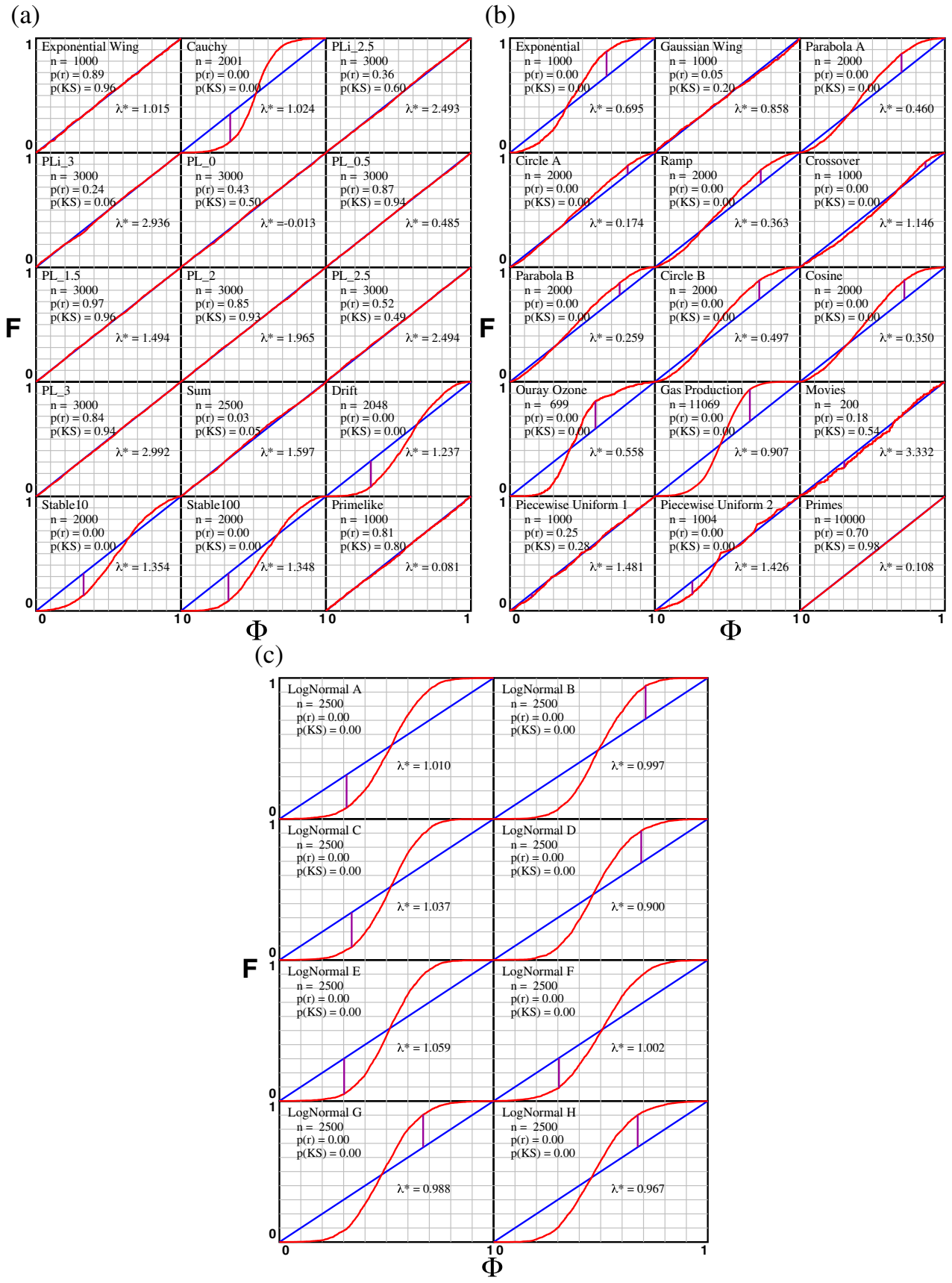

Fig. 2. F-Φ plots for the indicated datasets, prior to body-tail partitioning. Non-linearity of the red trace, $p(r)$ near 0, or $p(\mathrm{KS})$ near 0, all indicate that the power law MLD is a poor fit to the dataset. The vertical tie line between each red and blue trace denotes the maximum vertical distance between the two. Its length is the test statistic for the Kolmogorov-Smirnov p-test.

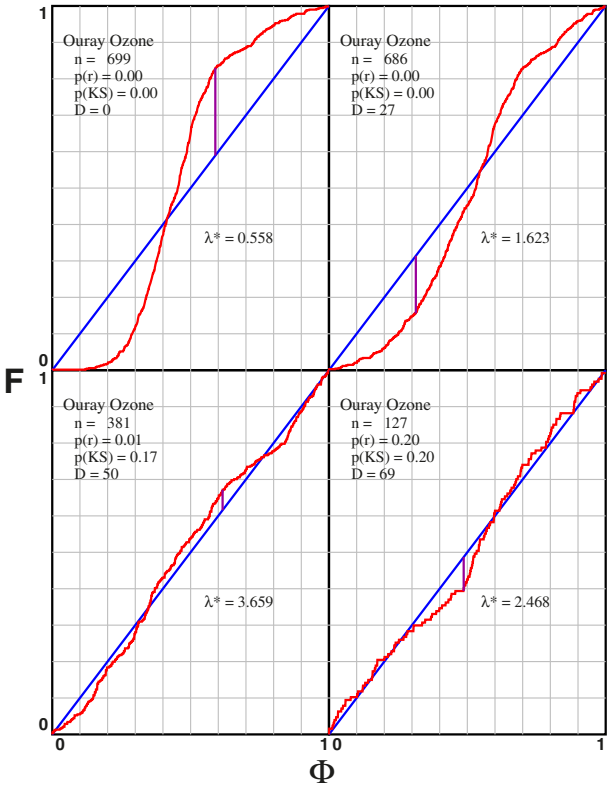

Fig. 3. F-Φ plots for the Ouray Ozone dataset, showing the effect of different body-tail partitions. When $D = 0$, the full dataset has been included in the analysis; when $D = 27$, only data points at or greater than 27 ppb are included, etc.

I have used the following approach. Each x_j in the dataset is considered as a candidate for the value D, meaning that all points greater than or equal to any one x_j are assigned to a tentative tail. I then accepted one of these that had a small D (and therefore a large number of elements) while giving a p_r^* greater than about 0.1 or 0.2, after confirming that it also produces a linear F-Φ plot. Figure 3 displays a few examples of how p_r or linearity of the F-Φ plot develop for the Ouray Ozone dataset. As D progresses from 0 to 27 to 50 to 69 ppb, linearity of the F-Φ plot improves, and the p_r^*-value increases to 0.2.

6 Results

The analysis discussed here was applied to all 38 benchmark datasets. Thirteen of the datasets, identified in Table 1, yield linear F-Φ plots and large p_r^*-values from the outset, and do not require separation into a body and a tail. For the remaining 25, it has always proved possible to assign the high end of the distribution to a tail that follows a power law. This occurs even for datasets from distributions having little or no resemblance to a power law. The F-Φ plots of the tails of these 25 datasets are shown in Figure 4 and the final results for all 38 datasets are summarized in Table 2. The p-values obtained by the ring test and by the Kolmogorov-Smirnov test generally agree, for example, at the 0.05 significance level (the odds are 1 in 20 that they do not agree, so the few instances when they do not are not statistically significant).

For all 38 datasets, we have obtained $p_r^* \gtrsim 0.2$ after partitioning the dataset into a body and a tail. This suggests that many arbitrary datasets will obey a power law, at least in their tails. Some readers may think I have employed an overabundance of examples, but these examples all serve to demonstrate how easily one may find power-law behavior, even when none has been programmed in from the outset.

The ubiquity of power-law behavior is not necessarily remarkable. Any differentiable function is locally linear. Because our analysis is based on a linearizing transformation, it is guaranteed to succeed for any differentiable distribution function if applied over a sufficiently narrow domain. Truly remarkable behavior only occurs when the dataset spans a broad data range. This agrees with Stumpf and Porter [5] who have cautioned against inferring true power law behavior unless, in our notation, $b^*/a^* \gtrsim 100$. In other words, it is usually possible, given an arbitrary dataset, to adjust D in order to see power-law behavior in the tail. This can be taken as evidence of true power-law behavior only if the tail is broad. However, as several examples below indicate, even $b^*/a^* \gtrsim 100$ may be misleading.

For all of the datasets of class AP and P, $\lambda^* \cong \lambda$ (design), and b^*/a^* is always large, indicating that the basic technique is able to correctly recognize true power law behavior and to select the correct λ.

None of the datasets of Classes A, E, or O were designed to show true power-law behavior. Therefore, those that give large b^*/a^* values demand special comment:

(A) The two Piecewise Uniform datasets, with b^*/a^* around 8000 and 700, respectively, were designed to show quasi-power-law behavior, and for them the analysis also yields $\lambda^* \cong \lambda$ (design).

(B) Primelike and Primes exhibit b^*/a^* of about 4000 and 50 000, respectively. They were generated, exactly and asymptotically respectively, from distributions $\propto (\ln x)^{-1}$. Since $\ln x$ is the integral of x^{-1}, perhaps it is not too surprising that these datasets yield power-law behavior with an effective λ near 0. Nevertheless, they are not true power laws, and provide the first of several examples of broad datasets that can be mistaken for power laws.

(C) Crossover has been designed to be in a slow crossover from $\lambda = 1$ to $\lambda = 2$. However, it is truncated before reaching the $\lambda = 2$ asymptote, and so does not have a true power-law tail. Nevertheless, it exhibits $b^*/a^* \approx 30$ (the Sum database also incorporates a crossover from $\lambda = 2$ to $\lambda = 1.5$, but its tail extends well beyond the crossover, so it has been assigned to class AP).

(D) Crossover, Gaussian Wing, and a majority of the log-normal datasets have b^*/a^* values greater than one or two orders of magnitude, in spite of the fact that none of these, including none of the log-normal datasets, were selected to have power-law tails. In these cases, the fit seems to come about because the distribution

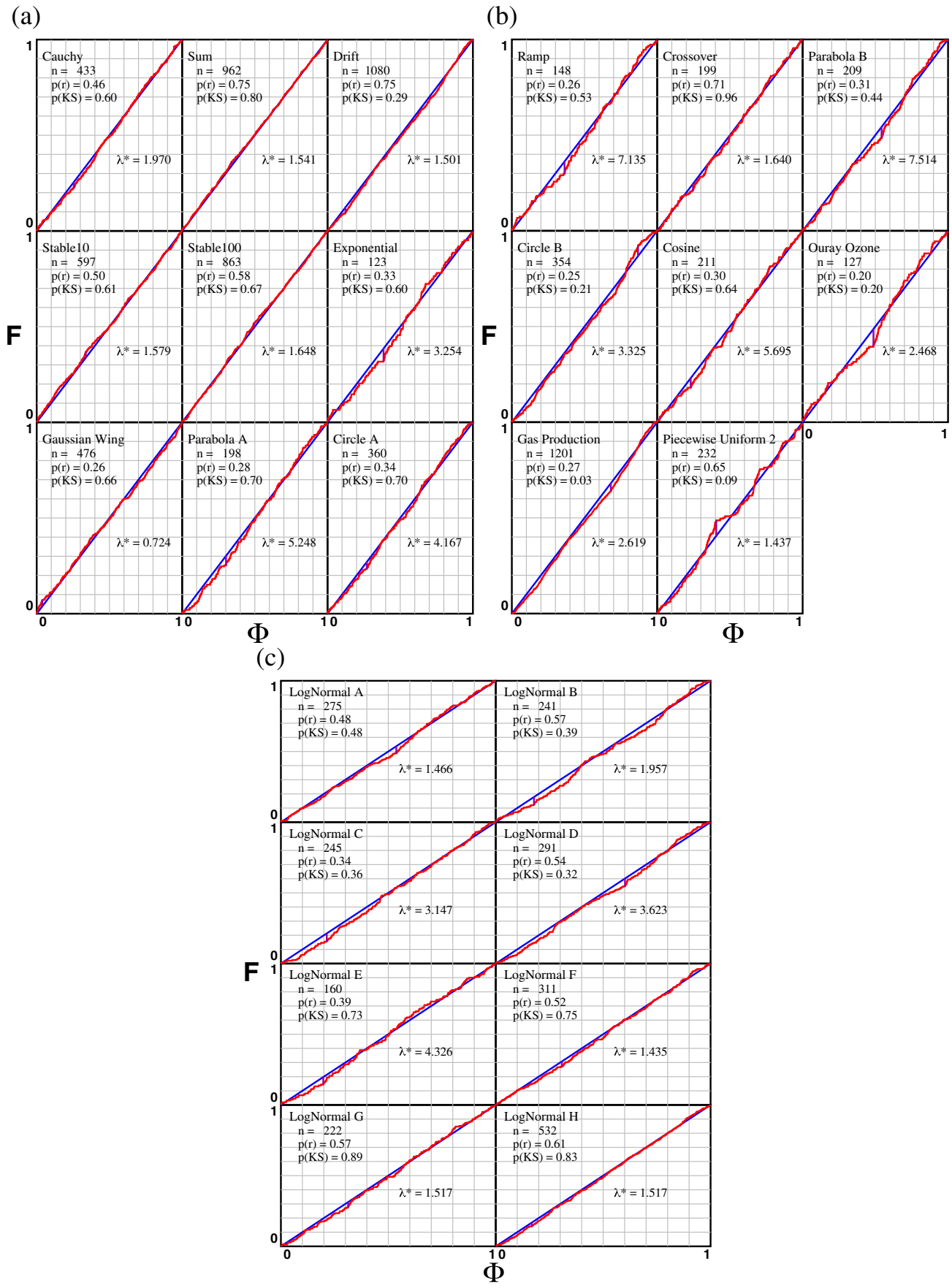

Fig. 4. F-Φ plots for the indicated datasets, after body-tail partitioning.

Table 3. Percent error committed when one estimates the mean of the tail of the indicated log-normal dataset from its power-law MLD.

Log-normal dataset	λ^*	% Error
F	1.435	+96%
A	1.466	–51%
H	1.517	–28%
G	1.517	+34%
B	1.957	+1%
C	3.147	–4%
D	3.623	+1%
E	4.326	+1%

in question, when truncated to the interval (a^*, b^*), is approximated by a power law. Then, the inherent statistical noise in a finite dataset is able to mask the differences between the two distributions. This finding indicates that it is possible to assign power-law behavior to empirical datasets even when the physical process follows some other distribution, and even over several orders of magnitude.

(E) Gas Production, with $b^*/a^* \approx 20$, may be an example of a power-law phenomenon.

Many of the sample datasets were generated from distribution functions that have no high-end cutoffs, but such datasets, being finite, always have a largest member, and the power-law MLD cutoff is, by definition, $b^* = x_{\max}$. When the analysis assigns some power law to a non-power-law dataset, the value of b^* has no other significance. When applied to a power-law dataset, correlations between b^* and b depend on the values of n and λ. In particular, b^* is close to b when $\lambda < 1$ and $n \gg 1$. On the other hand, as λ increase beyond 1, the likelihood that $x_{\max}$ is close to b decreases. In such cases, b^* again has no other significance than being the largest member of the set. For example, for the three datasets PL_0, PL_0.5, and Exponential Wing, with $\lambda = 0$, 0.5, and 1, respectively, b and b^* agree to within 2%. On the other hand, the four datasets PL_2.5, PL_3, PLi_2.5, and PLi_3 are drawn from the power laws (1, 10^4, 2.5), (1, 10^4, 3), (1, ∞, 2.5), and (1, ∞, 3), respectively, and their respective $b^* = x_{\max}$ values are 235, 93.7, 314, and 45.9.

The fact that power laws seem capable of approximating broad log-normal datasets may have serious repercussions, especially when the analysis arrives at $\lambda^* < 2$, which is the domain for which the mean of the power law is a strong function of the value of the cutoff. For example Table 3 displays the relative error when we estimate the mean of each of the log-normal tails from the MLD power law. Agreement is reasonable only when λ^* is larger than about 2.

7 Discussion and conclusions

In this paper, I have presented techniques for obtaining a power-law fit to an empirical dataset, as well as techniques for evaluating the quality of the fit. The analysis includes power laws with high-end cutoffs, since these are often necessary to model finite processes.

Above, I alluded to the "finite-sampling" problem, in which we wish to characterize an unknown distribution from only a finite sample. This can be a significant challenge for heavy-tailed datasets. A common fallacy is to assume that the MLD, or some other fitted distribution, is a close approximation to the true distribution function, and therefore assume that it effectively solves the finite sampling problem. There are two problems with such an assumption. First, in addition to the MLD power law, there are many other power laws lying close by in parameter space with likelihoods almost as large as the MLD. When the contributions from these are included, we may still have large measurement uncertainties, especially at small n. Second, depending on the value of λ the mean of a power law (Eq. (4)), may depend strongly on b and b^* can be a poor approximation to the true value of b. Table 3 provides examples of this behavior. In such cases, there is little or no information in the dataset itself that would permit us to estimate the true mean. Therefore, I regard the finite-sampling problem as still unresolved for heavy-tailed datasets. Determination of the MLD is an important first step, but an ability to place confidence limits on the mean of the MLD is still lacking.

Based on the results reported here, it is not uncommon to be able to fit a power law to the high-end members of a heavy-tailed dataset, but this is no guarantee that the underlying distribution is obeying a power law, especially if the fit encompasses a relatively small range of data. Stumpf and Porter [5] enjoin us to be suspicious unless the range spans two or more orders of magnitude. However, even this is not a complete guarantee, since I have been able to fit the tails of log-normal and other datasets to power laws over several orders of magnitude. Apparently this occurs because one function approximates the other closely enough that any differences disappear below the level of statistical noise.

Therefore, in the absence of a convincing mathematical model of a phenomenon, it seems difficult to confirm true power-law behavior. Nevertheless, I propose the pragmatic point of view that power-law analysis is useful, as long as the fit is good. The techniques presented here provide a straightforward method for analyzing heavy-tailed distributions in terms of power laws. Computer codes will be provided upon request.

This work was supported by grants from the Utah Science Technology and Research (USTAR) Initiative and by the US Bureau of Land Management.

Appendix: Statistics of the ring test

Here I demonstrate that p_r as defined in equation (24) represents the probability that the T-vector belonging to the dataset in question lies nearer to the origin than those belonging to datasets generated by the test distribution.

Suppose that we have a set of polar angles $\{\theta_j\}$ where each is assumed to be distributed uniformly between 0 and 2π. They define a set of points $\mathbf{z}_j = (\cos\theta_j, \sin\theta_j)$ distributed on the unit circle. The vector

$$\mathbf{T} = (T_x, T_y) = \frac{1}{n}\left(\sum \cos\theta_j, \sum \sin\theta_j\right) \tag{A.1}$$

is the center-of-mass of the n points. We begin by deriving the distribution function for $\mathbf{T}$. It can be written formally as

$$\begin{aligned} P(\mathbf{T}) = \frac{1}{(2\pi)^n}\int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \ldots \\ \times \int_0^{2\pi} d\theta_n \delta\left(T_x - \frac{1}{n}\sum \cos\theta_j\right) \\ \times \delta\left(T_y - \frac{1}{n}\sum \sin\theta_j\right). \end{aligned} \tag{A.2}$$

The integrals over θ_j sample uniformly all possible positions of the points on the unit circle, while the Dirac-δ functions select out only those combinations of θ_j that sum up to one specific value of the vector $\mathbf{T}$. Each factor $(2\pi)^{-1}$ normalizes one of the θ-integrals. By introducing the Fourier representation of the δ-functions,

$$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} dk e^{ikx} \tag{A.3}$$

we can write

$$\begin{aligned} P(\mathbf{T}) = \left(\frac{1}{2\pi}\right)^{n+2} \int_{-\infty}^{+\infty} dk_x \\ \times \int_{-\infty}^{+\infty} dk_y \left[\int_0^{2\pi} d\theta \exp\left(\frac{-i}{n}\mathbf{k}\cdot\mathbf{r}\right)\right]^n e^{i\mathbf{k}\cdot\mathbf{T}}, \end{aligned} \tag{A.4}$$

where the following vector notation is employed:

$$\mathbf{k} = (k_x, k_y), \quad \mathbf{r} = (\cos\theta, \sin\theta). \tag{A.5}$$

The θ-integral sums up phase-factors over the circumference of the unit circle and yields the result

$$\int_0^{2\pi} d\theta \exp\left(\frac{-i}{n}\mathbf{k}\cdot\mathbf{r}\right) = 2\pi I_0\left(\frac{ik}{n}\right), \tag{A.6}$$

where $k = (k_x^2 + k_y^2)^{1/2}$ and where I is a modified Bessel function [22]. When its argument is purely imaginary, I_0 achieves a maximum at $k = 0$, and decreases monotonically as k grows. This function is then raised to a power n, which magnifies the maximum at $k = 0$ relative to all other values of k. The result is that for n larger than about 10, only the small-k behavior is relevant, and I_0^n can be approximated from its series expansion to low order:

$$I_0(iz) \cong 1 - \frac{z^2}{4} \tag{A.7}$$

with the result that

$$\begin{aligned} \left[\int_0^{2\pi} d\theta \exp\left(\frac{-i}{n}\mathbf{k}\cdot\mathbf{r}\right)\right]^n &\cong (2\pi)^n \left(1 - \frac{k^2}{4n^2}\right)^n \\ &\cong (2\pi)^n \exp\left(\frac{-k^2}{4n}\right) \end{aligned} \tag{A.8}$$

and

$$P(\mathbf{T}) = \frac{1}{(2\pi)^2}\int_{-\infty}^{+\infty} dk_x \int_{-\infty}^{+\infty} dk_y \exp\left(-\frac{k^2}{4n} + i\mathbf{k}\cdot\mathbf{T}\right) \tag{A.9}$$

or

$$P(\mathbf{T}) = \frac{n}{\pi} e^{-nT^2}, \tag{A.10}$$

where $T = (T_x^2 + T_y^2)^{1/2}$ is the distance of $\mathbf{T}$ from the origin. The probability that $\mathbf{T}$ lies a distance greater than T_0 from the origin is

$$\int_{T0}^{\infty} (2\pi T dT)\frac{n}{\pi} e^{-nT^2} = \exp\left(-nT_0^2\right). \tag{A.11}$$

This expression therefore serves as a definition of the ring-value (Eq. (24)).

Note added in proof. The disadvantage of the ring test mentioned in Section 4, relating to non-uniformities of high symmetry, can be remedied by mapping not to only one, but to several circumferences of the unit circle. A complete report will be published elsewhere.

References

1. B.B. Mandelbrot, *The Fractal Geometry of Nature* (W.H. Freeman, New York, 1983)
2. M.E.J. Newman, Contemp. Phys. **46**, 323 (2005)
3. A. Clauset, C.R. Shalizi, M.E.J. Newman, SIAM Rev. **51**, 661 (2009)
4. E. Barkai, Y. Garini, R. Metzler, Phys. Today **65**, 29 (2012)
5. M.P.H. Stumpf, M.A. Porter, Science **335**, 665 (2012)
6. V.M. Zolotarev, in *One-Dimensional Stable Distributions*, Translations of Mathematical Monographs (American Mathematical Society, Providence, 1986), Vol. 65
7. M. Levy, S. Solomon, Physica A **242**, 90 (1997)
8. S. Redner, Eur. Phys. J. B **4**, 131 (1998)
9. A. Dragulescu, V.M. Yakovenko, Eur. Phys. J. B **17**, 723 (2000)
10. A. Dragulescu, V.M. Yakovenko, Physica A **299**, 213 (2001)
11. P. Kroupa, Science **295**, 82 (2002)
12. P. Lukasiewicz, A. Orlowski, Physica A **344**, 146 (2004)
13. S. Sinha, Physica A **359**, 555 (2006)
14. M.L. Mansfield, to be submitted
15. R.A. Alvarez, S.W. Pacala, J.J. Winebrake, W.L. Chameides, S.P. Hamburg, Proc. Natl. Acad. Sci. USA **109**, 6435 (2012)

16. A.S. Katzenstein, L.A. Doezema, I.J. Simpson, D.R. Blake, F.S. Rowland, Proc. Natl. Acad. Sci. USA **100**, 11975 (2013)
17. A.R. Brandt, G.A. Heath, E.A. Kort, F. O'Sullivan, G. Pétron, S.M. Jordaan, P. Tans, J. Wilcox, A.M. Gopstein, D. Arent, S. Wofsy, N.J. Brown, R. Bradley, G.D. Stucky, D. Eardley, R. Harriss, Science **343**, 733 (2014)
18. S.M. Miller, S.C. Wofsy, A.M. Michalak, E.A. Kort, A.E. Andrews, S.C. Biraud, E.J. Dlugokencky, J. Eluszkiewicz, M.L. Fischer, G. Janssens-Maenhout, B.R. Miller, J.B. Miller, S.A. Montzka, T. Nehrkorn, C. Sweeney, Proc. Natl. Acad. Sci. USA **110**, 20018 (2013)
19. A. Karion, C. Sweeney, G. Pétron, G. Frost, M. Hardesty, J. Kofler, B.R. Miller, T. Newberger, S. Wolter, R. Banta, A. Brewer, E. Dlugokencky, P. Lang, S.A. Montzka, R. Schnell, P. Tans, M. Trainer, R. Zamora, S. Conley, Geophys. Res. Lett. **40**, 4393 (2013)
20. M.L. Goldstein, S.A. Morris, G.G. Yen, Eur. Phys. J. B **41**, 255 (2004)
21. I.M. Charkravarti, R.G. Laha, J. Roy, in *Handbook of Methods of Applied Statistics* (Wiley, New York, 1967), Vol. 1
22. M. Abramowitz, I.A. Stegen, *Handbook of Mathematical Functions* (Dover, New York, 1965)