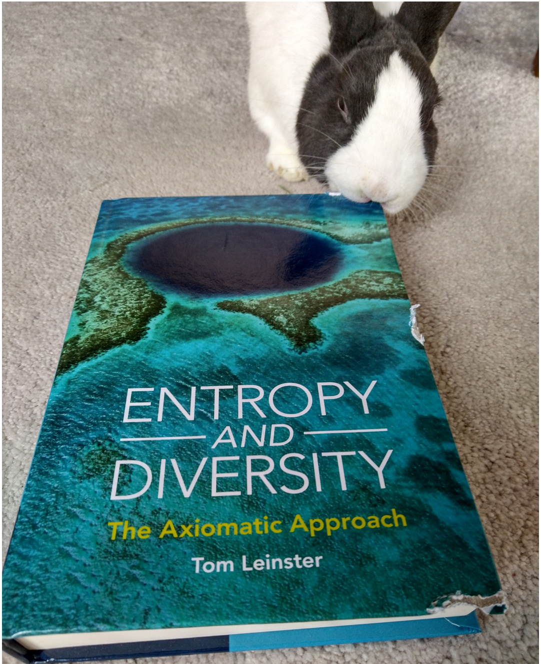
The mathematics of diversity

Tom Leinster

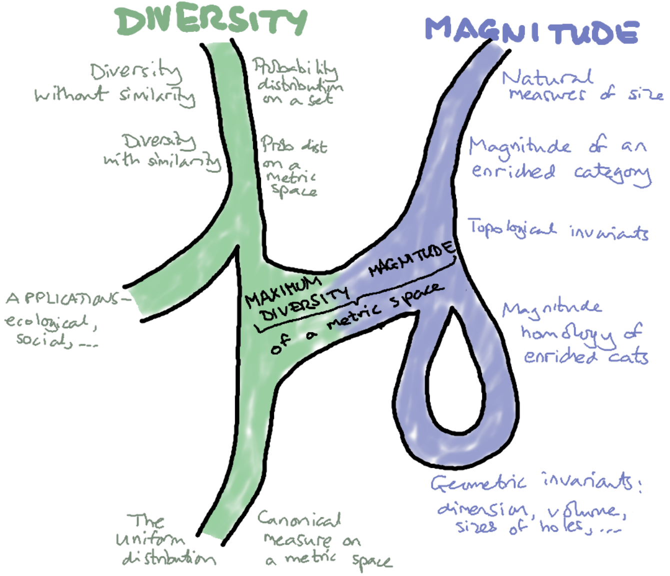
School of Mathematics
University of Edinburgh

Boyd Orr Centre
for Population and Ecosystem Health
University of Glasgow

Main reference



Plan



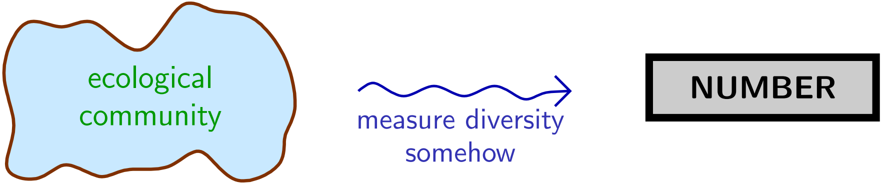
Just the left-hand half today. (Right-hand half here.)

Recurrent question

What is the $\left\{ \begin{array}{c} \text{unique} \\ \text{universal} \\ \text{canonical} \end{array} \right\}$ such-and-such?

What is the best/canonical way to measure diversity?

Basic challenge:



Take a community divided into species.

Crudest diversity measure: the number of species present.

But this is often misleading.

Example There are 8 species of great ape on the planet...

... but 99.99% of ape individuals are from a single species.

Ecologists have proposed and used many, many diversity measures. . .

Whittaker's index of association
Percentage difference (*alias* Bray–Curtis)
Wishart coefficient = (1−similarity ratio)
D = (1−Kulczynski coefficient)
Abundance-based Jaccard
Abundance-based Sørensen
Abundance-based Ochiai

Species richness $x \equiv \sum_{i=1}^{S} p_i^0$

Shannon entropy $x \equiv -\sum_{i=1}^{S} p_i \ln p_i$

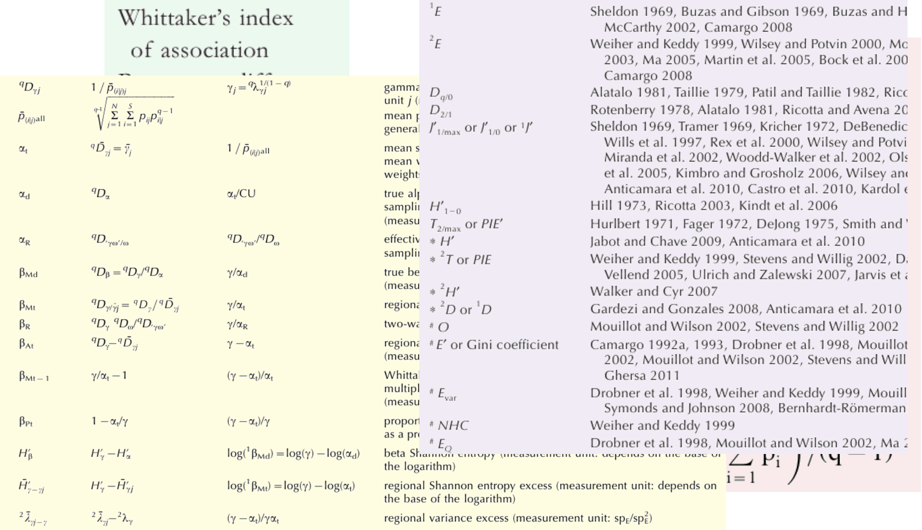
Simpson concentration $x \equiv \sum_{i=1}^{S} p_i^2$

Gini–Simpson index $x \equiv 1 - \sum_{i=1}^{S} p_i^2$

HCDT entropy $x \equiv \left(1 - \sum_{i=1}^{S} p_i^q\right)/(q-1)$

Renyi entropy $x \equiv \left(-\ln \sum_{i=1}^{S} p_i^q\right)/(q-1)$

Ecologists have proposed and used many, many diversity measures. . .



A very simple model of an ecological community

Take a community whose organisms are divided into n species.

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

Write $\boldsymbol{p} = (p_1, \ldots, p_n)$.

Mathematically: a community is a probability distribution on a finite set.

A spectrum of viewpoints

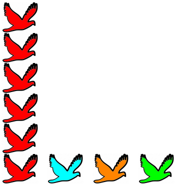
Species
are what matter

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 $\longrightarrow$

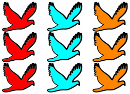


$\longleftarrow$ This

is more diverse than

is less diverse than

that $\longrightarrow$

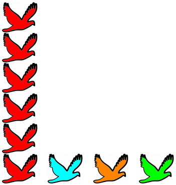


$\longleftarrow$ that

A spectrum of viewpoints

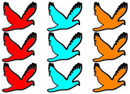
Rare species are important ——— *Rare species are unimportant*

This ⟶   ⟵ This



is more diverse than   is less diverse than

that ⟶   ⟵ that



How to acknowledge the spectrum of viewpoints

In 1973, the ecologist Mark Hill defined a family of diversity measures acknowledging the spectrum of viewpoints.

Let $0 \leq q \leq \infty$. The diversity or Hill number of order q of the community is

$$D_q(\boldsymbol{p}) = \left(\sum_i p_i^q \right)^{1/(1-q)}$$

(taking limits to get the definitions for $q = 1, \infty$).

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

Examples

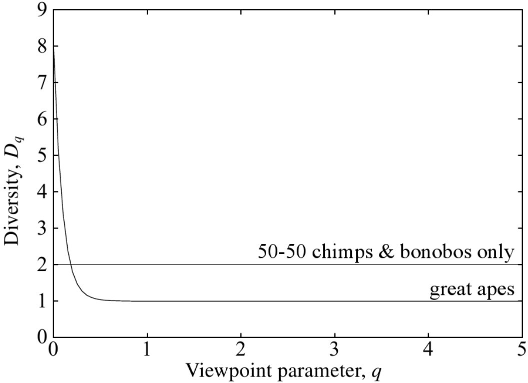
- $D_0(\boldsymbol{p})$ = number of species present.
- $D_q(1/n, \ldots, 1/n) = n$: 'there are n species in perfect balance'.

The role of q

In the definition of diversity $D_q(\boldsymbol{p})$, there is a parameter q. What does it do?

Example Take $\boldsymbol{p}$ to be the frequencies of the eight species of great ape on the planet.

Or take $\boldsymbol{p}$ to be the 50-50 distribution of chimpanzees and bonobos only.



Moral: You can't always say whether one distribution has higher diversity than another.

The answer may depend on q.

Digression: entropy

$$\text{entropy} = \log(\text{diversity})$$

The logarithm of the Hill number D_q is the Rényi entropy of order q:

$$H_q(\boldsymbol{p}) = \log D_q(\boldsymbol{p}).$$

When $q = 1$, this is the Shannon entropy.

For diversity, there are good reasons to use the exponential form.

For information theory, there are good reasons to use the logarithmic form.

Unique characterization of the Hill numbers

Any measure of diversity should behave in a logical way that reflects biological etc. intuition.

E.g. Consider a group of islands with disjoint species.

Increasing the diversity of one island should increase the diversity of the whole.

Unique characterization of the Hill numbers

Any measure of diversity should behave in a logical way that reflects biological etc. intuition.

E.g. Consider a group of islands with disjoint species.

Increasing the diversity of one island should increase the diversity of the whole.

Theorem *The only diversity measures satisfying seven sensible properties are the Hill numbers D_q ($0 \le q \le \infty$).*

A major shortcoming

Intuitively, diversity should reflect how *different* the species (etc.) are, not just their frequencies.

> '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*.

A slightly less simple model of an ecological community

Tom Leinster and Christina Cobbold,
Measuring diversity: the importance of species similarity,
Ecology 93 (2012), 477–89.

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 $n \times n$ matrix $Z = (Z_{ij})$.

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

- E.g. The naive model $Z = I$: different species have nothing in common.
- E.g. Given a metric d on $\{1, \ldots, n\}$, put $Z_{ij} = e^{-d(i,j)}$.

Roughly: a community is a probability distribution on a finite metric space.

Similarity-sensitive diversity

How typical is the ith species? Multiply the similarity matrix Z by the abundance vector $\boldsymbol{p}$:

$$(Z\boldsymbol{p})_i = \sum_j Z_{ij} p_j.$$

So the *atypicality* of the ith species can be quantified as

$$1/(Z\boldsymbol{p})_i.$$

Diversity is defined as the average atypicality of an individual.

Here 'average' *could* be the ordinary arithmetic mean

$$\sum_i p_i \cdot \frac{1}{(Z\boldsymbol{p})_i},$$

but it's useful to consider *all* power means. So for $0 \leq q \leq \infty$, define

$$D_q^Z(\boldsymbol{p}) = \left(\sum_i p_i \left(\frac{1}{(Z\boldsymbol{p})_i} \right)^{1-q} \right)^{1/(1-q)}$$

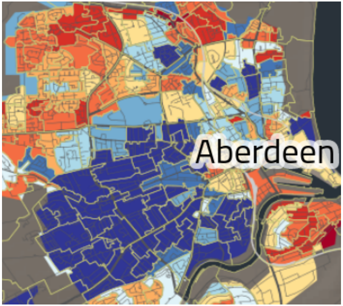
(taking limits at $q = 1, \infty$).

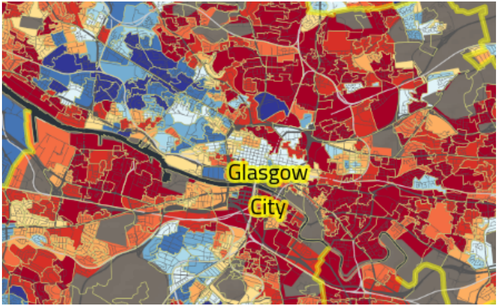
Features of the similarity-sensitive diversity measures

- This definition unifies many of the diversity measures used by ecologists (and elsewhere in the life sciences).
- When $Z = I$ (naive model), we recover the Hill numbers: $D_q^I = D_q$.
- They have been applied at all ecological scales, from microbial to large predators.
- Mathematically: taking logs, we get a notion of the entropy of a probability distribution on a metric space.

Digression: social applications

From the Scottish Index of Multiple Deprivation.
Red: high deprivation. Blue: low deprivation.





How would you quantify the concentration/spread of deprivation?
Can you measure how separated the poor and rich areas are?

A group of Edinburgh undergraduates is working on it. . .

Maximizing diversity

Tom Leinster and Mark Meckes,
Maximizing diversity in biology and beyond,
Entropy 18 (2016), article 18.

Fix a similarity matrix Z, i.e. a list of species with known similarities.

Or if you prefer: fix a finite metric space.

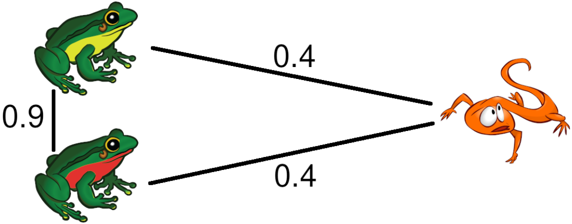
Allow the relative abundance distribution $\boldsymbol{p}$ to vary.

Which $\boldsymbol{p}$ achieves the maximum possible diversity? What *is* that maximum?

Example: frogs and newts

Take a three-species system with these similarities:

$$Z = \begin{pmatrix} 1 & 0.9 & 0.4 \\ 0.9 & 1 & 0.4 \\ 0.4 & 0.4 & 1 \end{pmatrix}$$



Which distribution maximizes diversity?

- Not $(1/3, 1/3, 1/3)$, because then we'd have $2/3$ frog and $1/3$ newt.
- Not $(1/4, 1/4, 1/2)$, because the frog species aren't quite identical.
- It should be somewhere in between.

In particular, the maximizing distribution should *not* be uniform.

The maximum diversity theorem

Fix a similarity matrix Z, i.e. a list of species with known similarities.

Or if you prefer: fix a finite metric space.

Allow the relative abundance distribution $\boldsymbol{p}$ to vary.

Which $\boldsymbol{p}$ achieves the maximum possible diversity? What *is* that maximum?

In principle, both answers depend on q.

Theorem (with Mark Meckes)

Both answers are independent of q. That is:

- there is a probability measure $\boldsymbol{p}$ maximizing $D_q^Z(\boldsymbol{p})$ for all $q \in [0, \infty]$ simultaneously
- $\sup_{\boldsymbol{p}} D_q^Z(\boldsymbol{p})$ is independent of q.

If $\boldsymbol{p}$ maximizes $D_q^Z(\boldsymbol{p})$ for all q, we call $\boldsymbol{p}$ a maximizing measure.

It is usually unique, making it a *canonical probability measure on a finite metric space*.

From finite to infinite spaces

Tom Leinster and Emily Roff,
The maximum entropy of a metric space,
Quarterly Journal of Mathematics, to appear.

All this can be done for a general *compact* metric space, not necessarily finite.

For a probability measure $\mathbb{P}$ on a compact metric space X, put

$$Z\mathbb{P}(x) = \int e^{-d(x,y)}\, d\mathbb{P}(y)$$

and

$$D_q^X(\mathbb{P}) = \left(\int \left(\frac{1}{Z\mathbb{P}(x)} \right)^{1-q} d\mathbb{P}(x) \right)^{1/(1-q)}.$$

The maximum diversity theorem holds for all compact metric spaces.

Example: real interval

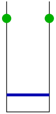
Take a real interval $[0, L]$ of length L with its usual metric.

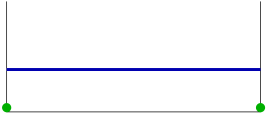
Which probability measure on $[0, L]$ maximizes diversity?

- One guess: $\frac{1}{2}(\delta_0 + \delta_L)$ (push all the mass to the ends).
- Another guess: the uniform distribution, i.e. the normalization of Lebesgue measure $\lambda_{[0,L]}$.
- In fact, it's a linear combination of these. It's the normalization of

$$\delta_0 + \lambda_{[0,L]} + \delta_L.$$

Warning The maximizing measure is scale-dependent!

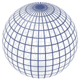
Small L: 

Large L: 

What is the uniform distribution?

Given a space X, which probability measure on X deserves to be called the 'uniform distribution' on X?

The answer is obvious for certain classes of space X. E.g.:

- finite spaces 
- suitably symmetric spaces  : the unique symmetric measure
- subsets of $\mathbb{R}^n$ with $0 < \mathrm{Vol}(X) < \infty$: normalized Lebesgue measure.

And clearly there's no sensible uniform distribution for some X, e.g. $\mathbb{R}$ or $\mathbb{Z}$.

Is there a good general answer?

What is the uniform distribution on a compact metric space?

- Old idea from statistics: the canonical choice of probability distribution is the one with the maximum entropy.

  Should we refer to the maximizing measure as the 'uniform distribution'?

  No! It's scale-dependent, and that's bad.

  It also gives the wrong answer in examples (e.g. the interval).

- But if we take the large-scale limit, it works...

The uniform measure

Let X be a compact metric space.

For $t > 0$, write tX for X with the metric scaled up by a factor of t.

Assume that for $t \gg 0$, there is only one maximizing measure $\mathbb{P}_t$ on tX.

Definition The uniform measure on X is

$$\mathbb{U}_X = \lim_{t \to \infty} \mathbb{P}_t,$$

if this limit exists (in the weak* topology).

This is scale-independent: $\mathbb{U}_{tX} = \mathbb{U}_X$ for all $t > 0$.

The uniform measure is a canonical scale-independent measure on a compact metric space.

Easy examples

- Finite spaces: The uniform measure on a finite space is the uniform measure in the usual sense.
- Symmetric spaces: Let X be a compact metric space such that for all $x, y \in X$, some self-isometry of X maps x to y.
  The Haar measure theorem implies that there is a unique isometry-invariant probability measure on X.
  And that's what the uniform measure on X is.

Slightly less easy example: the interval

Consider the real interval $[0, L]$.

We saw that the maximizing measure on $[0, L]$ is the normalization of $\delta_0 + \lambda_{[0,L]} + \delta_L$.

When we scale up by a large factor t, the point masses at the endpoints become negligible.

So the uniform measure on $[0, L]$ is the normalization of Lebesgue measure, $\lambda_{[0,L]}/L$.

That is, it's the uniform distribution in the usual sense.

Much harder case: subsets of $\mathbb{R}^n$

Let X be a compact subset of $\mathbb{R}^n$.

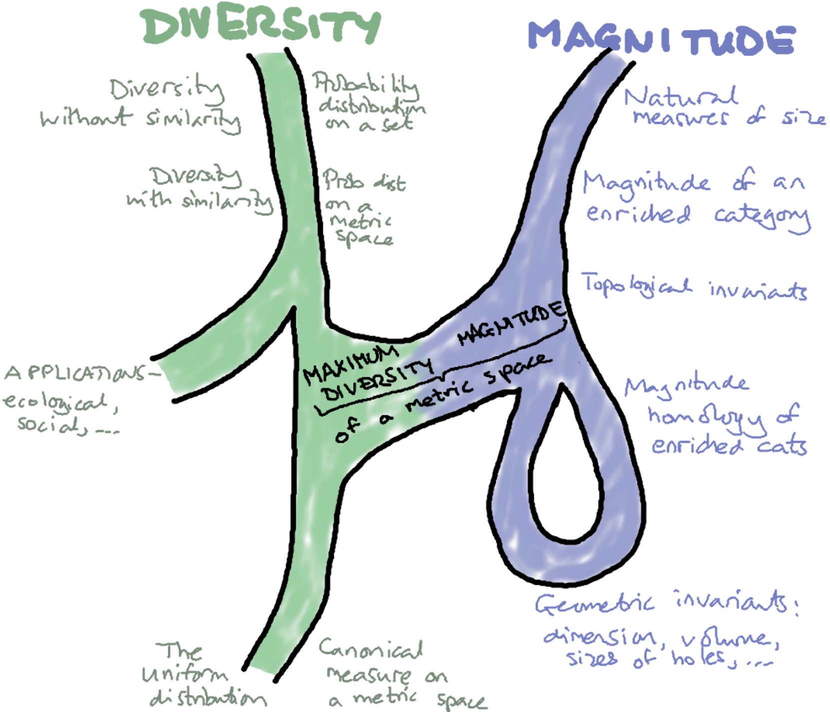
Problem Unlike for the interval, we have no description of the maximizing measure on X — even when X is a 2-dimensional disc!

So unlike for the interval, we *can't* find the uniform measure on X by finding the maximizing measure on tX for each finite t, then letting $t \to \infty$.

Nevertheless, assuming that X has nonzero measure...

Theorem *The uniform measure $\mathbb{U}_X$ is normalized Lebesgue measure on X.*

Proof Lots of analysis (paper with Emily Roff).


DIVERSITY
MAGNITUDE
Diversity without similarity
Probability distribution on a set
Natural measures of size
Diversity with similarity
Prob dist on a metric space
Magnitude of an enriched category
Topological invariants
MAGNITUDE
MAXIMUM DIVERSITY
of a metric space
APPLICATIONS – ecological, social, ...
Magnitude homology of enriched cats
Geometric invariants: dimension, volume, sizes of holes, ...
The uniform distribution
Canonical measure on a metric space

Thanks