An operadic characterization of Shannon entropy

Tom Leinster*

Rough notes, 30 April 2011

This is a recasting of Fadeev's theorem, as stated in Rényi's paper 'On measures of entropy and information'.

Plan: first I give the definition of lax map of operad algebras. This is a standard 2-categorical definition. Then I spend a while unwinding the definition. Finally, I show that when applied in a particular situation, the axioms in the definition are almost the same as Fadeev's axioms for entropy. This gives the operadic characterization of Shannon entropy.

This result arose from conversations with Steve Lack.

1 The definition of lax map

Let $\mathscr{E}$ be a category with finite limits and P be a symmetric operad in $\mathscr{E}$. Let X and Y be (strict) P-algebras in $\mathbf{Cat}(\mathscr{E})$. A **lax map** $(f, \phi) : X \to Y$ is a functor $f : X \to Y$ together with a natural transformation

$$\begin{tikzcd}
P_n \times X^n \arrow[r, "1\times f^n"] \arrow[d] \arrow[dr, phantom, "\swArrow \phi_n"] & P_n \times Y^n \arrow[d] \\
X \arrow[r, "f"'] & Y
\end{tikzcd}$$

for each $n \in \mathbb{N}$, satisfying the following three axioms. In all the diagrams, it's supposed to be obvious what the vertical arrows are.

i. For all $n, k_1, \ldots, k_n \in \mathbb{N}$, writing $k = \sum k_i$,

$$\begin{tikzcd}
P_n \times P_{k_1} \times X^{k_1} \times \cdots \times P_{k_n} \times X^{k_n} \arrow[r, "1\times 1\times f^{k_1}\times\cdots"] \arrow[d] \arrow[dr, phantom, "/\!/"] & P_n \times P_{k_1} \times X^{k_1} \times \cdots \times P_{k_n} \times Y^{k_n} \arrow[d] \\
P_k \times X^k \arrow[r, "1\times f^k"] \arrow[d] \arrow[dr, phantom, "\swArrow \phi_k"] & P_k \times Y^k \arrow[d] \\
X \arrow[r, "f"'] & Y
\end{tikzcd}$$

$$= \begin{tikzcd}
P_n \times P_{k_1} \times X^{k_1} \times \cdots \times P_{k_n} \times X^{k_n} \arrow[r, "1\times 1\times f^{k_1}\times\cdots"] \arrow[d] \arrow[dr, phantom, "\swArrow 1 \times \phi_{k_1} \times \cdots \times \phi_{k_n}"] & P_n \times P_{k_1} \times X^{k_1} \times \cdots \times P_{k_n} \times Y^{k_n} \arrow[d] \\
P_n \times X^n \arrow[r, "1\times f^n"] \arrow[d] \arrow[dr, phantom, "\swArrow \phi_n"] & P_n \times Y^n \arrow[d] \\
X \arrow[r, "f"'] & Y.
\end{tikzcd}$$

*School of Mathematics and Statistics, University of Glasgow, Glasgow G12 8QW, UK; Tom.Leinster@glasgow.ac.uk. Supported by an EPSRC Advanced Research Fellowship.

ii. The composite natural transformation

$$\begin{tikzcd}
X \arrow[r, "f"] \arrow[d] \arrow[dr, phantom, "\sslash"] & Y \arrow[d] \\
P_1 \times X^1 \arrow[r, "1\times f"] \arrow[d] \arrow[dr, phantom, "\Swarrow \phi_1"] & P_1 \times Y^1 \arrow[d] \\
X \arrow[r, "f"'] & Y
\end{tikzcd}$$

is the identity.

iii. For each $n \in \mathbb{N}$ and permutation $\sigma \in S_n$,

$$\begin{tikzcd}
P_n \times X^n \arrow[r, "1\times f^n"] \arrow[d, "\sigma_* \times 1"'] \arrow[dr, phantom, "\sslash"] & P_n \times Y^n \arrow[d, "\sigma_* \times 1"] \\
P_n \times X^n \arrow[r, "1\times f^n"] \arrow[d] \arrow[dr, phantom, "\Swarrow \phi_n"] & P_n \times Y^n \arrow[d] \\
X \arrow[r, "f"'] & Y
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
P_n \times X^n \arrow[r, "1\times f^n"] \arrow[d, "1 \times \sigma^*"'] \arrow[dr, phantom, "\sslash"] & P_n \times Y^n \arrow[d, "1 \times \sigma^*"] \\
P_n \times X^n \arrow[r, "1\times f^n"] \arrow[d] \arrow[dr, phantom, "\Swarrow \phi_n"] & P_n \times Y^n \arrow[d] \\
X \arrow[r, "f"'] & Y.
\end{tikzcd}$$

2 Unwinding the definition

There are other ways of phrasing this. If $\mathscr{E}$ has well-behaved colimits (e.g. if it's **Set** or **Top**) then the operad P induces a 2-monad on $\mathbf{Cat}(\mathscr{E})$. An algebra for this 2-monad is precisely a P-algebra in $\mathbf{Cat}(\mathscr{E})$, and a lax map of algebras for this 2-monad is precisely a lax map in the sense above.

If $\mathscr{E} = \mathbf{Set}$ then we can also say it explicitly. The algebra structure on X gives us for each operation $p \in P_n$ a functor $\overline{p} : X^n \to X$; similarly for Y. The natural transformation ϕ gives us, for each operation $p \in P_n$ and objects $x_1, \ldots, x_n \in X$, a map

$$\phi = \phi_{p;x_1,\ldots,x_n} : \overline{p}(f(x_1), \ldots, f(x_n)) \to f(\overline{p}(x_1, \ldots, x_n))$$

in the category Y. This is to satisfy naturality in $x_1, \ldots, x_n$ and the three axioms above. Axiom (i) says that the composite

$$\begin{aligned}
\overline{p}\big(\overline{p_1}(f(x_1^1), \ldots, f(x_1^{k_1})), \ldots, \overline{p_n}(f(x_n^1), \ldots, f(x_n^{k_n}))\big) &\xrightarrow{\overline{p}(\phi,\ldots,\phi)} \overline{p}\big(f(\overline{p_1}(x_1^1, \ldots, x_1^{k_1})), \ldots, f(\overline{p_n}(x_n^1, \ldots, x_n^{k_n}))\big) \\
&\xrightarrow{\phi} f\big(\overline{p}(\overline{p_1}(x_1^1, \ldots, x_1^{k_1}), \ldots, \overline{p_n}(x_n^1, \ldots, x_n^{k_n}))\big)
\end{aligned}$$

is equal to

$$\overline{p \circ (p_1, \ldots, p_n)}(f(x_1^1), \ \ldots \ , f(x_n^{k_n})) \xrightarrow{\phi} f(\overline{p \circ (p_1, \ldots, p_n)}(x_1^1, \ \ldots \ , x_n^{k_n})).$$

Axiom (ii) says that $\phi : \overline{1_P}(f(x)) \to f(\overline{1_P}(x))$ is the identity on $f(x)$. Axiom (iii) says that

$$\phi : \overline{\sigma p}(f(x_1), \ldots, f(x_n)) \to f(\overline{\sigma p}(x_1, \ldots, x_n))$$

is equal to

$$\phi : \overline{p}(f(x_{\sigma 1}), \ldots, f(x_{\sigma n})) \to f(\overline{p}(x_{\sigma 1}, \ldots, x_{\sigma n})).$$

When $\mathscr{E} = \mathbf{Top}$, the conditions on ϕ are the same except that $\phi_{p;x_1,\ldots,x_n}$ must also vary continuously (in the space of arrows of Y) with p and $x_1, \ldots, x_n$.

The terminal category 1 has a unique P-algebra structure. A **lax point** of a P-algebra Y in $\mathbf{Cat}(\mathscr{E})$ is a lax map $1 \to Y$. (These are also sometimes called internal P-algebras in Y: e.g. consider the case where P is the terminal operad.)

From now on, assume that $\mathscr{E}$ is **Set** or **Top**. A lax point of Y consists of a functor $f : 1 \to Y$, that is, an object $y \in Y$, together with a map

$$\phi = \phi_p : \overline{p}(y, \ldots, y) \to y$$

for each operation $p \in P_n$, satisfying axioms. The naturality axiom is always satisfied, trivially. Axiom (i) is that the composite

$$\overline{p}(\overline{p_1}(y, \ldots, y), \ldots, \overline{p_n}(y, \ldots, y)) \xrightarrow{\overline{p}(\phi,\ldots,\phi)} \overline{p}(y, \ldots, y) \xrightarrow{\phi} y$$

is equal to

$$\overline{p \circ (p_1, \ldots, p_n)}(y, \ \ldots \ , y) \xrightarrow{\phi} y.$$

Axiom (ii) is that $\phi : \overline{1_P}(y) \to y$ is the identity. Axiom (iii) is that $\phi_{\sigma p} = \phi_p$. When $\mathscr{E} = \mathbf{Top}$, we have the further axiom that ϕ_p varies continuously with p.

Suppose further that Y is a monoid $M = (M, \cdot, 1)$ in $\mathscr{E}$, viewed as a one-object category. For $(M, \cdot, 1)$ to be a P-algebra in $\mathbf{Cat}(\mathscr{E})$ means that the object M of $\mathscr{E}$ is a P-algebra and that each map $\overline{p} : M^n \to M$ is a monoid homomorphism. A lax point of M consists of an element $\phi_p \in M$ for each $n \in \mathbb{N}$ and $p \in P_n$, satisfying axioms. I will now write ϕ_p as $\phi(p)$. Axiom (i) is that

$$\phi(p \circ (p_1, \ldots, p_n)) = \phi(p) \cdot \overline{p}(\phi(p_1), \ldots, \phi(p_n)).$$

Axiom (ii) is that $\phi(1_P) = 1$. Axiom (iii) is that $\phi(\sigma p) = \phi(p)$. If $\mathscr{E} = \mathbf{Top}$, we have the further axiom that $\phi : P_n \to M$ is continuous for each $n \in \mathbb{N}$.

3 The result

Now take $\mathscr{E} = \mathbf{Top}$, and take P to be the operad Δ in which Δ_n is the standard topological $(n-1)$-simplex

$$\{p \in [0, \infty)^n \mid \sum p_i = 1\}.$$

The operadic composition is as follows: given $p \in \Delta_n, r_1 \in \Delta_{k_1}, \ldots, r_n \in \Delta_{k_n}$,

$$p \circ (r_1, \ldots, r_n) = (p_1 r_1^1, \ldots, p_1 r_1^{k_1}, \ \ldots, \ p_n r_n^1, \ldots, p_n r_n^{k_n}).$$

Like any convex subset of $\mathbb{R}^n$, the space $\mathbb{R}$ is naturally a Δ-algebra in **Top**. For each $p \in \Delta_n$, the map

$$\begin{array}{rccc} \overline{p}: & \mathbb{R}^n & \to & \mathbb{R} \\ & x & \mapsto & \sum_i p_i x_i \end{array}$$

is additive. Hence the monoid $(\mathbb{R}, +, 0)$ is a (one-object) P-algebra in $\mathbf{Cat}(\mathbf{Top})$, which I will just call $\mathbb{R}$.

A lax point of $\mathbb{R}$ consists of a function $H : \Delta_n \to \mathbb{R}$ for each $n \in \mathbb{N}$, satisfying the following axioms. Axiom (i) is that

$$H(p \circ (r_1, \ldots, r_n)) = H(p) + \overline{p}(H(r_1), \ldots, H(r_n))$$

$(p \in \Delta_n, r_i \in \Delta_{k_i})$, that is,

$$H(p \circ (r_1, \ldots, r_n)) = H(p) + \sum_i p_i H(r_i).$$

Axiom (ii) is that $H(1) = 0$, where the '(1)' on the left-hand side is the unique element of Δ_1. Axiom (iii) is that H is symmetric in its arguments. We also have the axiom that for each n, the function $H : \Delta_n \to \mathbb{R}$ is continuous.

A special case of axiom (i) is where $r_1 = (t, 1-t)$ and $r_2 = \cdots = r_n = (1)$; this is then one of the Fadeev axioms. Symmetry and continuity are also on Fadeev's list. So our axioms imply all the Fadeev axioms except normalization ($H(1/2, 1/2) = 1$). Conversely, the Fadeev axioms except normalization imply ours: axiom (i) follows by induction from the special case just mentioned, and axiom (ii) is redundant (by taking $n = k_1 = 1$ and $p = r_1 = (1)$ in axiom (i)).

Fadeev's theorem therefore states that the lax points of $\mathbb{R}$ are exactly the real multiples of Shannon entropy.