Measuring diversity: the importance of species similarity

Tom Leinster
Glasgow

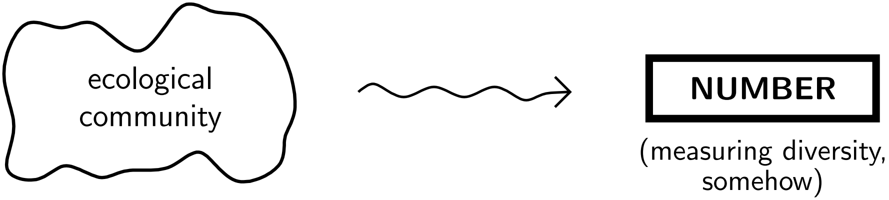
Joint with
Christina Cobbold
Glasgow

Ecology, in press

Paper and slides are available on my web page

One-slide overview of diversity measurement

Basic challenge:



Criticisms of subject:

- 'Diversity' can mean too many different things
- Too many diversity measures have been proposed
- Diversity measures produce meaningless numbers
- A single number carries little information
- Diversity measures are too dependent on the notion of species
- The varying differences between species are ignored.

Plan

1. Measuring diversity, ignoring species similarity

2. Measuring diversity, incorporating species similarity

1. Measuring diversity, ignoring species similarity

Mark Hill, 1973. Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**:427–432.

Lou Jost, 2006. Entropy and diversity. *Oikos* **113**:363–375.

A very simple model of an ecological community

Take a community whose organisms are divided into S species.

Let p_i be the relative abundance of the ith species. (So $p_1 + \cdots + p_S = 1$.)

Write $\mathbf{p} = (p_1, \ldots, p_S)$.

Assume that the community is fully censused: $\mathbf{p}$ is known exactly.

A spectrum of viewpoints

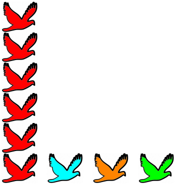
Species
are what matter

Rare species
count for as much
as common ones
—every species is precious

Communities
are what matter

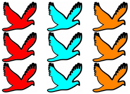
Common species
are the really
important ones
—they shape the community

This ⟶ ⟵ This



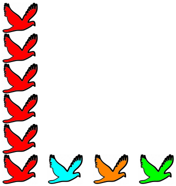
is more diverse than / is less diverse than

that ⟶ ⟵ that

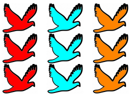


A spectrum of viewpoints

Rare species are important ⟵⟶ *Rare species are unimportant*

This ⟶  ⟵ This

is more diverse than | is less diverse than

that ⟶  ⟵ that

How to acknowledge the spectrum of viewpoints

Hill defined a family of diversity measures acknowledging the spectrum of viewpoints.

Let $0 \leq q < \infty$. The diversity of order q of the community is

$$ {}^{q}D(\mathbf{p}) = \begin{cases} \left(\sum_{i=1}^{S} p_i^q \right)^{1/(1-q)} & \text{if } q \neq 1 \\ 1/p_1^{p_1} p_2^{p_2} \cdots p_S^{p_S} & \text{if } q = 1. \end{cases} $$

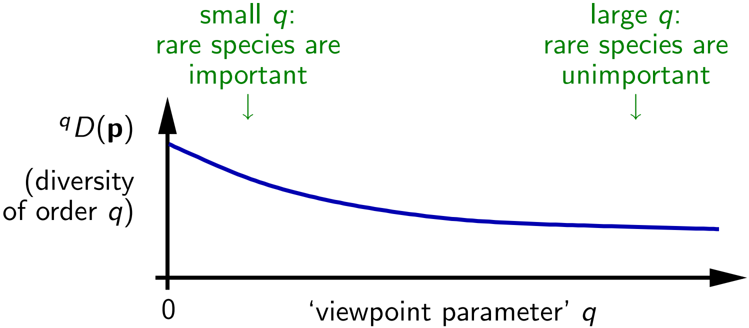
The parameter q controls the relative emphasis placed on rare and common species—in other words, where on the spectrum you are.

Diversity profiles

The belief (or superstition) of some ecologists that a diversity index provides a basis (or talisman) for reaching a full understanding of community structure is totally unfounded

—E.C. Pielou, *Ecological Diversity*, 1975.

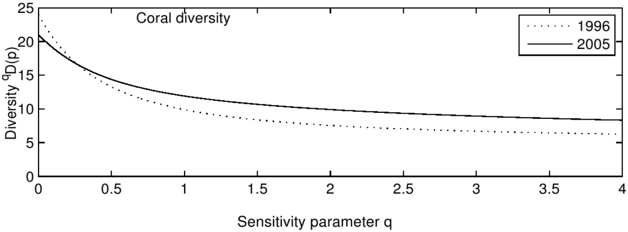
The diversity profile of a community is the graph of ${}^qD(\mathbf{p})$ against q.



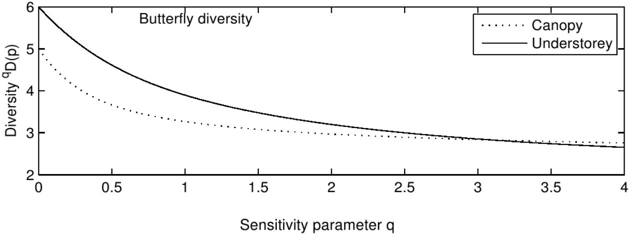
Comparing communities using diversity profiles

Plot the diversity profiles of two communities on the same axes.
They might cross. . .

Examples:



1996 community was more rich in species, but less even.



Understorey is more diverse, unless we care only about dominance.

Connections with other diversity measures

- ${}^0D(\mathbf{p}) =$ number of species
- ${}^1D(\mathbf{p}) = \exp(\text{Shannon entropy})$
- ${}^2D(\mathbf{p}) = 1/\sum_{i=1}^{S} p_i^2$: inverse Simpson concentration
- $\lim_{q\to\infty} {}^qD(\mathbf{p}) = 1/\max_i p_i$: Berger–Parker index

Effective numbers

(after Jost)

A meteorite hits an island of 10 000 equally abundant species.
90% of species are wiped out entirely, and the rest are untouched.

- The Simpson index $1 - \sum p_i^2$ drops by $< 0.1\%$
- The Shannon entropy $-\sum p_i \log p_i$ drops by 25%
- But all the diversities ${}^qD$ drop by exactly 90%.

This is because the diversities ${}^qD$ are all effective numbers:
the diversity of a community of S equally abundant species is S.
The others are not.

Principle

Every diversity measure can and should be
converted into an effective number.

2. Measuring diversity, incorporating species similarity

Tom Leinster and Christina Cobbold, 2012. Measuring diversity: the importance of species similarity. *Ecology*, in press.

What's wrong with the measures in part 1?

> 'Biological diversity' means the variability among living organisms

—UN Environment Programme definition (quoted in Magurran, *Measuring Biological Diversity*, p.6).

> . . . associated with the idea of diversity is the concept of '*distance*', i.e. some measure of the dissimilarity of the resources in question

—OECD *Handbook of Biodiversity Valuation: A Guide for Policy Makers*.

> . . . consider any *s*-species community with given proportions of its members in the several species. One would obviously regard its diversity as greater if the species belonged to several genera than if they were all congeneric, and as greater still if these genera belonged to several families than if they were confamilial.

—E.C. Pielou, *Ecological Diversity*.

A slightly less simple model of an ecological community

Continue to assume that the organisms in our community are divided into S species, with relative abundances $(p_1, \ldots, p_S) = \mathbf{p}$.

Assume we also have a measure of the similarity between the ith and jth species,

$$0 \leq Z_{ij} \leq 1.$$

Here $Z_{ij} = 0$ means total dissimilarity, and $Z_{ij} = 1$ means identical species.

This defines an $S \times S$ matrix $\mathbf{Z} = (Z_{ij})$.

The similarities Z_{ij} can be determined genetically, phylogenetically, functionally, morphologically, taxonomically, ... You choose!

Example

The naive model: different species never have anything in common.

Then $\mathbf{Z} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = \mathbf{I}$.

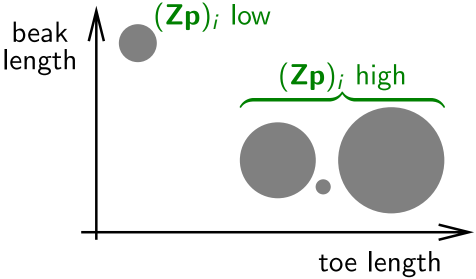
How ordinary is a species?

A species is 'ordinary' within its community if it is common, or species similar to it are common.

The ordinariness of the *i*th species is measured by

$$(\mathbf{Zp})_i = \sum_{j=1}^{S} Z_{ij} p_j.$$

Simple morphological example:



Similarity-sensitive diversity measures

The ordinariness of a species is $(\mathbf{Zp})_i = \sum_j Z_{ij} p_j$.

The average ordinariness within the community is

$$\sum_{i=1}^{S} p_i (\mathbf{Zp})_i.$$

This measures *lack* of diversity. So, one measure of diversity is

$$1 \Big/ \sum_{i=1}^{S} p_i (\mathbf{Zp})_i.$$

This is the case $q = 2$ of the following family of measures:

For $0 \leq q < \infty$, the diversity of order q of the community is

$$ {}^q D^{\mathbf{Z}}(\mathbf{p}) = \begin{cases} \left(\sum_{i=1}^{S} p_i (\mathbf{Zp})_i^{q-1} \right)^{1/(1-q)} & \text{if } q \neq 1 \\ 1/(\mathbf{Zp})_1^{p_1} (\mathbf{Zp})_2^{p_2} \cdots (\mathbf{Zp})_S^{p_S} & \text{if } q = 1. \end{cases} $$

As before, q controls the emphasis on rare/common species.

The landscape of diversity measures

This formula

$$
{}^{q}D^{\mathbf{Z}}(\mathbf{p}) = \begin{cases} \left(\sum_{i=1}^{S} p_i (\mathbf{Zp})_i^{q-1} \right)^{1/(1-q)} & \text{if } q \neq 1 \\ 1/(\mathbf{Zp})_1^{p_1} (\mathbf{Zp})_2^{p_2} \cdots (\mathbf{Zp})_S^{p_S} & \text{if } q = 1. \end{cases}
$$

unifies many existing diversity measures.

Examples:

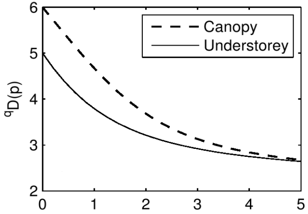
- Taking $\mathbf{Z} = \mathbf{I}$ — the 'naive model', in which different species are regarded as completely dissimilar — we get Hill's measures ${}^{q}D(\mathbf{p})$.
- Taking $q = 2$, we get Rao's quadratic entropy (transformed slightly).

A changed ecological judgement

Abundance of Charaxinae butterflies at a rainforest site in Ecuador (DeVries et al, 1997):

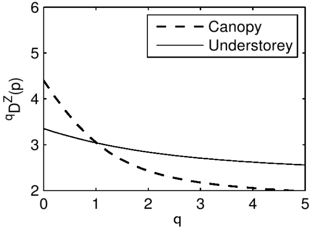
Species	Canopy	Understorey
Prepona laertes	15	0
Archaeoprepona demophon	14	37
Zaretis itys	25	11
Memphis arachne	89	23
Memphis offa	21	3
Memphis xenocles	32	8

With naive matrix ($\mathbf{Z} = \mathbf{I}$), canopy is more diverse than understorey:



But with taxonomic matrix

$$Z_{ij} = \begin{cases} 0 & \text{if of different genera} \\ 0.5 & \text{if different but congeneric} \\ 1 & \text{if } i = j, \end{cases}$$



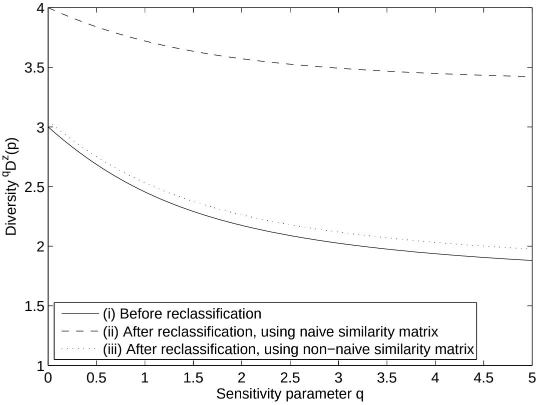
canopy looks less diverse. Reason: most of diversity of canopy is in one genus.

Reduced dependence on the concept of species

The division of organisms into species can be arbitrary or incomplete (e.g. in microbial ecology).

Similarity-sensitive diversity measures are more robust in the face of taxonomic change.

The graph shows what happens to the diversity profile of a three-species community if one of the species is reclassified into two.



What makes our diversity measures special?

It's easy to make up diversity measures...

...but harder to find measures that behave logically.

Diversity of order q (for any value of q) has the following properties:

- If two species are nearly identical, then merging them into one leaves the diversity nearly unchanged.
- It is an effective number: the diversity of a community of S equally abundant, totally dissimilar species is S.
- If m islands each have diversity d, and species on different islands are totally dissimilar, then the diversity of the whole is md.
- ...

No other general measures are known with the same good properties.

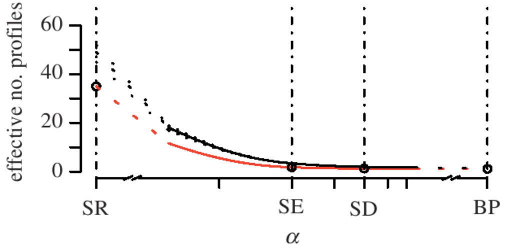
Conclusions

Answering the criticisms

- ‘Diversity’ can mean too many different things
  We can separate out the meanings by choosing different $\mathbf{Z}$s.
- Too many diversity measures have been proposed
  Many of them are unified under the umbrella of ${}^qD^{\mathbf{Z}}$.
- Diversity measures produce meaningless numbers
  Effective numbers are directly meaningful.
- A single number carries little information
  So, draw the whole diversity profile!
- Diversity measures are too dependent on the notion of species
  These ones behave proportionately when species boundaries are changed.
- The varying differences between species are ignored.
  Not here.

Unspoken statistical questions

Some work has been done on estimation of Hill's naive diversities ${}^{q}D(\mathbf{p})$.



(Figure from Mather et al., *Proc. R. Soc. B*, in press, doi:10.1098/rspb.2011.1975)

But apparently, nothing is known about the statistical properties of the diversities ${}^{q}D^{\mathbf{Z}}(\mathbf{p})$ for an arbitrary similarity matrix $\mathbf{Z}$.

Thanks

Dan Haydon

Lou Jost

Louise Matthews

Anne Chao

Richard Reeve

... and for listening.

Exploratory Conference on the Mathematics of Biodiversity

2–6 July

Centre de Recerca Matemàtica

Barcelona

www.crm.cat