The lognormal distribution is not an appropriate null hypothesis for the species–abundance distribution

MARK WILLIAMSON* and KEVIN J. GASTON†

Department of Biology, University of York, York YO10 5DD, UK; and †Biodiversity and Macroecology Group, Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

Summary

1. Of the many models for species–abundance distributions (SADs), the lognormal has been the most popular and has been put forward as an appropriate null model for testing against theoretical SADs. In this paper we explore a number of reasons why the lognormal is not an appropriate null model, or indeed an appropriate model of any sort, for a SAD.

2. We use three empirical examples, based on published data sets, to illustrate features of SADs in general and of the lognormal in particular: the abundance of British breeding birds, the number of trees > 1 cm diameter at breast height (d.b.h.) on a 50 ha Panamanian plot, and the abundance of certain butterflies trapped at Jatun Sacha, Ecuador. The first two are complete enumerations and show left skew under logarithmic transformation, the third is an incomplete enumeration and shows right skew.

3. Fitting SADs by χ^2 test is less efficient and less informative than fitting probability plots. The left skewness of complete enumerations seems to arise from a lack of extremely abundant species rather than from a surplus of rare ones. One consequence is that the logit-normal, which stretches the right-hand end of the distribution, consistently gives a slightly better fit.

4. The central limit theorem predicts lognormality of abundances within species but not between them, and so is not a basis for the lognormal SAD. Niche breakage and population dynamical models can predict a lognormal SAD but equally can predict many other SADs.

5. The lognormal sits uncomfortably between distributions with infinite variance and the log-binomial. The latter removes the absurdity of the invisible highly abundant half of the individuals abundance curve predicted by the lognormal SAD. The veil line is a misunderstanding of the sampling properties of the SAD and fitting the Poisson lognormal is not satisfactory. A satisfactory SAD should have a thinner right-hand tail than the lognormal, as is observed empirically.

6. The SAD for logarithmic abundance cannot be Gaussian.

Key-words: Barro Colorado trees, British breeding birds, central limit theorem, Ecuadorian butterflies, individual abundance curves, log-binomial, logit-normal, lognormal, niche apportionment, skewness, species–abundance distributions, veil line.

Introduction

The numbers of individuals of each of several species in a sample, assemblage or community produces a species–abundance distribution (SAD). There have been many suggestions as to which statistical distribution fits a particular species–abundance distribution or a set

thereof (for reviews see May 1975; Magurran 1988, 2004; Tokeshi 1993; Marquet, Keymer & Cofré 2003). Most suggested distributions have little theoretical justification.

Hubbell (2001) has developed a well-known neutral theory that leads to a new distribution, the zero sum multinomial (ZSM). This is based on continuum vegetation (Williamson 2003), where every point is occupied and occupied by precisely one individual. In its basic form the Hubbell theory applies, even roughly, to

Correspondence: K. J. Gaston, Tel: 0114 2220030; Fax: 0114 2220002; E-mail: k.j.gaston@sheffield.ac.uk

rather few animal communities (such as some coral reefs). Nevertheless, it is a SAD derived from theory rather than from an empirical fit and so there is interest in seeing how well it performs compared with empirical SADs. So far, the ZSM has been compared only with the lognormal distribution (in which the logarithms of the abundances of the different species follow a normal (Gaussian) distribution), which has been put forward as an appropriate null model for a SAD. McGill (2003a) talks of 'reasonable null models' and 'the null lognormal hypothesis', Nee & Stone (2003) of '[the log-normal] an older, simpler null model' and justify that by the central limit theorem. Harte (2003), too, says 'A hand-waving justification for the lognormal distribution is that it could arise from the central limit theorem' and, going back a little, Taylor (1978) said 'Common logic suggests that the frequency distribution for N individuals (which vary logarithmically) in S species (which is a Poisson variate) … is a Log-normal'.

Our object in this paper is to explore a number of reasons why the lognormal is not an appropriate null model, or indeed an appropriate model of any sort, for a SAD. Others before us have objected to the lognormal (Lambshead & Platt 1985; Hughes 1986; Dewdney 2003) primarily on features of empirical distributions. We also start with examples which show why the lognormal is popular, and we then consider three issues. Under fitting, we discuss not only the results of statistical tests but also a new point, that there is a closely related distribution, the logit-normal, which often fits better. Under causes, we look at the central limit theorem and other derivations and provide a new argument against the use of the former for SADs. Under mathematical consequences, we discuss the canonical hypothesis and a new suggestion, the log-binomial, and argue that the lognormal cannot, on theoretical grounds, be a SAD. Our examples will often, for convenience, be drawn from our own publications; we are not attempting a comprehensive review.

Three examples

First, we present three rank abundance graphs, also known as dominance diversity graphs, and three histograms (Fig. 1). Note that the left-hand end of the histograms is the right-hand end of the rank abundance graphs; we shall use 'left-hand end' to refer to the histogram form.

The first example is of British breeding birds (Fig. 1a), long thought to have an archetypal lognormal distribution (Fisher 1952; Williams 1964, pp. 45–47), but certainly unusual because it uses counts (for the rarer species) and estimates (for the more commoner) of the total population and so, in that sense, is not a sample but a complete enumeration (Lambshead & Platt 1985). The data are from Gaston & Blackburn's (2000)
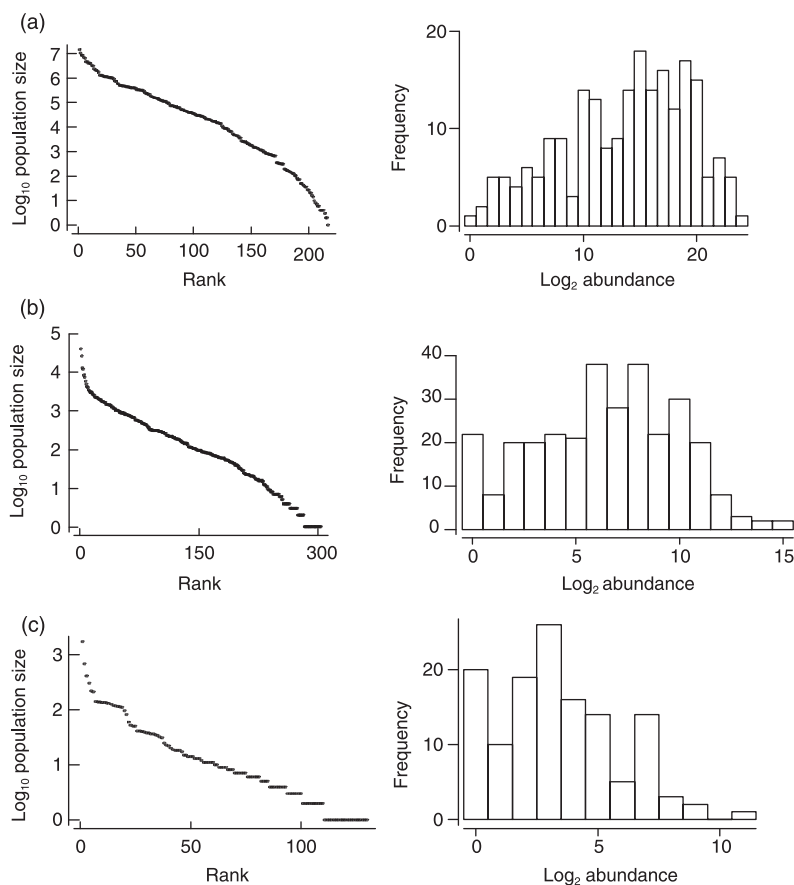

Fig. 1. Rank abundance plots and histograms for (a) British breeding birds (b) trees in the 50 ha plot on Barro Colorado Island, Panama, with d.b.h. > 1 cm and (c) butterflies from Jatun Sacha, Ecuador. Note the few singletons in (a) and the many in (c).

Appendix III, derived from Stone *et al.* (1997), which they regarded as 'the current British assemblage'. It comprises approximately 125 000 000 individuals. There are various other listings at other dates, and with different definitions of what constitutes this assemblage, that could have been used but the conclusions are the same.

Complete enumerations are rare. Our second example is another, an exact count of the trees greater than 1 cm diameter at breast height (d.b.h.) in a 50 ha plot on Barro Colorado Island (BCI), Panama in 1990 (Fig. 1b). Impressively, every such tree (243 704 of them) has been mapped to the nearest 10 cm (Condit 1998; Williamson 2003). There are three such publicly available surveys, differing only in their date (Center for Tropical Forest Science, Panama, http://ctfs.si.edu/datasets/BCIdata.html). Chave (2004) uses the average of these three, Plotkin & Muller-Landau (2002) another > 1 cm data set. Hubbell & Foster (1983) used trees with d.b.h. > 20 cm, McGill (2003a), Volkov *et al.* (2003) and Etienne & Olff (2004) all use enumerations of trees > 10 cm d.b.h. Hubbell & Foster (1983) say 'This [20 cm] lower cut-off is sufficiently small for all trees in the "canopy" to have been counted, at the risk of including some individuals of primarily subcanopy species. As is often the case, the canopy of the BCI forest is difficult to define rigorously.' Leigh (1999, p. 95) puts the BCI canopy at 35 m which, from well-known biophysical rules, implies a minimum d.b.h. of nearer 40 cm or the trees would buckle (McMahon & Bonner 1983). The 10 cm d.b.h. cut-off would seem to include many non-canopy trees, even though the canopy height in tropical forests is very variable (Leigh 1999), and so is a questionable set for testing Hubbell's theory and the ZSM. We prefer the > 1 cm set, using the latest such set publicly available, because it includes more individuals and so is more effective for examining the lognormal. It is, however, the least like continuum vegetation and so the least appropriate for testing the ZSM, but we introduce the ZSM only as it has been a recent prompt for declaring the lognormal an appropriate null hypothesis. The > 10 cm set can be found from Condit *et al.* (2002); there is yet another BCI data set, using biomass, from Condit, Hubbell & Foster (1996). The > 20 cm set seems not to be publicly available in electronic form.

Our third example is of the more common, sample, type of data set and is for a subset of butterflies from the Jatun Sacha Biological Station and Reserve, Napo Province, Ecuador (Fig. 1c; read from the graphs in DeVries, Murray & Lande 1997; see also Lande, Engen & Saether 2003). These are the trappable so-called fruit-feeding nymphalids, but we will refer to them as butterflies. It is a remarkably large sample of 6690 individuals, but a complete enumeration would include more species. This is often described as truncation, but this is misleading (Dewdney 1998). We will, to be neutral, call it incomplete enumeration. Incomplete enumerations are seldom random samples of anything definable. They may be systematic samples of some-

thing that can be specified, although environmental variation and differences in the detectability or observability of species may make that difficult.

As is common, all three rank abundance graphs in Fig. 1 are sigmoidal, with a central inflection, although the butterflies are scarcely so. The lognormal is also sigmoidal on such a plot, and sigmoidal rank abundance plots have been taken in the past as evidence of lognormality (e.g. Williamson 1972; McGowan & Walker 1993). In the butterflies, rather conspicuously, and the trees, to a lesser extent, the most common observation is of singletons, species represented by only one individual. This, too, is common (Pielou 1969), but contrary to Dewdney's (1998, 2000, 2003) claim it is far from universal [Fig. 1a; Williamson 1981 (Lepidoptera); McGowan & Walker (1993 (copepods and phytoplankton); Williamson 1987 (birds); Whittaker 1965 (Great Smoky plants); Whittaker 1969 (Brookhaven plants); Wilson 1991 (various plant communities)]. Incomplete enumeration leads to many singletons but, as the tree data show, the converse is not true.

Fitting the lognormal

Much of the early work on the lognormality of SADs simply plotted the data as a histogram and decided that they did (or did not) look normal. Williams (1964) used probit plots as well. Nowadays it is simple to compute better tests. As has often been said (e.g. McGill 2003c; Ulanowicz 2003), testing by fitting is not sufficient. If the data fail to fit a theoretical distribution, then that is evidence against the theory. However, if the data do fit, all that can be said is that the theory might be correct. Much more stringent tests come from considering the consequences of theory and seeing if those are matched by the data. We come to those consequences in the last section. Here, we make some points about fitting and some of the results that have been found.

METHODS OF FITTING

Nee & Stone (2003) say 'the lognormal is trivially easy to fit' but that is only true with the, rather rare, complete enumerations. With incomplete enumerations, as Gray (1987) warns, it is necessary to allow for the incompleteness. As will be seen, assuming a simple truncation, the standard method that has been used (O'Hara & Oksanen 2003) is not correct, but what is correct is still uncertain. Consequently, we will fit our three examples as if they were all complete enumerations even though only two of them are (or in the case of the British birds, is an estimated approximation). Even when the lognormal has been fitted, there is still an important choice of how to test the goodness of that fit.

The usual method of fitting has been to generate a histogram from the data, following Preston (1948), and then to use a χ^2 test on the deviation from expectation of the observed values for the bins (the classes into which species are grouped, e.g. McGill 2003a; Volkov

et al. 2003). This is weak. Each bin must have a sufficiently large expectation to make the χ^2 test valid. That loses information. The test also ignores any pattern in the deviations across the bins, again losing information. The situation is often made worse by following Preston's (1948) method of choosing the divisions between bins. This is described mathematically by Engen (1978) and used explicitly by Nee, Harvey & May (1991), Olszewski & Erwin (2004) and others. Preston used the powers of two, i.e. 1, 2, 4, 8, etc., as his bin boundaries, which involves splitting any observation that falls on a boundary, any species whose abundance is an exact power of two, between the two bins on either side. Half is put on each side, but the shape of the normal curve shows that that is not quite right. The effect is serious for the low numbers just mentioned; it is of no account at, for instance, 65 536. Some authors have avoided such splitting by moving the bin margins slightly. Magurran (1988, 2004) and others (see Lande *et al.* 2003) use 1–2 (or 1, 2), 3–4, 5–8, etc., while Gray (1987), Plotkin & Muller-Landau (2002), O'Hara & Oksanen (2003) and Chave (2004) use 1, 2–3, 4–7, etc. Both schemes mean that neither the bin boundaries nor their medians double exactly.

All these difficulties are overcome easily by using the powers of two as the centre of bins, not their edges. The boundaries then come at $2^n/2^{1/2}$ (or $2^{n-1/2}$), namely 0·71, 1·4, 2·8, 5·7, etc., avoiding all integers. This is used, without comment, by Hubbell (2001) and should generally be adopted when a histogram is wanted and is used in Fig. 1. It can be used equally for other powers, substituting any chosen number for 2. Using triplings rather than doublings, Williams (1964) and DeVries *et al.* (1997) used divisions at $3^n/2$, which is close to but less precisely geometric than $3^n/3^{1/2}$. Computer software may do this automatically, but just what it is doing should be watched carefully.

Rather than using histograms, it is better and relatively easy to use the individual observations of species abundance by producing a normal probability plot (Fig. 2). This is more efficient in the technical statistical sense, using all the information in the data. Normal probability plots are available in many statistical packages; the normal scores involved are usually available separately. Various tests are then available for these plots and indeed can be calculated without plotting. The ones we discuss below are Anderson–Darling, Shapiro–Wilks or the equivalent Ryan–Joiner, and Kolmogoroff–Smirnov. Lilliefor's probabilities are thought better with the Kolmogoroff–Smirnov test under some circumstances (STATISTICA 6·1, Statsoft, Inc.), but have not been used here.

RESULTS FROM FITTING

(i) General results

If the lognormal was a correct theory of assemblages, then the top two normal probability plots in Fig. 2 should show a good fit (lying along the straight line) being complete enumerations, but the third, the incom-
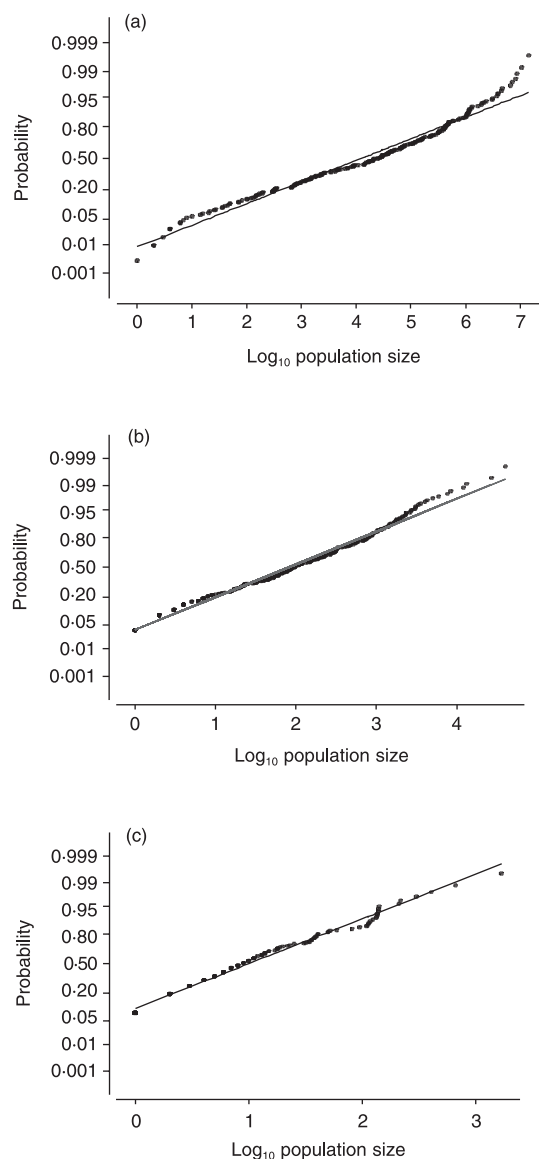

Fig. 2. Probability plots of the three rank abundance plots in Fig. 1. Note that in the top two cases the least good fit is at the right-hand end, not the left-hand one.

plete enumeration, should fit badly, particularly at low values at the left-hand end. In fact, the third plot seems to give the best fit. This is quantified in Table 1, which brings out the important point that different tests give different results. On the Anderson–Darling test all three data sets are non-normal. The Ryan–Joiner test, as so often in our experience, gives results that match what the eye sees, namely that the British birds are not normal while the Ecuadorian butterflies are. BCI trees are also just significantly not normal at $P \sim 0.03$. That changes again with the Kolmogoroff–Smirnov test, where only the British birds are not normal.

Another result that is clear to the eye in Fig. 2 but is not brought out by the tests is that the worst-fitting parts of the curves are at the right-hand ends of the British birds and BCI trees. In both cases, compared to the fitted line, there is a deficiency in the abundance of the extremely common species. The topmost points,

Table 1. Tests for fit to a lognormal of three species–abundance data sets, giving the number of species, probabilities for the Anderson–Darling, Ryan–Joiner (equivalent to Shapiro–Wilks test) and Kolmogoroff–Smirnov tests, and the correlation coefficient used in the Ryan–Joiner test

	No. spp.	Anderson–Darling test	Ryan–Joiner test	Kolmogoroff–Smirnov test	Correlation coefficient
British breeding birds	217	< 0·001	< 0·01	< 0·01	0·9839
BCI trees, > 1 cm 1990	304	< 0·001	0·033	> 0·15	0·9946
Ecuadorian butterflies	130	0·001	> 0·100	> 0·15	0·9947

the most abundant species, would have to move to the right, to be even more abundant, to lie on the fitted line. That brings us to the question of skewness.

(ii) Skewness

That the SAD for British birds is not a symmetrical curve, as it should be if it were lognormal, but is skewed to the left (left-skew, negative skew, longer tail to the left, shorter to the right) was first noted by Nee *et al.* (1991), omitting on rational grounds some of the species included here. This was confirmed by Gaston & Blackburn (2000) (this set) and by Gaston, Blackburn & Gregory (1997) and Hubbell (2001), with a set derived from Gibbons, Reid & Chapman (1993). Hubbell (2001) found also skewness in the BCI tree data. Gregory (2000), surveying the breeding bird data for 48 European regions or countries, found that 40 sets were left-skewed, one was symmetrical and seven right-skewed. He, as did Nee *et al.* (1991), used g_1, the standard third moment measure of skewness. The values of g_1 in our three examples are: British birds −0·435 ($\chi_1^2 = 6·938$, $P < 0·01$), BCI trees −0·138 ($\chi_1^2 = 0·976$, NS), Ecuadorian butterflies +0·464 ($\chi_1^2 = 4·772$, $P < 0·05$), all using test statistic 2 from http://www.xycoon.com. So our complete enumerations are left-skewed with the birds and the butterflies significantly skewed on this test.

As Nee *et al.* (1991) noted, Hutchinson (1967) had commented on this long ago, saying 'MacArthur (personal communication), however, has noted that the Preston curves, when sufficiently unveiled, generally have a somewhat asymmetrical form; if any value of the ordinate above which the entire curve is in the wholly positive quadrant is taken, the area to the left of the mode will be a little greater than the area to the right.' By taking areas above a nonzero ordinate value of a histogram, MacArthur neatly side-stepped the problem of detecting skewness in incomplete enumerations (so-called veiled data).

The usual description of this left skew (see Gaston & Blackburn 2000; Hubbell 2001) is to say that there is a surplus in the long left-hand tail. McGill (2003b) thinks this is a sampling effect, Magurran & Henderson (2003) think it comes from including casuals, and Gregory (2000) says 'driven by a very small number of vagrant breeders'. McGill's sampling effect seems special to his data, Magurran & Henderson have several definitions of casuals, all of which seem somewhat arbitrary, and Gregory's vagrant breeders include one endemic and several that seem definitely established. However, all that is probably
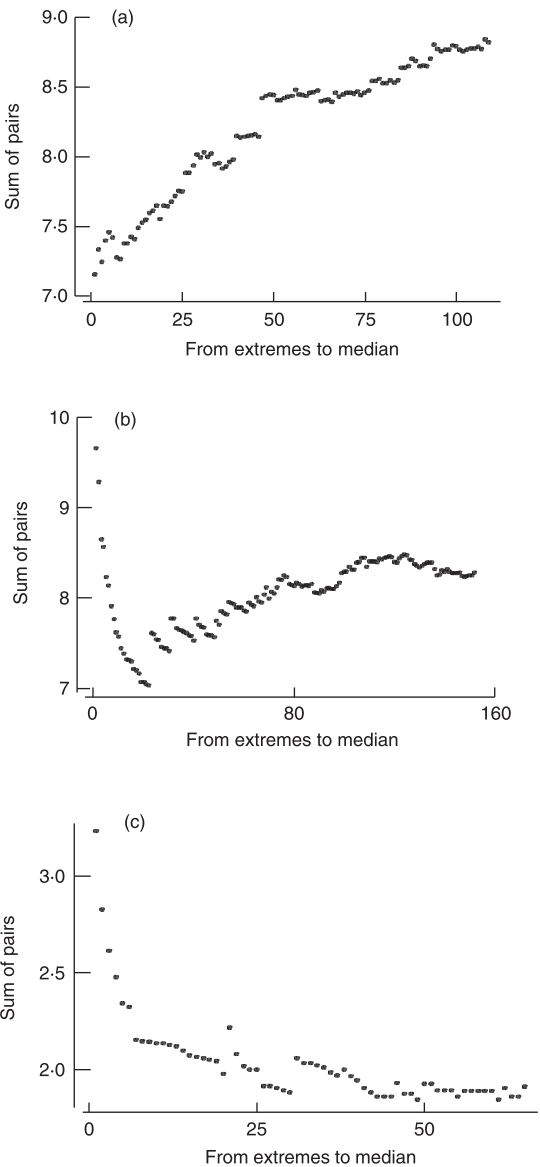

Fig. 3. Modified Tukey symmetry plots of the three rank abundance plots in Fig. 1. These show the left skew of the two complete enumerations (birds and trees) and the right skew of the incomplete enumeration (butterflies).

irrelevant, because Fig. 2 indicates that for both British birds and BCI trees the skewness comes from a deficiency in common species, not from a surplus of rare ones. It seems better to say that the complete enumerations have a short right-hand tail than a long left-hand one.

There are various ways of showing skewness, of which Fig. 3 provides one. The method is a modification of a

suggestion by Tukey in Wilk & Gnanadesikan (1968). Tukey suggested plotting $n_{1+j} + n_{max-j}$ on the ordinate against $n_{1+j} - n_{max-j}$ for $j = 0$ to max/2 on the abscissa, which gives a horizontal line if the distribution is symmetrical. It is a touch simpler to use the rank of the observation on the abscissa. In other words, the observations in pairs is summed from the two ends and worked inwards to the median. If the plot dips down to the left, as in the top two graphs of Fig. 3, then the distribution is left-skewed; if down to the right, as for the butterflies, then it is right-skewed. This plot shows the pattern of skewness throughout the graph, not just whether or not it is skewed. The kick from the singletons among the BCI trees is conspicuous and can also be seen in the butterfly data, but apart from that it is not possible to say from these graphs which end is producing an effect. This is because saying which end only makes sense in comparison with a particular function, as in Fig. 2.

(iii) Comparisons with other distributions

There is a great deal in the literature about which SAD function fits which types of data better. Gray (1987) showed that there was considerable inconsistency and dogma. As our intention is merely to show that the lognormal is not a suitable null hypothesis, that work would be relevant if some other SAD had been shown to be consistently better, or better in certain situations. The evidence seems not to be as strong as that. If one SAD were universally better there would be little argument that it should be preferred and all others discarded. That, however, is not what has been found. The lognormal sometimes fits better than the other SADs that have been suggested, sometimes worse.

The most recent comparisons of fit to the lognormal have been for the ZSM. McGill (2003a) claimed that the lognormal was better, but he had trouble simulating the ZSM. Volkov *et al.* (2003) produced a simpler algorithm for the ZSM, although still requiring numerical integration, and concluded that for BCI trees the ZSM gave the better fit but that 'the lognormal does slightly better than the neutral theory for the Pasoh data set, obtained in a tropical tree community in Malaysia.' Etienne & Olff (2004) used an ingenious new method to obtain the ZSM distribution and found, again with BCI > 10 cm d.b.h. trees, that the lognormal did better, but only weakly so. McKane *et al.* (2004) provide an analytical solution for the Hubbell local community. Olszewski & Erwin (2004) fitted that to four fossil brachiopod communities and find the ZSM is platykurtic compared to the lognormal and has similar likelihoods. The only safe conclusion from all this hard work would seem to be that the lognormal and the ZSM fit tropical forest tree (and possibly other) data about equally well. Plotkin & Muller-Landau (2002) fitted various SADs to BCI data and found that the gamma distribution (which we do not discuss in this paper) fitted best. Our discussion of the central limit theorem,

below, suggests why the ZSM and the lognormal might be expected to be similar. Before that we consider a SAD that seems to work consistently better than the lognormal (and so presumably the ZSM) in complete enumerations.

(iv) Comparison with the logit-normal distribution

Range sizes for species usually can vary between a well-defined minimum and a well-defined maximum. In Britain, using hectad recording (recording by 10 km × 10 km grid squares), the range size can vary from one hectad up to the total number of hectad grid squares that exist. This led Williamson & Gaston (1999) to show that the logit transformation of range size was appropriate, and that the logits of the ranges of sets of species were usually normally distributed. This transformation has proved useful in several contexts (Williamson 1998; Gaston 2002; Preston, Pearman & Dines 2002; Williamson, Preston & Telfer 2003).

The logit is $\log(p/q)$ where p is the proportion of the variable from 0 to 1, while $q = (1-p)$. Writing p as n/s, where n is the number of observations (hectads in a British range study) and s is the maximum, then q is $(s-n)/s$. Williamson & Gaston (1999) found it desirable to let s vary, to allow for hectads that had not been properly recorded. Extending that, if s becomes very large, then q ceases to vary and the logit becomes proportional to $\log(p)$. In that sense, the logarithm is a natural limit of the logit. If logit-normal distributions are considered with variable s, then their limit is the lognormal when s becomes indefinitely large.

The abundance of species is also limited, but it is not clear what that limit is for a particular assemblage or community. Should it be the maximum numbers any one species can attain, or should it be the maximum for the sum of all species, or some other value? If logits are fitted to abundances by finding an optimum s, are they more normally distributed than the logarithms of the abundances? Is the logit-normal a better fit than the lognormal?

Testing this idea with bird population data from a variety of European countries, i.e. complete enumerations (Gregory 2000), we found a curious result (Williamson, Gaston & Gregory, in preparation). Whether or not the fit was significantly good (using the Ryan–Joiner test), the logit-normal always fitted better than the lognormal. This effect can be seen for British birds and BCI trees, with the logit transformation improving the fit at the right-hand end (Fig. 4, Table 2) but otherwise having little effect. As the high abundances are now nearer the limit (the limit for the lognormal being infinity), there appears to be less of a shortage of abundant species. Table 2 shows that for these two sets the optimal fit comes with a maximum abundance not much greater than that of the most common species. The only difference in significance from Table 1 is that, under the Ryan–Joiner test, BCI trees move from being just significantly a bad fit to just not significantly so at the

Table 2. Tests for fit to a logit-normal of three species–abundance data sets, giving the number of species, the multiple that *s* (the right-hand limit) is of the maximum observed population size (see text for details), probabilities for the Anderson–Darling, Ryan–Joiner (equivalent to Shapiro–Wilks test), Kolmogoroff–Smirnov tests, and the correlation coefficient used in the Ryan–Joiner test

	No. spp.	Multiple	Anderson–Darling test	Ryan–Joiner test	Kolmogoroff–Smirnov test	Correlation coefficient
British breeding birds	217	1·01	< 0·001	< 0·01	< 0·01	0·9901
BCI trees, > 1 cm 1990	304	1·38	0·001	0·0542	> 0·15	0·9954
Ecuadorian butterflies	130	5·88+	0·001	> 0·100	> 0·15	0·9947

conventional 5% level but, as the correlation coefficients show, the British bird data also fit a touch better. The transformation has no effect on the Ecuadorian butterfly data, where there was no shortage of abundant species in the first place, compared to the fitted lognormal.

Although the effect is small and we have only two results to present here, we make bold to say that the logit-normal gives a consistently better fit to empirical SADs than the lognormal in complete enumeration data. We would expect that it will also perform better than the ZSM, as that is so close to the lognormal. Should that be confirmed, then, on the argument given earlier, neither the lognormal nor the ZSM are acceptable as SADs for complete enumerations.

Suggested causes of the distribution

CENTRAL LIMIT THEORY FOR LOGARITHMICALLY VARYING RANDOM PROCESSES

The quotations in the Introduction from Nee & Stone (2003), Harte (2003) and Taylor (1978) show that it is commonplace for biologists to think that a lognormal SAD follows from the central limit theorem applied to multiplicative processes. The central limit theorem says, approximately, that a variable subject to a (moderately large) number of independent additive effects will tend to be distributed normally. In that case, independent samples from the same population will be normally distributed. Le Cam (1986) gives the exact mathematics, some history and notes that Polyá dubbed the theorem central because of its importance. If the effects are multiplicative rather than additive then the variable will have a lognormal distribution. 'The central limit theorems … are the groundwork on which all existing theories of the genesis of the lognormal have been erected.' (Aitchison & Brown 1957, p. 52).

This argument is satisfactory for the populations of a single species but not for a set of species, as has been pointed out by Pielou several times (and by other authors following her, e.g. Williamson 1988). For example, 'The … argument justifies … the hypothesis that the sizes of separate and independent populations of some one species (in separate noninteracting communities) will be random variates from a lognormal distribution. … But it does *not* follow (as is sometimes assumed) that the distribution of population sizes of a number of
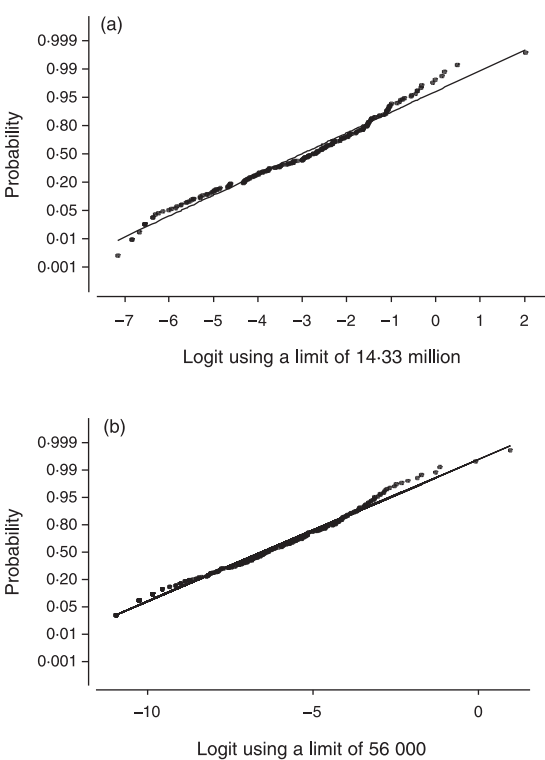

Fig. 4. Probability plots of the optimal logit transformation of plots (a) and (b) in Fig. 1. Compared with Fig. 2, the difference is primarily in the right-hand end, which now fits better.

different species, occurring together and forming a single community, must also be lognormal.' (Pielou 1975, p. 48). This reasoning can also be looked at the other way. Only if the populations are separate independent samples of the same entity will the central limit theorem hold. Samples from different places or different times of a single species may be from the same entity; samples of different species are definitely samples of different entities unless the species have identical demographic parameters. For SADs that suggests that the Hubbell model, which has almost those assumptions, should lead to a lognormal SAD. In practice, as noted above, the Hubbell ZSM is scarcely distinguishable from a lognormal with tropical tree data. Note that all the quotations given in the Introduction regarding the central limit theorem postdate Pielou, showing how difficult it is to overcome a widespread and attractive fallacy. Possibly, the assertion that the conditions of the theorem have not been met is not sufficiently compelling.

Here is a slightly more elaborate (and new) argument. Consider a data matrix of the logarithms of the abundances of several species. In this matrix, rows represent species, and columns independent samples of the set. The columns could be different years if the dynamics of all the species make the yearly abundances of each independent. Then, from the argument above, each row can be expected to tend to lognormality (Dennis & Patil 1988). The matrix as a whole then will be the sum of a set of lognormal distributions each, in general, with a different mean and variance. As Allen, Li & Charnov (2001) show, such a sum will not be lognormal. Each column is an independent sample of the total matrix and so, again will not be lognormal. Despite the application of the central limit theorem to each row, the SADs of the columns will not be lognormal. In practice, even the rows will only sometimes be lognormal as there will be other factors acting. Halley & Inchausti (2002) found that about half of the 544 population time series, all more than 30 years long, they examined appeared lognormal, although they did not allow for autocorrelation or trends in the data. Nor are they from the 'separate noninteracting communities' which Pielou (above) states correctly as being needed for the central limit theorem to apply properly.

It might be thought that the rather loose argument for applying the central limit theorem to SADs was just mildly inaccurate. In fact, it is clearly wrong and has no validity except in a purely neutral model.

NICHE BREAKAGE OR NICHE APPORTIONMENT

Niche apportionment models start with a totality of niche space and then subdivide that space according to various rules (Tokeshi 1999). What Tokeshi calls the random fraction model divides the space into two random parts; one part is chosen at random and again divided at random and so on. The sizes of the resulting set of random pieces of niche space will approach a lognormal distribution. That was shown first by Kolmogoroff (1941), and applied first to species abundance distributions by Bulmer (1974). There are some important assumptions, notably that the size of a niche fragment translates linearly into a population abundance, that species divide niche space into contiguous distinct volumes, and that niche space is saturated, without gaps. It is generally thought that none of those assumptions is exactly true or even true at all; but the model is the only plausible one, apart from Hubbell's neutral theory and the Engen & Lande theories discussed below, that predicts a distribution approaching the lognormal.

If the niche apportionment rules are different, the resulting distribution of niche sizes will be different. Sugihara (1980) proposed making the division always 3 : 1 instead of random and claimed to produce a canonical lognormal distribution. As Nee *et al.* (1991) showed by simulation, his rule produces a slightly left-skewed distribution and so not a lognormal and even less a canonical lognormal. He may, by chance, have produced something that actually matches the SADs of British birds and BCI trees; but other rules, such as Tokeshi's power fraction model, also produce a mild left skew.

None of the variety of rules suggested for niche apportionment models are particularly plausible from general ecological knowledge and none is to be obviously preferred to another. Without more work attaching the models to ecological reality, the best that can be said is that they produce plausible distributions that provide quite good fits, under some tests, to some real data. As this stage they can neither be dismissed nor accepted. Sugihara *et al.* (2003) construct a hierarchical framework for niches which may have some promise in getting to a (nonlognormal) species–abundance function.

MODELS FROM POPULATION DYNAMICS

A more satisfactory biological approach than either the central limit theorem or niche apportionment would be to build species abundance models from the collective dynamics of a set of species. In a sense, that is exactly what Hubbell's neutral model does, so the ZSM can be regarded as a population dynamic model. Chave (2004) shows why the ZSM may also fit in non-neutral equilibrial situations. Ecologists in general have reacted strongly against Hubbell's model on various grounds, including that there are obvious differences in the population dynamics of different species (e.g. Enquist, Sanderson & Weiser 2002; Nee 2003; Nee & Stone 2003). Curiously, many of them still revere the equally neutral MacArthur–Wilson model. Hubbell's model does produce impressive fits to data but the ZSM, as we have seen, is very close to the lognormal; indeed, effectively indistinguishable from it in the tropical forest data.

Engen, Lande and their colleagues are developing more conventional population dynamic models based on competition communities. These are sets of species at the same trophic level that, in the model, interact only by competition. Our three examples may approximate to that, the trees the most, the birds the least. In their earlier papers (Engen & Lande 1996a,b), where the models include speciation, the form of intraspecific density dependence is important. Logistic dependence ($d \ln n/dt = a - bn$) led to a gamma SAD while the lognormal SAD arose from Gompertz dependence ($d \ln n/dt = a \cdot \exp(-bt)$). The gamma SAD is the standard generalization of the log-series (Kempton & Taylor 1974). Developing these models purely ecologically (Lande *et al.* 2003), a set of stochastic differential equations with Gompertz density dependence, demographic and both general and specific environmental stochasticity can lead to a lognormal SAD, if the intrinsic rates of natural increase (r) are normally distributed. Other models involve migration (Engen 2001).

It is at least clear that community models can be built that will result in a variety of SAD functions, including the lognormal. Testing these theories is very difficult. Who has any idea of the functional shape (or shapes) of intraspecific density dependence or the distribution of intrinsic rates of increase across a set of species? Testing the consequences rather than the intrinsic details of these theories is easier, as is evident in Hubbell's approach. Altogether, population modelling seems to us a better approach than niche theories. All the more reason then not to develop models that have the unfortunate mathematical consequences of the lognormal discussed in the next section.

Consequences of the mathematical nature of the lognormal distribution

Only the consequences of a theory can produce good evidence in its favour. Here, the theory is that a mathematical distribution, the lognormal, describes the abundance of species. The nature of the distribution and its consequences determine whether it is a satisfactory description and may indicate the limits of its scope. In this section we look at six such features of the lognormal. The conclusions vary between condemning the theory and neutrality about it; none offer support for it.

(i) The lognormal distribution borders unsatisfactory distributions

The mathematics of the lognormal distribution are, of course, well known (Aitchison & Brown 1957; Shimizu & Crow 1988), and we will not revisit them here. Ulanowicz (2003) points out that the lognormal is at the interface between well-behaved negative exponentials and a power-law family of distributions that have infinite theoretical variances. He says 'it is sometimes not easy to decide whether data represent a lognormal distribution or whether they might better portray a closely related, but more poorly behaved, power-law formula'. Similarly, Schmoyer *et al.* (1996) note that the log-*t* distributions (t as in Student's *t*-test) have infinite variances and that the lognormal is the log-*t* with an infinite number of degrees of freedom. In both cases, of course, real finite samples must have finite variances, but could be drawn from theoretical distributions with infinite variances. The Cauchy distribution is the standard example of one such in textbooks.

The distributions with infinite variance all have thicker tails than the lognormal. As we will show below, the lognormal already has too thick a right-hand tail to be plausible. The same argument would show that these infinite-variance distributions are even less plausible.

(ii) The lognormal is not additive and is taxonomically restricted

The addition of two lognormal curves with different means or different variances or both leads to a distri-

bution that is not lognormal. So if the lognormal applies, for instance, to passeriform birds and to charadriiform birds, it cannot apply to all birds and, conversely, if it applies to all birds it cannot apply to individual orders of birds. In practice, the lognormal has been applied to taxonomically limited sets without examining subsets. Indeed, that is what we did with our three examples. It is far from clear at what level it should apply. This difficulty is common to all other SAD functions that have been proposed, which suggests that we are far from a satisfactory theory of SADs, irrespective of the validity of the lognormal.

(iii) The lognormal is continuous

It has been noted in the literature from time to time that SADs are intrinsically discrete but the lognormal is continuous (Kempton & Taylor 1974; Pielou 1975; Williamson 1981; Magurran 1988). Equally, it has been said that the discrete SAD might be the realization of an underlying continuous process (Pielou 1977). Most other SAD functions are discrete. However, as the argument in (v) below will bring out, in practice SADs are only evidently discrete at low numbers and having a discrete function for abundant species is not helpful. For instance, most British bird populations are estimated to only two significant figures. A logarithmic transformation for population size has long been recognized as desirable (Williams 1964; Williamson 1972; Gaston & McArdle 1994) and logarithms are in general not integers. The Poisson lognormal is one attempt to bridge the transition but has not been found satisfactory (see (vi) below).

One consequence of the continuous nature of the lognormal is that it goes to infinity in both directions or, converted back to an arithmetic scale, has a tail ending at zero on the left and a long thin tail going to infinity on the right. The former is regarded by Dewdney (1998, 2000, 2003) as showing the lognormal to be unrealistic as, in his survey, all empirical SADs had a high point at the left-hand side; the singleton class is the most common. The latter is the cause of an unsatisfactory individuals abundance curve (see below). A discrete distribution could perhaps circumvent both these difficulties, but that they are difficulties is a valid objection to the lognormal.

(iv) The right-hand tail is too thick

This objection is lethal to the theory that the lognormal describes the abundance of species yet, oddly, has led to a large literature about how interesting it is. The problem was discussed first in Preston (1948) and developed into the theory of canonical lognormals in Preston (1962), with exceptions noted in Preston (1980). By simple manipulation (Preston 1948; Aitchison & Brown 1957), the species–abundance distribution can produce an individuals–abundance distribution, the number of individuals of species with particular abundances. The

latter is also a normal (Gaussian) curve, right-shifted. Preston developed this into the canonical lognormal hypothesis, whereby the most abundant species had individuals at the maximum of the individuals abundance curve. This is shown in many books (e.g. Magurran 1988, 2004).

In Preston's canonical view, the whole of the right-hand half of the individuals abundance curve is missing. Taking a less rigid link between the two curves, as in Preston (1948) or May (1975), the missing part is still approximately a half. The lognormal SAD predicts the existence of many extremely abundant species that do not exist. It is standard that a theory that predicts an absurdity must be wrong. Surprisingly, this obvious argument against the feasibility of a lognormal SAD seems not to have been made before. The canonical lognormal is, then, not something interesting but something absurd. We agree with Dennis & Patil (1988) that 'The enthusiasm ecologists have for this [Preston's Canonical] hypothesis must be judged from a statistical standpoint as premature.'

It is the right-hand tail of the lognormal that generates the non-existent part of the individuals abundance curve. An appropriate theory should generate an individuals abundance curve that matches what is observed. It follows that such a theory must have a thinner right-hand tail than the lognormal. This is indeed what we noted is observed in fitting lognormals; the right-hand tail is empirically thinner than the lognormal.

(v) The log-binomial and the individuals abundance curve

As noted above, on one side the lognormal borders unsatisfactory distributions. On the other side it can be said to border the log-binomial. This is not a realistic distribution and has not appeared in the SAD literature before. Nevertheless, it makes a useful point about the relationship of the SAD and the individuals abundance curve. We develop the argument numerically. It is easily put into algebraic form.

Consider a small binomial distribution: 1, 6, 15, 20, 15, 6, 1. Suppose, in the spirit of Preston, that these are the numbers of species that have exactly 1, 2, 4, 8, 16, 32 and 64 individuals. Then, the individuals abundance distribution, from cross-multiplying, is 1, 12, 60, 160, 240, 192, 64. The binomial distribution, the SAD, is given by the coefficients in the expansion of $(x + y)^6$, while the individuals abundance distribution is given by the coefficients of $(x + 2y)^6$. The 2 in $2y$ comes from using a doubling between classes; it would be $3y$ had we used triplings as, for instance, do DeVries *et al*. (1997). The individuals abundance curve is significantly left-skewed ($\chi_1^2 = 10.27$, $P < 0.01$). As the normal is the limit of the binomial, it could be said that the symmetrical lognormal individuals curve borders asymmetrical, left-skewed, log-binomial individuals curves.

In the sense that it produces a realistic individuals abundance curve, echoing what is seen empirically, the

log-binomial is an improvement on the lognormal. By allowing species to have only exactly 2^n (or 3^n or any other such series) individuals it is clearly most unrealistic. No doubt the realism could be improved by regarding the 2^n classes as bin markers and allowing the species to have numbers included in the bin. The argument for the individuals abundance curve would then become approximate, but the result would still be a strong left skew. We do not consider this to be a line worth pursuing. The log-binomial has been introduced solely to show the exceptional nature of the symmetrical, normal, shape of the lognormal individuals abundance curve.

(vi) Preston's veil line is a misunderstanding

Preston (1948) developed the theory of the lognormal SAD using histograms in 'octaves', i.e. a doubling between bins. He thought that in samples, what we have called incomplete enumerations, the left-hand part of the histogram would be truncated. Some of the bins would not be observed. He also thought that doubling the sample would reveal exactly one more bin at the left-hand end and similarly for other increases in sample size.

This view has remained popular despite its neglect of proper sampling. Grundy (1951) was sceptical: 'Heuristically, the sample may be supposed to include most of the species to the right of the veil line, and few of those to the left', but suggested no improvement. He also noted the unfortunate nature of Preston's binning, which we have already discussed, and adds 'Preston does not, however, seem to use this [equation 5, a formula for using probits], since he adopts a method of grouping into octaves by which half the species appearing as singletons in the sample are assigned to the left of the veil line.'

From the literature, two points emerge. The first is that sampling does often appear to give an appearance of truncation (e.g. Taylor 1978, using different samples; Gaston & Blackburn 2000, using proper subsamples). The second is that increasing the sample does not just add new bins on the left, it changes the shape of the curve. Complete enumerations are left-skewed and incomplete enumerations are often right-skewed (Fig. 3). This can be seen, although it was not discussed, in Hutchinson's (1967) tabulation of Patrick's diatom data, and seen and discussed in Hubbell & Foster (1983) for trees at BCI.

Dewdney (1998) developed a new theory for sampling from SADs. This uses a Poisson approximation to the hypergeometric distribution involved in taking finite samples without replacement. For some SADs, such as the log-series, sampling does not change the shape. He did not produce an expression for the sampled lognormal but plots such a curve showing the increased right skew. All that can be said is that the shape will change and the change will not involve a veil line but a diminution growing stronger going leftwards. This

diminution looks like a truncation when the data are presented as histograms. In rank abundance plots it normally produces a great number of singletons (Fig. 1c) and log right skew (Fig. 3c). The problem of sampling from the lognormal is made more difficult by the samples being necessarily discrete. Nevertheless, there will always (*contra* Wilson 1991) be a sampling effect in incomplete enumerations.

One mathematical attempt to deal with the problems of sampling and discreteness was the development of the Poisson lognormal (Grundy 1951; Cassie 1962; Pielou 1969, 1975, 1977; Bulmer 1974; Slocomb, Stauffer & Dickinson 1977). The argument necessarily has several assumptions about sampling. It is not clear if it is these assumptions, or the more general assumption that it is a lognormal SAD that is being sampled, that leads to an unfortunate and clearly incorrect result. Taking the simpler truncation view, the Poisson lognormal gives an estimate of the number of species that have been truncated which, added to the species that have been observed, leads to an estimate of the total number of species. This, of course, only makes sense for incomplete enumerations. For our example of Ecuadorian butterflies, there certainly must be a figure for the total number of fruit-eating nymphalid species that occurred in Ecuador, or some part of it, during the sampling period and it would be interesting to estimate it. In other cases, for instance for Lepidoptera caught in a light trap, it is less clear what set of species would be estimated. The problem is one of extrapolation, always a risky process.

Slocomb *et al*. (1977) say 'Pielou (1975) points out that currently available estimates of N [the total number of species] are not satisfactory; this observation is certainly supported by the results presented here and in Bulmer (1974). This lack of confidence in estimates of N is unfortunate'. Pielou (1975) says 'estimates of s* [the total number of species] obtained by fitting the Poisson lognormal and the continuous lognormal … are discrepant. … estimates of s* rarely inspire confidence. The whole problem is ripe for further investigation.' Hughes (1986) and Magurran (1988, 2004) also note that such estimates are unsatisfactory. O'Hara & Oksanen (2003) note that different estimates are found for the number of unobserved species when fitting histograms from when fitting rank abundance curves, but they are uncertain why. It seems to us that the assumption of truncation is the central issue. Agreeing with Dewdney (1998) that sampling does not lead to truncation means that all calculations based on that assumption, including fitting Poisson lognormals, should be set aside.

Pielou (1969, 1977) notes that 'In many collections it is found that singleton species (those represented by one individual) are numerous, often the most numerous' and Dewdney (1998, 2000, 2003) confirms this. That is not true if the collections are large enough. Collections of many thousands of individuals are often still small relative to the actual size of the assemblage,

particularly at broader spatial scales. From our examples and two data sets (copepods and phytoplankton) in McGowan & Walker (1993) we suggest, tentatively, that the dominance of singletons will be lost when between 100 000 and 1 000 000 individuals have been studied. Range-size distributions for various groups in Europe do go down to a size of one hectad (100 km²), the sampling unit, but such singletons are not the most common observation, as is seen easily in area dominance (or rank range distribution) plots (Gaston *et al*. 1998; Williamson & Gaston 1999; Williamson 2002; Gaston 2003), and even so each hectad normally contains more than one individual, even of rare species. The BCI trees have numerous singletons but the BCI plot is only 50 ha (0·5 km²). Scaling up implies 200 individuals per hectad as a median, many thousands in Panama. The abundance of singletons in published data shows the effort required to obtain an adequate SAD rather than showing the shape of such distributions. A satisfactory theoretical SAD should have few singletons in large enumerations and be slightly log left-skewed at that size while smaller enumerations should have a dominance of singletons viewed arithmetically and be log right-skewed. The lognormal does not have these properties (Dewdney 1998).

Conclusions

The lognormal has been an attractive model for species–abundance distributions for two main reasons: many empirical rank abundance curves are sigmoidal, and the central limit theorem seems to give a simple and natural explanation. The situation has been clouded by working mostly with incomplete enumerations, usually called samples. Nevertheless, it is now certain that the lognormal is unacceptable.

First, empirically, fits to complete enumerations show a deficiency of observations at the right-hand tail leading to left skew, and it would seem that the logit-normal fits better.

Secondly, the central limit theorem leads to a lognormal distribution of different samples of one species, or of samples of sets of equivalent species, but fails to do so for sets of species with different biological characters. As the last is what is needed, it does not lead to a biologically realistic SAD. Other theories, either niche apportionment or based on population dynamics, that can give rise to a lognormal can also with minor changes yield other left-skewed distributions.

Thirdly, there are a number of theoretical objections to the lognormal. By far the most serious of these is that the lognormal SAD implies a lognormal individuals abundance curve that in turn implies that there are many abundant species that are not observed. The lognormal theory of SADs leads to a false conclusion. There are also serious problems with the effect of sampling and discreteness at the left-hand end of the distribution. These conclusions show that models that claim to lead to the lognormal SAD must be mistaken.

So, the distribution of the logarithm of species abundance remains unknown except, as Mandelbrot (1982, p. 278) said of turbulent fluids, it cannot be Gaussian. A true log species–abundance distribution must have a steeper, thinner, right-hand form than the lognormal. Much data show it to be bell-shaped but we know of no good reason to prefer, as of now, any one such mathematical function over any other. There are many suggestions in the literature, but what is needed is a distribution with a sound theoretical basis that leads to testable and correct consequences.

Acknowledgements

References

Aitchison, J. & Brown, J.A.C. (1957) *The Lognormal Distribution*. Cambridge University Press, Cambridge.

Allen, A.P., Li, B.-L. & Charnov, E.L. (2001) Population fluctuations, power laws and mixtures of lognormal distributions. *Ecology Letters*, **4**, 1–3.

Bulmer, M.G. (1974) On fitting the Poisson lognormal distribution to species–abundance data. *Biometrics*, **30**, 101–110.

Cassie, R.M. (1962) Frequency distribution models in the ecology of plankton and other organisms. *Journal of Animal Ecology*, **31**, 65–92.

Chave, J. (2004) Neutral theory and community ecology. *Ecology Letters*, **7**, 241–253.

Condit, R. (1998) *Tropical Forest Census Plots: Methods and Results from Barro Colorado Island, Panama and a Comparison with Other Plots*. Springer-Verlag, New York.

Condit, R., Hubbell, S.P. & Foster, R.B. (1996) Changes in tree species abundance in a neotropical forest: impact of climate change. *Journal of Tropical Ecology*, **12**, 231–256.

Condit, R., Pitman, N., Leigh, E.G. Jr, Chave, J., Terborgh, J., Foster, R.B., Núñez, P., Aguilar, S., Valencia, R., Villa, G., Muller-Landau, H., Losos, E. & Hubbell, S.P. (2002) Beta-diversity in tropical forest trees. *Science*, **295**, 666–669.

Dennis, B. & Patil, G.P. (1988) Applications in ecology. *Lognormal Distributions: Theory and Applications* (eds E.L. Crow & K. Shimizu), pp. 303–330. Marcel Dekker, New York.

DeVries, P.J., Murray, D. & Lande, R. (1997) Species diversity in vertical, horizontal, and temporal dimensions of a fruit-feeding butterfly community in an Ecuadorian rainforest. *Biological Journal of the Linnean Society*, **62**, 343–364.

Dewdney, A.K. (1998) A general theory of the sampling process with applications to the 'veil line'. *Theoretical Population Biology*, **54**, 294–302.

Dewdney, A.K. (2000) A dynamical model of communities and a new species–abundance distribution. *Biological Bulletin*, **198**, 152–165.

Dewdney, A.K. (2003) The stochastic community and the logistic-J distribution. *Acta Oecologia*, **24**, 221–229.

Engen, S. (1978) *Stochastic Abundance Models*. Chapman & Hall, London.

Engen, S. (2001) A dynamic and spatial model with migration generating the log-Gaussian field of population densities. *Mathematical Biosciences*, **173**, 85–102.

Engen, S. & Lande, R. (1996a) Population dynamic models generating the lognormal species abundance distribution. *Mathematical Biosciences*, **132**, 169–183.

Engen, S. & Lande, R. (1996b) Population dynamic models generating species abundance distributions of the gamma type. *Journal of Theoretical Biology*, **178**, 325–331.

Enquist, B.J., Sanderson, J. & Weiser, M.D. (2002) Neutral macroecology: a reply to G. Bell. *Science*, **295**, 1517–1520.

Etienne, R.S. & Olff, H. (2004) A novel genealogical approach to neutral biodiversity theory. *Ecology Letters*, **7**, 170–175.

Fisher, J. (1952) Bird numbers; a discussion on the breeding populations of inland birds of England and Wales. *South Eastern Naturalist*, **57**, 1–10.

Gaston, K.J. (2002) Abundance, occupancy and conservation biology. *Avian landscape ecology: pure and applied issues in the large-scale ecology of birds. Proceedings of the 2002 IALE (UK) Conference* (eds D.E. Chamberlain & A.M. Wilson), pp. 215–227. IALE (UK), Garstang, Lancs.

Gaston, K.J. (2003) *The Structure and Dynamics of Geographic Ranges*. Oxford University Press, Oxford.

Gaston, K.J. & Blackburn, T.M. (2000) *Pattern and Process in Macroecology*. Blackwell Science, Oxford.

Gaston, K.J., Blackburn, T.M. & Gregory, R.D. (1997) Abundance–range size relationships of breeding and wintering birds in Britain: a comparative analysis. *Ecography*, **20**, 569–579.

Gaston, K.J. & McArdle, B.H. (1994) The temporal variability of animal abundances: measures, methods and patterns. *Philosophical Transactions of the Royal Society, London B*, **345**, 335–358.

Gaston, K.J., Quinn, R.M., Blackburn, T.M. & Eversham, B.C. (1998) Species–range size distributions in Britain. *Ecography*, **21**, 361–370.

Gibbons, D.W., Reid, J.B. & Chapman, R.A., eds (1993) *The New Atlas of Breeding Birds in Britain and Ireland: 1988–1991*. Poyser, London.

Gray, J.S. (1987) Species–abundance patterns. *Symposia of the British Ecological Society*, **27**, 53–68.

Gregory, R.D. (2000) Abundance patterns in European breeding birds. *Ecography*, **23**, 201–208.

Grundy, P.M. (1951) The expected frequencies in a sample of an animal population in which the abundancies of species are lognormally distributed. Part 1. *Biometrika*, **38**, 427–434.

Halley, J. & Inchausti, P. (2002) Lognormality in ecological time series. *Oikos*, **99**, 518–530.

Harte, J. (2003) Tail of death and resurrection. *Nature*, **424**, 1006–1007.

Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton.

Hubbell, S.P. & Foster, R.B. (1983) Diversity of canopy trees in a neotropical forest and implications for conservation. *Tropical Rain Forest: Ecology and Management* (eds S.L. Sutton, T.C. Whitmore & A.C. Chadwick), pp. 25–41. Blackwell Scientific Publications, Oxford.

Hughes, R.G. (1986) Theories and models of species abundance. *American Naturalist*, **128**, 879–899.

Hutchinson, G.E. (1967) *A Treatise on Limnology*, Vol. 2. John Wiley & Sons, New York.

Kempton, R.A. & Taylor, L.R. (1974) Log-series and log-normal parameters as diversity discriminants for the Lepidoptera. *Journal of Animal Ecology*, **43**, 381–399.

Kolmogoroff, A.N. (1941) Über das logarithmisch Normale Verteilungsgesetz der Dimensionen der Teilchen bei Zerstückelung. *Comptes Rendu de l'Académie Des Sciences de l'Union Des Républiques Soviétiques Socialistes*, **31**, 99–101.

Lambshead, J. & Platt, H.M. (1985) Structural patterns of marine benthic assemblages and their relationships with empirical statistical models. *Proceedings of the 19th European Marine Biology Symposium* (ed. P.E. Gibbs), pp. 371–380. Cambridge University Press, Cambridge.

Lande, R., Engen, S. & Saether, B.-E. (2003) *Stochastic Population Dynamics in Ecology and Conservation*. Oxford University Press, Oxford.

Le Cam, L. (1986) The central limit theorem around 1935 (with discussion). *Statistical Science*, **1**, 78–96.

Leigh, E.G. Jr (1999) *Tropical Forest Ecology: a View from Barro Colorado Island*. Oxford University Press, New York.

Magurran, A.E. (1988) *Ecological Diversity and its Measurement*. Croom-Helm, London.

Magurran, A.E. (2004) *Measuring Biology Diversity*. Blackwell Publishing, Oxford.

Magurran, A.E. & Henderson, P.A. (2003) Explaining the excess of rare species in natural species abundance distributions. *Nature*, **422**, 714–716.

Mandelbrot, B.B. (1982) *The Fractal Geometry of Nature*. W.H. Freeman, San Francisco.

Marquet, P.A., Keymer, J.E. & Cofré, H. (2003) Breaking the stick in space: of niche models, metacommunities and patterns in the relative abundance of species. *Macroecology: Concepts and Consequences* (eds T.M. Blackburn & K.J. Gaston), pp. 64–84. Blackwell Science, Oxford.

May, R.M. (1975) Patterns of species abundance and diversity. *Ecology and Evolution of Communities* (eds M.L. Cody & J.M. Diamond), pp. 197–227. Harvard University Press, Cambridge, Massachusetts.

McGill, B.J. (2003a) A test of the unified neutral theory of biodiversity. *Nature*, **422**, 881–885.

McGill, B.J. (2003b) Does mother nature really prefer rare species or are left-log-skewed SADs a sampling artefact? *Ecology Letters*, **6**, 766–773.

McGill, B.J. (2003c) Strong and weak tests of macroecological theory. *Oikos*, **102**, 679–685.

McGowan, J.A. & Walker, P.M. (1993) Pelagic diversity patterns. *Species Diversity in Ecological Communities* (eds R.E. Ricklefs & D. Schluter), pp. 203–214. University of Chicago Press, Chicago.

McKane, A.J., Alonso, D. & Solé, R.V. (2004) Analytic solution of Hubbell's model of local community dynamics. *Theoretical Population Biology*, **65**, 67–73.

McMahon, T.A. & Bonner, J.T. (1983) *On Size and Life*. Scientific American Books, New York.

Nee, S. (2003) The unified phenomenological theory of biodiversity. *Macroecology* (eds T.M. Blackburn & K.J. Gaston), pp. 31–41. Blackwell Publishing, Oxford.

Nee, S., Harvey, P.H. & May, R.M. (1991) Lifting the veil on abundance patterns. *Proceedings of the Royal Society, London B*, **243**, 161–163.

Nee, S. & Stone, G. (2003) The end of the beginning for neutral theory. *Trends in Ecology and Evolution*, **18**, 433–434.

O'Hara, R.B. & Oksanen, J. (2003) fisherfit {vegan} Fit Fisher's logseries and Preston's lognormal model to abundance data. http://cc.oulu.fi/~jarioksa/softhelp/vegan/html/fisherfit.html.

Olszewski, T.D. & Erwin, D.H. (2004) Dynamic response of Permian brachiopod communities to long-term environmental change. *Nature*, **428**, 738–741.

Pielou, E.C. (1969) *An Introduction to Mathematical Ecology*. Wiley-Interscience, New York.

Pielou, E.C. (1975) *Ecological Diversity*. Wiley-Interscience, New York.

Pielou, E.C. (1977) *Mathematical Ecology*, 2nd edn. Wiley-Interscience, New York.

Plotkin, J.B. & Muller-Landau, H.C. (2002) Sampling the species composition of a landscape. *Ecology*, **83**, 3344–3356.

Preston, F.W. (1948) The commonness, and rarity, of species. *Ecology*, **29**, 254–283.

Preston, F.W. (1962) The canonical distribution of commonness and rarity. *Ecology*, **43**, 410–432.

Preston, F.W. (1980) Noncanonical distributions of commonness and rarity. *Ecology*, **61**, 88–97.

Preston, C., Pearman, D.A. & Dines, T.D., eds (2002) *New Atlas of the British and Irish Flora*. Oxford University Press, Oxford.

Schmoyer, R.L., Beauchamp, J.J., Brandt, C.C. & Hoffman, F.O. Jr (1996) Difficulties with the lognormal model in mean estimation and testing. *Environmental and Ecological Statistics*, **3**, 81–97.

Shimizu, K. & Crow, E.L. (1988) History, genesis and properties. *Lognormal Distributions: Theory and Applications* (eds E.L. Crow & K. Shimizu), pp. 1–25. Marcel Dekker, New York.

Slocomb, J., Stauffer, B. & Dickson, K.L. (1977) On fitting the truncated lognormal distribution to species–abundance data using maximum likelihood estimation. *Ecology*, **58**, 693–696.

Stone, B.H., Sears, J., Cranswick, P.A., Gregory, R.D., Gibbons, D.W., Rehfisch, M.M., Aebischer, N.J. & Reid, J.B. (1997) Population estimates of birds in Britain and in the United Kingdom. *British Birds*, **90**, 1–22.

Sugihara, G. (1980) Minimal community structure: an explanation of species abundance patterns. *American Naturalist*, **116**, 770–787.

Sugihara, G., Bersier, L.-F., Southwood, T.R.E., Pimm, S.L. & May, R.M. (2003) Predicted correspondence between species abundances and dendrograms of niche similarities. *Proceedings of the National Academy of Sciences USA*, **100**, 5246–5251.

Taylor, L.R. (1978) Bates, Williams, Hutchinson – a variety of diversities. *Symposia of the Royal Entomological Society*, **9**, 1–18.

Tokeshi, M. (1993) Species abundance patterns and community structure. *Advances in Ecological Research*, **24**, 111–186.

Tokeshi, M. (1999) *Species Coexistence*. Blackwell Science, Oxford.

Ulanowicz, R.E. (2003) Some steps towards a central theory of ecosystem dynamics. *Computational Biology and Chemistry*, **27**, 523–530.

Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2003) Neutral theory and relative species abundance in ecology. *Nature*, **424**, 1035–1037.

Whittaker, R.H. (1965) Dominance and diversity in land plant communities. *Science*, **147**, 250–260.

Whittaker, R.H. (1969) Evolution of diversity in plant communities. *Brookhaven Symposia in Biology*, **22**, 178–195.

Wilk, M.B. & Gnanadesikan, R. (1968) Probability plotting methods for the analysis of data. *Biometrika*, **55**, 1–17.

Williams, C.B. (1964) *Patterns in the Balance of Nature*. Academic Press, London.

Williamson, M. (1972) *The Analysis of Biological Populations*. Edward Arnold, London.

Williamson, M. (1981) *Island Populations*. Oxford University Press, Oxford.

Williamson, M. (1987) Are communities ever stable? *Symposia of the British Ecological Society*, **26**, 353–371.

Williamson, M. (1988) Relationship of species number to area, distance and other variables. *Analytical Biogeography* (eds A.A. Myers & P.S. Giller), pp. 91–115. Chapman & Hall, London.

Williamson, M. (1998) Measuring the impact of plant invaders in Britain. *Plant Invasions: Ecological Mechanisms and Human Responses* (eds U. Starfinger, K. Edwards, I. Kowarik & M. Williamson), pp. 57–68. Backhuys, Leiden.

Williamson, M. (2002) Alien plants in the British Isles. *Biological Invasions: Economic and Environmental Costs of Alien Plant, Animal and Microbe Species* (ed. D. Pimentel), pp. 91–112. CRC Press, Boca Raton.

Williamson, M. (2003) Species–area relationships at small scales in continuum vegetation. *Journal of Ecology*, **91**, 904–907.

Williamson, M. & Gaston, K.J. (1999) A simple transformation for sets of range sizes. *Ecography*, **22**, 674–680.

Williamson, M., Preston, C. & Telfer, M. (2003) On the rates of spread of alien plants in Britain. *Plant Invasions: Ecological Threats and Management Solutions* (eds L. Child, J.H. Brock, G. Brundu, K. Prach, P. Pysek, P.M. Wade & M. Williamson), pp. 63–74. Backhuys, Leiden.

Wilson, J.B. (1991) Methods for fitting dominance/diversity curves. *Journal of Vegetation Science*, **2**, 35–36.