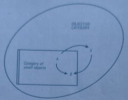
Entropy, Diversity and Magnitude

Emily Roff
University of Edinburgh

Count Me In
Research Experiences for Undergraduates
July 2024, Edinburgh

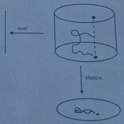
Slides at `https://www.maths.ed.ac.uk/~emilyroff/CountMeIn.pdf`

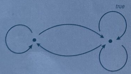


Conceptual Mathematics

A first introduction to categories

F. William Lawvere
Stephen H. Schanuel





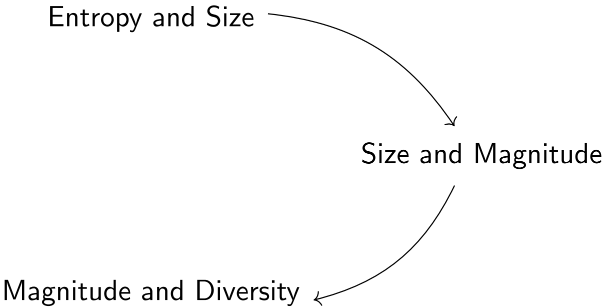
In this talk

Entropy and Size

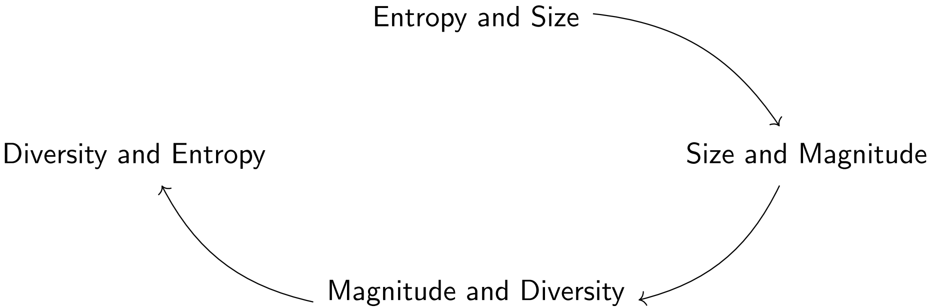
In this talk

Entropy and Size → Size and Magnitude

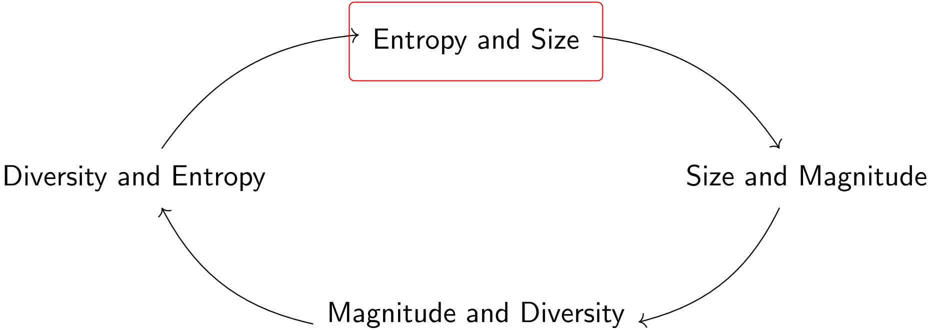
In this talk


Entropy and Size
Size and Magnitude
Magnitude and Diversity

In this talk


Entropy and Size
Size and Magnitude
Magnitude and Diversity
Diversity and Entropy

In this talk

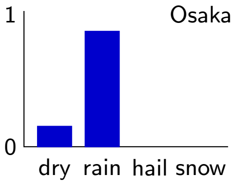

Entropy and Size
Size and Magnitude
Magnitude and Diversity
Diversity and Entropy

Entropy

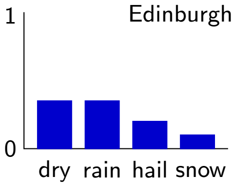
Fix a finite set X, and let $\mathcal{P}(X)$ denote the set of probability distributions on X.

Interpretation

Example The weather in June



$\mathbf{p}_O = (1/5, 4/5, 0, 0)$



$\mathbf{p}_E = (1/3, 1/3, 2/9, 1/9)$

Entropy

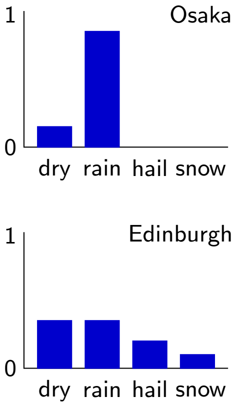
Fix a finite set X, and let $\mathcal{P}(X)$ denote the set of probability distributions on X.

The **Shannon entropy** of $\mathbf{p} \in \mathcal{P}(X)$ is

$$\mathrm{H}(\mathbf{p}) = \sum_{\substack{x \in X \\ \mathbf{p}(x) \neq 0}} \mathbf{p}(x) \log\left(\frac{1}{\mathbf{p}(x)}\right).$$

Interpretation

Example The weather in June



$\mathbf{p}_O = (1/5, 4/5, 0, 0)$

$\mathbf{p}_E = (1/3, 1/3, 2/9, 1/9)$

Entropy

Fix a finite set X, and let $\mathcal{P}(X)$ denote the set of probability distributions on X.

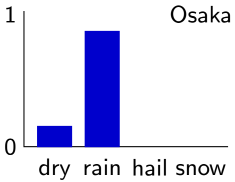
The **Shannon entropy** of $\mathbf{p} \in \mathcal{P}(X)$ is

$$\mathrm{H}(\mathbf{p}) = \sum_{\substack{x \in X \\ \mathbf{p}(x) \neq 0}} \mathbf{p}(x) \log\left(\frac{1}{\mathbf{p}(x)}\right).$$

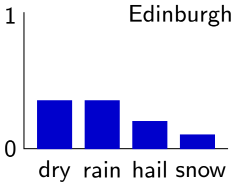
Interpretation

- $\log(1/\mathbf{p}(x))$ is the 'surprisal' of event x occurring: it's 0 if $\mathbf{p}(x) = 1$, and the less likely x is, the larger the surprisal.

Example The weather in June



$\mathbf{p}_O = (1/5, 4/5, 0, 0)$



$\mathbf{p}_E = (1/3, 1/3, 2/9, 1/9)$

Entropy

Fix a finite set X, and let $\mathcal{P}(X)$ denote the set of probability distributions on X.

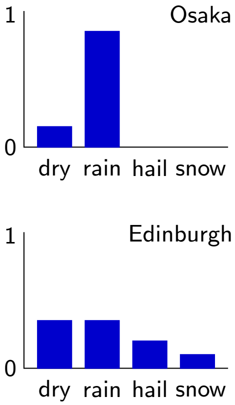
The **Shannon entropy** of $\mathbf{p} \in \mathcal{P}(X)$ is

$$H(\mathbf{p}) = \sum_{\substack{x \in X \\ \mathbf{p}(x) \neq 0}} \mathbf{p}(x) \log \left(\frac{1}{\mathbf{p}(x)} \right).$$

Interpretation

- $\log(1/\mathbf{p}(x))$ is the 'surprisal' of event x occurring: it's 0 if $\mathbf{p}(x) = 1$, and the less likely x is, the larger the surprisal.
- $H(\mathbf{p})$ is the average surprisal or expected suprisal of events in X according to $\mathbf{p}$.

Example The weather in June



$\mathbf{p}_O = (1/5, 4/5, 0, 0)$

$\mathbf{p}_E = (1/3, 1/3, 2/9, 1/9)$

Entropy

Fix a finite set X, and let $\mathcal{P}(X)$ denote the set of probability distributions on X.

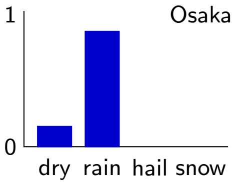
The **Shannon entropy** of $\mathbf{p} \in \mathcal{P}(X)$ is

$$\mathrm{H}(\mathbf{p}) = \sum_{\substack{x \in X \\ \mathbf{p}(x) \neq 0}} \mathbf{p}(x) \log\left(\frac{1}{\mathbf{p}(x)}\right).$$

Interpretation

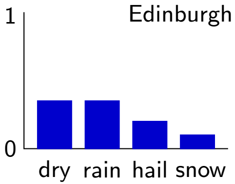
- $\log(1/\mathbf{p}(x))$ is the 'surprisal' of event x occurring: it's 0 if $\mathbf{p}(x) = 1$, and the less likely x is, the larger the surprisal.
- $\mathrm{H}(\mathbf{p})$ is the average surprisal or expected suprisal of events in X according to $\mathbf{p}$.

Example The weather in June



$\mathrm{H}(\mathbf{p}_O) \sim 0.217$

$\mathbf{p}_O = (1/5, 4/5, 0, 0)$



$\mathbf{p}_E = (1/3, 1/3, 2/9, 1/9)$

Entropy

Fix a finite set X, and let $\mathcal{P}(X)$ denote the set of probability distributions on X.

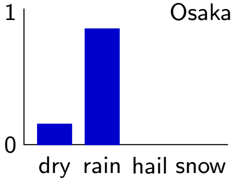
The **Shannon entropy** of $\mathbf{p} \in \mathcal{P}(X)$ is

$$\mathrm{H}(\mathbf{p}) = \sum_{\substack{x \in X \\ \mathbf{p}(x) \neq 0}} \mathbf{p}(x) \log\left(\frac{1}{\mathbf{p}(x)}\right).$$

Interpretation

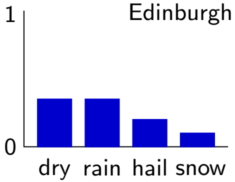
- $\log(1/\mathbf{p}(x))$ is the 'surprisal' of event x occurring: it's 0 if $\mathbf{p}(x) = 1$, and the less likely x is, the larger the surprisal.
- $\mathrm{H}(\mathbf{p})$ is the average surprisal or expected suprisal of events in X according to $\mathbf{p}$.

Example The weather in June



$\mathrm{H}(\mathbf{p}_O) \sim 0.217$

$\mathbf{p}_O = (1/5, 4/5, 0, 0)$



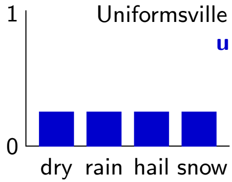
$\mathrm{H}(\mathbf{p}_E) \sim 0.569$

$\mathbf{p}_E = (1/3, 1/3, 2/9, 1/9)$

Entropy and Size

Shannon entropy is maximized by the uniform distribution, **u**. Its entropy is

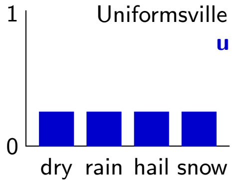
$$H(\mathbf{u}) = \sum_{x \in X} \mathbf{u}(x) \log \left(\frac{1}{\mathbf{u}(x)} \right)$$



Entropy and Size

Shannon entropy is maximized by the uniform distribution, $\mathbf{u}$. Its entropy is

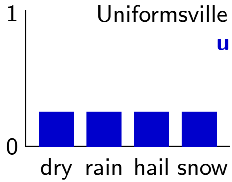
$$\begin{aligned} \mathrm{H}(\mathbf{u}) &= \sum_{x \in X} \mathbf{u}(x) \log\left(\frac{1}{\mathbf{u}(x)}\right) \\ &= \sum_{x \in X} \frac{1}{\#X} \log\left(\#X\right) \end{aligned}$$



Entropy and Size

Shannon entropy is maximized by the uniform distribution, $\mathbf{u}$. Its entropy is

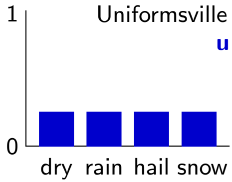
$$\begin{aligned}
\mathrm{H}(\mathbf{u}) &= \sum_{x \in X} \mathbf{u}(x) \log \left(\frac{1}{\mathbf{u}(x)} \right) \\
&= \sum_{x \in X} \frac{1}{\#X} \log (\#X) \\
&= \log (\#X) .
\end{aligned}$$



Entropy and Size

Shannon entropy is maximized by the uniform distribution, $\mathbf{u}$. Its entropy is

$$\begin{aligned} \mathrm{H}(\mathbf{u}) &= \sum_{x\in X} \mathbf{u}(x) \log\left(\frac{1}{\mathbf{u}(x)}\right) \\ &= \sum_{x\in X} \frac{1}{\#X} \log(\#X) \\ &= \log(\#X). \end{aligned}$$

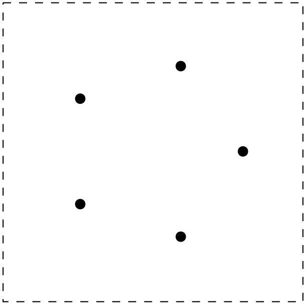


Alternatively This is telling us that the cardinality of X is determined by

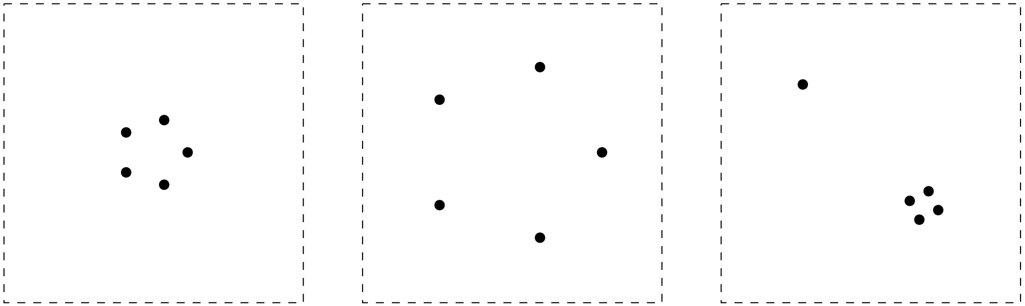
$$\#X = \max_{\mathbf{p}\in\mathcal{P}(X)} \left(\exp \mathrm{H}(\mathbf{p})\right).$$

This is a very simple variational principle for the size of a finite set.

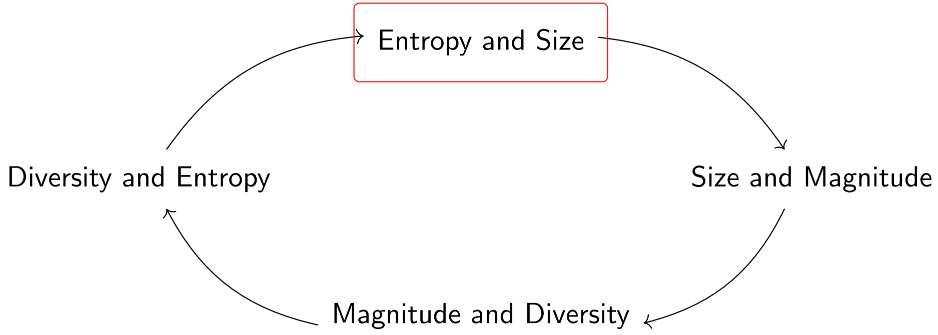
Size

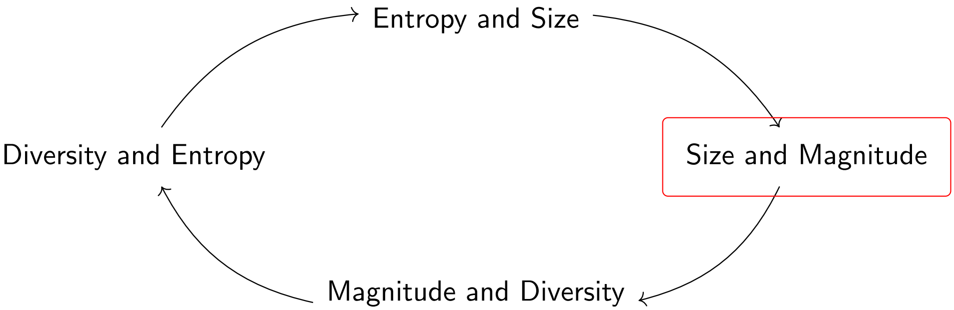


Size



Are these the same size?

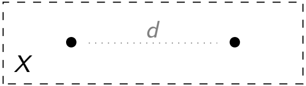

Entropy and Size
Size and Magnitude
Magnitude and Diversity
Diversity and Entropy


Entropy and Size
Size and Magnitude
Magnitude and Diversity
Diversity and Entropy

The Magnitude of a Finite Metric Space

Let X be a finite metric space with metric d.

Simplest Example



The Magnitude of a Finite Metric Space

Let X be a finite metric space with metric d.

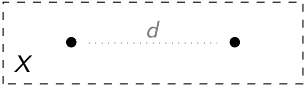
Recipe for Magnitude (Leinster, 2010)

1. Write down the $X \times X$ matrix Z with

$$Z(x, y) = e^{-d(x,y)}.$$

This is the **similarity matrix** of X.

Simplest Example



The Magnitude of a Finite Metric Space

Let X be a finite metric space with metric d.

Recipe for Magnitude (Leinster, 2010)

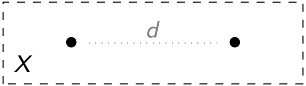
1. Write down the $X \times X$ matrix Z with

$$Z(x, y) = e^{-d(x,y)}.$$

   This is the **similarity matrix** of X.

2. Invert Z, if you can.*

Simplest Example



* For *almost every* metric space, Z is invertible.

The Magnitude of a Finite Metric Space

Let X be a finite metric space with metric d.

Recipe for Magnitude (Leinster, 2010)

1. Write down the $X \times X$ matrix Z with

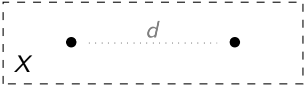
$$Z(x, y) = e^{-d(x,y)}.$$

   This is the **similarity matrix** of X.

2. Invert Z, if you can.*
3. Add up all the entries of Z^{-1}.

* For *almost every* metric space, Z is invertible.

Simplest Example



The Magnitude of a Finite Metric Space

Let X be a finite metric space with metric d.

Recipe for Magnitude (Leinster, 2010)

1. Write down the $X \times X$ matrix Z with

$$Z(x, y) = e^{-d(x,y)}.$$

   This is the **similarity matrix** of X.

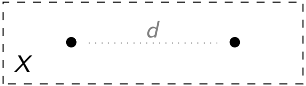
2. Invert Z, if you can.*
3. Add up all the entries of Z^{-1}.

The **magnitude** of X is the real number

$$|X| = \sum_{x,y \in X} Z^{-1}(x, y).$$

* For *almost every* metric space, Z is invertible.

Simplest Example



The Magnitude of a Finite Metric Space

Let X be a finite metric space with metric d.

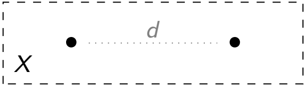
Recipe for Magnitude (Leinster, 2010)

1. Write down the $X \times X$ matrix Z with
$$Z(x, y) = e^{-d(x,y)}.$$
This is the **similarity matrix** of X.
2. Invert Z, if you can.*
3. Add up all the entries of Z^{-1}.

The **magnitude** of X is the real number
$$|X| = \sum_{x,y \in X} Z^{-1}(x, y).$$

* For *almost every* metric space, Z is invertible.

Simplest Example



The similarity matrix of this space is
$$Z = \begin{pmatrix} 1 & e^{-d} \\ e^{-d} & 1 \end{pmatrix}, \text{ so } |X| = \frac{2}{1 + e^{-d}}.$$

The Magnitude of a Finite Metric Space

Let X be a finite metric space with metric d.

Recipe for Magnitude (Leinster, 2010)

1. Write down the $X \times X$ matrix Z with

$$Z(x,y) = e^{-d(x,y)}.$$

   This is the **similarity matrix** of X.

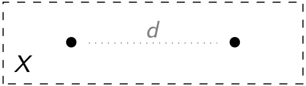
2. Invert Z, if you can.*
3. Add up all the entries of Z^{-1}.

The **magnitude** of X is the real number

$$|X| = \sum_{x,y \in X} Z^{-1}(x,y).$$

* For *almost every* metric space, Z is invertible.

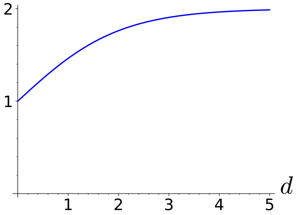
Simplest Example



The similarity matrix of this space is

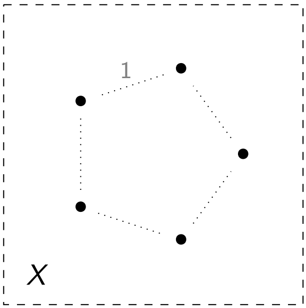
$$Z = \begin{pmatrix} 1 & e^{-d} \\ e^{-d} & 1 \end{pmatrix}, \text{ so } |X| = \frac{2}{1+e^{-d}}.$$

We can plot this as a function of d...



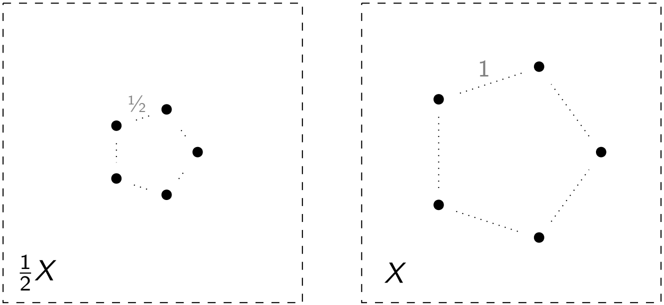
The Viewpoint Parameter

Given a metric space (X, d), for each $t \in (0, \infty)$ we denote by tX the space (X, td).



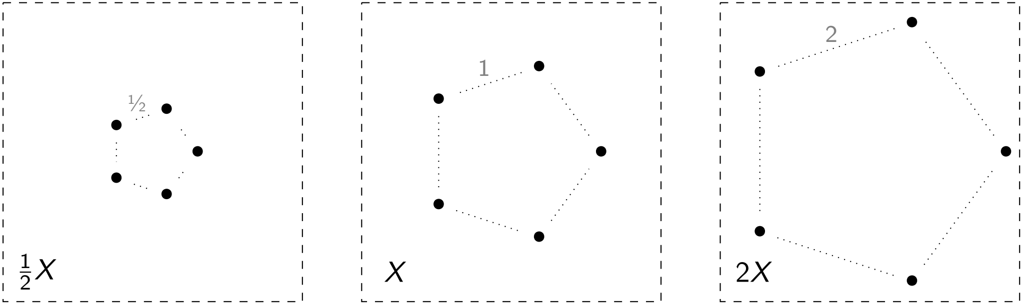
The Viewpoint Parameter

Given a metric space (X, d), for each $t \in (0, \infty)$ we denote by tX the space (X, td).



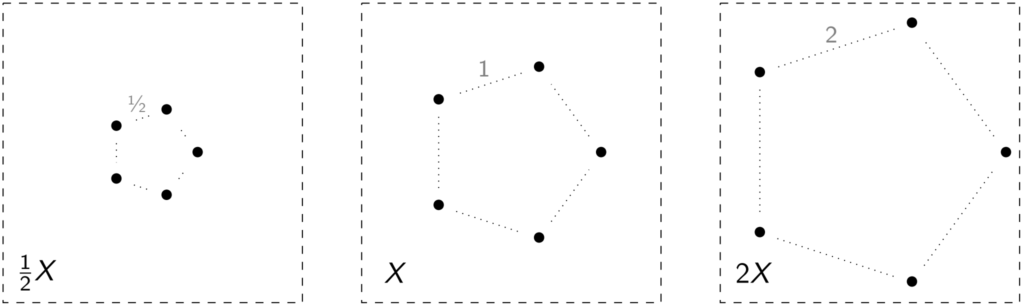
The Viewpoint Parameter

Given a metric space (X, d), for each $t \in (0, \infty)$ we denote by tX the space (X, td).



The Viewpoint Parameter

Given a metric space (X, d), for each $t \in (0, \infty)$ we denote by tX the space (X, td).

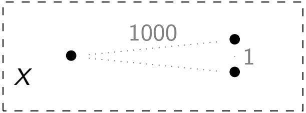


Think of t as controlling our **viewpoint**: it lets us zoom in and out on X.

The Effective Number of Points in a Space

The **magnitude function** of a finite metric space X is the function $(0, \infty) \to \mathbb{R}$ defined by $t \mapsto |tX|$.

Example



The Effective Number of Points in a Space

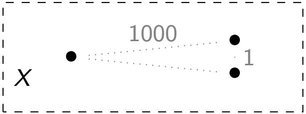
The **magnitude function** of a finite metric space X is the function $(0, \infty) \to \mathbb{R}$ defined by $t \mapsto |tX|$.

Basic Theorems

- (Leinster, 2010) For every finite metric space,

$$\lim_{t \to \infty} |tX| = \#X.$$

Example



The Effective Number of Points in a Space

The **magnitude function** of a finite metric space X is the function $(0, \infty) \to \mathbb{R}$ defined by $t \mapsto |tX|$.

Basic Theorems

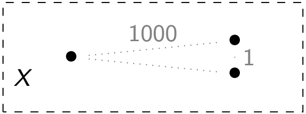
- (Leinster, 2010) For every finite metric space,

$$\lim_{t \to \infty} |tX| = \#X.$$

- (R.+Yoshinaga, 2023) For *almost* every finite space,

$$\lim_{t \to 0} |tX| = 1.$$

Example



The Effective Number of Points in a Space

The **magnitude function** of a finite metric space X is the function $(0, \infty) \to \mathbb{R}$ defined by $t \mapsto |tX|$.

Basic Theorems

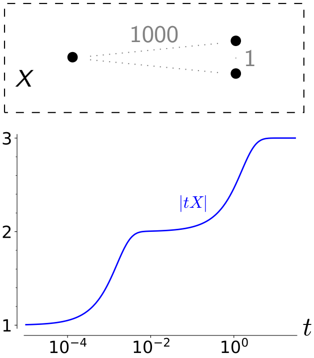
- (Leinster, 2010) For every finite metric space,

$$\lim_{t\to\infty} |tX| = \#X.$$

- (R.+Yoshinaga, 2023) For *almost* every finite space,

$$\lim_{t\to 0} |tX| = 1.$$

Example



The Effective Number of Points in a Space

The **magnitude function** of a finite metric space X is the function $(0, \infty) \to \mathbb{R}$ defined by $t \mapsto |tX|$.

Basic Theorems

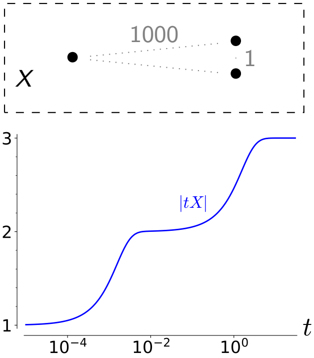
- (Leinster, 2010) For every finite metric space,

$$\lim_{t\to\infty} |tX| = \#X.$$

- (R.+Yoshinaga, 2023) For *almost* every finite space,

$$\lim_{t\to 0} |tX| = 1.$$

Example



Slogan Magnitude records the effective number of points in X as the scale varies.

Idea 1

For finite sets,

$\#$ of points = max(exp(entropy))

Idea 1

For finite sets,
$\#$ of points $=$ max(exp(entropy))

Idea 2

For finite metric spaces,
magnitude $=$ effective $\#$ of points

Idea 1

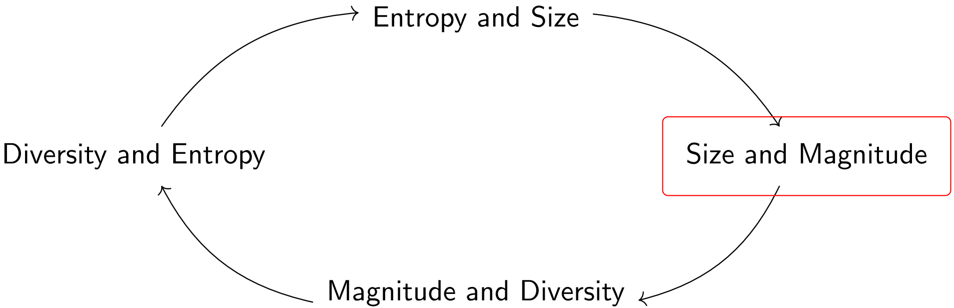
For finite sets,
\# of points = max(exp(entropy))

Idea 2

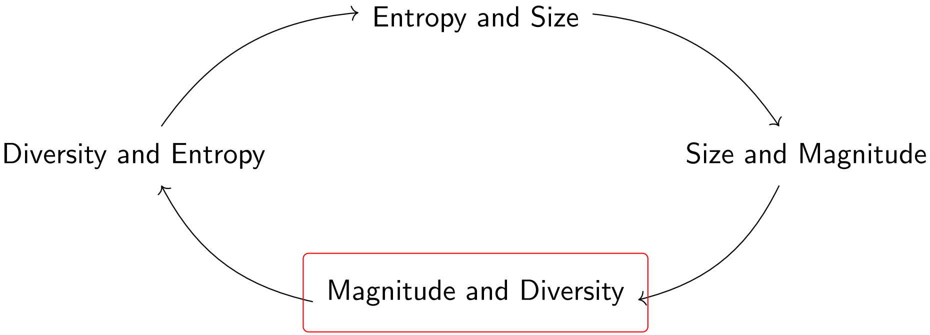
For finite metric spaces,
magnitude = effective # of points

Question

Is there a notion of
effective entropy
for probability distributions
on a finite metric space, such that
magnitude = max(exp(effective entropy))?



Entropy and Size
Diversity and Entropy
Size and Magnitude
Magnitude and Diversity

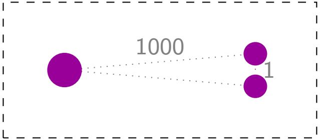

Entropy and Size
Diversity and Entropy
Size and Magnitude
Magnitude and Diversity

The Weighting on a Metric Space

Definition The **weighting** on a finite metric space X is the vector $\mathbf{w} \in \mathbb{R}^X$ defined by

$$\mathbf{w}(x) = \sum_{y \in X} Z^{-1}(x, y).$$

Example



On this space the weighting is

$$\mathbf{w} = \begin{pmatrix} 1 & 0.7 & 0.7 \end{pmatrix}.$$

The Weighting on a Metric Space

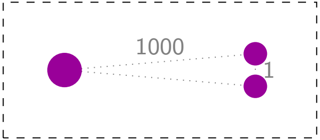
Definition The **weighting** on a finite metric space X is the vector $\mathbf{w} \in \mathbb{R}^X$ defined by

$$\mathbf{w}(x) = \sum_{y \in X} Z^{-1}(x, y).$$

Key Properties

1. The weighting is the vector of row sums in Z^{-1}, so $|X| = \sum_{x \in X} \mathbf{w}(x)$

Example



On this space the weighting is

$$\mathbf{w} = \begin{pmatrix} 1 & 0.7 & 0.7 \end{pmatrix}.$$

The Weighting on a Metric Space

Definition The **weighting** on a finite metric space X is the vector $\mathbf{w} \in \mathbb{R}^X$ defined by

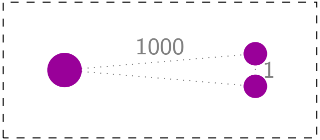
$$\mathbf{w}(x) = \sum_{y \in X} Z^{-1}(x, y).$$

Key Properties

1. The weighting is the vector of row sums in Z^{-1}, so $|X| = \sum_{x \in X} \mathbf{w}(x)$

and $Z\mathbf{w} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

Example



On this space the weighting is

$$\mathbf{w} = \begin{pmatrix} 1 & 0.7 & 0.7 \end{pmatrix}.$$

The Weighting on a Metric Space

Definition The **weighting** on a finite metric space X is the vector $\mathbf{w} \in \mathbb{R}^X$ defined by

$$\mathbf{w}(x) = \sum_{y \in X} Z^{-1}(x, y).$$

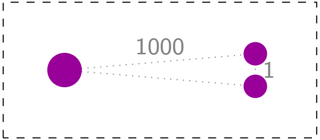
Key Properties

1. The weighting is the vector of row sums in Z^{-1}, so

$|X| = \sum_{x \in X} \mathbf{w}(x)$

and $Z\mathbf{w} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ so $|X| = \mathbf{w}^T Z \mathbf{w}$.

Example



On this space the weighting is

$$\mathbf{w} = \begin{pmatrix} 1 & 0.7 & 0.7 \end{pmatrix}.$$

The Weighting on a Metric Space

Definition The **weighting** on a finite metric space X is the vector $\mathbf{w} \in \mathbb{R}^X$ defined by

$$\mathbf{w}(x) = \sum_{y \in X} Z^{-1}(x, y).$$

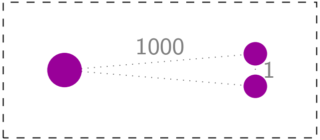
Key Properties

1. The weighting is the vector of row sums in Z^{-1}, so

$|X| = \sum_{x \in X} \mathbf{w}(x)$

and $Z\mathbf{w} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ so $|X| = \mathbf{w}^T Z \mathbf{w}$.

Example



On this space the weighting is

$$\mathbf{w} = \begin{pmatrix} 1 & 0.7 & 0.7 \end{pmatrix}.$$

2. If $\mathbf{w}(x) \geqslant 0$ for all $x \in X$ then $\frac{1}{|X|} \cdot \mathbf{w}$ is a probability distribution on X.

Magnitude as a Maximum

Let X be such that Z is positive definite, meaning $\mathbf{v}^T Z \mathbf{v} > 0$ for all $\mathbf{v} \in \mathbb{R}^X \backslash \{\mathbf{0}\}$.

In this case Z is *certain* to be invertible.

Magnitude as a Maximum

Let X be such that Z is positive definite, meaning $\mathbf{v}^T Z \mathbf{v} > 0$ for all $\mathbf{v} \in \mathbb{R}^X \backslash \{\mathbf{0}\}$.

Proposition Then the magnitude of X is given by

$$|X| = \sup_{\substack{\mathbf{v} \in \mathbb{R}^X \\ \mathbf{v} \neq \mathbf{0}}} \frac{\left(\sum_{x \in X} \mathbf{v}(x)\right)^2}{\mathbf{v}^T Z \mathbf{v}}$$

and a vector $\mathbf{v}$ attains the supremum if and only if $\mathbf{v} = \alpha \cdot \mathbf{w}$ for the weighting $\mathbf{w}$ on X and some $\alpha \in \mathbb{R} \backslash \{0\}$.

In this case Z is *certain* to be invertible.

Magnitude as a Maximum

Let X be such that Z is positive definite, meaning $\mathbf{v}^T Z \mathbf{v} > 0$ for all $\mathbf{v} \in \mathbb{R}^X \setminus \{\mathbf{0}\}$.

Proposition Then the magnitude of X is given by

$$|X| = \sup_{\substack{\mathbf{v} \in \mathbb{R}^X \\ \mathbf{v} \neq \mathbf{0}}} \frac{\left(\sum_{x \in X} \mathbf{v}(x)\right)^2}{\mathbf{v}^T Z \mathbf{v}}$$

and a vector $\mathbf{v}$ attains the supremum if and only if $\mathbf{v} = \alpha \cdot \mathbf{w}$ for the weighting $\mathbf{w}$ on X and some $\alpha \in \mathbb{R} \setminus \{0\}$.

In this case Z is *certain* to be invertible.

Proof idea. Since Z is positive definite it defines an inner product on $\mathbb{R}^X$, so we can apply the Cauchy–Schwartz inequality. □

Magnitude as a Maximum

Let X be such that Z is positive definite, meaning $\mathbf{v}^T Z \mathbf{v} > 0$ for all $\mathbf{v} \in \mathbb{R}^X \setminus \{\mathbf{0}\}$.

Proposition Then the magnitude of X is given by

$$|X| = \sup_{\substack{\mathbf{v} \in \mathbb{R}^X \\ \mathbf{v} \neq \mathbf{0}}} \frac{\left(\sum_{x \in X} \mathbf{v}(x)\right)^2}{\mathbf{v}^T Z \mathbf{v}}$$

and a vector $\mathbf{v}$ attains the supremum if and only if $\mathbf{v} = \alpha \cdot \mathbf{w}$ for the weighting $\mathbf{w}$ on X and some $\alpha \in \mathbb{R} \setminus \{0\}$.

Corollary If the weighting on X is non-negative, then

$$|X| = \max_{\mathbf{p} \in \mathcal{P}(X)} \frac{1}{\mathbf{p}^T Z \mathbf{p}} .$$

In this case Z is *certain* to be invertible.

Proof idea. Since Z is positive definite it defines an inner product on $\mathbb{R}^X$, so we can apply the Cauchy–Schwartz inequality. □

Magnitude as a Maximum

Let X be such that Z is positive definite, meaning $\mathbf{v}^T Z \mathbf{v} > 0$ for all $\mathbf{v} \in \mathbb{R}^X \setminus \{\mathbf{0}\}$.

Proposition Then the magnitude of X is given by

$$|X| = \sup_{\substack{\mathbf{v} \in \mathbb{R}^X \\ \mathbf{v} \neq \mathbf{0}}} \frac{\left(\sum_{x \in X} \mathbf{v}(x)\right)^2}{\mathbf{v}^T Z \mathbf{v}}$$

and a vector $\mathbf{v}$ attains the supremum if and only if $\mathbf{v} = \alpha \cdot \mathbf{w}$ for the weighting $\mathbf{w}$ on X and some $\alpha \in \mathbb{R} \setminus \{0\}$.

Corollary If the weighting on X is non-negative, then

$$|X| = \max_{\mathbf{p} \in \mathcal{P}(X)} \frac{1}{\mathbf{p}^T Z \mathbf{p}} .$$

In this case Z is *certain* to be invertible.

Proof idea. Since Z is positive definite it defines an inner product on $\mathbb{R}^X$, so we can apply the Cauchy–Schwartz inequality. □

Proof. The supremum in the Proposition is attained by the probability distribution $\frac{1}{|X|} \cdot \mathbf{w}$. □

Magnitude as a Maximum

Let X be such that Z is positive definite, meaning $\mathbf{v}^T Z \mathbf{v} > 0$ for all $\mathbf{v} \in \mathbb{R}^X \backslash \{\mathbf{0}\}$.

Proposition Then the magnitude of X is given by

$$|X| = \sup_{\substack{\mathbf{v} \in \mathbb{R}^X \\ \mathbf{v} \neq \mathbf{0}}} \frac{\left(\sum_{x \in X} \mathbf{v}(x)\right)^2}{\mathbf{v}^T Z \mathbf{v}}$$

and a vector $\mathbf{v}$ attains the supremum if and only if $\mathbf{v} = \alpha \cdot \mathbf{w}$ for the weighting $\mathbf{w}$ on X and some $\alpha \in \mathbb{R} \backslash \{0\}$.

Corollary If the weighting on X is non-negative, then

$$|X| = \max_{\mathbf{p} \in \mathcal{P}(X)} \boxed{\frac{1}{\mathbf{p}^T Z \mathbf{p}}}.$$

Question What does this value tell us about $\mathbf{p}$?

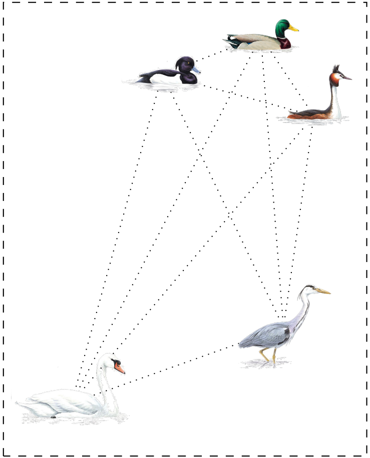
In this case Z is *certain* to be invertible.

Proof idea. Since Z is positive definite it defines an inner product on $\mathbb{R}^X$, so we can apply the Cauchy–Schwartz inequality. □

Proof. The supremum in the Proposition is attained by the probability distribution $\frac{1}{|X|} \cdot \mathbf{w}$. □

Change of Scene!

Let's suppose X is a set of biological species, and the metric d on X records differences among species.

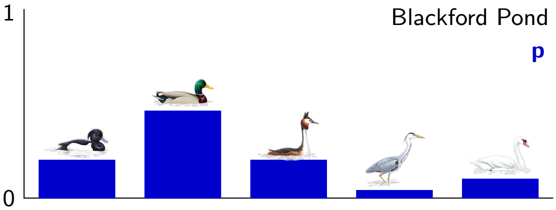


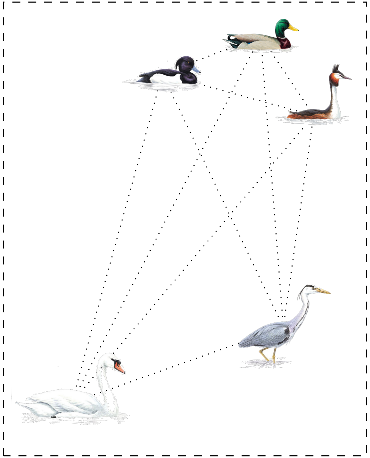
Bird illustrations from the RSPB.

Change of Scene!

Let's suppose X is a set of biological species, and the metric d on X records differences among species.

An ecological community comprising members of species in X can be modelled by a probability distribution $\mathbf{p}$ on X, where $\mathbf{p}(x)$ is the relative abundance of species x in the community.



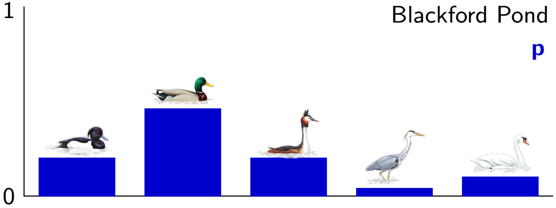


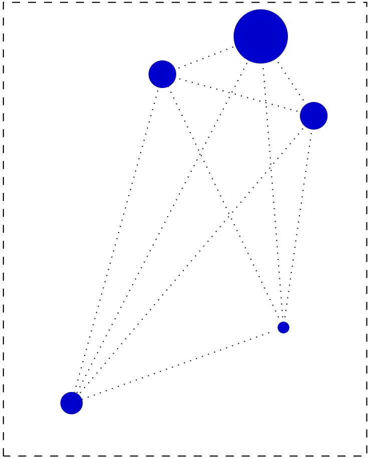
Bird illustrations from the RSPB.

Change of Scene!

Let's suppose X is a set of biological species, and the metric d on X records differences among species.

An ecological community comprising members of species in X can be modelled by a probability distribution $\mathbf{p}$ on X, where $\mathbf{p}(x)$ is the relative abundance of species x in the community.

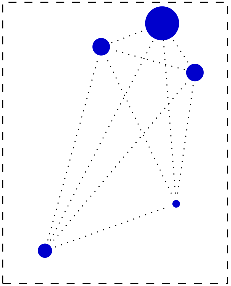




Bird illustrations from the RSPB.

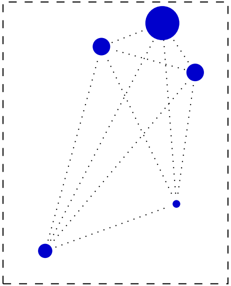
Ecological Interpretations

Quantity	Interpretation
$Z(a, b)$	The similarity between species a and b
$(Z\mathbf{p})(a)$	
$\mathbf{p}^T Z\mathbf{p}$	
$1/\mathbf{p}^T Z\mathbf{p}$	



Ecological Interpretations

Quantity	Interpretation
$Z(a,b)$	The similarity between species a and b
$(Z\mathbf{p})(a)$	The typicality of species a in community $\mathbf{p}$
$\mathbf{p}^T Z\mathbf{p}$	
$1/\mathbf{p}^T Z\mathbf{p}$	



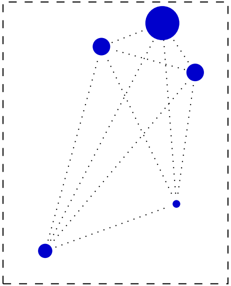
Interpretation For each species $a \in X$,

$$(Z\mathbf{p})(a) = \sum_{b\in X} \mathbf{p}(b) e^{-d(a,b)}$$

tells us the expected similarity between a member of species a and a random member of the community: the typicality of a in $\mathbf{p}$.

Ecological Interpretations

Quantity	Interpretation
$Z(a,b)$	The similarity between species a and b
$(Z\mathbf{p})(a)$	The typicality of species a in community $\mathbf{p}$
$\mathbf{p}^T Z\mathbf{p}$	The homogeneity of community $\mathbf{p}$
$1/\mathbf{p}^T Z\mathbf{p}$	



Interpretation For each species $a \in X$,

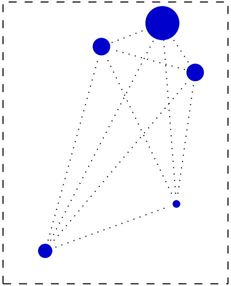
$$(Z\mathbf{p})(a) = \sum_{b \in X} \mathbf{p}(b) e^{-d(a,b)}$$

tells us the expected similarity between a member of species a and a random member of the community: the typicality of a in $\mathbf{p}$.

So $\mathbf{p}^T Z\mathbf{p} = \sum_a \mathbf{p}(a)(Z\mathbf{p})(a)$ records the average typicality of members of the community: the homogeneity of $\mathbf{p}$.

Ecological Interpretations

Quantity	Interpretation
$Z(a, b)$	The similarity between species a and b
$(Z\mathbf{p})(a)$	The typicality of species a in community $\mathbf{p}$
$\mathbf{p}^T Z\mathbf{p}$	The homogeneity of community $\mathbf{p}$
$1/\mathbf{p}^T Z\mathbf{p}$	The diversity of community $\mathbf{p}$



Interpretation For each species $a \in X$,

$$(Z\mathbf{p})(a) = \sum_{b \in X} \mathbf{p}(b) e^{-d(a,b)}$$

tells us the expected similarity between a member of species a and a random member of the community: the typicality of a in $\mathbf{p}$.

So $\mathbf{p}^T Z\mathbf{p} = \sum_a \mathbf{p}(a)(Z\mathbf{p})(a)$ records the average typicality of members of the community: the homogeneity of $\mathbf{p}$.

And the less homogeneous a community is, the more diverse we consider it to be

Progress

Objective

Define **effective entropy** such that

$$\text{magnitude} = \max_{\mathbf{p}\in\mathcal{P}(X)} \exp(\text{effective entropy}(\mathbf{p})).$$

So Far

If X has positive-definite similarity matrix and non-negative weighting, then

$$|X| = \max_{\mathbf{p}\in\mathcal{P}(X)} \frac{1}{\mathbf{p}^T Z \mathbf{p}} = \max_{\mathbf{p}\in\mathcal{P}(X)} (\text{diversity}(\mathbf{p})).$$

Progress

Objective

Define **effective entropy** such that

$$\text{magnitude} = \max_{\mathbf{p} \in \mathcal{P}(X)} \exp(\text{effective entropy}(\mathbf{p})).$$

"I really don't see what diversity has to do with entropy."

So Far

If X has positive-definite similarity matrix and non-negative weighting, then

$$|X| = \max_{\mathbf{p} \in \mathcal{P}(X)} \frac{1}{\mathbf{p}^T Z \mathbf{p}} = \max_{\mathbf{p} \in \mathcal{P}(X)} (\text{diversity}(\mathbf{p})).$$

Progress

Objective

Define **effective entropy** such that

$$\text{magnitude} = \max_{\mathbf{p} \in \mathcal{P}(X)} \exp(\text{effective entropy}(\mathbf{p})).$$

"I really don't see what diversity has to do with entropy."

"I'm an ecologist, and that's *not* what *I* mean by diversity!"

So Far

If X has positive-definite similarity matrix and non-negative weighting, then

$$|X| = \max_{\mathbf{p} \in \mathcal{P}(X)} \frac{1}{\mathbf{p}^T Z \mathbf{p}} = \max_{\mathbf{p} \in \mathcal{P}(X)} (\text{diversity}(\mathbf{p})).$$

A Spectrum of Perspectives on Diversity

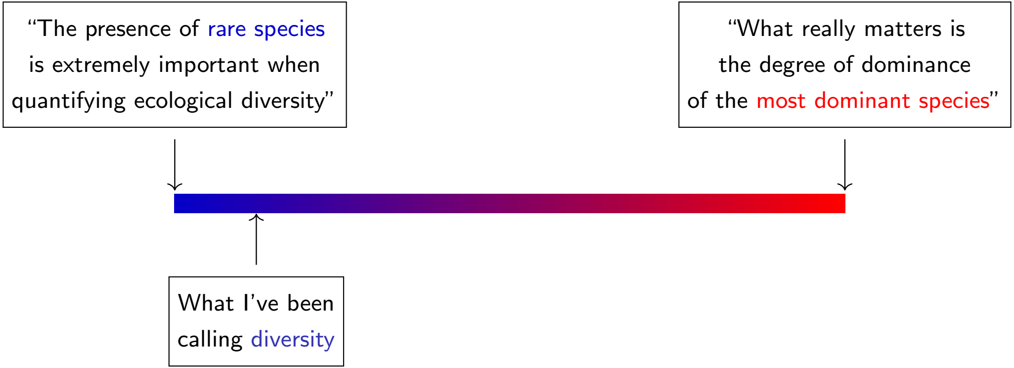
"The presence of rare species is extremely important when quantifying ecological diversity"

A Spectrum of Perspectives on Diversity

"The presence of rare species is extremely important when quantifying ecological diversity"

"What really matters is the degree of dominance of the most dominant species"

A Spectrum of Perspectives on Diversity



A Spectrum of Diversity Indices

Definition (Leinster & Cobbold, 2012) Let X be a finite metric space. For each $q \in \mathbb{R}_{\geqslant 0} \backslash \{1\}$, the **diversity of order q** of $\mathbf{p} \in \mathcal{P}(X)$ is the real number

$$D_q^Z(\mathbf{p}) = \left(\sum_{x \in X} \mathbf{p}(x)(Z\mathbf{p})(x)^{q-1} \right)^{\frac{1}{1-q}}.$$

The **diversity of order 1** is defined to make $D_q^Z(\mathbf{p})$ continuous in q:

$$D_1^Z(\mathbf{p}) = \lim_{q \to 1} D_q^Z(\mathbf{p}) = \prod_{x \in X} (Z\mathbf{p})(x)^{-\mathbf{p}(x)}$$

most sensitive to rare species

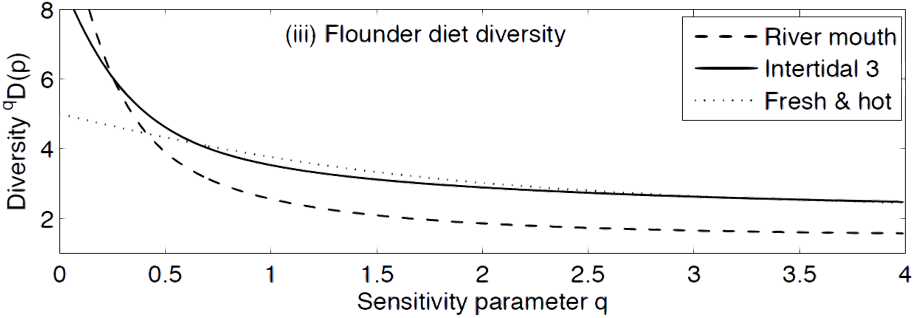
least sensitive to rare species

0 2 $q \to$ ∞

$1/\mathbf{p}^T Z \mathbf{p}$

Example

Diversity Profiles



Leinster & Cobbold, *Measuring Diversity...*, Ecology 93 (2012)

The Maximum Diversity Theorem

Theorem (Leinster & Meckes, 2015)

Let X be a finite metric space. Then:

1. There exists a probability distribution on X that maximizes $D_q^Z(-)$ for all $q \geqslant 0$.
2. The value $\max_{\mathbf{p}} D_q^Z(\mathbf{p})$ is independent of q.

The uniform distribution is usually *not* maximizing. Instead, maximizing distributions are **balanced**: they make all points in X equally typical.

Example

The Maximum Diversity Theorem

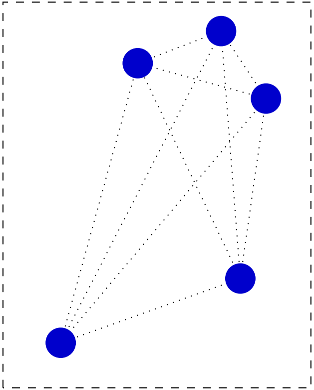
Theorem (Leinster & Meckes, 2015)

Let X be a finite metric space. Then:

1. There exists a probability distribution on X that maximizes $D_q^Z(-)$ for all $q \geqslant 0$.
2. The value $\max_{\mathbf{p}} D_q^Z(\mathbf{p})$ is independent of q.

The uniform distribution is usually *not* maximizing. Instead, maximizing distributions are **balanced**: they make all points in X equally typical.

Example



The uniform distribution.

The Maximum Diversity Theorem

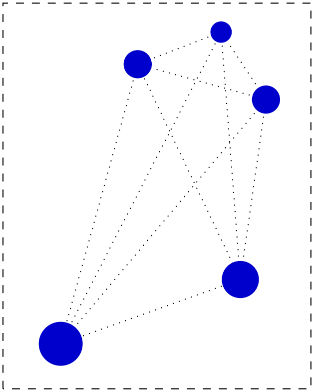
Theorem (Leinster & Meckes, 2015)

Let X be a finite metric space. Then:

1. There exists a probability distribution on X that maximizes $D_q^Z(-)$ for all $q \geqslant 0$.
2. The value $\max_{\mathbf{p}} D_q^Z(\mathbf{p})$ is independent of q.

The uniform distribution is usually *not* maximizing. Instead, maximizing distributions are **balanced**: they make all points in X equally typical.

Example



The maximizing distribution.

The Maximum Diversity Theorem

Theorem (Leinster & Meckes, 2015)

Let X be a finite metric space. Then:

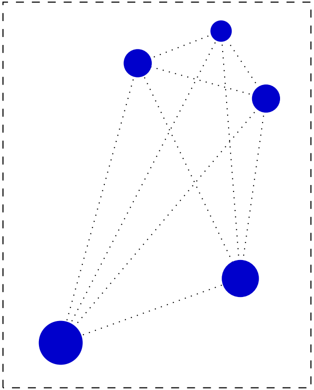
1. There exists a probability distribution on X that maximizes $D_q^Z(-)$ for all $q \geqslant 0$.
2. The value $\max_{\mathbf{p}} D_q^Z(\mathbf{p})$ is independent of q.

The uniform distribution is usually *not* maximizing. Instead, maximizing distributions are **balanced**: they make all points in X equally typical.

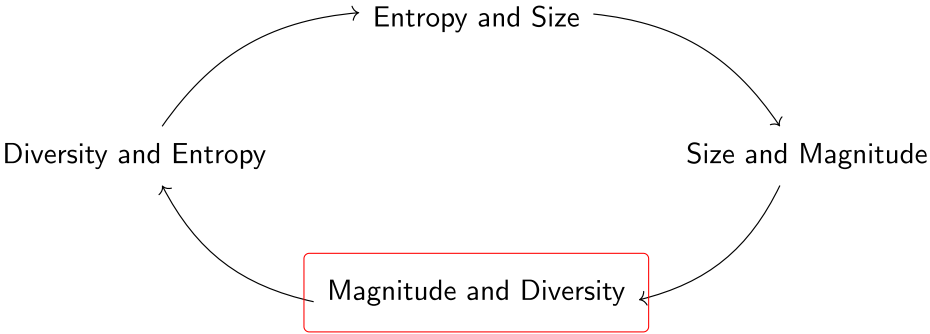
Corollary Suppose X is such that Z is positive definite and the weighting $\mathbf{w}$ is non-negative. Then

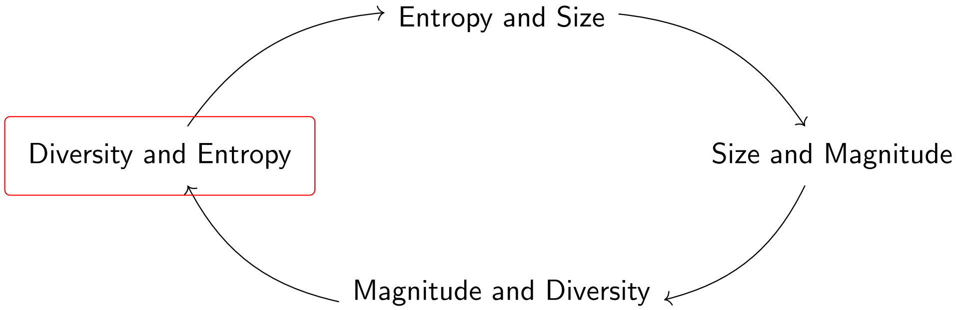
$$|X| = \max_{\mathbf{p} \in \mathcal{P}(X)} D_q^Z(\mathbf{p}) \text{ for } \underline{\text{every}}\ q \geqslant 0.$$

Example



The maximizing distribution.


Entropy and Size
Diversity and Entropy
Size and Magnitude
Magnitude and Diversity


Entropy and Size
Diversity and Entropy
Size and Magnitude
Magnitude and Diversity

Recovering Shannon Entropy from Order-1 Diversity

Let $Z_t(x, y) = e^{-td(x,y)}$; then $\lim_{t\to\infty} Z_t = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$.

Recovering Shannon Entropy from Order-1 Diversity

Let $Z_t(x,y) = e^{-td(x,y)}$; then $\lim_{t\to\infty} Z_t = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$.

Theorem (Leinster, 2010) For every finite space X,

$$\lim_{t\to\infty} |tX| = \#X.$$

Intuition By sending t to ∞, we zoom in so close that we forget the metric on X: all its points look isolated.

Proof idea. Matrix inversion is continuous, so $\lim_{t\to\infty} Z_t^{-1} = \mathbb{I}$ too, and $\sum \mathbb{I} = \#X$. □

Recovering Shannon Entropy from Order-1 Diversity

Let $Z_t(x,y) = e^{-td(x,y)}$; then $\lim_{t\to\infty} Z_t = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$.

Theorem (Leinster, 2010) For every finite space X,

$$\lim_{t\to\infty} |tX| = \#X.$$

Intuition By sending t to ∞, we zoom in so close that we forget the metric on X: all its points look isolated.

Theorem For every probability distribution $\mathbf{p}$ on X,

$$\lim_{t\to\infty} \log D_1^{Z_t}(\mathbf{p}) = \mathrm{H}(\mathbf{p}).$$

Proof idea. Matrix inversion is continuous, so $\lim_{t\to\infty} Z_t^{-1} = \mathbb{I}$ too, and $\sum \mathbb{I} = \#X$. □

Recovering Shannon Entropy from Order-1 Diversity

Let $Z_t(x,y) = e^{-td(x,y)}$; then $\lim_{t\to\infty} Z_t = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$.

Theorem (Leinster, 2010) For every finite space X,

$$\lim_{t\to\infty} |tX| = \#X.$$

Intuition By sending t to ∞, we zoom in so close that we forget the metric on X: all its points look isolated.

Theorem For every probability distribution $\mathbf{p}$ on X,

$$\lim_{t\to\infty} \log D_1^{Z_t}(\mathbf{p}) = \mathrm{H}(\mathbf{p}).$$

Proof idea. Matrix inversion is continuous, so $\lim_{t\to\infty} Z_t^{-1} = \mathbb{I}$ too, and $\sum \mathbb{I} = \#X$. □

Proof idea. $D_1^Z(\mathbf{p})$ is continuous in Z, so

$$\begin{aligned} &\lim_{t\to\infty} \log D_1^{Z_t}(\mathbf{p}) \\ &= \log D_1^{\mathbb{I}}(\mathbf{p}) \\ &= \log \left(\prod \mathbf{p}(x)^{-\mathbf{p}(x)}\right) \\ &= \mathrm{H}(\mathbf{p}). \end{aligned}$$

□

Recovering Shannon Entropy from Order-1 Diversity

Let $Z_t(x, y) = e^{-td(x,y)}$; then $\lim_{t\to\infty} Z_t = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$.

Theorem (Leinster, 2010) For every finite space X,

$$\lim_{t\to\infty} |tX| = \#X.$$

> *Proof idea.* Matrix inversion is continuous, so $\lim_{t\to\infty} Z_t^{-1} = \mathbb{I}$ too, and $\sum \mathbb{I} = \#X$. □

Intuition By sending t to ∞, we zoom in so close that we forget the metric on X: all its points look isolated.

Theorem For every probability distribution $\mathbf{p}$ on X,

$$\lim_{t\to\infty} \log D_1^{Z_t}(\mathbf{p}) = \mathrm{H}(\mathbf{p}).$$

> *Proof idea.* $D_1^Z(\mathbf{p})$ is continuous in Z, so
>
> $$\lim_{t\to\infty} \log D_1^{Z_t}(\mathbf{p}) = \log D_1^{\mathbb{I}}(\mathbf{p}) = \log\left(\prod \mathbf{p}(x)^{-\mathbf{p}(x)}\right) = \mathrm{H}(\mathbf{p}).$$
>
> □

Idea $\log D_1^Z(\mathbf{p})$ is the 'effective Shannon entropy' of $\mathbf{p}$.

A Variational Principle for Magnitude

Definition

The **effective entropy of order** q of $\mathbf{p} \in \mathcal{P}(X)$ is $\mathrm{H}_q^Z(\mathbf{p}) = \log D_q^Z(\mathbf{p})$.

A Variational Principle for Magnitude

Definition

The **effective entropy of order** q of $\mathbf{p} \in \mathcal{P}(X)$ is $\mathrm{H}_q^Z(\mathbf{p}) = \log D_q^Z(\mathbf{p})$.

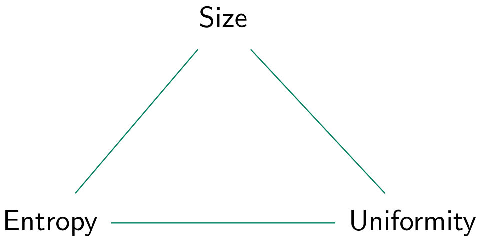
Theorem

Suppose X has positive definite similarity matrix and non-negative weighting. Then

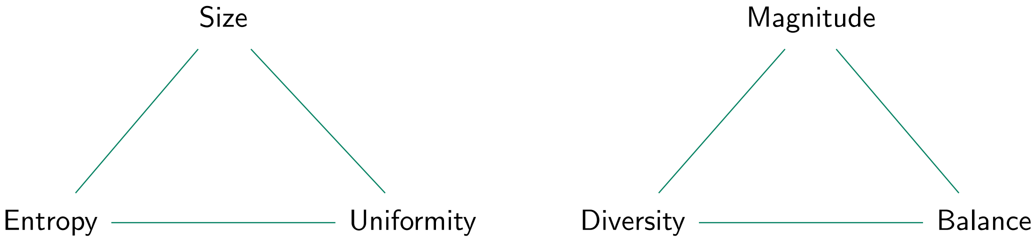
$$|X| = \max_{\mathbf{p} \in \mathcal{P}(X)} (\exp(\mathrm{H}_q^Z(\mathbf{p}))) \text{ for every } q \geqslant 0.$$

□

Summary



Summary



Thank you.

References

- Leinster. The magnitude of metric spaces. arXiv:1012.5857, 2010; *Documenta Mathematica* 18, 2013.
- Leinster and Cobbold. Measuring diversity: the importance of species similarity. *Ecology* 93, 2012.
- Leinster and Meckes. Maximizing diversity in biology and beyond. arXiv:1512.06314, 2015; *Entropy* 18, 2016.
- Leinster and Roff. The maximum entropy of a metric space. arXiv:1908.11184, 2019; *The Quarterly Journal of Mathematics* 72(4), 2021.
- Roff and Yoshinaga. The small-scale limit of magnitude and the one-point property. arXiv:2312.14497, 2023.