CHAPTER 10

Fitting and empirical evaluation of models for species abundance distributions

Sean R. Connolly and Maria Dornelas

Copyright 2011. OUP Oxford.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

10.1 Introduction

Identifying and explaining patterns in the commonness and rarity of species has been a fundamental concern of community ecology for nearly a century (Chapter 8). Such *species abundance distributions*, when considered without reference to species identity, are particularly useful to ecologists for two reasons. Firstly, all assemblages have abundance distributions to compare with one another and, secondly, those distributions contain more information than univariate statistics such as species richness or other diversity metrics (McGill et al. 2007). Consequently, ecologists frequently analyse abundance distributions to identify systematic regularities that hold across disparate assemblages and to test ecological theory that purports to explain such regularities.

Nearly a century of investigation has identified both broad-scale similarities and systematic differences in patterns of species abundance. Assemblages with more than a few species overwhelmingly have a frequency distribution of species' abundances that exhibits a 'hollow curve': a plurality of species are represented by one or a few individuals; the number of species represented by increasing numbers of individuals drops off sharply as abundance increases; and there is a very long tail of moderately to highly abundant species (Fig. 10.1a). However, when abundance is plotted on a logarithmic scale, visually identifiable qualitative differences emerge: for instance, in some assemblages, one or more internal modes are present, but in others, no mode is apparent (Fig. 10.1b). The extent to which these differences represent stochastic variation in the numbers of species in each abundance class, systematic distortion of abundance distributions by sampling effects, or differences in the abundance distributions of the underlying communities is controversial (Hubbell 2001; Lande et al. 2003).

Since Motomura (1932; in Whittaker (1965)) proposed the geometric series to characterize patterns of relative abundance in benthic lake habitats, numerous community models have been used to predict species abundance distributions. Some of these models purport to link abundance distributions to a broad range of other ecological quantities, such as numbers of unobserved species (Connolly et al. 2005), range-size frequency distributions (McGill & Collins 2003), species–area and species–time relationships (Adler 2004), body size distributions (Loehle 2006), spatial and temporal patterns in relative abundance (McGill et al. 2005; Dornelas et al. 2006), and niche similarity (Sugihara et al. 2003). Many such models predict some of these additional patterns, but not others, so species abundance distributions have become the data most commonly used to compare the performance of alternative community models. Using such data to compare models has been criticized recently because many attempts to assess model fit to species abundance distributions fail to identify significant discrepancies between models and data, or to clearly discriminate between the fit of alternative models. However, recent developments in model formulation, model fitting, goodness-of-fit testing and model selection have substantially increased

EBSCO Publishing : eBook Collection (EBSCOhost) - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM
AN: 462693 ; Anne E. Magurran, Brian J. McGill.; Biological Diversity : Frontiers in Measurement and Assessment
Account: uamster.main.ebooks

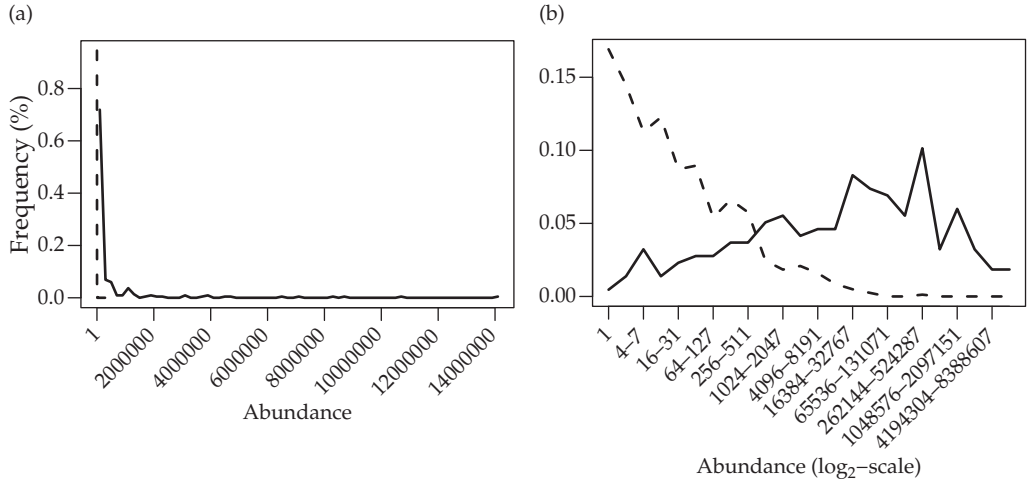

Figure 10.1 Species abundance distributions for the British Breeding Bird survey (Gaston & Blackburn (2000): solid lines) and the benthic fauna of the Ekofisk Oil Field (Gray et al. (1990): dashed lines) on (a) an arithmetic scale and (b) a $\log_2$ scale. Note the strong qualitative similarities between the data sets on an arithmetic scale and the marked differences on the logarithmic scale.

our ability to identify and interpret lack of model fit and to compare the fit of alternative models.

In this chapter, we review and assess traditional and recent techniques for fitting, and for evaluating the fit of, species abundance models. We discuss different approaches for fitting species abundance models to data, for assessing the models' (absolute) goodness of fit to those data and for comparing the (relative) fit of alternative models. Because there has been an increasing diversity of approaches in each of these areas, we describe and critically evaluate these different approaches. We conclude with some advice about the choice of approaches to fitting and testing species abundance models, highlight some important areas for further work, and offer some cautionary words about the ecological interpretation of the results of such analyses.

This chapter focuses on the analysis of numerical abundance (i.e. counts of individuals) because these data have been most intensively investigated and because much of the ecological theory and the more recent statistical modelling is tailored to such data. However, other abundance currencies, such as percentage cover, biomass and energy, have also been analysed (Chiarucci et al. 1999; Connolly et al. 2005). In particular, some theory for abundance distributions, such as niche-apportionment theory, is probably better suited to such alternative metrics (Tokeshi 1990). Recently, Morlon et al. (2009) proposed a general statistical framework for identifying causal relationships among different abundance metrics, and O'Dwyer et al. (2009) derived the relationships that arise for a neutral model with size-dependent demography. We will return to this topic briefly at the end of the chapter.

10.2 State of the field

10.2.1 Species abundance models

Species abundance models may be theoretical (i.e. derived explicitly from assumptions about the biological factors that generate variability in abundances) or phenomenological (i.e. chosen because they appear to resemble empirical distributions). Stochastic abundance models, such as neutral models (Bell 2000; Hubbell 2001) and the environmental stochasticity models of Engen and Lande (1996a,b) are examples of the former, while the logit-normal distribution (Williamson & Gaston 2005) is an example of the latter. However, most models incorporate both theoretical and phenomenological elements. For instance, niche apportionment models are derived by assuming that a species' abundance is proportional to a randomly determined share of resources along one or more niche axes (Tokeshi 1990). Such models reflect the biological reasoning that species differences in access to resources

EBSCOhost - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM. All use subject to https://www.ebsco.com/terms-of-use

are the principal drivers of species abundances, and reflect ecological concepts such as pre-emption of niche space. However, a species' share of the available resource pool is either assumed fixed or chosen randomly from a uniform distribution, even though (to our knowledge) a biological argument for these particular values or distributions has never been advanced in the literature. Similarly, Engen & Lande (1996b) show how a log-normal distribution of species abundances can arise in a stochastic model of community dynamics, but elements of the model (e.g. a log-normal distribution of intrinsic growth rates among species) appear to have been chosen more for mathematical tractability than on biological grounds.

In addition to reflecting the processes that generate the underlying distribution of species abundances in the community, abundance distributions in ecological data sets also reflect the processes by which the data are sampled from the community. Some of the earliest attempts to fit species abundance models to data explicitly considered both of these factors. For instance, Preston (1948) proposed that incomplete sampling would produce a truncated species abundance distribution. Subsequently, the role of incomplete sampling was formalised by mathematical models (e.g. Pielou 1975). Such models typically are compound distributions that account for a species' abundance in the community, and in the sample:

$$\Pr(r, n) = \Pr(r|n)\ \Pr(n) \tag{10.1a}$$

$$\Pr(r) = \sum_{n} \Pr(r, n) \tag{10.1b}$$

Equation 10.1a follows from the definition of conditional probability. The first term is the probability that a species has abundance r, given that its true abundance is n; this quantity depends on stochastic sampling effects that cause observed relative abundance in samples to vary around the species' true relative abundance in the community. The second term is the probability that a species' true abundance is n; this depends on the shape of the abundance distribution in the underlying community, and thus on the biological processes that determine that shape. The overall probability that a species has abundance r in a sample (equation 10.1b) is the sum (or integral, if true abundance is a continuous variable) of equation 10.1a over all possible true abundance values, n.

Models of the form given by equation 10.1 reveal that incomplete sampling can cause sample abundance distributions to differ markedly from the underlying community abundance distributions. These effects are somewhat more complicated than simple truncations of the true distribution, as proposed by Preston (1948), but they often qualitatively resemble such truncated distributions (Fig. 10.2). Most such models assume that the data are a random sample from the larger community (e.g. $\Pr(r|n)$ follows a Poisson distribution with a mean proportional to n), but some explicitly consider the effect of non-random sampling of species' abundances. That is, encounter rates for a species during ecological sampling may not be proportional to their true abundances in the community if species differ in detectability or in spatial distribution relative to the sampling scale (Chapter 3). For instance, Engen et al. (2002) examine changes in the shape of species abundance distributions when such non-randomness follows a Poisson log-normal or negative binomial distribution, rather than a Poisson distribution (also see Green and Plotkin (2007)).

10.2.2 Obtaining predicted abundances

Fitting a species abundance model to a sample of species abundances requires choosing a set of parameter values from their ranges of possible values. This requires a method for generating model predictions for each set of parameter values to be considered, and a statistic that quantifies the model's agreement with the data for those parameter values. For many species abundance models, predictions can be generated exactly, by plugging parameter values into a closed-form expression. For example, for a random sample from a gamma distribution of species abundances (Pielou 1975), or as a result of demographic stochasticity with a constant rate of immigration (Volkov et al. 2007), the probability that a species has abundance r follows a negative binomial distribution:

$$p_r = \frac{\Gamma(r+k)}{r!\Gamma(k)} \frac{m^k}{(1+m)^{r+k}} \tag{10.2}$$

EBSCOhost - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM. All use subject to https://www.ebsco.com/terms-of-use

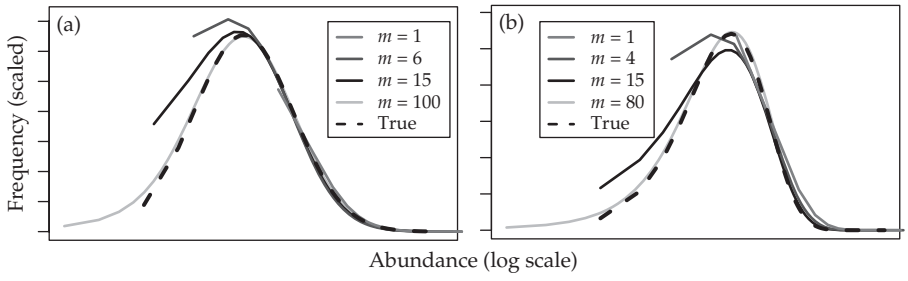

Figure 10.2 Illustration of veil-like effect of incomplete sampling on species abundance patterns. Dashed lines show (a) log-normal and (b) gamma relative abundances for an underlying community; the shaded lines show the (a) Poisson log-normal and (b) negative binomial species abundance pattern expected when sampling from these communities at different levels of intensity. Following Lande et al. (2003), each curve has been rescaled and shifted along the horizontal axis to highlight that, while the pattern resembles qualitatively an 'unveiling' of the underlying abundance distribution, the curves do not superimpose perfectly.

where m and k are the *scale* and *shape* parameters of the underlying gamma distribution.

In other cases, a mathematical expression for a model prediction exists, but a numerical approximation is required to generate a prediction. For instance, the Poisson log-normal distribution arises from Poisson sampling of individuals from a log-normal distribution of species abundances:

$$p_r = \int_{\lambda=0}^{\infty} \left[\frac{\exp(-\lambda)\ \lambda^r}{r!} \right] \left[\frac{1}{\lambda\ \sigma\ \sqrt{2\ \pi}} \exp\left(-\frac{1}{2\ \sigma^2} \left[\ln\left(\frac{\lambda}{m} \right) \right]^2 \right) \right] d\lambda$$

$$= \frac{1}{r!\ \sigma\ \sqrt{2\ \pi}} \int_{\lambda=0}^{\infty} \lambda^{r-1} \exp\left(-\lambda - \frac{1}{2\ \sigma^2} \left[\ln\left(\frac{\lambda}{m} \right) \right]^2 \right) d\lambda \qquad (10.3)$$

where $\ln(m)$ and σ are the mean and standard deviation of the natural logarithm of abundance, respectively. The first square brackets in the top line of equation 10.3 say that the probability that a species has abundance r in the sample is Poisson with rate parameter λ, and the second square brackets say that the rate parameter λ, which is proportional to relative abundance in the community, follows a log-normal distribution. In this case, the integral cannot be solved explicitly, so numerical methods for approximating its value must be applied. In some cases, the numerical calculations can be done (with the aid of a computer) almost as quickly as evaluating equation 10.2. In other cases, however, the necessary calculations are more cumbersome and can substantially slow down the process of parameter estimation because they may need to be made repeatedly for a large number of different combinations of parameter values.

Some models require stochastic simulation to generate model predictions. Most niche apportionment models fall into this category, as do some stochastic abundance models that do not have analytical solutions (e.g. Bell 2001). This greatly increases the amount of computational time required to generate model predictions. Because each simulated model outcome is subject to stochastic variation, a large number of simulations must be conducted and generally averaged in some way to produce the model's prediction. However, even this averaging process does not completely eliminate stochastic error in model predictions. When the model has parameters that are estimated from data, this complicates the process of finding the parameter values that yield the best fit to the data, as we explain below.

10.2.3 Choosing parameters

Determining how well a model can explain patterns in empirical data often requires that parameter values be chosen from a range of possible values, but this is not always the case. For instance, under the dominance pre-emption species abundance model, the proportion of the available resource pool

EBSCOhost - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM. All use subject to https://www.ebsco.com/terms-of-use

monopolized by one species is drawn from a uniform distribution: all proportions between 0.5 and 1 are equally likely (Tokeshi 1990). A second species then consumes a similarly determined proportion of the remaining resources, and so forth. There is no parameter that can be varied to make some resource proportions more likely than others.

For models that do have parameters, values for those parameters generally are estimated directly from the data being compared with the model. Several different approaches have been used to estimate parameters, which we review below. Regardless of the particular method, however, such estimates are based on some statistic that quantifies the discrepancies between data and model predictions for a particular set of parameter values. The values of model parameters are then changed until the value of this statistic is minimized. When a model has only one or two parameters, best-fit values can usually be approximated by brute force: calculating the model fit statistic for a range of parameter values and finding the statistic's minimum value (e.g. Fig. 10.3a). However, when calculating model predictions or model fit statistics is computationally demanding, or when many parameters must be estimated, this approach can be impractical; therefore, contemporary applications typically rely on optimization algorithms to find efficiently the maxima or minima of functions. These algorithms start from an initial set of parameter values and change those values systematically until further changes cease to improve the fit. This is why generating models by stochastic simulation can cause difficulties in parameter estimation: the model fit statistic will be subject to small ups and downs due to stochastic simulation error, and this can cause optimization algorithms to fail (Fig. 10.3b). There is no readily available fix for this problem: researchers must check that any parameter estimates obtained genuinely do give the best fit, for instance by plotting goodness-of-fit profiles (Hilborn & Mangel 1997). Failure to find the best fit can also occur when there are multiple 'local' peaks in model fit, meaning any small movement away from a combination of parameter values causes a reduction in model fit, even though a better fit is possible for a very different set of parameter values (see Etienne et al. (2006) for an example in species abundance

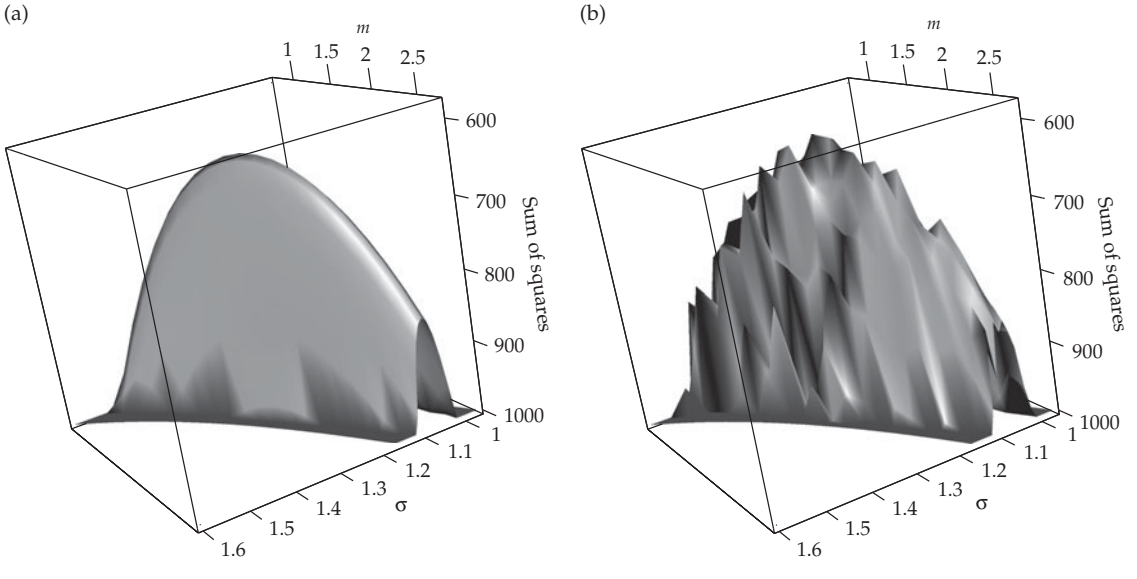

Figure 10.3 (a) A surface plot of the negative sum of squares for the Poisson log-normal, fitted to rank abundance data simulated from a Poisson log-normal distribution with $m = 1$, $\sigma = 1.5$ and $S = 200$. (b) A surface plot of the same model fitted to the same simulated data, but with the expected rank abundance distribution for each set of parameter values generated by averaging the rank abundance distributions for 1000 simulated data sets using those parameter values.

EBSCOhost - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM. All use subject to https://www.ebsco.com/terms-of-use

analysis). There is no foolproof way to avoid this, but it is good practice to re-start optimization algorithms from a range of different initial parameter values, to increase the likelihood that all of the local peaks in the model fit are found.

We now turn to several common statistics used in parameter estimation.

Maximum likelihood for species abundances

Contemporary studies frequently fit species abundance models by maximum likelihood methods. This requires deriving, or numerically approximating, the probability that a particular set of species abundance values will be observed, given specific values of model parameters. For instance, equation 10.2 gives the probability that a species will have abundance r in a sample, given that species abundances follow a negative binomial distribution with parameters m and k. The support that an observation provides for a particular set of parameter values, called the *likelihood*, is proportional to this probability:

$$\mathcal{L}(\theta \mid r) \propto \Pr(r \mid \theta) \tag{10.4}$$

where θ is a vector of all of the model's parameter values (e.g. $\theta = [m, k]$ for the negative binomial) and r is an observed species abundance. Note that for a species to appear in the likelihood, it must appear in the sample. If one is fitting a model in which some species in the underlying community may not appear in the sample, such as the Poisson log-normal or negative binomial, then the *zero-truncated* form of the model accounts for this:

$$\Pr(r \mid \theta) = \frac{p_r}{1 - p_0}, \tag{10.5}$$

where p_r is the probability that a species has abundance r (e.g. as in equations 10.2 and 10.3). Thus, the denominator is the probability that the species does not have abundance zero, and $\Pr(r|\theta)$ is the probability that a species has abundance r, *given* that it has appeared in the sample at least once. Most commonly, equation 10.5 is extended to account for the entire sample by making the simplifying assumption that species' abundances are statistically independent of one another, and thus the likelihoods for the individual species are multiplied together:

$$\mathcal{L}(\theta|\boldsymbol{r}) = \frac{S_{obs}!}{\prod_{n=1}^{N} \phi_n!} \prod_{s=1}^{S_{obs}} \mathcal{L}(\theta|r_s) \tag{10.6}$$

Here, S_{obs} is the total number of species in the data set, N is the total number of individuals sampled and ϕ_n is the number of species with abundance n. The fraction is a normalizing constant to account for the number of different ways that the observed species abundances can be divided up among the S_{obs} species sampled (usually ignored, because for any given sample of species abundances, it is independent of the model or parameter values).

In reality, species abundances will not be statistically independent of one another. For instance, in sampling a given area for sessile organisms, if one species is particularly abundant, then there is less space available for another species. One way to account for such constraints is to condition on the total number of individuals that are sampled. In other words, *given* that one has observed a particular set of species abundances that add up to a total sample size of N, what is the likelihood for a particular set of parameter values? This requires normalizing (i.e. dividing) by the probability that the species abundances add up to the observed sample size. Calculating this quantity can be very time-consuming because it requires calculating the likelihoods for all of the possible combinations of S_{obs} species abundances that give a total sample size of N, and adding them up (Etienne & Olff 2004).

Because accounting for such conditioning is potentially cumbersome, it is worth evaluating the robustness of parameter estimates obtained, using the simpler likelihood (equation 10.6), to violation of this independence assumption. To explore this, we compared the bias of parameter estimates obtained using this likelihood when data are simulated in two different ways: in the *Poisson* algorithm, individuals (and thus species) are sampled independently of one another, consistent with equation 10.6; in the *hypergeometric* algorithm, individuals are sampled from the community only until a fixed total sample size is reached (see Box 10.1 for details). We found no differences in the bias of parameter estimates for the Poisson log-normal or

EBSCOhost - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM. All use subject to https://www.ebsco.com/terms-of-use

negative binomial models (Fig. 10.4): Wilcoxon tests for paired observations confirm no significant differences in bias for any parameter values, for either model ($P > 0.4$ in every case).

Box 10.1 Analysis of parameter bias due to fixed sample size

For the Poisson log-normal distribution, we first simulated an underlying distribution of species abundances in the community for a particular total number of species, S_{true}, and with a particular standard deviation of log abundance, σ. For the hypergeometric algorithm, we sampled N individuals from this community abundance distribution, where the probability that an individual belonged to species i, p_i, was equal to the species' relative abundance in the community. We then fit the Poisson log-normal distribution to the resulting simulated sample abundance distribution, using the likelihood in equation 10.6. For the Poisson algorithm, for each species in the underlying community, we generated its numerical abundance in the sample by choosing a random number from a Poisson distribution with mean p_iN; thus, total sample size was a Poisson random variable with mean N. Note that, in both algorithms, the number of observed species in the sample can be (and generally is) less than S_{true} because some species have abundance zero. For each bootstrap algorithm, we repeated this simulation 1000 times for each of 27 combinations of parameter values and sample sizes: σ = [1,2,3], N = [100,1000,10 000] and S_{true} = [20,200,2000]. We calculated the mean standardized bias as the difference between the mean parameter estimate (across the 1000 simulations for a particular parameter combination) and the true value, divided by the standard deviation of parameter estimates among those simulations.

For the negative binomial, we followed the same procedure, except that the underlying community abundance distribution was gamma, rather than log-normal, and the parameter and sample size combinations were k = [0.01, 0.1, 1], N = [1000, 10 000, 100 000] and S_{true} = [20,200,2000]. The larger values of N are required for this model to ensure that sufficiently many species appear in the sample for adequate converge of parameter estimation algorithms.

See Connolly et al. (2009) for a detailed description of all algorithms.

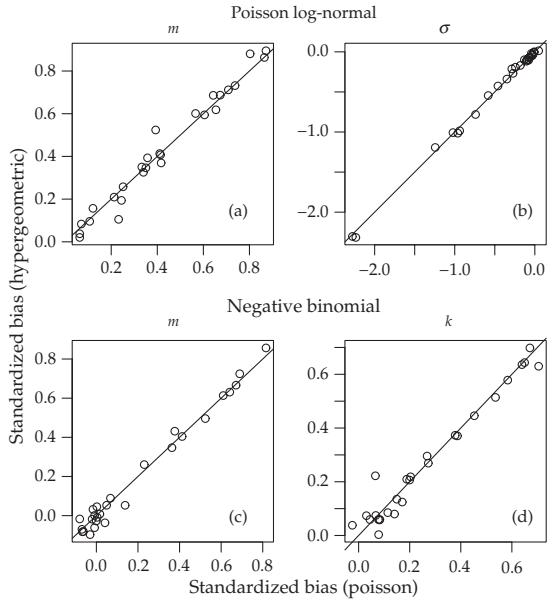

Figure 10.4 Comparison of the standardized bias of parameter values for data simulated according to the hypergeometric (fixed sample size) and Poisson (variable sample size) bootstrap algorithms. Each point represents the mean standardized bias for one of 27 parameter combinations for (a, b) the Poisson log-normal and (c, d) the negative binomial. See Box 10.1 for a detailed description of the bootstrap simulations. The large standardized biases (i.e. magnitudes > 1, lower left of panel b) correspond to cases with $\sigma = 3$ and sample sizes smaller than total species richness: in these cases, the proportion of the total number of species that actually appeared in the simulated sample was very small.

These analyses suggest that failing to condition on sample size is unlikely to substantially affect the biases in parameter estimates, at least for these two models. Moreover, it illustrates how this problem can be examined for other models.

Maximum likelihood for abundance classes

Several analyses first place species' abundances into abundance classes and then use maximum likelihood methods to fit model predictions of the number of species in each abundance class against the corresponding observed values (e.g. Hubbell 2001). In general, this approach is likely to be inferior to using the actual abundance values because it removes information from the data that can provide important information about model fit. Because such abundance classes are usually logarithmic in nature (i.e. they get wider as abundance increases),

EBSCOhost - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM. All use subject to https://www.ebsco.com/terms-of-use

this loss of information will be particularly pronounced for the most abundant species. This is potentially quite important because the thickness of the tail of highly abundant species is one of the major differences between species abundance models. For these reasons, fitting models to unbinned data should tend to lead to narrower confidence limits on model parameters (and for similar reasons should lead to stronger goodness-of-fit tests and model selection results). However, we have no particular reason to suspect that parameter estimates would tend to be biased when data are binned. Consequently, we suspect that the use of abundance classes probably entails weaker statistical inferences, rather than erroneous inferences.

There is one circumstance in which use of abundance classes for model fitting may be preferable and that is when model predictions must be generated by stochastic simulation (e.g. Hubbell 2001). To calculate likelihood, each observed species abundance must have a corresponding model prediction for the probability that a species has that abundance. When model predictions are generated by simulation, such probabilities must be approximated by running a large number of simulations for a particular set of parameter values and then calculating the average proportion of species that ended up with that abundance in those simulations. However, when the true probability that a species will end up with a particular abundance value is very low, stochastic variation among ensembles of simulations can have a substantial effect on parameter estimation. For instance, substantially different likelihoods may be obtained, depending on whether one, two or three out of 1000 simulations produce a species with a particular species abundance that is observed in the data. Making predictions about abundance classes instead of specific abundance values will tend to make the outcomes of simulations more consistent with one another across replicate runs and therefore stabilize parameter estimates.

Maximum likelihood for rank abundances

Rank abundance distributions differ from normal species abundance distributions by associating a rank with each species' abundance (i.e. the most abundant species has rank 1, next most abundant has rank 2, etc.). This introduces an additional layer of statistical complexity on top of those relevant to unranked abundance distributions: a species with abundance rank j cannot be more abundant than the one with rank $j-1$, and cannot be less abundant than the species with rank $j+1$. We know of only one maximum likelihood approach to rank abundance distributions (Foster & Dunstan 2009). This approach conditions on the total number of species, the total sample size and the abundances of all of the species with higher rank:

$$\Pr(n_j) = \Pr(N)\ \Pr(S|N)\ \Pr(n_1|S, N)\ \Pr(n_2|S, N, n_1) \ldots \Pr(n_S|S, N, n_1, n_2, \ldots, n_{S-1}). \quad (10.7)$$

In cases where the probability distribution of abundances, for a species with a particular abundance rank, has a simpler form than the full species abundance distribution itself, this may be a more tractable approach to parameter estimation than one based on fitting to the frequency distribution of species abundances. For instance, under the geometric series model, the abundance of a species is drawn from a probability distribution that depends on the abundances of all species with higher abundances, but not on species with lower abundances. Such models lend themselves naturally to approaches like this one. However, for many models (e.g. the Poisson log-normal, as far as we are aware), no likelihood of the form of equation 10.7 has been formulated for the rank abundance distribution.

Least-squares approaches

Least-squares approaches are often used as an alternative to maximum likelihood in the literature. Most commonly, these methods are applied to rank abundances by finding the parameter values that minimize

$$\sum_{j=1}^{S} \left(n_j - \widehat{n_j}\right)^2, \quad (10.8)$$

where n_j is the observed abundance of the species with rank j, and $\widehat{n_j}$ is the predicted abundance of a species with abundance rank j (Loehle & Hansen 2005; Woodcock et al. 2007). However, least-squares methods have also been used to fit species abun-

EBSCOhost - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM. All use subject to https://www.ebsco.com/terms-of-use

dance data categorized into abundance classes (Gray et al. 2005; Pueyo 2006; Volkov et al. 2007).

There are reasons to treat least-squares fitting of species abundance data with considerable caution, particularly when applied to rank abundance data. The statistical theory that supports least-squares estimation holds if three conditions are met: observed values are statistically independent of one another, the variance of this distribution is the same for all predicted values, and the residual variation has a mean of zero. For rank abundance data, the first two of these assumptions are certainly violated, often quite severely, and there is no particular reason to expect the third assumption to be met either. Specifically, every rank abundance is constrained to be no larger than the abundance of the next more highly ranked species, and no smaller than the next lower-ranked species. These constraints make rank abundances highly non-independent. Moreover, the width of this range of possible abundance values tends to narrow substantially from the most abundant to the least abundant species, so the residual variation in abundances will also change substantially as a function of rank. For instance, many species abundance distributions have a large number of 'singleton' species (species represented by only one individual) in an assemblage. In such communities, the lowest-ranked species will be guaranteed to have an abundance of exactly one individual because the next-higher ranked species will also have an abundance of one.

For least-squares fits to abundance classes, the situation is somewhat different. The 'observed' and 'predicted' values are not the abundances of species with particular ranks, but rather the numbers of species falling in a particular abundance class. For this latter case, any statistical non-independence among species is likely to be weaker. Moreover, under random sampling, the stochastic variability around the number of species in an abundance class should approximately follow a Poisson distribution. Provided the predicted number of species in each abundance class is large enough (around five species or more), this Poisson variability should closely approximate a normal distribution. This does not resolve the problem of unequal variances, which violates an assumption of the least-squares approach. However, because the variance and mean are equal for Poisson random variables, the residual variance can be homogenized by dividing each residual by the corresponding expected mean value. This approach is implicit in approaches that minimize Pearson's X^2 statistic (e.g. Doroghazi & Buckley 2008):

$$X^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}, \tag{10.9}$$

where O_i and E_i are the observed and expected numbers of species in abundance class i, respectively. Thus, parameter estimation by minimizing X^2 should have similar properties to maximum likelihood fitting to abundance classes, provided that expected values are not too small.

Separately estimated parameters

Although most species abundance analyses estimate parameters by finding the values that maximize model fit, the parameters of some models can be estimated independently. Generally, this requires that the parameter represent a directly measurable biological quantity. For example, most neutral models explicitly include a parameter to characterize the proportion of new recruits in a local community that arrive from elsewhere (e.g. the migration parameter m; Hubbell (2001)). If there is empirical information about dispersal in a group of species (e.g. the shape of the dispersal kernel), then m can be estimated based on this information (Etienne 2005). Because this approach constrains the possible shapes that a model can take, it makes good agreement between a model and data less likely to occur by chance, and thus more likely to indicate that the model does, in fact, approximate the ecological processes that gave rise to the data. The closest example of this of which we are aware is a study of intertidal dynamics, in which mortality and immigration probabilities were estimated from observed transitions in space occupancy from experimental plots, then used to predict the species abundance distribution from independent transects at the same site (Wootton 2005).

10.2.4 Goodness-of-fit testing

Goodness-of-fit testing involves determining whether a model, once it has been fitted to data,

EBSCOhost - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM. All use subject to https://www.ebsco.com/terms-of-use

adequately captures the patterns in those data. Often, in species abundance analysis, such assessments are made subjectively, based (for instance) on the apparent visual concordance between observed and predicted values, or on the percentage of variance explained by a model. Formal goodness-of-fit tests determine whether the model's fit to the data is significantly worse than one would expect if the model were true, and, if so, how much worse. In species abundance analysis, such tests have included several conventional tests for frequency distributions (e.g. Kolmogorov–Smirnov, Pearson's X^2 or G-tests). More recently, however, computationally intensive methods have been applied that appear to be better at identifying lack of model fit. The last few years have also seen techniques for unpacking information about model fit to determine how model predictions depart systematically from empirical data. Finally, although not goodness-of-fit tests *sensu stricto*, model fit also can be evaluated by making predictions about patterns other than those used in the fitting procedure. This latter approach is a way of moving beyond a model's ability to describe the data to gain insight into whether the processes that generate the patterns in the data are consistent with those included in the model.

Goodness-of-fit tests are null hypothesis tests: they determine whether or not one can reject the hypothesis that the data were actually generated by the specified model. However, there is a growing recognition among ecologists that models are, by definition, idealizations of nature and therefore that all models can be rejected, given sufficient data and a powerful enough goodness-of-fit test. This recognition has been partly responsible for sparking interest in model selection statistics, which aim to compare model fit relative to other models, rather than relative to the possibility that the model is somehow 'true', as in goodness-of-fit tests. However, the way in which a model fails to fit empirical data can sometimes suggest particular processes, omitted from the model, that may be important in nature (e.g. Dornelas et al. 2006). In addition, some model selection statistics are invalid if none of the alternative models provide a good approximation to the data. Finally, a model's utility for predicting unobserved data depends critically on how well it characterizes the true underlying distribution. Consequently, the analysis of goodness of fit remains important in the study of species distributions, but there is an emerging shift away from statistical significance per se, and towards quantifying the magnitude of lack of fit and pinpointing the particular features of the data that are poorly approximated by a model.

Subjective methods

Graphical assessment of model fit goes back to the earliest analyses of species abundance distributions (e.g. Fisher et al. 1943; Preston 1948) and is still widespread in the species abundance literature (Ford & Lancaster 2007; Engen et al. 2008). Visual assessments are important for identifying what it is about the data that is responsible for their lack of agreement with a model, but they can be uninformative or even misleading when used in isolation, as can other ad hoc methods such as inspecting r^2 values. The predictions of a species abundance model, except as an extraordinary coincidence, will never match an empirical pattern exactly. Such discrepancies may be due to stochastic effects that are consistent with the model. For instance, a model may predict that the most abundant species includes 60% of the individuals in a community, and this may, in fact, be true of the actual community, but the species may be slightly over- or under-represented in the sample being analysed. Alternatively, discrepancies may be due to the fact that the model poorly approximates the 'true' species abundance distribution. Subjective methods cannot be used to distinguish rigorously between these two possibilities.

Asymptotic tests

Goodness-of-fit tests compare a model fit statistic with a null distribution. If the data are consistent with the model, the model fit statistic should be a random draw from that distribution. Statistical theory has been used extensively to identify statistics whose null distributions have the same basic shape regardless of details of the model, such as parameter values. Several such test statistics have been used in the analysis of species abundance data. For instance, the log-normal and logit-normal distributions have been tested by conducting standard

EBSCOhost - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM. All use subject to https://www.ebsco.com/terms-of-use

normality tests on transformed abundances (Connolly et al. 2005; Williamson & Gaston 2005); however, this approach is useful only for distributions that can be made Gaussian by transformation. More general tests involve the use of contingency tables comparing observed and predicted frequencies of species in different abundance classes via Pearson's X^2 or similar statistics (Yin et al. 2005; Caruso & Migliorini 2006; Pueyo 2006; Sitran et al. 2009), for which the null distribution is chi-squared. Other analyses have used the Kolmogorov–Smirnov test, but this test is very weak when model parameters are estimated from data, as in most species abundance analyses (Zar 1996).

Parametric bootstrapping

Asymptotic tests are based on mathematical approximations of what the distribution of a particular statistic should be, given particular assumptions about the data. Often, these simplifying assumptions are not well met. For example, the chi-squared null distribution is appropriate only when there are minimum frequencies in each abundance category, so, in practice, species must be placed in abundance classes that often span a large range of abundance values. This removes information from the data and can lead to tests with low statistical power. An alternative approach is to estimate the null distribution directly, by simulating data sets that accord with the assumptions of a particular fitted model and then computing a model fit statistic for each simulation. The frequency distribution of this test statistic across simulations is then used as the null distribution for the model's fit to the empirical data. Such approaches have been termed 'parametric bootstrapping' because one re-samples data from a fitted model to estimate the uncertainty distribution of a statistic, rather than re-sampling from the data itself (Efron & Tibshirani 1993). The disadvantage of this approach is that this null distribution is sample and model specific, and so must be conducted anew for each model and data set under investigation. However, advances in computing power are making such analyses increasingly practical, and its versatility makes it a compelling alternative to more traditional approaches. In biology, for instance, parametric bootstrapping is becoming commonplace in capture–mark–recapture analysis (White et al. 2001) and phylogenetics (Holmes 2003).

In the analysis of species abundances, parametric bootstrapping has been used relatively infrequently, but several different test statistics have been proposed. Tokeshi (1990) first proposed testing model fit by simulation by comparing empirical rank abundance data with comparable data simulated from niche apportionment models, and Bersier and Sugihara (1997) subsequently proposed a test statistic for such analyses. Volkov et al. (2003) and Etienne et al. (2006) use the maximum log-likelihood of the model as the test statistic. Connolly et al. (2005; 2009) used deviance, a test statistic based on the log-likelihood, but normalized to facilitate comparability across sites with different sample sizes and parameter values.

Sample size: species or abundance?

In most species abundance analyses, goodness-of-fit tests make an implicit assumption that species abundance values are sampled, rather than individuals. For instance, the null distribution of Pearson's X^2 test statistic is based on an assumption that the stochastic variability around predicted frequencies of species in an abundance category follows a Poisson distribution. This assumed variability corresponds to a situation in which a certain number of species are sampled, each of which has an abundance label attached to it. Similarly, in some parametric bootstrap analyses, one samples an abundance value for each species observed in an empirical sample (Diserud and Engen 2000; Connolly et al. 2005; Engen et al. 2008). However, in most ecological sampling, a particular area or volume of habitat is sampled, and this sample contains some number of individuals, which in turn have species identities. In particular, a replicate sample might not yield the same number of species because some species that are present in the community do not appear in a given sample, and that number is subject to stochastic sampling variation.

It is possible to develop parametric bootstrap algorithms that simulate this process of re-sampling more realistically than the species sampling approach assumed by most goodness-of-fit tests. Etienne (2007) provides such an algorithm for a neutral model, for which dispersal limitation

EBSCOhost - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM. All use subject to https://www.ebsco.com/terms-of-use

determines a species' presence and abundance in a sample, and Connolly et al. (2009) provide an algorithm for models that assume random sampling from an underlying community abundance distribution (e.g. Poisson log-normal, negative binomial, etc.). These algorithms lead to goodness-of-fit statistics whose null distributions are substantially narrower, and therefore more powerful at detecting lack of model fit, than analogues based on species sampling assumptions (Connolly et al. 2009: also see Box 10.1). Such differences suggest that, in general, goodness-of-fit tests based on the sampling of individuals may be more powerful at detecting departures from species abundance models than approaches that implicitly or explicitly assume sampling of species.

Multi-pattern testing

Most assessments of model fit compare a model's predicted values with the same data that were used to fit the model. Such assessments tell us whether the model adequately describes the species abundance data. However, frequently the goal of species abundance analysis is to go beyond this and determine whether that good fit indicates that the model approximates well the biological processes that generated the data. One way to do this is to test whether the parameters implied by fit to species abundance data can also explain other characteristics of the assemblage being investigated (McGill 2003c). Most examples of this approach have involved tests of neutral models. For instance, Adler (2004) fitted the neutral model to species abundance distributions of local grassland assemblages, used those parameter estimates, to determine the species–area relationship implied for that assemblage and then compared that prediction with the empirical species–area relationship. Wootton (2005) estimated neutral model parameters from temporal transitions in space occupancy in a benthic community and then tested the model's ability to predict community change when a hypothesized competitive dominant was removed. Dornelas et al. (2006) fitted the neutral model to species abundance distributions of scleractinian reef corals, predicted the frequency distribution of community similarity implied by those estimates, and compared that with the community similarity distribution exhibited by the data.

10.2.5 Model selection

In species abundance analysis, we often wish to know which of several alternative models provides the best approximation for the data, in addition to (or instead of) assessing their absolute goodness of fit. Many studies have compared graphically empirical data with the predictions of different species abundance models. For instance, many of the early analyses that suggest multimodality in species abundance distributions were based on graphical inspection of species abundance distributions (Gray & Mirza 1979; Ugland & Gray 1982). Other approaches are essentially comparisons of goodness-of-fit statistics for different models, such as r^2 values calculated from rank abundance distributions, Pearson's X^2 statistics calculated from frequency distributions of species abundance or statistics calculated from parametric bootstrapping (McGill 2003b; Volkov et al. 2007; Harte et al. 2008; Connolly et al. 2009).

Although approaches like those just described can be informative, they have important limitations. Alternative models may differ in how much uncertainty is associated with their predicted abundance patterns. Models with very flexible forms due to, for example, large numbers of parameters, are more likely to be able to provide good fits to data, even if they would do a poor job of predicting what a new sample might look like. In addition, the model with the lowest r^2 value, or which appears visually to give the best fit, may depend critically on the scale on which species abundances are represented. For instance, for rank abundances considered on an arithmetic scale, apparent model fit will be much more dominated by the model's fit to highly abundant species, compared to when rank abundances are represented logarithmically. Moreover, different models make different assumptions about the sources of variability in species abundance data and thus differ in terms of how much a given discrepancy between the model and data is consistent with a model's assumptions.

To overcome the subjectivity inherent in such ad hoc model selection approaches, biologists in many fields are increasingly shifting towards the use of model selection statistics that have a stronger foundation in statistical theory. Although this practice

EBSCOhost - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM. All use subject to https://www.ebsco.com/terms-of-use

has not yet become widespread in the analysis of species abundance data, several model selection statistics have been used, including Akaike's information criterion (AIC; Connolly et al. 2005), Bayes factors (Etienne & Olff 2005), Bayesian (or Schwarz) information criterion (BIC; Dornelas & Connolly 2008) and deviance information criterion (DIC; Golicher et al. 2006; Mac Nally 2007). Of these statistics, AIC and BIC are straightforward to calculate, once one has maximum likelihood estimates for parameters, whereas calculation of Bayes factors and DIC are more complicated, sometimes considerably so, because they incorporate prior beliefs about model parameter values. All such model selection statistics, however, quantify the trade-off between the increased uncertainty associated with more complex models and the increased bias (i.e. systematic discrepancies between model predictions and data) associated with simpler models. Moreover, because they are based on likelihood, discrepancy is measured on a natural scale, and thus these statistics do not have the arbitrariness of, for instance, choosing whether to plot rank abundances arithmetically or logarithmically.

Akaike's information criterion

AIC estimates the amount of information that is lost when a model is used as an approximation for the true distribution from which the data have been drawn (Burnham & Anderson 1998). Thus, it provides a relative measure of a model's ability to predict new data sampled from the same true distribution. In many ecological applications, where sample sizes are relatively small, a modified form of this criterion, called AIC_c, is used:

$$AIC_c = -2log(\mathcal{L}_{\max}) + 2k + \frac{2\,k\,(k+1)}{n-k-1} \qquad (10.10)$$

where $\mathcal{L}_{\max}$ is the model's maximum likelihood, k is the number of parameters in the model, and n is the sample size. The first term measures the model's lack of fit to the data in hand, while the second two terms account for the fact that models with more parameters tend to predict the values of new data with greater uncertainty. Notice that the third term becomes very small as sample size (n) gets large, so that, for moderately large sample sizes, AIC depends only on the model's overall fit and the number of model parameters. Uncertainty about which model is best can also be estimated with AIC by calculating Akaike weights, according to which the probability that a model is the best model is proportional to $e^{-\mathrm{AIC}/2}$.

Like all model selection statistics, AIC is controversial. Proponents of AIC typically emphasize the fact that it is an estimate of the information lost when the model is used as an approximation for the true distribution from which the data come, and thus is defensible on objective grounds as an aim of model selection. Critics of AIC most frequently cite the fact that it is not *consistent*, which means that, when one of the models under consideration was actually used to generate the data, the probability that AIC selects this true model does not approach 100% as sample size increases. For extensive discussion of the merits and shortcomings of AIC, readers should consult the relevant statistical literature directly (see Burnham & Anderson (1998) and Taper and Lele (2004) for ecologically oriented discussion).

Finally, it is worth noting that the derivation of AIC at several points uses the model as an approximation for the true distribution from which the data come, and therefore AIC is usually recommended only for models that are good approximations to the truth. Thus, goodness-of-fit testing is an indispensable part of model selection using AIC.

Bayesian approaches

To understand Bayesian approaches to model selection, it is important to recognize that all such approaches involve the application of Bayes' theorem to calculate probability distributions for model parameters:

$$p\,(\boldsymbol{\theta}|y, M) = \frac{\Pr\,(y|\boldsymbol{\theta}, M)\;\; p\,(\boldsymbol{\theta}|M)}{c} \qquad (10.11)$$

$\Pr(y|\boldsymbol{\theta}, M)$ is equal to the likelihood: the probability of observing the data y, if the model, M, with a particular set of parameter values, $\boldsymbol{\theta}$, were true. $p(\boldsymbol{\theta}|M)$ is the probability distribution for the parameters, prior to the data being collected (the *prior*). This prior may be based on previously available independent data, it may reflect subjective belief or (most commonly) is chosen so that its effect on $p(\boldsymbol{\theta}|y, M)$, called the *posterior* distribution, is very

EBSCOhost - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM. All use subject to https://www.ebsco.com/terms-of-use

small relative to the likelihood. The posterior distribution is an updated probability distribution for the model parameters, in light of the data. Thus, if a particular set of parameter values produces an extremely good model fit compared to another set of parameter values, then the first set will have a much larger posterior probability, relative to the second set, than in the prior distribution. c is a normalizing constant chosen so that the posterior distribution, $p(\boldsymbol{\theta}|y, M)$, is a probability distribution (i.e. so that integrating $p(\boldsymbol{\theta}|y, M)$ over all possible combinations of parameter values yields 1.0). One can think of the prior and posterior distributions as estimating the probability that a particular set of parameter values are the 'true' ones. Alternatively, they may be interpreted as estimating the probability that a particular set of parameter values are the ones that yield the best approximation for truth that is possible for a particular model (sensu AIC; see, for example, Spiegelhalter et al. (2002)).

The most intuitive Bayesian model selection statistic is the Bayes factor (also called the posterior odds ratio) and it is applicable where only two competing models are under consideration:

$$B = \frac{\Pr(y|M_1)\ \Pr(M_1)}{\Pr(y|M_2)\ \Pr(M_2)} = \frac{\left[\int_{\theta_1} \Pr(y|\theta_1, M_1)\ p(\theta_1|M_1)\, d\theta_1\right]\ \Pr(M_1)}{\left[\int_{\theta_2} \Pr(y|\theta_2, M_2)\ p(\theta_2|M_2)\, d\theta_2\right]\ \Pr(M_2)} \tag{10.12}$$

$\Pr(y|M_1)$ is the probability of observing the data, y, if model M_1 were true. To calculate this, we must first calculate the probability of observing the data, given this model and a particular set of parameter values (θ_1) for that model: $\Pr(y|\theta_1, M_1)$. Then, we average this probability over all of the possible sets of parameter values for the model, weighted according to the prior probabilities for the model parameters. That weighted average is the integral in the numerator and it, in turn, is multiplied by $\Pr(M_1)$, which is the prior probability for the model. Often this quantity is assumed to be equal for all competing models, in which case the Bayes factor just estimates the support provided by the data for the competing models, analogous to a traditional likelihood ratio. In practice, the integrals in equation 10.12 are very hard to calculate, so Bayes factors are typically only used where a computationally intensive numerical technique, Markov Chain Monte Carlo (MCMC), is applied; this approach can be used to estimate such integrals.

The BIC (Schwarz 1978), although derived from Bayes factors, is much simpler to calculate:

$$\text{BIC} = 2\left[-\log(\mathcal{L}_{\max}) + \frac{k\ \log(n)}{2}\right] \tag{10.13}$$

Here, the penalty term for extra parameters, $k\log(n)/2$, arises in the derivation of BIC as an approximation of the difference between a model's maximum log-likelihood and the weighted average likelihood that appears in equation 10.12. In contrast to the penalty term in AIC, it gets larger as sample size increases. In other words, if sample size is large, then a model with more parameters will need to exhibit a larger improvement in fit to be selected as the best model, compared to what is required if sample size is small. Unlike AIC, BIC is consistent: it selects the true generating model with increasing certainty as sample size increases. Thus, if one of the models under consideration is believed to be true, then BIC should be preferred. However, if one instead believes that the true distribution from which the data come is more complex than any of the models being compared, then there is some disagreement about whether or not it is appropriate to use BIC as a model selection statistic (Burnham & Anderson 1998; Spiegelhalter et al. 2002; Boik 2004). An additional caveat with BIC is that its derivation assumes a very large sample size. There is no correction for small sample size, as with AIC, because if sample size is not large, then, in the Bayesian framework, prior beliefs about the values of model parameters have an impact on model selection.

Finally, DIC is a Bayesian approach based on model *deviance* (Spiegelhalter et al. 2002). Deviance estimates lack of fit as the difference between the log-likelihood of a model and the log-likelihood of a hypothetical model that fits the data perfectly. Formally, DIC is:

$$\text{DIC} = D(\bar{\theta}) + \left[\overline{D(\theta)}\ - D(\bar{\theta})\right], \tag{10.14}$$

where $\overline{D(\theta)}$ is an average model deviance, weighted according to the *posterior* distribution of the model

EBSCOhost - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM. All use subject to https://www.ebsco.com/terms-of-use

parameters, and $D(\bar{\theta})$ is the model deviance, evaluated at the posterior mean parameter values. Here, the term in square brackets is analogous to the penalty terms in AIC and BIC because the difference between $\overline{D(\theta)}$ and $D(\bar{\theta})$ will tend to be larger for models with more parameters. When the prior distribution on model parameters is very flat, or there is a very large amount of data, then DIC estimates the same quantity as AIC (Spiegelhalter et al. 2002). However, when parameters are estimated by MCMC methods, DIC can be calculated from the posterior distribution of model parameter values, therefore DIC may be preferable to AIC when the investigator wishes to incorporate prior distributions for parameter values or where the number of parameters or the sample size is not easy to count (as in some hierarchical models and data structures), and thus the value of AIC is ambiguous.

10.3 Prospectus

The analysis of species abundance patterns has been a staple of macroecology for nearly a century. This spans a time period from just after the invention of the likelihood concept, to the invention of computers, of information theory, of the emergence of 'frequentist' and 'Bayesian' approaches as competing schools of thought in statistics, of bootstrapping, and of numerical methods of parameter estimation that make shortcuts such as least-squares unnecessary. Consequently, the tools available for fitting and evaluating models have expanded enormously since geometric, log-normal and log-series distributions were first compared with species abundance data. Having reviewed the range of tools applied in contemporary analysis, we offer some guidelines about how these tools may be utilized productively in the analysis of species abundance patterns and highlight some important areas for further research.

10.3.1 Sampling theory for species abundance models

Sampling distorts the shape of species abundance distributions, often substantially. Therefore, whenever possible, compound distributions that explicitly include both a model for the 'true' underlying species abundance distribution and a model characterizing sampling from that distribution should be used. However, for some species abundance models, such as niche-apportionment models, such compound distributions have not yet been formulated, and this is an important area for further research. In addition, compound distributions that explicitly characterize the effects of local aggregation at the sampling site, or differences in species detectability, on species abundance data have only been explored in a few studies. Most ecologists do not sample randomly at the scale for which they wish to draw inferences; rather, they are more likely to intensively sample at multiple small sites spread over the broader area of interest. When local aggregation occurs, the abundance distributions at study sites will differ from that of the area as a whole. Consequently, an accessible and relatively general framework for non-random sampling is needed.

In addition, most sampling theory developed to date has focused on counts of individuals and thus is appropriate for numerical abundance data. Accounting for sampling effects in other currencies of abundance (e.g. cover, biomass) is generally lacking, but could proceed in at least two ways. One would be to consider sampling effects at the level of individuals, alongside the infinitely many ways that the continuously varying biomasses of individuals could sum up to a particular observed biomass. The other would be to apply a distribution analogous to equation 10.1, where the first term characterizes stochastic variation in biomass sampled given a particular area, volume or total community biomass sampled. However, we know of no theory, or even empirically documented regularities, that might be used to characterize biomass aggregation, as exists for individuals (e.g. Harte et al. 1999; He & Gaston 2003). This, too, stands as an important challenge for future research.

10.3.2 Parameter estimation

We favour approaches to parameter estimation that utilize maximum likelihood methods wherever possible. In our view, fitting models by eye is unnecessarily subjective and, given the accessibility of software to fit frequency distributions, no longer justifiable. Moreover, except where stochastic simulation is used, binning species abundances before fitting models is not necessary

EBSCOhost - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM. All use subject to https://www.ebsco.com/terms-of-use

and probably reduces the precision of parameter estimates. More generally, we are wary of least-squares methods for fitting species abundance models because they are supported by statistical theory only under approximate normality and homoscedasticity. In general, these conditions will not be met for rank abundance data, and will only be met for species abundance distributions when abundances are categorized, probably quite coarsely so for the tail of very highly abundant species, and the data or residual variances are appropriately transformed. Nevertheless, despite its strong theoretical foundations, maximum likelihood is not perfect. In particular, maximum likelihood estimates can be biased, particularly for distributions that, like species abundances, exhibit substantial skew (Fig. 10.4; also see Diserud and Engen (2000) and Connolly et al. (2009)). Consequently, we believe that a rigorous, comparative analysis of different approaches to parameter estimation would be informative.

Many species abundance models allow species to have abundances that can be arbitrarily large, but in real ecological samples, quadrat sizes or transect lengths place upper bounds on the total number of individuals that can appear in a sample. To account for this, some researchers have argued for the use of likelihoods that condition on the total number of individuals actually sampled, rather than treating each species' abundance as completely independent of all others, thus implicitly allowing for unrealistically large or small sample sizes (e.g. Etienne & Olff 2005). Although we have not found this statistical independence assumption to increase the bias in parameter estimates (Fig. 10.4; also see Connolly et al. (2009)), we do not know whether this robustness is generally true for species abundance models or what impact it may have on model selection statistics. Further work on this question may help to identify if, and under what conditions, constraints on total sample size have important effects on the results of species abundance analysis.

For species abundance models that are explicitly derived from some ecological theory (e.g. stochastic abundance models), the use of external independent information about the values of model parameters has great potential to strengthen the inferences that can be drawn from fits to species abundance models (Wootton 2005). This kind of approach can be further developed by explicitly incorporating uncertainty about model parameter values from such independent information. Thus, for a given set of parameter values, there would be two likelihoods: one for the species abundance data, and a second one for the independent data. These likelihoods would be multiplied together to obtain the overall likelihood for the parameters given both data sets. Similarly, in a Bayesian approach, the independent information could be used to construct a prior probability distribution for the model parameter(s). To our knowledge, such approaches have yet to be explored in species abundance analysis.

10.3.3 Goodness-of-fit testing

Failure to find statistically significant lack of model fit indicates only that the investigator has either not sampled enough or has used an insufficiently powerful goodness-of-fit test to detect it. Claims that a model fits data well should be supported by evidence that any lack of fit is small in magnitude, in some meaningful sense, regardless of statistical significance. Similarly, the use of model selection by AIC should be accompanied by evidence that any lack of fit that is present is likely to be small (Burnham & Anderson 1998). Conversely, finding statistically significant lack of fit is a first step to determining what it is about the data that the model fails to capture, and whether this information implicates particular processes, omitted from the model, as important determinants of relative abundance patterns. In this context, we believe graphical approaches to assessing model fit are most useful because they can indicate where discrepancies between models and data are large. When compared with the discrepancies that are consistent with the model's stochastic elements (e.g. as produced in parametric bootstrap simulations), such graphical assessments can be made more rigorously (Connolly et al. 2009). Graphical assessments are also more likely to bear fruit if a variety of representations of model fit are used, including species abundance distributions and rank abundance distributions, and using a variety of transformations of the relevant axes (Pueyo et al. 2006; Etienne et al. 2007b).

EBSCOhost - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM. All use subject to https://www.ebsco.com/terms-of-use

Species abundance models make predictions not only about the expected number of species with a particular abundance, but also about the site-to-site variability in relative abundance patterns. This information can also be used in goodness-of-fit testing, when multiple species abundance distributions are available. This idea was first used for replicate samples from the same community (Bersier & Sugihara 1997), but can be extended to sites from different locations that have different species abundance model parameters (Connolly et al. 2009). For data that include multiple species abundance samples, predictions about the variability in species abundance statistics among samples can provide additional information that individual fits cannot.

10.3.4 Model selection

Cross-validation is a well-established, widespread model selection technique that has not been used at all in the analysis of species abundances, to our knowledge. Because it is such an intuitive approach to model selection, it probably deserves some attention. However, the best way to approach cross-validation in the context of species abundance analysis is not immediately obvious. For instance, if one simply eliminates a species, chosen at random, from a dataset and re-fits a species abundance model, then one makes the assumption that the species is the unit that is sampled. This assumption has substantial effects on the results of goodness-of-fit testing (Connolly et al. 2009), so we are wary of giving it critical importance in model selection. Alternatively, however, it may be possible to subsample individuals from within species abundance distributions, or to subsample sites from a collection of replicate sites. For instance, one could randomly sample half of the individuals from a data set, fit a species abundance distribution to those individuals and then test the fitted model against the abundance distribution for the other half. Similarly, where multiple sites are distributed randomly within a larger area, and the species abundance distribution is expected to have the same parameters in each site, then a species abundance model fitted to some of the sites could be tested against the abundance distributions from the other sites.

In general, there is considerable controversy about what approach to model selection is the best one, in ecological or other applications. Therefore, rather than recommending the use of a particular statistic or method, we encourage analysts to understand and take appropriate account of the assumptions that underlie the model selection statistics that they do use. Many applications of AIC, for instance, are undertaken without any attempt to evaluate goodness of fit, despite the fact that one must assume, in order to derive AIC, that the model in question is such a good approximation that it can be substituted for the truth. Similarly, when Bayesian model selection statistics are used, and the prior distribution is chosen for mathematical convenience or on the basis of subjective belief, an assessment of the robustness of a study's conclusions to those priors is important. Finally, graphical analysis of model fit can be an informative complement to formal model selection, just as it is to goodness-of-fit testing.

10.3.5 Conclusions

Parameter estimation, goodness-of-fit testing and model selection are tools for obtaining as good a fit of a model to data as possible—either the data in hand, or a new sample of the same kind of data from the same sampling universe. For most ecologists, however, the goal of species abundance analysis is not to describe variation in relative abundances, but rather to use information about model fit to draw inferences about the processes that drive community structure or dynamics. Rigorous statistical techniques help ecologists quantify how well ecological models characterize patterns in nature, but they are an aide to, and not a substitute for, thinking about the processes that give rise to those natural patterns. Philosophers, historians and sociologists of science have increasingly recognized that the evaluation of models in science often depends on much more than goodness of fit or predictive accuracy. In practice, scientists often evaluate models based in part on how consistent they are with existing, well-supported theory, or on their explanatory power—their ability to provide unified explanations for a large number of patterns that previously required separate explanations (Kosso 1992).

EBSCOhost - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM. All use subject to https://www.ebsco.com/terms-of-use

Consider the evaluation of neutral models as an example. Neutral theory offers considerable explanatory power, purporting to explain not only patterns of species abundance, but species–area relationships and even patterns of speciation and extinction in the fossil record (Hubbell 2001). On the other hand, there is considerable empirical evidence for niche differences and life history tradeoffs among organisms. There is also an extensive body of ecological theory, supported by experiments, that links such patterns to species coexistence. Consequently, neutral models, by explaining species coexistence as transient, driven by demographic stochasticity and with species differences playing a negligible role, appears to contradict a large body of ecological theory and data about species coexistence (Abrams 2001; Mazancourt 2001). This may explain why some of the more influential tests of neutral theory have examined how well-known ecological realities omitted from neutral theory drive the very patterns that neutral models seek to explain (Fargione et al. 2003) or assessed the extent to which neutral models really can explain multiple patterns simultaneously (Adler 2004). Those two kinds of tests address these important (albeit subjective) criteria for evaluating models: potential explanatory power and consistency with established and well-supported theory.

Bayesian approaches offer a way to combine these different kinds of considerations by means of prior probability distributions on parameters or prior probabilities assigned to models based (for instance) on their mechanistic plausibility or their explanatory power. However, there will probably never be ways to objectively determine how much to weight to give these different considerations, alongside model fit, when evaluating models. Thus, model evaluation in the broadest sense—the scientific community's judgment about a model's potential to be a productive framework for future research—will probably always involve subjective elements. Indeed, for precisely this reason, we believe that the use of rigorous and powerful statistical tools for assessing model fit is crucial: by providing objective information important to model evaluation, it can help to make the dialogue between theory and data as fruitful as possible.

10.4 Key points

1. Fitting models to species abundance distributions entails considerations about four main issues: obtaining model predictions, estimating parameters, testing the goodness of fit and selecting among competing models.
2. Because sampling affects the shape of species abundance distributions, often substantially, species abundance models should incorporate a sampling theory.
3. We favour maximum likelihood methods for parameter estimation because commonly used alternatives are problematic on statistical grounds. Maximum likelihood will probably perform best when applied to abundances that have not been grouped into abundance classes, unless model predictions must be obtained by stochastic simulation. However, a rigorous comparative analysis of different approaches to parameter estimation has not yet been undertaken.
4. Rigorously quantifying goodness of fit is an important part of species abundance analysis, even if model selection is used instead of classical null hypothesis testing. Parametric bootstrapping is a robust and versatile alternative to conventional goodness-of-fit testing, particularly useful where conventional tests have low statistical power. Graphical analyses complement formal goodness-of-fit testing by helping to identify the features of the data most responsible for lack of fit.
5. Model selection statistics compare models according to different criteria, reflecting philosophical differences about the nature and purpose of statistical analysis, and they should be used and interpreted with these differences in mind. However, all involve a trade-off between how well a model fits the data in hand, and how flexible the model is (i.e. how readily it might fit noise in the data as if it were part of the pattern). Consequently, they offer a stronger basis for comparing the extent to which models capture real patterns in particular empirical data sets, relative to traditional approaches like graphical inspection and comparative analysis of goodness-of-fit statistics.

EBSCOhost - printed on 4/8/2024 9:35 AM via UNIVERSITEIT VAN AMSTERDAM. All use subject to https://www.ebsco.com/terms-of-use