View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of the Western Cape Research Repository

Mon. Not. R. Astron. Soc. **397,** 495–505 (2009) doi:10.1111/j.1365-2966.2009.14956.x

Fitting power-law distributions to data with measurement errors

C. Koen★ and L. Kondlo

Department of Statistics, University of the Western Cape, Private Bag X17, Bellville, 7535 Cape, South Africa

Accepted 2009 April 21. Received 2009 April 9; in original form 2008 December 22

ABSTRACT

If X, which follows a power-law distribution, is observed subject to Gaussian measurement error e, then $X + e$ is distributed as the convolution of the power-law and Gaussian distributions. Maximum-likelihood estimation of the parameters of the two distributions is considered. Large-sample formulae are given for the covariance matrix of the estimated parameters, and implementation of a small-sample method (the jackknife) is also described. Other topics dealt with are tests for goodness of fit of the posited distribution, and tests whether special cases (no measurement errors or an infinite upper limit to the power-law distribution) may be preferred. The application of the methodology is illustrated by fitting convolved distributions to masses of giant molecular clouds in M33 and the Large Magellanic Cloud (LMC), and to H I cloud masses in the LMC.

Key words: methods: statistical – ISM: clouds – galaxies: ISM.

1 INTRODUCTION

The assumption of power-law probability distributions is ubiquitous in astronomy. Examples include stellar initial mass functions (e.g. Kroupa 2001); masses of molecular clouds (e.g. Rosolowsky 2005); luminosity functions for various classes of objects, such as galaxies (Misgeld, Hilker & Mieske 2009), star clusters (Gieles et al. 2006; van den Bergh 2006) and Kuiper-belt objects (Fraser & Kavelaars 2008); sizes of a variety of astronomical objects, as diverse as bolides (Brown et al. 2002) and circumstellar discs (Vicente & Alves 2005); the column densities implied by C IV systems in Lyα forests (Ellison et al. 2000); the intensities of giant radio pulses from pulsars (Kinkhabwala & Thorsett 2000); numerous properties of the sun, such as the burst energies, durations and interburst intervals in the solar wind (Freeman, Watkins & Riley 2000); etc. Often power-law indices, and other distributional parameters, are estimated by fitting the power-law form to a set of observations $\{y_1, y_2, \ldots, y_N\}$. This is not entirely appropriate if the measurements are contaminated by substantial measurement errors. Denote by x_j the true (i.e. error-free) values of the quantity of interest and by e_j the errors, so that the measured values $y_j = x_j + e_j$. If the errors have the probability density function (PDF) f_e, and the power-law PDF is denoted by f_x, then

$$f_y(y) = \int_{-\infty}^{\infty} f_x(x) f_e(y - x)\,\mathrm{d}x, \tag{1}$$

i.e. the measurements y_j have a convolved PDF (e.g. Mood, Graybill & Boes 1974). For the truncated power law

$$f_x(x) = \frac{\gamma x^{-(\gamma+1)}}{L^{-\gamma} - U^{-\gamma}} \qquad L \le x \le U \tag{2}$$

and zero mean Gaussian errors,

$$f_y(y) = \frac{1}{\sqrt{2\pi}\sigma} \frac{\gamma}{L^{-\gamma} - U^{-\gamma}} \int_L^U x^{-(\gamma+1)} \exp\left[-\frac{1}{2}\left(\frac{y-x}{\sigma}\right)^2\right] \mathrm{d}x \tag{3}$$

follows. This PDF could differ substantially from (2).

One simple way of seeing the difference between the distributions (2) and (3) is to compare probability–probability (PP) plots based on the two forms. The calculations required for the PP plots are based on cumulative distribution functions (CDFs). The CDFs corresponding to (1)–(3) are

$$F_y(y) = \int_{-\infty}^{\infty} F_x(x) f_e(y - x)\,\mathrm{d}x,$$
$$F_x(x) = \frac{L^{-\gamma} - x^{-\gamma}}{L^{-\gamma} - U^{-\gamma}} \qquad L \le x, \tag{4}$$

$$F_y(y) = \Phi\left(\frac{y-U}{\sigma}\right) + \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{L^{-\gamma} - U^{-\gamma}} \times \int_L^U \left(L^{-\gamma} - x^{-\gamma}\right) \exp\left[-\frac{1}{2}\left(\frac{y-x}{\sigma}\right)^2\right] \mathrm{d}x, \tag{5}$$

where

$$\Phi(v) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{v} \mathrm{e}^{-t^2/2}\,\mathrm{d}t$$

is the standard normal CDF. The PP plots are graphs of the empirical CDFs

$$\widehat{F}(y) = \frac{(y_j < y)}{N+1}$$

against the theoretical counterparts (4) or (5). The empirical CDF is a step function, with a step of size $1/(N+1)$ at each measurement

★E-mail: ckoen@uwc.ac.za

© 2009 The Authors. Journal compilation © 2009 RAS

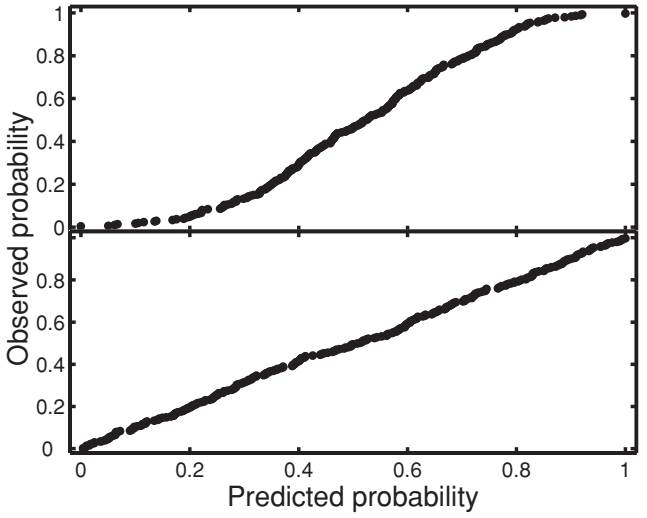

Figure 1. PP plots for a simulated data set, consisting of 300 values distributed according to equation (3). The plot in the top panel is based on the (incorrect) assumption that there is no measurement error; the plot in the bottom panel incorporates Gaussian measurement errors.

value. The theoretical, or predicted, distribution values are obtained by first estimating the parameters occurring in the theoretical CDFs (i.e. L, U, γ and σ); F_y can then be explicitly evaluated in each observation value y_j. If the data are indeed drawn from the proposed distribution, then the plotted points should lie close to the straight line $\widehat{F} = F$. The interested reader is referred to D'Agostino & Stephens (1986) for a more extensive discussion.

A simulated example with $\gamma = 1.5$, $L = 3$, $U = 6$, $\sigma = 0.4$ can be seen in Fig. 1. For both plots, values of the parameters estimated by maximum likelihood have been substituted in order to calculate the CDF (4 or 5). The non-linear form in the top panel, corresponding to the pure power law, convincingly demonstrates that the distribution (4) does *not* describe the data. On the other hand, (5) (bottom panel) is clearly a good model for the data, since observed and predicted distributions are closely similar.

A more direct qualitative demonstration of some of the pitfalls caused by the presence of measurement errors is given in Fig. 2, based on the same data used to generate Fig. 1. The top panel shows a histogram of a sample from an error-free power-law distribution. A histogram of the same data, with added measurement errors, is given in the bottom panel. Two effects are clearly visible: the contaminated data extend beyond the interval $[L, U]$ over which the error-free data occur, and the shape of the distribution is changed. The first effect will clearly bias estimates of the lower and upper limits, while the second will lead to biased estimates of the power-law exponent: in particular, since the data are spread over a wider interval, the value of γ estimated from error contaminated will generally be too small. Maximum-likelihood (see below) estimates of γ, L and U for the data with measurement errors are 1.19, 2.27 and 7.12, respectively, if the model (2) is fitted to the data. On the other hand, fitting the correct model (3) to the data gives estimates 1.46, 3.05, 6.03 – which compare favourable with the true parameter values $\gamma = 1.5$, $L = 3$ and $U = 6$.

Power-law distributions without upper limits, i.e. $U \to \infty$, are also commonly of interest. If there are no measurement errors,

$$f_x(x) = \gamma L^\gamma x^{-(\gamma+1)} \qquad x \geq L,$$
$$F_x(x) = 1 - L^\gamma x^{-\gamma}. \tag{6}$$

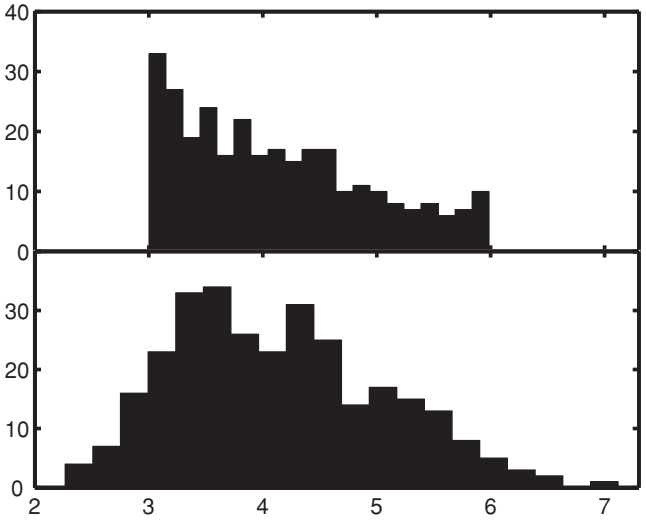

Figure 2. The top histogram is for 300 simulated data elements from a power-law distribution with $L = 3$, $U = 6$ and $\gamma = 1.5$. Zero mean Gaussian measurements errors with $\sigma = 0.4$ were added to give the convolved distribution in the bottom panel. The error-contaminated data 'spill' out of the interval $[L, U]$, most notably for data near the lower limit L.

If the observations are subject to measurement errors, then

$$f_y(y) = \frac{\gamma L^\gamma}{\sqrt{2\pi}\sigma} \int_L^\infty x^{-(\gamma+1)} \exp\left[-\frac{1}{2}\left(\frac{y-x}{\sigma}\right)^2\right] \mathrm{d}x,$$
$$F_y(y) = \Phi\left(\frac{y-L}{\sigma}\right) - \frac{L^\gamma}{\sqrt{2\pi}\sigma} \int_L^\infty x^{-\gamma} \exp\left[-\frac{1}{2}\left(\frac{y-x}{\sigma}\right)^2\right] \mathrm{d}x. \tag{7}$$

2 STATISTICAL THEORY

2.1 Maximum-likelihood estimation

Given that it is assumed that the data follow the statistical distribution (3), determination of the unknown parameters by maximum-likelihood estimation (MLE) is indicated. Provided the observations $y_j (j = 1, 2, \ldots, N)$ are independent, the likelihood is simply the product of the N PDFs of the form (3), with $y = y_j$ in the jth term. The log likelihood (which is easier to work with) is then

$$\mathcal{L} = -\frac{N}{2}\log 2\pi - N\log\sigma + N\log\gamma - N\log(L^{-\gamma} - U^{-\gamma}) + \sum_{j=1}^{N} \log \int_L^U x^{-(\gamma+1)} \exp\left[-\frac{1}{2}\left(\frac{y_j - x}{\sigma}\right)^2\right] \mathrm{d}x. \tag{8}$$

If the error variance σ^2 is specified, σ is treated as a constant; otherwise, it is a fourth parameter to be estimated along with γ, L and U. MLE entails maximization of (8) with respect to the unknown parameters, which can be done numerically. Alternatively, the first partial derivatives of $\mathcal{L}$ with respect to the unknown parameters could be set equal to zero, giving four simultaneous non-linear algebraic equations which could be solved for the unknowns. The former procedure is simpler, and therefore the one adopted in this paper. The partial derivatives are nonetheless given in Appendix A, as these are required in what follows.

The notation

$$\boldsymbol{\phi} = [\gamma \;\; L \;\; U \;\; \sigma]' \tag{9}$$

for the vector of parameters, and

$$\widehat{\boldsymbol{\phi}} = [\widehat{\gamma}\ \ \widehat{L}\ \ \widehat{U}\ \ \widehat{\sigma}]' \tag{10}$$

for the vector of estimates, is a useful shorthand.

The analogous derivatives of $\mathcal{L}$ for the untruncated distribution (7) are in Appendix B. In this case,

$$\boldsymbol{\phi} = [\gamma\ \ L\ \ \sigma]' \,. \tag{11}$$

2.2 The accuracy of the estimated parameters

For data sets which are not too large, non-parametric computer-intensive methods such as bootstrapping or the jackknife can be used to find, for example, standard errors of the estimated parameters. Bootstrapping (e.g. Efron & Tibshirani 1993) entails repeating the estimation for pseudo-samples of size N, drawn with replacement from the original N data elements. The number of pseudo-samples would typically be of the order of a thousand or more. In the case of the jackknife, pseudo-samples of size $N-1$ are produced by excluding each of the N observations in turn. The estimation is repeated for each of the N pseudo-samples, giving the set $\{\widehat{\boldsymbol{\phi}}_{(1)}, \widehat{\boldsymbol{\phi}}_{(2)}, \ldots, \widehat{\boldsymbol{\phi}}_{(N)}\}$ of N parameter vector values. The corresponding covariance matrix is

$$\mathbf{C}_J = \frac{N-1}{N}\sum_{j=1}^{N}(\widehat{\boldsymbol{\phi}}_{(j)} - \widehat{\boldsymbol{\phi}}_{(*)})(\widehat{\boldsymbol{\phi}}_{(j)} - \widehat{\boldsymbol{\phi}}_{(*)})' \quad \widehat{\boldsymbol{\phi}}_{(*)} \equiv \frac{1}{N}\sum_{j=1}^{N}\widehat{\boldsymbol{\phi}}_{(j)}. \tag{12}$$

Bissell & Ferguson (1975) is a readable introduction to the jackknife.

For large samples, the standard asymptotic theory of maximum-likelihood estimates can be used: as N increases, the covariance matrix of the parameter estimates approaches the inverse of the Fisher information matrix

$$\mathcal{F} = \left[-E\frac{\partial^2\mathcal{L}}{\partial\phi_i\partial\phi_j}\right] \quad i, j = 1, 2, 3, 4, \tag{13}$$

where E indicates the expected value (e.g. Rice 2007). The required second derivatives of $\mathcal{L}$ are listed in Appendix A (or in Appendix B for the untruncated distribution). Obtaining the expected values of these is daunting, hence the usual alternative is followed of substituting data values for expectations (giving the 'empirical' information matrix). The covariance matrix is then

$$\mathbf{C}_I = -\left[\frac{\partial^2\mathcal{L}}{\partial\phi_i\partial\phi_j}\right]^{-1} \quad i, j = 1, 2, 3, 4. \tag{14}$$

Although calculation of $\mathbf{C}_I$ requires a fair amount of careful computer programming, the calculation time is trivial. Furthermore, its approximation of the covariance matrix of $\widehat{\boldsymbol{\phi}}$ improves with increasing sample size. These facts make it an attractive alternative to $\mathbf{C}_J$ for large N.

2.3 Goodness of fit of the model

Fig. 1 demonstrates the importance of verifying that the statistical model fitted to the data is appropriate. PP (or quantile–quantile) plots are useful informal goodness-of-fit tests. More formally, there are a number of statistical tests which can be performed – the Kolmogorov–Smirnov, Anderson–Darling and Cramér–von-Mises are probably the best known. All three of these statistics measure discrepancies between the theoretical and observed CDFs, and in that sense are quantitative (rather than visual) methods to evaluate deviations from linearity in plots such as Fig. 1. Unfortunately, none of these statistics is distribution free in the present context, hence percentage points would need to be found by simulation or bootstrapping. For large N (a few hundred or more), this is currently still tedious due to the excessive amount of computer time needed (primarily required for the maximization of $\mathcal{L}$ in 8).

A very simple alternative is to use a χ^2 statistic S which compares the observed and predicted numbers of data elements in P selected intervals:

$$S = \sum_{j=1}^{P}\frac{(O_j - E_j)^2}{E_j} \sim \chi^2_{P-K-1}. \tag{15}$$

In (15), O_j and E_j are, respectively, the observed and expected numbers of data elements in the jth of the P bins and K is the number of estimated parameters (four for the full model, if σ is unspecified; three if σ is given). Let the bins be (u_0, u_1); (u_1, u_2); ... ; (u_{P-1}, u_P), with $u_0 = -\infty$ and $u_P = \infty$. Then,

$$E_j = N[F_y(u_j) - F_y(u_{j-1})] \,. \tag{16}$$

2.4 Testing for specific power-law distributional forms

Specific forms of (3) are often of interest. Two types of null hypotheses could be of interest. In the first, there may be theoretical models, or other data sets, which suggest particular parameter values, and the researcher may want to test for conformity with these. A common example is to test whether the exponent $\gamma = c$, for specified c. The second type of null hypothesis involves testing whether a simplified form, rather than the full form (3), provides an adequate description of the data. Two examples are: $U \to \infty$, for which (3) reduces to

$$f_y(y) = \frac{1}{\sqrt{2\pi}\sigma}\gamma L^\gamma \int_L^\infty x^{-(\gamma+1)}\exp\left[-\frac{1}{2}\left(\frac{y-x}{\sigma}\right)^2\right]\mathrm{d}x; \tag{17}$$

and $\sigma \to 0$, for which (3) reduces to (2). A further special case of both (17) and (2) is $U \to \infty$, $\sigma = 0$, which gives the standard power-law distributional form (6).

The first type of hypothesis can be tested by likelihood ratio test statistics (e.g. Mood et al. 1974):

$$\Lambda = 2[\max\mathcal{L}(H1) - \max\mathcal{L}(H0)] \sim \chi^2_d \,. \tag{18}$$

The null and alternative hypotheses are denoted by $H0$ and $H1$, respectively, and the maxima of the log likelihood are determined under each of these. The statistics has an asymptotic χ^2_d distribution, with degrees of freedom d equal to the number of constraints imposed by $H0$. (For example, if the null hypothesis specifies values for γ and either L or U, then $d = 2$.)

The same procedure cannot be used for the second type of hypothesis, since required regularity conditions are not satisfied. An extensive discussion of broad classes of such problems can be found in Andrews (2001). The two examples above are of the type in which there are parameters (U or σ) which lie on the boundaries of their parameter spaces under the null hypothesis. (This is easy to see in the case $H0$: $\sigma = 0$; in the case $U \to \infty$, it is more obvious if one considers the equivalent hypothesis $H0$: $U^{-1} = 0$.)

Fortunately, there is a simple alternative way of viewing the problem, namely as one of *model selection*. In that case, information criteria such as

$$\mathrm{AIC} = -2\mathcal{L} + 2K + \frac{2K(K+1)}{N-K-1} \ \text{(Akaike information criterion)},$$
$$\mathrm{BIC} = -2\mathcal{L} + K\log N \qquad \text{(Bayes information criterion)} \tag{19}$$

(K being the number of model parameters) for the competing models can be compared. The likelihood term in these criteria measures how well the model fits the data; since it appears as the negative of the likelihood, the term is small for a good fit. The remaining term(s) are a measure of the model complexity – simple models (i.e. small values of K) are preferred. It is therefore desirable to have both $-\mathcal{L}$ and the model complexity terms as small as possible, i.e. the 'best' model is that which minimizes the information criterion.

Burnham & Anderson (2004) provide an excellent discussion of the two information criteria in (19). The AIC is a bias-corrected estimator for the Kullback–Leibler information (the information lost when a model is used to summarize the information in a set of observations). In the case of the BIC, it is implicitly assumed that one of the candidate models *fully* describes the observations; as the number N of observations increases, so the probability approaches unity that the BIC selects this model. There is thus a philosophical difference between the two criteria. Burnham & Anderson (2004) also point out a Bayesian interpretation, in which the two information criteria assign different sets of prior probabilities to the collection of candidate models. In practice, the AIC is usually best when increased model complexity leads to incrementally better fits, while the BIC performs best for data which can be modelled very well with simple models (i.e. small K).

It is also possible to attach probabilities to each of the models (Burnham & Anderson 2004): let

$$\Delta_i = \exp\left[-\frac{1}{2}(IC_i - IC_{\min})\right],$$

where i indexes the model and $IC_{\min}$ is the minimum value of the information criterion (either AIC or BIC). Then, the model probabilities are

$$p_i = \frac{\Delta_i}{\sum_i \Delta_i}, \tag{20}$$

and obviously the model with the largest probability is selected.

The maximum-likelihood parameter estimates for the zero measurement error models (6) and (2) are known:

$$\widehat{L} = \min_i(x_i),$$
$$\widehat{\gamma} = N\left[\sum_i(\log x_i - \log \widehat{L})\right]^{-1} \tag{21}$$

for $U \to \infty$, and

$$\widehat{L} = \min_i(x_i),$$
$$\widehat{U} = \max_i(x_i),$$
$$\frac{N}{\widehat{\gamma}} + \frac{Nr^{\widehat{\gamma}}\log r}{1 - r^{\widehat{\gamma}}} = \sum_{i=1}^{N}(\log x_i - \log \widehat{L}), \quad r \equiv \min_i(x_i)/\max_i(x_i) \tag{22}$$

for finite U. The last of equations (22), for $\widehat{\gamma}$, is implicit (Aban, Meerschaert & Panorska 2006).

3 A COMPUTATIONAL DETAIL

Evaluation of the likelihood function is computationally expensive, since N convolution integrals need to be calculated. It is therefore important that the integration domain be restricted to a subinterval of $[L, U]$ over which the integrand is significantly non-zero. Set

$$I(y_j) = x^{-(\gamma+1)}\exp\left[-\frac{1}{2}\left(\frac{y_j - x}{\sigma}\right)^2\right]$$

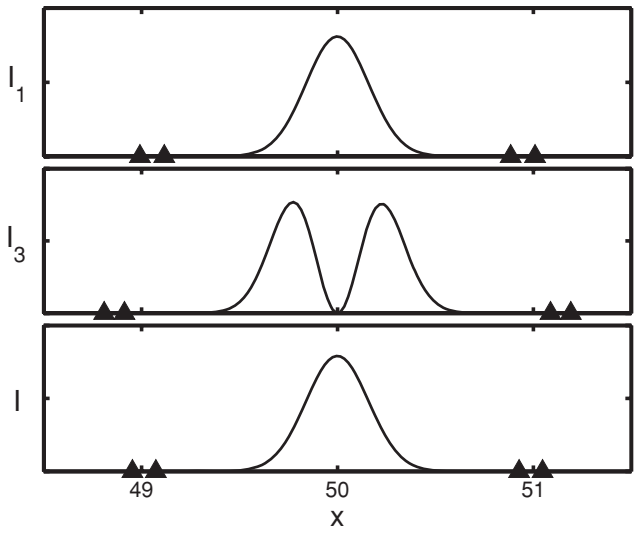

Figure 3. Examples of three of the integrands in equation (A1), calculated in the point $x = 50$. The power-law exponent is $\gamma = 0.82$, and $\sigma = 0.16$. The triangles on the horizontal axes show the points where the functions are equal to 10^{-6} (inner pairs) or 10^{-8} (outer pairs). See Section 3 for further discussion.

then, since $x \geq L$,

$$I(y_j) \leq L^{-(\gamma+1)}\exp\left[-\frac{1}{2}\left(\frac{y_j - x}{\sigma}\right)^2\right].$$

It is then not difficult to show that $I(y_j) \geq \epsilon$ only over the interval $[y_j - \delta, y_j + \delta]$, where

$$\delta = \sigma\sqrt{-2\left[\log\epsilon + (\gamma+1)\log L\right]}. \tag{23}$$

More generally, the arguments of the integrals in (A1) are of the form

$$I_i(j) = f_i(x, y_j)\, x^{-(\gamma+1)}\exp\left[-\frac{1}{2}\left(\frac{y_j - x}{\sigma}\right)^2\right],$$

where, for example, $f_5(x, y_j) = (y_j - x)^2 \log x$ for the integrals $I_5(y_j)$. The corresponding general form of (23) is

$$\delta = \sigma\sqrt{-2\left[\log\epsilon + (\gamma+1)\log L - \log\kappa_j\right]}, \tag{24}$$

where

$$\kappa_j = \max_x f(x, y_j) \quad L \leq x \leq U.$$

Specifically, letting $d_j = \max(|y_j - L|, |y_j - U|)$,

$$\kappa_j = \begin{cases} \log U & \left[f_1(x, y_j) = \log x\right] \\ \log^2 U & \left[f_2(x, y_j) = \log^2 x\right] \\ d_j^2 & \left[f_3(x, y_j) = (y_j - x)^2\right] \\ d_j^4 & \left[f_4(x, y_j) = (y_j - x)^4\right] \\ d_j^2 \log U & \left[f_5(x, y_j) = (y_j - x)^2 \log x\right]. \end{cases} \tag{25}$$

Fig. 3 is an illustration, based on the parameter values $\gamma = 0.82$, $L = 0.45$, $U = 100.5$ and $\sigma = 0.16$ (see the example in Section 4.3). The figure contains plots of the integrands of $I(y)$, $I_1(y)$ and $I_3(y)$ as defined in (A1), calculated in the point $y = 50$. The inner pairs of filled triangles on the horizontal axes indicate points corresponding to $\epsilon = 10^{-6}$, and the outer triangles denote points corresponding to $\epsilon = 10^{-8}$. Computation time is considerably reduced by evaluating integrals only over the approximate range [49, 51], rather than

$[L, U] = [0.45, 100.5]$. Clearly, the smaller $\sigma/(U - L)$, the greater the time saved by this device.

Limiting the integration domain is particularly pertinent in the case of the non-truncated power law (7), for which the upper limit $U \to \infty$. Equation (23) for the integral in the expression (8) for the likelihood, i.e. $J_0(i)$ in (B1), still applies. Approximations for the remainder of the integrals in (B1) are not as easy, as the functions $f(x, y_j)$ are all monotonically increasing. Fortunately, the numerical determination of the interval over which the integrand is non-negligible is straightforward.

4 AN APPLICATION

In this section, the theory is applied to the estimation of the distributions of the masses of giant molecular clouds (GMCs) in M33 and the Large Magellanic Cloud (LMC), and to the mass distribution of H I clouds in the LMC. The intention is to demonstrate the application of the theory, rather than to derive definitive results. Therefore, for example, questions regarding the quality of published data are not addressed (Rosolowsky & Leroy 2006).

The importance of the mass spectra of GMCs has been emphasized in recent reviews by Rosolowsky (2005) and Blitz et al. (2007) to whom the interested reader is referred. A satisfactory power-law fit to GMC masses in M33 is demonstrated in Section 4.1, while problems with the fitting of similar observations in the LMC are shown in Section 4.3. In Section 4.2, mass spectra fitted to H I data from the LMC are discussed.

4.1 GMCs in M33

The data analysed in the subsection were taken from Engargiola et al. (2003). In particular, the 'corrected' masses, denoted $M^*_{\rm CO}$ are analysed. There are $N = 148$ GMCs in the catalogue.

Results of the distribution fitting are summarized in Table 1. The optimal model is the full distribution in equation (5), with the untruncated power law (7) a distant second. The estimated exponent $\widehat{\gamma} = 1.3$ is in good agreement with the Engargiola et al. (2003) estimate of 1.6 ± 0.3, but their lower mass limit $L = 4$ is rather different from the value 6.9 ± 0.7 obtained here.

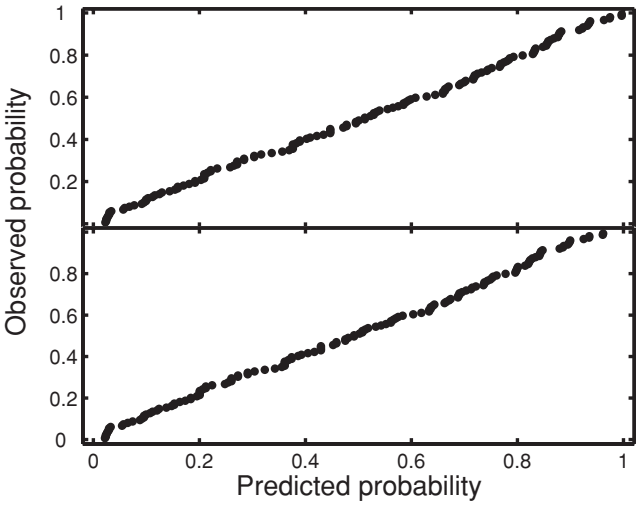

Figure 4. PP plots for the M33 data. The plot in the top panel is for the CDF (5) (i.e. a finite upper limit U) while the plot in the bottom is for the CDF (7) (infinite U).

Fig. 4 compares the fits of the two best distributional models. Focussing on the bottom panel ($U \to \infty$), it appears that there are slight deficiencies at both the low-mass (corresponding to small F_y) and large-mass (corresponding to large F_y) ends. The deficiency at the large-mass end implies that there are not enough large masses to justify the assumption that $U \to \infty$. The lack of very small masses – which is also visible in the top panel – is most easily explained as being due to a small data incompleteness at the low-mass end.

The two estimated covariance matrices (14) and (12) are

$$\mathbf{C}_I = \begin{bmatrix} 0.069 & 0.13 & 0.26 & 0.081 \\ 0.13 & 0.42 & 0.47 & 0.23 \\ 0.26 & 0.47 & 24.69 & 0.38 \\ 0.081 & 0.23 & 0.38 & 0.32 \end{bmatrix},$$

$$\mathbf{C}_J = \begin{bmatrix} 0.075 & 0.17 & -0.033 & 0.10 \\ 0.17 & 0.55 & 0.042 & 0.32 \\ -0.033 & 0.063 & 6.49 & 0.062 \\ 0.10 & 0.32 & 0.062 & 0.31 \end{bmatrix}. \quad (26)$$

Table 1. The results of fitting power-law distributions to the masses of GMCs in M33.

	Distribution model			
Equation number	Full (5)	$\sigma = 0, U$ finite (4)	$\sigma \neq 0, U \to \infty$ (7)	$\sigma = 0, U \to \infty$ (6)
	Model probabilities			
AIC	0.916	0	0.084	0
BIC	0.721	0	0.279	0
Number of bins	Significance level of goodness-of-fit statistic S			
$P = 10$	0.50	$p < 0.001$	0.50	$p < 0.001$
$P = 15$	0.66	$p < 0.001$	0.70	$p < 0.001$
$P = 20$	0.28	$p < 0.001$	0.08	$p < 0.001$
	Estimated parameters and standard errors for the full model			
	$\widehat{\gamma}$	$\widehat{L}$	$\widehat{U}$	$\widehat{\sigma}$
Estimate	1.33	6.94	77.7	3.48
Asymptotic S.E.	0.26	0.65	4.97	0.57
Jackknife S.E.	0.27	0.73	2.55	0.55

Note. The first part of the table compares the fitting results for the four different power-law forms discussed in Section 2, while the last few lines give the estimated parameters with associated standard errors for the optimal model. The unit of mass is $10^4\,{\rm M}_\odot$.

© 2009 The Authors. Journal compilation © 2009 RAS, MNRAS **397,** 495–505

The agreement is reasonable, except for covariances involving $\widehat{U}$. We speculate that this is due to the structure of the power-law fitting problem: many more data are required for the accurate estimation of U, since only the largest (and therefore most scanty) data determine the value of $\widehat{U}$. This means that large sample approximations (such as $\mathbf{C}_I$) of covariances of $\widehat{U}$ will be poorer than covariances involving other parameters. If this is correct, then the jackknife covariance matrix is preferred for data sets of this order.

Comparison of the computation times of $\mathbf{C}_I$ and $\mathbf{C}_J$ is also of interest: these were, respectively, 3 s and 1.4 h (Pentium 4 processor, clock speed 3.2 GHz). Of course, the computation of $\mathbf{C}_J$ could be speeded up considerably by relaxing the convergence criteria invoked when maximizing likelihoods, but it would still be orders of magnitude larger than the time required to calculate $\mathbf{C}_I$. For larger N, the computation time of $\mathbf{C}_I$ rises slowly (most of it expended on the calculation of the integrals in A1 or B1), while the computation of $\mathbf{C}_J$ becomes prohibitive for data set sizes of the order of a few hundred.

The estimated measurement error $\sigma = 0.56$ is bound to be a severe underestimate of the errors in the largest clouds. Since there are many more small clouds than large, it seems safe to assume that $\widehat{\sigma}$ will be determined primarily by masses close to L. The model for the measurement errors is obviously very restrictive, a point which is further discussed in Section 5.

4.2 H I clouds in the LMC

Kim et al. (2007) presented a catalogue with three different sets of H I cloud masses, for three different brightness temperature thresholds. Acceptable models could only be fitted to the $T_b = 64$ K data, and those results are presented in Table 2. There are $N = 195$ masses.

The preferred model, according to both information criteria, is again the form (5), with (7) the second choice. Interestingly, there is a larger difference between the asymptotic and the jackknife standard error estimates than in Table 1, despite the fact that the data set is somewhat larger. A possible contributory factor is the extent of the high-mass tail of the LMC H I cloud distribution: the 190 lowest masses are in the interval 2.2–550.4 $\times$ $10^3\,\mathrm{M}_\odot$, while the remaining five masses range from 867 $\times$ 10^3 to 2913 $\times$ $10^3\,\mathrm{M}_\odot$.

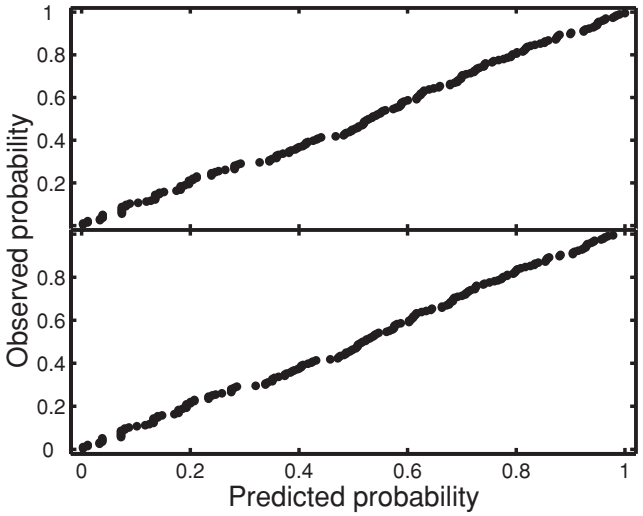

Figure 5. PP plots for the LMC H I data. The plot in the top panel is for the CDF (5) (i.e. a finite upper limit U) while the plot in the bottom is for the CDF (7) (infinite U).

The isolation of the largest mass value (the second largest mass is 1496 $\times$ $10^3\,\mathrm{M}_\odot$) accounts for the very large standard errors on U in Table 2. Removing the largest mass does not affect the estimates of γ and L by much ($\widehat{\gamma} = 0.60, \widehat{L} = 3.66$), but the estimated measurement error is increased to $\widehat{\sigma} = 0.98$, and $\widehat{U}$ is dramatically reduced to 1501 $\times$ $10^3\,\mathrm{M}_\odot$.

The estimated power-law exponent $\widehat{\gamma} = 0.55$ is in reasonable agreement with the value 0.68 found by Kim et al. (2007) for the same data set. The PP plots (Fig. 5) show no evidence of data incompleteness: this is in agreement with the very small completeness limit of 14.6 $\mathrm{M}_\odot$ estimated by Kim et al. (2007). Note also that the discrepancy between the observed probabilities and those predicted by the $U \to \infty$ model (bottom panel) is small. Both finite and infinite U have been claimed in earlier analyses of GMC data for various galaxies; in fact, different types of distribution have been found to

Table 2. The results of fitting power-law distributions to the masses of H I clouds in the LMC (threshold temperature 64 K).

	Distribution model			
	Full	$\sigma = 0, U$ finite	$\sigma \neq 0, U \to \infty$	$\sigma = 0, U \to \infty$
Equation number	(5)	(4)	(7)	(6)
	Model probabilities			
AIC	0.938	0	0.062	0
BIC	0.755	0	0.245	0
Number of bins	Significance level of goodness-of-fit statistic S			
$P = 10$	0.30	$p < 0.001$	0.40	$p < 0.001$
$P = 15$	0.60	$p < 0.001$	0.67	$p < 0.001$
$P = 20$	0.63	$p < 0.001$	0.28	$p < 0.001$
	Estimated parameters and standard errors for the full model			
	$\widehat{\gamma}$	$\widehat{L}$	$\widehat{U}$	$\widehat{\sigma}$
Estimate	0.55	3.47	2915.9	0.82
Asymptotic S.E.	0.054	0.20	16.0	0.18
Jackknife S.E.	0.053	0.19	1410	0.18

Note. The first part of the table compares the fitting results for the four different power-law forms discussed in Section 2, while the last few lines give the estimated parameters with associated standard errors for the optimal model. The unit of mass is $10^3\,\mathrm{M}_\odot$.

Table 3. The results of fitting power-law distributions to the masses of GMCs in the LMC.

	Distribution model			
Equation number	Full (5)	$\sigma = 0, U$ finite (4)	$\sigma \neq 0, U \to \infty$ (7)	$\sigma = 0, U \to \infty$ (6)
	Model probabilities			
AIC	0.781	0	0.219	0
BIC	0.398	0	0.602	0
	Estimated parameters of the two optimal models			
	$\widehat{\gamma}$	$\widehat{L}$	$\widehat{U}$	$\widehat{\sigma}$
Estimated full model	0.82	0.45	100.5	0.16
Estimated $U \to \infty$ model	0.88	0.46	∞	0.17

Note. The first part of the table compares the fitting results for the four different power-law forms discussed in Section 2, while the last few lines give the estimated parameters with associated standard errors for the two optimal models. The unit of mass is $10^5\,\mathrm{M}_\odot$.

apply to different parts of the same galaxy (e.g. Rosolowsky 2005; Rosolowsky et al. 2007). It would be interesting to see whether the $U \to \infty$ model would supplant the finite-U model if realistic measurement errors for the high-mass end of the distribution were introduced.

4.3 GMCs in the LMC

A feature of the catalogue of GMC masses in the LMC (Fukui et al. 2008) is that values are only given to one significant digit. There are consequently only 21 distinct values amongst the $N = 230$ masses in the catalogue, ranging from 0.1 to $100 \times 10^5\,\mathrm{M}_\odot$. Since the mass distribution is therefore effectively discrete, with, for example, 39 GMCs assigned masses of $1 \times 10^5\,\mathrm{M}_\odot$, it is not possible to sensibly compare the observed and predicted distributions using statistics such as those discussed in Section 2.3. Table 3 therefore only compares the performances of the different models, while Fig. 6 shows the PP plots for the two optimal models.

It is noteworthy that different models are selected by the AIC and BIC, respectively. It is generally accepted that the AIC is more liberal than the BIC in terms of the number of model parameters it allows, and the results in Table 3 conform to this expectation: the full model, selected by the AIC, is described by four parameters, while the BIC-selected model requires only three.

Fukui et al. (2008) found $\gamma = 0.75$, similar to the values in Table 3.

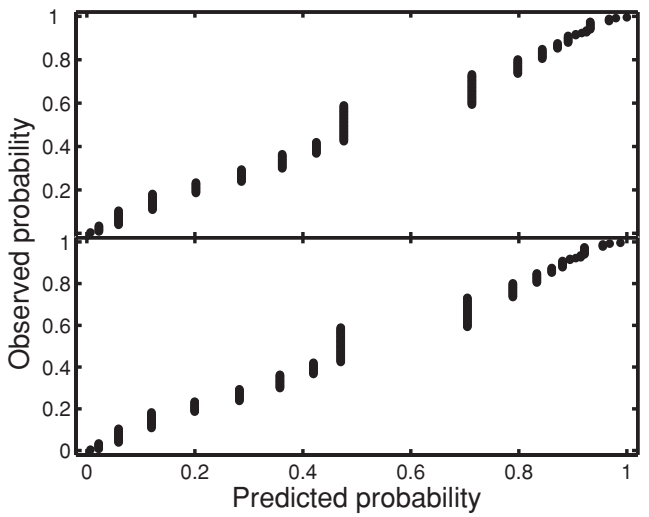

Figure 6. PP plots for the LMC GMC data. The plot in the top panel is for the CDF (5) (i.e. a finite upper limit U) while the plot in the bottom is for the CDF (7) (infinite U).

5 CONCLUSIONS

The comparison of the estimated parameters for models which ignore the measurement errors with parameters of the optimal models is instructive. In all three data sets analysed in Section 4, estimated power-law exponents are smaller if $\sigma = 0$ is assumed: 0.73 (1.33), 0.46 (0.55) and 0.49 (0.82) for the data discussed in Sections 4.1, 4.2 and 4.3, respectively (full model estimates are given in the brackets). Similarly for the lower limits L, estimates of 2.5 (6.9), 2.18 (3.47) and 0.20 (0.45) were found for the three data sets (optimal model estimates in brackets). Interestingly, the $\sigma \neq 0$ model estimates of U are larger than those obtained assuming no measurement error, although the difference is negligible in the last two data sets. For the M 33 data analysed in Section 4.1, $\widehat{U} = 77.7, 69.2$ for the models with and without measurement errors, respectively.

The differences between the $\sigma = 0$ and $\sigma \neq 0$ models can be quantified in terms of the estimated standard errors in Tables 1 and 2. It is 2.2, 5.9 and 3.3 for $\widehat{\gamma}$, $\widehat{L}$ and $\widehat{U}$ (M 33 data) and 1.7, 6.8 and 0.0 for $\widehat{\gamma}$, $\widehat{L}$ and $\widehat{U}$ (H I clouds in the LMC).

The examples in the previous section point to two extensions to the theory which are required in order to enhance the practical applicability of the theory in this paper. First, it seems very likely that in most settings measurement errors will depend on the true values of variables. For example, in the case of GMCs a model, such as

$$\sigma = a + bx$$

with a and b constant, seems reasonable (e.g. Rosolowsky 2005). This complicates the analysis, since σ can no longer be treated as a constant in (3) – it must be included in the integrand. Secondly, the issue of data completeness has been ignored in the analysis above. Mathematically, for incomplete data the PDF (2) is replaced by

$$f_x(x) = g(x)\frac{\gamma x^{-(\gamma+1)}}{L^{-\gamma} - U^{-\gamma}} \qquad L \leq x \leq U\ ,$$

where the function g, constrained by

$$\int_{-\infty}^{\infty} g(x)\,\mathrm{d}x = 1 \qquad 0 \leq g(x) \leq 1,$$

gives the fractional data completeness. An estimate of the function g can often be made by consideration of the properties of the experimental equipment and the measurement techniques.

The identification of incompleteness for data such as those considered in this paper is an interesting issue. Measurements errors cause the data to 'spill' from the interval $[L, U]$, particularly near the lower limit L where the probability density $f_x(x)$ is largest. This causes a tailing off of data with decreasing x ($x < L$), giving the impression of data incompleteness. The point is illustrated in Fig. 2, which shows histograms for simulated data with, and without, measurement errors. The data in the bottom panel appear to be complete over [3, 6.5], and incomplete for smaller y. In actual fact, there is no incompleteness.

The determination of completeness limits is therefore not entirely straightforward. A brute force way of dealing with this is to select a conservative completeness interval, and to ignore all data outside the interval. The price paid is that the analysis is more complicated – furthermore, if the completeness interval is too small, it may no longer be possible to determine L and/or U.

ACKNOWLEDGMENTS

The authors thank the referee for constructive comments which lead to an improved paper. LK was supported by postgraduate bursary grants from Square Kilometre Array (SKA) South Africa and the South African National Research Foundation (NRF), via the University of the Western Cape.

REFERENCES

Aban I. B., Meerschaert M. M., Panorska A., 2006, J. Am. Stat. Assoc., 101, 270
Andrews D. W. K., 2001, Econometrica, 9, 683
Bissell A. F., Ferguson R. A., 1975, The Statistician, 24, 79
Blitz L., Fukui Y., Kawamura A., Leroy A., Mizuno N., Rosolowsky E., 2007, in Reipurth B., Jewitt D., Keil K., eds, Protostars and Planets V. Univ. of Arizona Press, Tucson, p. 81
Brown P., Spalding R. E., ReVelle D. O., Tagliaferri E., Worden S. P., 2002, Nat, 420, 294
Burnham K. P., Anderson D. R., 2004, Sociol. Methods Res., 33, 261
D'Agostino R. B., Stephens M. A., eds., 1986, Goodness-of-Fit Techniques. Marcel Dekker, New York
Efron B., Tibshirani R. J., 1993, An Introduction to the Bootstrap. Chapman & Hall, London
Ellison S. L., Songaila A., Schaye J., Pettini M., 2000, AJ, 120, 1175
Engargiola G., Plambeck R. L., Rosolowsky E., Blitz L., 2003, ApJS, 149, 343
Fraser W. C., Kavelaars J. J., 2008, Icarus, 198, 452
Freeman M. P., Watkins N. W., Riley D. J., 2000, Phys. Rev. E, 62, 8794
Fukui Y. et al., 2008, ApJS, 178, 56
Gieles M., Larsen S. S., Bastian N., Stein I. T., 2006, A&A, 450, 129
Kim S. et al., 2007, ApJS, 171, 419
Kinkhabwala A., Thorsett S. E., 2000, ApJ, 535, 365
Kroupa P., 2001, in Deiters S., Fuchs B., Spurzem R., Just A., Wielen R., eds. ASP Conf. Ser. Vol. 228, Dynamics of Star Clusters and the Milky Way. Astron Soc. Pac., San Francisco, p. 187
Misgeld I., Hilker M., Mieske S., 2009, A&A, 496, 683
Mood A. M., Graybill F. A., Boes D. C., 1974, Introduction to the Theory of Statistics, 3rd edn. McGraw-Hill, Tokyo
Rice J. A., 2007, Mathematical Statistics and Data Analysis, 3rd edn. Thomson, Belmont, CA
Rosolowsky E., 2005, PASP, 117, 1403
Rosolowsky E., Leroy A., 2006, PASP, 118, 590
Rosolowsky E., Keto E., Matsushita S., Willner S. P., 2007, ApJ, 661, 830
van den Bergh S., 2006, AJ, 131, 304
Vicente S. M., Alves J., 2005, A&A, 441, 195

APPENDIX A: DERIVATIVES OF THE LOG-LIKELIHOOD FUNCTION, TRUNCATED POWER LAW

The following definitions are useful:

$$
\begin{aligned}
E(i,x) &= \exp -\frac{1}{2}\left(\frac{y_i - x}{\sigma}\right)^2,\\
I_0(i) &= \int_L^U x^{-(\gamma+1)} E(i,x)\,\mathrm{d}x,\\
I_1(i) &= \int_L^U x^{-(\gamma+1)} \log x\, E(i,x)\,\mathrm{d}x,\\
I_2(i) &= \int_L^U x^{-(\gamma+1)} (\log x)^2 E(i,x)\,\mathrm{d}x,\\
I_3(i) &= \int_L^U x^{-(\gamma+1)} (y_i - x)^2 E(i,x)\,\mathrm{d}x,\\
I_4(i) &= \int_L^U x^{-(\gamma+1)} (y_i - x)^4 E(i,x)\,\mathrm{d}x,\\
I_5(i) &= \int_L^U x^{-(\gamma+1)} (y_i - x)^2 \log x\, E(i,x)\,\mathrm{d}x.
\end{aligned}
\tag{A1}
$$

It is assumed throughout that the error variance σ^2 needs to be estimated: if it is known, then derivatives with respect to σ can be ignored. The first derivatives of the log-likelihood function in (8) are

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \gamma} &= \frac{N}{\gamma} - \frac{N}{L^{-\gamma} - U^{-\gamma}}\left(U^{-\gamma}\log U - L^{-\gamma}\log L\right) - \sum_i \frac{I_1(i)}{I_0(i)},\\
\frac{\partial \mathcal{L}}{\partial L} &= \frac{N\gamma L^{-(\gamma+1)}}{L^{-\gamma} - U^{-\gamma}} - L^{-(\gamma+1)} \sum_i \frac{E(i,L)}{I_0(i)},
\end{aligned}
$$

$$\frac{\partial\mathcal{L}}{\partial U} = -\frac{N\gamma U^{-(\gamma+1)}}{L^{-\gamma}-U^{-\gamma}} + U^{-(\gamma+1)}\sum_i \frac{E(i,U)}{I_0(i)},$$
$$\frac{\partial\mathcal{L}}{\partial\sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_i \frac{I_3(i)}{I_0(i)}. \tag{A2}$$

The second derivatives, required in the information matrix, are

$$\frac{\partial^2\mathcal{L}}{\partial\gamma^2} = -\frac{N}{\gamma^2} + \frac{NU^{-\gamma}L^{-\gamma}}{[L^{-\gamma}-U^{-\gamma}]^2}(\log U - \log L)^2 + \sum_i\left[\frac{I_2(i)}{I_0(i)} - \frac{I_1^2(i)}{I_0^2(i)}\right],$$
$$\frac{\partial^2\mathcal{L}}{\partial L^2} = \frac{N\gamma^2L^{-2(\gamma+1)}}{(L^{-\gamma}-U^{-\gamma})^2} - \frac{N\gamma(\gamma+1)L^{-(\gamma+2)}}{L^{-\gamma}-U^{-\gamma}}$$
$$- L^{-2(\gamma+1)}\sum_i\left[\frac{E(i,L)}{I_0(i)}\right]^2 - L^{-(\gamma+2)}\sum_i\frac{E(i,L)}{I_0(i)}\left[L\frac{y_i-L}{\sigma^2} - (\gamma+1)\right],$$
$$\frac{\partial^2\mathcal{L}}{\partial U^2} = \frac{N\gamma^2U^{-2(\gamma+1)}}{(L^{-\gamma}-U^{-\gamma})^2} + \frac{N\gamma(\gamma+1)U^{-(\gamma+2)}}{L^{-\gamma}-U^{-\gamma}}$$
$$- U^{-2(\gamma+1)}\sum_i\left[\frac{E(i,U)}{I_0(i)}\right]^2 + U^{-(\gamma+2)}\sum_i\frac{E(i,U)}{I_0(i)}\left[U\frac{y_i-U}{\sigma^2} - (\gamma+1)\right],$$
$$\frac{\partial^2\mathcal{L}}{\partial\sigma^2} = \frac{N}{\sigma^2} - \frac{1}{\sigma^6}\sum_i\left\{\left[\frac{I_3(i)}{I_0(i)}\right]^2 - \frac{I_4(i)}{I_0(i)}\right\} - \frac{3}{\sigma^4}\sum_i\frac{I_3(i)}{I_0(i)},$$
$$\frac{\partial^2\mathcal{L}}{\partial\gamma\partial L} = -\frac{N\gamma L^{-(\gamma+1)}}{[L^{-\gamma}-U^{-\gamma}]^2}(U^{-\gamma}\log U - L^{-\gamma}\log L) - \frac{NL^{-(\gamma+1)}}{L^{-\gamma}-U^{-\gamma}}(\gamma\log L - 1)$$
$$+ L^{-(\gamma+1)}\log L\sum_i\frac{E(i,L)}{I_0(i)} - L^{-(\gamma+1)}\sum_i\frac{I_1(i)E(i,L)}{I_0^2(i)},$$
$$\frac{\partial^2\mathcal{L}}{\partial\gamma\partial U} = \frac{N\gamma U^{-(\gamma+1)}}{[L^{-\gamma}-U^{-\gamma}]^2}(U^{-\gamma}\log U - L^{-\gamma}\log L) + \frac{NU^{-(\gamma+1)}}{L^{-\gamma}-U^{-\gamma}}(\gamma\log U - 1)$$
$$- U^{-(\gamma+1)}\log U\sum_i\frac{E(i,U)}{I_0(i)} + U^{-(\gamma+1)}\sum_i\frac{I_1(i)E(i,U)}{I_0^2(i)},$$
$$\frac{\partial^2\mathcal{L}}{\partial\gamma\partial\sigma} = \frac{1}{\sigma^3}\sum_i\left[\frac{I_1(i)I_3(i)}{I_0^2(i)} - \frac{I_5(i)}{I_0(i)}\right],$$
$$\frac{\partial^2\mathcal{L}}{\partial L\partial U} = -\frac{N\gamma^2L^{-(\gamma+1)}U^{-(\gamma+1)}}{(L^{-\gamma}-U^{-\gamma})^2} + L^{-(\gamma+1)}U^{-(\gamma+1)}\sum_i\frac{E(i,L)E(i,U)}{I_0^2(i)},$$
$$\frac{\partial^2\mathcal{L}}{\partial L\partial\sigma} = \frac{L^{-(\gamma+1)}}{\sigma^3}\sum_i\left[\frac{I_3(i)E(i,L)}{I_0^2(i)} - \frac{(y_i-L)^2E(i,L)}{I_0(i)}\right],$$
$$\frac{\partial^2\mathcal{L}}{\partial U\partial\sigma} = -\frac{U^{-(\gamma+1)}}{\sigma^3}\sum_i\left[\frac{I_3(i)E(i,U)}{I_0^2(i)} - \frac{(y_i-U)^2E(i,U)}{I_0(i)}\right]. \tag{A3}$$

Maximum-likelihood parameter estimates are obtained by setting the derivatives in (A2) equal to zero; the results

$$\sum_i\frac{I_1(i)}{I_0(i)} = \frac{N}{\gamma} - \frac{N}{L^{-\gamma}-U^{-\gamma}}\left(U^{-\gamma}\log U - L^{-\gamma}\log L\right)$$
$$\sum_i\frac{E(i,L)}{I_0(i)} = \frac{N\gamma}{L^{-\gamma}-U^{-\gamma}}$$
$$\sum_i\frac{E(i,U)}{I_0(i)} = -\frac{N\gamma}{L^{-\gamma}-U^{-\gamma}}$$
$$\sum_i\frac{I_3(i)}{I_0(i)} = N\sigma^2 \tag{A4}$$

follow. Note, though, that the last of equations (A4) only applies if σ is unspecified.

The relations (A4) can be used to simplify some of the expressions in (A3) a little:

$$\begin{aligned}
\frac{\partial^2\mathcal{L}}{\partial L^2} &= \frac{N\gamma^2 L^{-2(\gamma+1)}}{(L^{-\gamma}-U^{-\gamma})^2} - L^{-2(\gamma+1)}\sum_i\left[\frac{E(i,L)}{I_0(i)}\right]^2 - L^{-(\gamma+1)}\sum_i\frac{E(i,L)}{I_0(i)}\frac{(y_i-L)}{\sigma^2},\\
\frac{\partial^2\mathcal{L}}{\partial U^2} &= \frac{N\gamma^2 U^{-2(\gamma+1)}}{(L^{-\gamma}-U^{-\gamma})^2} - U^{-2(\gamma+1)}\sum_i\left[\frac{E(i,U)}{I_0(i)}\right]^2 + U^{-(\gamma+1)}\sum_i\frac{E(i,U)}{I_0(i)}\frac{y_i-U}{\sigma^2},\\
\frac{\partial^2\mathcal{L}}{\partial\gamma\partial L} &= -\frac{N\gamma L^{-(\gamma+1)}}{(L^{-\gamma}-U^{-\gamma})^2}(U^{-\gamma}\log U - L^{-\gamma}\log L) + \frac{NL^{-(\gamma+1)}}{L^{-\gamma}-U^{-\gamma}} - L^{-(\gamma+1)}\sum_i\frac{I_1(i)E(i,U)}{I_0^2(i)},\\
\frac{\partial^2\mathcal{L}}{\partial\gamma\partial U} &= \frac{N\gamma U^{-(\gamma+1)}}{(L^{-\gamma}-U^{-\gamma})^2}(U^{-\gamma}\log U - L^{-\gamma}\log L) - \frac{NU^{-(\gamma+1)}}{L^{-\gamma}-U^{-\gamma}} + U^{-(\gamma+1)}\sum_i\frac{I_1(i)E(i,L)}{I_0^2(i)},\\
\frac{\partial^2\mathcal{L}}{\partial\sigma^2} &= -\frac{2N}{\sigma^2} - \frac{1}{\sigma^6}\sum_i\left\{\left[\frac{I_3(i)}{I_0(i)}\right]^2 - \frac{I_4(i)}{I_0(i)}\right\}.
\end{aligned} \tag{A5}$$

APPENDIX B: DERIVATIVES OF THE LOG-LIKELIHOOD FUNCTION, UNTRUNCATED POWER LAW

The first of the following definitions is the same as in (A1); the remainder reflect the fact that the power-law extends to infinity:

$$\begin{aligned}
E(i,x) &= \exp -\frac{1}{2}\left(\frac{y_i-x}{\sigma}\right)^2,\\
J_0(i) &= \int_L^\infty x^{-(\gamma+1)}E(i,x)\,\mathrm{d}x,\\
J_1(i) &= \int_L^\infty x^{-(\gamma+1)}\log x\,E(i,x)\,\mathrm{d}x,\\
J_2(i) &= \int_L^\infty x^{-(\gamma+1)}(\log x)^2E(i,x)\,\mathrm{d}x,\\
J_3(i) &= \int_L^\infty x^{-(\gamma+1)}(y_i-x)^2E(i,x)\,\mathrm{d}x,\\
J_4(i) &= \int_L^\infty x^{-(\gamma+1)}(y_i-x)^4E(i,x)\,\mathrm{d}x,\\
J_5(i) &= \int_L^\infty x^{-(\gamma+1)}(y_i-x)^2\log x\,E(i,x)\,\mathrm{d}x.
\end{aligned} \tag{B1}$$

It is assumed throughout that the error variance σ^2 needs to be estimated: if it is known, then derivatives with respect to σ can be ignored. The first derivatives of the log-likelihood function in (8), with $U\to\infty$, are

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial\gamma} &= \frac{N}{\gamma} - N\log L - \sum_i\frac{J_1(i)}{J_0(i)},\\
\frac{\partial\mathcal{L}}{\partial L} &= \frac{N\gamma}{L} - L^{-(\gamma+1)}\sum_i\frac{E(i,L)}{J_0(i)},\\
\frac{\partial\mathcal{L}}{\partial\sigma} &= -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_i\frac{J_3(i)}{J_0(i)}.
\end{aligned} \tag{B2}$$

The second derivatives, required in the information matrix, are

$$\begin{aligned}
\frac{\partial^2\mathcal{L}}{\partial\gamma^2} &= -\frac{N}{\gamma^2} + \sum_i \left[\frac{J_2(i)}{J_0(i)} - \frac{J_1^2(i)}{J_0^2(i)}\right], \\
\frac{\partial^2\mathcal{L}}{\partial L^2} &= \frac{N\gamma^2}{L^2} - L^{-2(\gamma+1)} \sum_i \left[\frac{E(i,L)}{J_0(i)}\right]^2 - L^{-(\gamma+1)} \sum_i \frac{E(i,L)}{J_0(i)} \frac{(y_i - L)}{\sigma^2}, \\
\frac{\partial^2\mathcal{L}}{\partial\sigma^2} &= -\frac{2N}{\sigma^2} - \frac{1}{\sigma^6} \sum_i \left\{ \left[\frac{J_3(i)}{J_0(i)}\right]^2 - \frac{J_4(i)}{J_0(i)} \right\}, \\
\frac{\partial^2\mathcal{L}}{\partial\gamma\partial L} &= \frac{N}{L}(1 + \gamma \log L) - L^{-(\gamma+1)} \sum_i \frac{J_1(i)E(i,L)}{J_0^2(i)}, \\
\frac{\partial^2\mathcal{L}}{\partial\gamma\partial\sigma} &= \frac{1}{\sigma^3} \sum_i \left[\frac{J_1(i)J_3(i)}{J_0^2(i)} - \frac{J_5(i)}{J_0(i)}\right], \\
\frac{\partial^2\mathcal{L}}{\partial L\partial\sigma} &= \frac{L^{-(\gamma+1)}}{\sigma^3} \sum_i \left[\frac{J_3(i)E(i,L)}{J_0^2(i)} - \frac{(y_i - L)^2 E(i,L)}{J_0(i)}\right].
\end{aligned} \tag{B3}$$

Relations analogous to (A4) are

$$\begin{aligned}
\sum_i \frac{J_1(i)}{J_0(i)} &= \frac{N}{\gamma} + N \log L, \\
\sum_i \frac{E(i,L)}{J_0(i)} &= N\gamma L^\gamma, \\
\sum_i \frac{J_3(i)}{J_0(i)} &= N\sigma^2.
\end{aligned} \tag{B4}$$

Note that the last of equation (B4) only applies if σ is unspecified. These equations were used to simplify (B3).

This paper has been typeset from a TEX/LATEX file prepared by the author.

© 2009 The Authors. Journal compilation © 2009 RAS, MNRAS **397,** 495–505