What is the uniform distribution?

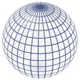
Tom Leinster

Edinburgh

Prelude

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?

Plan

1. Measuring diversity

2. Maximizing diversity

3. The uniform measure

Conclusion

1. Measuring diversity

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

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

Spaces with similarities

Motivating examples:

- Given a finite set X of species, there are various ways to quantify the similarity $K(x, y)$ between species $x, y \in X$: genetic, phylogenetic, ... Or very crudely, can put

$$K(x, y) = \delta(x, y) = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{otherwise.} \end{cases}$$

- Given a metric space $X = (X, d)$, define the 'similarity' between points x and y to be

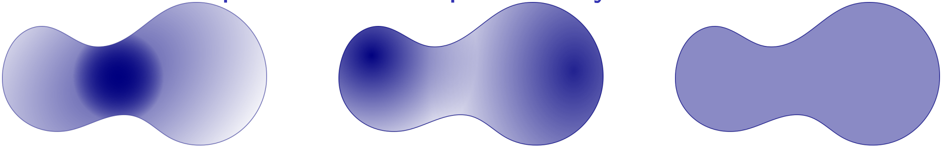
$$K(x, y) = e^{-d(x,y)}.$$

Definition A similarity kernel K on a topological space X is a continuous symmetric function $K \colon X \times X \to \mathbb{R}^+$ such that $K(x, x) > 0$ for all $x \in X$.

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

We'll always assume that X is compact Hausdorff.

How spread out is a probability distribution?



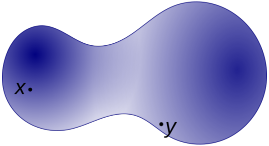
Let (X, K) be a space with similarities and $\mathbb{P}$ a probability measure on X.

The typicality of a point $x \in X$ is

$$K\mathbb{P}(x) = \int_X K(x, y)\, d\mathbb{P}(y).$$

It measures how much mass is nearby.

Here x is more typical than y:



The atypicality of x is $\dfrac{1}{K\mathbb{P}(x)}$.

How spread out is a probability distribution?

Let (X, K) be a space with similarities and $\mathbb{P}$ a probability measure on X.

Spread can be quantified as the average atypicality of a point in X.

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

$$\int_X \frac{1}{K\mathbb{P}(x)}\, d\mathbb{P}(x),$$

but it's useful to consider *all* power means:

Definition Let $q \in [0, \infty]$. The diversity of order q of $\mathbb{P}$ is

$$D_q(\mathbb{P}) = \left(\int_X \left(\frac{1}{K\mathbb{P}(x)} \right)^{1-q} d\mathbb{P}(x) \right)^{\frac{1}{1-q}}.$$

This is undefined for $q = 1, \infty$, but can take limits in q to get the right definitions there.

The entropy of order q of $\mathbb{P}$ is $H_q(\mathbb{P}) = \log D_q(\mathbb{P})$.

Example: finite probability distributions

Let X be a finite set with a probability distribution $\mathbb{P}$.

Use the crude similarity kernel $K = \delta$.

Then $H_q(\mathbb{P})$ is the Rényi entropy of order q of $\mathbb{P}$.

In particular, $H_1(\mathbb{P})$ is the Shannon entropy of $\mathbb{P}$.

Example: biological diversity

Take X to be a finite set of species and $K(x, y)$ to be the similarity between species x and y.

Then $D_q(\mathbb{P})$ measures the diversity of a community with species abundance distribution $\mathbb{P}$.

This encompasses and unifies many of the biodiversity measures used by ecologists.

Example: the case $q = 2$

For any space with similarities (X, K), and any probability measure $\mathbb{P}$ on it,

$$D_2(\mathbb{P}) = \frac{1}{\int_X \int_X K(x, y)\, d\mathbb{P}(x)\, d\mathbb{P}(y)}.$$

The denominator is the expected similarity between a random pair of points.

Intuition High expected similarity means high concentration, or low diversity.

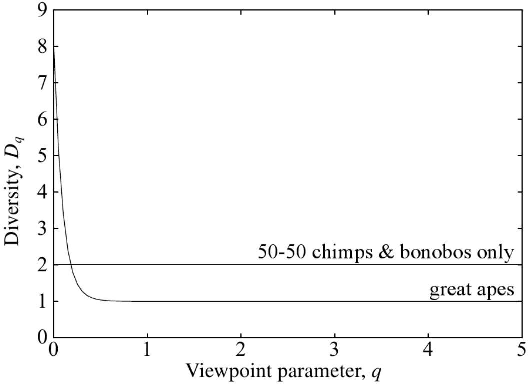
The role of q

In the definition of diversity $D_q(\mathbb{P})$ and entropy $H_q(\mathbb{P}) = \log D_q(\mathbb{P})$, there is a real parameter q. What does it do?

Example Take $X = \{1, \ldots, 8\}$ with the crude similarity kernel $K = \delta$.

Take $\mathbb{P}$ to be the frequencies of the eight species of great ape on the planet.

Let $\mathbb{Q}$ be the 50-50 distribution of chimpanzees and bonobos only.



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

The answer may depend on q.

2. Maximizing diversity

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

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

The maximum diversity theorem

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

Which probability measure on X achieves the maximum possible diversity (or entropy) on X? 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 $\mathbb{P}$ maximizing $D_q(\mathbb{P})$ for all $q \in [0, \infty]$ simultaneously
- $\sup_{\mathbb{P}} D_q(\mathbb{P})$ is independent of q

(and the same for entropy H_q).

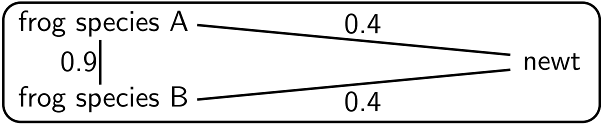
If $\mathbb{P}$ maximizes $D_q(\mathbb{P})$ for all q, we call $\mathbb{P}$ a maximizing measure. It is usually unique.

Warning Even when X is finite, the maximizing measure is not usually uniform!

Example: frogs and newts

Take a three-species system ($X = \{1, 2, 3\}$) with these similarities:

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



Which distribution on X 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. And it is: it's $(0.261, 0.261, 0.478)$.

In particular, the maximizing distribution is not uniform.

Example: real interval

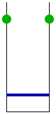
Take a real interval $[0, L]$ of length L with its usual metric, so that $K(x, y) = e^{-|x-y|}$.

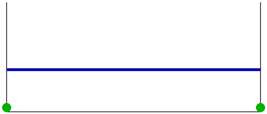
Which probability measure on $[0, L]$ maximizes diversity (or equivalently, entropy)?

- 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: 

3. The uniform measure

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

Overview

Goal Given a space X of a suitable kind, assign to it a 'uniform probability measure' $\mathbb{U}_X$ in some canonical way.

Wishlist:

- It should agree with the existing notions of uniform measure when X is finite, or symmetric, or $\subseteq \mathbb{R}^n$.
- It should be scale-independent: scaling the metric on X up or down should not affect $\mathbb{U}_X$.

We'll achieve this by combining two ideas:

- Maximize diversity ($\Leftrightarrow$ entropy). Intuition: maximum diversity distributions are maximally spread out, or maximally uniform.
- Scale up to infinity. This removes the scale-dependence.

The definition

Let X be a compact metric space. For $t > 0$, we write tX for X equipped with the similarity kernel

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

So t acts as a scale factor.

Assume that for each $t > 0$, the maximizing measure $\mathbb{P}_t$ on tX is unique, and that $\lim_{t\to\infty} \mathbb{P}_t$ exists (in the weak* topology).

Definition The uniform measure on X is $\mathbb{U}_X = \lim_{t\to\infty} \mathbb{P}_t$.

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

Easy case 1: finite spaces

Let X be a finite metric space. Then the uniform measure $\mathbb{U}_X$ on X is the uniform measure in the usual sense.

Why?

For $t < \infty$, the maximizing measure $\mathbb{P}_t$ on tX is *not* uniform (remember the frogs and newts).

But as $t \to \infty$, all the similarities $K^t(x, y) = e^{-td(x,y)} \to 0$ for $x \neq y$. Hence $K \to \delta$. And the maximizing measure for $K = \delta$ is uniform.

Easy case 2: symmetric spaces

Let X be a compact metric space that's symmetric in the sense that for all $x, y \in X$, there is some distance-preserving symmetry of X that maps x to y.

The Haar measure theorem implies that there is a unique symmetry-invariant probability measure on X. And that's what $\mathbb{U}_X$ is.

Why? Because $\mathbb{P}_t$ is this unique symmetric measure for each finite t.

Slightly less easy case: the interval

Let $L > 0$ and 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 $\mathbb{U}_{[0,L]}$ is the normalization of Lebesgue measure, $\lambda_{[0,L]}/L$.

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

Way 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 directly calculating the maximizing measure on tX for each finite t, then letting $t \to \infty$.

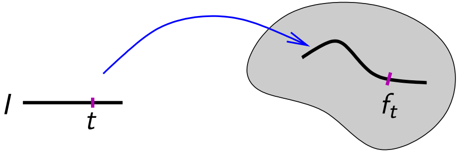
Nevertheless...

Theorem *Assuming that X has nonzero measure, the uniform measure $\mathbb{U}_X$ is normalized Lebesgue measure on X.*

Proof Lots of analysis.

Example: the Jeffreys prior

Let M be a measure space and let $(f_t)_{t \in I}$ be a smooth family of probability density functions on M, indexed over a real interval I.



Question Is there an 'objective prior' on I, i.e. a 'canonical' probability measure on I associated with this family?

One answer: take the distribution on I proportional to the square root of the Fisher information. This Jeffreys prior has excellent invariance properties.

It also comes out of our framework...

Put the Fisher metric on the set of PDFs on M.Give I the metric

$$d(t_1, t_2) = \text{arclength from } f_{t_1} \text{ to } f_{t_2} \text{ (wrt Fisher metric on } M\text{)}.$$

So (I, d) is a metric space, and we can ask what its uniform measure is.

It's the Jeffreys prior!

Conclusion

Summary

- For a probability measure $\mathbb{P}$ on a metric space (or more generally, on a space with some notion of 'similarity' between points), there is a one-parameter family $(D_q(\mathbb{P}))_{q>0}$ of diversities, quantifying how spread out the measure is.
- These diversities D_q are used in ecology to quantify biodiversity, and $\log D_q$ is a kind of entropy.
- Miraculously, there is a single probability measure on the space that maximizes D_q for all parameters q simultaneously!
- This maximizing measure changes as you scale the space up.
- Taking the large-scale limit gives a probability measure $\mathbb{U}_X$ on our space X, which deserves to be called the uniform measure on X.
- For spaces X where there's already an obvious notion of uniform measure, this definition agrees.

Open questions

- What properties does the uniform measure have?
- What properties would you *want* something with that name to have?
- What other examples of the uniform measure are there?
- What other examples would you want there to be?
  Are there any spaces that you'd like to be able to put some kind of canonical probability distribution on, but don't know how?

Thanks