Fractals, Vol. 9, No. 2 (2001) 209–222

UPPER-TRUNCATED POWER LAW DISTRIBUTIONS

STEPHEN M. BURROUGHS* and SARAH F. TEBBENS
College of Marine Science, University of South Florida, St. Petersburg, Florida 33701, USA
**Department of Chemistry and Physics, The University of Tampa, Tampa, Florida 33606, USA*

Received October 12, 2000; Accepted November 20, 2000

Abstract

Power law cumulative number-size distributions are widely used to describe the scaling properties of data sets and to establish scale invariance. We derive the relationships between the scaling exponents of non-cumulative and cumulative number-size distributions for linearly binned and logarithmically binned data. Cumulative number-size distributions for data sets of many natural phenomena exhibit a "fall-off" from a power law at the largest object sizes. Previous work has often either ignored the fall-off region or described this region with a different function. We demonstrate that when a data set is abruptly truncated at large object size, fall-off from a power law is expected for the cumulative distribution. Functions to describe this fall-off are derived for both linearly and logarithmically binned data. These functions lead to a generalized function, the upper-truncated power law, that is independent of binning method. Fitting the upper-truncated power law to a cumulative number-size distribution determines the parameters of the power law, thus providing the scaling exponent of the data. Unlike previous approaches that employ alternate functions to describe the fall-off region, an upper-truncated power law describes the data set, including the fall-off, with a single function.

1. INTRODUCTION

Cumulative number-size distributions have been employed in a wide variety of applications including seismology,[1] forest fire area,[2] and fault lengths[3] and offsets.[4] When a cumulative number-size distribution of data follows a power law, the data set is often considered fractal since both power laws and fractals are scale-invariant. A number of factors should be considered when fitting a function to a cumulative number-size distribution, such as whether the data has been binned and, if binned, where to plot the binned values. Interpretation of the results takes further consideration, especially if there is a deviation from a power law (straight line on a log-log plot) at either the upper or lower limits of the data.[5–7]

In Sec. 2, we present the continuous non-cumulative and cumulative power law functions. In Sec. 3, we derive the functions that describe discrete power law distributions that have been linearly and logarithmically binned. We demonstrate that the scaling exponent of a cumulative number-size distribution of *logarithmically* binned data is equal to the scaling exponent of the non-cumulative number-size distribution of the same data. In contrast, the scaling exponent of a cumulative number-size distribution of *linearly* binned data differs by one from the scaling exponent of the non-cumulative number-size distribution for the same data. In Sec. 3, we also derive the functions for cumulative power law distributions that are abruptly truncated at large object size. In Sec. 4, we present the function that describes the cumulative distribution independent of binning method, the upper-truncated power law. In Sec. 5, we discuss how to plot binned data as non-cumulative number-size distributions. In Sec. 6, we apply the upper-truncated power law to a synthetic data set. In the final sections, we discuss the results and summarize our conclusions.

2. CONTINUOUS NUMBER-SIZE FUNCTIONS

Non-cumulative and cumulative number-size functions are the basis of all equations used in this work and are defined below. The terms *power function* and *power law* are used interchangeably and refer to continuous functions. The term *distribution* is used to describe a set of discrete objects. *Objects* are the items measured to construct a data set, such as earthquakes, faults and forest fires. Each measured object has a characteristic *size* such as earthquake *magnitude*, fault *offset length*, or forest fire *area*. *Non-cumulative* distributions include only the number of objects within each data bin. These distributions sometimes are referred to as density distributions and may be plotted as histograms. *Cumulative* distributions include the number of objects within each data bin plus all larger objects. Cumulative number-size distributions sometimes are plotted as rank-order distributions or Zipf plots.[8]

2.1 Non-Cumulative Number-Size Functions

For a set of data points where the number-size distribution follows a power law, the number of objects, $n(r)$, with characteristic size r, is

$$n(r) = cr^{-d}\,. \tag{1}$$

On a log-log plot of $n(r)$ vs. r, this equation is a straight line where $-d$ equals the slope and c is a constant equal to the number of objects with size $r = 1$.

2.2 Cumulative Number-Size Functions

Cumulative number-size functions are commonly used in fractal analysis of data sets to find the fragmentation dimension. The number-size distribution for a large number of objects may be fractal if the cumulative number of objects, $N(r)$, with characteristic size greater than or equal to r satisfies the power law relation

$$N(r) = Cr^{-D}\,. \tag{2}$$

Equation (2) is a straight line on a log-log plot, where $-D$ is the slope and C is a constant equal to the number of objects with size $r \geq 1$. This cumulative relation applies when r is a continuous set of values. The term D is the scaling exponent of the power law. When r represents a one-dimensional quantity, such as height, D is the fractal dimension of the distribution. If r represents a two-dimensional or three-dimensional quantity, such as area or volume, the fractal dimension is $2D$ or $3D$, respectively. Equations (1) and (2) are in a form consistent with the work of many authors.[9,10] In Sec. 3, we examine relationships between these equations for linearly and logarithmically binned data and the effects of upper truncation.

3. ANALYZING DISCRETE DATA

Discrete data sets are comprised of individual, non-continuous objects. To derive the cumulative distribution from the non-cumulative distribution for a discrete data set, we must consider the binning interval used to record the object size. We will consider two types of binning: linear and logarithmic. Unless otherwise noted, all bins contain data.

3.1 Linear Binning Interval

For linearly binned data, each bin has the same width, Δr. This bin width may represent the width of a histogram bar. We will derive the equations that describe the non-cumulative and cumulative distributions of linearly binned data.

3.1.1 *Linear binning interval with discrete data*

If object sizes are sorted in rank order with r_i the ith largest value, then for any object size r,

$$r_i = r + i\Delta r\,. \tag{3}$$

The objects in Eq. (3) are discrete, so the sum of the objects will not be equivalent to the continuous function $N(r)$ in Eq. (2). We call the sum of discrete objects that are binned linearly $N_{\text{DLIN}}(r)$. When the sum is taken from any object size, r, to the maximum object size in the data set, we call this sum $N_{\text{DLIN}}^{(\max)}(r)$. $N_{\text{DLIN}}^{(\max)}(r)$ is calculated for any object size, r, by summing all individual objects of size greater than or equal to r, thus

$$N_{\text{DLIN}}^{(\max)}(r) = \sum_{i=0}^{\max} c r_i^{-d}\,. \tag{4}$$

The term $r_{\max}$ is the largest value of r in the data set (see Sec. 3.1.2). If we take the sum from any value of r to infinity, we call this sum $N_{\text{DLIN}}^{(\infty)}(r)$. Changing the upper limit of the sum in Eq. (4) to infinity and substituting Eq. (3) for r_i into Eq. (4) gives

$$N_{\text{DLIN}}^{(\infty)}(r) = \sum_{i=0}^{\infty} c(r + i\Delta r)^{-d}\,. \tag{5}$$

Equation (5) may be rewritten

$$N_{\text{DLIN}}^{(\infty)}(r) = cr^{-d} + \sum_{i=1}^{\infty} c(r + i\Delta r)^{-d}\,. \tag{6}$$

If Δr is small relative to the range of the values of r, then the sum in Eq. (6) may be approximated by the integral

$$\sum_{i=1}^{\infty} c(r + i\Delta r)^{-d} \approx \frac{c}{\Delta r}\int_r^{\infty} x^{-d}dx\,. \tag{7}$$

For the sum in Eq. (7) to converge to a finite value, we must consider only the case where $d > 1$. The right-hand-side of Eq. (7) may be evaluated as

$$\frac{c}{\Delta r}\int_r^{\infty} x^{-d}dx = \left.\frac{cx^{1-d}}{\Delta r(1-d)}\right|_r^{\infty} = \frac{cr^{1-d}}{\Delta r(d-1)}, \quad \text{for } d > 1\,. \tag{8}$$

Substituting the final term of Eq. (8) for the sum in Eq. (6), we have

$$N_{\text{DLIN}}^{(\infty)}(r) \approx cr^{-d} + \frac{cr^{1-d}}{(d-1)\Delta r}\,. \tag{9}$$

Factoring the right-hand-side of Eq. (9), we obtain

$$N_{\text{DLIN}}^{(\infty)}(r) \approx cr^{-d}\left(1 + \frac{r}{(d-1)\Delta r}\right). \tag{10}$$

If Δr is small relative to r, then $(\frac{r}{\Delta r})$ is much greater than 1, and

$$1 + \frac{r}{(d-1)\Delta r} \approx \frac{r}{(d-1)\Delta r}\,. \tag{11}$$

The error introduced by this approximation will be demonstrated graphically in Sec. 3.1.3. Substituting the approximation from Eq. (11) into Eq. (10), and re-arranging, we have

$$N_{\text{DLIN}}^{(\infty)}(r) \approx \left(\frac{cr^{1-d}}{(d-1)\Delta r}\right). \tag{12}$$

With the above approximations, the sum of individual objects approaches the continuous function of Eq. (2) so $N_{\text{DLIN}}^{(\infty)}(r) \approx N(r)$. Comparing Eq. (12), written in terms of c and d, to Eq. (2), written in terms of C and D, for linearly binned data, with Δr small relative to the range of values of r, we have

$$N(r) = \left(\frac{cr^{1-d}}{(d-1)\Delta r}\right) \tag{13}$$

where

$$C = \frac{c}{(d-1)\Delta r} \tag{14}$$

and

$$D = d - 1\,. \tag{15}$$

Equations (14) and (15) give the relationship between the constants and coefficients for the non-cumulative and cumulative distributions of the same linearly binned data. The cumulative distribution scaling exponent, D, differs by 1 from the non-cumulative distribution scaling exponent, d.

3.1.2 *Linear binning interval with discrete data truncated at large r*

When the cumulative power law is applied to individual, discrete objects, Eq. (2) predicts that an object size exists such that $N(r) = 1$. The object size for which the cumulative function equals 1, we call r_{N1}, so that $Cr_{\mathrm{N1}}^{-D} = 1$. When the largest object found in a data set, $r_{\max}$, is less than r_{N1}, we consider the data set to be upper-truncated.

We now consider a linearly binned upper-truncated data set. The function that describes the result of summing the discrete objects of Eq. (4) evaluated from r to $r_{\max}$, we call $M_{\mathrm{DLIN}}(r)$. If in Eq. (8) we integrate from r to $r_{\max}$, we have

$$M_{\mathrm{DLIN}}(r) = cr^{-d} + \frac{cr_{\max}^{1-d}}{(1-d)\Delta r} - \frac{cr^{1-d}}{(1-d)\Delta r}\,. \quad (16)$$

Equation (16) may be re-written as

$$M_{\mathrm{DLIN}}(r) = cr^{-d} + \frac{c}{(d-1)\Delta r}(r^{1-d} - r_{\max}^{1-d})\,. \quad (17)$$

Equation (17) is an approximation of the cumulative number of objects greater than or equal to size r, binned linearly, with a maximum observed value $r_{\max}$.

Equation (17) is a good approximation to Eq. (4), $N_{\mathrm{DLIN}}^{(\max)}(r)$, for values of r close to $r_{\max}$, as will be shown graphically in Sec. 3.1.3.

Substituting from Eqs. (14) and (15) into Eq. (17) yields

$$M_{\mathrm{DLIN}}(r) = CD\Delta r(r^{-(D+1)}) + C(r^{-D} - r_{\max}^{-D})\,. \quad (18)$$

Equation (18) describes a linearly binned cumulative number-size distribution with a maximum observed value for r. It is useful to consider one bin larger than $r_{\max}$, which we call r_{T}, so that $r_{\mathrm{T}} = r_{\max} + \Delta r$. Replacing $r_{\max}$ with $r_{\mathrm{T}} - \Delta r$, when Δr is small, Eq. (18) becomes

$$M_{\mathrm{DLIN}}(r) \approx C(r^{-D} - r_{\mathrm{T}}^{-D})\,. \quad (19)$$

Equation (19) describes a linearly binned power law distribution that is upper-truncated. Notice that $M_{\mathrm{DLIN}}(r_{\mathrm{T}}) = 0$. Thus, there are no objects of size r_{T} or larger.

Equation (19) represents the power law, Cr^{-D}, with the objects of size r_{T} and larger removed. Fitting Eq. (19) to a set of cumulative data points provides the coefficient, C, scaling exponent, D, and truncation term, r_{T}. Since there will probably be a scatter of the data points about the best fit of Eq. (19), the determined value of r_{T} may not be larger than the largest object size in the data set.

3.1.3 *Graphical illustration for linear binning interval*

We now graphically illustrate the functions derived in Sec. 3.1.2. Starting with a cumulative power function in the form of Eq. (2), we arbitrarily choose $C = 120$ and $D = 0.4$ so $N(r) = 120r^{-0.4}$ (top line in Fig. 1). To simulate linear binning, we sample this function by selecting object sizes in the range $0.1 \leq r \leq 3$ with $\Delta r = 0.1$ [see Eq. (3)]. The solid boxes along the top line in Fig. 1 represent the number of objects greater than or equal to each selected object size. With our chosen cumulative parameters C and D, and binning interval Δr, Eqs. (14) and (15) yield the non-cumulative parameters $c = 4.8$ and $d = 1.4$. Therefore Eq. (1) becomes $n(r) = 4.8r^{-1.4}$ (bottom line in Fig. 1). The solid diamonds on this line represent the number of objects of each selected size.

The cumulative distribution, $N_{\mathrm{DLIN}}^{(\max)}(r)$, is calculated for each value of r by adding together all objects of size greater than or equal to r in the chosen range $0.1 \leq r \leq 3$. This process is the sum in Eq. (4) with a maximum object size, $r_{\max}$, equal to 3. This cumulative distribution, $N_{\mathrm{DLIN}}^{(\max)}(r)$, is shown by solid triangles in Fig. 1. The function that approximates this cumulative distribution, $M_{\mathrm{DLIN}}(r)$ [Eq. (18)], is calculated from the chosen values for the cumulative parameters and plotted as open boxes in Fig. 1. $M_{\mathrm{DLIN}}(r)$ provides a good approximation to the cumulative distribution, $N_{\mathrm{DLIN}}^{(\max)}(r)$, when r is near $r_{\max}$. As predicted by the derivations in Sec. 3.1.2, the approximation is less accurate for smaller values of r. Notice the prominent fall-off of $N_{\mathrm{DLIN}}^{(\max)}(r)$ from a power law as r approaches $r_{\max}$ in the range $\frac{1}{2}\log r_{\max} < \log r \leq \log r_{\max}$.

This section demonstrates relationships between non-cumulative and cumulative linearly binned functions. Methods for applying these functions to discrete values in data sets will be presented in Sec. 5.

3.2 Logarithmic Binning Interval

Logarithmic binning means that bin width increases as the object size increases so that the ratio between successive bin widths is constant. Logarithmic binning is commonly used in many scientific disciplines, such as studies of earthquake magnitude[1] and sediment grain size.[11]

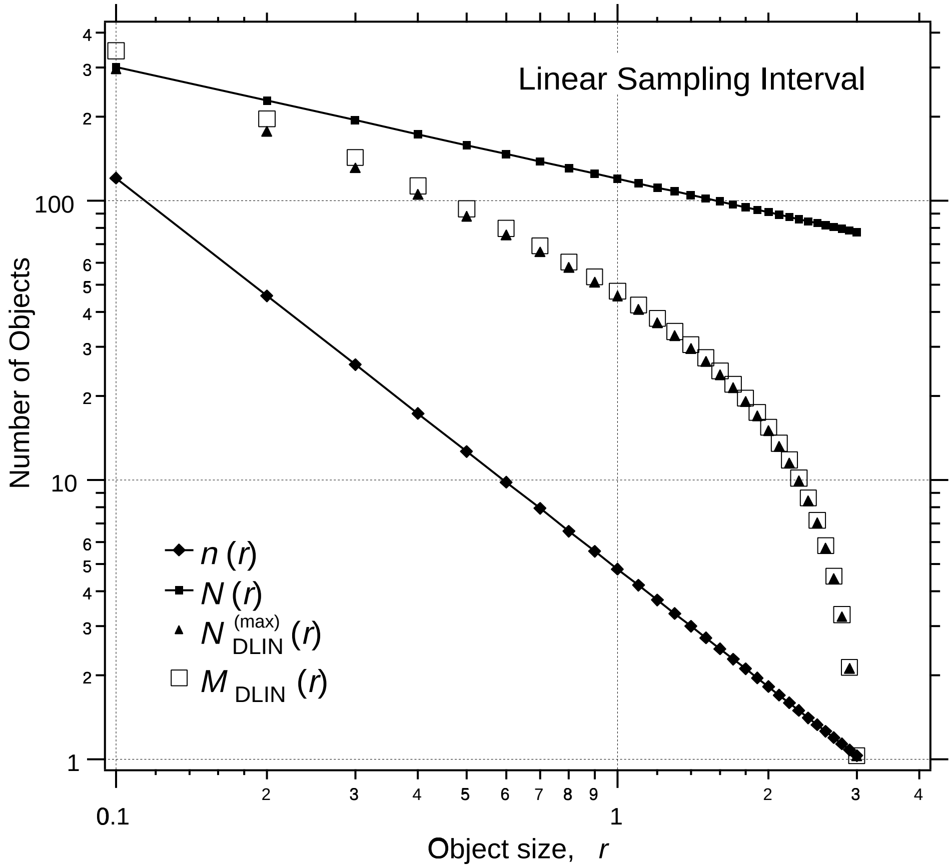


Fig. 1 Behavior of cumulative functions with linear binning. Note that this figure illustrates the behavior of functions, not data sets. In this example $N(r) = 120r^{-0.4}$, so $n(r) = 4.8r^{-1.4}$ and their slopes differ by 1. Values of r are chosen in the range $0.1 \le r \le 3$ at intervals of $\Delta r = 0.1$. Each plotted point represents the value of a function at these selected values of r. Evaluating $n(r)$ at each selected r and adding these values from each r to $r_{\max} = 3.0$ produces $N_{\mathrm{DLIN}}^{(\max)}(r)$. $M_{\mathrm{DLIN}}(r)$ is the approximation of $N_{\mathrm{DLIN}}^{(\max)}(r)$ as calculated from Eq. (18). $M_{\mathrm{DLIN}}(r)$ approximates $N_{\mathrm{DLIN}}^{(\max)}(r)$ very well when r is near $r_{\max}$. $N_{\mathrm{DLIN}}^{(\max)}(r)$ exhibits a significant fall-off from the straight line trend of a power law in the largest half-magnitude ($r > 1$ in this example). $N_{\mathrm{DLIN}}^{(\max)}(r)$ approximates a straight line below the largest half-magnitude ($r \le 1$ in this example), but the slope of $N_{\mathrm{DLIN}}^{(\max)}(r)$ in this region is steeper than the slope of the power law, $N(r)$. Attempts to fit a power law to $N_{\mathrm{DLIN}}^{(\max)}(r)$ in the nearly straight portion of the graph will not yield the correct exponent for the power law.

3.2.1 *Logarithmic binning interval with discrete data*

To compute the cumulative distribution for object size r with a logarithmic binning interval, let

$$r_j = r10^{ja} \tag{20}$$

where there are j measured sizes, r_j, greater than r, and where a is the $\log_{10}$ of the ratio of successive bin widths.

We will call the sum of discrete objects that are binned logarithmically $N_{\mathrm{DLOG}}(r)$. When the sum is taken from any object size, r, to the maximum object size in the data set, we will call this sum $N_{\mathrm{DLOG}}^{(\max)}(r)$. $N_{\mathrm{DLOG}}^{(\max)}(r)$ is calculated for any object size, r, by summing all discrete objects of size greater than or equal to r as follows

$$N_{\mathrm{DLOG}}^{(\max)}(r) = \sum_{j=0}^{\max} c r_j^{-d} \,. \tag{21}$$

As in the case of linear binning, $r_{\max}$ is the largest value of r in the data set. If we take the sum from any value of r to infinity, we call this sum $N_{\mathrm{DLOG}}^{(\infty)}(r)$. Changing the upper limit of the sum in Eq. (21) to infinity and substituting Eq. (20) for r_j into Eq. (21) gives

$$N_{\mathrm{DLOG}}^{(\infty)}(r) = \sum_{j=0}^{\infty} c(r10^{ja})^{-d} \,. \tag{22}$$

The right-hand-side of Eq. (22) is a geometric series. Evaluating this series gives

$$N_{\mathrm{DLOG}}^{(\infty)}(r) = \frac{cr^{-d}}{1-10^{-ad}}\,. \tag{23}$$

The right sides of Eqs. (23) and (2) have the same form, so for logarithmically binned data, $N_{\mathrm{DLOG}}^{(\infty)}(r) = N(r)$ where

$$C = \frac{c}{1-10^{-ad}} \tag{24}$$

and

$$D = d\,. \tag{25}$$

Note that $N(r)$ is not the integral of $n(r)$ for logarithmically binned data. The exponents of both the cumulative function $N(r)$ and the non-cumulative function $n(r)$ are the same, $d = D$.

3.2.2 *Logarithmic binning with discrete data truncated at large r*

Using an approach similar to that taken for linearly binned data (Sec. 3.1.2), we now consider a logarithmically binned data set, truncated at $r_{\max}$, where $r_{\max}$ is less than r_{N1}. We wish to find the function that describes the result of summing the discrete objects of Eq. (21). We will call this function $M_{\mathrm{DLOG}}(r)$.

Equation (23) evaluated at $r_{\max}$ is

$$N_{\mathrm{DLOG}}^{(\infty)}(r_{\max}) = \frac{cr_{\max}^{-d}}{1-10^{-ad}}\,. \tag{26}$$

Equation (26) represents the cumulative distribution from $r_{\max}$ to infinity, exclusive of $r_{\max}$. The cumulative distribution evaluated from r to infinity, minus the cumulative distribution evaluated from $r_{\max}$ to infinity, is obtained by subtracting Eq. (26) from Eq. (23). This difference excludes the value of $n(r_{\max})$. Adding to this difference the value of $n(r_{\max})$, where $n(r_{\max}) = cr_{\max}^{-d}$, produces the cumulative distribution from r to $r_{\max}$ inclusive. Following these steps, we obtain

$$M_{\mathrm{DLOG}}(r) = \left(\frac{c}{1-10^{-ad}}\right)(r^{-d} - r_{\max}^{-d}) + cr_{\max}^{-d}\,. \tag{27}$$

Re-writing Eq. (27) in terms of C and D from Eqs. (24) and (25), gives

$$M_{\mathrm{DLOG}}(r) = C(r^{-D} - r_{\max}^{-D}) + C(1-10^{-aD})r_{\max}^{-D}\,. \tag{28}$$

Equation (28) describes a logarithmically binned cumulative number-size distribution with a maximum value for r. For logarithmic binning, $r_{\max} = r_{\mathrm{T}}10^{-a}$, so Eq. (28) becomes

$$M_{\mathrm{DLOG}}(r) = C(r^{-D} - r_{\mathrm{T}}^{-D})\,. \tag{29}$$

Equation (29) describes a logarithmically binned power law distribution that is upper-truncated with no objects of size r_{T} or larger.

3.2.3 *Graphical illustration for logarithmic binning interval*

The equations derived above for logarithmically binned data will now be shown graphically. We start with the same cumulative power function we used for the linear binning example in Sec. 3.1.3, $N(r) = 120r^{-0.4}$. To simulate logarithmic binning, we choose a logarithmic sampling interval with $a = 0.1$. Equations (24) and (25) yield the values $c = 10.6$ and $d = 0.4$. The non-cumulative function is therefore $n(r) = 10.6r^{-0.4}$. This is not the non-cumulative function we obtained when the same cumulative function was binned linearly in Sec. 3.1.3. The non-cumulative function depends on the binning method. In Fig. 2, the non-cumulative function is plotted as the lower line (solid dots) and the cumulative function is plotted as the upper line (solid squares).

The cumulative distribution, $N_{\mathrm{DLOG}}^{(\max)}(r)$, is calculated for each value of r, using the approach in Sec. 3.1.3. Here we choose the range $0.1 \le r \le 3.2$ corresponding to $10^{-1.0} \le r \le 10^{0.5}$ with the exponent changing in steps of $a = 0.1$. This cumulative distribution is plotted as solid triangles in Fig. 2. The mathematical evaluation of the sum, $M_{\mathrm{DLOG}}(r)$, in the range $0.1 \le r \le 3.2$ is given by Eq. (28) and plotted as open boxes in Fig. 2 (using the chosen values of a, C, D, r, and $r_{\max}$). Since no approximations were made in evaluating the sum in Eq. (22), $N_{\mathrm{DLOG}}^{(\max)}(r)$ equals $M_{\mathrm{DLOG}}(r)$ and the solid triangles and open boxes plot at the same points in Fig. 2.

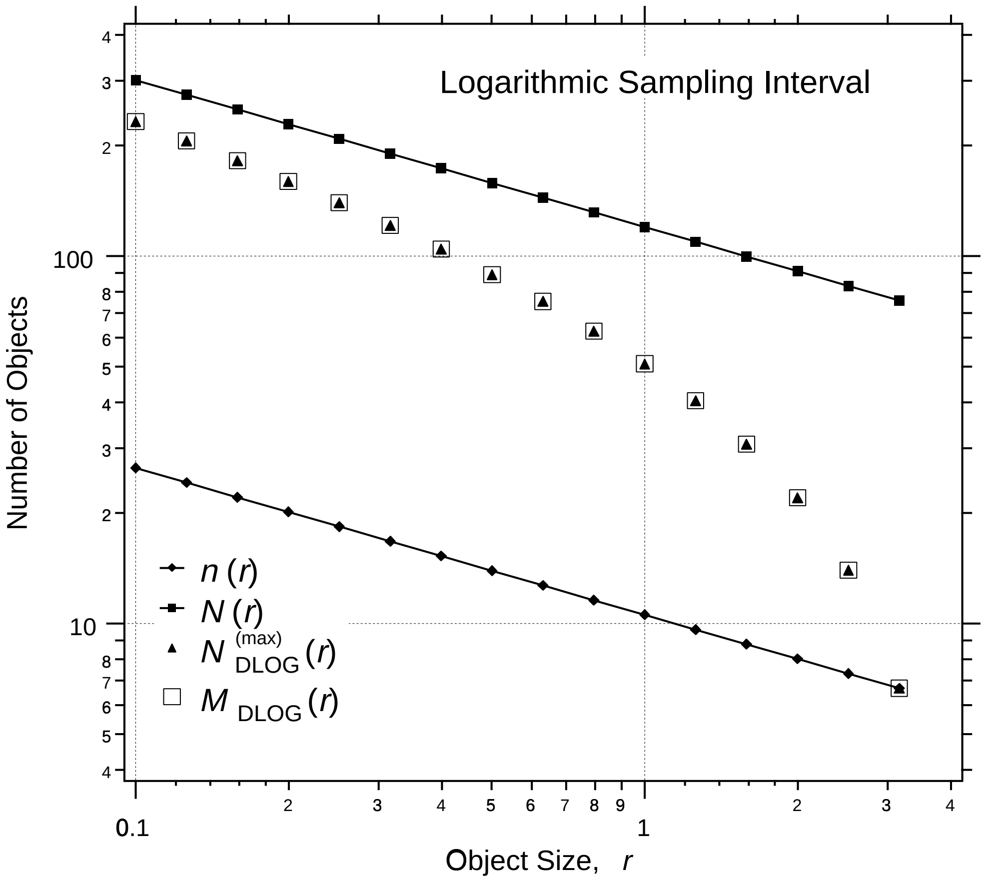


Fig. 2 Behavior of cumulative functions with logarithmic binning. In this example $N(r) = 120r^{-0.4}$, so $n(r) = 10.6r^{-0.4}$ and they have the same slope. Values of r are chosen in the range $10^{-1.0} \leq r \leq 10^{0.5}(0.1 \leq r \leq 3.2)$, with the exponent changing in steps of $a = 0.1$. Each plotted point represents the value of a function at these selected values of r. Evaluating $n(r)$ at each selected r and adding these values from each r to $r_{\max} = 3.2$ produces $N_{\mathrm{DLOG}}^{(\max)}(r)$. $M_{\mathrm{DLOG}}(r)$ gives the result of this summation as calculated from Eq. (28). $M_{\mathrm{DLOG}}(r)$ and $N_{\mathrm{DLOG}}^{(\max)}(r)$ plot at the same points since no approximations are made in deriving $M_{\mathrm{DLOG}}(r)$. $N_{\mathrm{DLOG}}^{(\max)}(r)$ exhibits a significant fall-off from the straight line trend of a power law in the largest half-magnitude ($r > 1$ in this example). $N_{\mathrm{DLOG}}^{(\max)}(r)$ approximates a straight line below the largest half-magnitude ($r \leq 1$ in this example), but the slope of $N_{\mathrm{DLOG}}^{(\max)}(r)$ in this region is steeper than the slope of the power law, $N(r)$. Attempts to fit a power law to $N_{\mathrm{DLOG}}^{(\max)}(r)$ in the nearly straight portion of the graph will not yield the correct exponent for the power law.

4. THE UPPER-TRUNCATED POWER LAW

When the bin width is small relative to the range of r, the cumulative distributions are the same for both linear binning [$M_{\mathrm{DLIN}}(r)$, Eq. (19)] and logarithmic binning [$M_{\mathrm{DLOG}}(r)$, Eq. (29)], since under these conditions $M_{\mathrm{DLIN}}(r) = M_{\mathrm{DLOG}}(r)$. We call this equation the upper-truncated power law, $M(r)$, where

$$M(r) = C(r^{-D} - r_{\mathrm{T}}^{-D})\,. \tag{30}$$

Equation (30) is appropriate for fitting a cumulative distribution governed by an underlying power law, regardless of the binning method. To find the values of C and D for the cumulative power function, we find the values of C, D, and r_{T} that provide the best fit of the upper-truncated power law [Eq. (30)] to a cumulative number-size distribution (see Sec. 5.2).

5. ANALYTICAL TECHNIQUES

5.1 Analysis of Synthetic Discrete Data

The points plotted on the cumulative and non-cumulative lines in Figs. 1 and 2 were chosen to illustrate the effect of the binning interval. The distribution of these points is not like that of any

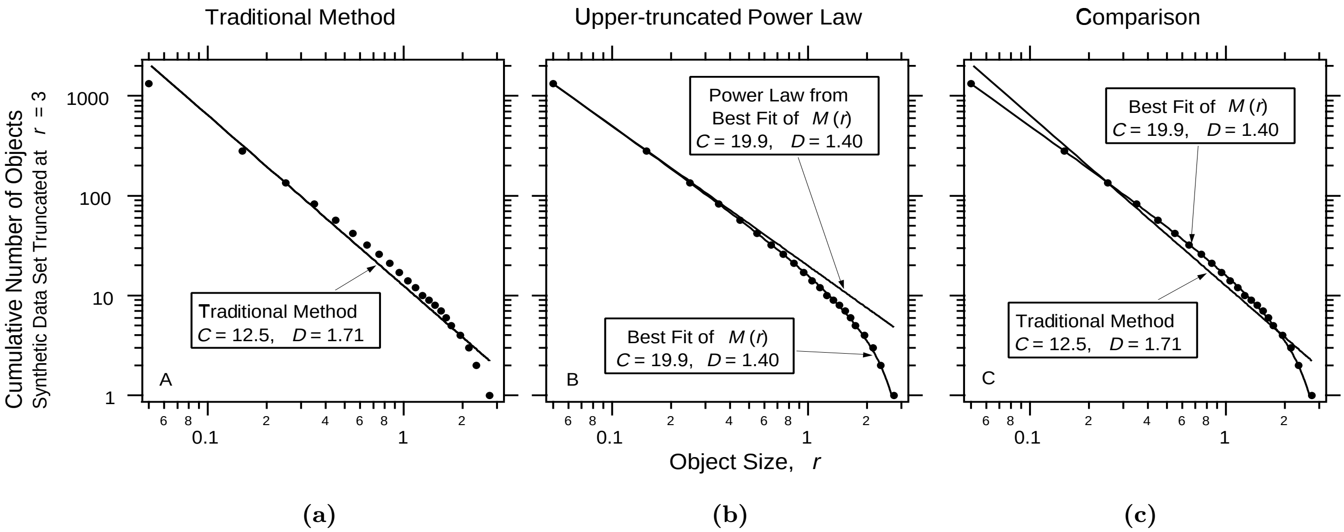


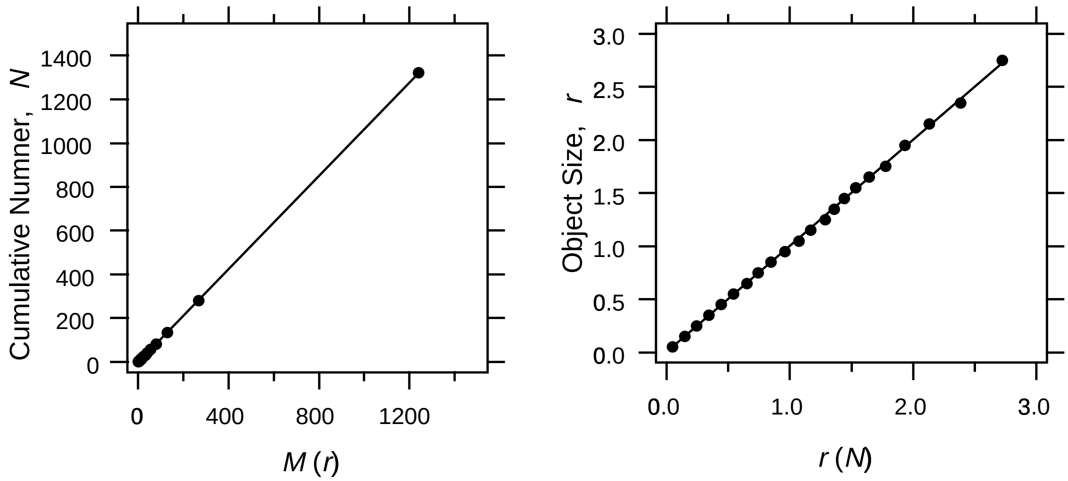
Fig. 3 Synthetic cumulative data with chosen values of $C = 20$ and $D = 1.4$ truncated at $r_{\max} = 3$ for linear binning with $\Delta r = 0.1$. Cumulative data points are plotted at the beginning of each data bin. **(a)** Traditional method. A power law fit to all data points provides incorrect values for C and D. **(b)** The upper-truncated power law. The upper-truncated power law fit to all data points produces $C = 19.9$ and $D = 1.4$, in good agreement with the chosen values. **(c)** Comparison of traditional and upper-truncated power law methods. The traditional method overestimates the value of D whereas the upper-truncated power law provides a better fit to the data and yields values of C and D in close agreement with the chosen values.

real data points we have observed for naturally occurring power distributions. For example, as the object size becomes large, natural data sets tend to have fewer points than shown in Figs. 1 and 2 and sometimes there are empty data bins.

To demonstrate the method of fitting an upper-truncated power law to a distribution, we create a synthetic data set that replicates characteristics similar to those observed for real discrete cumulative data. Starting with a cumulative power law, we calculate a set of values for r such that for each successive r, the cumulative function decreases by one. We arbitrarily choose $C = 20$, $D = 1.4$, and truncate the data set at a maximum object size of $r_{\max} = 3$ [Fig. 3(a)]. To create a synthetic cumulative data set, we bin the r values linearly with $\Delta r = 0.1$. To fit a power law to a cumulative data set, previous authors have fit a power function *directly* to some portion of the cumulative distribution. We call this approach the traditional method and show the results for the synthetic data in Fig. 3(a). The traditional method over-estimates the value of D in the cumulative power function. A more accurate approach is to fit the cumulative distribution with an upper-truncated power law. The best fit of $M(r)$ to the cumulative distribution is shown in Fig. 3(b). Figure 3(c) compares the results of this fitting method to the results of the traditional method. Fitting the distribution with an upper-truncated power law accurately determines C and D of the power law.

5.2 Fitting the Upper-Truncated Power Law to a Data Set

To find the best fit of the upper-truncated power law [Eq. (30)] to a data set, we need to simultaneously solve for the parameters C, D, and r_{T}. It is useful to plot a graph that is expected to yield a straight line so that least-squares fitting can be employed. There are two ways to do this. First, for each object size in the data set, the actual value of the cumulative number, N, is plotted against the cumulative number calculated from the upper-truncated power law, $M(r)$ [Fig. 4(a)]. If the data set follows an upper-truncated power law and the parameters are chosen correctly, the resulting graph will be a straight line of slope 1 and intercept 0. Adjusting the parameters of the upper-truncated power law to give the best fit to this straight line will yield the appropriate values for C, D and r_{T}. This procedure emphasizes the largest values of N, and therefore the smallest values of object size, when



(a) (b)

Fig. 4 Demonstration of how to fit the upper-truncated power law to a cumulative distribution. **(a)** For each object size in the data set, the actual value of the cumulative number, N, is plotted against the cumulative number calculated from the upper-truncated power law, $M(r)$ [Eq. (30)]. **(b)** For each value of N in the data set, the actual object size, r, is plotted against the calculated value $r(N)$ [Eq. (31)]. The parameters C, D, and r_T are adjusted to provide the best fit to a straight line of slope one and intercept zero for both **(a)** and **(b)**. In this case, using the data shown in Fig. 3, we find $C = 19.9$, $D = 1.4$, and $r_T = 2.96$ which is in close agreement with the chosen values.

determining the three parameters. This technique may not place enough emphasis on the largest objects and thus may fail to represent the fall-off. A second method places more emphasis on the largest objects by re-writing the upper-truncated power law as $r(N)$, where

$$r(N) = \left(\frac{N}{C} + r_T^{-D}\right)^{-\frac{1}{D}} . \qquad (31)$$

For each value of N in the data set, the actual object size, r, is plotted against the calculated value $r(N)$ [Fig. 4(b)]. This procedure yields a straight line of slope 1 and intercept 0 if the distribution follows an upper truncated power law and if the parameters are correctly chosen. Alternating between these two procedures will converge on the parameters of C, D and r_T that yield the best fit of the upper-truncated power law to all values in the data set.

5.3 Plotting Binned Data

The value of r chosen as the plotting location for binned data may significantly affect interpretation of the data. For cumulative distributions, each plotted point represents all values of the plotted size or larger. To correctly represent all the objects in the interval, the data should be plotted at the beginning (lower end) of the bin. In contrast, for non-cumulative linearly binned data, plotting the number of objects in a bin at the bin's center is a good approximation as long as the width of the bin is small relative to the object size (see Appendix A.1). For non-cumulative logarithmically binned data, the number of objects in a bin should be plotted at the beginning of the bin (see Appendix A.2). Failure to follow these plotting guidelines may lead to misinterpretation of the data.

6. COMPARISON TO OTHER METHODS

Several methods have been employed to analyze cumulative number-size distributions that exhibit a fall-off at large object size. Common approaches either ignore the fall-off and fit a power law to the entire distribution,[12] or exclude the largest event sizes where the fall-off is observed and fit a power law to the remaining smaller event sizes.[13,14] If the fall-off is caused by upper truncation, these two methods over-estimate the fractal dimension of the distribution (see Sec. 5.1).

Methods that account for the fall-off include: fitting different functions to different portions of the same distribution; fitting multiple power laws to different sections of the same distribution; differentiating the cumulative distribution to determine a non-cumulative distribution that is then analyzed;[2,15] and estimating the number of large objects missing from the upper end of the data set.[6,7] The first two methods suggest that different functions, or different forms of the same function, describe different portions of the distribution. If an upper-truncated power law is found to fit the entire distribution, then this single function provides a simpler interpretation than fitting multiple functions to the data set. The third method,[2,15] differentiating the cumulative distribution, has been shown to yield the same results as fitting an upper-truncated power law directly to the cumulative distribution for forest fire areas in the Australian Capital Territory.[16] To obtain each point in the non-cumulative distribution, the cumulative distribution is differentiated over a chosen range of points. The resulting non-cumulative distribution depends on the choice of range. The chosen range sometimes must be adjusted to use fewer points where the data are sparse, such as at large object size. An upper-truncated power law is simpler to apply, as it is fit directly to all the data points in the cumulative distribution and does not require choosing and adjusting a range of points for differentiation. The fourth method[6,7] accounts for the fall-off by estimating the number of objects missing above the upper end of the data set, called the censoring correction, then adding the censoring correction to the cumulative number associated with each object size in the data set. This process is repeated until fall-off is no longer observed in the resulting corrected cumulative distribution. Fitting a power law to this corrected distribution yields a value for D. However, the value for D can be obtained directly by fitting an upper-truncated power law, $M(r)$, to the uncorrected cumulative distribution. These two methods should produce the same results because the term Cr_{T}^{-D} in $M(r)$ [Eq. (30)] is equivalent to the censoring correction approximated by iteration. Fitting an upper-truncated power law directly to the cumulative distribution is a simpler approach to finding D.

7. CONCLUSIONS

Many cumulative number-size distributions of natural data exhibit fall-off from a power law at large object size. These distributions may be described with a single function, the upper truncated power law [Eq. (30)]. Fitting the upper-truncated power law to a cumulative number-size distribution yields the values C and D of the power law, and thus provides the scaling exponent that describes the data.

NOTATIONS

a	log of ratio between successive bin widths for logarithmically binned data
c	coefficient of non-cumulative power law
C	coefficient of cumulative power law
d	exponent of non-cumulative power law is $(-d)$
D	exponent of cumulative power law is $(-D)$
i	index for summing linearly binned objects to obtain the cumulative distribution
j	index for summing logarithmically binned objects to obtain the cumulative distribution
$M(r)$	Upper-truncated power law
$M_{\mathrm{DLIN}}(r)$	continuous function describing a cumulative number-size distribution of linearly binned data truncated at large object size
$M_{\mathrm{DLOG}}(r)$	continuous function describing a cumulative number-size distribution of logarithmically binned data truncated at large object size
$n(r)$	continuous function describing a non-cumulative number-size distribution of data
$N(r)$	continuous function describing a cumulative number-size distribution of data
$N_{\mathrm{DLIN}}(r)$	cumulative number-size distribution of discrete, linearly binned object sizes
$N_{\mathrm{DLIN}}^{(\infty)}(r)$	cumulative number-size distribution for linearly binned sizes found by counting to a very large object size
$N_{\mathrm{DLIN}}^{(\max)}(r)$	cumulative number-size distribution for linearly binned sizes found by counting to the largest object size in the data set
$N_{\mathrm{DLOG}}(r)$	cumulative number-size distribution of discrete, logarithmically binned object sizes

$N_{\mathrm{DLOG}}^{(\infty)}(r)$	cumulative number-size distribution for logarithmically binned sizes found by counting to a very large object size
$N_{\mathrm{DLOG}}^{(\max)}(r)$	cumulative number-size distribution for logarithmically binned sizes found by counting to the largest object size in the data set
r	object size
Δr	sampling interval or bin width for linear binning
r_b	object size at the small end of a linear bin
r_c	object size at bin center
r_e	object size at the large end of a linear bin
r_s	object size at the small end of a logarithmic bin
r_f	object size at the large end of a logarithmic bin
$r_{\max}$	largest object size in a data set
r_{N1}	object size where the cumulative power law equals 1
r_{PLIN}	plotting point for linearly binned non-cumulative data
r_{PLOG}	plotting point for logarithmically binned non-cumulative data
r_{T}	truncation object size, in an upper-truncated distribution there are no objects of this size or larger

ACKNOWLEDGMENTS

Kevin Dove and Emilio Toro of the University of Tampa provided invaluable help reviewing the mathematical analysis. This manuscript benefited from discussions with Robert Byrne and Chris Barton. SB was supported in part by a Knight Research Fellowship awarded by the USF College of Marine Science.

REFERENCES

1. B. Gutenberg and C. F. Richter, *Seismicity of the Earth and Associated Phenomena* (Princeton University Press, Princeton, 1949).
2. B. D. Malamud, G. Morein and D. L. Turcotte, *Science* **281**, 1840–1842 (1998).
3. C. H. Scholz and J. C. Contreras, *Geology* **26**, 967–970 (1998).
4. T. Villemin, J. Angelier and C. Sunwoo, in *Fractals in the Earth Sciences*, eds. C. Barton and P. R. LaPointe (Plenum Press, New York, 1995), pp. 205–226.
5. B. A. Brooks and R. W. Allmendinger, *J. Geophys. Res.* **101**, 13633–13653 (1996).
6. G. Pickering, J. M. Bull, D. J. Sanderson and P. V. Harrison, in *Fractals and Dynamic Systems in Geoscience*, ed. J. H. Kruhl (Springer-Verlag, Berlin, 1994), pp. 105–118.
7. G. Pickering, J. M. Bull and D. J. Sanderson, *Tectonophysics* **248**, 1–20 (1995).
8. B. B. Mandelbrot, *Fractals and Scaling in Finance* (Springer, New York, 1997), and references therein.
9. D. L. Turcotte, *Fractals and Chaos in Geology and Geophysics*, 2nd ed. (Cambridge Press, Cambridge, 1997).
10. J. Feder, *Fractals* (Plenum Press, New York, 1988).
11. C. K. Wentworth, *J. Geology* **27**, 507–521 (1922).
12. M. Matsushita, *J. Phys. Soc. Jpn.* **54**, 857–860 (1985).
13. P. Jackson and D. J. Sanderson, *Tectonophysics* **210**, 179–190 (1992).
14. A. Sornette, P. Davy and D. Sornette, *J. Geophys. Res.* **98**, 12111–12139 (1993).
15. C. H. Scholz and P. A. Cowie, *Nature* **346**, 837–839 (1990).
16. S. M. Burroughs and S. F. Tebbens, *Pure Appl. Geophys.* (in press).

APPENDIX A

A.1 Plotting Non-Cumulative Linearly Binned Data

When analyzing and plotting linearly binned non-cumulative data, a single object size must be chosen for each bin to represent all objects within the bin. The choice of effective object size for each bin affects the results of any fitting equation. If objects are measured with a much better resolution than the width of the data bin, the average object size within each bin can be calculated and used as the effective size of all objects in the bin. If the resolution of measurement is comparable to the width of the data bin, the position of objects within the bin will not be known, so calculating an average size in the bin may not be possible. However, if the bin width is small (small Δr), the fractional change in object size within the bin will be small and the average object size within a bin will be close to the size at bin center. The proper effective size to use in plotting the non-cumulative number of objects in a bin can be found by setting the non-cumulative function equal to the number of objects in the bin.

The proper plotting location we are seeking for linear binning we call $r_{\rm PLIN}$. We use the symbols r_b and r_e to represent the small end and large end of a linear bin so $\Delta r = r_e - r_b$. The number of objects in a bin is the difference between the cumulative number at the beginning of the bin and the cumulative number at the end of the bin, or $N(r_b) - N(r_e)$. We wish to find $r_{\rm PLIN}$ such that

$$n(r_{\rm PLIN}) = N(r_b) - N(r_e)\,. \qquad \text{(A.1)}$$

Substituting the appropriate power functions into Eq. (A.1) yields

$$cr_{\rm PLIN}^{-d} = Cr_b^{-D} - Cr_e^{-D}\,. \qquad \text{(A.2)}$$

From Eqs. (14) and (15), Eq. (A.2) can be rewritten as

$$CD(\Delta r)r_{\rm PLIN}^{-(D+1)} = Cr_b^{-D} - Cr_e^{-D}\,. \qquad \text{(A.3)}$$

Solving for $r_{\rm PLIN}$ gives

$$r_{\rm PLIN} = \left(\frac{D(\Delta r)}{r_b^{-D} - r_e^{-D}}\right)^{\frac{1}{(D+1)}}\,. \qquad \text{(A.4)}$$

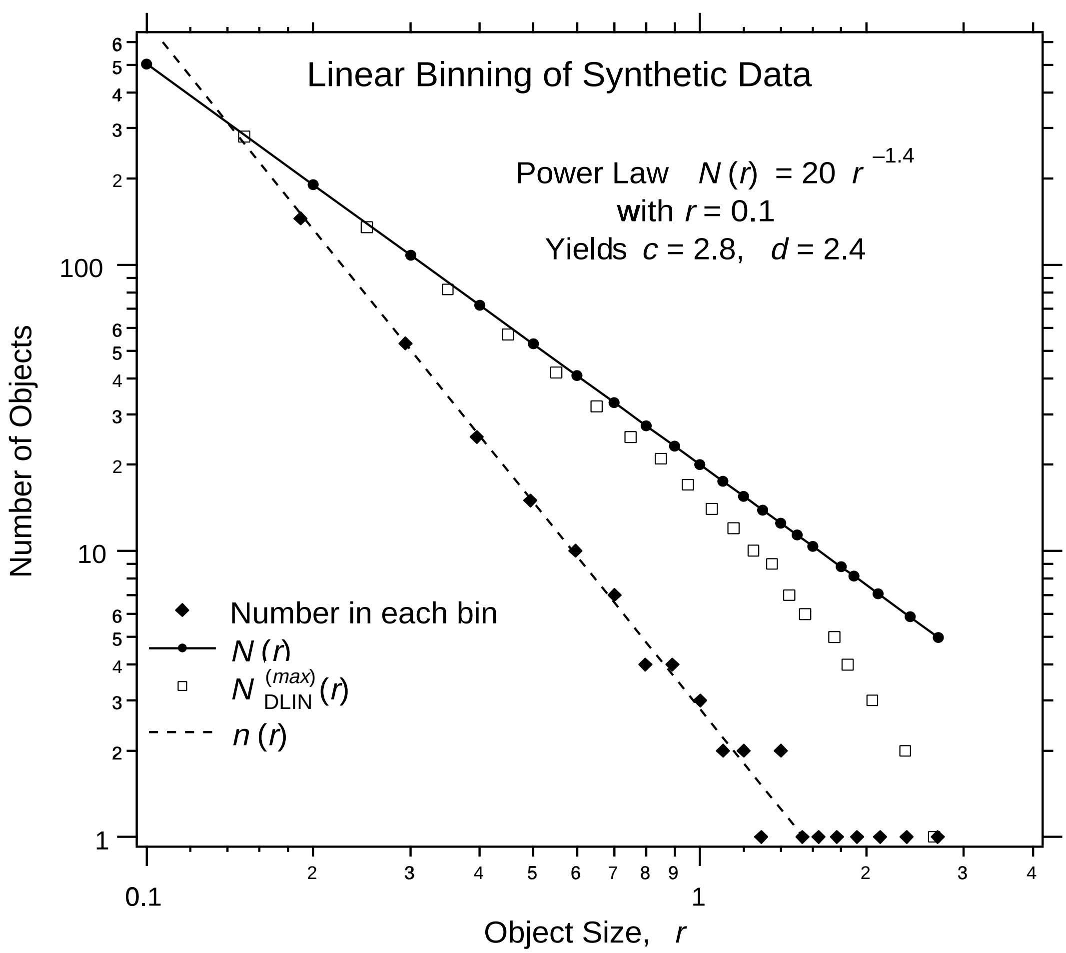


Fig. A.1 Example for linearly binned data. A synthetic data set is generated from a power law truncated at $r_{\max} = 3$ and linearly binned with $\Delta r = 0.1$. This yields the same cumulative distribution as shown in Fig. 3. Dots on the line representing the cumulative power function, $N(r)$, are located at the center of each occupied data bin (upper line). The spacing between these dots increases for large values of r, indicating that some bins contain no data. For this example, the cumulative number, $N_{\rm DLIN}^{(\max)}(r)$, is determined from the power law, $N(r)$ (see Sec. 5.1). For each data bin centered on r_c, the cumulative number, $N_{\rm DLIN}^{(\max)}(r)$, is plotted at the beginning of the bin. The non-cumulative number within each bin is plotted at the average of the r values in the bin. The parameters of $n(r)$ are obtained by solving Eqs. (14) and (15) using the chosen parameters for $N(r)$. Following these plotting guidelines, the non-cumulative distribution (solid diamonds) is well-represented by the function for $n(r)$ (dashed line).

If the bin width and exponent of the cumulative power function are known, Eq. (A.4) gives the object size for which the non-cumulative function is equal to the number of objects in the bin.

If the cumulative power law is not known, the approximate value of r_{PLIN} can still be found for small (but nonzero) bin width. If we call the center of the bin r_c, then $r_b = r_c - \frac{\Delta r}{2}$ and $r_e = r_c + \frac{\Delta r}{2}$, and the denominator of Eq. (A.4) becomes $(r_c - \frac{\Delta r}{2})^{-D} - (r_c + \frac{\Delta r}{2})^{-D}$. Expanding these terms, and ignoring second-order terms and higher (powers of Δr greater than one), we obtain the following approximation:

$$\left(r_c - \frac{\Delta r}{2}\right)^{-D} - \left(r_c + \frac{\Delta r}{2}\right)^{-D} \approx D(\Delta r) r_c^{-(D+1)} . \tag{A.5}$$

Equation (A.4) therefore becomes

$$r_{\text{PLIN}} = \left(\frac{D(\Delta r)}{r_b^{-D} - r_e^{-D}}\right)^{\frac{1}{(D+1)}} \approx \left(\frac{D(\Delta r)}{D(\Delta r) r_c^{-(D+1)}}\right)^{\frac{1}{(D+1)}} \tag{A.6}$$

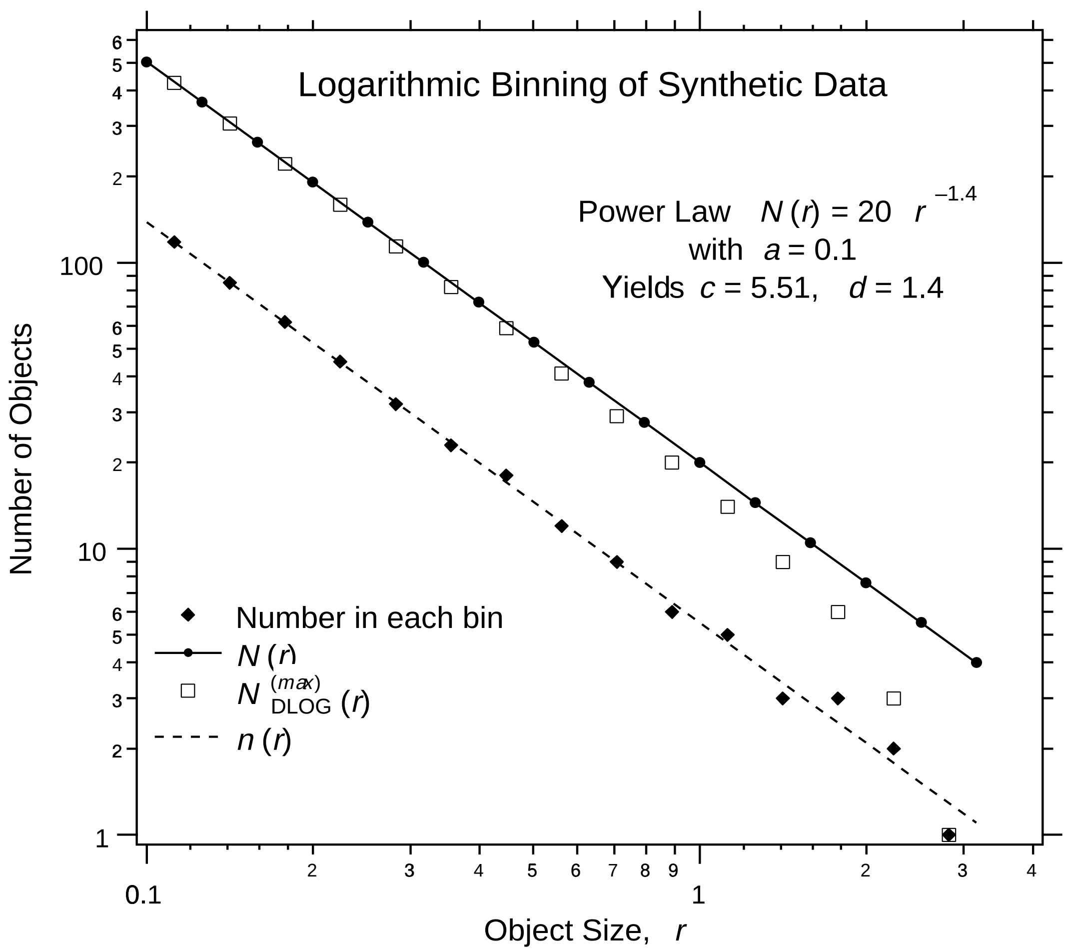


Fig. A.2 Example for logarithmically binned data. A synthetic data set is generated from a power law truncated at $r_{\max} = 3.2$ and logarithmically binned with $a = 0.1$. Dots on the graph of the cumulative power function, $N(r)$, indicate the center of each occupied data bin. In this example there are no unoccupied bins within the range of the data. For each data bin centered on r_c, the cumulative number, $N_{\text{DLOG}}^{(\max)}(r)$, is plotted at the beginning of the bin, not at r_c. The non-cumulative number for each bin is also plotted at the beginning of each bin since it can be shown that, for logarithmic binning, the number of points within each bin is equal to the value of the associated non-cumulative function at the start of the bin (see Appendix A.2). The parameters of $n(r)$ are obtained by solving Eqs. (24) and (25) using the chosen parameters for $N(r)$. Following these plotting guidelines, the non-cumulative distribution (solid diamonds) is well-represented by the function for $n(r)$ (dashed line).

or, for nonzero bin width ($\Delta r \neq 0$),

$$r_{\rm PLIN} \approx r_c\,. \tag{A.7}$$

When plotting non-cumulative linearly binned data, plotting the number of objects in a bin at the bin's center is a good approximation as long as the width of the bin is small relative to the size of objects in the bin. This approximation is not valid for two cases. First, at small object sizes, the width of the bin may not be small relative to the object size being measured. In this case, the fractional change in object size within a bin may be significant, producing a significant concentration of objects toward the small end of the bin. Second, when there are only a few objects in a bin, their sizes may be scattered about within the bin range and not necessarily have an average size near the bin's center. This may occur for both the largest and smallest objects in the data set. For small objects, the data set may be incomplete. For large objects, there may be only a few objects in the bin (or one object, or none at all). The largest and smallest object sizes should not be considered when fitting a power law to a non-cumulative distribution of data points (see Fig. A.1).

A.2 Plotting Non-Cumulative Logarithmically Binned Data

For logarithmically binned non-cumulative data, the choice of effective object size for each bin affects the coefficient of the function obtained when a power law is fit to the data. The proper object size to use for plotting logarithmically binned non-cumulative data we call $r_{\rm PLOG}$. We use r_c to represent the center of the bin in log space and the symbols r_s and r_f to represent the small end and large end of a bin. For logarithmic binning, $r_s = r_c 10^{-(\frac{a}{2})}$ and $r_f = r_c 10^{+(\frac{a}{2})}$, so we may write $r_f = r_s 10^a$. The number of objects in a bin is the difference between the cumulative number at the start of the bin and the cumulative number at the finish of the bin, or $N(r_s) - N(r_f)$. We wish to find the value of $r_{\rm PLOG}$ that makes the non-cumulative function evaluated at $r_{\rm PLOG}$ equal to the number of objects in the bin, so

$$n(r_{\rm PLOG}) = N(r_s) - N(r_f)\,. \tag{A.8}$$

Substituting the appropriate power functions into Eq. (A.8) yields

$$c r_{\rm PLOG}^{-d} = C r_s^{-D} - C r_f^{-D}\,. \tag{A.9}$$

From Eqs. (24) and (25) and the above relationships between r_s and r_f, Eq. (A.9) can be rewritten as

$$C(1 - 10^{-aD}) r_{\rm PLOG}^{-D} = C r_s^{-D} - C(r_s 10^a)^{-D} \tag{A.10}$$

so, for nonzero bin width ($a \neq 0$),

$$r_{\rm PLOG} = r_s\,. \tag{A.11}$$

For logarithmic binning, the number of objects in a bin equals the value of the non-cumulative function at the start of the bin (see Fig. A.2).