Operads and Entropy

Tai-Danae Bradley

Review

Given a probability distribution p on a set $X = \{1, \ldots, n\}$, the **Shannon entropy of** p is given by

$$H(p) = -\sum_{i=1}^{n} p_i \ln(p_i)$$

where $p_i := p(i)$. Intuitively, $H(p)$ is a measure of "surprise."

Here's a slightly tidier way to write $H(p)$.

Define a function $d\colon [0,1] \to \mathbb{R}$ by

$$d(a) = \begin{cases} -a\ln(a) & \text{if } a > 0, \\ 0 & \text{if } a = 0. \end{cases}$$

We can use d to rewrite entropy as

$$H(p) = \sum_{i=1}^{n} d(p_i).$$

Define a function $d\colon [0,1] \to \mathbb{R}$ by

$$d(a) = \begin{cases} -a\ln(a) & \text{if } a > 0, \\ 0 & \text{if } a = 0. \end{cases}$$

We can use d to rewrite entropy as

$$H(p) = \sum_{i=1}^{n} d(p_i).$$

It's easy to check d is a nonlinear **derivation**:

$$d(ab) = d(a)b + ad(b) \qquad \text{for all } a, b \in [0,1].$$

Is entropy itself equal to "d of something"? That is, do we have the following?

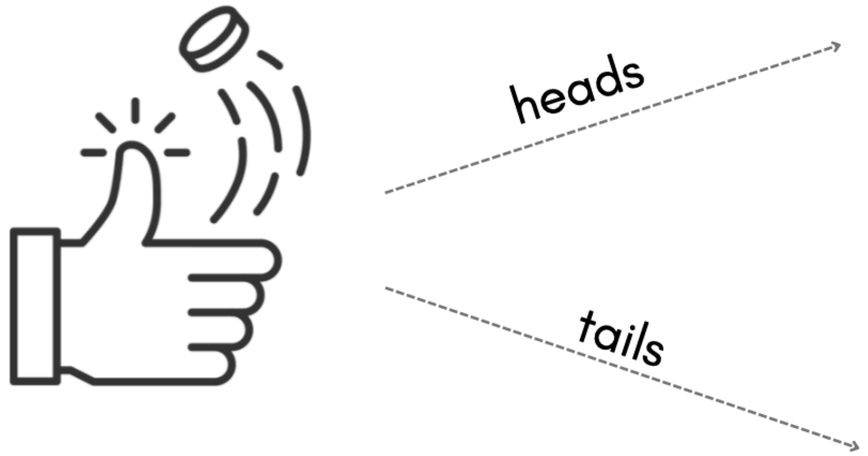
$$H(p) \stackrel{?}{=} d\left(\sum_{i=1}^{n} p_i\right)$$

Is entropy itself equal to "d of something"? That is, do we have the following?

$$H(p) \stackrel{?}{=} d\left(\sum_{i=1}^{n} p_i\right)$$

No, d is not linear. Still, we may wonder if there is something "more" going on. Let's look at some other facets of entropy.

The Chain Rule



$$p = \left(\frac{1}{2}, \frac{1}{2}\right)$$

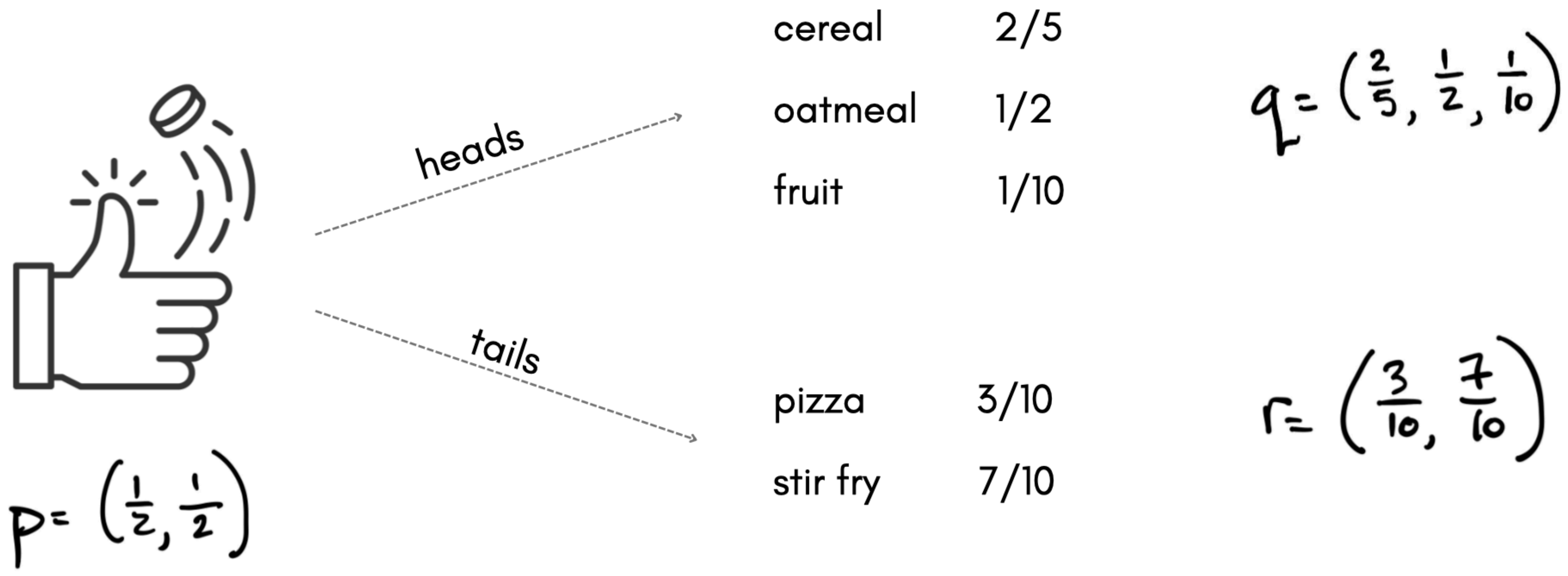
cereal	2/5
oatmeal	1/2
fruit	1/10

$$q = \left(\frac{2}{5}, \frac{1}{2}, \frac{1}{10}\right)$$

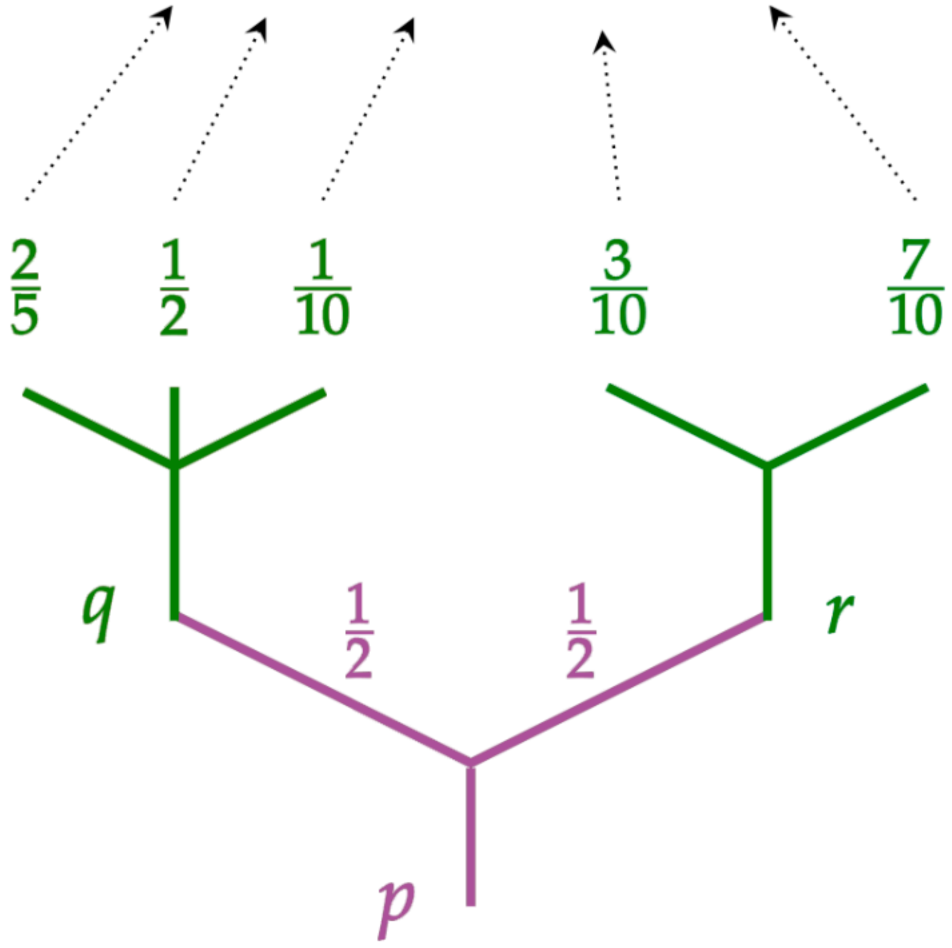
pizza	3/10
stir fry	7/10

$$r = \left(\frac{3}{10}, \frac{7}{10}\right)$$

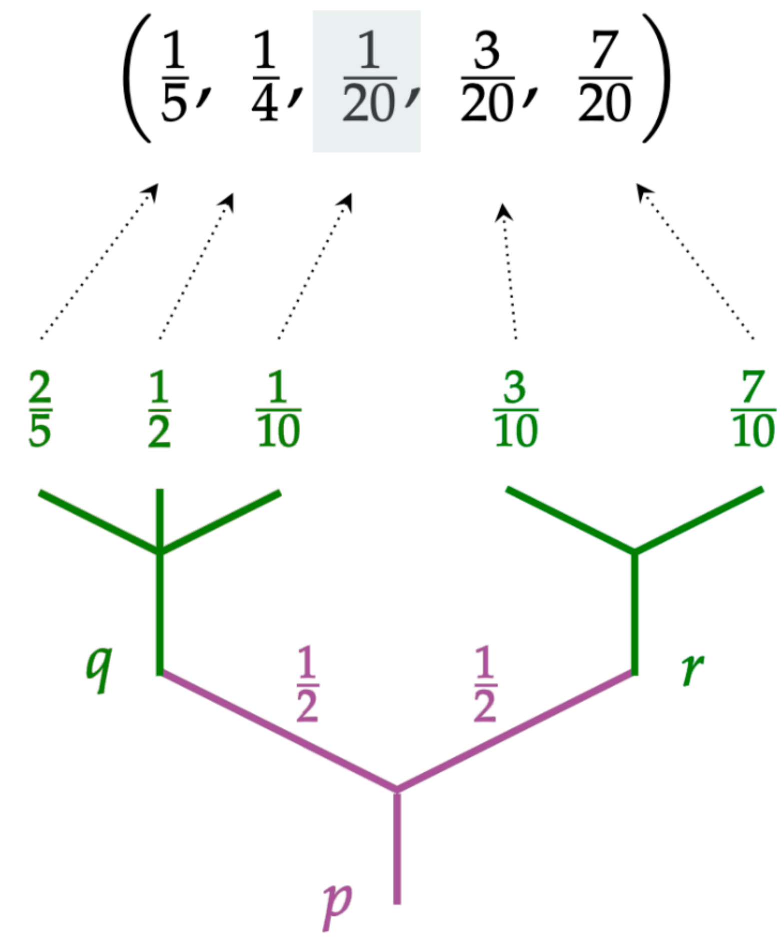
Suppose we flip a fair coin, then choose a meal for breakfast or dinner. This two-step process defines a probability distribution on five food options. For example, the probability of flipping **heads** and choosing **fruit** is $\left(\frac{1}{2}\right)\left(\frac{1}{10}\right) = \frac{1}{20}$.



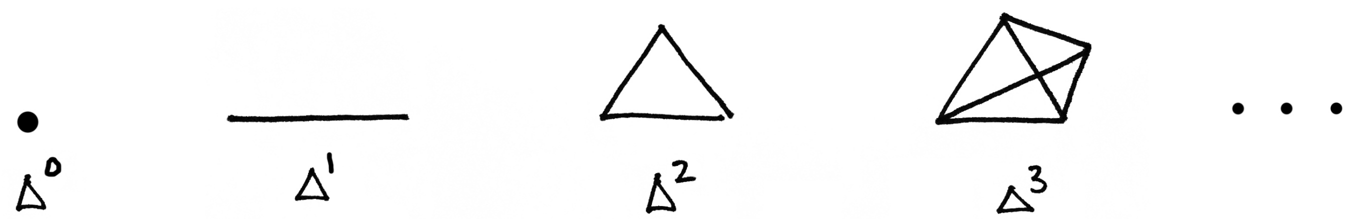
$$\left(\frac{1}{5}, \frac{1}{4}, \frac{1}{20}, \frac{3}{20}, \frac{7}{20}\right)$$

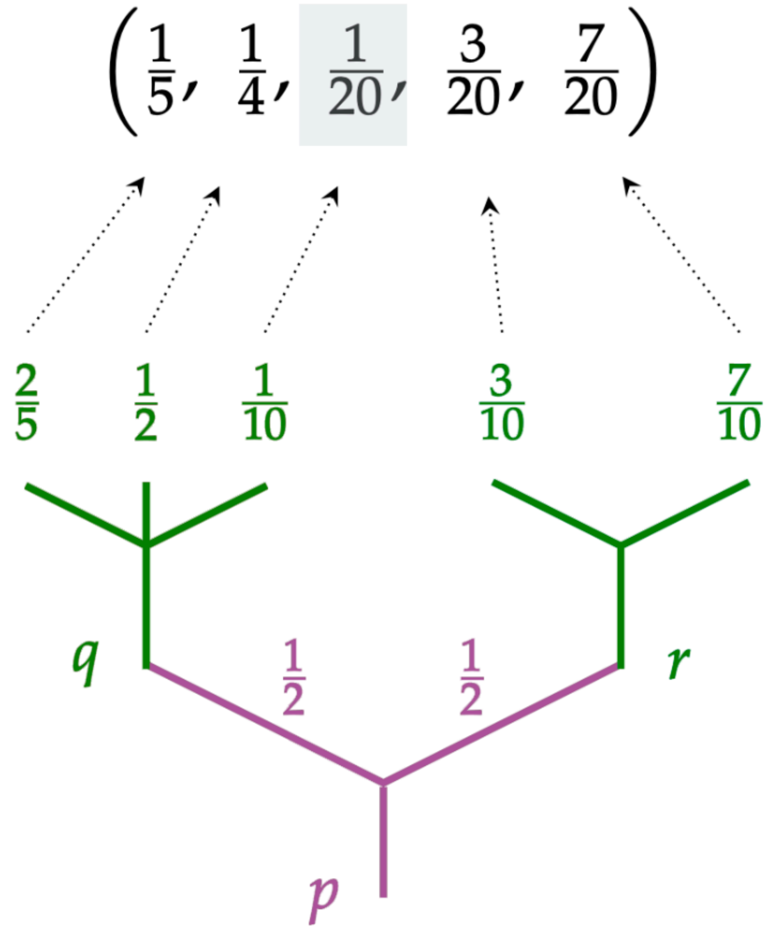


$$p \circ (q, r)$$

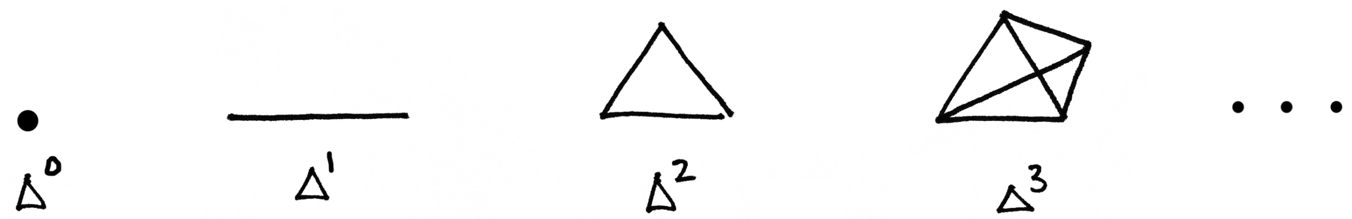


A probability distribution $p = (p_1, \ldots, p_n)$ is a point the topological simplex $\Delta^{n-1} \subseteq \mathbb{R}^n$. Let's reindex and write $\Delta_n := \Delta^{n-1}$ to denote the space of probability distributions on n elements.



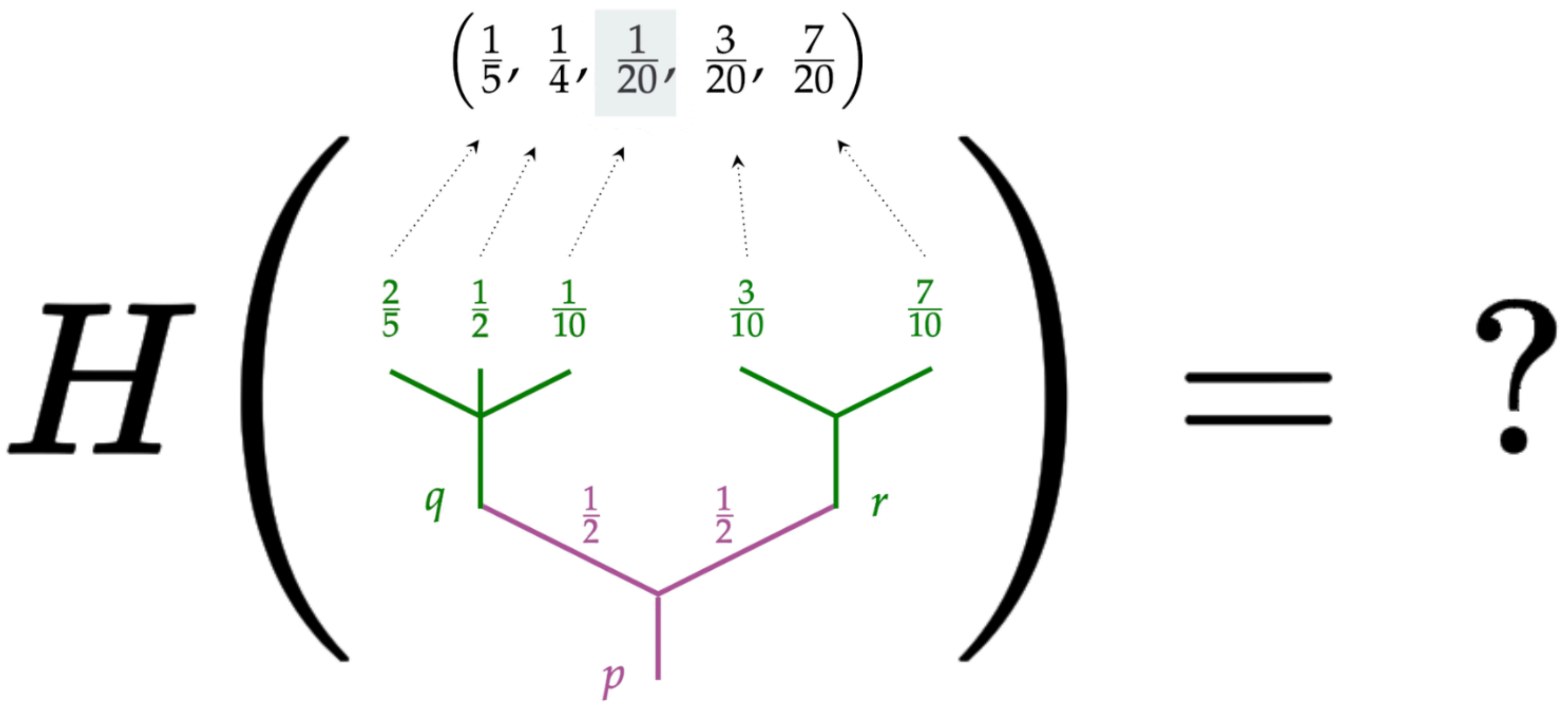


A probability distribution $p = (p_1, \ldots, p_n)$ is a point the topological simplex $\Delta^{n-1} \subseteq \mathbb{R}^n$. Let's reindex and write $\Delta_n := \Delta^{n-1}$ to denote the space of probability distributions on n elements.

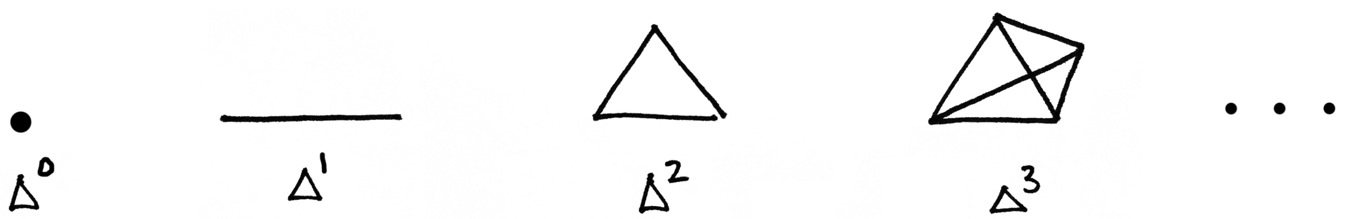


Multiplying probabilities gives us a **composition** function

$$\begin{aligned} \Delta_2 \times \Delta_3 \times \Delta_2 &\to \Delta_5 \\ (p, q, r) &\mapsto p \circ (q, r) \end{aligned}$$



A probability distribution $p = (p_1, \ldots, p_n)$ is a point the topological simplex $\Delta^{n-1} \subseteq \mathbb{R}^n$. Let's reindex and write $\Delta_n := \Delta^{n-1}$ to denote the space of probability distributions on n elements.



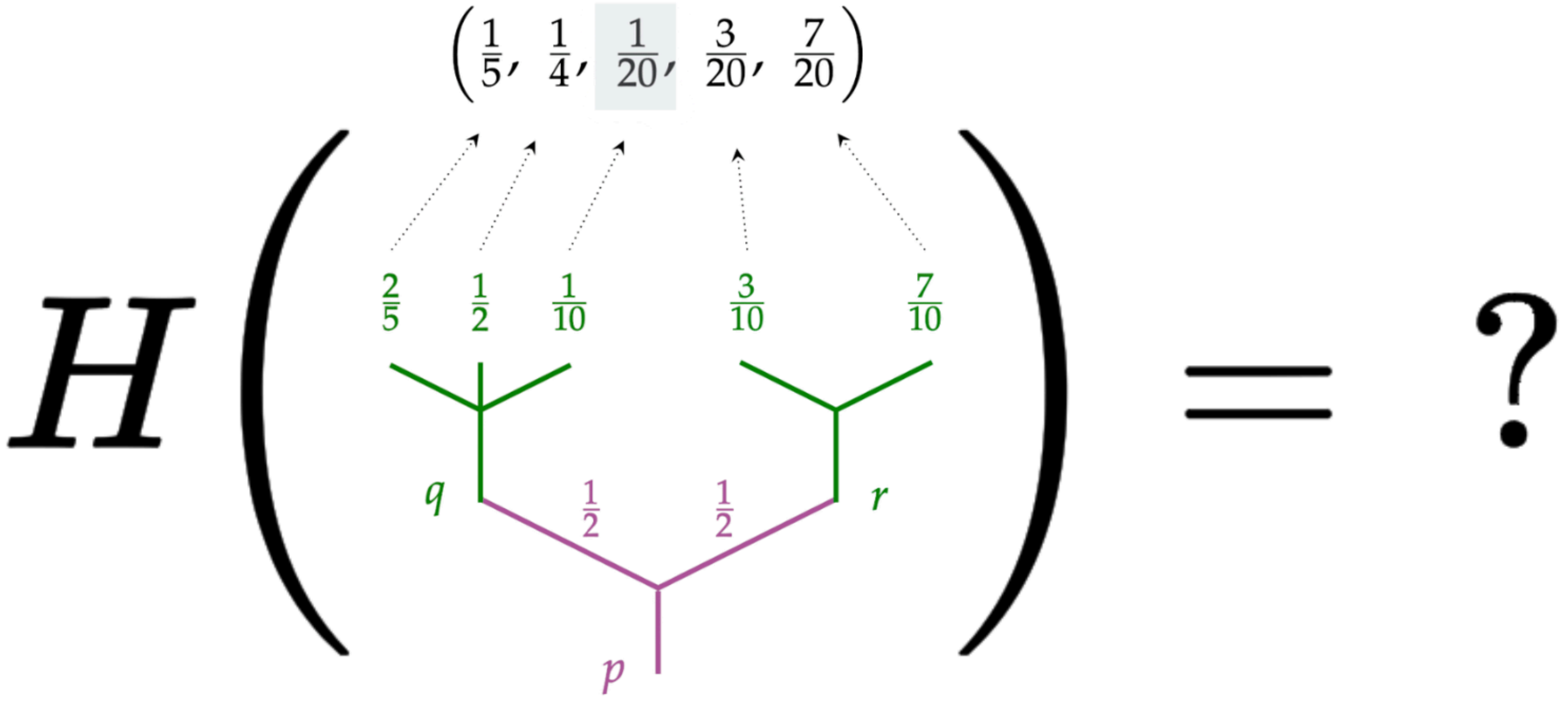
Multiplying probabilities gives us a **composition** function

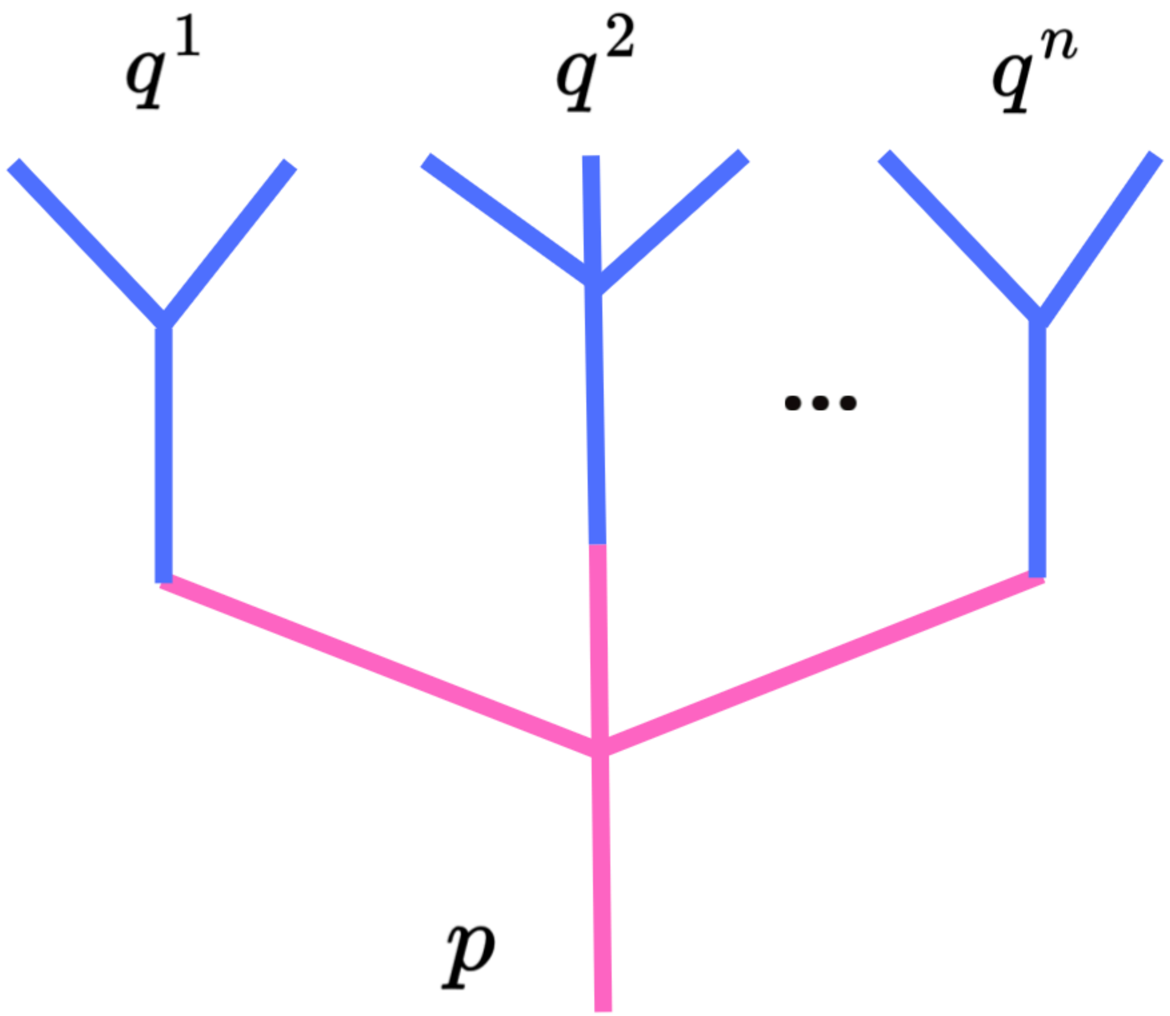
$$
\begin{aligned}
\Delta_2 \times \Delta_3 \times \Delta_2 &\to \Delta_5 \\
(p, q, r) &\mapsto p \circ (q, r)
\end{aligned}
$$

What is the Shannon entropy of this composite distribution $p \circ (q, r)$?

The chain rule,

$$H(p \circ (q, r)) = H(p) + \tfrac{1}{2} H(q) + \tfrac{1}{2} H(r).$$





The chain rule,

$$H(p \circ (q, r)) = H(p) + \tfrac{1}{2} H(q) + \tfrac{1}{2} H(r).$$

More generally, we can compose n probability distributions $q^1, \ldots, q^n$ with $p = (p_1, \ldots, p_n)$ to obtain a new distribution $p \circ (q^1, \ldots, q^n)$, whose entropy is equal to

$$H(p \circ (q^1, \ldots, q^n)) = H(p) + \sum_{i=1}^{n} p_i H(q^i).$$

Proof: Arithmetic. Recall that $H(p) = \sum_i d(p_i)$ where $d(p_i) = -p_i \ln(p_i)$ is a derivation.

Putting all of this together, Shannon entropy defines a *sequence* of *continuous* functions

$$(H \colon \Delta_n \to \mathbb{R})_{n \in \mathbb{N}}$$

Putting all of this together, Shannon entropy defines a *sequence* of *continuous* functions

$$(H\colon \Delta_n \to \mathbb{R})_{n\in\mathbb{N}}$$

that satisfies the chain rule:

$$H(p \circ (q^1, \ldots, q^n)) = H(p) + \sum_{i=1}^{n} p_i H(q^i).$$

Putting all of this together, Shannon entropy defines a *sequence* of *continuous* functions

$$(H\colon \Delta_n \to \mathbb{R})_{n\in\mathbb{N}}$$

that satisfies the chain rule:

$$H(p \circ (q^1, \ldots, q^n)) = H(p) + \sum_{i=1}^{n} p_i H(q^i).$$

$$H(\bullet\circ) = H(\bullet) + \bullet H(\circ)$$

Putting all of this together, Shannon entropy defines a *sequence* of *continuous* functions

$$(H\colon \Delta_n \to \mathbb{R})_{n\in\mathbb{N}}$$

that satisfies the chain rule:

$$H(p \circ (q^1, \ldots, q^n)) = H(p) + \sum_{i=1}^{n} p_i H(q^i).$$

$$H(\bullet\circ) = H(\bullet) + \bullet H(\circ)$$

$$H(\bullet\circ) = H(\bullet)\circ + \bullet\, H(\circ)$$

The chain rule is "the most important algebraic property of Shannon entropy."

In 2020, Tom Leinster proved that the chain rule together with continuity are enough to characterize entropy. This is a variant of a 1956 characterization of entropy by Dmitry Faddeev.

It's also used in the proof of an *operadic* characterization Leinster gave around 2010.[1]

Theorem 2.5.1 (Faddeev) *Let* $(I\colon \Delta_n \to \mathbb{R})_{n\geq 1}$ *be a sequence of functions. The following are equivalent:*

i. *the functions I are continuous and satisfy the chain rule*

$$I(\mathbf{w} \circ (\mathbf{p}^1, \ldots, \mathbf{p}^n)) = I(\mathbf{w}) + \sum_{i=1}^{n} w_i I(\mathbf{p}^i)$$

$(n, k_1, \ldots, k_n \geq 1,\ \mathbf{w} \in \Delta_n,\ \mathbf{p}^i \in \Delta_{k_i})$;

ii. $I = cH$ *for some* $c \in \mathbb{R}$.

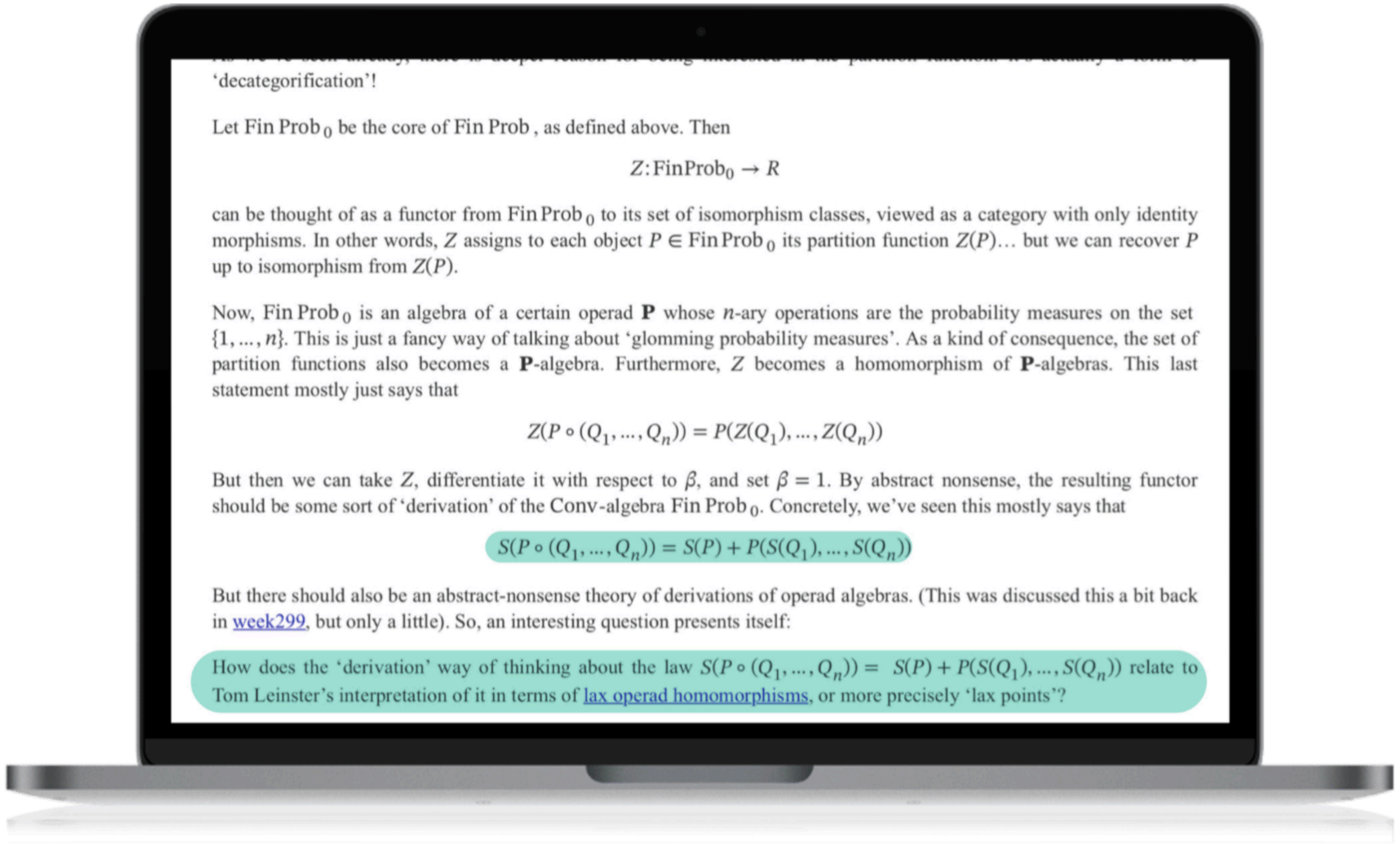
[1] Tom Leinster, *Entropy and Diversity* (Theorem 12.3.1), 2020.

The chain rule is "the most important algebraic property of Shannon entropy."

In 2020, Tom Leinster proved that the chain rule together with continuity are enough to characterize entropy. This is a variant of a 1956 characterization of entropy by Dmitry Faddeev.

It's also used in the proof of an *operadic* characterization Leinster gave around 2010.[1] At the same time, John Baez noticed the **chain rule for entropy looks like the Leibniz rule** and asked if the two perspectives were related.

- John Baez, "Entropy as a Functor" (2010) www.ncatlab.org/johnbaez/show/Entropy+as+a+functor



[1] Tom Leinster, *Entropy and Diversity* (Theorem 12.3.1), 2020.

Yes, the chain rule is the Leibniz rule in disguise.

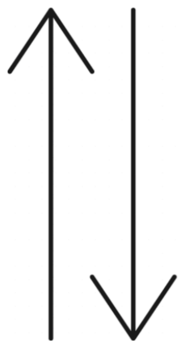
There is a correspondence between Shannon entropy and derivations of a certain operad.

- B., "Entropy as a Topological Operad Derivation," *Entropy* (2021)

- What is an operad?
- What is a derivation of one?
- What is the correspondence?
- Why care?

Here are *some* of the highlights.

entropy

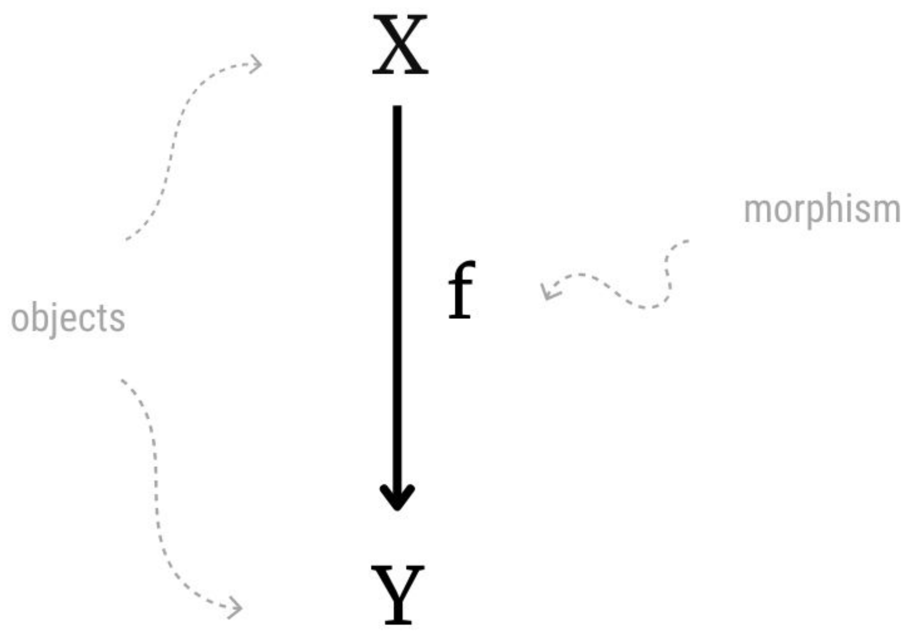


derivations

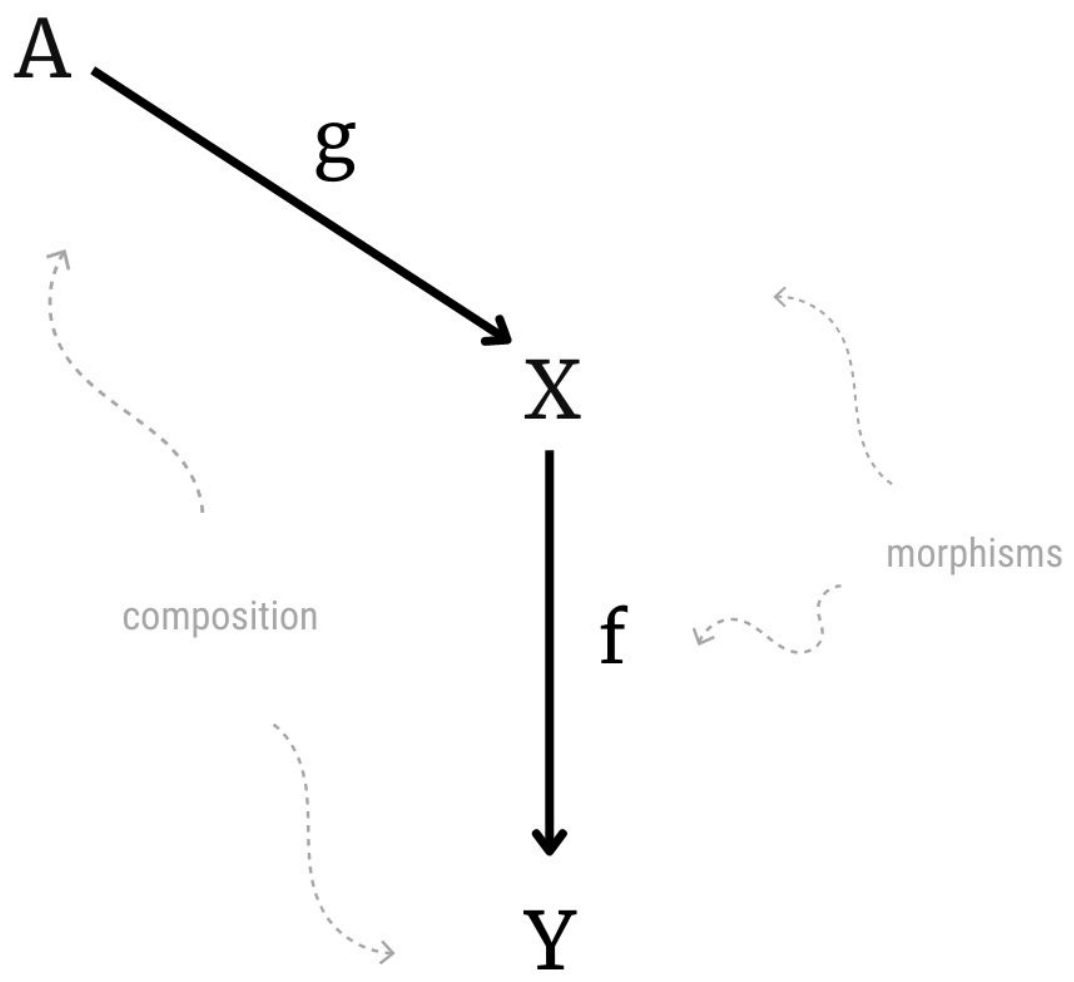
What is an operad?

Let's start with categories.

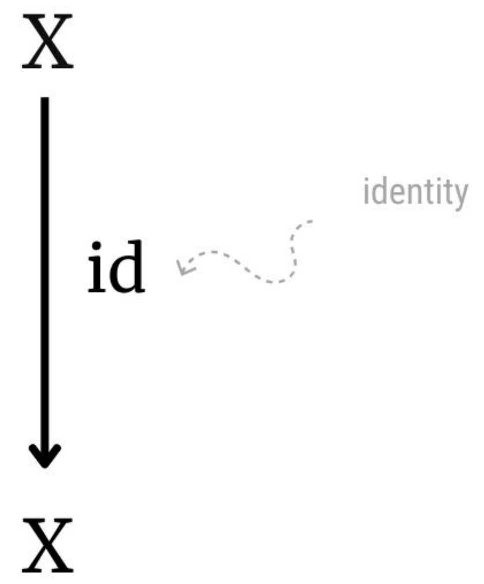
A category consists of objects and morphisms



A category consists of objects and morphisms, together with composition



A category consists of objects and morphisms, together with composition and identity morphisms.

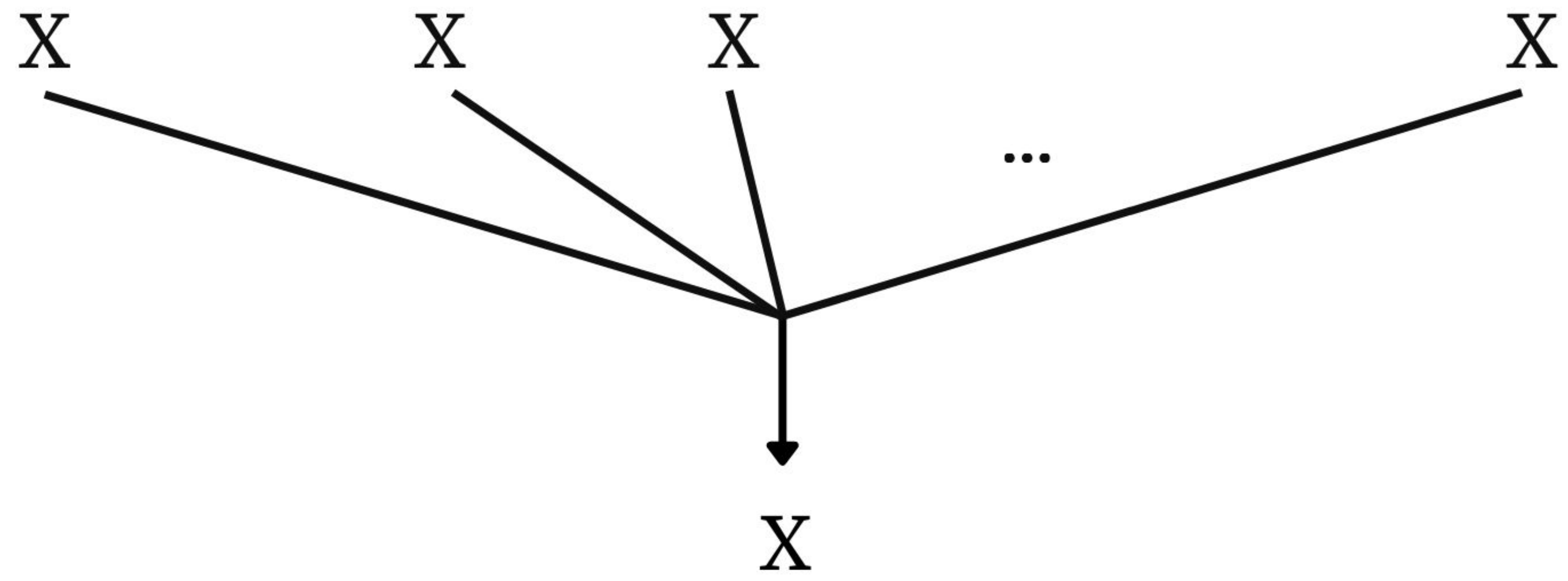


X

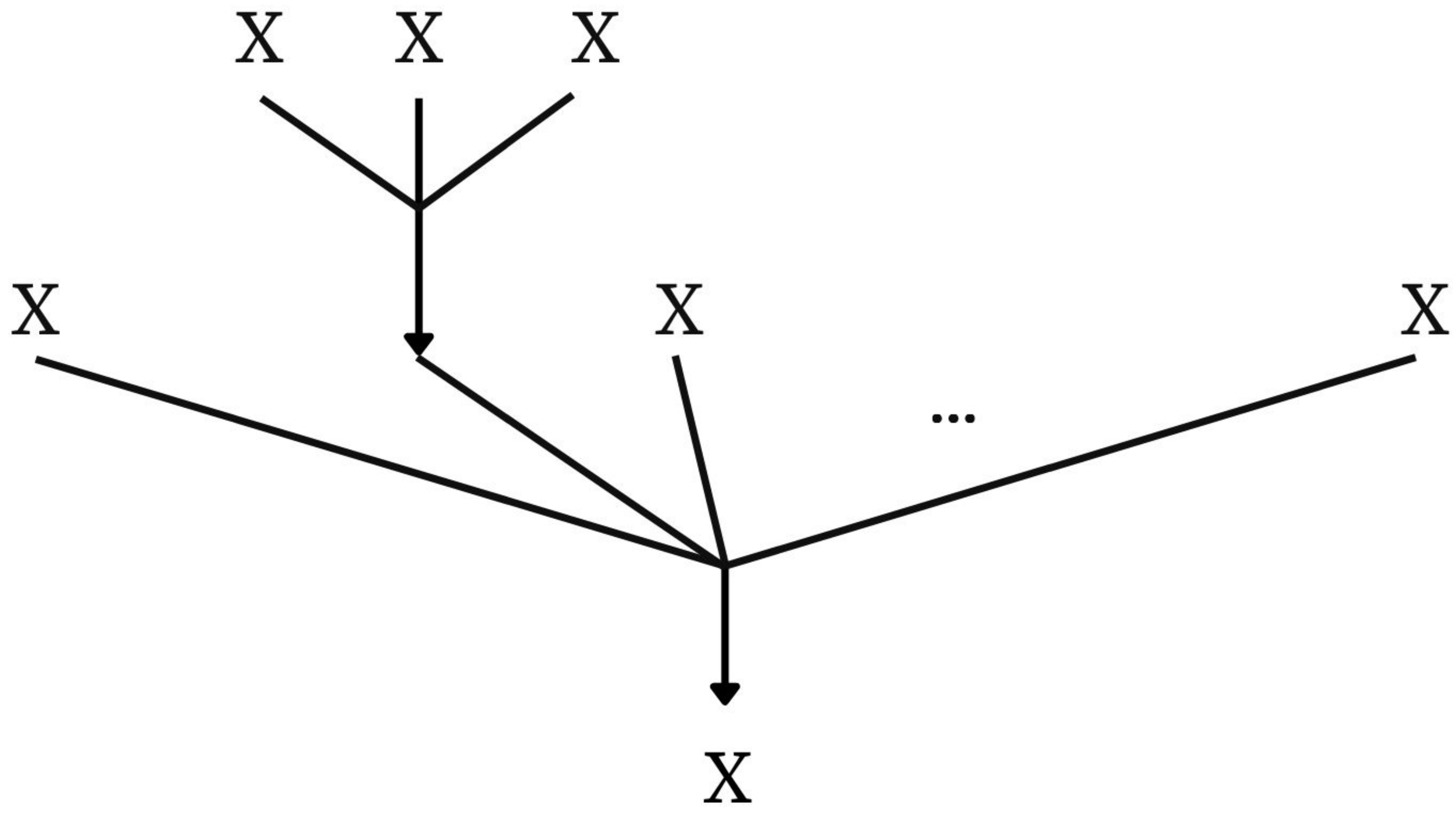
↓

X

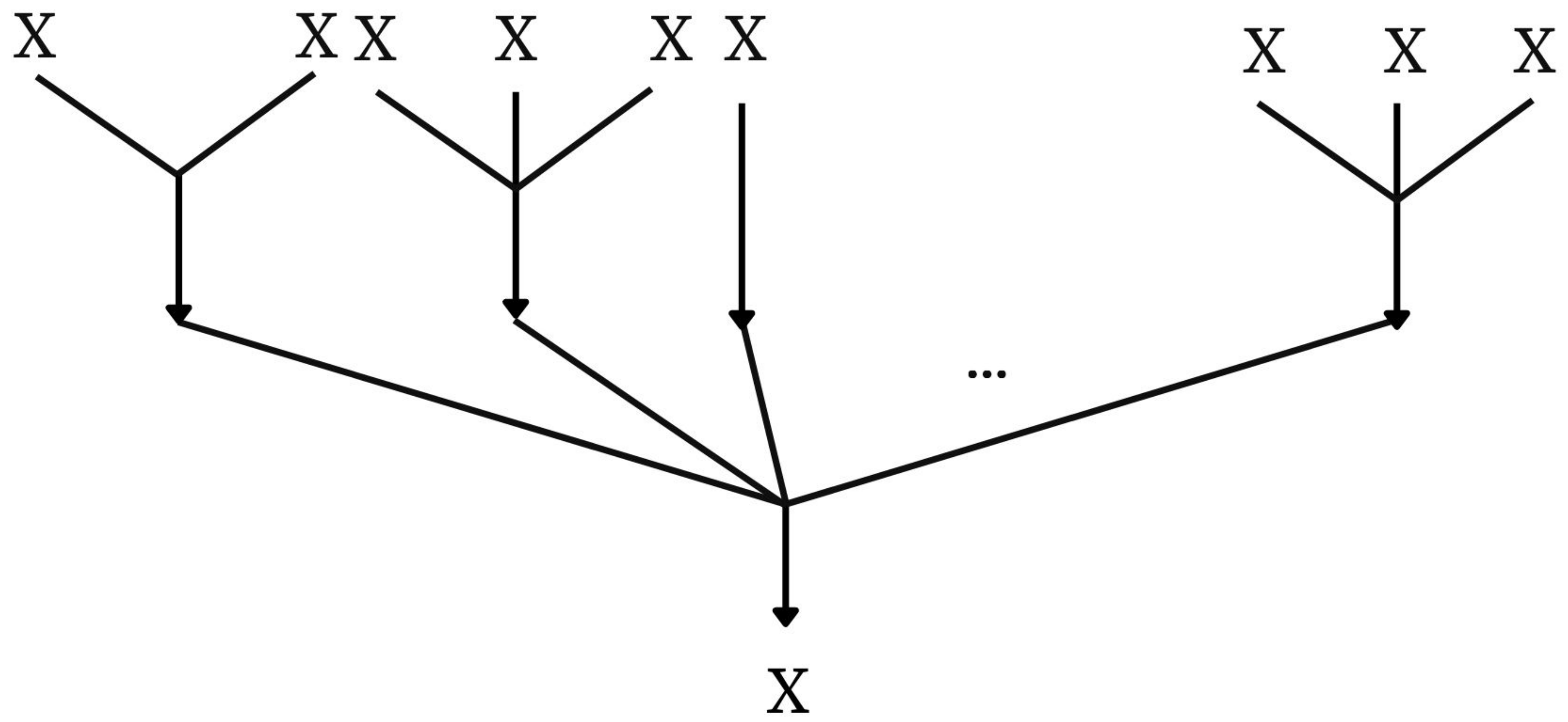
If the category has finite products, then we can consider maps $X \times \cdots \times X \to X$,



If the category has finite products, then we can consider maps $X \times \cdots \times X \to X$, which can be composed with other maps $X \times \cdots \times X \to X$.



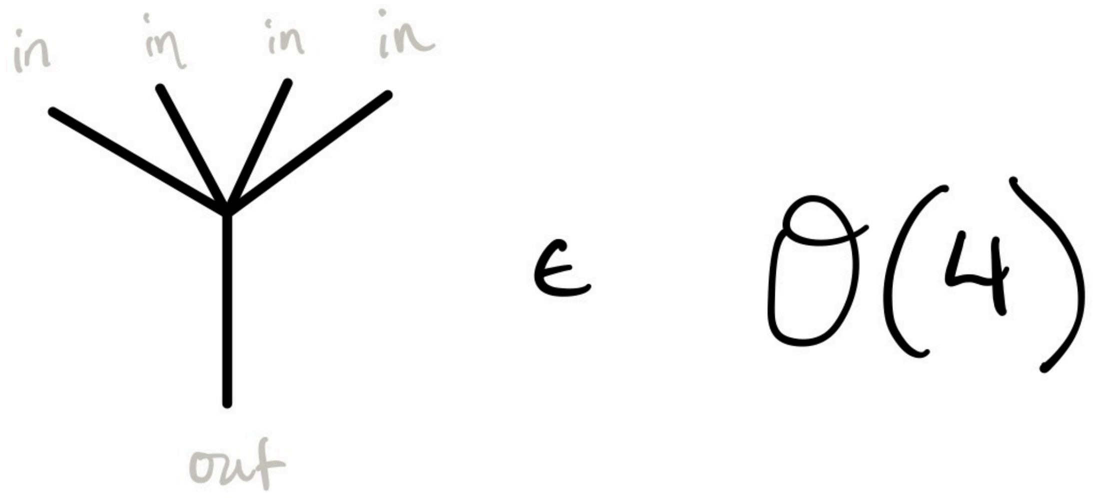
If the category has finite products, then we can consider maps $X \times \cdots \times X \to X$, which can be composed with other maps $X \times \cdots \times X \to X$.



Operads generalize this idea.

concrete maps $\rightsquigarrow$ abstract operations

An **operad** $\mathcal{O}$ consists of a collection of sets $\mathcal{O}(1), \mathcal{O}(2), \ldots$ (whose elements are thought of as "abstract n-ary operations")



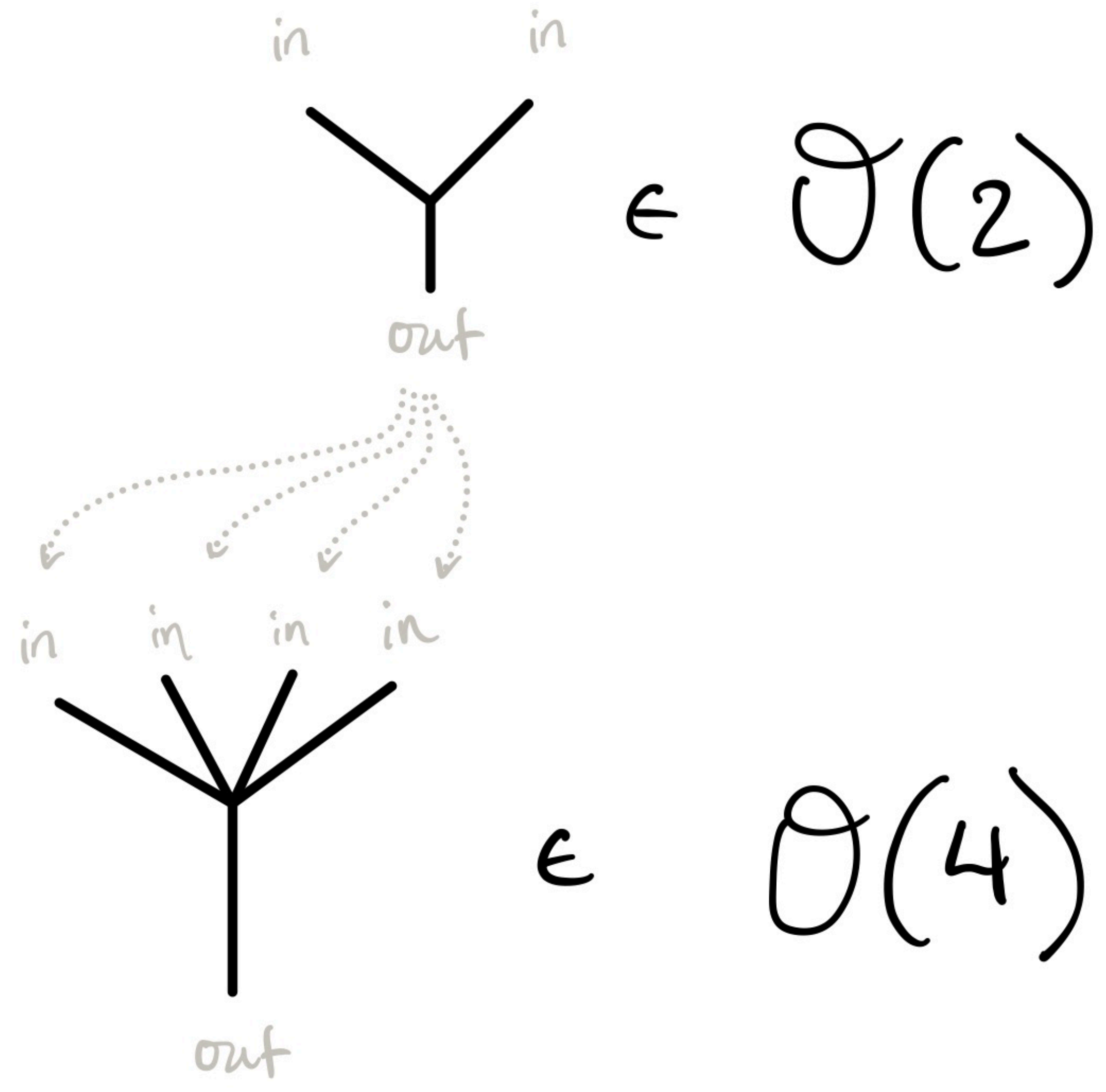
Operads generalize this idea.

concrete maps $\rightsquigarrow$ abstract operations

An **operad** $\mathcal{O}$ consists of a collection of sets $\mathcal{O}(1), \mathcal{O}(2), \ldots$ (whose elements are thought of as "abstract n-ary operations") together with **composition** functions

$$\circ_i \colon \mathcal{O}(n) \times \mathcal{O}(m) \to \mathcal{O}(n+m-1)$$

for all $n, m \geq 1$ and $1 \leq i \leq n$



Operads generalize this idea.

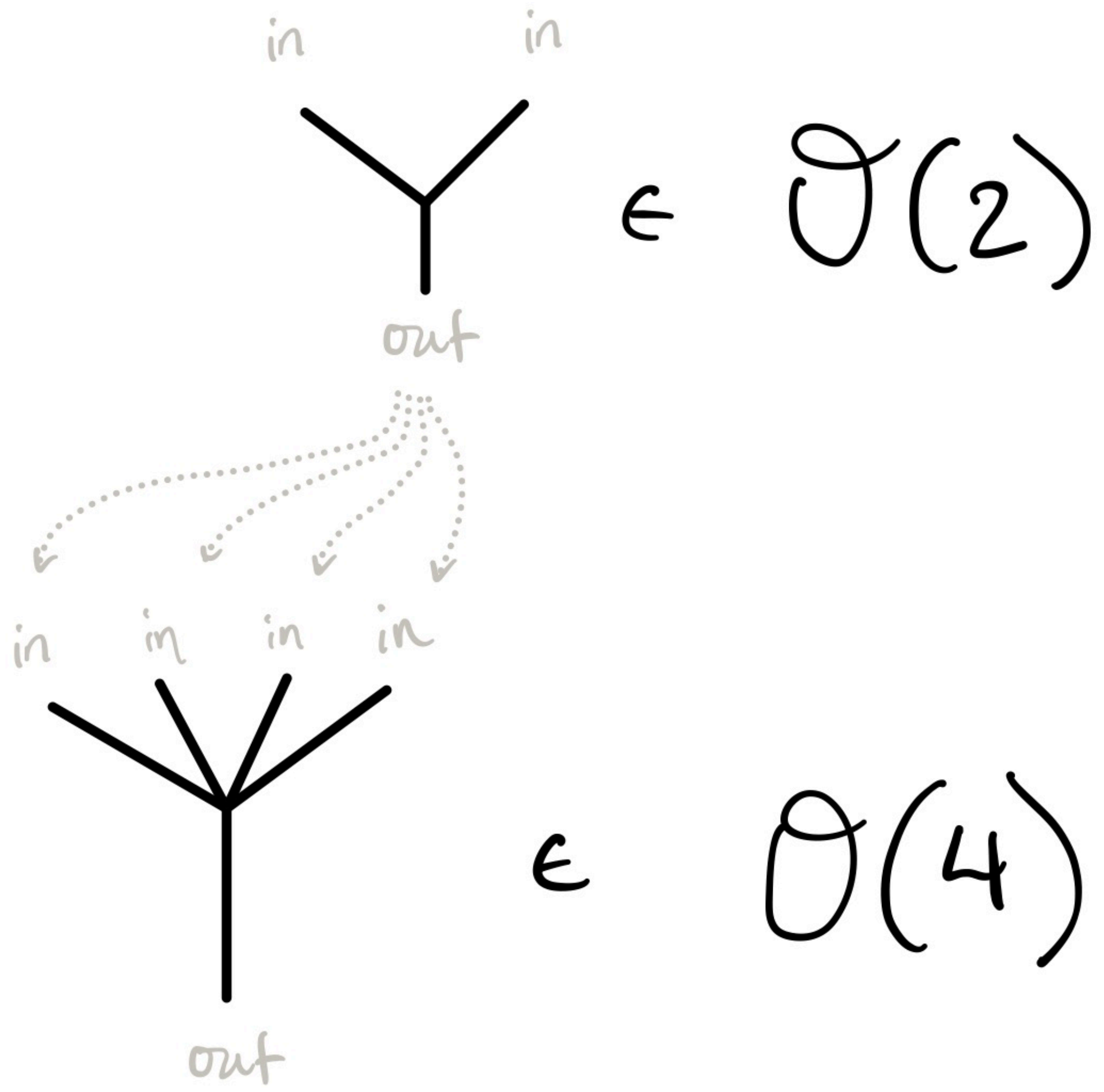
concrete maps $\rightsquigarrow$ abstract operations

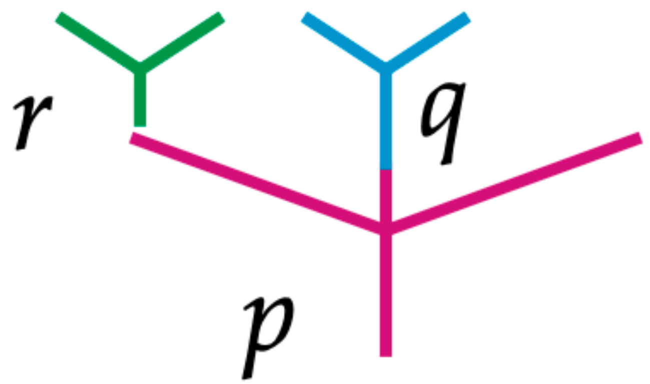
An **operad** $\mathcal{O}$ consists of a collection of sets $\mathcal{O}(1), \mathcal{O}(2), \ldots$ (whose elements are thought of as "abstract n-ary operations") together with partial **composition** functions

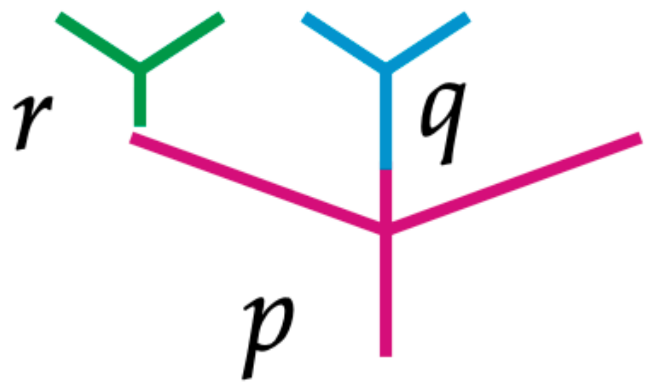
$$\circ_i \colon \mathcal{O}(n) \times \mathcal{O}(m) \to \mathcal{O}(n+m-1)$$

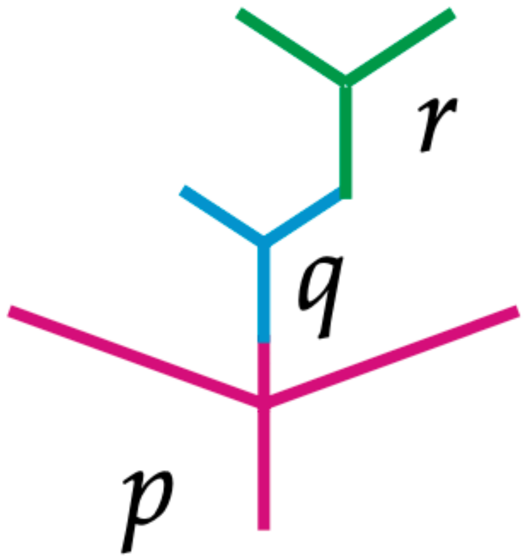
for all $n, m \geq 1$ and $1 \leq i \leq n$ and an element $1 \in \mathcal{O}(1)$ called the **identity**, satisfying associativity and unital axioms.

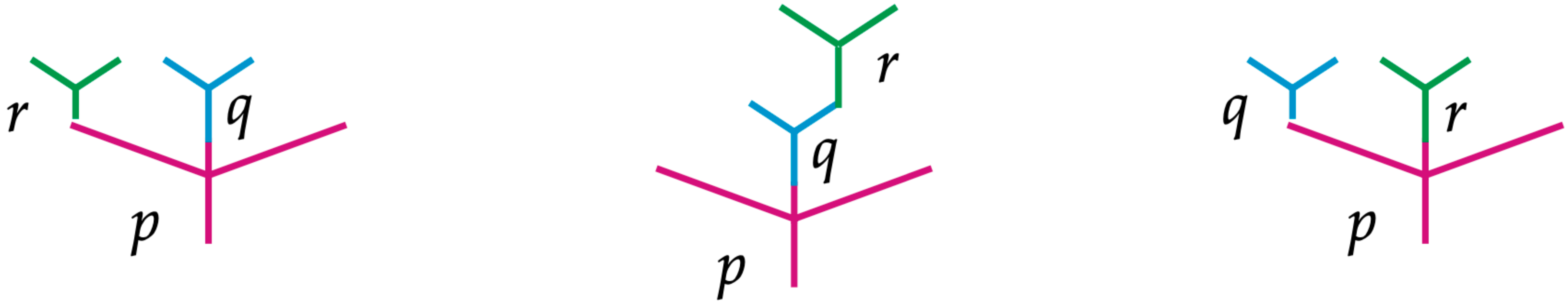
Think of operations in $\mathcal{O}(n)$ as planar rooted trees with n leaves.




r
q
p


r
q
p


r
q
p

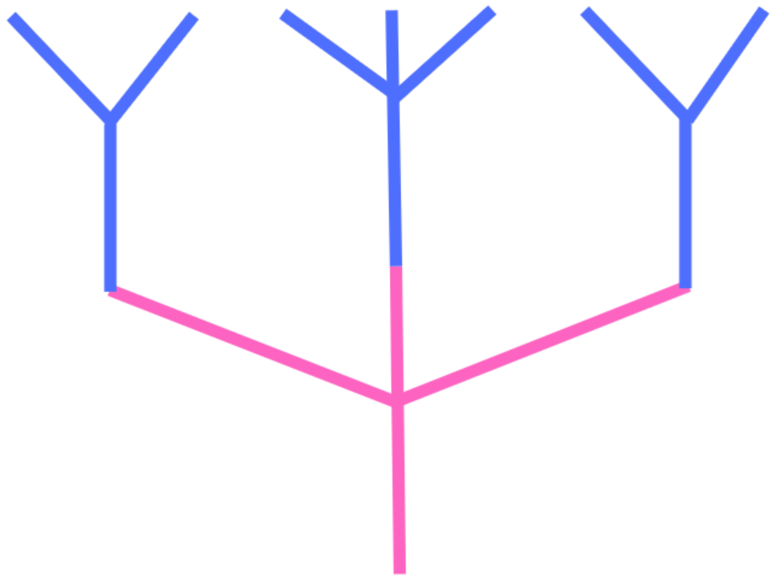

r
q
p
r
q
p
q
r
p

Some housekeeping

- Sometimes it's convenient to define operads with simultaneous composition $\circ$ rather than partial maps $\circ_i$. We'll use both.

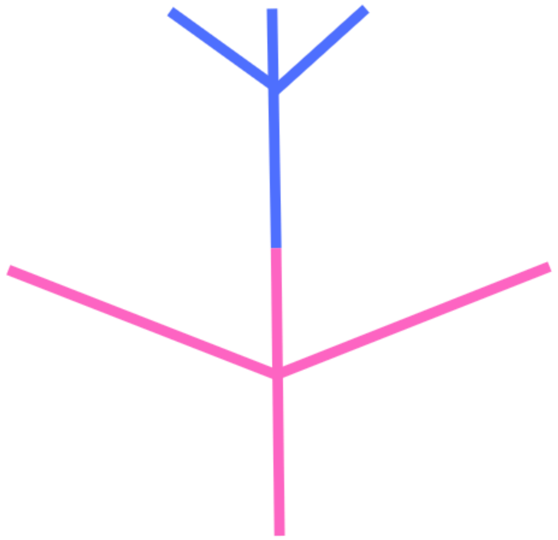
simultaneous

$$p \circ (q^1, q^2, q^3)$$



partial

$$p \circ_i q$$

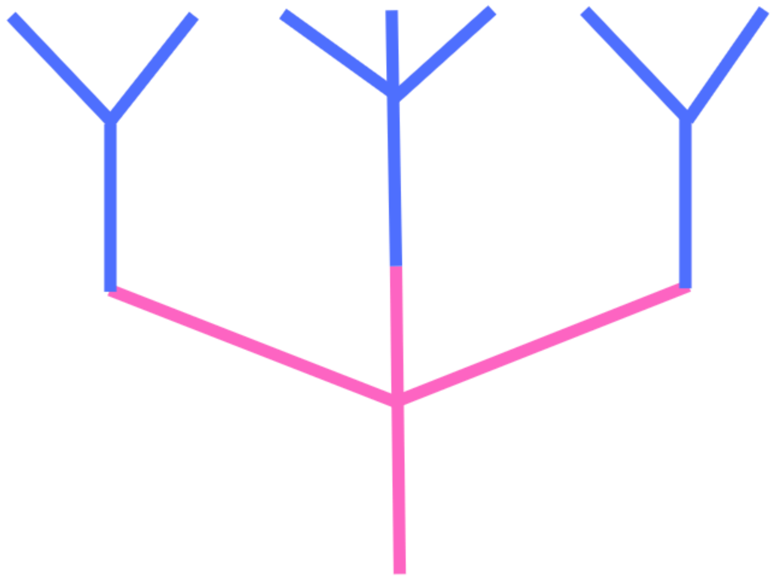


Some housekeeping

- Sometimes it's convenient to define operads with simultaneous composition $\circ$ rather than partial maps $\circ_i$. We'll use both.
- The $\mathcal{O}(n)$ may be other objects besides sets.
  - If they're vector spaces, then $\mathcal{O}$ is sometimes called a **linear operad.**
  - If they're topological spaces, then $\mathcal{O}$ is sometimes called a **topological operad.**

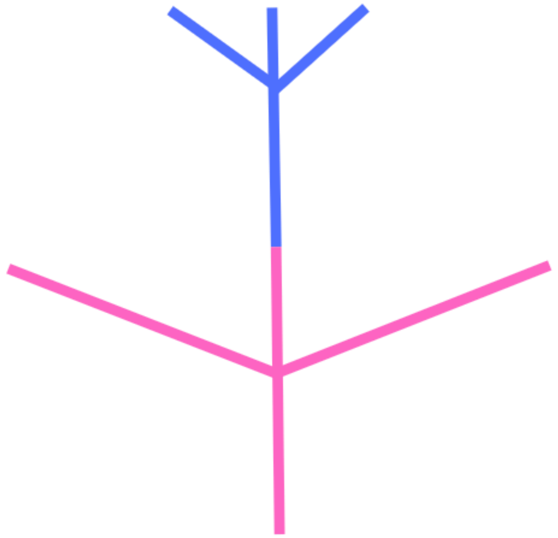
simultaneous

$$p \circ (q^1, q^2, q^3)$$



partial

$$p \circ_i q$$

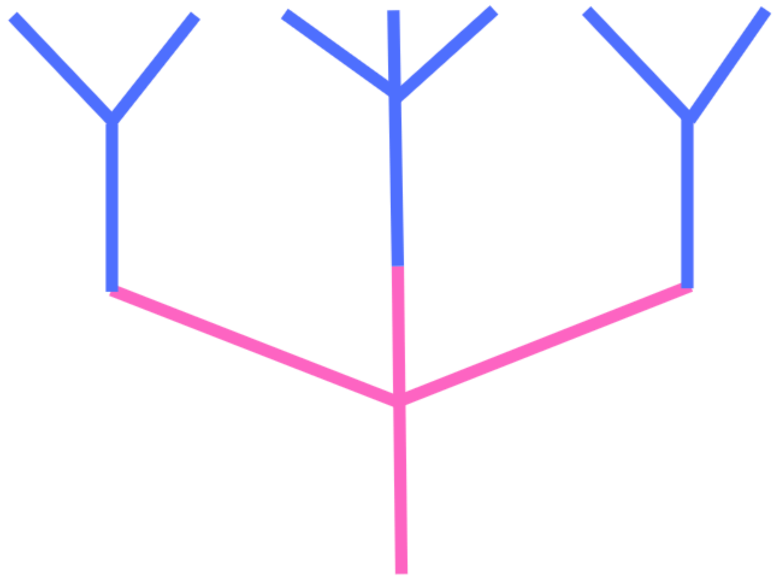


Some housekeeping

- Sometimes it's convenient to define operads with simultaneous composition $\circ$ rather than partial maps $\circ_i$. We'll use both.
- The $\mathcal{O}(n)$ may be other objects besides sets.
  - If they're vector spaces, then $\mathcal{O}$ is sometimes called a **linear operad.**
  - If they're topological spaces, then $\mathcal{O}$ is sometimes called a **topological operad.**
- More generally, operads can be defined in any (symmetric) **monoidal category**; the composition maps are morphisms in the category.

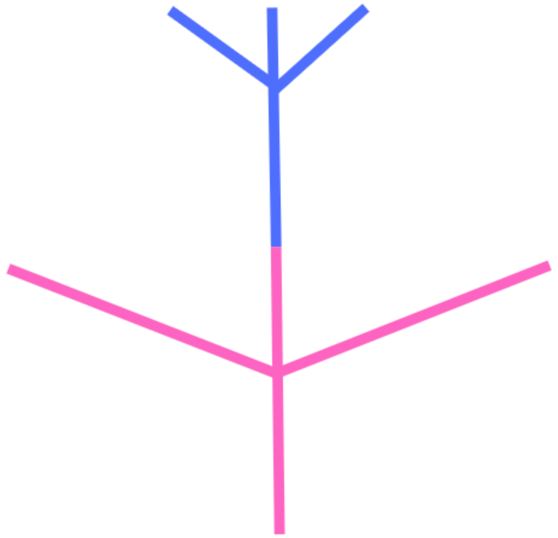
simultaneous

$$p \circ (q^1, q^2, q^3)$$

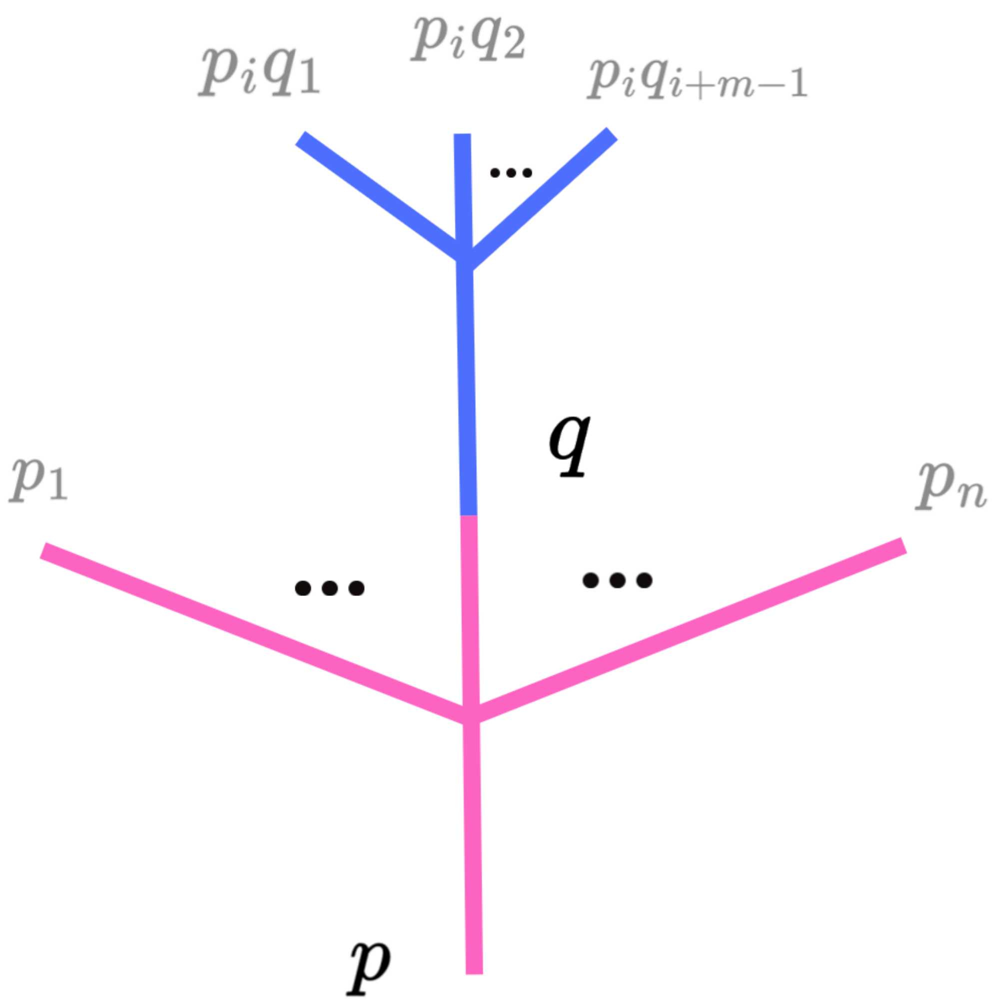


partial

$$p \circ_i q$$



Operad Δ of Topological Simplices



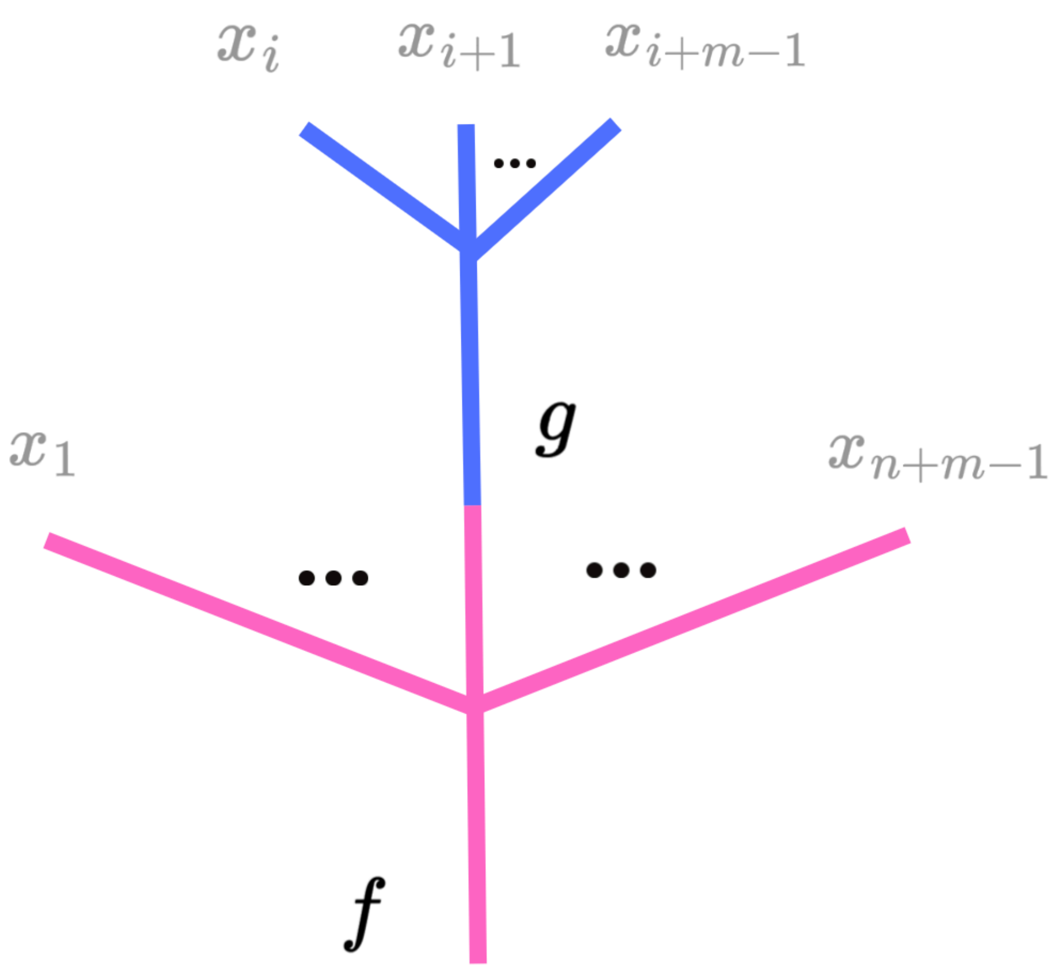
Topological simplices $\Delta_1, \Delta_2, \ldots$ form an operad. An n-ary operation is a probability distribution $p \in \Delta_n$.

- **Composition** $\circ_i$ is multiplication of probabilities from earlier. Given $p \in \Delta_n$ and $q \in \Delta_m$, we have

$$p \circ_i q = (p_1, \ldots, p_i q_1, \ldots, p_i q_m, \ldots, p_n) \in \Delta_{n+m-1}$$

- The **identity** is the probability distribution on one element $(1) \in \Delta_1$.

Endomorphism Operad, $\mathrm{End}(X)$



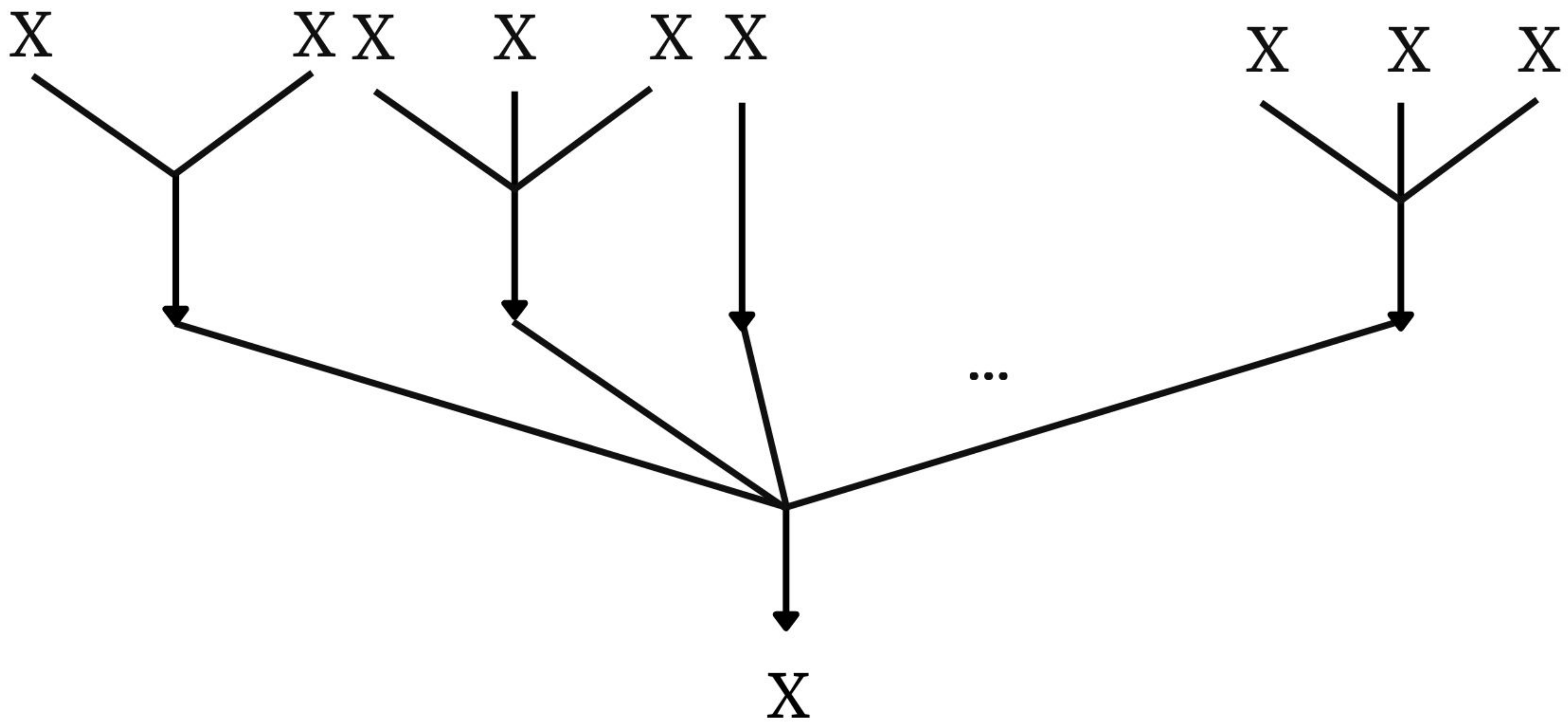
Given a fixed set X, sets of functions $\mathrm{End}_X(n) := \{X^n \to X\}$ form an operad.

- **Composition** $\circ_i$ is usual function composition. Given functions $f\colon X^n \to X$ and $g\colon X^m \to X$, use the output of g as the i^{th} input of f to obtain a new function

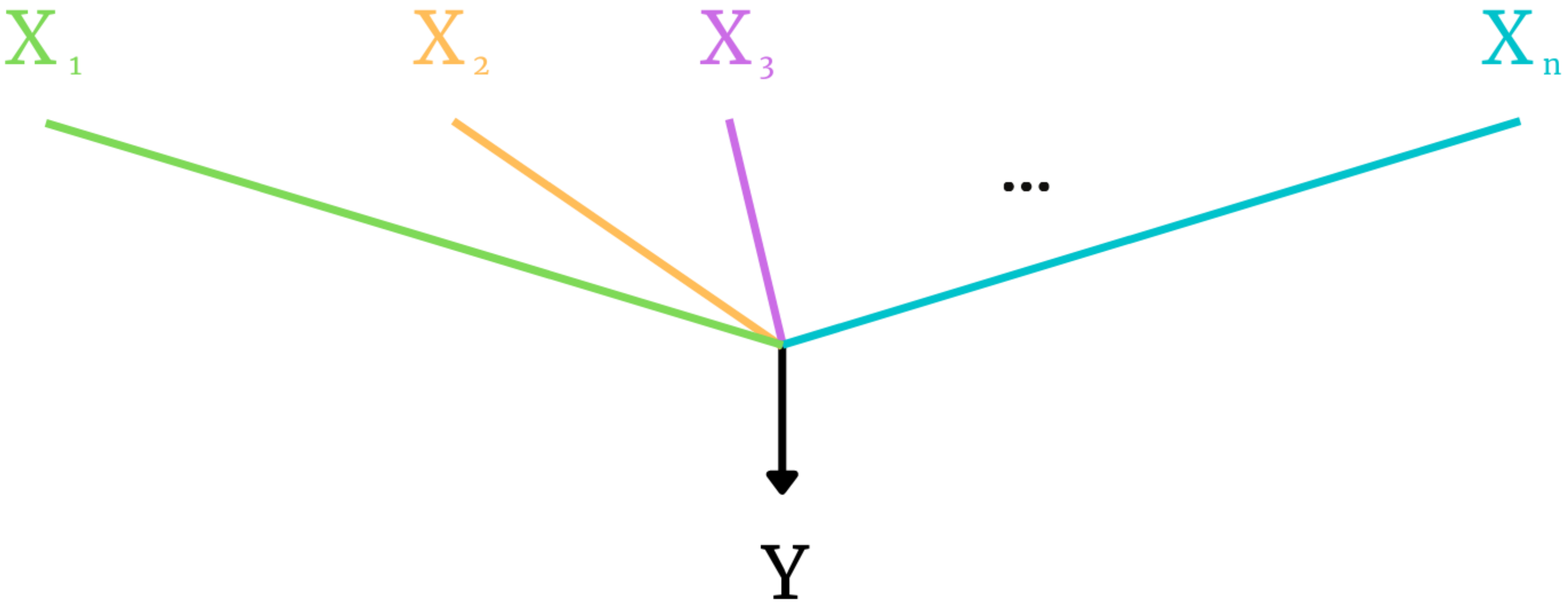
$$f \circ_i g \in \mathrm{End}_X(n+m-1).$$

- The **identity** is the identity function $\mathrm{id}_X\colon X \to X$.

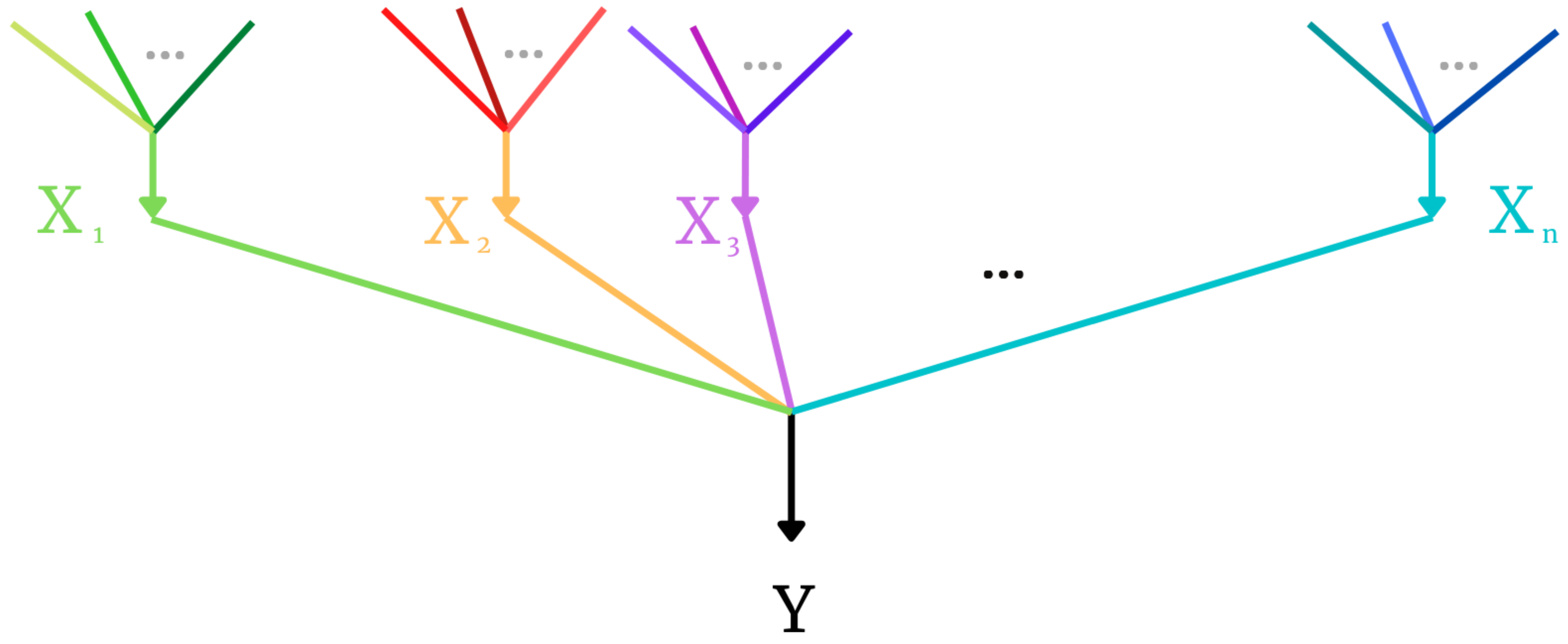
Aside: Think back to category theory. Instead of a single object



Aside: Think back to category theory. Instead of a single object we can also consider multiple objects or "types"

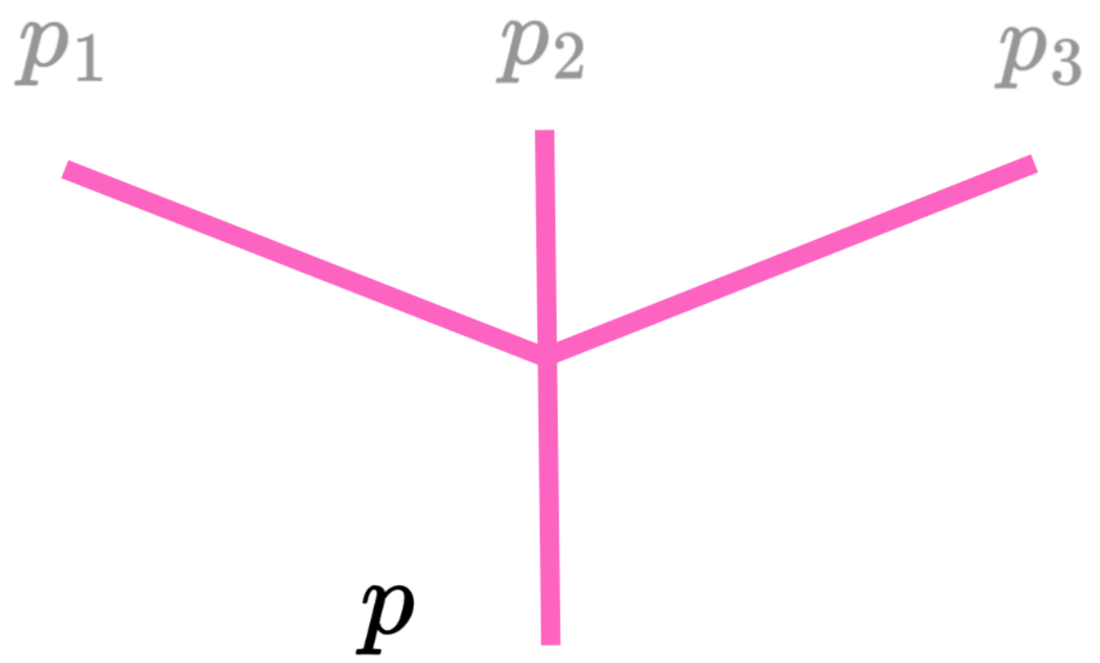


Aside: Think back to category theory. Instead of a single object we can also consider multiple objects or "types" and ways to compose them. This generalization of operads is called a **multicategory.**



Sometimes "operad" means "multicategory" (with symmetric group action), but not in this talk.

How is a probability distribution an "operation"?

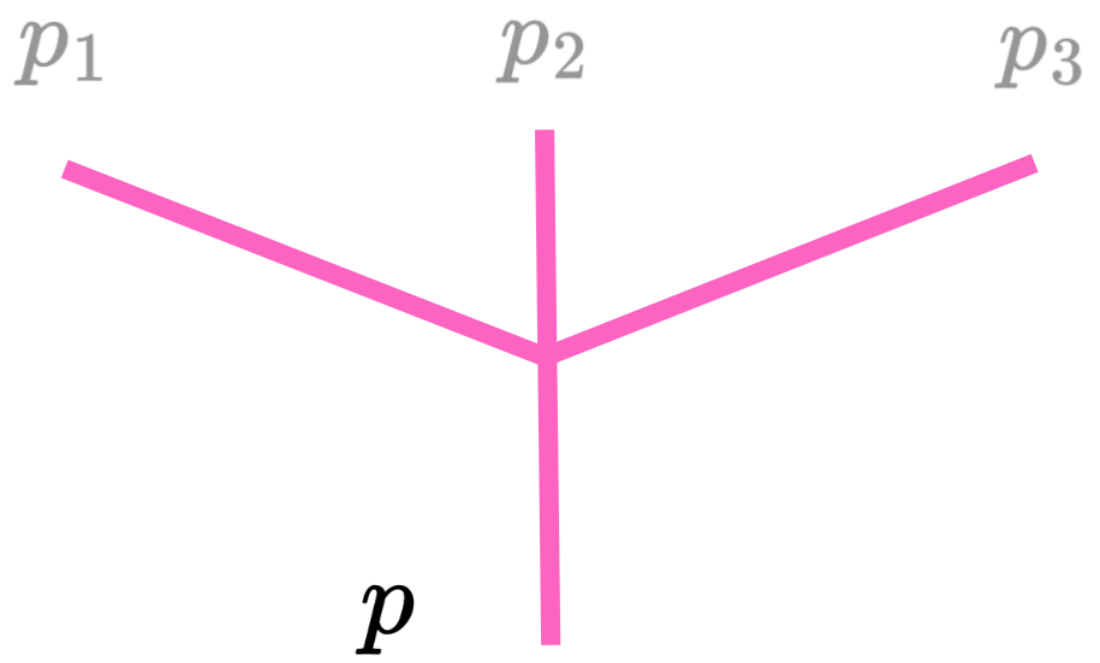


How is a probability distribution an "operation"? It's not.

It's helpful to represent elements of Δ_n by *actual* operations on some object X.

$$\Delta_n \to \{\text{maps } X^n \to X\}$$

This should be compatible with compositions $\circ_i$ and identities.



How is a probability distribution an "operation"? It's not.

It's helpful to represent elements of Δ_n by *actual* operations on some object X.

$$\Delta_n \to \{\text{maps } X^n \to X\}$$

This should be compatible with compositions $\circ_i$ and identities.

Example: The Real Line

Let $X = \mathbb{R}$ and let $\mathrm{End}_{\mathbb{R}}(n) := \mathsf{Top}(\mathbb{R}^n, \mathbb{R})$ denote the space of continuous functions $\mathbb{R}^n \to \mathbb{R}$ equipped with the product topology.

Then convex combinations work:

$$\Delta_n \to \mathrm{End}_{\mathbb{R}}(n)$$

$$p \mapsto \left(x \mapsto \sum_i p_i x_i \right).$$

Algebras Over Operads

Given an operad $\mathcal{O}$ of sets, an $\mathcal{O}$**-algebra** (or $\mathcal{O}$**-representation**) is a set X together with a sequence of functions

$$(\mathcal{O}(n) \to \mathrm{End}_X(n))_{n \in \mathbb{N}}$$

compatible with the operad composition and unit.

Analogously, "the real line $\mathbb{R}$ is a Δ-algebra."

More generally, any convex subset of Euclidean space is an algebra of the operad of topological simplices.

What about entropy?

"How does the 'derivation' way of thinking about [the chain rule] relate to Tom Leinster's [operadic] interpretation of it...?"

- Baez (2010)

Let's work backwards.

A **derivation** of an algebra A with values in an A-bimodule M is a linear map $d\colon A \to M$ satisfying

$$d(ab) = d(a)b + ad(b)$$

for all $a, b \in A$.

Let's work backwards.

$$d(p \circ q) = dp \circ q + p \circ dq$$

Let's work backwards.

$$d(p \circ_i q) = dp \circ_i q + p \circ_i dq$$

Let's work backwards.

$$d(p \circ_i q) = dp \circ_i^R q + p \circ_i^L dq$$

Bimodules over an operad

Our desiderata suggests the following definitions. The results then fall into place.

An **abelian bimodule over the operad** Δ is a sequence of topological spaces $M(1), M(2), \ldots$ together with left/right actions, where each $M(n)$ is also an abelian monoid.

$$\circ_i^L \colon \Delta_n \times M(m) \to M(n+m-1)$$

$$\circ_i^R \colon M(n) \times \Delta_m \to M(n+m-1)$$

By way of analogy

Just as every algebra A is a bimodule over itself, so every Δ-algebra is an abelian bimodule over Δ in a straightforward way.

$$d\colon \Delta_n \to \mathbf{End}_{\mathbb{R}}(n)$$

$$p \in \Delta_n \qquad \mapsto \qquad dp\colon \mathbb{R}^n \to \mathbb{R}$$

Derivation of an operad

A **derivation** of the operad of topological simplices is a sequence of continuous functions $(d\colon \Delta_n \to M(n))_{n\in\mathbb{N}}$ satisfying the Leibniz rule.

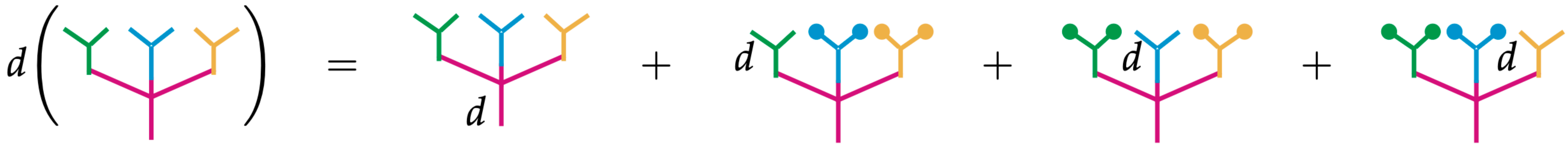
Derivation of an operad

A **derivation** of the operad of topological simplices is a sequence of continuous functions $(d\colon \Delta_n \to M(n))_{n\in\mathbb{N}}$ satisfying the Leibniz rule.

Before:	**After:**
entropy	**derivation**
chain rule	**Leibniz rule**
$(H\colon \Delta_n \to \mathbb{R})_{n\in\mathbb{N}}$	$(d\colon \Delta_n \to \mathrm{End}_{\mathbb{R}}(n))_{n\in\mathbb{N}}$

Proposition (B., 2021). When $\mathrm{End}_{\mathbb{R}}$ is equipped with a *particular* abelian Δ-bimodule structure, every derivation of the operad of topological simplices satisfies the chain rule.

$$d(p \circ (q^1, \ldots, q^n)) \overset{\text{basically}}{=} d(p) + \sum_{i=1}^{n} p_i \, d(q^i)$$



Theorem (B., 2021). Shannon entropy defines a derivation $(d\colon \Delta_n \to \mathbf{End}_{\mathbb{R}}(n))$ of the operad $\mathbf{\Delta}$, and every derivation of $\mathbf{\Delta}$ is a constant multiple of Shannon entropy when evaluated at the origin.

$$\text{“}d = cH.\text{”}$$

Proof. Arithmetic + the Proposition and Leinster/Faddeev.

Theorem (B., 2021). Shannon entropy defines a derivation $(d\colon \Delta_n \to \mathbf{End}_{\mathbb{R}}(n))$ of the operad Δ, and every derivation of Δ is a constant multiple of Shannon entropy when evaluated at the origin.

$$\text{``}d = cH.\text{''}$$

Proof. Arithmetic + the Proposition and Leinster/Faddeev.

Corollary. The chain rule for entropy:

$$H(p \circ (q^1, \ldots, q^n)) = H(p) + \sum_{i=1}^{n} p_i H(q^i).$$

A few other facets

for the record

Information cohomology

Given discrete random variables X and Y, conditional entropy satisfies

$$H(X, Y) = H(X) + H(Y|X).$$

sometimes called the **chain rule**.

Information cohomology

Given discrete random variables X and Y, conditional entropy satisfies

$$H(X, Y) = H(X) + H(Y|X).$$

sometimes called the **chain rule**.

Now let $XH(Y) := H(Y|X)$ to rewrite this as

$$H(X, Y) = H(X) + XH(Y).$$

Information cohomology

Given discrete random variables X and Y, conditional entropy satisfies

$$H(X,Y) = H(X) + H(Y|X).$$

sometimes called the **chain rule**.

Now let $XH(Y) := H(Y|X)$ to rewrite this as

$$H(X,Y) = H(X) + XH(Y).$$

This looks like the **1-cocycle condition** $\partial H = 0$ in group cohomology. (Consistent with the previous theorem, "$d = cH$.")

Information cohomology

Given discrete random variables X and Y, conditional entropy satisfies

$$H(X,Y) = H(X) + H(Y|X).$$

sometimes called the **chain rule**.

Now let $XH(Y) := H(Y|X)$ to rewrite this as

$$H(X,Y) = H(X) + XH(Y).$$

This looks like the **1-cocycle condition** $\partial H = 0$ in group cohomology. (Consistent with the previous theorem, "$d = cH$.")

In 2015, Baudot and Bennequin teased out this analogy in full detail.

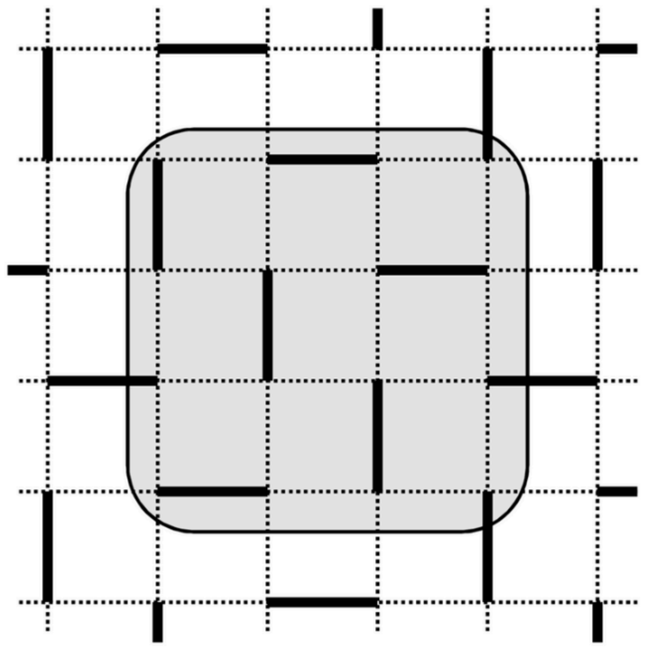
- P. Baudot, D. Bennequin, "Homological Nature of Entropy," *Entropy* (2015)

Entanglement entropy

A quantum state exhibits an "area law" if the **entanglement entropy** S between a region A and its complement is proportional to the size of the boundary ∂A,

$$S(A) \sim c|\partial A|.$$

Area laws are commonly observed in ground states of quantum-many body systems, but not random quantum states.



source: Adrian E. Feiguin, "The Density Matrix Renormalization Group and its time-dependent variants," *AIP Conference Proceedings* 1419, 5 (2011)

Entanglement entropy

A quantum state exhibits an "area law" if the **entanglement entropy** S between a region A and its complement is proportional to the size of the boundary ∂A,

$$S(A) \sim c|\partial A|.$$

Area laws are commonly observed in ground states of quantum-many body systems, but not random quantum states.

Black hole entropy

The **Bekenstein-Hawking entropy** of a black hole is proportional to the area "$|\partial A|$" of its event horizon.

$$S_{BH} = c|\partial A|$$

In holographic duality, the Ryu-Takayanagi formula has a similar flavor.

Other facets of entropy

- Tom Mainiero: entropy appears in the **Euler characteristic** of a cochain complex associated to a quantum state
  - "Homological Tools for the Quantum Mechanic," 2019
- P. Elbaz-Vincent and H. Gangl: showed information functions of degree 1 behave "**a lot like certain derivations**"
  - "Finite Polylogarithms, Their Multiple Analogues and the Shannon Entropy," 2015
- J. Baez, T. Fritz, T. Leinster: **category theoretical** characterization
  - "A characterterization of entropy in terms of information loss," 2011