REVIEW

Measurement scale in maximum entropy models of species abundance

S. A. FRANK

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA, USA

Abstract

The consistency of the species abundance distribution across diverse communities has attracted widespread attention. In this paper, I argue that the consistency of pattern arises because diverse ecological mechanisms share a common symmetry with regard to measurement scale. By symmetry, I mean that different ecological processes preserve the same measure of information and lose all other information in the aggregation of various perturbations. I frame these explanations of symmetry, measurement, and aggregation in terms of a recently developed extension to the theory of maximum entropy. I show that the natural measurement scale for the species abundance distribution is log-linear: the information in observations at small population sizes scales logarithmically and, as population size increases, the scaling of information grades from logarithmic to linear. Such log-linear scaling leads naturally to a gamma distribution for species abundance, which matches well with the observed patterns. Much of the variation between samples can be explained by the magnitude at which the measurement scale grades from logarithmic to linear. This measurement approach can be applied to the similar problem of allelic diversity in population genetics and to a wide variety of other patterns in biology.

It is better to be vaguely right than exactly wrong (Read, 1909).

Introduction

The species abundance distribution (SAD) describes the number of individuals of each species observed in a sample. The SAD shape is remarkably consistent across communities. Many distinct rare species each have only a single individual in the sample. The number of different species declines as the count of individuals per species rises (Fisher *et al.*, 1943; Preston, 1948; MacArthur, 1957, 1960; Whittaker, 1965; May, 1975; Hubbell, 2001; Magurran, 2004; McGill *et al.*, 2007; Ulrich *et al.*, 2010).

Such a consistent pattern naturally leads to widespread interest. What is the best description of the pattern? What theory best explains the consistency across such diverse habitats? How should we think of the differences

Correspondence: Steven A. Frank, Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697-2525, USA.
Tel.: +1 949 824 2244; fax: +1 949 824 2181; e-mail: safrank@uci.edu

in distribution that do occur between certain types of habitat? A vast literature is devoted to these questions. Here, I focus on connecting two points of view in the recent debate.

The first point of view argues that different processes can lead to the same pattern. Thus, a fit between the observed SAD and a particular mechanistic theory must be treated with caution, because other equally plausible mechanisms lead to the same pattern (May, 1975; Pueyo *et al.*, 2007; McGill, 2010). No one argues directly against this point of view. Nonetheless, the tendency of different processes to lead to the same pattern is sometimes ignored, because we do not have a fully convincing theory for why widely different processes would in fact lead consistently to the narrow range of observed SAD patterns.

The second point of view uses maximum entropy theory to explain the observed SAD. In maximum entropy, the most likely probability distribution is the one that is most random, or has highest entropy, subject to certain minimal constraints. A constraint might, for example, be that the average population size of species is

set by the habitat. By maximum entropy, the abundances of species would be the most random pattern such that the overall average abundance is fixed (Shipley *et al.*, 2006; Pueyo *et al.*, 2007; Harte *et al.*, 2008; Banavar *et al.*, 2010; Haegeman & Etienne, 2010; He, 2010).

Maximum entropy could, in principle, explain why different mechanistic hypotheses lead to the same SAD pattern. Two different mechanisms have the same SAD shape if they both constrain the average abundance and otherwise produce various perturbations that ultimately tend to cancel in the aggregate. However, there is at present no general understanding of the relation between maximum entropy and the tendency for different mechanistic hypotheses to converge to the same SAD pattern. Thus, maximum entropy is often viewed as an alternative theory for SAD patterns (McGill, 2010), rather than a more fundamental principle about probability that necessarily plays a central role in translating process into pattern.

In this paper, I use recent advances in maximum entropy theory to strengthen the argument that many different underlying mechanistic hypotheses lead to the same common SAD pattern. The advances follow from my work with Eric Smith, showing the importance of information invariance and measurement scale in understanding the fundamental ways in which different probability patterns arise (Frank & Smith, 2010, 2011).

The first section describes the common theoretical SADs that have been used to fit observed patterns. In that section, I show that the gamma distribution subsumes the log series, power law, and geometric distributions as special cases. The gamma distribution can also take on shapes very close to the widely used lognormal distribution. The gamma often fits observed SAD patterns better than the lognormal.

The second section argues that the gamma distribution arises as the natural expression of pattern on a log-linear measurement scale. A log-linear scale is logarithmic at small magnitudes and continuously grades into a linear scale at large magnitudes. In terms of SADs, the match to the gamma pattern means that the information one obtains from an observed species abundance in a sample scales logarithmically at low abundance and linearly at high abundance. I use recent advances in maximum entropy theory to derive this relation between log-linear measurement scale and observed SAD patterns (Frank & Smith, 2010, 2011).

From the first two sections, I conclude that SADs often follow a gamma distribution and the gamma distribution arises naturally as the expression of pattern on a log-linear measurement scale. Those conclusions leave us with the question: Why do ecological mechanisms often lead to log-linear scaling? My main goal is to establish that question, which the first two sections accomplish. In the third section, I explore possible answers by examining the way in which specific ecological mechanisms associate with log-linear scaling.

The discussion analyzes the position of maximum entropy among the various approaches to understanding biological pattern. McGill (2010) recently classed maximum entropy as an approach that makes particular hidden assumptions about mechanism. By this view, maximum entropy is a testable hypothesis that can be evaluated by observation. By contrast, I argue that maximum entropy is like the calculus. One does not test the calculus by comparing predictions with data. Rather, both the calculus and maximum entropy provide analytical tools that help in understanding the logical relations between assumptions and observations.

In the appendix, I note that Pueyo *et al.* (2007) originally established the basic approach of maximum entropy and invariance for species abundance problems. I then describe specific limitations in the way that Pueyo *et al.* (2007) framed the maximum entropy problem and how my measurement theory approach resolves those problems. My resolution connects the maximum entropy method to a broader framework of measurement and information, providing a deeper understanding that is essential for interpreting ecological pattern.

With regard to ecological pattern, maximum entropy can be used in two ways. First, from a consistently observed pattern, such as the SAD, one can induce the necessary and sufficient attributes that various ecological mechanisms must have to match observed pattern. Second, one can deduce the predicted pattern generated by a wide class of ecological mechanisms that share common attributes. Those shared attributes determine the measurement scale and thus define pattern. I will emphasize that the way in which maximum entropy has been applied to ecology needs to be revised to relate the method to the structure of the biological problem.

Common measurements of genetic diversity are often analogous to the species abundance problem. In genetics, one may classify genetic variants into distinct alleles and then measure the abundance of each allele. The distribution of allelic abundances in a sample has the same structure as the distribution of species in a sample. The equivalence of species and allelic sampling problems has been discussed often (Watterson, 1974; Hubbell, 2001; Leigh, 2007; Johnson *et al.*, 1997, chapter 41). In this paper, I focus on the ecological problem of species, because that subject has developed more fully the particular issues that I will analyze.

Maximum entropy methods in relation to ecological and genetic patterns illustrate a deeper problem in biology (Frank, 2009). How do we separate commonly observed patterns generated by typical processes of measurement and aggregation from those special patterns that provide information about the underlying biological mechanisms? Current biological analysis has largely ignored this central problem. Consistent progress depends on clearer understanding. In particular, we must have some sense of what is surprising and informative vs. what is unsurprising and uninformative. Otherwise,

much analysis devotes attention to what is in fact expected based on the simplest notions of aggregation and measurement.

SADs follow the gamma distribution

Among S species, the probability p_y is the fraction of species each with y individuals and Sp_y is the number of species each with y individuals. The distribution of p_y defines the SAD. I focus on the underlying distribution of species abundances, ignoring the inevitable fluctuations caused by sampling. Sampling fluctuations are important, but in this paper, I wish to isolate the forces that shape the underlying distribution from sampling fluctuations.

In this section, I argue that a gamma probability distribution is a simple and general description of observed SAD patterns. Ulrich *et al.* (2010) recently conducted a meta-analysis of 558 SADs derived from 306 publications. Their conclusions match the broad consensus in the literature that SADs typically fit best to either a log series distribution or a lognormal distribution. They also found that several observed SADs fit best to a power law distribution.

One can always quibble about the methods of fitting and the choice of alternative distributions to compare. My point of view in this paper does not depend on the fine points of fitting alternative distributions. Rather, we can simply take the qualitative conclusion that observed distributions usually fit reasonably well to either the log series, lognormal, or power law pattern. Different observed SADs vary in which of these three distributions fits best.

Log series and power law distributions

The log series distribution has the form

$$p_y = k\frac{\theta^y}{y}, \tag{1}$$

where $k = -1/\log(1 - \theta)$, $0 < \theta < 1$, and $y = 1, 2, \ldots$ is the number of individuals in the sample for each species of class y. I use 'log' for the natural logarithm. This distribution has its highest value (mode) at $y = 1$, so that the rarest species, each represented by a single individual, occur most frequently in the sample. Species with increasingly large populations in the sample occur at a decreasing frequency.

The power law distribution has the form

$$p_y = ky^{-\gamma}, \tag{2}$$

where $\gamma > 1$ and k is chosen so that the total probability is one over all values of y. The power law also has a mode at one, and frequency declines steadily for species with increasing population sizes. The rate of decline in frequency with increasing population size is slower for the power law than for the log series.

For discrete distributions, such as the forms of the log series and power law given here, we may truncate the distribution so that the greatest value of y is not higher than some upper bound, B, resetting k so that the total probability is one.

Normal and lognormal distributions

To describe SADs by the lognormal, one must first transform the underlying measurement scale. The traditional approach follows the Preston plot method, in which one forms bins on a $\log_2$ scale for abundance values y, such that $y = 1$ maps to the 2^0 or $\log_2 = 0$ bin, $y = 2$ maps to the $\log_2 = 1$ bin, the combination of $y = 3, 4$ maps to the $\log_2 = 2$ bin, the combination of $y = 5, 6, 7, 8$ maps to the $\log_2 = 3$ bin, and so on. This binning creates a discrete distribution on the logarithmic scale.

On the logarithmic scale, the lognormal has the symmetric shape of a normal distribution. For fits to observed SADs, one matches a continuous normal distribution to the discrete binned logarithmic distribution.

In the next section, I will describe the gamma distribution on both the linear and logarithmic scales. To compare those forms of the gamma with the lognormal, it is useful to describe the transformations between the log and linear scales for the lognormal.

Here, I use y for the linear scale and $x = \log y$ for the log scale. I use the standard notation for continuous probability, in which $p_x dx$ is the probability that x falls in the interval between x and $x + dx$ for a small increment dx. The magnitude of the increment dx may change in ways that define the measurement scale, as shown in the following.

On the log scale, the normal distribution is

$$p_x dx = ke^{-\frac{(x-\mu)^2}{2\sigma^2}} dx, \tag{3}$$

where the measure dx is for the log scale, x. To get the linear scale measure, we make the substitution $x = \log y$ to obtain

$$p_{\log y} d\log y = ke^{-\frac{(\log y - \mu)^2}{2\sigma^2}} d\log y,$$

where k is always taken to adjust so that the total probability is one. We get the form of p_y on the linear scale by changing the measure $d\log y = dy/y$ and noting that $p_y = p_{\log y}/y$, yielding

$$p_y dy = ky^{-1}e^{-\frac{(\log y - \mu)^2}{2\sigma^2}} dy.$$

Thus, the distribution here is normal on the log scale and lognormal on the linear scale.

To compare the lognormal and gamma distributions with log-binned SAD data on the logarithmic scale, we need to express the gamma on the logarithmic scale.

Gamma and exponential-gamma distributions

In this section, I first describe the gamma distribution on the linear scale and then present what I call the exponential-gamma on the log scale.

The discrete gamma distribution has the form

$$p_y = ky^{\alpha-1}e^{-\lambda y} \qquad (4)$$

for $y = 1, 2, \ldots$. Define $\theta = e^{-\lambda}$, which allows us to write the discrete gamma as

$$p_y = ky^{\alpha-1}\theta^y. \qquad (5)$$

The discrete gamma contains the discrete log series, power law, and geometric distributions as special cases. For $\alpha = 0$, we obtain the log series distribution; for $\theta = 1$, we obtain the power law distribution; for $\alpha = 1$, we obtain the geometric distribution. The gamma distribution subsumes the other classic distributions because it derives from a generalization of the measurement scales of the other distributions, as explained later.

Because the two-parameter discrete gamma distribution contains the one-parameter log series, power law, and geometric distributions as special cases, the gamma must always fit any observed SAD at least as well as the special cases.

In fitting observed distributions, it is widely known that the gamma and lognormal distributions can take on similar shapes and are often hard to distinguish in practice (Cho *et al.*, 2004; Silva & Lisboa, 2007). For SADs, the main weakness of the lognormal is that its symmetry on the log scale usually underestimates the frequency of rare species in the lower tail of the distribution (Diserud & Engen, 2000; Hubbell, 2001; Wilson & Lundberg, 2004). A greater weight in the lower tail of the gamma is exactly the main difference between the gamma and lognormal distributions. I show an example below.

To examine the gamma distribution on the log scale, we first need the continuous form on the linear scale. The continuous linear form has the same expression as the discrete form but with a continuous interpretation

$$p_y\mathrm{d}y = ky^{\alpha-1}e^{-\lambda y}\mathrm{d}y,$$

where $y > 0$ and $k = \lambda^{\alpha}/\Gamma(\alpha)$, and I write $\mathrm{d}y$ to emphasize the continuous measure of probability on the linear scale. To transform this linear scale version to the log scale, we will once again use the substitution $x = \log y$. However, in this case, we are going in the reverse direction, so we need the inverse substitution $y = e^x$ yielding

$$p_{e^x}\mathrm{d}e^x = k(e^x)^{\alpha-1}e^{-\lambda e^x}\mathrm{d}e^x$$
$$= ke^{(\alpha-1)x-\lambda e^x}\mathrm{d}e^x.$$

Now change the measure to the log scale in terms of x by the substitution $\mathrm{d}e^x = e^x\mathrm{d}x$ and note that $p_x = p_{e^x}e^x$, yielding what I call the exponential-gamma distribution

$$p_x\mathrm{d}x = ke^{\alpha x-\lambda e^x}\mathrm{d}x. \qquad (6)$$

Because $x = \log y$ and $y > 0$, the domain is $-\infty < x < \infty$. However, it is often useful to study the truncated form with $x \geq 0$, corresponding to $y \geq 1$ for species abundance counts that have a lower bound of one. In the truncated form, $k = \lambda^{\alpha}/\Gamma(\alpha,\lambda)$, where the denominator is the incomplete gamma function evaluated from a lower bound of λ.

In the case of the lognormal, that distribution is normal on the logarithmic scale and lognormal on the linear scale. In the case here, the distribution is gamma on the linear scale and exponential-gamma on the logarithmic scale.

Example comparison of lognormal and gamma distributions

In this paper, rather than fitting data to distributions, I emphasize the gamma distribution as a natural expression for the diversity of observed SAD forms. However, it is useful to look at a plot of some data to get a feel for the shapes. Figure 1 shows data for a typical SAD. I fit by eye the matching normal and exponential-gamma distributions on the log scale, corresponding to the lognormal and gamma distributions on the linear scale.

This figure illustrates the commonly observed excess of rare species compared with the lognormal pattern. The gamma pattern differs most strongly from the lognormal by allowing a higher probability weighting of small values; otherwise, the lognormal and gamma distributions are similar.

Plotkin & Muller-Landau (2002) commented on the good fit to SADs provided by the gamma distribution.
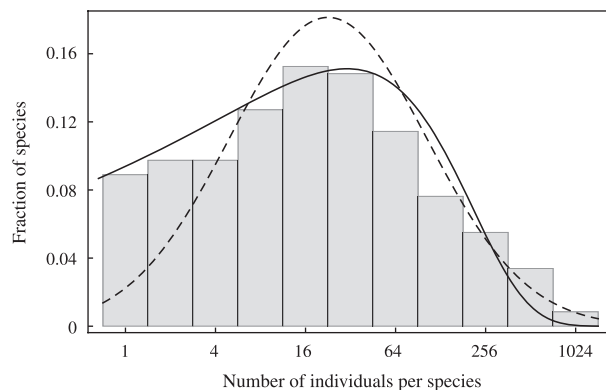
Fig. 1 Preston plot of species abundance distribution for tree species with diameter at breast height (dbh) greater than 10 cm from a 50 ha plot on Barro Colorado Island, Panama. Data extracted from fig. 5.7 of Hubbell (2001). The numbers on the abcissa show the linear counts, y, scaled logarithmically so that $x = \log y$. The dashed line shows a matching normal distribution on the log scale (lognormal on the linear scale) from eqn (3). The solid line shows a matching exponential-gamma distribution on the log scale (gamma on the linear scale) from eqn (6).

They also noted that the gamma distribution is not commonly used to fit SADs, in spite of the generally good match and some clear precedents (Fisher *et al.*, 1943; Dennis & Patil, 1984; Engen & Lande, 1996). The gamma may lack popularity because explicit models of ecological process and sampling often lead to other distributions.

In my view, those explicit models make the mistake of setting too many exact assumptions about the generation of pattern. Those exact assumptions can never be matched by the heterogeneous reality of nature. Instead, the dominant aspects of pattern may arise from very general aspects of the way in which heterogeneous perturbations combine in the aggregate. The central limit theorem is the most obvious example, in which linearly scaled perturbations lead to a normal distribution. When perburbations follow other measurement scales, different distributions may arise.

Other distributions may sometimes provide a better fit than the gamma, for example, Hubbell's (2001) zero-sum multinomial distribution of the neutral theory. However, my point goes beyond the particular fit by various distributions with different numbers of parameters. Rather, I emphasize that the gamma distribution subsumes as special cases several classic distributions commonly used for SADs, the gamma typically outperforms the lognormal in fitting, and the gamma arises naturally from very general aspects of measurement and information. I am particularly interested in this last aspect of the gamma distribution as a simple expression of likely probability patterns in relation to natural changes in measurement scale with magnitude.

Gamma from log-linear measurement and maximum entropy

In this section, I show that the gamma distribution arises as the natural expression of pattern on a log-linear measurement scale. A log-linear scale is logarithmic at small magnitudes and continuously grades into linear at larger magnitudes. The following section illustrates why SADs may tend to associate with log-linear scaling, in which low population abundances carry information in relation to a log scale, and large population abundances carry information in relation to a linear scale.

To explain the claim that the gamma distribution expresses log-linear scaling, I first review the standard method of maximum entropy to derive probability distributions. I then describe recent extensions to maximum entropy to incorporate the role of measurement scale. My review of maximum entropy and description of the extensions for measurement scale are condensed summaries of Frank & Smith (2010, 2011).

Maximum entropy

The method of maximum entropy defines the most likely probability distribution as the distribution that maximizes a measure of entropy (randomness) subject to various information constraints (Jaynes, 2003). The idea is that the many random perturbations that affect pattern mostly tend to cancel each other in the aggregate, leaving the aggregate completely random except for any constraints that restrict the pattern.

For example, the average number of individuals per species may be constrained by the productivity of the habitat, and that average will be maintained in spite of the wide variety of other processes that perturb species distributions. So the final pattern must reflect that constraint. Particular processes may tend to push species abundances in one direction, but other processes will push in the other direction. Only the constraints remain in the aggregate, all else tends to maximum randomness or entropy. Maximum randomness is equivalent to minimum information. Thus, maximizing entropy is equivalent to minimizing the information expressed in the final pattern subject to any constraints that cause information to be retained.

The power of maximum entropy is that aggregate patterns almost always seem to converge to a few simple patterns that express maximum randomness subject to just a few informational constraints. To analyze a problem by maximum entropy, one first identifies the informational constraints that define a particular problem. Then, by maximizing randomness subject to those constraints, one obtains the predicted form of the probability distribution that describes the pattern. Going the other way, each common probability distribution is an exact expression of a few informational constraints with all else maximally random.

To derive a probability distribution by maximum entropy, we write the quantity to be maximized as

$$\Lambda = \mathscr{E} - \kappa C_0 - \sum_{i=1}^{n} \lambda_i C_i, \qquad (7)$$

where $\mathscr{E}$ measures entropy, the C_i are the constraints to be satisfied, and κ and the λ_i are the Lagrange multipliers to be found by satisfying the constraints. Let $C_0 = \sum p_y - 1$ be the constraint that the probabilities must total one, where p_y is the probability distribution function of y. The other constraints are usually written as $C_i = \sum p_y f_i(y) - \bar{f}_i$, where the $f_i(y)$ are various transformed measurements of y and the overbar denotes mean value. A mean value is either the average of some function applied to each of a sample of observed values or an *a priori* assumption about the average value of some function with respect to a candidate set of probability laws. If $f_i(y) = y^i$, then $\bar{f}_i$ are the moments of the distribution—either the moments estimated from observations or *a priori* values of the moments set by assumption. The moments are often regarded as 'standard' constraints, although from a mathematical point of view, any properly formed constraint can be used.

Here, I confine the analysis to a single constraint of measurement. I express that constraint with a more general notation, $C_1 = \sum p_y T(f_y) - \bar{T}_f$, where $f_y \equiv f(y)$, and $T(f_y) \equiv T_f$ is a transformation of f_y. I could, of course, express the constraining function for y directly through f_y. However, I wish to distinguish between an initial function f_y that can be regarded as a standard measurement, in any sense in which one chooses to interpret the meaning of standard, and a transformation of standard measurements denoted by T_f that arises from information about the measurement scale.

The maximum entropy distribution is obtained by solving the set of equations

$$\frac{\partial \Lambda}{\partial p_y} = \frac{\partial \mathscr{E}}{\partial p_y} - \kappa - \lambda T_f = 0, \qquad (8)$$

where one checks the candidate solution for a maximum and obtains κ and λ by satisfying the constraint on total probability and the constraint on $\bar{T}_f$. For continuous probability distributions, I assume that I can treat the entropy measures, the contraints, and the maximization procedure by the continuous limit of the discrete case.

In the standard approach, one defines entropy by extension of Shannon information

$$\mathscr{E} = -\int p_y \log\left(\frac{p_y}{m_y}\right) dy. \qquad (9)$$

For discrete distributions, m_y is a prior probability distribution that sets the default pattern of randomness in the absence of any additional informational constraints, yielding an expression that can be interpreted as relative entropy, sometimes called the Kullback-Leibler divergence (Cover & Thomas, 2006). Alternatively, in the case of continuous distributions, m_y may be an adjustment to maintain an invariant measure of information under changes in scale (Jaynes, 2003; Frank & Smith, 2011).

With these definitions, the solution of eqn (8) is

$$p_y \propto m_y e^{-\lambda T_f}, \qquad (10)$$

where λ satisfies the constraint C_1, and the proportionality is adjusted so that the total probability is one by choosing the parameter κ to satisfy the constraint C_0. In the applications in this paper, I will use $m_y \propto 1$ so that the default distribution is uniform, the most random pattern with the highest entropy, and the pattern that lacks any information. I include in $T(f_y)$ any attributes of measurement that may deform the default distribution (Frank & Smith, 2010, 2011). The general maximum entropy solution with $m_y \propto 1$ is

$$p_y \propto e^{-\lambda T_f}. \qquad (11)$$

Information invariance and measurement scale

Maximum entropy must capture all of the available information about a particular problem. One form of information concerns transformations to the measurement scale that leave the most likely probability distribution unchanged (Jaynes, 2003; Frank, 2009; Frank & Smith, 2010). Here, it is important to distinguish between measurements and measurement scale. In my notation, I start with measurements, f_y, made on the measurement scale y. For example, one may have measures of squared deviations about zero, $f_y = y^2$, with respect to the measurement scale y, such that $\bar{f}_y$ is the second moment of the measurements with respect to the underlying measurement scale.

Suppose that one obtains the same information about the underlying probability distribution from measurements of f_y or transformed measurements, $G(f_y)$. Put another way, if one has access only to measurements $G(f_y)$, one has the same information that would be obtained if the measurements were reported as f_y. One may say that the measurements f_y and $G(f_y)$ are equivalent with respect to information, or that the transformation $f_y \rightarrow G(f_y)$ is an information invariance that describes a symmetry of the measurement scale.

To capture this information invariance in maximum entropy, we must express measurements so that

$$T(f_y) = \delta + \phi T[G(f_y)] \qquad (12)$$

for some arbitrary constants δ and ϕ. Putting this definition of $T(f_y) \equiv T_f$ into eqn (11) shows that the same maximum entropy solution arises from the observations f_y or the transformed observations, $G(f_y)$, because the κ and λ parameters of eqn (8) will adjust to the constants δ and ϕ so that the distribution remains unchanged.

Intuitive aspects of invariance and measurement

Intuitively, one can think of information invariance and measurement scale in the following way. On a linear scale, each incremental change of fixed length yields the same amount of information or surprise independently of magnitude. Thus, if we change the scale by multiplying all magnitudes by a constant, we obtain the same pattern of information relative to magnitude. In other words, the linear scale is invariant to multiplication by a constant factor so that, within the framework of maximum entropy subject to constraint, we get the same information about probability distributions from an observation y or $G(y) = cy$.

On a logarithmic scale, each incremental change in proportion to the current magnitude yields the same amount of information or surprise. Information is scale dependent. We obtain the same information at any point on the scale by comparing ratios. For example, we gain the same information from the increment $dy/y = d \log(y)$ independently of the magnitude of y. Thus, we achieve information invariance with respect to ratios by measuring increments on a logarithmic scale. Within the framework of maximum entropy subject to constraint, we get the same information about probability distributions from an observation y or $G(y) = y^c$, corresponding

to informationally equivalent measurements $T(y) = \log(y)$ and $T(y^c) = c \log(y)$ (see Frank & Smith, 2010).

Nearly all of the common probability distributions arise from a simple form of information invariance, measurement scale, and a constraint on measured values of y for mean values or measured values of y^2 for variances (Frank & Smith, 2010, 2011). The main measurement scales in the common distributions express a change of information with magnitude. For example, the linear-log scale grades from linear at small magnitudes to logarithmic at large magnitudes, expressed as $T(y) = \log(1 + by)$, where b is a parameter that determines the magnitude at which the scale changes from linear to logarithmic. When by is small, the scaling is linear, and when by is large, the scaling is logarithmic. By contrast, the log-linear scale grades from logarithmic at small magnitudes to linear at large magnitudes, expressed as $T(y) = y + b \log(y)$.

Gamma distribution from log-linear scaling

When we use the log-linear scale, $T(y) = y + b \log(y)$, in the general maximum entropy solution of eqn (11), we obtain the gamma distribution

$$
\begin{aligned}
p_y &= ke^{-\lambda T_f} \\
&= ke^{-\lambda(y+b\log(y))} \\
&= ky^{\alpha-1}e^{-\lambda y},
\end{aligned}
$$

where $\alpha - 1 = -\lambda b$ and k is the proportionality constant needed for the total probability to be one. This distribution matches the discrete gamma first presented in eqn (4). We obtain the same expression of the distribution from maximum entropy for discrete and continuous cases, with proper definition of the domain as $y = 1, 2, \ldots$ for the discrete case and $y > 0$ for the continuous case, and using the proper normalization for k to guarantee that the total probability is one in each case.

The constraint associated with log-linear scaling is $\bar{T} = \hat{\mu} + b \log(\hat{y})$, where $\hat{\mu}$ is the mean of y and $\hat{y}$ is the geometric mean for which $\log(\hat{y})$ is the mean of $\log(y)$. The means may either be estimated from a sample or assumed *a priori* to take on particular values. It is well known in the maximum entropy literature that one can derive a gamma distribution by constraining the mean and geometric mean (Kapur, 1989; Frank, 2009). However, the measurement theory approach to maximum entropy derives the joint constraint on the mean and geometric mean as an outcome of a general method to analyze the relation between information and magnitude (Frank & Smith, 2010, 2011). The earlier studies simply invoked those constraints as a sufficient description of the gamma distribution (Kapur, 1989; Frank, 2009).

The form of a probability distribution under maximum entropy can be read directly as an expression of how the measurement scale changes with magnitude (Frank & Smith, 2011). From the general solution in eqn (11),

linear scales $T(y) \propto y$ yield distributions that are exponential in y, whereas logarithmic scales $T(y) \propto c \log(y)$ yield distributions that are linear in y^c. Exponential distributions of the form $e^{-\lambda y}$ arise from underlying linear scales, whereas power law distributions of the form y^{-c} arise from underlying logarithmic scales.

The gamma distribution has form $y^{-c}e^{-\lambda y}$. When the magnitude of y is small, the shape of the distribution is dominated by the power law component, y^{-c}. As the magnitude of y increases, the shape of the distribution is dominated by the exponential component, $e^{-\lambda y}$. Thus, the underlying measurement scale grades from logarithmic at small magnitudes to linear at large magnitudes. Indeed, the gamma distribution is exactly the expression of an underlying measurement scale that grades from logarithmic to linear as magnitude increases.

We can now state the key observation about measurement and ecological pattern. Empirically, the evidence strongly demonstrates that SADs almost always follow an approximately log-linear scaling. Variation in the transition between the logarithmic and linear regime describes nearly all of the variation in observed pattern.

Different ecological mechanisms lead to log-linear scaling

What are the set of underlying ecological mechanisms and aspects of measurement that lead in the aggregate to log-linear scaling of SADs?

My main goal for this paper is to reformulate the problem of SADs in terms of this question about log-linear scaling. I cannot answer this question at present, because there is no general understanding of the different kinds of processes that, in the aggregate, lead to particular information invariances and measurement scales. Future progress in understanding biological pattern depends strongly on progress in understanding aggregation, invariance, and scale.

The following section presents a preliminary example. That example hints at the range of processes leading to log-linear scaling.

Stochastic models of population growth

Dennis & Patil (1984) showed that stochastic fluctuations of population growth often lead to a gamma distribution of population abundance. Their work generalized mathematical results from previous studies (May, 1974; May *et al.*, 1978). Costantino & Desharnais (1981) supplemented their mathematical derivation of gamma population abundance with supporting data from several laboratory studies of the flour beetle.

In this section, I highlight the essential aspect of population dynamics that leads to the gamma distribution. The essence reduces to log-linear scaling of perturbations, supporting my claim that a clear understanding of the proper scale of measurement leads to a clear

understanding of biological pattern. Measurement also helps to explain why some patterns are so common, because the most common patterns associate with the most common measurement scales.

By studying the example of population growth, we can learn how aspects of dynamics associate with scale. However, one must remember that this particular model of population growth is just one example of a process that leads to log-linear scaling and the gamma pattern. I give a simplified version of Dennis & Patil (1984), adding my own interpretation with regard to log-linear scaling.

A deterministic model of population growth can often be expressed as

$$\frac{dy}{dt} = yg(y),$$

where y is population size, and $g(y)$ is the growth rate of the population as a function of its size. An equilibrium occurs when $g(y^*) = 0$, and the equilibrium is stable when $g'(y^*) < 0$. A model with stochastic fluctuations can be written as

$$\frac{dy}{dt} = yg(y) + yz(t),$$

where $z(t)$ is a Gaussian perturbation with mean zero and variance σ^2. The magnitude of the perturbation scales linearly with population size, which is equivalent to a constant magnitude of perturbation per individual independent of population size.

Suppose we approximate the growth rate by a linear expression, $g(y) \approx a - cy$. This linear approximation gives the standard logistic growth equation, usually with notation $a = r$ and $c = r/K$, where $-r/K$ is the slope of the growth rate as the population size increases, and $K = y^*$ is the equilibrium. It is common to call r the intrinsic rate of increase and K the carrying capacity, although here those parameters simply arise from a linearization of the general function for growth rate, $g(y)$. Putting the pieces together, we obtain a model with Gaussian stochastic fluctuations in proportion to population size and a linear approximation of population growth

$$\frac{dy}{dt} = y(a - cy) + yz(t).$$

Turelli (1977) analyzed various interpretations of this stochastic equation. For my purposes, the following simplified presentation captures the essential aspect of implicit log-linear measurement scale that leads to a gamma distribution of population sizes.

We can think of perturbations of population size at each point in time as having two components. First, a direction component arises from the tendency for the population to follow the expected instantaneous growth rate $a - cy$, leading to a mean directional tendency on population size of $m(y) = y(a - cy)$. The stochastic perturbations, $z(t)$, have a mean value of zero and do not contribute to a directional tendency. Second, the term

$z(t)$ contributes a directionally unbiased fluctuation with variance σ^2 and the combination $yz(t)$ contributes variance $v(y) = y^2\sigma^2$. The stochastic fluctuations, σ^2, are of the same order of magnitude or less than the maximal directional tendency, a.

Because the population has a randomly fluctuating component, the population size never settles to a single value. Instead, a steady state may be reached, such that for each population size, y, the tendency of the population to move to another size is balanced by the overall tendency of other population sizes to change to y. This balance allows us to calculate the steady state probability distribution for population sizes, which is

$$p_y = ke^{-\psi(y)}, \qquad (13)$$

where k is chosen so that the total probability is one. From a general solution of stochastic differential equations (May, 1974; May et al., 1978), we have

$$-\psi(y) = -\log v(y) + 2\int^y \frac{m(\tilde{y})}{v(\tilde{y})} d\tilde{y}. \qquad (14)$$

Using the expressions above for the mean, m, and variance, v, of perturbations and carrying out the integration, we obtain

$$-\psi(y) = (\alpha - 1)\log y - \lambda y + C, \qquad (15)$$

where $\lambda = 2c/\sigma^2$ and $\alpha - 1 = 2(a/\sigma^2 - 1)$, and C is a constant that will be absorbed by k in the following step. Using these expressions in eqn (13), we obtain

$$p_y = ke^{(\alpha-1)\log y - \lambda y}$$
$$= ky^{\alpha-1}e^{-\lambda y},$$

which is a gamma distribution.

My main argument is that to interpret a probability distribution with respect to underlying process, we must focus on the measurement scale expressed by $T(y)$. By comparing eqn (13) for the stochastic solution of the probability distribution with eqn (11) for the maximum entropy probability distribution in relation to measurement scale, we have that $\psi(y) = \lambda T(y)$, where in eqn (11) I use $f(y) = y$ and $T_f = T(y)$. From eqn (14), moving $\log v(y)$ inside the integral yields

$$-\lambda T(y) = -\psi(y) = 2\int^y \frac{m(\tilde{y}) - v'(\tilde{y})/2}{v(\tilde{y})} d\tilde{y},$$

where v' is the derivative with respect to y. Differentiating and using the shorthand notation $m \equiv m(y)$ and $v \equiv v(y)$, we obtain

$$dT(y) \propto \frac{m - v'/2}{v} dy.$$

The directional and stochastic perturbations, m and v, change the effective measurement scale in the manner given by this expression. One may think of the scale dT as expressing information in relation to magnitude. Stronger stochastic fluctuations, v, effectively reduce the precision or information with regard to directional

tendency, whereas stronger directional fluctuations, m, in relation to v, effectively provide more information about directional tendency.

Returning to the specific problem of population growth that leads to the gamma distribution, we can translate the stochastic solution for the probability distribution based on eqn (15) into an expression in terms of the measurement function $T(y)$, yielding

$$-\lambda T(y) = -\psi(y) = (\alpha - 1)\log y - \lambda y,$$

allowing us to write the measurement function in the generic form of log-linear scaling,

$$T(y) = y + b \log y, \tag{16}$$

or, equivalently,

$$dT(y) \propto \left(1 + \frac{b}{y}\right) dy.$$

This measure scales logarithmically when abundance, y, is low and linearly when abundance is high.

The parameter $b = -(\alpha-1)/\lambda$ determines the magnitude of abundance at which the scale grades from logarithmic to linear. We can express b in terms of the parameters of the growth equation and the magnitude of stochastic perturbations

$$b = -K\left(1 - \frac{\sigma^2}{r}\right),$$

where I have used the traditional parameters of logistic growth, r for intrinsic rate of increase, and K for carrying capacity.

This expression for b shows that higher carrying capacity, K, increases the domain of the logarithmic regime to higher magnitude, causing pattern to follow a power law over a wider span. This control of the relative logarithmic and linear domains is easy to understand. At low abundance relative to carrying capacity, the deterministic component of population growth perturbs abundance exponentially. As abundance approaches the carrying capacity, deterministic perturbations become linear in abundance.

The term σ^2/r reflects the relative scaling of the linear stochastic perturbations, σ^2, to the logarithmic deterministic perturbations, r. Greater relative stochastic fluctuations reduce b and shift scaling toward the linear domain, because noise in this model is added as a linear perturbation.

Discussion

The gamma distribution expressed in terms of the log-linear measurement scale can be viewed in two alternative ways. First, the derivation of the gamma distribution from a model of stochastic population growth can be thought of as the primary line of reasoning. By that view, the log-linear interpretation follows secondarily as a description of stochastic population growth. Second, one may think of the log-linear measurement scaling as a primary argument for why a gamma distribution is likely to be a common pattern. By that view, the derivation from a particular stochastic model of population growth arises secondarily as a special instance of a much wider class of problems that share the common log-linear scaling.

The current literature promotes the first view: primary, specific derivations from underlying mechanistic models. In my opinion, that approach is certain to be exactly wrong. The exact part arises because the models derive exact expectations from explicit mechanistic assumptions. The wrong part arises because nature will certainly not be the outcome of exactly those specific assumptions. Natural pattern will almost certainly be dominated by the aggregation of various distinct processes.

In thinking about such aggregation, we will never be able to specify exactly the various ecological mechanisms and their relative contributions. We can only vaguely analyze how such aggregation may consistently shape pattern. However, we know that aggregation of heterogeneous processes can sometimes attract strongly to particular outcomes, as in the central limit theorem. My conjecture is that species abundance is dominated in the aggregate by a log-linear measurement scale, reflecting relatively consistent informational invariances in relation to magnitude. Such invariance leads to the gamma pattern. By this route of analysis, one can be vaguely but consistently right about pattern and its causes.

The ecological and genetical literature has devoted little effort to search for general principles such as measurement invariance. Instead, we have a vast catalog of specific assumptions leading to exactly specified outcomes. The neutral theories are typically posed as outcomes that follow exactly from precise assumptions about process, rather than expectations that follow vaguely from general principles of aggregation and measurement. These fields will never mature fully until they develop a clearer sense of the relations between theory and pattern.

My emphasis on aggregation's tendency to attract to particular patterns is of course not new. Many authors, such as May (1975) and Levin & Pacala (1997), have advocated the perspective of the central limit theorem or mean field approximations to explain consistent pattern. More recently, several papers have discussed how different ecological mechanisms attract to the same SADs (summarized by Alonso *et al.*, 2008).

However, these studies, on how different mechanisms attract to the same outcome, are limited in scope. Such studies rarely develop a clear sense of what kinds of processes do attract to a particular outcome, such as a gamma distribution, and what kinds of processes do not attract to that particular pattern. In this regard, measurement theory sets the important questions. What

essential aspects of ecological mechanisms and aggregation preserve key information invariances? What aspects of process do not matter, because the details of those processes do not alter the fundamental information invariances?

The words *invariance* and *symmetry* can be used interchangeably, in the sense stated by Cantwell (2002, p. 4, attributed to Hermann Weyl): *An object is symmetrical if one can subject it to a certain operation and it appears exactly the same after the operation. The object is then said to be invariant with respect to the given operation.*

Thus, the causes of pattern reduce to the fundamental symmetries of ecological process. Here, the symmetries may often have to do with scale in relation to dynamics. For example, in the log-linear scaling that arises in the particular model of stochastic population growth, the invariant transformations that define the symmetries of the measurement scale change with magnitude, because relative scaling of the deterministic and stochastic perturbations change with magnitude. Many problems must reduce to such descriptions of symmetries, capturing the essence of process in relation to pattern. What we need is more work devoted to the special aspects of ecological and genetic processes, so that we can readily reduce particular, complex situations to their essential symmetries. That view, although relatively unused in biology, is not so far-fetched. As Anderson (1972) noted: *It is only slightly overstating the case to say that physics is the study of symmetry.*

Acknowledgments

References

Alonso, D., Ostling, A. & Etienne, R.S. 2008. The implicit assumption of symmetry and the species abundance distribution. *Ecol. Lett.* **11**: 93–105.

Anderson, P. 1972. More is different. *Science* **177**: 393–396.

Banavar, J.R., Maritan, A. & Volkov, I. 2010. Applications of the principle of maximum entropy: from physics to ecology. *J. Phys. Condens. Matter* **22**, 063101.

Cantwell, B. 2002. *Introduction to Symmetry Analysis*. Cambridge University Press, Cambridge, UK.

Cho, H.K., Bowman, K.P. & North, G.R. 2004. A comparison of gamma and lognormal distributions for characterizing satellite rain rates from the tropical rainfall measuring mission. *J. Appl. Meteorol.* **43**: 1586–1597.

Costantino, R.F. & Desharnais, R.A. 1981. Gamma distributions of adult numbers for *Tribolium* populations in the region of their steady states. *J. Anim. Ecol.* **50**: 667–681.

Cover, T.M. & Thomas, J.A. 2006. *Elements of Information Theory*, 2nd edn. Wiley, Hoboken, NJ.

Dennis, B. & Patil, G.P. 1984. The gamma distribution and weighted multimodal gamma distributions as models of population abundance. *Math. Biosci.* **68**: 187–212.

Diserud, O.H. & Engen, S. 2000. A general and dynamic species abundance model, embracing the lognormal and the gamma models. *Am. Nat.* **155**: 497–511.

Engen, S. & Lande, R. 1996. Population dynamic models generating species abundance distributions of the gamma type. *J. Theor. Biol.* **178**: 325–331.

Fisher, R.A., Corbet, A.S. & Williams, C.B. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**: 42–58.

Frank, S.A. 2009. The common patterns of nature. *J. Evol. Biol.* **22**: 1563–1585.

Frank, S.A. & Smith, D.E. 2010. Measurement invariance, entropy, and probability. *Entropy* **12**: 289–303.

Frank, S.A. & Smith, E. 2011. A simple derivation and classification of common probability distributions based on information symmetry and measurement scale. *J. Evol. Biol.*, doi: 10.1111/j.1420-9101.2010.02204.x.

Haegeman, B. & Etienne, R.S. 2010. Entropy maximization and the spatial distribution of species. *Am. Nat.* **175**: E74–E90.

Hand, D. 2004. *Measurement Theory and Practice*. Arnold, London.

Harte, J., Zillio, T., Conlisk, E. & Smith, A.B. 2008. Maximum entropy and the state-variable approach to macroecology. *Ecology* **89**: 2700–2711.

He, F. 2010. Maximum entropy, logistic regression, and species abundance. *Oikos* **119**: 578–582.

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

Jaynes, E.T. 2003. *Probability Theory: The Logic of Science*. Cambridge University Press, New York.

Jeffreys, H. 1957. *Scientific Inference*, 2nd edn. Cambridge University Press, London.

Johnson, N.L., Kotz, S. & Balakrishnan, N. 1997. *Discrete Multivariate Distributions*. Wiley, New York.

Kapur, J.N. 1989. *Maximum-Entropy Models in Science and Engineering*. Wiley, New York.

Leigh, Jr., E.G. 2007. Neutral theory: a historical perspective. *J. Evol. Biol.* **20**, 2075–2091.

Levin, S.A. & Pacala, S.W. 1997. Theories of simplification and scaling of spatially distributed processes. In: *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions* (D. Tilman & P.M. Kareiva, eds), pp. 271–295. Princeton University Press, Princeton, NJ.

MacArthur, R.H. 1957. On the relative abundance of bird species. *Proc. Natl Acad. Sci. USA* **43**: 293–295.

MacArthur, R.H. 1960. On the relative abundance of species. *Am. Nat.* **94**: 25–36.

Magurran, A.E. 2004. *Measuring Biological Diversity*. Wiley-Blackwell, New York.

May, R.M. 1974. *Stability and Complexity in Model Ecosystems*, 2nd edn. Princeton Univ Press, Princeton, NJ.

May, R.M. 1975. Patterns of species abundance and diversity. In: *Ecology and Evolution of Communities* (M.L. Cody & J.M. Diamond eds), pp. 81–120. Harvard University Press, Cambridge, MA.

May, R.M., Beddington, J.R., Horwood, J.W. & Shepherd, J.G. 1978. Exploiting natural populations in an uncertain world. *Math. Biosci.* **42**: 219–252.

McGill, B.J. 2010. Towards a unification of unified theories of biodiversity. *Ecol. Lett.* **13**: 627–642.

McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K. *et al.* 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.* **10**: 995–1015.

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

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

Pueyo, S. 2006. Self-similarity in species–area relationship and in species abundance distribution. *Oikos* **112**: 156–162.

Pueyo, S., He, F. & Zillio, T. 2007. The maximum entropy formalism and the idiosyncratic theory of biodiversity. *Ecol. Lett.* **10**: 1017–1028.

Read, C. 1909. *Logic, Deductive and Inductive*. A. Moring, London.

Shipley, B., Vile, D. & Garnier, E. 2006. From plant traits to plant communities: a statistical mechanistic approach to biodiversity. *Science* **314**: 812–814.

Silva, E.L. & Lisboa, P. 2007. Analysis of the characteristic features of the density functions for gamma, Weibull and log-normal distributions through RBF network pruning with QLP. In: *Proceedings of the 6th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases*, pp. 223–228.

Turelli, M. 1977. Random environments and stochastic calculus. *Theor. Popul. Biol.* **12**: 140–178.

Ulrich, W., Ollik, M. & Ugland, K.I. 2010. A meta-analysis of species–abundance distributions. *Oikos* **119**: 1149–1155.

Watterson, G.A. 1974. Models for the logarithmic species–abundance distributions. *Theor. Popul. Biol.* **6**: 217–250.

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

Wilson, W.G. & Lundberg, P. 2004. Biodiversity and the Lotka–Volterra theory of species interactions: open systems and the distribution of logarithmic densities. *Proc. R. Soc. Lond. B Biol. Sci.* **271**: 1977–1984.

Appendix

Pueyo *et al.* (2007) used maximum entropy to derive the log series distribution for species abundances. Their maximum entropy method applied Jaynes' (2003) concepts of invariance and symmetry to derive the log series pattern. From the log series, Pueyo *et al.* (2007) added an additional constraint to obtain the gamma distribution. They showed that the gamma subsumes the log series, geometric, and power law distributions and is close to the lognormal. These points are the same ones that I have emphasized throughout my paper. In this regard, Pueyo *et al.* (2007) deserve full credit for the origins of these ideas and their application to species abundances. Given this clear precedent, my presentation in this paper extends the topic in two ways.

First, Pueyo *et al.* (2007) begin with the same concepts of invariance and symmetry from Jaynes (2003) on which I based my own approach. The Jaynesian approach is itself a slight extension of Jeffreys' (1957) Bayesian notion of prior distributions, in which one maximizes a measure of entropy that is taken relative to a prior description of what is most random for a given problem.

Pueyo *et al.* (2007) emphasize that their particular explanation of the prior distribution for ecological problems is their most important contribution. On page 1023, they say

> We have shown that common shapes of SADs can be predicted from extremely general assumptions. This conclusion is extensive to common shapes of SARs [species area relations], because these shapes are mathematically related to the SADs we found (Pueyo, 2006). We expect more findings to follow, because we think we have correctly identified the prior distribution (eqn 7), which is the Rosetta Stone that allows translating concepts between statistical physics and macroecology.

The way in which Pueyo *et al.* developed the prior following Jaynes was indeed the state of the art in 2007. However, the particular derivation they gave and their particular result was just a long way of arriving exactly at the famous Jeffreys prior, in which the 'most random' expression of a variable, n, is a probability distribution proportional to n^{-1}.

Pueyo *et al.* (2007) present various specific derivations of the Jeffreys prior based on particular ecological assumptions. In one example that they highlight, spatial and geometric aspects play a role in their derivation. Although they emphasize that the Jeffreys prior is a very general expression of invariance and symmetry, at the same time, they tie their expression of their key result to rather specific ecological descriptions.

I followed the same broad concepts, but the limitations in the approach of Pueyo *et al.* (2007) are important. The Jeffreys approach is based on a Bayesian notion of relative entropy with a prior notion of randomness. Our work has shown that view to be too limiting in understanding probability (Frank & Smith, 2010, 2011). In the particular case of the Jeffreys prior, all that is involved is the assumption of ratio invariance for measurements, leading to the natural scale for information as logarithmic. In this regard, Pueyo *et al.* (2007) are too specific in claiming the association between their particular ecological motivations and logarithmic scaling, giving the false impression that they have found the 'Rosetta Stone' for ecological pattern, when in fact all that they have done is give some specific examples in which information is properly measured on a logarithmic scale, but without a clear notion of the general role of measurement.

The understanding of ecological pattern will remain confused as long as one conflates the general issue of measurement scale with the specifics of certain ecological examples. That confusion will prevent the full conceptual power of maximum entropy and measurement from being appreciated in its application to ecology. In this paper, I worked toward joining the principles of measurement theory to ecological pattern.

My second extension to Pueyo *et al.* (2007) concerns how one can derive the gamma distribution, the general expression of SAD pattern. Pueyo *et al.* (2007) first derived, too specifically, the Jeffreys prior. From that prior, and a constraint on mean abundance, they obtain the log series. They then recognize that the log series is too limited to describe SAD pattern. To extend to the gamma, they note that the log series constrains the geometric mean when given arithmetic mean abundance. To get the gamma from the log series, they realize that the geometric mean must be allowed to vary independently of the arithmetic mean, so, ad hoc, they allow the geometric mean to vary independently.

Once the arithmetic and geometric means can vary independently, one has the well-known description of the gamma distribution as arising when arithmetic and geometric means are sufficient statistics within a context of maximum entropy subject to constraint (Kapur, 1989). Although a correct expression of the relation between constraint and pattern, that ends up just being a description of the gamma rather than a derivation of the pattern from a principled way of understanding how pattern arises. By that approach, any probability distribution can be obtained by tautologically invoking the sufficient statistics that are the constraints under maximum entropy. One loses any claim to deriving pattern from fundamental principles.

By contrast, Eric Smith and I (Frank & Smith, 2010, 2011) have replaced the limited Jeffreys–Jaynes approach to relative entropy priors with a principled notion of measurement scale based on the fundamental concepts of measurement theory (Hand, 2004). We started with the standard principles of measurement theory and then developed a novel extension of measurement to show the structural relations between a variety of common measurement scales. Those structurally related measurement scales combined with maximum entropy encompass essentially all of the common families of probability distributions. By contrast, Pueyo *et al.* (2007), by following the standard implementation of the Jeffreys–Jaynes approach, found that their log series expression had to be supplemented with an ad hoc assumption to get to the desired gamma pattern. In that one arbitrary step, they are already outside of a coherent framing of invariance, symmetry, and pattern.

All of this may seem more important for probability and mathematics than for ecology. However, understanding ecological pattern depends on understanding how aggregation, measurement, and randomness set the basic contours of pattern in nature. In this regard, the recent influx of maximum entropy concepts into ecology is a welcome step. But if those concepts are developed in ecology in a limited way or, more commonly, in a way that confuses the specific and the general, the net result will be the tendency to reject maximum entropy as a failed or confused approach. That would be a mistake, because ecological understanding will necessarily be limited if the field does not properly incorporate the fundamental principles of measurement and information.