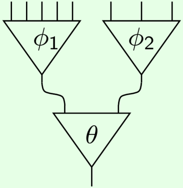
Entropy modulo a prime

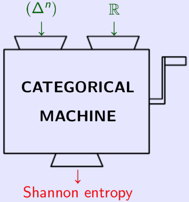
Tom Leinster

University of Edinburgh

Three talks

Sunday: Operads





Monday: An algebraic view of entropy

Thursday: Entropy modulo a prime

Trajectory of these three talks

SUNDAY MONDAY

I will explain how operads—a cousin of algebraic theories—lead to the notion of entropy (which might seem to belong to other branches of science).

Then I will show how this story leads to a mysterious construction in number theory.

THURSDAY

Confession

This talk doesn't have much to do with universal algebra, or lattice theory, or indeed category theory.

It's not clear *where* it fits into the mathematical landscape. No one knows! It's a mystery.

But this talk *does* follow on from the algebraic, axiomatic approach to entropy described in my last talk.

It channels the spirit of the previous talks, rather than the details.

Recap of Monday (just the bits we need)

Shannon entropy takes as input a finite probability distribution

$$\mathbf{p} = (p_1, \ldots, p_n) \qquad (p_i \geq 0, \sum p_i = 1)$$

and produces as output a real number,

$$H(\mathbf{p}) = -\sum_{i=1}^{n} p_i \log p_i \qquad (\text{where } 0 \log 0 = 0).$$

It satisfies a recursivity rule:

$$H\big(\mathbf{p} \circ (\mathbf{q}_1, \ldots, \mathbf{q}_n)\big) = H(\mathbf{p}) + \sum_{i=1}^{n} p_i H(\mathbf{q}_i),$$

where $\mathbf{p} \circ (\mathbf{q}_1, \ldots, \mathbf{q}_n)$ is the composite distribution

$$(p_1 q_1^1, \ldots, p_1 q_1^{k_1}, \ \ldots, \ p_n q_n^1, \ldots, p_n q_n^{k_n}).$$

H is the *only* continuous functional obeying this rule, up to a scalar factor.

Today's idea

$\mathbb{R}$ appears twice in the definition of Shannon entropy:

- the probabilities p_i are in $\mathbb{R}$
- the entropy $H(\mathbf{p})$ is in $\mathbb{R}$.

> Thinking algebraically: let's try to replace $\mathbb{R}$ by another field.

To make this generalization, how should we look at real entropy?

- As uniformity, or disorder, or genericity, or diversity, ...? No **x**
- As defined by the formula $-\sum p_i \log p_i$? Maybe helpful?
- As characterized by the recursivity rule? Yes ✓

We won't do general fields: just $\mathbb{Z}/p\mathbb{Z}$ for primes p.

What we'll do

Fix a prime p.

For each $n \geq 1$, write

$$\Pi_n = \{\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n) \in (\mathbb{Z}/p\mathbb{Z})^n \ : \ \sum \pi_i = 1\}$$

(the mod p analogue of the simplex Δ_n).

We'll define an "entropy mod p" functional

$$H_p \colon \Pi_n \to \mathbb{Z}/p\mathbb{Z}.$$

We'll see that it satisfies the recursivity rule, and that up to a scalar multiple, this characterizes it uniquely. So it's the right definition!

Then we'll start exploring...

Plan

1. Logarithms and derivations mod p
2. The definition of entropy mod p
3. The characterization theorem (or: why is this the right definition?)
4. Residues (or: what is the residue mod 7 of $\log \sqrt{8}$?)
5. Entropy as a polynomial (or: looking over the horizon)

1. Logarithms and derivations mod p

Logarithms mod p

We can try to take the real function $-\sum \pi_i \log \pi_i$ and imitate it mod p. But what is the analogue of "log"?

Problem The real logarithm is a group homomorphism $(\mathbb{R}_{>0}, \cdot) \to (\mathbb{R}, +)$.

But there is no nontrivial homomorphism $((\mathbb{Z}/p\mathbb{Z})^\times, \cdot) \to (\mathbb{Z}/p\mathbb{Z}, +)$.

Solution There is a next best thing: a homomorphism

$$q_p \colon ((\mathbb{Z}/p^2\mathbb{Z})^\times, \cdot) \to (\mathbb{Z}/p\mathbb{Z}, +),$$

namely, the Fermat quotient

$$q_p(a) = \frac{a^{p-1} - 1}{p}.$$

Up to a constant factor, it is the *only* such homomorphism.

How derivations come into it

It turns out that the function $x \mapsto -x \log x$ is just as important as the logarithm function.

If we define $\partial_{\mathbb{R}} \colon [0, \infty) \to \mathbb{R}$ by

$$\partial_{\mathbb{R}}(x) = \begin{cases} -x \log x & \text{if } x > 0 \\ 0 & \text{if } x = 0, \end{cases}$$

then real entropy $H_{\mathbb{R}}$ is given by

$$H_{\mathbb{R}}(\boldsymbol{\pi}) = \sum_{i=1}^{n} \partial_{\mathbb{R}}(\pi_i) \qquad (\boldsymbol{\pi} \in \Delta_n).$$

And $\partial_{\mathbb{R}}$ behaves like differentation!

$$\partial_{\mathbb{R}}(xy) = x\partial_{\mathbb{R}}(y) + \partial_{\mathbb{R}}(x)y, \qquad \partial_{\mathbb{R}}(1) = 0.$$

Except. . . it's not additive.

In fact, entropy measures the *failure of $\partial_{\mathbb{R}}$ to be additive:*

$$H_{\mathbb{R}}(\boldsymbol{\pi}) = \sum \partial_{\mathbb{R}}(\pi_i) - \partial_{\mathbb{R}}\Big(\sum \pi_i\Big).$$

Derivations mod p

We've already seen that the closest mod p analogue of log is the Fermat quotient

$$\begin{array}{rccc} q_p\colon & (\mathbb{Z}/p^2\mathbb{Z})^\times & \to & \mathbb{Z}/p\mathbb{Z} \\ & a & \mapsto & \frac{a^{p-1}-1}{p}. \end{array}$$

The mod p analogue of $\partial_{\mathbb{R}}$ is

$$\begin{array}{rccc} \partial_p\colon & \mathbb{Z}/p^2\mathbb{Z} & \to & \mathbb{Z}/p\mathbb{Z} \\ & a & \mapsto & \frac{a-a^p}{p}. \end{array}$$

This function ∂_p also behaves like differentiation:

$$\partial_p(ab) = a\partial_p(b) + \partial_p(a)b, \qquad \partial_p(1) = 0$$

for all $a, b \in \mathbb{Z}/p^2\mathbb{Z}$.

Except it's also not additive...

2. The definition of entropy mod p

The definition

Idea Real entropy satisfies

$$H_{\mathbb{R}}(\boldsymbol{\pi}) = \sum \partial_{\mathbb{R}}(\pi_i) - \partial_{\mathbb{R}}\Big(\sum \pi_i\Big) \qquad (\boldsymbol{\pi} \in \Delta_n).$$

To get the mod p analogue, where $\pi_i \in \mathbb{Z}/p\mathbb{Z}$, we can't just replace $\partial_{\mathbb{R}}$ by ∂_p: the domain of ∂_p is $\mathbb{Z}/p^2\mathbb{Z}$, whereas π_i is only defined mod p.

But we can do this instead:

Definition Let $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n) \in (\mathbb{Z}/p\mathbb{Z})^n$ with $\sum \pi_i = 1$. Its entropy mod p is

$$H_p(\boldsymbol{\pi}) = \sum_{i=1}^n \partial_p(a_i) - \partial_p\Big(\sum_{i=1}^n a_i\Big) \quad \in \mathbb{Z}/p\mathbb{Z},$$

where $a_i \in \mathbb{Z}$ is a representative of $\pi_i \in \mathbb{Z}/p\mathbb{Z}$.

One can show that $H_p(\boldsymbol{\pi})$ is independent of the choice of representatives a_i: it depends only on $\boldsymbol{\pi}$.

The definition, more explicitly

Substituting the definition of ∂_p into the definition of H_p gives a direct formula:

$$H_p(\boldsymbol{\pi}) = \tfrac{1}{p}\left(1 - \sum a_i^p\right) \quad \in \mathbb{Z}/p\mathbb{Z} \qquad (\boldsymbol{\pi} \in \Pi_n)$$

where again, $a_i \in \mathbb{Z}$ represents $\pi_i \in \mathbb{Z}/p\mathbb{Z}$.

Examples

- Let $p = 7$. Then $(2, 2, 4) \in \Pi_3$, and

$$H_7(2, 2, 4) = \tfrac{1}{7}\left(1 - [2^7 + 2^7 + 4^7]\right) \equiv 3 \pmod 7.$$

- Take any prime p and integer n with $p \nmid n$. Then $(1/n, \ldots, 1/n) \in \Pi_n$. We have $H_p(1/n, \ldots, 1/n) = q_p(n)$, just like $H_{\mathbb{R}}(1/n, \ldots, 1/n) = \log n$.
- The case $n = 2$: one can show that

$$H_p(\pi, 1 - \pi) = \sum_{0<r<n} \frac{\pi^r}{r}$$

  for all $\pi \in \mathbb{Z}/p\mathbb{Z}$ (assuming $p \neq 2$).
  There are no representatives a_i in this formula!

3. The characterization theorem

Or: *Why is this the right definition?*

The characterization thoerem

"Probability distributions" mod p can be composed like real ones: given

$$\boldsymbol{\pi} \in \Pi_n, \quad \boldsymbol{\gamma}_1 \in \Pi_{k_1}, \ldots, \boldsymbol{\gamma}_n \in \Pi_{k_n},$$

we get a composite

$$\boldsymbol{\pi} \circ (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_n) = (\pi_1\gamma_1^1, \ldots, \pi_1\gamma_1^{k_1}, \ \ldots, \ \pi_n\gamma_n^1, \ldots, \pi_n\gamma_n^{k_n}) \in \Pi_{k_1+\cdots+k_n}.$$

Entropy mod p obeys the same recursivity rule as real entropy:

$$H_p(\boldsymbol{\pi} \circ (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_n)) = H_p(\boldsymbol{\pi}) + \sum_i \pi_i H_p(\boldsymbol{\gamma}_i).$$

Theorem *H_p is the only sequence of functions $(\Pi_n \to \mathbb{Z}/p\mathbb{Z})_{n\geq 0}$ obeying this recursivity rule, up to a constant factor.*

This is exactly like the characterization theorem for real entropy, except without continuity. *So, we've got the right definition!*

4. Residues

Or: *What is the residue mod* 7 *of* $\log\sqrt{8}$*?*

Where this all started

The idea of "entropy mod p" was proposed by Maxim Kontsevich in 1995 in a $2\frac{1}{2}$-page unpublished note.

This talk elaborates on his ideas.

Kontsevich wrote:

> If we have a random variable ξ which takes finitely many values with all probabilities in $\mathbb{Q}$ then we can define not only the transcendental number $H(\xi)$ but also its "residues modulo p" for almost all primes p!

What did he have in mind? Maybe the following. . .

Kontsevich's residue proposal

Take a probability distribution with rational probabilities:

$$\mathbf{r} = \left(\frac{s_1}{t}, \ldots, \frac{s_n}{t}\right).$$

- On the one hand, $\mathbf{r}$ has a real entropy $H_{\mathbb{R}}(\mathbf{r})$, typically transcendental.
- On the other, for all primes p except the factors of t, we can interpret $\mathbf{r}$ as a "probability distribution" with probabilities in $\mathbb{Z}/p\mathbb{Z}$, so it has an entropy $H_p(\mathbf{r}) \in \mathbb{Z}/p\mathbb{Z}$.
- Kontsevich says: view $H_p(\mathbf{r}) \in \mathbb{Z}/p\mathbb{Z}$ as the residue class mod p of $H_{\mathbb{R}}(\mathbf{r}) \in \mathbb{R}$!

Does this proposal make sense?

Kontsevich's suggestion only makes sense if

$$H_{\mathbb{R}}(\mathbf{q}) = H_{\mathbb{R}}(\mathbf{r}) \Rightarrow H_p(\mathbf{q}) = H_p(\mathbf{r})$$

whenever $\mathbf{q}, \mathbf{r}$ are rational probability distributions with denominators not divisible by p.

Theorem *This is true.*

So, there's a well-defined residue map $x \mapsto [x]$ from

$$\{\text{real numbers arising as the entropy of a rational distribution with denominators not divisible by } p\}$$

to $\mathbb{Z}/p\mathbb{Z}$.

Moreover, it's additive: $[x + y] = [x] + [y]$, as a "residue" should be.

Example

Take the rational distribution $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$.

It has real entropy

$$-\left(\frac{1}{4}\log\frac{1}{4} + \frac{1}{4}\log\frac{1}{4} + \frac{1}{2}\log\frac{1}{2}\right) = \log\sqrt{8}.$$

For any $p \neq 2$, we can calculate $H_p(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}) \in \mathbb{Z}/p\mathbb{Z}$, and this is the residue mod p of $\log\sqrt{8}$.

Example Let $p = 7$. Then $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}) = (2, 2, 4)$, and $H_7(2, 2, 4) = 3$ (calculated earlier).

So, the residue mod 7 of $\log\sqrt{8}$ is 3.

5. Entropy as a polynomial

Or: *Looking over the horizon*

Entropy as a polynomial

Our "direct" definition of entropy mod p wasn't so direct...

$$H_p(\boldsymbol{\pi}) = \tfrac{1}{p}\left(1 - \sum a_i^p\right)$$

...as we had to choose integers a_i representing π_i.

But an equivalent definition expresses $H_p(\boldsymbol{\pi})$ as a (ferocious) polynomial in $\pi_1, \ldots, \pi_n$ themselves:

$$H_p(\boldsymbol{\pi}) = - \sum_{\substack{0 \leq j_1, \ldots, j_n < p \\ j_1 + \cdots + j_n = p}} \frac{\pi_1^{j_1} \cdots \pi_n^{j_n}}{j_1! \cdots j_n!}.$$

Write $h(\pi_1, \ldots, \pi_n)$ for the polynomial on the right-hand side.

Then h satisfies a 2-cocycle condition:

$$h(x, y) - h(x, y + z) + h(x + y, z) - h(y, z) = 0.$$

This leads into deep waters, such as "information cohomology"...

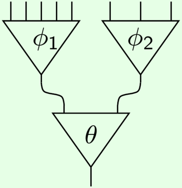
Open questions

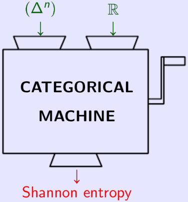
Open questions

- We've seen how to define entropy over $\mathbb{R}$ and also over $\mathbb{Z}/p\mathbb{Z}$. What about other fields, such as $\mathbb{Q}_p$? Is there a single unified theory?
- What does entropy mod p "mean"? Can we develop any intuition for it? Are there any applications?
- Do other entropic quantities—such as relative entropy, conditional entropy, mutual information, or Rényi entropies—have analogues mod p?

Three talks

Sunday: Operads

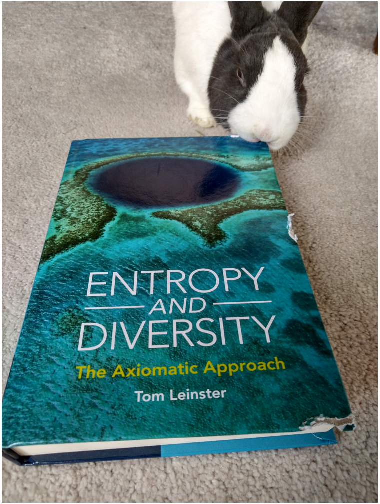




Monday: An algebraic view of entropy

Thursday: Entropy modulo a prime

Main reference for all three talks



. . . and citations therein.

Thank you very much for listening