Intermediate category theory reading group: *An Operadic Tale of Entropy*

Speaker: Emmanouil Sfinarolakis
Notetaker: Maia Woolf*

2025-11-26 (week 8)

Algebraic information theory

In particular today entropy, using operads

Quantifier is not in the logic sense

Entropy/complexity

Goal: quantify "surprise" of event x

Both sides of coin heads should give no surprise

Educated guess: surprise is reciprocal of probability

But should have surprise 0 for probability 1

So can't be reciprocal $\rightsquigarrow$ fix this by taking logorithm.

Look at limit $\rightsquigarrow$ fix this by multiplying by probability

Requirements for entropy

Continuity: distribution deformed by a bit (e.g. coin landing on edge)

Symmetry: names of outcomes shouldn't affect complexity

Maximality: fair distribution (uniform) maximises complexity

Monotonicity: more outcomes means more complexity

Expansibility: adding an outcome with probability zero does not change complexity. The system practically has not changed at all.

Disintegration (physicists might call this the grouping axiom or chain rule (it has a lot of different names)): combining macroscopic distribution and a family of microscopic distributions $\rightsquigarrow$ formula for calculating overall complexity

Uniqueness of entropy functions theorems: Khinchin, Faddeev, Leinster. Each showed that in a certain situation respectively, Shannon entropy is the only entropy function. Each used slightly different axioms, trying to find which axioms are the 'core' important ones. Tom Leinster managed to do everything in the language of category theory.

Information

Very similar requirements to entropy.

Shannon showed that Shannon information is the only function that satisfies certain axioms. This function is strongly related to the Shannon entropy function.

Theorem: Shannon entropy measures the optimal average number of binary questions needed to identify an element from a distribution.

*Email: `maia.woolf@ed.ac.uk`

Recall: derivation theory. Derivation = linear map + Leibniz's rule (aka. chain rule). This is where the 'algebraic' part of algebraic information theory appears.

In what sense, in what framework, under what conditions is entropy a derivation?

Definition: topological n-simplex. Thinking about our probability distributions as points in $\mathbb{R}^n$. In this presentation, the point is the 1-simplex Δ^1 (as opposed to Δ^0), the line is Δ^2, etc.

What does composing probability distributions mean in this geometric sense? Example: unfair coin, composing with (doing nothing, a fair 6-sided die).

Can do *pointwise evaluation*.

Almost derivation. Need to consider an operad of functions.

Operads

Think of n-ary operations as planar rooted trees with n leaves

Think of composition as grafting tree on ith leaf. Need to specify which leaf we are using for the composition.

We have a collection of operations, with composition. We have a 1-ary identity of composition. We have 3 associativity axioms, differing based on where the three trees are in relation to each other. These give us a *non-symmetric operad*.

We have an operad $\Delta = (\Delta^1, \Delta^2, \dots)$ of topological simplices. Composition is as we saw before for topological spaces as probability spaces.

Endomorphism operad $\mathrm{End}_{\mathbb{R}}$.

Want to connect the above two operads in order to make sense of Shannon entropy. To do this, define some *action maps* $\varphi_n : \Delta^n \to \mathrm{End}_{\mathbb{R}}(n), P \mapsto \langle P, \bullet \rangle$, where $\varphi_n(P)(x) := \langle P, x \rangle = \sum_{i=1}^n P_i x_i$. $\mathrm{End}_{\mathbb{R}}$ takes a probability distribution and a point and gives us a number.

Goal: define map d out of Δ obeying Leibnitz's rule. There are some key observations here. Will define a left and a right composition. Need an abelian bimodule $(\mathrm{End}_{\mathbb{R}}, +, \circ_i^R, \circ_i^L)$ over Δ.

These compositions are very weird! Think of right composition as a local weighed sum, and take a convex combination of the coordinates. Think of left composition as forgetting the irrelevant leaves from the probability distribution.

Now we can define a derivation of Δ in the bimodule $\mathrm{End}_{\mathbb{R}}$ over Δ.

Now we can find a derivation of simultaneous composition, via iterated application.

Final theorem for today: Shannon entropy is a derivation $d : \Delta \to \mathrm{End}_{\mathbb{R}}, P \mapsto dP, dP(x) \equiv H(P)$ of the operad Δ of topological simplices. Furthermore (and this is powerful), every derivation on the operad Δ, $d : \Delta \to \mathrm{End}_{\mathbb{R}}$, is a scalar multiple of the Shannon entropy (only!) when evaluated at the origin: $dP(0) = \alpha H(P)$.

Entropy is lots of things! Including some unexpected things, such as a functional over geometric objects. Most importantly: entropy is a derivation of the operad of topological simplices.