Parameter estimation for power-law distributions by maximum likelihood methods

H. Bauke[a]

Rudolf Peierls Centre for Theoretical Physics, University of Oxford, 1 Keble Road, Oxford, OX1 3NP, UK

Abstract. Distributions following a power-law are an ubiquitous phenomenon. Methods for determining the exponent of a power-law tail by graphical means are often used in practice but are intrinsically unreliable. Maximum likelihood estimators for the exponent are a mathematically sound alternative to graphical methods.

PACS. 02.50.Tt Inference methods – 89.75.-k Complex systems

1 Introduction

The distribution of a discrete random variable is referred to as a distribution with a power-law tail if it falls as

$$p(k) \sim k^{-\gamma} \tag{1}$$

for $k \in \mathbb{N}$ and $k \geq k_{\min}$. Power-laws are ubiquitous distributions that can be found in many systems from different disciplines, see [1,2] and references therein for some examples.

Experimental data of quantities that follow a power-law are usually very noisy; and therefore obtaining *reliable* estimates for the exponent γ is notoriously difficult. Estimates that are based on graphical methods are certainly used most often in practice. But simple graphical methods are intrinsically unreliable and not able to establish a reliable estimate of the exponent γ.

For that reason, the authors of [3] introduced an alternative approach based on a maximum likelihood estimator for the exponent γ. Unfortunately the authors concentrate on a rather idealized type of power-law distributions, namely

$$p_1(k; \gamma) = \frac{k^{-\gamma}}{\zeta(\gamma, 1)} \tag{2}$$

with $k \in \mathbb{N}$, where the normalization constant $\zeta(\gamma, 1)$ is given by the Hurwitz-ζ-function which is defined for $\gamma > 1$ and $a > 0$ by

$$\zeta(\gamma, a) = \sum_{i=0}^{\infty} \frac{1}{(i+a)^{\gamma}}. \tag{3}$$

The distribution (2) is characterized by one parameter only, and therefore all properties of this distribution (e.g.

---

[a] e-mail: `heiko.bauke@physics.ox.ac.uk`

its mean) are determined solely by the exponent γ. In many applications the power-law (2) is too restrictive.

If one states that a quantity follows a power-law, then this means usually that the *tail* ($k \geq k_{\min}$) of the distribution $p(k)$ falls proportionally to $k^{-\gamma}$. Probabilities $p(k)$ for $k < k_{\min}$ may differ from the power-law and admit the possibility to tune the mean or other characteristics independently of γ. In some situations probabilities $p(k)$ may differ from a power-law for $k \geq k_{\max}$ as well, e.g. the distribution may have an exponential cut-off.

Therefore, I will generalize the maximum likelihood approach introduced in [3] to distributions that follow a power-law within a certain range $k_{\min} \leq k < k_{\max}$ but differ from a power-law outside this range in an arbitrary way. Furthermore, I will give some statements about the large sample properties of the estimate of the power-law exponent and present a numerical procedure to identify the power-law regime of the distribution $p(k)$. But first, let us see what is wrong with popular graphical methods.

2 Trouble with graphical methods

All graphical methods for estimating power-law exponents are based on a linear least squares fit of some empirical data points $(x_1, y_1), (x_2, y_2), \ldots, (x_M, y_M)$ to the function

$$y(x) = a_0 + a_1 x. \tag{4}$$

The linear least squares fit minimizes the residual

$$\Delta = \sum_{i=1}^{M} (y_i - a_0 - a_1 x_i)^2. \tag{5}$$

Estimates $\hat{a}_0$ and $\hat{a}_1$ of the parameters a_0 and a_1 are given by [4]

$$\hat{a}_0 = \frac{\left(\sum_{i=1}^M y_i\right)\left(\sum_{i=1}^M x_i^2\right) - \left(\sum_{i=1}^M x_i\right)\left(\sum_{i=1}^M y_i x_i\right)}{M\left(\sum_{i=1}^M x_i^2\right) - \left(\sum_{i=1}^M x_i\right)^2}$$
(6)

and

$$\hat{a}_1 = \frac{M\left(\sum_{i=1}^M y_i x_i\right) - \left(\sum_{i=1}^M x_i\right)\left(\sum_{i=1}^M y_i\right)}{M\left(\sum_{i=1}^M x_i^2\right) - \left(\sum_{i=1}^M x_i\right)^2}.$$
(7)

The ansatz for the residual (5) and derivation of (6) and (7) are based on several assumptions regarding the data points (x_i, y_i). It is assumed that there are no statistical uncertainties in x_i, but y_i may contain some statistical error. The errors in different y_i are independent identically distributed random variables with mean zero. In particular the standard deviation of the error is independent of x_i. For various graphical methods for the estimation of the exponent of a power-law distribution these conditions are not met, leading to the poor performance of these methods.

To illustrate the failure of graphical methods by a computer experiment $N = 10\,000$ random numbers m_i had been drawn from distribution (2) with $\gamma = 2.5$ and an estimate $\hat{\gamma}$ for the exponent γ was determined by various graphical methods. The estimator is a random variable and its distribution depends on the method that has been used to obtain the estimate. Important measures of the quality of an estimator are its mean and its standard deviation. If the mean of the estimator equals the true exponent γ then the estimator is unbiased and estimators with a distribution that is concentrated around γ are desirable. For each graphical method a histogram of the distribution of the estimator was calculated to rate the quality of the estimator by repeating the numerical experiment 500 times.

The most straight forward (and most unreliable) graphical approach is based on a plot of the empirical probability distribution $\hat{p}(k)$ on a double-logarithmic scale. Introducing the indicator function $\mathbb{I}[\cdot]$, which is one if the statement in the brackets is true and else zero, the empirical probability distribution is given by

$$\hat{p}(k) = \frac{1}{N}\sum_{i=1}^N \mathbb{I}[m_i = k].$$
(8)

An estimate $\hat{\gamma}$ for the power-law exponent γ is established by a least squares fit to

$$(x_i, y_i) = (\ln k, \ln \hat{p}(k)) \quad \text{for all } k \in \mathbb{N} \text{ with } \hat{p}(k) > 0,$$
(9)

$\hat{\gamma}$ equals the estimate (7) for the slope, see Figure 1a. Because the lack of data points in the tail of the empirical distribution this procedure underestimates systematically the exponent γ, see Table 1.

There are two ways to deal with the sparseness in the tail of the empirical distribution, logarithmic binning and considering the empirical cumulative distribution $\hat{P}(k)$ instead of $\hat{p}(k)$. The cumulative probability distribution of (2) is defined by

$$P(k) = \sum_{i=k}^\infty \frac{i^{-\gamma}}{\zeta(\gamma, 1)}.$$
(10)

If $p(k)$ has a power-law tail with exponent γ then $P(k)$ follows approximately a power-law with exponent $\gamma - 1$ because for $k \gg 1$ the distribution $P(k)$ can be approximated by

$$P(k) \approx \int_k^\infty \frac{i^{-\gamma}}{\zeta(\gamma, 1)}\, \mathrm{d}i = \frac{k^{1-\gamma}}{(\gamma - 1)\zeta(\gamma, 1)}.$$
(11)

The empirical cumulative probability distribution is given by

$$\hat{P}(k) = \frac{1}{N}\sum_{i=1}^N \mathbb{I}[m_i \geq k].$$
(12)

It is less sensitive to the noise in the tail of the distribution and therefore a fit of

$$(x_i, y_i) = (\ln k, \ln \hat{P}(k)) \quad \text{for all } k \in \mathbb{N} \text{ with } \hat{P}(k) > 0$$
(13)

to a straight line gives much better estimates for the exponent, see Figure 1b. But there is still a small bias to too small values and the distribution of this estimate is rather broad, see Table 1.

Logarithmic binning reduces the noise in the tail of the empirical distributions $\hat{p}(k)$ and $\hat{P}(k)$ by merging data points into groups. By introducing the logarithmically scaled boundaries

$$b_i = \text{round}\, c^i \quad \text{with some } c > 1$$
(14)

(The function $\text{round}\, x$ rounds x to the nearest integer.) a linear least squares fit is performed to

$$(x_i, y_i) = \left(\ln \frac{b_i + b_{i+1} - 1}{2}, \ln \sum_{k=b_i}^{b_{i+1}-1} \frac{\hat{p}(k)}{b_{i+1} - b_i}\right)$$
(15)

or

$$(x_i, y_i) = \left(\ln \frac{b_i + b_{i+1} - 1}{2}, \ln \sum_{k=b_i}^{b_{i+1}-1} \frac{\hat{P}(k)}{b_{i+1} - b_i}\right),$$
(16)

respectively. As a consequence of the binning the width of the distribution of the estimate $\hat{\gamma}$ of the power-law exponent γ is reduced, see Figures 1c, 1d and Table 1. According to the numerical experiments a fit of the logarithmically binned cumulative distribution gives the best results among graphical methods. It shows the smallest systematic bias.

All the methods that have been considered so far have a common weakness. In the deviation of (6) and (7) it was assumed that the standard deviation of the distribution of
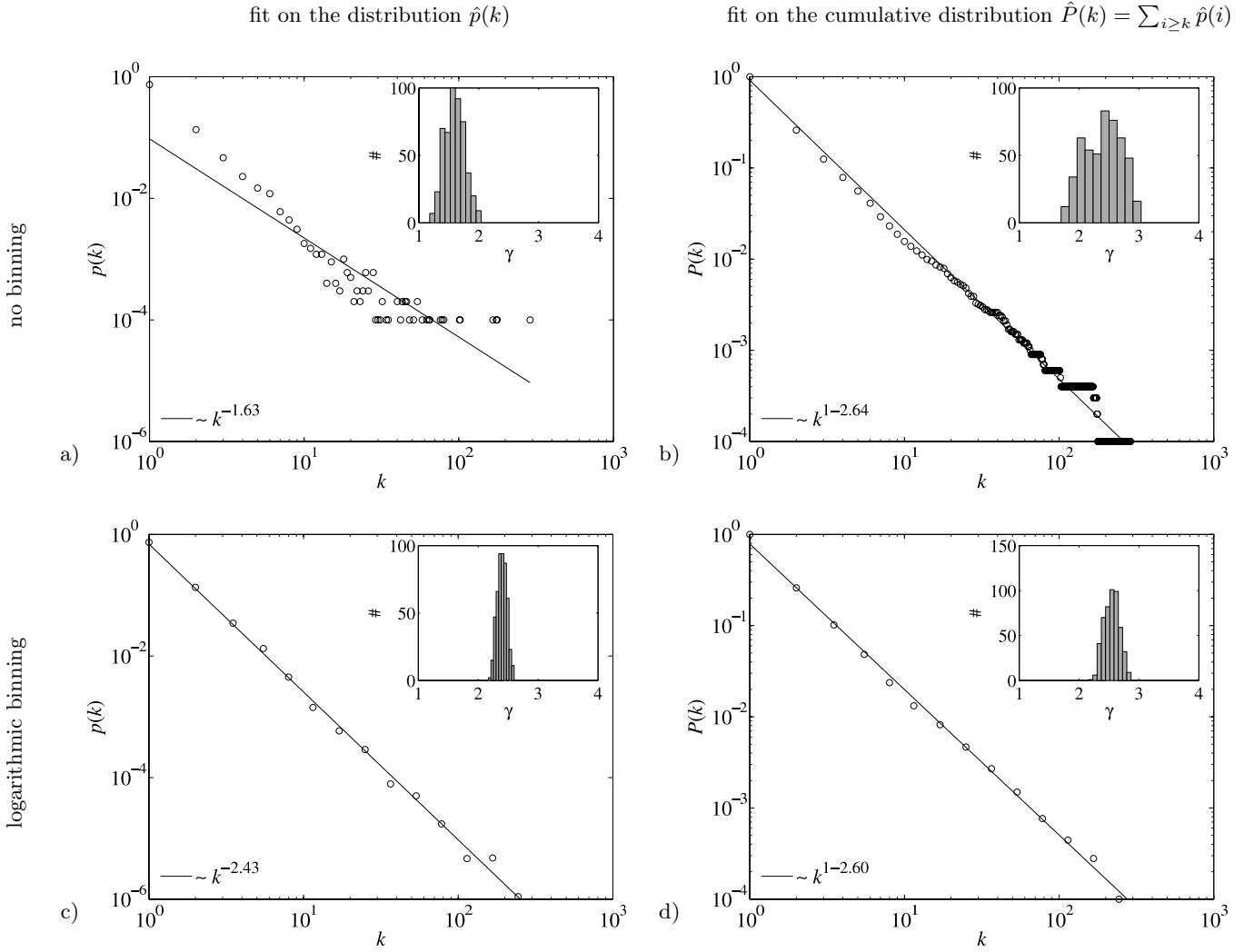
fit on the distribution $\hat{p}(k)$ fit on the cumulative distribution $\hat{P}(k) = \sum_{i \geq k} \hat{p}(i)$

Fig. 1. Comparison of various methods for estimating the exponent of a power-law. Each figure shows data for a single data set of $N = 10\,000$ samples drawn from distribution (2) with $\gamma = 2.5$. Insets present histograms of estimates for γ for 500 different data sets.

the error in y_i is the same for all data points (x_i, y_i). But this is obviously not the case. For fixed k the empirical distribution $\hat{p}(k)$ is a random variable with mean $p_1(k; \gamma)$ and standard deviation $\sqrt{p_1(k; \gamma)(1 - p_1(k; \gamma))/N}$. For the corresponding data on a logarithmic scale the standard deviation is approximately given by the quotient

$$\frac{\sqrt{p_1(k; \gamma)(1 - p_1(k; \gamma))/N}}{p_1(k; \gamma)} = \sqrt{\frac{1 - p_1(k; \gamma)}{N p_1(k; \gamma)}}. \quad (17)$$

A power-law distribution $p_1(k; \gamma)$ is a monotonically decreasing function and therefore (17) is an increasing function of k. Because the variation of the statistical error is not taken into account, the distribution of the estimate $\hat{\gamma}$ is very broad.

Methods that deal with the cumulative distribution have an additional weakness. Cumulation has the side-effect that the statistical errors in y_i are not independent any more, which violates another assumption of the deviation of (6) and (7).

To sum up, estimates of exponents of power-law distributions based on a linear least squares fit are intrinsically inaccurate and lack a sound mathematical justification.

3 Maximum likelihood estimators

Maximum likelihood estimators offer a solid alternative to graphical methods. Let $p(k; \theta)$ denote a single parameter probability distribution. The maximum likelihood estimator $\hat{\theta}_N$ for the unknown parameter based on a sample $m_1, m_2, \ldots, m_N$ of size N is given by

$$\hat{\theta}_N = \underset{\theta}{\operatorname{argmax}}[L(\theta)] = \underset{\theta}{\operatorname{argmax}}[\ln L(\theta)], \quad (18)$$

where

$$L(\theta) = \prod_{i=1}^{N} p(m_i; \theta) \quad (19)$$

Table 1. Mean and standard deviation of the distribution of the estimate for the power-law exponent γ for various methods. All methods have been applied to the same data sets of random numbers from distribution (2) with $\gamma = 2.5$. See text for details.

method	mean estimate	standard deviation of estimate
fit on empirical distribution	1.597	0.167
fit on cumulative empirical distribution	2.395	0.304
fit on empirical distribution with logarithmic binning[a]	2.397	0.080
fit on cumulative empirical distribution with logarithmic binning	2.544	0.127
maximum likelihood	2.500	0.016

[a] In [3] a similar experiment is reported. For a fit of the logarithmically binned empirical probability distribution the authors find a systematical bias of 29 %. I cannot reconstruct such a strong bias, instead I get a bias of 5 % only. Probably the quality of this method depends on the details of the binning procedure.

denotes the likelihood function. In the limit of asymptotically large samples and under some regularity conditions maximum likelihood estimators share some desirable features [5,6].

- The estimator $\hat{\theta}_N$ exists and is unique.
- The estimator $\hat{\theta}_N$ is consistent, that means for every $\varepsilon > 0$

$$\lim_{N \to \infty} \mathbb{P}\left[|\hat{\theta}_N - \theta| < \varepsilon\right] = 1\,, \qquad (20)$$

  where $\mathbb{P}\left[|\hat{\theta}_N - \theta| < \varepsilon\right]$ denotes the probability that the difference $|\hat{\theta}_N - \theta|$ is less than ε.
- The estimator $\hat{\theta}_N$ is asymptotically normal with mean θ and variance

$$(\Delta\hat{\theta}_N)^2 \doteq \left(N\,\mathbb{E}\left[\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\ln p(k;\theta)\right)^2\right]\right)^{-1}\,, \qquad (21)$$

  where $\mathbb{E}\left[\cdot\right]$ indicates the expectation value of the quantity in the brackets.
- Maximum likelihood estimators have asymptotically minimal variance among all asymptotically unbiased estimators. One says, they are asymptotically efficient.

4 Maximum likelihood estimators for genuine power-laws

The most general discrete genuine power-law distribution has a lower as well as an upper bound and is given by

$$p_{k_{\min}, k_{\max}}(k; \gamma) = \frac{k^{-\gamma}}{\zeta(\gamma, k_{\min}, k_{\max})} \qquad (22)$$

for $k \in \mathbb{N}$ with $k_{\min} \leq k < k_{\max}$. Where the non-standard notation

$$\zeta(\gamma, k_{\min}, k_{\max})\zeta(\gamma, k_{\min}) - \zeta(\gamma, k_{\max}) \qquad (23)$$

has been introduced. If the upper bound is missing the distribution

$$p_{k_{\min}}(k; \gamma) = \frac{k^{-\gamma}}{\zeta(\gamma, k_{\min})} \qquad (24)$$

has to be considered for $k \in \mathbb{N}$ with $k \geq k_{\min}$. The distributions (22) and (24) are generalizations of (2) and will be useful for the analysis of more general distributions that show a power-law behavior only in a certain range but have an arbitrary profile outside the power-law regime. This kind of distributions will be considered in Section 5.

The maximum likelihood estimator $\hat{\gamma}_N$ for the parameter γ of the distribution (22) follows from (18) and is given by

$$\hat{\gamma}_N = \operatorname*{argmax}_\gamma \left[-\gamma\left(\sum_{i=1}^N \ln m_i\right) - N\ln\zeta(\gamma, k_{\min}, k_{\max})\right] \qquad (25)$$

or equivalently by the implicit equation

$$\frac{\zeta'(\hat{\gamma}_N, k_{\min}, k_{\max})}{\zeta(\hat{\gamma}_N, k_{\min}, k_{\max})} + \frac{1}{N}\sum_{i=1}^N \ln m_i = 0\,, \qquad (26)$$

which has to be solved numerically. The prime denotes the derivative with respect to γ. The asymptotic variance of this estimator $\hat{\gamma}_N$ follows from (21) and equals

$$(\Delta\hat{\gamma}_N)^2 \doteq \frac{1}{N}$$
$$\times \frac{\zeta(\gamma, k_{\min}, k_{\max})^2}{\zeta''(\gamma, k_{\min}, k_{\max})\zeta(\gamma, k_{\min}, k_{\max}) - \zeta'(\gamma, k_{\min}, k_{\max})^2}\,. \qquad (27)$$

In the limit $k_{\max} \to \infty$ equations (25), (26), and (27) give the maximum likelihood estimator and the asymptotic variance of this estimator for power-law distributions lacking an upper cut-off (24). A graphical representation of the standard deviation (27) in the limit $k_{\max} \to \infty$ is given in Figure 2. For each fixed $k_{\min}$ the quantity $\Delta\hat{\gamma}_N\sqrt{N}$ grows faster than linear with γ. Therefore the larger the exponent γ the larger the sample size that is necessary to get an estimate within a given error bound.

If the maximum likelihood method is applied (assuming a distribution (24)) to the same data as in Section 2, numerical experiments show that the estimates for the exponent are much more precise. The estimate has no identifiable systematic bias, the standard deviation of the distribution of the estimate is smaller by an order of magnitude compared to graphical methods, see Table 1 and Figure 3.

5 Maximum likelihood method for general power-law distributions

The maximum likelihood procedure outlined in Section 4 can be generalized further to distributions $p(k)$ that are no
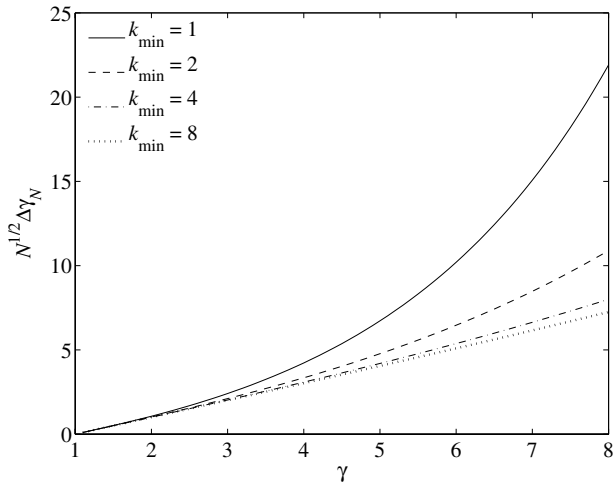
Fig. 2. Asymptotic standard deviation for maximum likelihood estimators for the exponent of a power-law distribution (24).
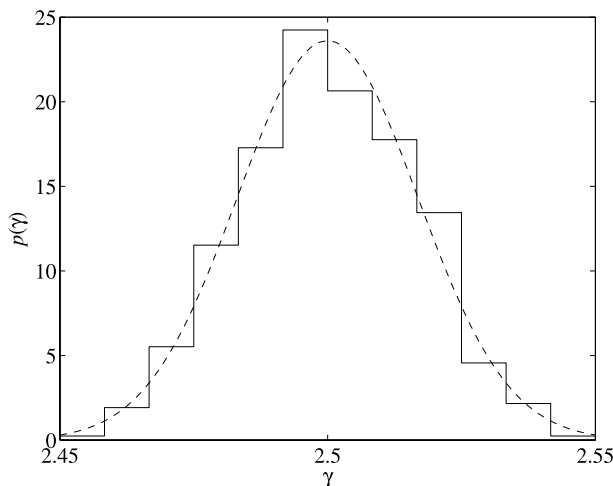

Fig. 3. Empirical distribution of the maximum likelihood estimator (histogram) versus its theoretical asymptotic distribution, which is given by a normal distribution with mean $\gamma = 2.5$ and variance (27). The histogram has been obtained from the same data as in Figure 1.

pure power-laws (22) or (24) but follow a power-law within a certain finite range or follow a power-law in the whole tail of the distribution and have an arbitrary profile outside the power-law regime. The popurse of this section is to establish methods for identifying the power-law regime and for estimating the exponent of the power-law regime without making special assumptions about the profile of the probability distribution beyond the power-law regime.

The main problem for a generalization of the maximum likelihood approach is that there might be no good hypothesis for the profile of the probability distribution beyond the power-law regime. To overcome this difficulty the empirical data set is restricted to a window $k_{c\,min} \leq m_i < k_{c\,max}$. (The following discussion covers the case of a power-law tail distributions as well by setting $k_{c\,max} = \infty$.) Assuming that $p(k)$ has a power-law profile for $k_{c\,min} \leq$

$k < k_{c\,max}$ then the probability distribution of the restricted data set is given by (22) with $k_{min} = k_{c\,min}$ and $k_{max} = k_{c\,max}$ (or by (24) with $k_{min} = k_{c\,min}$) and some unknown exponent γ. This allows to estimate the power-law exponent by the application of the maximum likelihood method on the restricted data set of size N' as presented in Section 4 *without* making a hypothesis about the profile of the probability distribution beyond the power-law regime.

In order to apply the maximum likelihood method one has to determine the cut-off points $k_{c\,min}$ and $k_{c\,max}$ first. Here it has to be taken into account that if the window $k_{c\,min} \leq m_i < k_{c\,max}$ is chosen too large the estimate $\hat{\gamma}$ is systematically biased, but on the other hand if it is too small the statistical error is larger than necessary. In some cases one can make conservative estimates for $k_{c\,min}$ and $k_{c\,max}$ by plotting the empirical probability distribution (8) on a double-logarithmic scale. An appropriate window can also be found by determining estimates $\hat{\gamma}_{N'}(k_{c\,min})$ as a function of the window and a χ^2-test.

Assuming the empirical data is drawn from a distribution with a power-law tail (no upper cut-off) the lower cut-off point $k_{c\,min}$ can be determined in the following systematic way. By varying the parameter $k_{c\,min}$ the maximum likelihood approach gives a sequence of estimates $\hat{\gamma}_{N'}(k_{c\,min})$. If $k_{c\,min}$ is very large the estimate will be quite inaccurate because only a tiny fraction of the experimental data is taken into account; but the smaller the cut-off $k_{c\,min}$ the more accurate the estimate of the exponent. If $k_{c\,min}$ approaches the point from above (but is still above) where the probability distribution starts do differ from a power-law $\hat{\gamma}_{N'}(k_{c\,min})$ will give a very precise estimate for the exponent γ. On the other hand, if $k_{c\,min}$ is too small the hypothesis that the (restricted) empirical data is drawn from a power-law distribution is violated which causes a significant change of the estimate of the power-law exponent.

If the empirical data is drawn from a distribution having both a lower crossover point as well as an upper crossover point a sequence of estimates $\hat{\gamma}_{N'}(k_{c\,min})$ is determined by restricting the data to a sliding window $k_{c\,min} \leq m_i < wk_{c\,min} = k_{c\,max}$ with $w > 1$. As long as the window lies completely within the power-law regime the maximum likelihood estimate obtained from the restricted data set will give a reliable estimate of the power-law exponent. If the window lies at least partly outside the power-law regime the estimate is systematically biased.

To illustrate the procedures outlined above I generated two data sets from two distributions having a power-law regime. The first data set of $N = 10\,000$ samples was drawn from a distribution with a power-law tail which is given by

$$p(k) \sim \begin{cases} 5^{-2.5} & \text{for } 1 \leq k \leq 5 \\ k^{-2.5} & \text{for } k > 5. \end{cases} \qquad (28)$$

Plotting the sequence of estimates $\hat{\gamma}_{N'}(k_{c\,min})$ against the parameter $k_{c\,min}$ reveals the exponent $\gamma = 2.5$ as wells as the crossover point $k = 5$ very clearly, see Figure 4. The second data set of $N = 100\,000$ samples was drawn from
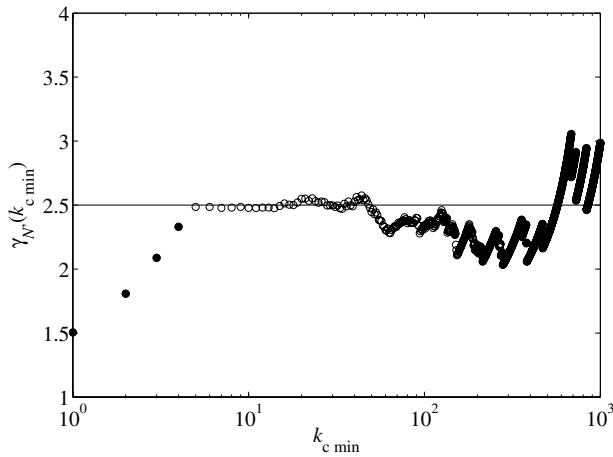
Fig. 4. Sequence of estimates $\hat{\gamma}_{N'}(k_{c\,min})$ as a function of the cut-off $k_{c\,min}$ for a data set of $N = 10\,000$ samples from distribution (28). Filled symbols mark where the χ^2-test has rejected the hypothesis that the restricted data follows a power-law (24) with exponent $\gamma = \hat{\gamma}_{N'}(k_{c\,min})$. An error probability of $\alpha = 0.001$ was chosen.

a distribution with two crossover points, viz.

$$p(k) \sim \begin{cases} 5^{-2.25} & \text{for } 1 \le k \le 5 \\ k^{-2.25} & \text{for } 5 \le k \le 100 \\ 100^{-2.25}\mathrm{e}^{-0.05(k-100)} & \text{for } k > 100. \end{cases} \quad (29)$$

Figure 5 shows the sequence of estimates $\hat{\gamma}_{N'}(k_{c\,min})$ that had been determined from restricted data sets of samples within the sliding window $k_{c\,min} \le m_i < 5k_{c\,min}$. This sequence exhibits a broad plateau that corresponds to the power-law exponent $\gamma = 2.25$. If the window does not lie completely inside the power-law regime the estimate $\hat{\gamma}_{N'}(k_{c\,min})$ deviates systematically from the known exponent.

Apart from a visual inspection of the $\hat{\gamma}_{N'}(k_{c\,min})$ plot the crossover point(s) to the power-law regime can be determined by means of a χ^2-test. To apply a χ^2-test the data set has to be divided into some bins and the following binning turned out to be appropriate: The data is partitioned into a small number b, say $b = 6$, of bins. In the case of a distribution with a power-law tail this means each bin j collects n_j items such that

$$n_1 = N'\hat{p}(k_{c\,min}) \qquad q_1 = p_{k_{c\,min}}(k_{c\,min}; \hat{\gamma}_{N'}(k_{c\,min}))$$
$$(30)$$

$$n_2 = N'\hat{p}(k_{c\,min} + 1) \quad q_2 = p_{k_{c\,min}}(k_{c\,min} + 1; \hat{\gamma}_{N'}(k_{c\,min}))$$
$$(31)$$

$$\vdots \qquad\qquad\qquad \vdots$$

and finally

$$n_b = N' \sum_{k=k_{c\,min}+b-1}^{\infty} \hat{p}(k) \quad q_b = \sum_{k=k_{c\,min}+b-1}^{\infty} p_{k_{c\,min}}(k; \hat{\gamma}_{N'}(k_{c\,min})),$$
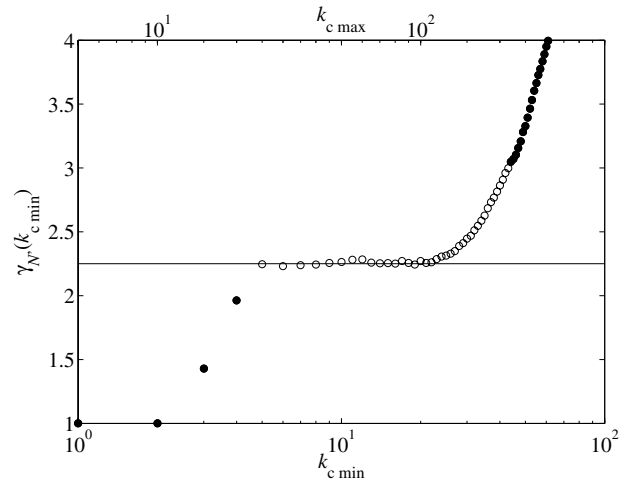$$(32)$$

Fig. 5. Sequence of estimates $\hat{\gamma}_{N'}(k_{c\,min})$ as a function of the lower cut-off $k_{c\,min}$ for a data set of $N = 100\,000$ samples from distribution (29). Filled symbols mark where the χ^2-test has rejected the hypothesis that the restricted data follows a power-law (22) with exponent $\gamma = \hat{\gamma}_{N'}(k_{c\,min})$. An error probability of $\alpha = 0.001$ and the window width $w = 5$ had been chosen.

where q_j denotes the probability that a data point falls into bin j under the assumption that the (restricted) data follows the power-law (24) with $k_{min} = k_{c\,min}$ and the exponent $\gamma = \hat{\gamma}_{N'}(k_{c\,min})$. For distributions with a finite power-law regime the binning procedure can be carried out in a similar way. In this case the summation index in (32) is bounded by $k_{c\,min} + b - 1 \le k < k_{c\,max}$ and the probability $p_{k_{c\,min},k_{c\,max}}(k; \hat{\gamma}_{N'}(k_{c\,min}))$ has to be considered instead of $p_{k_{c\,min}}(k; \hat{\gamma}_{N'}(k_{c\,min}))$.

The test statistic of the χ^2-test is given by

$$c^2 = \sum_{j=1}^{b} \frac{(n_j - N'q_j)^2}{N'q_j}. \quad (33)$$

If the to $k_{c\,min} \le m_i < k_{c\,max}$ restricted data is given by the power-law (22) or (24) with $k_{min} = k_{c\,min}$, $k_{max} = k_{c\,max}$, and the exponent $\gamma = \hat{\gamma}_{N'}(k_{c\,min})$ then the statistic c^2 follows asymptotically a χ^2-distribution with $\nu = (b-1)$ degrees of freedom, which is given by

$$p_{\chi^2}(x, \nu) = \frac{x^{\nu/2-1}\mathrm{e}^{-x/2}}{2^{-\nu/2}\,\Gamma(\nu/2)}. \quad (34)$$

Let χ_α^2 be the $(1 - \alpha)$-quantile of the distribution (34). The hypothesis that the restricted data is given by the power-law (22) or (24), respectively, with $k_{min} = k_{c\,min}$, $k_{max} = k_{c\,max}$, and the exponent $\gamma = \hat{\gamma}_{N'}(k_{c\,min})$ is accepted with the error probability α if $c^2 \le \chi_\alpha^2$. If the window $k_{c\,min} \le m_i < k_{c\,max}$ lies not completely within the power-law regime this hypothesis will be rejected by the χ^2-test and one can detect the upper crossover point as well as the lower crossover point (where the power-law loses its validity) in a reliable way, see Figures 4 and 5.

6 Computational remarks

The normalizing factors of the probability distributions (22) and (24) are given by the Hurwitz-ζ-function. This function is less common than other special functions and may not be available in the reader's favorite statistical software package but the GNU Scientific Library [7] offers an open source implementation of this function. A direct calculation of the Hurwitz-ζ-function by truncating the sum (3) gives unsatisfactory results.

The maximum likelihood estimator of the exponent can be computed numerically either by solving (25) or (26). Equation (25) has the advantage that it can be solved without calculating derivatives of the Hurwitz-ζ-function [8], whereas the solution of (26) involves its first derivative (e.g. bisection method) or even higher derivatives (e.g. Newton-Raphson method). An explicit implementation of these derivatives is often not available but may be calculated numerically.

7 Conclusion

Methods based on a least squares fit are not suited to establish estimates for power-law distribution exponents because least squares fits rely on assumptions about the data set that are not fulfilled by empirical data from power-law distributions. In this paper maximum likelihood estimators have been introduced as a reliable alternative to graphical methods. These estimators are asymptotically efficient and can be applied to data from a wide class of distributions having a power-law regime. The crossover points that separate the power-law regime from the rest of the distribution can be determined by a procedure based on a χ^2-test.

Finally I would like to mention that the idea to plot a sequence of estimates $\hat{\gamma}_{N'}(k_{c\,min})$ as shown in Figure 4 is related to so-called Hill plots [9,10]. The Hill estimator is a maximum likelihood estimator for the inverse of the exponent of the *continuous* Pareto distribution $p(k) = (\gamma - 1)(k/k_{min})^{-\gamma}/k_{min}$, see [10] for a detailed discussion.

References

1. M.E.J. Newman, Contemp. Phys. **46**, 323 (2005)
2. E.F. Keller, BioEssays **27**, 1060 (2005)
3. M.L. Goldstein, S.A. Morris, G.G. Yen, Eur. Phys. J. B **41**, 255 (2004)
4. F. Ramsey, D. Schafer, *The Statistical Sleuth: A Course in Methods of Data Analysis*, 2nd edn. (Duxbury Press, Pacific Grove, CA, 2002)
5. L.J. Bain, M. Engelhardt, *Introduction to Probability and Mathematical Statistics*, Duxbury Classic Series, 2nd edn. (Duxbury Press, 2000)
6. Y. Pawitan, *In all likelihood: statistical modelling and inference using likelihood*, Oxford science publications (Oxford University Press, 2001)
7. *GNU Scientific Library*, http://www.gnu.org/software/gsl/
8. R.P. Brent, *Algorithms for Minimization without Derivatives* (Prentice-Hall, Englewood Cliffs, New Jersey, 1973)
9. B.M. Hill, Ann. Stat. **3**, 1163 (1975)
10. H. Drees, L. de Haan, S. Resnick, Ann. Stat. **28**, 254 (2000)