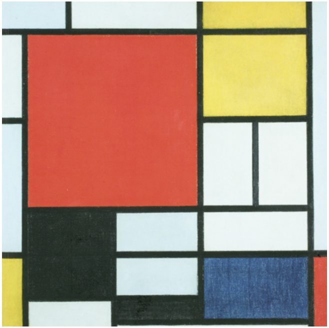
Maximum entropy, uniform measure

Emily Roff
The University of Edinburgh

ML@CL Seminar
Cambridge Computer Laboratory
13th November 2020

Turner

Turner



Mondrian

Plan

I. Quantifying diversity

II. Diversity and entropy

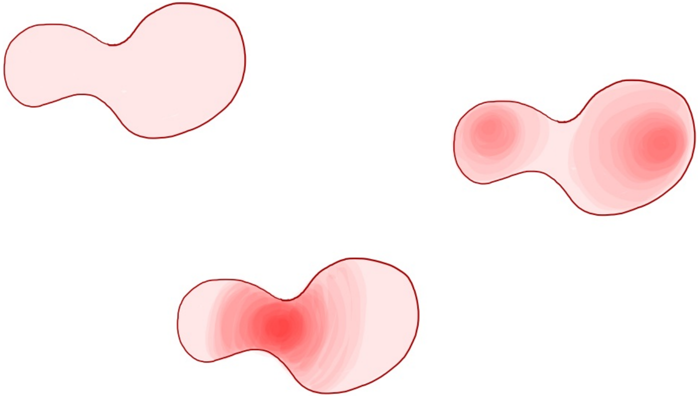
III. Maximizing entropy

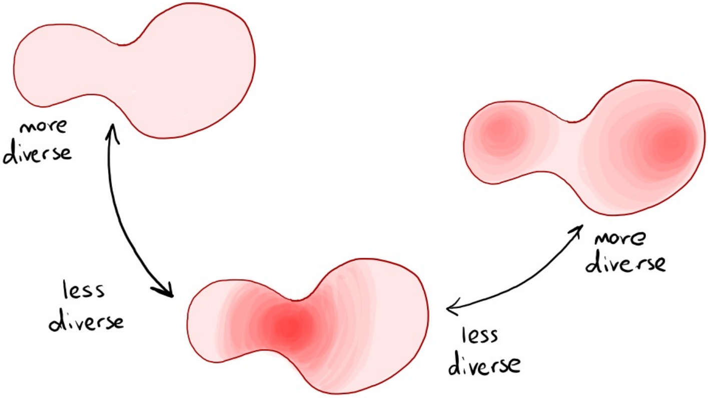
IV. Uniform measure

V. Categorical connections

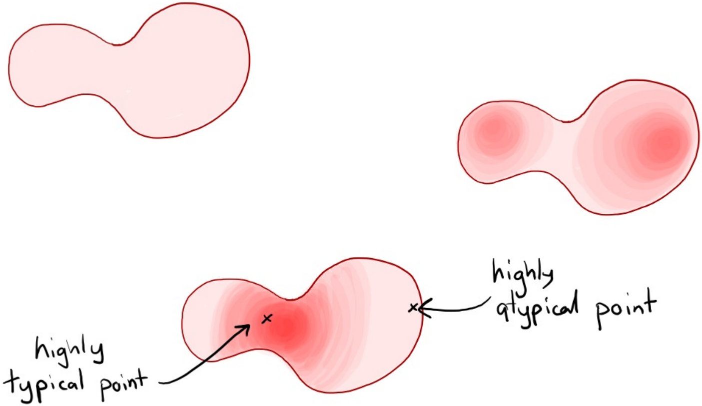
Part I

Quantifying Diversity

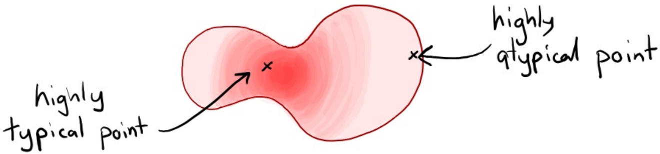




more
diverse
less
diverse
less
diverse
more
diverse


highly
atypical point
highly
typical point

"Diversity is average atypicality."



Spaces with similarities

Definition

Let X be a compact Hausdorff topological space.

A **similarity kernel** on X is a continuous function $K : X \times X \to [0, \infty)$ satisfying $K(x, x) > 0$ for all $x \in X$.

The pair (X, K) is called a **space with similarities**.

It's **symmetric** if $K(x, y) = K(y, x)$ for all $x, y \in X$.

Spaces with similarities

Definition

Let X be a compact Hausdorff topological space.

A **similarity kernel** on X is a continuous function $K : X \times X \to [0, \infty)$ satisfying $K(x, x) > 0$ for all $x \in X$.

The pair (X, K) is called a **space with similarities**.

It's **symmetric** if $K(x, y) = K(y, x)$ for all $x, y \in X$.

Example

A compact metric space with metric d has similarity kernel

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

When $X = \mathbb{R}^d$ this is the Laplace kernel.

Typicality functions

Definition

Let (X, K) be a space with similarities.

For each probability distribution μ on X, and each $x \in X$, define

$$(K\mu)(x) = \int K(x, -)\, \mathrm{d}\mu \in [0, \infty).$$

The function $K\mu : X \to [0, \infty)$ is the **typicality function** of μ.

The **atypicality function** of μ is $1/K\mu$.

Typicality functions

Definition

Let (X, K) be a space with similarities.

For each probability distribution μ on X, and each $x \in X$, define

$$(K\mu)(x) = \int K(x, -) \, \mathrm{d}\mu \in [0, \infty).$$

The function $K\mu : X \to [0, \infty)$ is the **typicality function** of μ.

The **atypicality function** of μ is $1/K\mu$.

Example

If X is a compact metric space, the typicality function of μ is given by

$$(K\mu)(x) = \int e^{-d(x,y)} \, \mathrm{d}\mu(y).$$

Diversity

Definition

Let (X, K) be a space with similarities, and μ a probability distribution on X. For $q \in [0, \infty)$ not equal to 1, the **diversity of order q** of μ is

$$D_q^K(\mu) = \left(\int \left(\frac{1}{K\mu} \right)^{1-q} \mathrm{d}\mu \right)^{1/(1-q)}.$$

At $q = 1, \infty$ this expression takes its limiting values.

Diversity

Definition

Let (X, K) be a space with similarities, and μ a probability distribution on X. For $q \in [0, \infty)$ not equal to 1, the **diversity of order q** of μ is

$$D_q^K(\mu) = \left(\int \left(\frac{1}{K\mu} \right)^{1-q} \mathrm{d}\mu \right)^{1/(1-q)} .$$

At $q = 1, \infty$ this expression takes its limiting values.

Example

If X is a compact metric space, then

$$D_q(\mu) = \left(\int \left(\int e^{-d(x,y)} \, \mathrm{d}\mu(x) \right)^{q-1} \mathrm{d}\mu(y) \right)^{1/(1-q)} .$$

Part II

Diversity and Entropy

Diversity on finite sets

Equip the set $X = \{x_1, \ldots, x_n\}$ with the similarity kernel K (a matrix). Let $\mu = (\mu_1, \ldots, \mu_n)$ be a probability distribution on X.

The diversity of order q of μ is

$$D_q^K(\mu) = \left(\sum_{\mathrm{supp}\mu} (K\mu)_i^{q-1} \mu_i \right)^{1/(1-q)} .$$

Diversity on finite sets

Equip the set $X = \{x_1, \ldots, x_n\}$ with the similarity kernel $K = I$.
Let $\mu = (\mu_1, \ldots, \mu_n)$ be a probability distribution on X.

The diversity of order q of μ is

$$D_q^I(\mu) = \left(\sum_{\mathrm{supp}\mu} \mu_i^q \right)^{1/(1-q)} = \exp\left(H_q(\mu)\right)$$

where H_q is the **Rényi entropy** of order q.

Diversity on finite sets

Equip the set $X = \{x_1, \ldots, x_n\}$ with the similarity kernel $K = I$.
Let $\mu = (\mu_1, \ldots, \mu_n)$ be a probability distribution on X.

The diversity of order q of μ is

$$D_q^I(\mu) = \left(\sum_{\mathrm{supp}\mu} \mu_i^q \right)^{1/(1-q)} = \exp\left(H_q(\mu)\right)$$

where H_q is the **Rényi entropy** of order q.

In particular,

$$D_1^I(\mu) = \exp\left(-\sum \mu_i \log \mu_i \right) = \exp(\textbf{Shannon}(\mu)).$$

Entropy in ecology

To ecologists, $\exp(H_q(\mu))$ is known as the **Hill number** of order q.

The Hill numbers are used as measures of ecological diversity.

Strategy Model an ecological community by a set of species X and a distribution μ on X, representing the relative abundances of species.

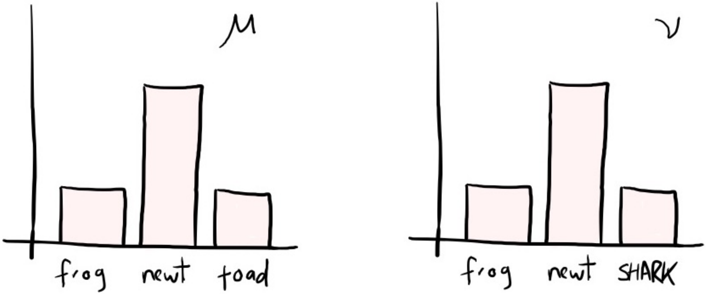
Then take the Hill number

$$\exp(H_q(\mu)) = D_q^I(\mu)$$

to quantify the 'diversity' of the community.

Entropy in ecology

Problem The Hill numbers don't see similarities between species.



$$H_q(\mu) = H_q(\nu)$$

Similarity-sensitive diversity

Solution (Cobbold and Leinster, 2012)

Record pairwise similarities of the species in a matrix, K. Define the **similarity-sensitive diversity of order** q to be

$$D_q^K(\mu) = \left(\sum_{\text{supp}\mu} (K\mu)_i^{q-1} \mu_i \right)^{1/(1-q)}.$$

This is where our diversity measures originate.

Similarity-sensitive diversity

Solution (Cobbold and Leinster, 2012)

Record pairwise similarities of the species in a matrix, K. Define the **similarity-sensitive diversity of order q** to be

$$D_q^K(\mu) = \left(\sum_{\text{supp}\mu} (K\mu)_i^{q-1} \mu_i \right)^{1/(1-q)}.$$

This is where our diversity measures originate.

For example,

$$D_2^K = \frac{1}{\text{expected similarity of two individuals chosen at random}}$$

while

$$D_2^I = \frac{1}{\text{probability that they're of the same species}}.$$

Similarity-sensitive entropy

In the general setting of a space with similarities, we define

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

Definition

Let (X, K) be a space with similarities, and μ a distribution on X. For $q \in [0, \infty]$, the **entropy of order q** of μ is $H_q^K(\mu) = \log D_q^K(\mu)$.

Similarity-sensitive entropy

In the general setting of a space with similarities, we define

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

Definition

Let (X, K) be a space with similarities, and μ a distribution on X. For $q \in [0, \infty]$, the **entropy of order** q of μ is $H_q^K(\mu) = \log D_q^K(\mu)$.

Example

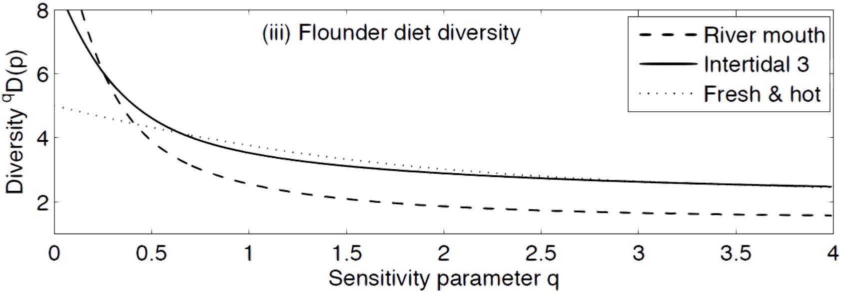
If X is a compact metric space, then

$$H_1(\mu) = -\int \log \left(\int e^{-d(x,y)} \, \mathrm{d}\mu(x) \right) \mathrm{d}\mu(y).$$

Part III

Maximizing Diversity and Entropy

The parameter q matters!



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

A maximum entropy theorem

Theorem (Leinster and Roff, 2019)

Let (X, K) be a symmetric space with similarities.

1. *There exists a probability distribution μ on X that maximises $D_q^K(\mu)$ and $H_q^K(\mu)$ for all $q \in [0, \infty]$ at once.*
2. *The maximum diversity $\sup_\mu D_q^K(\mu)$ and maximum entropy $\sup_\mu H_q^K(\mu)$ are independent of $q \in [0, \infty]$.*

A maximum entropy theorem

Theorem (Leinster and Roff, 2019)

Let (X, K) be a symmetric space with similarities.

1. *There exists a probability distribution μ on X that maximises $D_q^K(\mu)$ and $H_q^K(\mu)$ for all $q \in [0, \infty]$ at once.*
2. *The maximum diversity $\sup_\mu D_q^K(\mu)$ and maximum entropy $\sup_\mu H_q^K(\mu)$ are independent of $q \in [0, \infty]$.*

- If μ maximises H_q^K for one q, it maximises for all q.
- In general μ need not be unique, but if K is positive-definite, it is.
- So every compact subset of $\mathbb{R}^n$ has a unique distribution of maximum entropy. For almost all sets, we don't know what it is!

New invariants

Definition

Let (X, K) be a symmetric space with similarities.

The **maximum diversity** of X is

$$D_{\max}(X) = \sup_{\mu} D_q(\mu) \text{ for any } q.$$

The **maximum entropy** of X is

$$H_{\max}(X) = \log D_{\max}(X).$$

A distribution attaining the supremum is called **maximising**.

Part IV

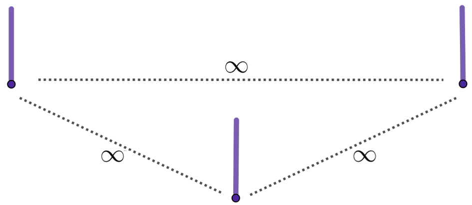
Uniform Distributions

A maximising measure

Take a finite set X, and $K = I$. Then

$$D_{\max}(X) = \sup D_1^I = \sup(\exp(\text{Shannon}))$$

which is uniquely attained by the uniform distribution.

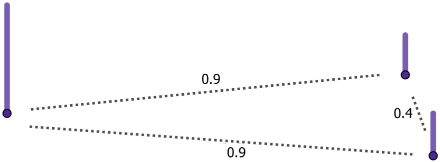


A maximising measure

Take a finite set X, and $K = I$. Then

$$D_{\max}(X) = \sup D_1^I = \sup(\exp(\text{Shannon}))$$

which is uniquely attained by the uniform distribution.



This no longer holds when $K \neq I$.

Balance

Maximising distributions possess a different sort of 'evenness', which is responsive to the geometry of the space.

Balance

Maximising distributions possess a different sort of 'evenness', which is responsive to the geometry of the space.

A distribution μ on X is **balanced** if $K\mu$ is constant on $\mathrm{supp}(\mu)$.

Lemma

Any maximising distribution is balanced.

Balance

Maximising distributions possess a different sort of 'evenness', which is responsive to the geometry of the space.

A distribution μ on X is **balanced** if $K\mu$ is constant on $\text{supp}(\mu)$.

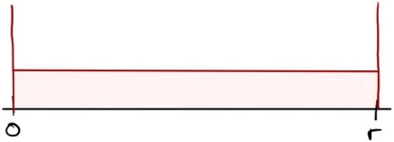
Lemma

Any maximising distribution is balanced.

Example

If an ecological community is maximally diverse, then all the species present must be equally typical.

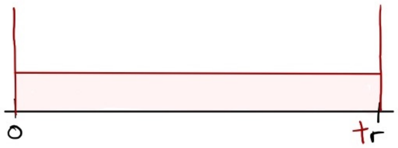
Balance is not uniformity



Consider $[0, r] \subset \mathbb{R}$. Its maximising distribution is

$$\mu = \frac{\delta_0 + \lambda_{[0,r]} + \delta_r}{2 + r}.$$

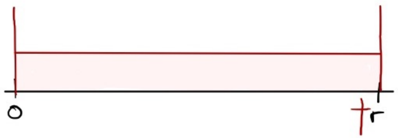
Balance is not uniformity



Scale the space by $t > 0$. The maximising distribution on $[0, tr]$ is

$$\mu_t = \frac{\delta_0 + t\lambda_{[0,tr]} + \delta_r}{2 + tr}.$$

Balance is not uniformity



Scale the space by $t > 0$. The maximising distribution on $[0, tr]$ is

$$\mu_t = \frac{\delta_0 + t\lambda_{[0,tr]} + \delta_r}{2 + tr}.$$

Balance is not uniformity

Scale the space by $t > 0$. The maximising distribution on $[0, tr]$ is

$$\mu_t = \frac{\delta_0 + t\lambda_{[0,tr]} + \delta_r}{2 + tr}.$$

Balance is not uniformity

Scale the space by $t > 0$. The maximising distribution on $[0, tr]$ is

$$\mu_t = \frac{\delta_0 + t\lambda_{[0,tr]} + \delta_r}{2 + tr}.$$

Uniform distribution

Given a metric space X and any $t \in [0, \infty)$, write tX for the space X after its distances have been scaled by t.

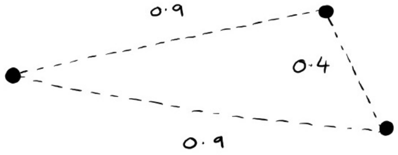
Definition

Let X be a compact metric space. Suppose tX has a unique maximising distribution μ_t for all $t \gg 0$, and that $\lim_{t \to \infty} \mu_t$ exists in $P(X)$.

Then the **uniform distribution** on X is

$$\mu_X = \lim_{t \to \infty} \mu_t.$$

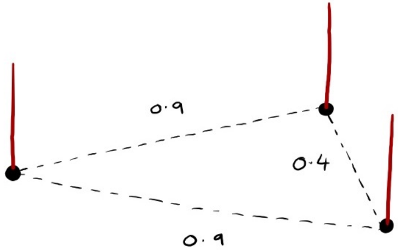
Example



Proposition

On a finite metric space, the uniform measure is the uniform measure.

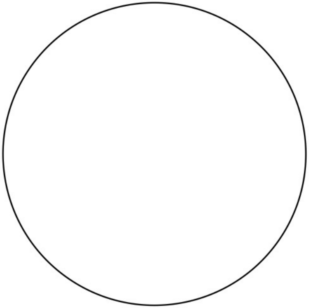
Example



Proposition

On a finite metric space, the uniform measure is the uniform measure.

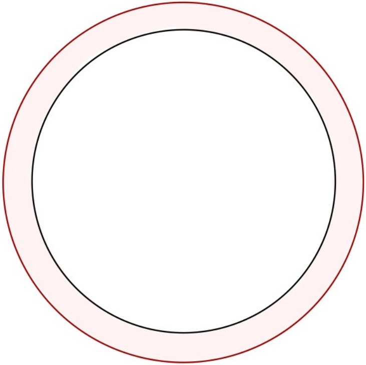
Example



Proposition

On a homogeneous space, the uniform measure is the Haar measure.

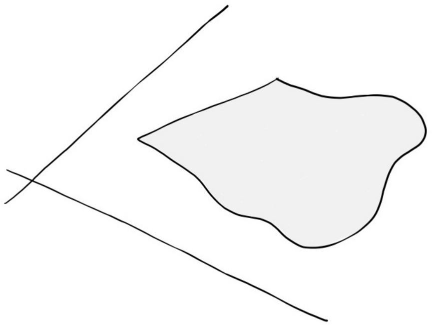
Example



Proposition

On a homogeneous space, the uniform measure is the Haar measure.

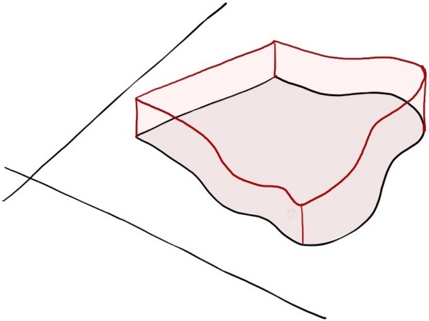
Example



Proposition

On a compact subset of $\mathbb{R}^n$ with nonzero volume, the uniform measure is normalised Lebesgue measure.

Example



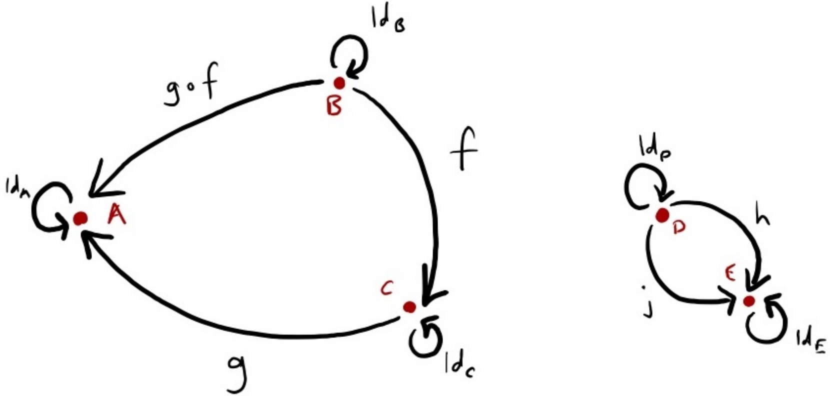
Proposition

On a compact subset of $\mathbb{R}^n$ with nonzero volume, the uniform measure is normalised Lebesgue measure.

Part V

Categorical Connections

Ordinary categories



Ordinary categories

A **category** $\mathbf{A}$ consists of

- (Objects) A set $\mathrm{ob}\,(\mathbf{A})$
- (Morphisms) For each $A, B \in \mathrm{ob}\,(\mathbf{A})$ a set $\mathrm{Hom}(A, B)$
- (Identities) For each $A \in \mathrm{ob}\,(\mathbf{A})$ a function $\{*\} \to \mathrm{Hom}(A, A)$
- (Composition) For each $A, B, C \in \mathrm{ob}\,(\mathbf{A})$ a function

$$\mathrm{Hom}(A, B) \times \mathrm{Hom}(B, C) \to \mathrm{Hom}(A, C).$$

Ordinary categories

A **category** $\mathbf{A}$ consists of

- (Objects) A set $\text{ob}\,(\mathbf{A})$
- (Morphisms) For each $A, B \in \text{ob}\,(\mathbf{A})$ a set $\text{Hom}(A, B)$
- (Identities) For each $A \in \text{ob}\,(\mathbf{A})$ a function $\{*\} \to \text{Hom}(A, A)$
- (Composition) For each $A, B, C \in \text{ob}\,(\mathbf{A})$ a function

$$\text{Hom}(A, B) \times \text{Hom}(B, C) \to \text{Hom}(A, C).$$

Idea

Why not replace sets and functions with something more interesting?

All we really need is a 'multiplication' like $\times$ with a 'unit' like $\{*\}$.

Enriched categories

Let $\mathcal{V}$ be a category with a monoidal product $\otimes$ and unit I.

A **$\mathcal{V}$-category** $\mathbf{A}$ consists of

- (Objects) A set $\text{ob}\,(\mathbf{A})$
- (Morphisms) For each $A, B \in \text{ob}\,(\mathbf{A})$ an object of $\mathcal{V}$, $\text{Hom}(A, B)$
- (Identities) For each $A \in \text{ob}\,(\mathbf{A})$ an arrow in $\mathcal{V}$, $I \to \text{Hom}(A, A)$
- (Composition) For each $A, B, C \in \text{ob}\,(\mathbf{A})$ an arrow in $\mathcal{V}$,

$$\text{Hom}(A, B) \otimes \text{Hom}(B, C) \to \text{Hom}(A, C).$$

Enriched categories

Let $\mathcal{V}$ be a category with a monoidal product $\otimes$ and unit I.

A **$\mathcal{V}$-category** $\mathbf{A}$ consists of

- (Objects) A set $\text{ob}(\mathbf{A})$
- (Morphisms) For each $A, B \in \text{ob}(\mathbf{A})$ an object of $\mathcal{V}$, $\text{Hom}(A, B)$
- (Identities) For each $A \in \text{ob}(\mathbf{A})$ an arrow in $\mathcal{V}$, $I \to \text{Hom}(A, A)$
- (Composition) For each $A, B, C \in \text{ob}(\mathbf{A})$ an arrow in $\mathcal{V}$,

$$\text{Hom}(A, B) \otimes \text{Hom}(B, C) \to \text{Hom}(A, C).$$

Examples

- If $\mathcal{V} = (\bot \to \top, \wedge)$, a $\mathcal{V}$-category is a preorder.
- If $\mathcal{V} = (\mathbf{Vect}, \otimes)$, a one-object $\mathcal{V}$-category is an associative algebra.

$[0, \infty)$-categories

Let $\mathcal{V} = [0, \infty)$. It's a category: there's an arrow $x \to y$ if and only if $x \geq y$. It has a monoidal product $+$ with unit 0.

A **$[0, \infty)$-category** $\mathbf{X}$ consists of

- (Objects) A set $\mathbf{X}$
- (Morphisms) For each $x, y \in \mathbf{X}$ a number $d(x, y) \in [0, \infty)$
- (Identities) For each $x \in \mathbf{X}$ an inequality $0 \geq d(x, x)$
- (Composition) For each $x, y, z \in \mathbf{X}$ an inequality,

$$d(x, y) + d(y, z) \geq d(x, z).$$

$[0,\infty)$-categories

Let $\mathcal{V} = [0,\infty)$. It's a category: there's an arrow $x \to y$ if and only if $x \geq y$. It has a monoidal product $+$ with unit 0.

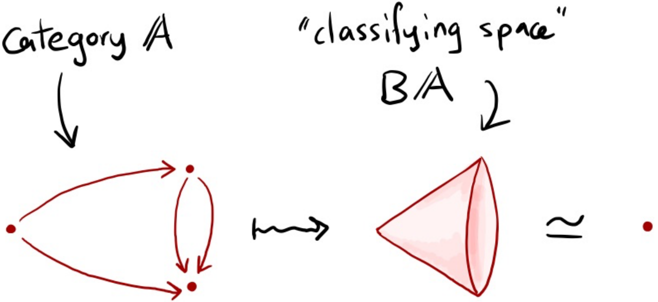
A **$[0,\infty)$-category** $\mathbf{X}$ consists of

- (Objects) A set $\mathbf{X}$
- (Morphisms) For each $x, y \in \mathbf{X}$ a number $d(x,y) \in [0,\infty)$
- (Identities) For each $x \in \mathbf{X}$ an inequality $0 \geq d(x,x)$
- (Composition) For each $x, y, z \in \mathbf{X}$ an inequality,

$$d(x,y) + d(y,z) \geq d(x,z).$$

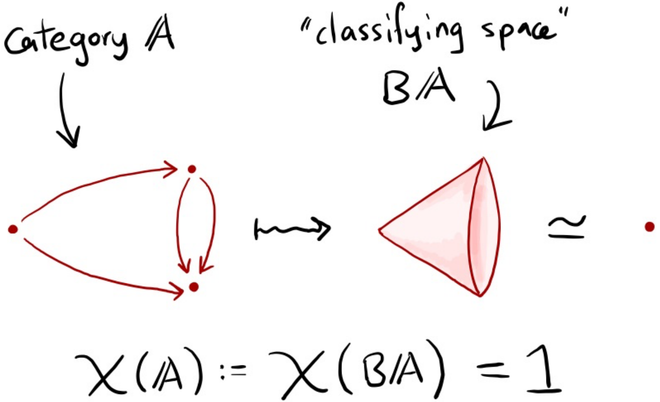
Moral A $[0,\infty)$-category is a generalized metric space.

The Euler characteristic of a category



$$\chi(\mathbb{A}) := \chi(B\mathbb{A}) = 1$$

The Euler characteristic of a category



Questions

- What is the Euler characteristic of an enriched category?
- What is the Euler characteristic of a metric space?

The magnitude (or Euler characteristic) of a metric space

Let X be a compact metric space.

Definition

A **weighting** on X is a signed measure ν such that $K\nu \equiv 1$.

If X possesses a weighting ν, the **magnitude** of X is

$$\chi(X) := \nu(X).$$

The magnitude (or Euler characteristic) of a metric space

Let X be a compact metric space.

Definition

A **weighting** on X is a signed measure ν such that $K\nu \equiv 1$.

If X possesses a weighting ν, the **magnitude** of X is

$$\chi(X) := \nu(X).$$

Now let μ be a maximum entropy distribution on X.

We know it's balanced: $K\mu|_{\text{supp }\mu} \equiv c$ for some constant c.

So its restriction to supp μ is proportional to a weighting, $\hat{\mu} = \frac{1}{c}\mu$.

Theorem

$D_{\max}(X) = \chi(\text{supp }\mu)$ *for any maximising measure* μ*.*

Bibliography

- Leinster and Cobbold, *Measuring diversity: the importance of species similarity*. Ecology 93 (2012).
- Leinster and Meckes, *Maximizing diversity in biology and beyond*. Entropy 18(3) (2016).
- Leinster and Meckes, *The magnitude of a metric space: from category theory to geometric measure theory*. Nicola Gigli (ed.), Measure Theory in Non-Smooth Spaces, de Gruyter Open, (2017).
- Leinster and Roff, *The maximum entropy of a metric space*. arXiv:1908.11184 (2019).