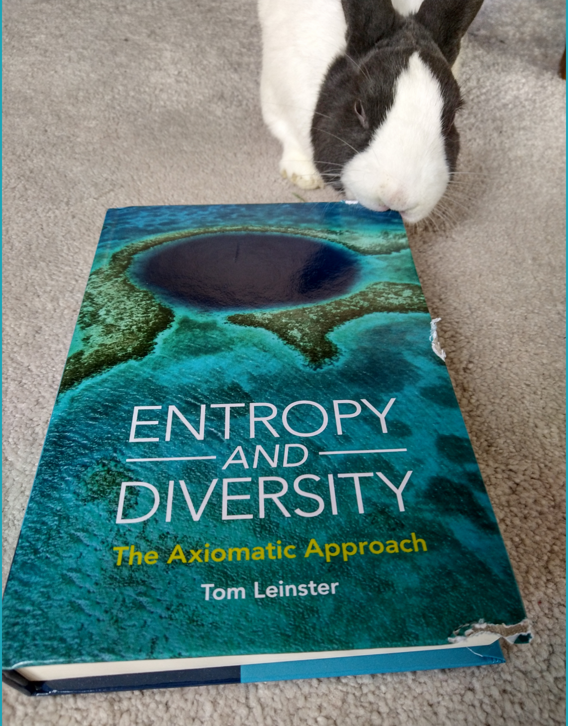

ENTROPY
AND
DIVERSITY
The Axiomatic Approach
Tom Leinster

Plan

1. Magnitude: the big picture
2. The magnitude of a metric space
3. Magnitude homology
4. (Bio)diversity
5. Maximizing diversity

1. Magnitude: the big picture

The idea

For many types of mathematical object, there is a canonical notion of size.

- Sets have cardinality. It satisfies

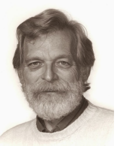
$$|S \cup T| = |S| + |T| - |S \cap T|$$
$$|S \times T| = |S| \times |T|.$$

- Subsets of $\mathbb{R}^n$ have volume. It satisfies

$$\mathrm{vol}(S \cup T) = \mathrm{vol}(S) + \mathrm{vol}(T) - \mathrm{vol}(S \cap T)$$
$$\mathrm{vol}(S \times T) = \mathrm{vol}(S) \times \mathrm{vol}(T).$$

- Topological spaces have Euler characteristic. It satisfies

$$\chi(S \cup T) = \chi(S) + \chi(T) - \chi(S \cap T) \quad \text{(under hypotheses)}$$
$$\chi(S \times T) = \chi(S) \times \chi(T).$$



Stephen Schanuel:
Euler characteristic is the topological analogue of cardinality.

The idea

For many types of mathematical object, there is a canonical notion of size.

- Sets have cardinality. It satisfies

$$|S \cup T| = |S| + |T| - |S \cap T|$$
$$|S \times T| = |S| \times |T|.$$

- Subsets of $\mathbb{R}^n$ have volume. It satisfies

$$\mathrm{vol}(S \cup T) = \mathrm{vol}(S) + \mathrm{vol}(T) - \mathrm{vol}(S \cap T)$$
$$\mathrm{vol}(S \times T) = \mathrm{vol}(S) \times \mathrm{vol}(T).$$

- Topological spaces have Euler characteristic. It satisfies

$$\chi(S \cup T) = \chi(S) + \chi(T) - \chi(S \cap T) \quad \text{(under hypotheses)}$$
$$\chi(S \times T) = \chi(S) \times \chi(T).$$

Challenge Find a general definition of 'size', including these and other examples.

One answer The magnitude of an enriched category.

The magnitude of a matrix

Let Z be a matrix.

If Z is invertible, the magnitude of Z is

$$|Z| = \sum_{i,j} (Z^{-1})_{ij}$$

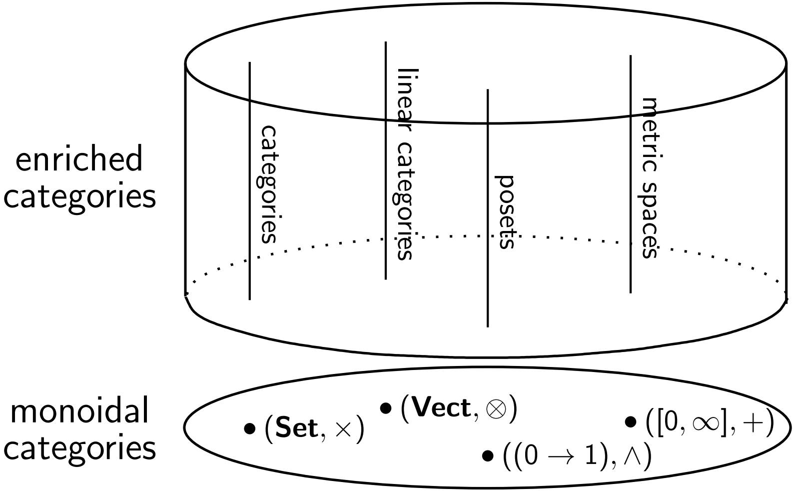
—the sum of all the entries of Z^{-1}.

(The definition can be extended to many non-invertible matrices. . .
but we won't need this refinement today.)

Enriched categories

A monoidal category is a category $\mathbf{V}$ equipped with some kind of product.

A category enriched in $\mathbf{V}$ is like an ordinary category, with a set/class of objects, but the 'hom-sets' $\mathrm{Hom}(A, B)$ are now objects of $\mathbf{V}$.



The magnitude of an enriched category

Let **V** be a monoidal category.

Suppose we have a notion of the 'size' of each object of **V**:
a multiplicative function $|\cdot|$ from $\mathrm{ob}\,\mathbf{V}$ to some field k.

E.g. $\mathbf{V} = \mathbf{FinSet}$, $k = \mathbb{Q}$, $|\cdot| =$ cardinality;
$\mathbf{V} = \mathbf{FDVect}$, $k = \mathbb{Q}$, $|\cdot| =$ dimension.

Then we get a notion of the 'size' of a category **A** enriched in **V**:

- write $Z_{\mathbf{A}}$ for the matrix $\big(|\mathrm{Hom}(A, B)|\big)_{A,B \in \mathrm{ob}\,\mathbf{A}}$ over k
- define the magnitude of the enriched category **A** to be

$$|\mathbf{A}| = |Z_{\mathbf{A}}| \in k$$

—i.e. the magnitude of the matrix $Z_{\mathbf{A}}$.

(Here assume **A** has only finitely many objects and $Z_{\mathbf{A}}$ is invertible.)

Examples not involving metric spaces

Ordinary finite categories (i.e. $\mathbf{V} = \mathbf{FinSet}$):

- For a finite category $\mathbf{A}$ satisfying mild conditions, $|\mathbf{A}|$ is $\chi(B\mathbf{A}) \in \mathbb{Z}$, the Euler characteristic of the classifying space of $\mathbf{A}$.
- For a finite group G seen as a one-object category, $|G| = 1/\text{order}(G)$.
- For a finitely triangulated manifold X, its poset $\mathbf{A}$ of simplices has magnitude $|\mathbf{A}| = \chi(X) \in \mathbb{Z}$.
- For a finitely triangulated *orbifold* X, its *category* $\mathbf{A}$ of simplices has magnitude $|\mathbf{A}| = \chi(X) \in \mathbb{Q}$. (Joint result with Ieke Moerdijk.)

Linear categories (i.e. $\mathbf{V} = \mathbf{Vect}$):

- For a suitably finite associative algebra E, let $\mathbf{IP}(E)$ denote the linear category of indecomposable projective E-modules.
  Then the magnitude of $\mathbf{IP}(E)$ is a certain Euler form associated with E.
  (Joint result with Joe Chuang and Alastair King.)

Metric spaces as enriched categories

There's at least an *analogy* between categories and metric spaces:

A category has:	A metric space has:
objects $a, b, \ldots$	points $a, b, \ldots$
sets $\mathrm{Hom}(a, b)$	numbers $d(a, b)$
composition operation	triangle inequality
$\mathrm{Hom}(a, b) \times \mathrm{Hom}(b, c) \to \mathrm{Hom}(a, c)$	$d(a, b) + d(b, c) \geq d(a, c)$

In fact, both are special cases of the concept of enriched category.

(A metric space is a category enriched in the poset $([0, \infty], \geq)$ with $\otimes = +$.)

2. The magnitude of a metric space

The magnitude of a finite metric space (concretely)

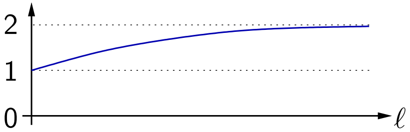
To compute the magnitude of a finite metric space $A = \{a_1, \ldots, a_n\}$:

- write down the $n \times n$ matrix with (i, j)-entry $e^{-d(a_i, a_j)}$
- invert it
- add up all n^2 entries.

And that's the magnitude $|A|$.

The magnitude of a finite metric space: first examples

- $|\emptyset| = 0$.
- $|\bullet| = 1$.
- $\left|\bullet\overset{\leftarrow\ \ell\ \rightarrow}{}\bullet\right| = \text{sum of entries of } \begin{pmatrix} e^{-0} & e^{-\ell} \\ e^{-\ell} & e^{-0} \end{pmatrix}^{-1} = \dfrac{2}{1+e^{-\ell}}$



- If $d(a,b) = \infty$ for all $a \neq b$ then $|A| = \text{cardinality}(A)$.

Slogan: Magnitude is the 'effective number of points'

Magnitude functions

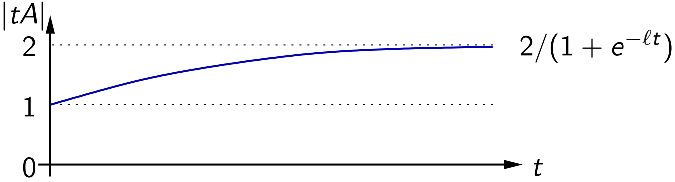
Magnitude assigns to each metric space not just a *number*, but a *function*.

For $t > 0$, write tA for A scaled up by a factor of t.

The magnitude function of a metric space A is the partial function

$$\begin{array}{ccc} (0,\infty) & \to & \mathbb{R} \\ t & \mapsto & |tA|. \end{array}$$

E.g.: the magnitude function of $A = (\bullet \overset{\leftarrow\ \ell\ \rightarrow}{} \bullet)$ is



A magnitude function has only finitely many singularities (none if $A \subseteq \mathbb{R}^n$).

It is increasing for $t \gg 0$, and $\lim_{t\to\infty} |tA| = \text{cardinality}(A)$.

The magnitude of a compact metric space

In principle, magnitude is only defined for enriched categories *with finitely many objects* — here, *finite* metric spaces.

Can the definition be extended to, say, compact metric spaces?

Theorem (Mark Meckes)

All sensible ways of extending the definition of magnitude from finite metric spaces to compact 'positive definite' spaces are equivalent.

Proof Uses some functional analysis.

Positive definite spaces include all subspaces of $\mathbb{R}^n$ with Euclidean or ℓ^1 (taxicab) metric, and many other common spaces.

The magnitude of a compact positive definite space A is

$$|A| = \sup\{|B| \ : \ \text{finite } B \subseteq A\}.$$

First examples

E.g. Line segment: $|t[0,\ell]| = 1 + \frac{1}{2}\ell \cdot t$.

E.g. Let $A \subseteq \mathbb{R}^2$ be an axis-parallel rectangle with the ℓ^1 (taxicab) metric. Then

$$|tA| = \chi(A) + \frac{1}{4}\text{perimeter}(A) \cdot t + \frac{1}{4}\text{area}(A) \cdot t^2.$$

Magnitude encodes geometric information

Theorem (Meckes) *Let A be a compact subset of $\mathbb{R}^n$, with Euclidean metric. From the magnitude function of A, you can recover its Minkowski dimension.*

Proof Uses a deep theorem from potential analysis, plus the notion of maximum diversity.

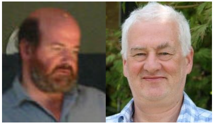
Theorem (Willerton) *Let A be a homogeneous Riemannian n-manifold. Then as $t \to \infty$,*

$$|tA| = a_n \mathrm{vol}(A) \cdot t^n + b_n \mathrm{tsc}(A) \cdot t^{n-2} + O(t^{n-4}),$$

where a_n and b_n are constants and tsc *is total scalar curvature.*

Proof Uses some asymptotic analysis.

Magnitude encodes geometric information



Theorem (Barceló and Carbery) *From the magnitude function of A, you can recover the volume of A.*

Proof Uses PDEs and Fourier analysis.

Theorem (Barceló and Carbery) *For odd n, the magnitude function of the Euclidean ball B^n is a rational function over $\mathbb{Q}$.*

Examples

$$\left|tB^1\right| = 1 + t$$

$$\left|tB^3\right| = 1 + 2t + t^2 + \tfrac{1}{6}t^3$$

$$\left|tB^5\right| = \frac{360 + 1080t + 1080t^2 + 525t^3 + 135t^4 + 18t^5 + t^6}{120(3 + t)}$$

Magnitude encodes geometric information

Theorem (Gimperlein, Goffeng and Louca) *Let A be a sufficiently regular subset of $\mathbb{R}^n$. From the magnitude function of A, you can recover the surface area of A.*

Proof Uses heat trace asymptotics (techniques related to heat equation proof of Atiyah–Singer index theorem) and treats t as a *complex* parameter.

Theorem (Gimperlein and Goffeng) *Let A and B be nice subsets of $\mathbb{R}^n$. Then*

$$|t(A \cup B)| + |t(A \cap B)| - |tA| - |tB| \to 0$$

as $t \to \infty$.

Magnitude of metric spaces doesn't *literally* obey inclusion-exclusion, as that would make it trivial. But it *asymptotically* does.

3. Magnitude homology

The idea in brief

Find a homology theory for enriched categories that categorifies magnitude.

This was first done for graphs (seen as metric spaces via shortest paths) by Hepworth and Willerton in 2015: given a graph G,

- they defined a group $H_{n,\ell}(G)$ for all integers $n, \ell \geq 0$ (a *graded* homology theory);
- writing $\chi_\ell(G) = \sum_n (-1)^n \operatorname{rank}(H_{n,\ell}(G))$, the magnitude function of G equals

$$t \mapsto \sum_\ell \chi_\ell(G) e^{-\ell t}.$$

So: the Euler characteristic of magnitude homology is magnitude.

The definition was extended to enriched categories in work with Mike Shulman in 2017.

Definition omitted...

Properties of magnitude homology of metric spaces

For metric spaces, magnitude homology is a $[0, \infty)$-graded homology theory.

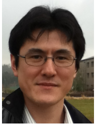
- For finite metric spaces, magnitude homology categorifies magnitude:

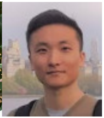
$$|tA| = \sum_{\ell \in [0,\infty)} \chi_\ell(A) e^{-\ell t}$$

(interpreted suitably), where $\chi_\ell(A) = \sum_n (-1)^n \operatorname{rank}(H_{n,\ell}(A))$ as before.

- Magnitude homology detects convexity: for closed $A \subseteq \mathbb{R}^n$,

$$A \text{ is convex} \iff H_{1,\ell}(A) = 0 \text{ for all } \ell > 0.$$

-  While ordinary homology detects the *existence* of holes, magnitude homology detects the *diameter* of holes (Kaneta and Yoshinaga).

-  There is a precise relationship between magnitude homology and persistent homology—but they detect different information (Otter; Cho).

4. (Bio)diversity

joint with Christina Cobbold

What is diversity?

Conceptual question Given an ecological community, consisting of individuals grouped into species, how can we reasonably quantify its 'diversity'?

Simplest answer Count the number n of species present.

Better answer Use the relative abundance distribution $\mathbf{p} = (p_1, \ldots, p_n)$ of species. 'Relative' means that $\sum p_i = 1$.

For any choice of parameter $q \in [0, \infty]$, can quantify diversity as

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

(taking limits for the values $q = 1, \infty$ where this is undefined).

Example If $\mathbf{p} = (1/n, \ldots, 1/n)$ then $D_q(\mathbf{p}) = n$.

Ecologists call $D_q(\mathbf{p})$ the Hill number of order q.

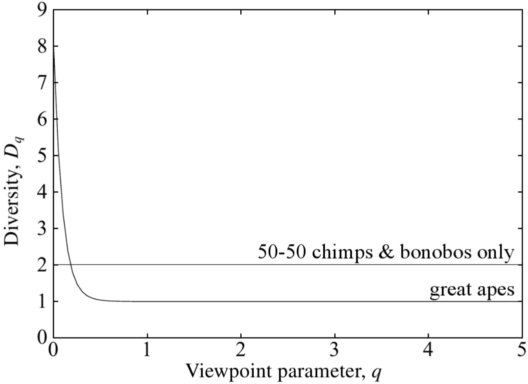
Information theorists call it the exponential of the Rényi entropy of order q. The case $q = 1$ is Shannon entropy.

The role of q

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

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

Let $\mathbf{p}'$ be the 50-50 distribution of chimpanzees and bonobos only.



Moral: You can't ask whether one probability distribution has higher diversity than another.

The answer may depend on q.

The axiomatic approach

What's so special about the Hill numbers D_q?

Why use them, rather than one of the many other diversity measures?

Theorem (*Entropy and Diversity*, Theorem 7.4.3)
Let D be a function $\{$finite probability distributions$\} \to \mathbb{R}^+$.
The following are equivalent:

- *D has seven desirable properties for a diversity measure*;
- *$D = D_q$ for some $q \in [0, \infty]$.*

Interpretation:

> The Hill numbers are the only sensible diversity measures

...at least, for this very simple model of a community.

The importance of species similarity

Intuitively, diversity should reflect the *differences* between species, not just their relative abundances.

We want measures that reflect biological reality!

Measures of diversity that ignore the varying differences between species are inadequate.

Write Z_{ij} for the similarity between species i and species j.

Interpretation:

- If $Z_{ij} = 0$, species i and j are completely dissimilar—nothing in common.
- If $Z_{ij} = 1$, species i and j are identical. (So normally $Z_{ii} = 1$.)

This gives an $n \times n$ similarity matrix $Z = (Z_{ij})_{i,j=1}^{n}$.

How do we measure similarity?

However we like! Examples:

- The naive model, where different species have nothing at all in common:

$$Z_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

  Then Z is the identity matrix $I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$.

- Taxonomically, e.g.

$$Z_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0.5 & \text{if } i \neq j \text{ but species } i \text{ and } j \text{ are in the same genus,} \\ 0 & \text{otherwise.} \end{cases}$$

- Genetically, phylogenetically, functionally, morphologically, ...
- If we have a metric d on the set $\{1, \ldots, n\}$ of species, we can define the similarities by $Z_{ij} = e^{-d(i,j)}$.

Similarity-sensitive diversity measures

Take a community of n species, with relative abundances $\mathbf{p} = (p_1, \ldots, p_n)$ and similarity matrix Z.

Its diversity of order q (for $0 \leq q \leq \infty$) is

$$D_q^Z(\mathbf{p}) = \left(\sum_{i : p_i \neq 0} p_i (Z\mathbf{p})_i^{q-1} \right)^{\frac{1}{1-q}}$$

($q \neq 1, \infty$), with the exceptional values defined by taking limits:

$$D_1^Z(\mathbf{p}) = \frac{1}{(Z\mathbf{p})_1^{p_1} (Z\mathbf{p})_2^{p_2} \cdots (Z\mathbf{p})_n^{p_n}},$$

$$D_\infty^Z(\mathbf{p}) = \frac{1}{\max\limits_{i : p_i \neq 0} (Z\mathbf{p})_i}.$$

I've skipped the story of how this definition can be motivated. But...

Unifying role

The formula

$$D_q^Z(\mathbf{p}) = \left(\sum_{i : p_i \neq 0} p_i (Z\mathbf{p})_i^{q-1} \right)^{\frac{1}{1-q}}$$

unifies many existing diversity measures.

- Take $Z = I$: the naive model, in which different species are seen as completely dissimilar. Then we recover the Hill numbers D_q.

  In particular, $Z = I$, $q = 1$ gives the exponential of Shannon entropy.

- Taking any similarity matrix Z and $q = 2$, we essentially get Rao's quadratic entropy:

$$D_2^Z(\mathbf{p}) = \frac{1}{\sum_{i,j} p_i Z_{ij} p_j} = \frac{1}{\text{mean similarity between individuals}}.$$

An analysis perspective on diversity measures

Naturally enough, we assumed the set of species was finite.

But these diversity measures can be defined for any probability measure μ on a compact Hausdorff space A equipped with a suitable 'similarity kernel' $Z \colon A \times A \to \mathbb{R}$.

E.g. When A is a compact *metric* space, take $Z(a, b) = e^{-d(a,b)}$.

For $q \in [0, \infty]$, the diversity measure $D_q(\mu)$ is defined as follows:

$$(Z\mu)(a) = \int_A Z(a, b)\, d\mu(b), \qquad D_q(\mu) = \left(\int_A \left(\frac{1}{Z\mu} \right)^{1-q} d\mu \right)^{\frac{1}{1-q}}$$

(joint work with Emily Roff).

This quantifies how spread out the probability measure μ is across the space.

5. Maximizing diversity

joint with Mark Meckes and Emily Roff

The maximum diversity theorem

Let A be a compact Hausdorff space with a similarity kernel Z (e.g. a finite set of species with known interspecies similarities).

What is the maximum possible diversity achievable by a probability measure (abundance distribution) on A?

What *is* that maximum?

In principle, both answers depend on q.

Theorem (with Mark Meckes (finite case) and Emily Roff (general case))
Both answers are independent of q. That is:

- *there is a probability measure μ maximizing $D_q(\mu)$ for all $q \in [0, \infty]$ simultaneously*
- $\sup_\mu D_q(\mu)$ *is independent of q.*

Geometric corollary: on a compact metric space A, we have:

- a canonical probability measure (the maximizer, which is usually unique)
- a canonical real number, $D_{\max}(A) = \sup_\mu D_q(\mu)$.

A little on maximum diversity

Maximum diversity is closely related to magnitude. In fact, for any compact metric space A,

$$D_{\max}(A) = |B|$$

for some closed $B \subseteq A$.

As for magnitude, the asymptotic behaviour of $t \mapsto D_{\max}(tA)$ for large t *encodes geometric information* about the space A. E.g.:

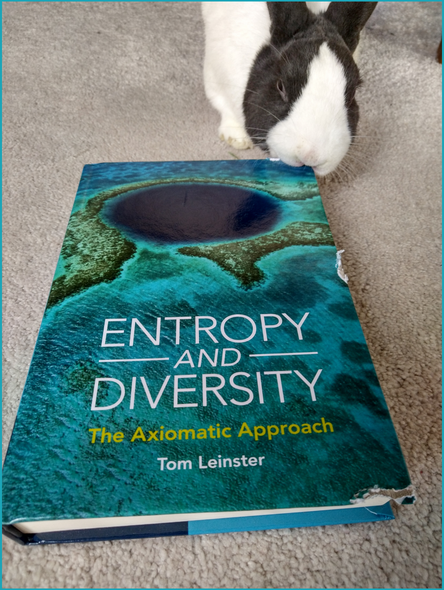
- The growth rate of $D_{\max}(tA)$ is the *Minkowski dimension* of A (Meckes).
- For $A \subseteq \mathbb{R}^n$,

  $$\mathrm{vol}(A) = c_n \lim_{t\to\infty} \frac{D_{\max}(tA)}{t^n},$$

  where c_n is a known constant.

But the maximum diversity of even some very simple spaces is unknown, e.g. Euclidean balls of dimension > 1.

References



The magnitude bibliography: www.maths.ed.ac.uk/∼tl/magbib