bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Invariance in ecological pattern

Steven A. Frank[1] and Jordi Bascompte[2]

The abundance of different species in a community often follows the log series distribution. Other ecological patterns also have simple forms. Why does the complexity and variability of ecological systems reduce to such simplicity? Common answers include maximum entropy, neutrality, and convergent outcome from different underlying biological processes. This article proposes a more general answer based on the concept of invariance, the property by which a pattern remains the same after transformation. Invariance has a long tradition in physics. For example, general relativity emphasizes the need for the equations describing the laws of physics to have the same form in all frames of reference. By bringing this unifying invariance approach into ecology, we show that the log series pattern dominates when the consequences of processes acting on abundance are invariant to the addition or multiplication of abundance by a constant. The lognormal pattern dominates when the processes acting on net species growth rate obey rotational invariance (symmetry) with respect to the summing up of the individual component processes. Recognizing how these invariances connect pattern to process leads to a synthesis of previous approaches. First, invariance provides a simpler and more fundamental maximum entropy derivation of the log series distribution. Second, invariance provides a simple derivation of the key result from neutral theory: the log series at the metacommunity scale and a clearer form of the skewed lognormal at the local community scale. The invariance expressions are easy to understand because they uniquely describe the basic underlying components that shape pattern.

"It was Einstein who radically changed the way people thought about nature, moving away from the mechanical viewpoint of the nineteenth century toward the elegant contemplation of the underlying symmetry [invariance] principles of the laws of physics in the twentieth century" (ref. 1, p. 153).

Introduction

Ecologists have been interested in species abundance distributions (SADs) since the classic papers by Fisher[2] and Preston[3]. Two major patterns have been identified depending on the size of the community. In a large community, abundances often follow the log series distribution[4]. Specifically, the probability that a species has a population size of n individuals follows p^n/n. Communities differ only in their average population size, described by the parameter, p. At smaller spatial scales, the species abundance pattern often follows a skewed lognormal (a random variable is lognormally distributed when its logarithm is normally distributed)[5,6].

It is intriguing that the species abundance distribution follows these simple patterns irrespective of the particular group (birds, insects, mammals) and region considered. Other ecological patterns also follow simple probability distributions[7–9]. Those patterns have attracted a lot of attention. Why does the variability and complexity of biology reduce to such a small range of simple distributions? How can we understand the relations between complex processes and simple patterns?

Approaches such as Harte's maximum entropy formalism[9] and Hubbell's neutral theory[5] have attempted to explain the generality of the log series and skewed lognormal patterns in species abundance distributions. Maximum entropy describes probability distributions that are maximally random subject to satisfying certain constraints[10–12]. This approach has a long tradition in physics, both in statistical me-

[1] Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697–2525, USA

[2] Department of Evolutionary Biology and Environmental Studies, University of Zurich, 8057 Zurich, Switzerland

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

chanics and information theory. An early maximum entropy approach in ecology derived the biomass pattern of populations[13–15].

Neutral theory derives probability distributions by assuming that all individuals are equivalent[16]. Variation arises by random processes acting on the mechanistically identical individuals. Put another way, the mechanistic processes are "neutral" apart from random processes. Both maximum entropy and neutral theory have been shown to provide a good fit to the empirical patterns of species abundance distributions. In this article, we subsume these two different ways of understanding the log series and skewed lognormal patterns with a more general perspective based on the concept of invariance[17].

Invariance can be defined as the property by which a system remains unchanged under some transformation. For example, a circle is the same (invariant) before and after rotation (Fig. 1a). In ecology, pattern often depends on the ways in which form remains invariant to changes in measurement. Some patterns retain the same form after uniformly stretching or shrinking the scale of measurement (Fig. 2b). Measures of length provide a common example of stretch invariance. One can measure lengths equivalently in millimeters or centimeters without loss of information. As we will see, that kind of invariance often determines the form of observed pattern.

To give another example, consider the common and widely familiar pattern of the normal distribution. By the central limit theorem, when independent random variables are added, their properly normalized sum tends toward a normal distribution, even when the component variables themselves are not normally distributed. The central limit theorem and the normal distribution are often considered as unique aspects of pattern that stand apart from other commonly observed patterns.

The invariance perspective that we promote shows how the normal distribution is in fact a specific example of a wider framework in which to understand the commonly observed patterns of nature. In particular, the normal distribution arises from the rotational invariance of the circle[18]. For two variables, x and y, with a given squared length, $x^2 + y^2 = r^2$, all combinations of the variables with the same radius, r, lie along the circumference of a circle (Fig. 1a). When each combination is equally likely, the rotationally invariant radius is sufficient to describe the probability pattern.

It is this rotational invariance that gives the particular mathematical form of the normal distribution, in which the average squared radius sets the variance of the distribution. By this perspective, the mathematical forms of all commonly observed distributional patterns express their unique invariances[18].

The perspective of invariance was the basis for most of the great conceptual advances of physics in the twentieth century[1]. For example, Gell-Mann's pioneering theoretical work on the fundamental particles of nature derived from invariance (symmetry) properties that unified understanding of known particles and predicted new particles such as quarks, which were subsequently observed. By contrast, general aspects of invariance have not been used consistently as the fundamental basis for understanding patterns in ecology. One exception concerns scale invariance, which is often discussed in ecology[19–21]. But scale invariance is typically limited to special kinds of patterns rather than forming a unified approach to diverse patterns.

The point of this paper is that invariance is the most general way in which to understand commonly observed patterns. Species abundance distributions provide an excellent illustration. For species abundances, we show that maximum entropy and neutral models can succeed in certain cases because they derive from invariance principles. However, maximum entropy and neutrality are often difficult to interpret because they hide their underlying basis in invariance.

Our unifying invariance analysis clarifies why seemingly different conceptual approaches have the same consequences for pattern. Similarly, seemingly different biological processes may often lead to the same observed pattern, because those different processes share a common basis in invariance. That deeper understanding suggests a more insightful way to think about alternative mechanistic models. It also suggests the kinds of empirical tests that may differentiate between alternative causal processes.

This manuscript is organized as follows. First, we highlight new theoretical results for species abundance distributions. Second, we review how invariance defines probability patterns in a general

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

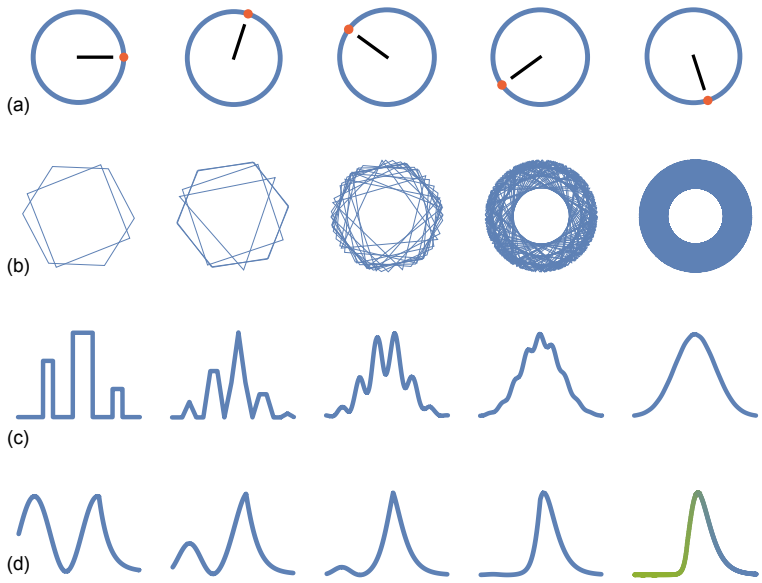

Figure 1: Rotational and asymptotic invariance. (a) Transforming a circle by rotation leaves the circle unchanged (invariant), with an invariant radial distance at all points along the circumference. (b) Rotating regular polygons changes pattern. However, as more rotated polygons are added, the form converges asymptotically to a rotationally invariant circle, in which adding another rotated polygon does not change the pattern. Many common patterns of nature are asymptotically invariant. In this case, aggregation causes loss of all information except invariant radial distance. (c) The normal distribution is asymptotically invariant. The left curve describes an arbitrary probability pattern. The second curve expresses the sum of two randomly chosen values from the first curve. The height is the relative probability of the summed values. The third, fourth, and fifth curves express the sum of 4, 8, and 16 randomly chosen values from the first curve. Each curve width is shrunk to match the first curve. In this case, aggregation smooths the curve, causing loss of all information except the average squared distance from the center (the variance), which is equivalent to the average squared radial distance of rotationally invariant circles. (d) Extreme value distributions are asymptotically invariant. The left curve is an arbitrarily chosen probability pattern. The second curve expresses the probability of the largest value in a sample of two randomly chosen values from the first curve. The third, fourth, and fifth curves show the probability of the largest value of 4, 8, and 16 randomly chosen values. The asymptotically invariant curve on the right expresses exponential scaling at small values and linear scaling at large values, labeled in green and blue. Commonly observed probability distributions often express simple combinations of linear, logarithmic, and exponential scaling. Panels (a-c) modified from ref. 17.

way[18,22,23]. The log series distribution[24] and the new gamma-lognormal distribution for species abundances follow directly from the universal invariance expression of probability patterns. Third, we show that maximum entropy and neutrality can easily be understood as special examples of invariance principles. Finally, we discuss the broad role of invariance in the analysis of ecological pattern.

New results

This article develops two new theoretical results. We highlight those results before starting on the general overview of invariance and pattern.

First, we present a simple maximum entropy derivation of the log series pattern. We show that constraining the average abundance per species is sufficient when analyzing randomness and entropy on the proper demographic scale.

The simplicity of our maximum entropy derivation contrasts with Harte's more complicated maximum entropy model[9,25]. Harte had to assume an additional unnecessary constraint on energy usage. He required that unnecessary constraint because he evaluated randomness on the scale of measured abundances rather than on the scale of demographic process. This will be made explicit below.

We use this new result to demonstrate that max-

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

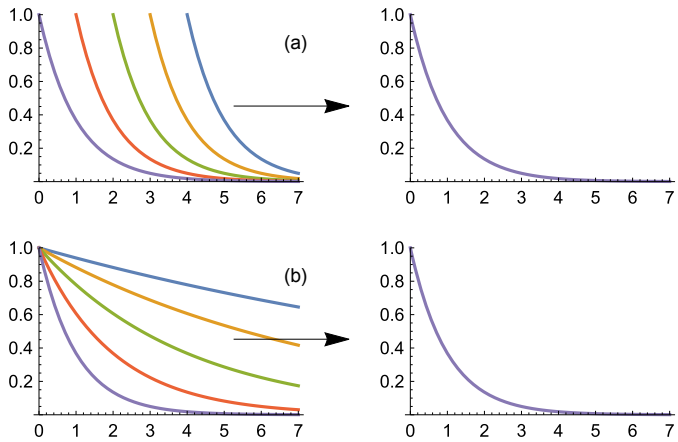

Figure 2: Shift and stretch invariance of the exponential distribution. (a) The left panel shows $e^{-(x+a)}$ for $a = 0, -1 \ldots, -4$. Decreasing values of a shift the curve to the right, which is equivalent to shifting the x axis by resetting the zero point. For probability patterns, the total probability must be normalized to one, which means that all curves must have the same area under the curve for values of x between 0 and ∞. To normalize the curves, the right panel plots $k_a e^{-(x+a)}$ with $k_a = e^a$. Thus, all curves become e^{-x} invariantly with respect to different shift values, a. (b) The left panel shows e^{-bx} for $b = 2^0, 2^{-1} \ldots, 2^{-4}$. Decreasing values of b stretch the x axis by a factor of 2 for each halving of b. To normalize the average value of each probability curve to be the same, the right panel shows $e^{-\lambda_b bx}$ for $\lambda_b = 1/b$. Thus, all curves become e^{-x} invariantly with respect to different stretch values, b.

imum entropy is the outcome of deeper underlying principles of invariance and pattern. By working at the deeper level of invariance, one obtains a simpler and more powerful understanding of pattern.

The second new result shows that Hubbell's[5] neutral model is the simple expression of three basic invariances. Hubbell's full range of log series and skewed lognormal (zero sum multinomial) results follows immediately from those three underlying invariances.

The three invariances correspond to a maximum entropy model that constrains the average abundance of species and the average and variance of the demographic processes influencing abundance. The three invariances lead to a simple gamma-lognormal distribution that matches the neutral theory pattern for species abundances. The gamma-lognormal is a product of the standard gamma and lognormal distributions.

Invariance

This section reviews how invariance considerations lead to the log series distribution[24]. We delay discussion of the gamma-lognormal until the later section on Hubbell's neutral model.

Canonical form of probability distributions

We can rewrite almost any probability distribution as

$$q_z = k e^{-\lambda T_z}, \tag{1}$$

in which $T(z) \equiv T_z$ is a function of the variable, z, and k and λ are constants. For example, Student's t-distribution, usually written as

$$q_z = k\left(1 + z^2/\nu\right)^{-(\nu+1)/2}$$

can be written in the form of eqn 1 with $\lambda = (\nu + 1)/2$ and $T_z = \log\left(1 + z^2/\nu\right)$.

The probability pattern, q_z, is invariant to a constant shift, $T_z \mapsto a + T_z$, because we can write the transformed probability pattern in eqn 1 as

$$q_z = k_a e^{-\lambda(a+T_z)} = k e^{-\lambda T_z},$$

with $k = k_a e^{-\lambda a}$ (Fig. 2a). We express k in this way because k adjusts to satisfy the constraint that the total probability be one. In other words, conserved total probability implies that the probability pattern is shift invariant with respect to T_z (see ref. 18).

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Now consider the consequences if the average of some value over the distribution q_z is conserved. For example, the average of z is the mean, μ, and the average of $(z-\mu)^2$ is the variance. A constraint causes the probability pattern to be invariant to a multiplicative stretching (or shrinking), $T_z \mapsto bT_z$, because

$$q_z = ke^{-\lambda_b b T_z} = ke^{-\lambda T_z},$$

with $\lambda = \lambda_b b$ (Fig. 2b). We specify λ in this way because λ adjusts to satisfy the constraint of conserved average value. Thus, invariant average value implies that the probability pattern is stretch invariant with respect to T_z.

Conserved total probability and conserved average value cause the probability pattern to be invariant to an affine transformation of the T_z scale, $T_z \mapsto a + bT_z$, in which "affine" means both shift and stretch.

The affine invariance of probability patterns with respect to T_z induces significant structure on the form of T_z and the associated form of probability patterns. Understanding that structure provides insight into probability patterns and the processes that generate them[18,22,23].

In particular, Frank & Smith[22] showed that the invariance of probability patterns to affine transformation, $T_z \mapsto a + bT_z$, implies that T_z satisfies the differential equation

$$\frac{\mathrm{d}T_z}{\mathrm{d}w} = \alpha + \beta T_z,$$

in which $w(z)$ is a function of the variable z. The solution of this differential equation expresses the scaling of probability patterns in the generic form

$$T_z = \frac{1}{\beta}\left(e^{\beta w} - 1\right), \tag{2}$$

in which, because of the affine invariance of T_z, we have added and multiplied by constants to obtain a convenient form, with $T_z \to w$ as $\beta \to 0$.

By writing T_z in this way, w expresses a purely shift-invariant aspect of the fundamental affine-invariant scale, because the shift transformation $w \mapsto a + w$ multiplies T_z by a constant, and probability pattern is invariant to constant multiplication of T_z. Thus, eqn 2 dissects the anatomy of a probability pattern (eqn 1) into its component invariances.

With this expression for T_z, we may write probability patterns generically as

$$q_z = ke^{-\lambda\left(e^{\beta w}-1\right)/\beta}. \tag{3}$$

This form has the advantage that $w(z)$ expresses the shift-invariant structure of a probability pattern. Most of the commonly observed probability patterns have a simple form for w[23,26]. That simplicity of the shift-invariant scale suggests that focus on w provides insight into common patterns.

Proportional processes and species abundances

To understand the log series, we must consider the relation $n = e^r$ between the observed pattern of abundances, n, and the processes, r. Here, r represents the total of all proportional processes acting on abundance[24].

A proportional process simply means that the number of individuals or entities affected by the process increases in proportion to the number currently present, n. Demographic processes, such as birth and death, act proportionally.

The sum of all of the proportional processes on abundance over some period of time is

$$r = \int_0^\tau m(t)\mathrm{d}t.$$

Here, $m(t)$ is a proportional process acting at time t to change abundance. Birth and death typically occur as proportional processes. The value of $r = \log n$ is the total of the m values over the total time, τ. For simplicity, we assume $n_0 = 1$.

The log series follows as a special case of the generic probability pattern in eqn 3. To analyze abundance, focus on the process scale by letting the variable of interest be $z \equiv r$, with the key shift-invariant scale as simply the process variable itself, $w(r) = r$. Then eqn 3 becomes

$$q_r\mathrm{d}r = ke^{-\lambda\left(e^{\beta r}-1\right)/\beta}\,\mathrm{d}r, \tag{4}$$

in which $q_r\mathrm{d}r$ is the probability of a process value, r, in the interval $r + \mathrm{d}r$.

Using $w(r) = r$ sets the the shift-invariant scale as the variable itself. Substituting this simplest form for the shift-invariant scale into the canonical equation

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

for common probability patterns in eqn 3 yields the simplest generic expression of probability pattern as eqn 4.

We can generalize the relation between abundance and process, $n = e^r$, by writing $n^\beta = e^{\beta r}$, which uses an additional parameter β to allow comparison with the canonical form of probability distributions in the previous subsection. When we focus on standard models of species abundances, we use $\beta = 1$.

We can change from the process scale, r, to the abundance scale, n, by noting that $\beta \log n = \beta r$, and so, for any β, we have $r = \log n$. Thus, we can use the substitutions $r \mapsto \log n$ and $\mathrm{d}r \mapsto n^{-1}\mathrm{d}n$ in eqn 4, yielding the identical probability pattern expressed on the abundance scale

$$q_n \mathrm{d}n = kn^{-1}e^{-\lambda(n^\beta - 1)/\beta}\,\mathrm{d}n. \tag{5}$$

The value of k always adjusts to satisfy the constraint of invariant total probability, and the value of λ always adjusts to satisfy the constraint of invariant average value.

For proportional processes and species abundances, $\beta = 1$, as noted above. For that value of β, we obtain the log series distribution[24]

$$q_n = kn^{-1}e^{-\lambda n}, \tag{6}$$

replacing $n - 1$ by n in the exponential term which, because of affine invariance, describe the same probability pattern. The log series is often written with $e^{-\lambda} = p$, and thus $q_n = kp^n/n$. One typically observes discrete values $n = 1, 2, \ldots$. See the Appendix for the general relation between discrete and continuous distributions. The continuous analysis here is sufficient to understand pattern.

We can also write the log series on the process scale, r, from eqn 4, as[24]

$$q_r = ke^{-\lambda e^r}. \tag{7}$$

This form shows that the log series is the simplest expression of generic probability patterns in eqn 3. The log series arises from $\beta = 1$, associated with $n = e^r$, and from the base shift-invariant scale as $w \equiv r$ for proportional processes, r.

Invariances of the log series

This subsection summarizes a few technical points about invariance. These technical points provide background for our simpler and more general derivation in the following section of maximum entropy models for species abundances. Those prior models focused only on abundances, n, without considering the underlying process scale, r.

We begin with invariance on the process scale, r. On that scale, the log series in eqn 7 is the pure expression of additive shift invariance to r and lack of multiplicative stretch invariance to r. For example, note in eqn 7 that an additive change, $r \mapsto r + a$, is compensated by a change in λ to maintain the overall invariance, whereas a multiplicative change, $r \mapsto br$, cannot be compensated by a change in one of the constants. For example, if r is net reproductive rate, then an improvement in the environment that adds a constant to everyone's reproductive rate does not alter the log series pattern. By contrast, multiplying reproductive rates by a constant does alter pattern.

To understand the parameter, β, from eqn 2, consider that

$$T = \frac{1}{\beta}\left(e^{\beta r} - 1\right) = \frac{1}{\beta}\left(n^\beta - 1\right),$$

in which β is the relative curvature of the measurement scale for abundance, n, with respect to the scale for process, r. The relative curvature is $\beta = T''/T'$, with the primes denoting differentiation with respect to r.

For the log series, the curvature of $\beta = 1$ describes the amount of bending of the abundance scale, $n = e^r$, with respect to multiplying the process scale, r, by a constant—the departure from stretch invariance.

The simple invariances with respect to process, r, become distorted and more difficult to interpret when we focus only on the observed scale for abundance, n, associated with the log series in eqn 6. In that form of the distribution, the canonical scale is

$$T = \frac{1}{\lambda}\log n + n. \tag{8}$$

In this expression, purely in terms of abundances, the logarithmic term dominates when n is small, and the linear term dominates when n is large. Thus, the scale changes from stretch but not shift invariant at small magnitudes to both shift and stretch invariant at large magnitudes[24].

Without the simple insight provided by the process

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

scale, r, we are left with a complicated and nonintuitive pattern that is separated from its simple cause. That difficulty has led to unnecessary complications in maximum entropy theories of pattern.

Maximum entropy

Maximum entropy describes probability distributions that are maximally random subject to satisfying certain constraints[10–12]. In eqn 1, with the generic description for distributions as

$$q_z \mathrm{d}z = k e^{-\lambda T_z} \mathrm{d}z,$$

maximum entropy interprets this form as the expression of maximum randomness with respect to the scale z, subject to the constraint that the average of T_z is fixed[23].

This section begins with a maximum entropy derivation for the log series based on our separation between the scales of process, r, and observed abundance, n.

We then discuss Harte's[9,25] alternative maximum entropy derivation of the log series. Harte's derivation emphasizes mechanistic aspects of energy constraints rather than our emphasis on the different scales of process and abundance.

Constraint of average abundance on process scale

The log series in eqn 7 is

$$q_r \mathrm{d}r = k e^{-\lambda e^r} \mathrm{d}r.$$

Here, $T = e^r = n$. This distribution expresses maximum entropy with respect to the process scale, r. The constraint is the ecological limitation on average abundance

$$\langle T \rangle_r = \langle e^r \rangle_r = \langle n \rangle_r, \tag{9}$$

in which $\langle \cdot \rangle_r$ denotes average value with respect to the process scale, r.

In this case, process values, r, are maximally random, subject to the ecological constraint that limits abundance, n. Thus, maximizing entropy with respect to the process scale, r, subject to a constraint on the observed pattern scale, n, leads immediately to the log series.

Relating the process scale, r, to the scale of ecological constraint, n, often makes sense. Typically, environmental perturbations associate with changes in demographic variables, such as birth and death rates. Such demographic factors typically act proportionally on populations, consistent with our interpretation of r as the aggregate of proportionally acting processes. The perturbations, acting on demographic variables, associate the process scale with the scale of randomness.

In contrast with the process scale of perturbation and randomness for the demographic variables, the scale of constraint naturally arises with respect to a limit on the number of individuals, n. Thus, randomness happens on the r scale and constraint happens on the n scale.

It is, of course, possible to formulate alternative models in which randomness and constraint happen on scales that differ from our interpretation. Different formulations are not intrinsically correct or incorrect. Instead, they express different assumptions about the relations between process, randomness, and invariance. The next section considers an alternative formulation.

Harte's joint constraints of abundance and energy

Harte developed comprehensive maximum entropy models of ecological pattern. He tested those theories with the available data. His work synthesizes many aspects of ecological pattern[9].

For species abundances, Harte[9,25] analyzed maximum randomness with respect to the scale of abundance values, n. Maximum entropy derivations commonly evaluate randomness on the same scale as the observations. In this case, with observations for the probabilities of abundances, p_n, entropy on the same scale is the sum or integral of $-p_n \log p_n$.

However, there is no a priori reason to suppose that the scale of observation is the same as the scale of randomness. The fact that observation, randomness, and process may occur on different scales often makes maximum entropy models difficult to develop and difficult to interpret. For example, we may observe the probabilities of abundances, p_n, but randomness may be maximized on the scale of process, as the sum or integral of $-p_r \log p_r$.

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In the final part of this section, we argue that invariance provides a truer path to the natural scale of analysis and to the mechanistic processes that generate pattern than does maximum entropy. Before comparing invariance and maximum entropy, it is useful to sketch the details of Harte's maximum entropy model for species abundances.

The simplest maximum entropy model analyzes entropy with respect to abundance, n, subject to a constraint on the average abundance, $\langle n \rangle$. That analysis yields an exponential distribution

$$q_n \mathrm{d}n = ke^{-\lambda n}\mathrm{d}n.$$

The exponential pattern differs significantly from the observed log series pattern. Thus, maximizing entropy with respect to the scale of abundance, n, and constraining the average abundance is not sufficient.

From our invariance perspective, it is natural to think of the scale of randomness in terms of $\mathrm{d}r$, the scale of proportional processes, rather than in terms of $\mathrm{d}n$, the scale of abundance. Maximizing randomness with respect to $\mathrm{d}r$ leads directly to the log series, as shown in the previous section.

Harte did not consider the distinction between the exponential and log series patterns with respect to the scale of randomness. Instead, to go from the default exponential pattern of maximum entropy to the log series, his maximum entropy analysis required additional assumptions. He proceeded in the following way.

Suppose that the total quantity of some variable, ϵ, is constrained to be constant over all individuals of all species. The average value per individual is $\langle \epsilon \rangle$. It does not matter what the variable ϵ is. All that matters is that the constraint exists. Harte assumed that ϵ is energy, but that assumption is unnecessary with regard to the species abundance distribution.

The value ϵ is distributed over individuals independently of their species identity. Thus, the variable $\delta|n = n\epsilon$ is the total value in a species with n individuals, with average value $\langle \delta|n \rangle = n\langle \epsilon \rangle$.

The joint distribution of n and δ is

$$q_{n,\delta} = q_n q_{\delta|n}.$$

The explicit form of this joint distribution can be obtained by maximizing entropy subject to the constraints on the average abundance per species, $\langle n \rangle$, and the average total value in a species with n individuals, $\langle \delta|n \rangle$, yielding

$$q_{n,\delta} = ke^{-\lambda n}e^{-\lambda'\delta}.$$

We obtain the form presented by Harte[25] using the equivalence $\delta = n\epsilon$, yielding

$$q_{n,\epsilon} = ke^{-\lambda n}e^{-\lambda' n\epsilon}.$$

The species abundance distribution is obtained by

$$q_n = \int q_{n,\epsilon}\mathrm{d}\epsilon = ke^{-\lambda n}\int e^{-\lambda' n\epsilon}\mathrm{d}\epsilon.$$

Noting that $\int e^{-\lambda' n\epsilon} = 1/\lambda' n$, and absorbing the constant λ' into k, we obtain the log series for the species abundance distribution

$$q_n = kn^{-1}e^{-\lambda n}.$$

Maximum entropy and invariance

Harte's maximum entropy derivation of the log series assumes joint constraints of abundance, n, and some auxiliary variable, ϵ, which he labeled as energy. He evaluated entropy on the scales of n and ϵ.

By contrast, our invariance derivation arises from a constraint on abundance plus evaluation of invariance or entropy on the scale $r = \log n$. On that scale, the log series arises in a simple and clear way. There is no need for constraint of a second auxiliary variable.

Without an invariance argument, nothing compels us to analyze with respect to the r scale. Harte, without focus on invariance, followed the most natural approach of using n as the scale for maximization of randomness and for constraint. That approach required an auxiliary constraint on a second scale to arrive at the log series.

Harte's approach was a major step in unifying the analysis of empirical pattern. But, in retrospect, his approach was unnecessarily complicated.

One might say that Harte's approach provided a richer theory because it led to predictions about both abundance and energy. However, the data on abundance patterns match very closely to the log series, whereas the data for different proxies of energy vary considerably[9].

Our invariance approach strips away the unnecessary auxiliary variable. The invariance theory therefore provides a much simpler way to derive and to

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

understand abundance patterns.

Maximum entropy can be thought of purely as a basic invariance method of analysis. Maximum entropy distributions have the form in eqn 1 as

$$q_z \mathrm{d}z = ke^{-\lambda T_z} \mathrm{d}z,$$

in which T_z is the affine-invariant scale that defines the probability pattern. Thus, the method of maximum entropy is simply a method for deriving the affine-invariant expression, T_z. In practice, maximum entropy has three limitations.

First, maximum entropy is silent with respect to the proper choice for the scale on which entropy is maximized and the constraints that set the affine-invariant expression, T_z. By contrast, focus on invariance led us to the shift invariance of the process scale, r. That scale provided a much simpler analysis, in which r is the incremental scale with respect to invariance and the measurement scale with respect to entropy.

In other words, maximum entropy is a blind application of the most basic invariance principles, without any guidance about the proper scales for invariance, randomness, and constraint. By contrast, an explicit invariance approach takes advantage of the insight provided by the analysis of invariance.

Second, by focusing on invariance, we naturally obtain the full invariance (symmetry) group expression in eqn 3 as the generic form of probability patterns

$$q_z = ke^{-\lambda\left(e^{\beta w}-1\right)/\beta}.$$

That generic expression leads us to a generalization of the log series in eqn 5 as[24]

$$q_n \mathrm{d}n = kn^{-1}e^{-\lambda\left(n^{\beta}-1\right)/\beta}\,\mathrm{d}n,$$

which is a two parameter distribution for abundances with respect to λ and β. The log series is a special case with $\beta = 1$.

Third, invariance leads to a deeper understanding of the relation between observed pattern and alternative mechanistic models of process. The following section provides an example.

Neutrality

Here, we analyze Hubbell's[5] neutral model of species abundances in the light of our invariance perspective. With that example in mind, we then discuss more generally how neutral models relate to invariance and maximum entropy.

Hubbell's neutral model

The strong recent interest in Hubbell's neutral model follows from the match of the theory to the contrasting patterns of species abundance distributions (SADs) that have been observed at different spatial scales. In the theory, many local island-like communities are connected by migration into a broader metacommunity. Sufficiently large metacommunities follow the log series pattern of species abundances. Each local community follows a distribution that Hubbell called the zero-sum multinomial[27], which is similar to a skewed lognormal. As noted by Rosindell et al.[6], it is this flexibility of the classic neutral model to reconcile the log series and lognormal distributions that allows it to fit empirical data well[28].

Invariance and the gamma-lognormal distribution

Broad consensus suggests that species abundances closely follow the log series pattern at large spatial scales. Extensive data support that conclusion[4].

Observed pattern at small spatial scales differs from the log series. Consensus favors a skewed lognormal pattern. The data typically show an excess of rare species, causing a skew relative to the symmetry of the lognormal when plotted on a logarithmic scale.

At small spatial scales, most recent analyses focus on data from a single long-term study of tree species in Panama[5,27]. Thus, some ambiguity remains about the form and consistency of the actual pattern at small scales.

The blue curve of Fig. 3 shows Chisholm and Pacala's[27] fit of the neutral theory to the Panama tree data for species abundances at small spatial scales. The gold curve shows the close match to the neutral theory pattern by a simple probability distribution derived from the analysis of invariance.

To obtain the matching distribution derived by invariance, we begin with the canonical form for probability distributions in eqn 3. That canonical form expresses pattern in terms of the shift-invariant scale, w. Next, we need to find the specific form of the scale

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

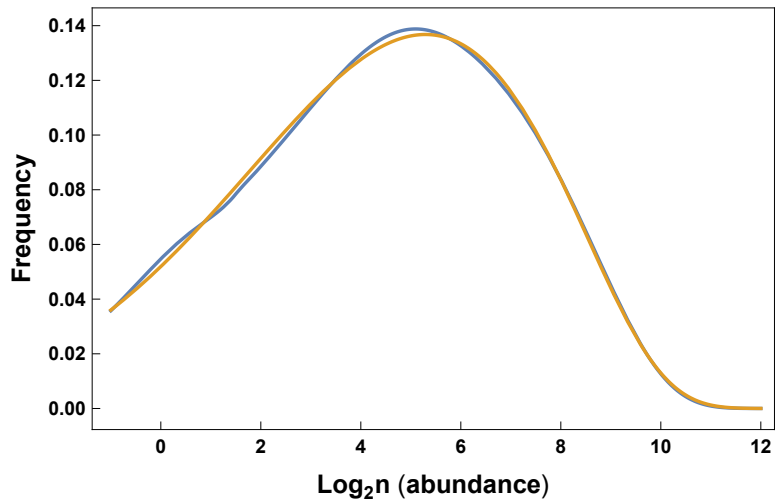

Figure 3: Match of the gamma-lognormal pattern in gold to the neutral theory fit in blue for Panama tree species abundances. The neutral theory fit to the data comes from Chisholm and Pacala's[27] analysis in their Fig. 1. They used Hubbell's neutral theory model with parameters $J = 21,060$, $m = 0.075$, and $\theta = 52.1$ in their eqn 3, originally from Alonso and McKane[29]. The gamma-lognormal model in eqn 11 produces essentially the identical pattern with parameters $\lambda = 0.00205$, $a = 0.491$, and $\alpha = 0.0559$. The abundance scale can be expressed equivalently on the process scale, $\log_2 n = r/\log 2$. See the Data accessibility statement for the calculations used to produce this plot.

w that relates this canonical form for probability distributions to the neutral theory. Because the neutral theory derives abundance, n, as an outcome of demographic processes, r, the fundamental shift-invariant scale for neutral theory is expressed in terms of the demographic process variable as

$$w = \log\left(e^r - \frac{a}{\lambda}r + \frac{\alpha}{\lambda}r^2\right). \tag{10}$$

Below, we discuss why this is a natural shift-invariant scale for neutral theory. For now, we focus on the details of the mathematical expressions. Recall that $n = e^r$ relates measured abundances, n, to the demographic process scale, r. If we assume that $\beta = 1$ in eqn 3 and use w from eqn 10, we obtain

$$q_r = ke^{-\lambda e^r + ar - \alpha r^2}, \tag{11}$$

with parameters λ, a, and α. We can write this distribution equivalently on the n scale for abundance as

$$q_n = kn^{\tilde{a}-1}e^{-\lambda n}e^{-\alpha(\log n - \mu)^2}. \tag{12}$$

In the second distribution, $\mu = (a - \tilde{a})/2\alpha$. Thus, both distributions have the same three parametric degrees of freedom.

The right-hand exponential term of eqn 12 is a lognormal distribution with parameters μ and $\sigma^2 = 1/2\alpha$. The remaining terms are a gamma distribution with parameters $\tilde{a}$ and λ. We call this product of the gamma and lognormal forms the gamma-lognormal distribution.

Figure 3 showed that the gamma-lognormal distribution matches the neutral theory fit for the Panama tree data. Fig. 4 shows that the shape of the gamma-lognormal matches the shape of the neutral theory predictions for various mechanistic parameters of the neutral theory.

In summary, the neutral theory distribution appears to be nearly identical to a gamma-lognormal distribution when compared over realistic parameter values. Both distributions have the same three parametric degrees of freedom.

Maximum entropy and the gamma-lognormal

The constraints on pattern can be seen most clearly by rewriting eqn 11 as

$$q_r = ke^{-\lambda T_r} = ke^{-\lambda e^r + \tilde{a}r - \alpha\tilde{r}^2}, \tag{13}$$

in which $\tilde{r}^2 = (r - \mu)^2$ is the squared deviation from μ, in which μ is the average value of r. This expression remains a three-parameter distribution because, as noted above, $\mu = (a - \tilde{a})/2\alpha$.

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

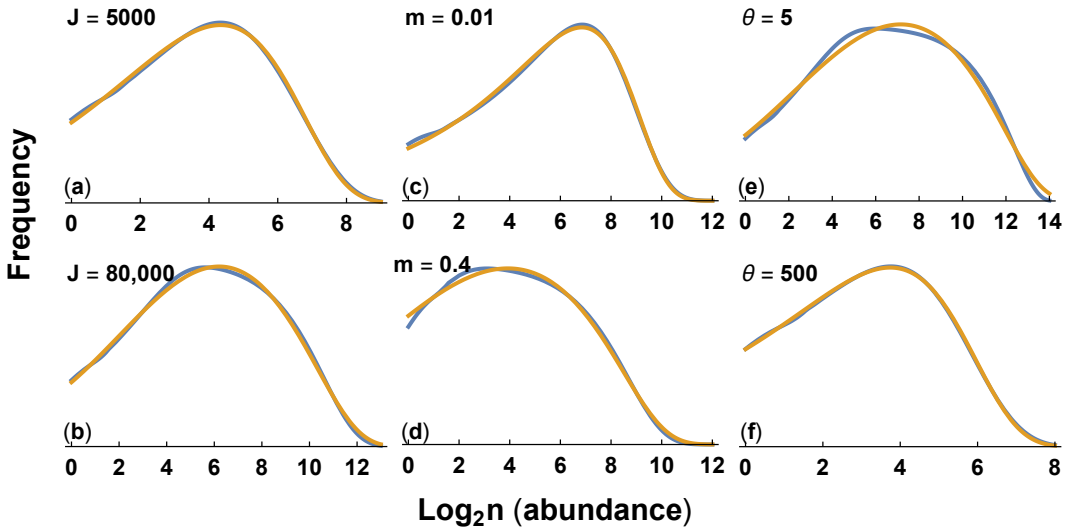

Figure 4: Match of Hubbell's neutral theory to the gamma-lognormal distribution. The blue curve for the neutral theory and the gold curve for the gamma-lognormal are calculated as described in Fig. 3. The parameters for the neutral theory are the same as in Fig. 3, except as shown in each panel. I fit the parameters for the gamma-lognormal to each neutral theory curve, with values for each panel: (a) $\lambda = 0.01115$, $a = 0.4452$, and $\alpha = 0.03660$; (b) $\lambda = 0.0004209$, $a = 0.4622$, and $\alpha = 0.05014$; (c) $\lambda = 0.002765$, $a = 0.3182$, and $\alpha = 0$; (d) $\lambda = 0.001777$, $a = 0.2217$, and $\alpha = 0.03576$; (e) $\lambda = 0.0001509$, $a = 0.3851$, and $\alpha = 0.03667$; (f) $\lambda = 0.02519$, $a = 0.3726$, and $\alpha = 0.006900$. See the Data accessibility statement for the calculations used to produce these plots.

With this set of parameters, the affine-invariant scale is

$$T_r = e^r - \frac{\tilde{a}}{\lambda} r + \frac{\alpha}{\lambda} \tilde{r}^2. \tag{14}$$

Note that T and w are related by eqn 2. We are using w from eqn 10 and $\beta = 1$, as noted below eqn 10. We ignore the extra -1 term in T of eqn 2, because the canonical form of probability distributions is invariant to adding a constant to T. The tilde parameters of the distribution in eqn 13 are interchangeable with the nontilde parameters of the identical distribution in eqn 12. The tilde expressions focus on the invariances that will help us to interpret ecological pattern. The nontilde expressions describe pattern in terms of the classic forms for the gamma and lognormal distributions.

By the standard theory of maximum entropy, q_r maximizes entropy on the incremental scale dr subject to a constraint on the average value of the defining affine-invariant scale, $\langle T \rangle_r$. That constraint is the linear combination of three constraints: the average abundance on the process scale, $\langle n = e^r \rangle_r$, the average demographic process value, $\langle r \rangle$, and the variance in the demographic process values, $\langle \tilde{r}^2 \rangle$.

By maximum entropy, all of the information in Hubbell's mechanistic process theory of neutrality and the matching gamma-lognormal pattern reduces to maximum randomness subject to these three constraints.

However, it is very unlikely that we would have derived the correct form by maximum entropy without knowing the answer in advance. This limitation emphasizes that maximum entropy provides deep insight into process and pattern, but often we need an external theory to guide our choice among various possible maximum entropy formulations.

Put another way, maximum entropy and process oriented theories, such as Hubbell's model, often work together synergistically to provide deeper insight than either approach alone.

Invariance, information and scale

Before turning to invariance and the gamma-lognormal pattern of neutral theory, it is useful to consider some basic properties of invariance and information[30,31]. In particular, this subsection develops our claim that the affine-invariant scale provides the deepest insights into the relations between pattern and process.

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We start by noting that, in the general expression for probability distributions

$$q_z = ke^{-\lambda T_z},$$

the affine-invariant scale, T_z, is equivalent to a common expression for the information content in a measurement, z, as

$$I_z = -\log q_z.$$

This expression follows from assuming that: information depends on the probability, q_z, of observing the measured value and not on the value itself; rarely observed values provide more information than commonly observed values; and the information in two independent measurements is the sum of the information in each measurement.

From the general expression for probability distributions

$$I_z = -\log q_z = -\log k + \lambda T_z.$$

Thus, an incremental change in information is equal to an incremental change in the affine-invariant scale

$$\mathrm{d}I_z = -\mathrm{d}\log q_z = \lambda \mathrm{d}T_z.$$

Equivalently, the change in information with respect to a change in the affine-invariant scale,

$$\frac{\mathrm{d}I_z}{\mathrm{d}T_z} = \lambda, \tag{15}$$

is constant at all magnitudes of the measurement, z. Every measured increment on the T_z scale provides the same amount of information about pattern. Constancy of information at all magnitudes is the ideal for a measurement scale. Thus, affine invariance provides the ideal scale on which to evaluate the pattern in measurements[23]. Figure 5 illustrates some key properties of the affine-invariant scale.

Information is sometimes thought of as a primary concept. However, it is important to understand that, in this context, information and affine invariance are the same thing. Neither is intrinsically primary.

We prefer to emphasize invariance, because it is an explicit description of the properties that pattern and process must obey[17,26,32]. Further analysis of invariant properties leads to deeper insight. For example, only through invariance can we obtain the group theory expression for the canonical form of probability patterns (eqn 3).

By contrast, "information" is just a vague word that associates with underlying invariances. Further analysis of information requires unwinding the definitions to return to the basic invariances.

Invariance interpretation of the gamma-lognormal

We turn now to the neutral theory model for abundances at local spatial scales. We showed that all of the information about pattern and process in the neutral theory is captured by the gamma-lognormal pattern in eqn 13 as

$$q_r = ke^{-\lambda T_r} = ke^{-\lambda e^r + \tilde{a}r - \alpha \tilde{r}^2},$$

which defines the affine-invariant scale in eqn 14 as

$$\lambda T_r = \lambda e^r - \tilde{a}r + \alpha \tilde{r}^2. \tag{16}$$

On this scale, changes in r provide the same amount of information about pattern at all magnitudes. Shifting the scale by a constant does not change the information about pattern in measurements. In other words, it does not matter where we set the zero point for T_r. Similarly, uniformly stretching or shrinking the scale, T_r, does not change the information in measurements of r.

We can parse the terms of eqn 16 with respect to constraint and invariance. When r is large, the term $\lambda e^r = \lambda n$ dominates the shape of the distribution in the upper tail, which decays as

$$q_r \mathrm{d}r = ke^{-\lambda e^r}\mathrm{d}r = ke^{-\lambda n}\mathrm{d}r$$

for sufficiently large $e^r = n$. The smaller the value of λ relative to $\tilde{a}$ and α, the greater e^r must be for this pattern to dominate. When λ is relatively large compared with $\tilde{a}$ and α, this pattern dominates at all magnitudes and leads to the log series.

With respect to constraint, for large values of abundance, n, the constraint on average abundances dominates the way in which altered process influences pattern. With respect to invariance, a process that additively shifts or multiplicatively stretches the $e^r = n$ values does not alter the pattern in the upper tail. Similarly, pattern is invariant to a process that additively shifts process values, r, but processes that multiplicatively change r alter pattern. Thus, we can evaluate the role of particular processes by considering

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

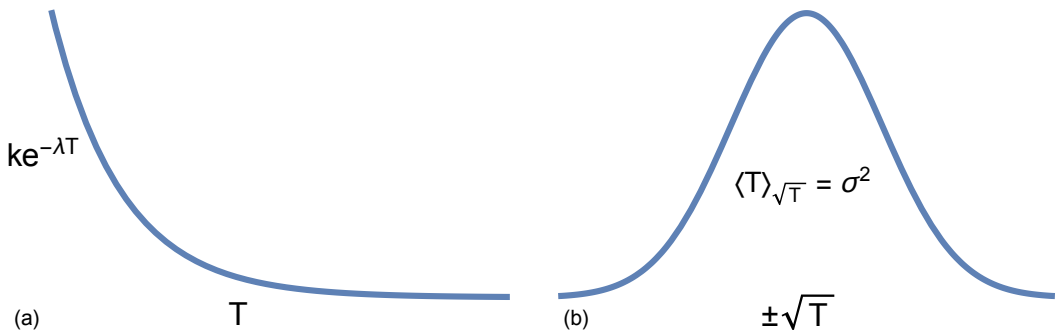

Figure 5: Continuous probability distributions can often be expressed as exponential or normal distributions with respect to the affine-invariant scale. A continuous distribution typically can be written as $q_z = ke^{-\lambda T_z}$, from eqn 1. In the figure, $T \equiv T_z$. (a) A parametric plot of q_z vs T_z is exponential. All of differences between probability distributions are contained in the form of the affine-invariant scale, T_z. The change in information for each increment of the affine-invariant scale is λ, as in eqn 15. (b) A parametric plot of q_z vs $\pm\sqrt{T_z}$ is normally distributed when describing the deviations from a unimodal peak of q_z. The average of the deviations on the affine-invariant scale, $\langle T \rangle$, relative to measurements on the square root of that scale, $\sqrt{T_z}$, is the variance, σ^2. For the normal distribution, we can think of a deviation from the central location on the affine-invariant scale, $T_z = R_z^2$, as the squared radial deviations along the circumference of a circle with radius R_z, describing the squared vector length for an aggregation of variables. The variance is the average of the squared radial deviations relative to the scale of radial measures, $\sqrt{T_z} = R_z$. Most continuous unimodal distributions are, in this way, equivalent to a normal distribution when scaled with respect to the square root of the affine-invariant measure. See refs. 17,18 for details.

how they change n or r.

The pattern at small and intermediate values of r depends on the relative sizes of the parameters. If the $\tilde{a}r$ term dominates, then the constraint, $\langle r \rangle$, on the average process value is most important. With respect to invariance when $\tilde{a}r$ dominates, a process that additively shifts or multiplicatively stretches the r values does not alter the pattern in the lower tail. That lower tail is a rising exponential shape, $e^{\tilde{a}r}$, as in Fig. 4c.

When the $\alpha\tilde{r}^2$ term is negligible at all magnitudes, the combination of the dominance by $\tilde{a}r$ in the lower tail, and the dominance by λe^r in the upper tail, yields the gamma distribution pattern on the abundance scale, n.

Finally, for magnitudes of r at which the $\alpha\tilde{r}^2 = \alpha(r-\mu)^2$ term dominates, the constraint, $\sigma^2 = \langle r - \mu^2 \rangle$, on the variance in process values is most important. In this case, pattern follows a normal distribution, $e^{-\alpha(r-\mu)^2}$, on the r scale, which is a lognormal distribution on the abundance scale, n.

When combining numerous process values to obtain an overall net r value, approximate rotational invariance is sufficient for the pattern to be very close to a perfect normal curve (see Introduction). When measuring net squared deviations from the mean, which is the squared radial distance, the pattern is invariant to shift and stretch of the squared radial measures, $(r-\mu)^2$.

In practice, the lognormal pattern of abundance dominates when a constraint on r dominates and net values of r obey rotational invariance (symmetry) with respect to the summing up of the individual processes acting on abundance.

Any theory of process that leads to those three basic invariances will follow the gamma-lognormal pattern. The great unsolved puzzle is how specific mechanistic processes combine such that the structure of pattern is fully expressed by these particular invariances of pattern or, equivalently, by constraints on the average values of certain quantities in the context of maximum entropy. Our work opens the way for a more direct attack on this great puzzle by clarifying the anatomy of a pattern, thereby clarifying the puzzle that must be solved.

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The anatomy of pattern

> *[J]ust as the physiologist divides the animal world, according to anatomy, into families and classes, so the ornamentist is able to classify all pattern-work according to its structure [invariance]. Like the scientist, he is able even to show the affinity between groups to all appearance dissimilar; and, indeed, to point out how few are the varieties of skeleton upon which all this variety of effect is framed (ref. 33, pp. 3–4). . . . The fact of the matter is, the characteristic lines of time-honoured patterns are mainly the direct result of the restrictions under which the craftsman was working (ref. 33, p. 47).*

Invariances comprise the structural components in the anatomy of pattern. Commonly observed patterns almost always dissect completely into a few simple invariances. Our primary goal has been to introduce into ecological study the anatomy of pattern and the methods of dissection.

Identifying and naming the parts does not tell one how those parts came to be. In fact, common patterns are widespread exactly because so many different underlying mechanistic processes give rise to the same simple invariances.

Roughly speaking, one can think of a common pattern as an attractor. Each different underlying mechanistic process that develops into the generic form traces a distinctive path from some starting point to the generic endpoint of the attractor. All of the different mechanistic processes and starting points that end up at the same attractor form the basin of attraction for that pattern.

Our work characterized the anatomy of pattern—the anatomy of the attractors. The next step requires understanding how various combinations of mechanistic processes lead to one attractor or another. Equivalently, one can think of a mechanistic process as something that transforms inputs into outputs[34]. Three questions follow.

How do particular cascades of input-output transformations ultimately combine to produce overall transformations that associate with simple invariances? What separates some cascades from others with regard to association with different invariances? In other words, how can we assign different mechanistic cascades to one basin of attraction or another?

If we could answer those questions, then we could predict whether different mechanistic processes lead to the same pattern or to different patterns.

The fact that different processes can attract to the same pattern has been widely discussed in ecology[27,35–43]. However, that past work typically did not explain common patterns in terms of invariance. Without invariance, one does not have a basis for describing the anatomy of common patterns or the reasons why certain processes attract to a particular pattern and others do not.

Invariance may provide a way to compare different models of process that lead to the same pattern. Among the many complex component processes that may occur in a model, which truly matter? In other words, which component processes shape the defining invariances and which are irrelevant? For the focal component processes of each model that matter, which empirical tests would tell us which of the alternative mechanistic models is the more likely match to natural processes?

Conclusions

The apparent simplicity of invariance can mislead about its ultimate power. For example, probability patterns express a shift and stretch invariant scaling. That affine-invariant scaling provides a constant measure of information at all magnitudes.

Shift and stretch invariance seem almost trivially simple. Yet, by analyzing how repeated transformations of shift and stretch retain invariance, we obtain the most general form that expresses various affine-invariant scales (eqn 2). That affine symmetry group defines the simple, general structure of probability patterns and their uniform measurement scales.

Knowing the general invariant form of probability patterns reveals the relations between different approaches. Invariance provides powerful methods to analyze pattern and process.

To sum up, our invariance approach is not just another one among various alternatives. Rather, it is the only way to relate process to pattern, because

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the essence of pattern is invariance. Only by understanding what pattern actually is and how it generally arises can one begin to formulate testable hypotheses about mechanism.

Put another way, pattern is always the interaction between, on the one hand, the generic aspects of invariance and scale that arise in all cases and, on the other hand, the particular aspects of biology that operate in each case. Without a clear view of that duality between the generic and the particular, it is easy to mistakenly attribute generic aspects of observed pattern to particular causes. To properly understand the role of specific mechanistic aspects in shaping pattern, one must evaluate pattern simultaneously from the perspectives of the generic and the particular.

Acknowledgments

The Donald Bren Foundation (SAF) and the Swiss National Science Foundation (grant 31003A_169671 to JB) support our research. SAF completed this work while on sabbatical in the Theoretical Biology group of the Institute for Integrative Biology at ETH Zürich.

References

1. Lederman, L. M. & Hill, C. T. *Symmetry and the Beautiful Universe* (Prometheus Books, Amherst, N.Y., 2004). URL http://www.loc.gov/catdir/toc/ecip0417/2004008558.html.
2. Fisher, R. A., Corbet, A. S. & Williams, C. B. The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology* **12**, 42–58 (1943).
3. Preston, F. W. The commonness, and rarity, of species. *Ecology* **29**, 2554–283 (1948).
4. Baldridge, E., Harris, D. J., Xiao, X. & White, E. P. An extensive comparison of species-abundance distribution models. *PeerJ* **4**, e2823 (2016). URL https://doi.org/10.7717/peerj.2823.
5. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton University Press, Princeton, NJ, 2001).
6. Rosindell, J., Hubbell, S. P. & Etienne, R. S. The unified neutral theory of biodiversity and biogeography at age ten. *Trends in Ecology and Evolution* **26**, 340–348 (2011).
7. Brown, J. H. *Macroecology* (University of Chicago Press, Chicado, 1995).
8. Gaston, K. J. & Blackburn, T. M. *Pattern and Process in Macroecology* (Blackwell Science, Oxford, UK, 2000).
9. Harte, J. *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics* (Oxford University Press, New York, 2011).
10. Jaynes, E. T. Information theory and statistical mechanics. II. *Phys. Rev.* **108**, 171–190 (1957).
11. Jaynes, E. T. Information theory and statistical mechanics. *Phys. Rev.* **106**, 620–630 (1957).
12. Jaynes, E. T. *Probability Theory: The Logic of Science* (Cambridge University Press, New York, 2003).
13. Wagensberg, J. & Valls, J. The extended maximum entropy formalism and the statistical structure of ecosystems. *Bulletin of Mathematical Biology* **49**, 531–538 (1987).
14. Wagensberg, J., Lopez, D. & Valls, J. Statistical aspects of biological organization. *Journal of Physics and Chemistry of Solids* **49**, 695–700 (1988).
15. Sherwin, W. B. & Prat i Fornells, N. The introduction of entropy and information methods to ecology by Ramon Margalef. *Entropy* **21**, 794 (2019). URL https://www.mdpi.com/1099-4300/21/8/794.
16. Alonso, D., Ostling, A. & Etienne, R. S. The implicit assumption of symmetry and the species abundance distribution. *Ecology Letters* **11**, 93–105 (2008).
17. Frank, S. A. The invariances of power law size distributions. *F1000Research* **5**, 2074 (2016).
18. Frank, S. A. Common probability patterns arise from simple invariances. *Entropy* **18**, 192 (2016). URL http://www.mdpi.com/1099-4300/18/5/192.
19. Solé, R. V., Manrubia, S. C., Benton, M., Kauffman, S. & Bak, P. Criticality and scaling in evolutionary ecology. *Trends in Ecology & Evolution* **14**, 156–160 (1999). URL https://doi.org/10.1016/S0169-5347(98)01518-3.
20. Jordano, P., Bascompte, J. & Olesen, J. M. Invariant properties in coevolutionary networks of

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

plant-animal interactions. *Ecology Letters* **6**, 69–81 (2003).

21. Stouffer, D. B., Camacho, J., Guimerà, R., Ng, C. A. & Amaral, L. A. N. Quantitative patterns in the structure of model and empirical food webs. *Ecology* **86**, 1301–1311 (2005).
22. Frank, S. A. & Smith, E. A simple derivation and classification of common probability distributions based on information symmetry and measurement scale. *Journal of Evolutionary Biology* **24**, 469–484 (2011).
23. Frank, S. A. How to read probability distributions as statements about process. *Entropy* **16**, 6059–6098 (2014).
24. Frank, S. A. The common patterns of abundance: the log series and Zipf's law. *F1000Research* **8**, 334 (2019).
25. Harte, J. Maximum entropy and the state-variable approach to macroecology. *Ecology* **89**, 2700–2711 (2008).
26. Frank, S. A. Invariant death. *F1000Research* **5**, 2076 (2016).
27. Chisholm, R. A. & Pacala, S. W. Niche and neutral models predict asymptotically equivalent species abundance distributions in high-diversity ecological communities. *Proceeding of the National Academy of Sciences USA* **107**, 15821–15825 (2010).
28. Etienne, R. S. & Rosindell, J. The spatial limitations of the current neutral models of biodiversity. *PLoS ONE* **6**, e14717 (2011).
29. Alonso, D. & McKane, A. J. Sampling Hubbell's neutral theory of biodiversity. *Ecology Letters* **7**, 901–910 (2004).
30. Tribus, M. *Thermostatics and Thermodynamics: An Introduction to Energy, Information and States of Matter, with Engineering Applications* (Van Nostrand, New York, 1961).
31. Cover, T. M. & Thomas, J. A. *Elements of Information Theory* (Wiley, New York, 1991).
32. Frank, S. A. Measurement invariance explains the universal law of generalization for psychological perception. *Proceedings of National Academy of Sciences USA* **115**, 9803–9806 (2018).
33. Day, L. F. *The Anatomy of Pattern* (B. T. Batsford, London, 1887).
34. Frank, S. A. Input-output relations in biological systems: measurement, information and the Hill equation. *Biology Direct* **8**, 31 (2013).
35. Cohen, J. E. Alternative derivations of a species abundance relation. *American Naturalist* **152**, 165–172 (1968).
36. Chave, J., Alonso, D. & Etienne, R. S. Comparing models of species abundance. *Nature* **441**, E1–E2 (2006).
37. Adler, P. B., Hillerislambers, J. & Levine, J. M. A niche for neutrality. *Ecology Letters* **10**, 95–104 (2007).
38. Zillo, T. & Condit, R. The impact of neutrality, niche differentiation and species input on diversity and abundance distributions. *Oikos* **116**, 931–940 (2007).
39. Etienne, R. S., Alonso, D. & McKane, A. J. The zero-sum assumption in neutral biodiversity theory. *Journal of Theoretical Biology* **248**, 522–536 (2007).
40. Frank, S. A. The common patterns of nature. *Journal of Evolutionary Biology* **22**, 1563–1585 (2009).
41. McGill, B. J. Towards a unification of unified theories of biodiversity. *Ecology Letters* **13** (2010).
42. Matthews, T. J. & Whittaker, R. J. Neutral theory and the species abundance distribution: recent developments and prospects for unifying niche and neutral perspectives. *Ecology and Evolution* **4**, 2263–2277 (2014).
43. Matthews, T. J. & Whittaker, R. J. Fitting and comparing competing models of the species abundance distribution: assessment and prospect. *Frontiers of Biogeography* **6**, 2 (2014).
44. Au, C. & Tam, J. Transforming variables using the Dirac generalized function. *The American Statistician* **53**, 270–272 (1999).

Appendix

This appendix provides details of how to analyze changes of variable and changes of scale in a consistent way for discrete and continuous random variables. The material here was originally published at https://doi.org/10.5281/zenodo.2597895 under a CC-BY Creative Commons 4.0 license as a supplement to ref. 24.

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discrete and continuous probability distributions are usually analyzed differently, which prevents a general understanding of scale. These notes present a technique by which a change of variable or a change of scale can be done in a consistent way for both discrete and continuous distributions[44]. The final section relates discrete and continuous scales, illustrated by the log series.

For example, suppose initial measurements are in terms of abundance, n, and we wish to analyze the data on the transformed logarithmic scale, $r = \log n$. How can we make the change of variable, $n \mapsto e^r$, consistently for discrete and continuous cases?

The Dirac delta function provides the basis for a consistent method. The next section introduces the basic aspects and notation for the Dirac delta function. The following section shows how to use this method to obtain a consistent approach for transforming scale by change of variable. The final sections consider transformations between discrete and continuous variables and the specification of the domains of variables.

Dirac delta function

The Dirac delta function, δ, provides the key. The function is defined such that

$$\int_{-\epsilon}^{\epsilon} \delta(z)\mathrm{d}z = 1$$

for any real value $\epsilon > 0$. In other words, for any region of integration containing 0, the integral of $\delta(z)$ is one. Then we also have

$$\int_{z-\epsilon}^{z+\epsilon} f(x)\delta(x-z)\mathrm{d}x = f(z).$$

In other words, the integral picks out the function evaluated at the point $x = z$, at which $\delta(x-z) = \delta(0)$.

With that definition, we can write a discrete probability distribution at the set of points $\Omega = \{x_i\}$ as a continuous probability density function

$$f(x) \sum_{x_i \in \Omega} \delta(x - x_i) = f(x)\delta_x, \tag{17}$$

because the cumulative distribution function, $F(x)$, of the continuous density, $f(x)\delta_x$, has the form of a discrete probability distribution

$$F(a) = \int_{-\infty}^{a+\epsilon} f(x)\delta_x \mathrm{d}x = \sum_{x_i \in \Omega} f(x_i),$$

in which $x_i < a + \epsilon$ for an infinitesimal positive value, ϵ. We need the extra ϵ so that δ_x integrates to one around a point $x_i = a$,

For continuous distributions, let the density of points in Ω increase to fill the interval $(-\infty, \infty)$ continuously. Then

$$\delta_x = \sum_{x_i \in \Omega} \delta(x - x_i) \to \int_{-\infty}^{\infty} \delta(x - x_i)\mathrm{d}x_i = 1,$$

because, whatever the value of x, there will be some point $x_i \in \Omega$ for which $x = x_i$, and any integral over the region including that point is one. With $\delta_x = 1$, the continuous probability density function is $f(x)$, and the cumulative distribution function is

$$F(a) = \int_{-\infty}^{a+\epsilon} f(x)\mathrm{d}x = \int_{-\infty}^{a} f(x)\mathrm{d}x,$$

because $\int_a^{a+\epsilon} f(x)\mathrm{d}x = 0$ for infinitesimal ϵ and finite $f(x)$.

Change of variable

We seek a consistent method for doing a change of variable in both continuous and discrete probability distributions. The approach arises from always considering the probability associated with a value as the area of a rectangle.

For a continuous distributions, we write $f(x)\mathrm{d}x$, which is the product of the probability function, f, as the height, and the infinitesimal interval, $\mathrm{d}x$, as the width. Thus, the probability in the interval $(a, a + \epsilon)$, with small width ϵ, is

$$\int_{a}^{a+\epsilon} f(x)\mathrm{d}x \approx f(a)\epsilon,$$

the product of the height, $f(a)$, and the width, ϵ. When we change variables, $x \mapsto g(x) \equiv y$, we obtain both a new height, $f(y)$, and a new width, $\mathrm{d}y$, and so we must compensate appropriately, as shown below.

For discrete distributions, we may write $f(x)\Delta x$, in which Δx is the width associated with a discrete point, x. Typically, we assume that $\Delta x = 1$ for all x, and write the discrete probability as $f(x)$. We can think of this as the area of a rectangle with implicit width of one.

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

When we change variables, $x \mapsto g(x) \equiv y$, traditionally one keeps the interval widths, $\Delta y = 1$, as one on the new scale, y, and the probabilities are simply $f(y)$ at the new points, y. However, by changing scales, the spacing between the discrete points on the y scale differs from the spacing between points on the original x scale.

This change of spacing can be interpreted as a change in the widths associated with discrete probability points, or as a change in the density of probability points in intervals along the y scale. Thus, as in the continuous case, we may wish to keep track of how both the heights change, $f(x) \mapsto f(y)$, and how the widths change with a change of scale, $\Delta x \mapsto \Delta y$. Doing so provides a consistent way of changing variables for continuous and discrete cases.

We begin with the standard method for continuous variables. We then demonstrate an analogous method for discrete variables based on the Dirac delta function.

For a continuous distribution, $f(x)$, we make the change $x \mapsto g(x) \equiv y$. With that transformation, we have

$$\mathrm{d}y/\mathrm{d}x = g'(x).$$

Define

$$m_y = \frac{1}{|g'(x)|},$$

in which the absolute value arises because we are using $\mathrm{d}x$ and $\mathrm{d}y$ as positive probability measures. Thus,

$$\mathrm{d}x = m_y \mathrm{d}y.$$

Then the standard result for the change of variable $x \mapsto y$ in a continuous distribution yields

$$f(x)\mathrm{d}x = f(y)m_y\mathrm{d}y. \tag{18}$$

For discrete distributions, we will derive the analogous change of variable expression

$$f(x)\delta_x\mathrm{d}x = f(y)m_y\delta'_y\mathrm{d}y = f(y)\delta_y\mathrm{d}y. \tag{19}$$

To obtain this result, we need to show that the change of variable $x \mapsto g(x) \equiv y$ leads to

$$\delta_x\mathrm{d}x \mapsto m_y\delta'_y\mathrm{d}y = \delta_y\mathrm{d}y,$$

which follows if

$$\delta_x \mapsto m_y^{-1}\delta_y \equiv \delta'_y.$$

To obtain this expression for δ'_y, we need the general change of variable rule for the Dirac delta function

$$\begin{aligned}\delta(x - x_i) &\mapsto |g'(x_i)|\delta[g(x) - g(x_i)] \\ &= m_y^{-1}\delta(y - y_i).\end{aligned}$$

Thus, with $\Omega' = \{g(x_i)\} = \{y_i\}$, we have

$$\delta_x = \sum_{x_i \in \Omega} \delta(x - x_i) \mapsto m_y^{-1} \sum_{y_i \in \Omega'} \delta(y - y_i) = m_y^{-1}\delta_y.$$

The gamma and log series distributions

The gamma distribution is given by the probability function

$$f(x) = kx^{\alpha-1}e^{-\lambda x},$$

in which the constant k normalizes the total probability to be one. With $\alpha = 0$ and $x > x_0 > 0$ for x_0 not too close to zero, this has the same mathematical form as the log series distribution.

Consider the change in variable $r = \log x = g(x)$, which corresponds to $x \mapsto e^r$. Then,

$$|g'(x)| = \mathrm{d}\log x/\mathrm{d}x = 1/x = e^{-r} = m_r^{-1}.$$

If we consider $f(x)$ as a continuous distribution, then we can apply the formula for change of variable in eqn 18 to obtain

$$\begin{aligned}f(x)\mathrm{d}x &= f(r)m_r\mathrm{d}r \\ &= ke^{r(\alpha-1)}e^{-\lambda e^r}e^r\mathrm{d}r \\ &= ke^{r\alpha-\lambda e^r}\mathrm{d}r \\ &= h(r)\mathrm{d}r,\end{aligned}$$

in which

$$h(r) = f(r)m_r = ke^{r\alpha-\lambda e^r}, \tag{20}$$

for $r > \log x_0$. For $\alpha = 0$, transforming the log series form

$$f(x) = kx^{-1}e^{-\lambda x}$$

by $r = \log x$ yields the equivalent distribution on the r scale as

$$h(r) = ke^{-\lambda e^r}. \tag{21}$$

Now consider $f(x)$ as a discrete distribution. Then, by eqn 19, we immediately have

$$f(x)\delta_x\mathrm{d}x = f(r)m_r\delta'_r\mathrm{d}r = h(r)\delta'_r\mathrm{d}r,$$

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in which the full form of $h(r)$ is given in eqn 20, and we also have the log series form with $\alpha = 0$ in eqn 21.

Thus, by using the measure $\mathrm{d}r$ for the widths in the continuous case and the measure $\delta'_r \mathrm{d}r$ for the widths in the discrete case, we obtain the identical probability function $h(r)$ for the continuous and discrete cases.

For the discrete case, we must keep in mind that

$$h(r)\delta'_r \mathrm{d}r = h(r)m_r^{-1}\delta_r \mathrm{d}r,$$

in which the right side is the traditional expression that picks out the probability mass, $h(r)m_r^{-1}$, as the heights at the points defined by $\delta_r \mathrm{d}r$, implicitly using constant widths of one on all scales, because at a discrete point, r^*, at which the probability is nonzero,

$$\int_{r^*-\epsilon}^{r^*+\epsilon} \delta_r \mathrm{d}r = 1.$$

In the gamma example, $m_r^{-1} = e^{-r}$, with $x = 1, 2, \ldots$ and $r = \log 1, \log 2, \ldots$, we have the traditional expression for a discrete change of variable with constant widths of one as

$$\begin{aligned} f(x)\delta_x \mathrm{d}x &\mapsto h(r)m_r^{-1}\delta_r \mathrm{d}r \\ &= ke^{r(\alpha-1)-\lambda e^r}\delta_r \mathrm{d}r. \end{aligned}$$

For the log series case, $\alpha = 0$, this becomes

$$h(r)m_r^{-1}\delta_r \mathrm{d}r = ke^{-r-\lambda e^r}\delta_r \mathrm{d}r.$$

We can go back to the classic log series expression by reversing the change, $e^r \mapsto n$, yielding the discrete distribution $f(n)\delta_n \mathrm{d}n$ for $n = 1, 2, \ldots$, with

$$f(n) = kn^{-1}e^{-\lambda n}.$$

In these examples, we have assumed that the distributions on the r and $n \equiv x$ scales are either both discrete or both continuous. In application, it will usually make sense to think of process, r, as a continuous variable, and abundance, n, as a discrete variable. Therefore, we need to consider transformations between continuous and discrete variables. We discuss that topic in the final section, after a brief summary of the discrete transformations.

Summary of alternative discrete expressions

We have two different ways of expressing transformed discrete variables, in which the initial distribution is given by $f(x)\delta_x \mathrm{d}x$, and we transform $x \mapsto y$.

In the first expression, the transformed distribution is

$$f(y)m_y\delta'_y \mathrm{d}y = h(y)\delta'_y \mathrm{d}y,$$

in which $h(y) = f(y)m_y$ is the same expression as obtained when transforming continuous variables. The measure for the y scale, $\delta'_y \mathrm{d}y$, stretches or shrinks in relation to the measure for the x scale, altering the widths associated with each height. This form has the advantage of retaining the same expressions for the probability functions in the discrete and continuous cases.

In the second expression, the transformed distribution is

$$f(y)\delta_y \mathrm{d}y = h(y)m_y^{-1}\delta_y \mathrm{d}y,$$

in which this expression highlights the difference between the standard form of the probability function obtained in the discrete case, $f(y) = h(y)m_y^{-1}$, and the standard form of the probability function obtained in the continuous case, $h(y) = f(y)m_y$. Here, we assume the widths associated with each probability point remain one on all scales.

For the transformation $x \mapsto g(x) \equiv y$, the value $m_y^{-1} = |g'(x)|$ determines the distinction between the discrete and continuous cases, associated with the change in widths between scales.

Transforming between continuous and discrete variables

In the prior cases, we transformed from one discrete variable to another discrete variable or from one continuous variable to another continuous variable. The expressions for transformation followed without any further assumptions.

In the case of the log series and the distribution of abundances, it often make sense to consider process, r, as a continuous variable, and abundance, n, as a discrete variable. We usually think of process as causing abundance. So we should begin with the continuous distribution for process, q_r, and seek the corresponding discrete distribution for abundance, q_n.

The probability mass, q_n, at a particular value of

git • preprint @ preprint-2.0-0::19462b2-2019-08-20 (2019-08-20 07:02Z) • safrank

bioRxiv preprint doi: https://doi.org/10.1101/673590; this version posted August 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

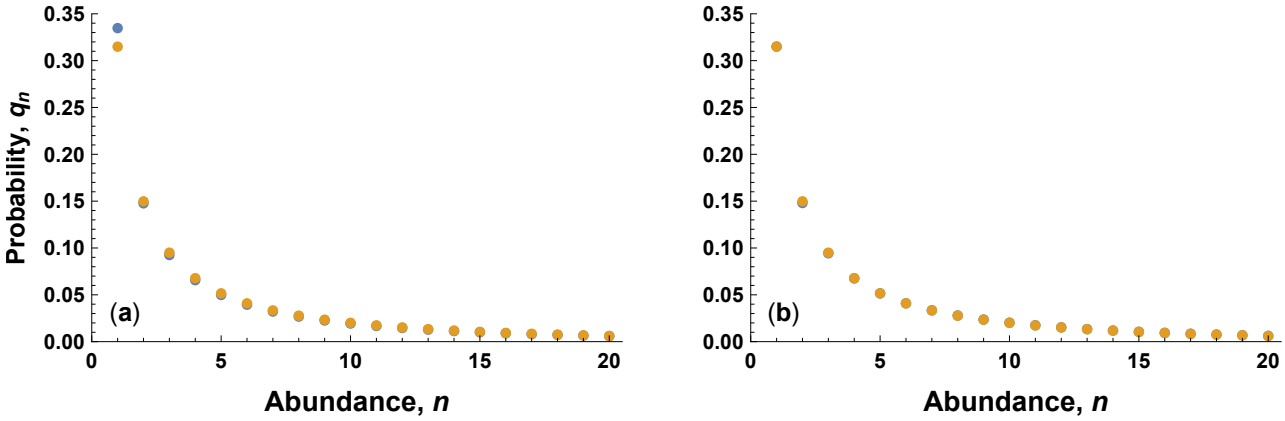

Figure A1: Close match between the continuous distribution for process, r, and the discrete log series for abundance, n. The blue circles show the probabilities of the discrete values of n obtained from the continuous distribution in r, calculated by eqn 22. The gold circles show the actual values of q_n for the log series in eqn 24. For most points, the values are nearly identical, causing the gold circles to hide the underlying blue circles. (a) When using no offset for continuous intervals, $\gamma = 0$, a slight mismatch occurs, particularly at $n = 1$. The nonlinearity of q_r causes the mismatch. (b) When using an offset of $\gamma = 0.1$, the continuous distribution of the process, r, maps almost perfectly onto the discrete log series of abundance, n. For all calculations, $\lambda = 0.05$.

$n = 1, 2, \ldots$, should map to the total probability for a matching range of growth rates, such that

$$q_n = \int_{a_n}^{b_n} q_r \mathrm{d}r, \tag{22}$$

in which $r > a_1$.

We need the particular form of q_r, which we take as the fundamental shift-invariant distribution in the main text

$$q_r = ke^{-\lambda e^r}. \tag{23}$$

We also need, for each n, the interval of growth rates, (a_n, b_n), that maps to the abundance, n. In particular, we need a sequence of contiguous intervals, $\{(a_n, b_n)\}$, that associate each value of n to an interval for r, such that $a_{n+1} = b_n$.

The problem concerns how to pick the sequence of intervals. The simplest approach is to use a standard rounding procedure, such that

$$a_n = \log(n - 0.5 + \gamma)$$
$$b_n = \log(n + 0.5 + \gamma),$$

in which $\gamma < 0.5$ is a correction for centering intervals to account for the nonlinearity in the mapping between the continuous and discrete probability expressions.

If we use $\gamma = 0.1$, the transformation from the continuous scale r to the discrete scale n in eqn 22 yields a distribution that closely matches the log series

$$q_n \approx kn^{-1}e^{-\lambda n}. \tag{24}$$

Figure A1 shows the match between the continuous distribution for r and the associated discrete distribution for n.