Species Abundance Distributions

Describing the way individuals in communities are distributed among the present species

A look at Species Abundance Patterns

Empirical distributions of relative abundance (two graphical representations)

1. Frequency histogram (Preston plot)
2. Dominance-diversity (or rank-abundance) diagram, after Whittaker (1975)

Not only at species scales, but also throughout taxonomic scales we see the same patterns

Also at larger scales

The importance of Species Abundance Distributions

"No other general attribute of ecological communities besides species richness has commanded more theoretical and empirical attention than relative species abundance."

"Commonness, and especially rarity, have long fascinated ecologists... and species abundance is of central theoretical and practical importance in conservation biology..."

"In particular, understanding the causes and consequences of rarity is a problem of profound significance because most species are uncommon to rare, and rare species are generally at greater risk to extinction."

Provides insight, not only at ecological but also on evolutionary scales!

Understanding patterns in Species Abundance

What can we do to better understand patterns in species abundance distributions?

1. Descriptive (inductive):
   Fitting curves to empirical distributions

2. Mechanistic (deductive):
   Creating mechanistic models that predict the shapes of distributions

A combined approach: create simple, mechanistic models that predict relative abundance distributions; adjust the model to maximize goodness-of-fit of predicted distributions to empirical distributions

Contents

1. Looking at empirical distributions of abundance
2. The importance of abundance models
3. Flavors of species distribution models
4. The story of never ending species (logseries)
5. Model commonness serves the greater good
6. Neutral theory and the zero sum multinomial
7. Take home message

Empirical distributions of relative abundance (two graphical representations)

Data from BCI 50-ha Forest Dynamics Plot, Panama; 229,069 individual trees of 300 species; most common species \( n_{\text{common}} = 36.081 \); several rarest species \( n_{\text{RARE}} = 1 \)

«
No other general attribute of ecological communities besides species richness has commanded more theoretical and empirical attention than relative species abundance.
Commonness, and especially rarity, have long fascinated ecologists... and species abundance is of central theoretical and practical importance in conservation biology...
In particular, understanding the causes and consequences of rarity is a problem of profound significance because most species are uncommon to rare, and rare species are generally at greater risk to extinction.

Provide insight, not only at ecological but also on evolutionary scales!

Flavors of species abundance models
We can use different models (mechanistic or descriptive) to fit species abundance distributions:

- Logarithmic series: rare species never end
- Log-Normal models: modal commonness
- Broken-stick: random division of resources
- Pre-emption model: hierarchical ordering
- Zipf-Mandelbrot: rare species larger costs
- Hubbell’s ZSM: commonness from zero-sum

The Logarithmic series
R.A. Fisher in 1942: \( S_k = \alpha x^k / n \)

- Most species are rare and even singletons (occur only once)
- More sampling may give more individuals for the rare species, but also adds on the “tail”

Log series and sample size
Larger samples makes rare species more common... But it also adds more rare species

Log series: the shape and Fishers alpha remain the same but there are infinitely rare species

Fisher’s log-series distribution
Let \( k \) = number of individuals for a species, \( k = 1, 2, 3, ... \)

Let \( S_k = \frac{\alpha x^k}{k} \) = number of species having \( k \) individuals.

with \( \alpha = \text{Fisher’s alpha diversity index} \) (to be calculated)
\( x = \text{fixed parameter} \) (to be calculated)

Then, for \( k=1, 2, 3, ... \)

\[ S_1 = \alpha \quad S_2 = \frac{\alpha x^2}{2} \quad S_3 = \frac{\alpha x^3}{3} \quad S_4 = \frac{\alpha x^4}{4} \quad \ldots \]

Where we, biologists, might think about:
- To characterize ‘diversity’ we have Simpson and Shannon index.
- Is ‘Fisher’ \( \alpha \) just another index, or does it tell something extra.
- So, what does it mean that \( \alpha \) is small or large.
How to find parameters $\alpha$ and $x$ for the Fisher log-series (given total number of species $S_T$ and total number of individuals $N_T$)

Step 1: Total number of species given by

$$S_T = S_1 + S_2 + S_3 + S_4 + \ldots$$

$$\Rightarrow S_T = \alpha x + \alpha x^2 + \alpha x^3 + \alpha x^4 + \ldots = -\alpha \log(1-x)$$

Step 2: Total number of individuals given by

$$N_T = S_1 + 2S_2 + 3S_3 + 4S_4 + \ldots$$

$$\Rightarrow N_T = \alpha x + \alpha x^2 + \alpha x^3 + \alpha x^4 + \ldots = \alpha x \left(1 + x + x^2 + x^3 + \ldots\right)$$

$$\Rightarrow N_T = \alpha x \left(\frac{1}{1-x}\right)$$

If $x$, the common ratio, is in (0,1) then sum of geometric series is given by $\frac{1}{1-x}$

So, from step1, step2 two equations in which $S_T$ and $N_T$ known and $\alpha$ and $x$ to be found

We want to know $\alpha$ because it might tell us something about diversity. We expect that $\alpha$ can be calculated from $S_T$ and $N_T$.

Why is it called a log series

For $x$ in (0,1), Taylor expansion of $-\log(1-x)$ (assume $\alpha = 1$) given by

$$-\log(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \ldots$$

"Taylor series is a representation of a function as an infinite sum of terms that are calculated from the values of the function's derivatives at a single point."

Preston’s Log-Normal

Preston did not believe Fisher’s assumption of never ending rare species

Hubbell 2001, pg 37:

As observed by Preston, if one samples long enough also log normal distributions will appear.

Preston’s Log-Normal

Preston plotted the number of species against “octaves”: doubling classes of abundance

Modal class in higher octaves and not so many rare species
The log-normal distribution and Preston's binning trick

Introduce the categories R (with number of individuals)

<table>
<thead>
<tr>
<th>Category (R)</th>
<th>nr of individuals per species</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(2^0)</td>
</tr>
<tr>
<td>1</td>
<td>(2^1)</td>
</tr>
<tr>
<td>2</td>
<td>(2^2)</td>
</tr>
<tr>
<td>3</td>
<td>(2^3)</td>
</tr>
<tr>
<td>4</td>
<td>(2^4)</td>
</tr>
</tbody>
</table>

Preston's log normal species distribution is described by

\[ S_R = S_0 e^{-\alpha R} \]

where \( \alpha \) is a 'Gaussian' - form parameter (Normal distribution).

Log-normal distribution:

Log of species abundances are normally distributed

Why might this be so?

May (1975) suggested that it arises from the statistical properties of large numbers and the Central Limit Theorem

Central Limit Theorem: When a large number of factors combine to determine the value of a variable (number of individuals per species), random variation in each of those factors (e.g., competition, predation, etc.) will result in the variable being normally distributed.


"... just as normal distributions are produced by additive combinations of random variables, lognormal distributions are produced by multiplicative combinations of random variables (May 1975)"

So Preston's idea was:

Making sample larger and larger will show more and more about lognormal relation between number of individuals and number of species.

So, sampling long enough, the rare species will be seen with a low number of species at a low number of individuals.

So, the left-hand portion of the curve may simply be missing beyond the "veil line" of a small sample.

So, the left-hand portion of the curve may simply be missing beyond the "veil line" of a small sample.
Sample size and the lognormal model

Mode moves to the right. More sampling “unveils” new rare species (lower octaves). Shape however does not change.

Difference is in the eye of the beholder

Flavors of species abundance models

- Logarithmic series: rare species never end
- Log-Normal models: modal commonness
- Broken-stick: random division of resources
- Pre-emption model: hierarchic ordering
- Zipf-Mandelbrot: rare species larger costs
- Hubbell’s ZSM: commonness from zero-sum
- ............

The Broken Stick hypothesis

Species “break” a community (e.g. stick) simultaneously into 5 pieces

Initial niche is broken at random and succesive niches are chosen with p relative to size

This results in a somewhat more even distribution of abundances among species than the other models

This suggests that it should occur when an important resource is shared more or less equitably among species

Niche Preemption (geometric)

Closely related to the broken stick hypothesis:
- Species 1 takes a certain percentage of the resources and prevents others from using them
- Species 2 then takes the same percentage of the remainder of all resources
- This continuous with the other species until all resources are used and all species are included

Assumes competitive exclusion and resource exhaustion

Minimal cooperation in ecosystem
Niche Preemption model
Assumes that the species abundance is roughly proportional to total resource use
- Linear increase in abundance leads to a linear increase in resource use
- Interspecific per-individual resource use is comparable
Mostly commonly found in species poor communities
- Early succession
- Degraded ecosystems (Enriched, Invaded)
- Harsh Ecosystems

Zipf-Mandelbrot: rare species, larger costs
The abundance of each species is dependent on:
- unspecified conditions (e.g. food consumption, mortality, reproduction rate, etc.) AND
- A number of preconditions that must be satisfied by the environment
The larger the number of conditions the less likely it is that the species in question will be present, since the probabilities are multiplicative hence, rare species have larger costs

Species Abundance Models
Dominance of any one species decreases from Logseries -> Zipf Mandelbrot
and thus:
Evenness increases from Logseries -> Zipf/Mandelbrot

Hubbell’s Neutral Theory
An attempt to predict relative-abundance distributions from neutral models of birth, death, immigration, extinction, and speciation
Assumptions: individuals play a zero-sum game within a community and have equivalent per capita demographic rates
Immigration from a source pool occurs at random; otherwise species composition in the community is governed by community drift

J = 36  D = 6  Recruits = 6

Flavors of species abundance models
• Logarithmic series: rare species never end
• Log-Normal models: modal commonness
• Broken-stick: random division of resources
• Pre-emption model: hierarchic ordering
• Zipf-Mandelbrot: rare species larger costs
• Hubbell’s ZSM: commonness from zero-sum

Hubbell’s Neutral Theory
Local communities are surrounded by a larger metacommunity: E.g. a single hectare plot is surrounded by an enormous forest which can supply new individuals (and maybe new species)
Each time an individual dies in the LC, recruitment can be either from the MC (m) or from the LC itself with (1-m)
Hubbell's Neutral Theory

An attempt to predict relative-abundance distributions from neutral models of birth, death, immigration, extinction, and speciation.

Assumptions:
- Individuals play a zero-sum game within a community and have equivalent per capita demographic rates.
- Immigration from a source pool occurs at random; otherwise species composition in the community is governed by community drift.
- The source pool (metacommunity) relative abundance is governed by its size, speciation and extinction rates.
- Local community relative abundance is additionally governed by the immigration rate.

By adjusting these (often unmeasurable) parameters, one is able to fit predicted relative abundance distributions to those observed in empirical datasets.

Hubbell's abundance model: ZSM

At large sample sizes, predicted relative species simply describe the metacommunity and becomes identical to Fisher's logseries.

$\theta$ then becomes equal to Fisher's alpha and x to the ratio of birth / death…

Model predictions can be fit to the observed relative-abundance (to the left: distribution of trees on BCI).

$\Theta(\theta) = 2Jm\nu$  
$J = \text{metacommunity size}$  
$\nu = \text{speciation rate}$  
$m = \text{immigration rate}$

Hubbell's neutral theory:

The model provides predictions that match nearly all observed distributions of relative abundance.

$\Theta(\theta) = 2Jm\nu$

$J = \text{metacommunity size}$

$\nu = \text{speciation rate}$

$m = \text{immigration rate}$

Figure from Hubbell (2001)

Does the good fit of the model to real data mean that we now understand relative abundance distributions mechanistically?

• Describing patterns in species abundance distributions can help to understand dynamics.
• There are different ways of viewing the same data, depending on the questions you have.
• Very different theories can explain similar patterns… and
• As a result: fitting a model does not mean the mechanistic metaphor is correct!
• Mathematics can aid (but sometimes also hinder) biological interpretation.
• New theories can spark discussion, even though you might not always agree with it (e.g., Neutral Theory).

Figure from Hubbell (2001)

Take home message