Polynomial functors and Shannon entropy

David I. Spivak

Abstract

Past work shows that one can associate a notion of Shannon entropy to a Dirichlet polynomial, regarded as an empirical distribution. Indeed, entropy can be extracted from any $d \in$ Dir by a two-step process, where the first step is a rig homomorphism out of Dir, the *set* of Dirichlet polynomials, with rig structure given by standard addition and multiplication. In this short note, we show that this rig homomorphism can be upgraded to a rig *functor*, when we replace the set of Dirichlet polynomials by the *category* of ordinary (Cartesian) polynomials.

In the Cartesian case, the process has three steps. The first step is a rig functor $\mathbf{Poly}^{\mathbf{Cart}} \to \mathbf{Poly}$ sending a polynomial p to $\dot{p}y$, where $\dot{p}$ is the derivative of p. The second is a rig functor $\mathbf{Poly} \to \mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}$, sending a polynomial q to the pair $(q(1), \Gamma(q))$, where $\Gamma(q) = \mathbf{Poly}(q, y)$ can be interpreted as the global sections of q viewed as a bundle, and $q(1)$ as its base. To make this precise we define what appears to be a new distributive monoidal structure on $\mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}$, which can be understood geometrically in terms of rectangles. The last step, as for Dirichlet polynomials, is simply to extract the entropy as a real number from a pair of sets (A, B); it is given by $\log A - \log \sqrt[A]{B}$ and can be thought of as the log aspect ratio of the rectangle.

1 Introduction

In practice, a probability distribution on a set of *outcomes* arises from considering finite samples. A sample consists of a set of observations, or *draws*, each corresponding to one of the outcomes. For example, the following is a sample with five (5) outcomes and eight (8) draws:

$$\tag{1}$$
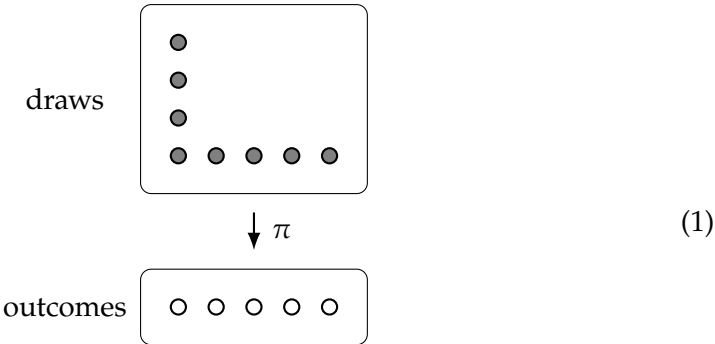
This corresponds to the probability distribution $P = (\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$. But the sample itself can be encoded in the form of a polynomial, namely $p := y^4 + 4y$. Note that $p(1) = 5$ is the number of outcomes and that $\dot{p}(1) = 8$ is the number of draws, where $\dot{p} = 4y^3 + 4$ is the derivative of p. The map π itself is somehow inherent in p: one of its summands has an exponent of 4, whereas its other four summands each have an exponent of 1. Yet one may wonder: is this polynomial encoding really meaningful, or is it just a bizarre packaging of the sample? Our goal in this paper is to show that it is meaningful, at least when it comes to considering the Shannon entropy $H(P)$.

The Shannon entropy of a distribution [Sha48] is a measure of how much information is transmitted when outcomes are selected according to the distribution. For example, if one repeatedly chooses an element of the 8 draws in diagram (1) uniformly at random but only reports the outcome, then the first outcome will show up four-times more often than any other. As we will explain, Shannon's information theory says that this distribution has entropy $H(P) = 2$, i.e. it transmits the same amount of information as if it were a uniform distribution on only 4 outcomes.

In this paper we will give a category-theoretic account of the Shannon entropy of the probability distribution corresponding to a sample encoded as a polynomial p, or more precisely a *polynomial functor* $p \in$ **Poly**. Polynomial functors are ubiquitous: they show up in type theory [ACH19; AN18], dynamical systems theory [Spi20; SN22], database theory [SW15; Spi21], programming language theory [BD96; AAG03], and higher category theory [TCM19; Sha21].

The category **Poly** of polynomial functors in one variable has an enormous amount of structure. For example, it has at least eight distinct monoidal structures, of which two will be relevant to us. One is the coproduct: given two polynomials p, q, we may add them to form $p + q$. In terms of samples, this operation simply takes the disjoint union of two samples: both the sets of outcomes and the sets of draws. The other is the *Dirichlet product*, denoted $\otimes$. We will give the precise formula for $p \otimes q$ in Section 2.3, but the idea is that it runs the two samples independently: an outcome in $p \otimes q$ is a pair consisting of an outcome from p and an outcome from q, and a draw is also a pair consisting of a draw from p and a draw from q.

These two operations make **Poly** a *distributive monoidal category*, because $\otimes$ distributes over +. The goal of this paper is to show that most of the process for taking the Shannon entropy of a sample is fully category-theoretic. Indeed, we will factor the process into three stages, the first two of which are completely categorical, and the last of which extracts a real number that will be the entropy.

The first stage is to define a rig functor $T: \mathbf{Poly^{Cart}} \rightarrow \mathbf{Poly}$, which sends a polynomial p to $T(p) := \dot{p}y$, where $\dot{p}$ is the derivative of p. The second stage is to define a rig functor $R: \mathbf{Poly} \rightarrow \mathbf{Set} \times \mathbf{Set^{op}}$, which sends a polynomial q to $R(q) := (q(1), \Gamma(q))$, where $\Gamma(p) = \mathbf{Poly}(p, y)$ can be construed as the set of global sections of p, viewed as a bundle.

The fact that both T and R are rig functors means that each preserves both the coproduct and the $\otimes$-product, a surprising amount of structure. But to say this, we need to define what appears to be a novel symmetric monoidal product $\otimes$ on

$\mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}$. It is given by

$$(A_1, B_1) \otimes (A_2, B_2) := \left(A_1 A_2 \, , \, B_1^{A_2} B_2^{A_1} \right).$$

This monoidal product $\otimes$ distributes over the coproduct, which is given by

$$(A_1, B_1) + (A_2, B_2) := (A_1 + A_2 \, , \, B_1 \times B_2),$$

hence making $\mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}$ a distributive monoidal category, and in particular a *rig category*. We will explain these two rig functors T and R in Section 3. We denote their composite—the result of the first and second stages—by

$$\hbar := (R \circ T) \colon \mathbf{Poly}^{\mathbf{Cart}} \to \mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}.$$

It contains the categorical aspect of the entropy in a given sample $p \in \mathbf{Poly}^{\mathbf{Cart}}$.

Before we discuss the third stage, we need a bit of intuition. Namely, we can think of an object $(A, B) \in \mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}$ as encoding a rectangle that has length A and width $\sqrt[A]{B}$. The coproduct of two rectangles is given by adding their lengths and taking the geometric mean of their widths. The $\otimes$-product of two rectangles is given by multiplying both their lengths and their widths. It is in these terms that we can understand the third and final stage, which is simply to take the *log aspect ratio* (the log of the quotient of length divided by width) of a given rectangle:

$$L(A, B) = \log A - \log \sqrt[A]{B}.$$

That is, we will prove that for any polynomial p with an associated probability distribution P, the Shannon entropy $H(P)$ can be computed by first applying the rig-functorial operation to obtain $\hbar(p) \in \mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}$, and then by extracting the log aspect ratio:

$$H(P) = L(\hbar(p)).$$

We will conclude by returning to our original example, after giving the full composite: the function that takes a polynomial p and returns the entropy of the corresponding empirical distribution is given by

$$L(\hbar(p)) := \log \dot p(1) - \frac{\log \Gamma(\dot p y)}{\dot p(1)}$$

Note that $\log \sqrt[A]{B} = \frac{\log B}{A}$.

So consider again the polynomial $p = y^4 + 4y$, depicted in (1). Then we calculate

$$\dot p y = 4y^4 + 4y, \quad \dot p(1) = 8, \quad \Gamma(\dot p y) = 4^4 * 1^4 = 2^8, \quad \text{and} \quad L(\hbar(p)) = \log 8 - \frac{\log 2^8}{8} = 2$$

which agrees with our former calculation: its entropy is $H(P) = L(\hbar(p)) = 2$.

The remainder of this note is divided into two sections: Section 2 gives background on polynomial functors, including the definition of $\mathbf{Poly}^{\mathbf{Cart}} \subseteq \mathbf{Poly}$ as well as the $+$ and $\otimes$ structures. Section 3 gives the main results: explaining the seemingly novel distributive monoidal structure on $\mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}$, providing a rig monoidal

3

functor $\hbar\colon \mathbf{Poly^{Cart}} \to \mathbf{Set} \times \mathbf{Set^{op}}$, showing how to extract the entropy via a partial function $L\colon \mathrm{Ob}(\mathbf{Set} \times \mathbf{Set^{op}}) \to \mathbb{R}$, and finally proving the main theorem: that $H(P) = L(\hbar(p))$.

There have been other categorical approaches to entropy, most notably [BFL11], [BF14], [Lei21], and [Par22]. Our presentation here has almost nothing in common with those.

However, this work is closely aligned with [SH21]. There, the authors—myself and Tim Hosgood—use Dirichlet polynomials rather than ordinary (Cartesian)[1] polynomials. At the time, we seemed to have a choice of whether to use Dirichlet or Cartesian polynomials, and the Dirichlet route seemed cleaner and more intuitive for talking about the bundles. However, we were missing a few key ideas at the time. Whereas there we only factored out from H a rig homomorphism (a function) Dir $\to$ Rect to a somewhat ad hoc rig we called Rect, the presentation here factors out from H a rig *functor* $\mathbf{Poly^{Cart}} \to \mathbf{Set} \times \mathbf{Set^{op}}$. Thus it is a significant categorical upgrade.

Acknowledgments

2 Background on polynomial functors

Readers familiar with the rig category $(\mathbf{Poly}, 0, +, y, \otimes)$ should skip to Section 3.1.

2.1 Basics

The main purpose of this section is to fix notation and provide a brief overview of polynomial functors in one variable. More extensive background material can be found in [SN22] and [GK12].

Definition 2.1 (Polynomial functor)**.** Given a set S, we denote the corresponding representable functor by

$$y^S := \mathbf{Set}(S, -)\colon \mathbf{Set} \to \mathbf{Set},$$

e.g. $y^S(X) := X^S$. In particular $y = y^1$ is the identity and $y^0 = 1$ is constant singleton.

A *polynomial functor* is a functor $p\colon \mathbf{Set} \to \mathbf{Set}$ that is isomorphic to a sum of representables, i.e. for which there exists a set T, a set $p[t] \in \mathbf{Set}$ for each $t \in T$, and an isomorphism of functors

$$p \cong \sum_{t \in T} y^{p[t]}.$$

---

[1]René Descartes at least invented the notation, e.g. $y^2 + 3y + 2$, for polynomials; hence we refer to them as *Cartesian polynomials* when we need to distinguish them from Dirichlet polynomials.

We refer to T as the set of *p-types*, and for each type $t \in T$ we refer to $p[t]$ as the set of *p-terms of type t*.

A *morphism* $\varphi \colon p \to p'$ of polynomial functors is simply a natural transformation between them. It is called *cartesian* if for every map of sets $f \colon S \to S'$, the naturality square

$$
\begin{array}{ccc}
p(S) & \xrightarrow{\ p(f)\ } & p(S) \\
{\scriptstyle \varphi(S)}\downarrow & \lrcorner & \downarrow{\scriptstyle \varphi(S')} \\
p'(S') & \xrightarrow[\ p'(f)\ ]{} & p'(S')
\end{array}
$$

is a pullback of sets. We denote the category of polynomial functors by **Poly** and the wide subcategory of polynomials and cartesian maps by $\mathbf{Poly}^{\mathbf{Cart}} \subseteq \mathbf{Poly}$. ◊

For any polynomial $p = \sum_{t \in T} y^{p[t]}$, we have a canonical isomorphism $p(1) \cong T$; hence from now on we will denote p by

$$
p = \sum_{I \in p(1)} y^{p[I]} \tag{2}
$$

so that each p-types is written with an upper-case letter, e.g. $I \in p(1)$, and its terms are written with corresponding lower-case letters, e.g. $i \in p[I]$.

Remark 2.2. Using the Yoneda lemma, the fact that a morphism in **Poly** is just a natural transformation, and the fact that a polynomial is a coproduct of representables, we derive

$$
\mathbf{Poly}(p, q) = \mathbf{Poly}\left(\sum_{I \in p(1)} y^{p[I]}, \sum_{J \in p(1)} y^{q[J]} \right)
$$

$$
\cong \prod_{I \in p(1)} \mathbf{Poly}\left(y^{p[I]}, \sum_{J \in p(1)} y^{q[J]} \right)
$$

$$
\cong \prod_{I \in p(1)} \sum_{J \in q(1)} \mathbf{Set}(q[J], p[I]).
$$

Thus we can understand a morphism $p \to q$ in **Poly** to consist of two parts $(\varphi_1, \varphi^\sharp)$ as follows:

$$
\varphi_1 \colon p(1) \to q(1) \qquad \text{and} \qquad \varphi_I^\sharp \colon q[J] \to p[I], \tag{3}
$$

where $J := \varphi_1(I)$. That is, φ_1 is a function from p-types to q-types, and $\varphi_i^\sharp$ is a function on terms that *depends on a choice of position $I \in p(1)$*. We refer to φ_1 as the *on-types function* and to $\varphi^\sharp$ as the *backwards on-terms* function.

One can check that a map $\varphi \colon p \to q$ is cartesian iff the backwards-on-terms function $\varphi_I^\sharp$ is a bijection $p[I] \cong q[\varphi_1 I]$ for each type $I \in p(1)$. ◊

Example 2.3 (Types and global sections, $p(1)$ and $\Gamma(p)$)**.** For any polynomial p, we will be particularly interested in two sorts of maps: $y \to p$ and $p \to y$. The former is easy: a map $y \to p$ is given on types by choosing a single type $I \in p(1)$ to be the

image of the unique type $! \in y(1)$ and given backward on terms using the unique choice of function $p[I] \to 1 = y[!]$. Thus we have $p(1) \cong \mathbf{Poly}(y, p)$.

More interesting are the maps $\gamma \colon p \to y$. This time γ is trivial on types: each type $I \in p(1)$ is sent to the unique type $! \in y(1)$. However on terms, we need a map $\varphi_I^\sharp \colon 1 \to p[I]$ for each I, meaning a choice of term $i \in p[I]$ for each $I \in p(1)$. In other words, writing $\Gamma(p) := \mathbf{Poly}(p, y)$, we have

$$\Gamma(p) \cong \prod_{I \in p(1)} p[I]. \tag{4}$$

We refer to $\Gamma(p)$ as the set of *global sections* of p, as is justified by the bundle terminology the next section.

Note that $-(1) \colon \mathbf{Poly} \to \mathbf{Set}$ and $\Gamma \colon \mathbf{Poly} \to \mathbf{Set}^{\mathrm{op}}$ are functorial, as they are represented and corepresented by $y \in \mathbf{Poly}$. We will be very interested in the functor

$$R \colon \mathbf{Poly} \to \mathbf{Set} \times \mathbf{Set}^{\mathrm{op}} \tag{5}$$

given by $R(p) := (p(1), \Gamma(p))$. In fact, R is a left adjoint, but we do not need that for this paper. In Remark 3.6 we will explain that $R(p)$ can be viewed as the *rectangular aspect* of the polynomial p, hence the name R. ◊

2.2 Derivatives and bundles

We can understand polynomial functors in terms of bundles, using the derivative. For any polynomial p, its derivative $\dot{p}$ is defined as follows:

$$\dot{p} := \sum_{I \in p(1)} \sum_{i \in p[I]} y^{p[I] - \{i\}} \tag{6}$$

where $p[I] - \{i\}$ denotes the set-difference. Note that $\dot{p}(1) \cong \sum_{I \in p(1)} p[I]$ is the set of all p-terms, and it comes with a map $\dot{p}(1) \to p(1)$ to the set of p-types. Often in the literature, this map of sets—which we call a bundle—is taken to be the polynomial itself. A map of polynomials $\varphi \colon p \to q$ can be written in terms of these bundles:

$$
\begin{array}{ccccc}
\dot{p}(1) & \xleftarrow{\ \varphi^\sharp\ } & p(1) \times_{q(1)} \dot{q}(1) & \longrightarrow & \dot{q}(1) \\
\downarrow & & \downarrow \quad \lrcorner & & \downarrow \\
p(1) & =\!=\!=\!= & p(1) & \xrightarrow{\ \varphi_1\ } & q(1)
\end{array}
$$

Just as in Remark 2.2, one provides a forward map on types $\varphi_1 \colon p(1) \to q(1)$, at which point one takes the pullback of that map with $\dot{q}(1) \to q(1)$, and then one provides a backward map $\varphi^\sharp \colon p(1) \times_{q(1)} \dot{q}(1) \to \dot{p}(1)$ on directions. Again, φ is cartesian iff $\varphi^\sharp$ is a bijection.

We write $pq = p \times q$ for the usual product of two polynomials, e.g. $\dot{p}y = \dot{p} \times y$.

Proposition 2.4. *The assignment $p \mapsto \dot{p}y$ is a functor $\mathbf{Poly}^{\mathbf{Cart}} \to \mathbf{Poly}^{\mathbf{Cart}}$.*

Proof. We can think of $\dot{p}y$ as follows:

$$\dot{p}y \cong \sum_{I \in p(1)} \sum_{i \in p[I]} y^{p[I]} \tag{7}$$

Given a cartesian map $\varphi \colon p \to q$, the bijection $\varphi^{\sharp} \colon q[\varphi_1(I)] \cong p[I]$ lets us define a map $\dot{p}y \to \dot{q}y$ in an obvious way. $\qquad\square$

Remark 2.5. In fact, the assignment $(p \mapsto \dot{p}y) \colon \mathbf{Poly}^{\mathbf{Cart}} \to \mathbf{Poly}^{\mathbf{Cart}}$ extends to a comonad on $\mathbf{Poly}^{\mathbf{Cart}}$. The counit map $\epsilon_p \colon \dot{p}y \to p$ is cartesian and is given on types by $(I, i) \mapsto I$. The comultiplication $\delta_p \colon \dot{p}y \to \ddot{p}y^2 + \dot{p}y$ is given by the coproduct inclusion.

A coalgebra for this comonad is a polynomial p equipped with a map $\gamma \colon p \to \dot{p}y$ such that $\epsilon_p \circ \gamma = \mathrm{id}_p$; it is not hard to check that the other condition holds for free. Hence a coalgebra structure on p can be identified with a choice a global section $p \to y$, i.e. an element $\gamma \in \Gamma(p)$. Of course the map $p \to y$ is not cartesian in general, so the only way it can be encoded in $\mathbf{Poly}^{\mathbf{Cart}}$ is via this coalgebra structure. A map of coalgebras is a cartesian map $\varphi \colon p \to p'$ that commutes with the global sections: $\Gamma(\varphi)(\gamma') = \gamma$.

The above is intriguing in that $\Gamma(p)$ is a major player in the story of this paper, but we currently know of no further connection between entropy and this comonad. $\quad\Diamond$

2.3 Rig monoidal structure on Poly

The category $\mathbf{Poly}$ has coproducts $p + q$ and products $p \times q$ given by usual polynomial arithmetic. We will be more interested in the former:[2] coproducts constitute a symmetric monoidal product with unit 0. A type in $p + q$ is a type in p or disjointly a type in q, and a term of that type is as specified in p or q, respectively.

We will also be interested in another monoidal product called *Dirichlet product* and denoted $- \otimes -$; the types and terms of $p \otimes q$ are given by the following formula:

$$\left(\sum_{I \in p(1)} y^{p[I]} \right) \otimes \left(\sum_{J \in q(1)} y^{q[J]} \right) := \sum_{(I,J) \in p(1) \times q(1)} y^{p[I] \times q[J]}. \tag{8}$$

This gives a symmetric monoidal structure $(\mathbf{Poly}, y, \otimes)$. A type in $p \otimes q$ is just a pair of types $(I, J) \in p(1) \times q(1)$ and a term of it is just a pair of terms $(i, j) \in p[I] \times q[J]$.

In the language of bundles, $p + q$ and $p \otimes q$ are respectively given by

$$
\begin{array}{ccc}
\dot{p}(1) + \dot{q}(1) & \qquad & \dot{p}(1) \times \dot{q}(1) \\
\downarrow & & \downarrow \\
p(1) + q(1) & \qquad & p(1) \times q(1)
\end{array}
$$

i.e. $(p \stackrel{\cdot}{+} q)(1) \cong \dot{p}(1) + \dot{q}(1)$ and $(p \stackrel{\cdot}{\otimes} q)(1) \cong \dot{p}(1) \times \dot{q}(1)$.

---

[2]The only reason we introduce $\times$ for $\mathbf{Poly}$ is to explain that the polynomial product $\dot{p}y$ is in fact the categorical product $\dot{p}y \cong \dot{p} \times y$.

The ⊗-structure distributes over the + structure:

$$p \otimes (q_1 + q_2) \cong (p \otimes q_1) + (p \otimes q_2),$$

thus making $(\mathbf{Poly}, 0, +, y, \otimes)$ a distributive monoidal category, and in particular a rig monoidal category.

Remark 2.6 (Leibniz and chain rules)**.** Some readers may be interested in the Leibniz rule and chain rule, that

$$\dot{\overline{(p \times q)}} \cong \dot{p} \times q + p \times \dot{q}$$
$$\dot{\overline{(p \triangleleft q)}} \cong (\dot{p} \triangleleft q) \times \dot{q}$$

where $\times$ is the categorical product and $\triangleleft$ is the composition product in $\mathbf{Poly}$. These hold, but we will not need them in this paper. ◊

3 Main results

We divide this section into two parts. Section 3.1 is the category theory part, in which we provide what seems to be a novel symmetric monoidal structure on $\mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}$ and show that both $p \mapsto \dot{p}y$ and $q \mapsto (q(1), \Gamma(q))$ are rig functors. At the end of this section, we will have a rig functor $\hbar \colon \mathbf{Poly}^{\mathbf{Cart}} \to \mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}$ that does the categorical work of Shannon entropy.

Section 3.2 is the finishing step, providing a function $\mathrm{Ob}(\mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}) \to \mathbb{R}$ and showing that when it is combined with the above, the map $H \colon \mathrm{Ob}(\mathbf{Poly}^{\mathbf{Cart}}) \to \mathbb{R}$ sends an appropriately finite polynomial p to the Shannon entropy of the empirical distribution defined by p.

3.1 Categorical entropy of a polynomial

Below we will often denote products of sets by juxtaposition, $AB := A \times B$. Recall the functor $p \mapsto \dot{p}y$ from Proposition 2.4.

Proposition 3.1. *The functor $p \mapsto \dot{p}y$ is a rig functor $\mathbf{Poly}^{\mathbf{Cart}} \to \mathbf{Poly}^{\mathbf{Cart}}$.*

Proof. Clearly $\dot{0} = 0$ and $\dot{\overline{(p + q)}} \cong \dot{p} + \dot{q}$, and by multiplying both sides by y we see that the functor $p \mapsto py$ preserves the coproduct structure. There is an isomorphism $\dot{y}y \cong y$, and for any $p, q \in \mathbf{Poly}^{\mathbf{Cart}}$ there is also an isomorphism $\dot{\overline{(p \otimes q)}}y \cong (\dot{p}y) \otimes (\dot{q}y)$, as follows from (7) and (8); thus $p \mapsto \dot{p}y$ preserves the ⊗-structure. All of these isomorphisms are natural in $p, q \in \mathbf{Poly}^{\mathbf{Cart}}$, completing the proof. □

The following corollary is straightforward, since $\mathbf{Poly}^{\mathbf{Cart}}$ inherits + and ⊗ from the forgetful functor $\mathbf{Poly}^{\mathbf{Cart}} \to \mathbf{Poly}$.

Corollary 3.2. *The functor $T(p) := \dot{p}y$ is a rig functor $T \colon \mathbf{Poly}^{\mathbf{Cart}} \to \mathbf{Poly}$.*

Remark 3.3 (Total polynomial). Note that for any p we have $(\dot{p}y)(1) \cong \dot{p}(1)$. We think of $\dot{p}y$ as the *total polynomial* of p, akin to the total space of a bundle, where p is playing the role of the base. To justify this intuition, note that $\dot{p}y$ comes with a "projection" map $\epsilon \colon \dot{p}y \to p$ and that a section $p \to \dot{p}y$ of ϵ can be identified with a section $\gamma \in \Gamma(p)$ of p as a bundle; see Remark 2.5 and Example 2.3. ◊

Example 3.4. For any polynomial p, we have

$$\Gamma(\dot{p}y) \cong \prod_{I \in p(1)} p[I]^{p[I]}.$$

This formula—which follows directly from Eqs. (4) and (7)—will be relevant when connecting the category theory to Shannon entropy later on. ◊

Proposition 3.5. *The category* $\mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}$ *has a distributive monoidal structure:*

$$(A_1, B_1) + (A_2, B_2) := (A_1 + A_2 \,, B_1 B_2) \tag{9}$$

$$(A_1, B_1) \otimes (A_2, B_2) := (A_1 A_2 \,, B_1^{A_2} B_2^{A_1}) \tag{10}$$

The units are $(0, 1)$ *and* $(1, 1)$ *respectively.*

Proof. Coproducts in $\mathbf{Set}^{\mathrm{op}}$ are products in $\mathbf{Set}$, justifying the first line; these clearly form a symmetric monoidal structure. For the $\otimes$-monoidal structure, note that the formula is functorial in $A \in \mathbf{Set}$ and $B \in \mathbf{Set}^{\mathrm{op}}$. It is also symmetric as well as unital: $(1, 1) \otimes (A_2, B_2) \cong (A_2, B_2)$. Associativity is justified as follows:

$$(A_1, B_1) \otimes ((A_2, B_2) \otimes (A_3, B_3)) \cong (A_1 A_2 A_3, B_1^{A_2 A_3} B_2^{A_1 A_3} B_3^{A_1 A_2})$$
$$\cong ((A_1, B_1) \otimes (A_2, B_2)) \otimes (A_3, B_3).$$

There is an absorption map $(0, 1) \otimes (A, B) \cong (0, B) \to (0, 1)$, and the distributivity of $\otimes$ over $+$ is justified as follows:

$$(A, B) \otimes ((A_1, B_1) + (A_2, B_2)) \cong (A(A_1 + A_2), B^{A_1 + A_2}(B_1 B_2)^A)$$
$$\cong (AA_1 + AA_2, B^{A_1} B^{A_2} B_1^A B_2^A)$$
$$\cong ((A, B) \otimes (A_1, B_1)) + ((A, B) \otimes (A_2, B_2)).$$

We leave the remaining details to the interested reader. □

Remark 3.6 (Formal roots and rectangular aspect). One can think of an object $(A, B) \in \mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}$ as formally representing the Ath root of B, i.e. the number $\sqrt[A]{B} = B^{\frac{1}{A}}$, keeping track of the base A as well. It is helpful to think of (A, B) as a rectangle with length A and width $\sqrt[A]{B}$. From this perspective, the sum from (9) adds the lengths and takes the geometric mean of the widths, and the monoidal product from (10) takes the product of both lengths and widths:

$$(B_1 B_2)^{\frac{1}{A_1 + A_2}} = \left(\left(\sqrt[A_1]{B_1} \right)^{A_1} \times \left(\sqrt[A_2]{B_2} \right)^{A_2} \right)^{\frac{1}{A_1 + A_2}} \qquad \text{and} \qquad (B_1^{A_2} B_2^{A_1})^{\frac{1}{A_1 A_2}} = \sqrt[A_1]{B_1} \sqrt[A_2]{B_2}.$$

For any polynomial p, the functor $R(p) := (p(1), \Gamma(p))$ from (5) is consonant with this interpretation. We may say that $R(p)$ denotes the *rectangular aspect* of p in the sense that $p(1)$ represents the length and $\sqrt[p(1)]{\Gamma(p)}$, the geometric mean of the fiber cardinalities, represents the width. For example, the polynomial $p = y^4 + 4y$, depicted in Diagram (1), has length $p(1) = 5$ and width $\sqrt[5]{4} \approx 1.3$. $\diamond$

Remark 3.7. The $\otimes$ operation (10) on $\mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}$ in fact has a closure

$$[(A_1, B_1), (A_2, B_2)] := \left(A_2^{A_1} B_1^{B_2}, A_1 B_2 \right).$$

We will not need this, but it is interesting. $\diamond$

Proposition 3.8. *The functor* $R\colon \mathbf{Poly} \to \mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}$ *from* (5) *is a rig functor.*

Proof. Recall from (5) that $R(p) := (p(1), \Gamma(p))$. Clearly $0(1) = 0$ and $(p + q)(1) \cong p(1) + q(1)$. Also $\Gamma(0) = 1$ and $\Gamma(p + q) \cong \Gamma(p) \times \Gamma(q)$; hence R preserves the $(0, +)$ monoidal structure. Moreover, we have $y(1) = 1$ and $(p \otimes q)(1) \cong p(1) \times q(1)$ and $\Gamma(y) = 1$, so to show that R preserves the $(y, \otimes)$ monoidal structure, it remains only to provide an isomorphism

$$\Gamma(p \otimes q) \cong \Gamma(p)^{q(1)} \times \Gamma(q)^{p(1)}.$$

It is given as follows:

$$
\begin{aligned}
\Gamma(p \otimes q) &\cong \prod_{(I,J) \in p(1) \times q(1)} p[I]q[J] \\
&\cong \left(\prod_{(I,J) \in p(1) \times q(1)} p[I] \right) \times \left(\prod_{(I,J) \in p(1) \times q(1)} q[J] \right) \\
&\cong \prod_{J \in q(1)} \prod_{I \in p(1)} p[I] \times \prod_{I \in p(1)} \prod_{J \in q(1)} q[J] \\
&\cong \Gamma(p)^{q(1)} \times \Gamma(q)^{p(1)}
\end{aligned}
$$

$\square$

We summarize the above section before we go on to the final one. Namely, the functors $T\colon \mathbf{Poly}^{\mathbf{Cart}} \to \mathbf{Poly}$ and $R\colon \mathbf{Poly} \to \mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}$ from Corollary 3.2 and Proposition 3.8 compose to form a rig functor $\hbar := R \circ T$ given by

$$
\begin{aligned}
\mathbf{Poly}^{\mathbf{Cart}} &\xrightarrow{\hbar} \mathbf{Set} \times \mathbf{Set}^{\mathrm{op}} \\
p &\mapsto (\dot{p}(1), \Gamma(\dot{p}y)).
\end{aligned}
\tag{11}
$$

We refer to $\hbar(p) \in \mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}$ as *the categorical entropy* of the polynomial p. This pair of sets leaves behind any semblance of the probability distribution associated with p, but it retains the data necessary to compute p's entropy—as we'll see in Theorem 3.13—and it is rig-functorial in p.

3.2 Shannon entropy

Writing log to denote $\log_2$, we define a partial function $L\colon \mathrm{Ob}(\mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}) \to \mathbb{R}$ by

$$(A, B) \mapsto \log A - \frac{\log B}{A}. \tag{12}$$

Equivalently, $L(A, B) = \log A - \log \sqrt[A]{B}$. When $A = 0$ and $B = 1$, we define this function to be $L(0, 1) := 0$; for all cases where $A = 0$, or $B = 0$, or either A or B is infinite, we leave $L(A, B)$ undefined. We will be only interested in this map when it is composed with the categorical entropy $\hbar$ from (11), and Lemma 3.9 below says that we do not need to worry about the undefined cases.

Lemma 3.9. *Let $p \in \mathbf{Poly}^{\mathbf{Cart}}$ with categorical entropy $(A, B) := \hbar(p)$, and suppose that $^{\#}\dot{p}(1) < \infty$. Then we have that*
  i. $B \neq 0$,
 ii. *if $A = 0$ then $B = 1$, and*
iii. *both A and B are finite.*

Proof. By definition of $\hbar$, we have that $A := \dot{p}(1)$ and $B := \Gamma(\dot{p}y)$.
  i. One easily checks using (4) that for any $q \in \mathbf{Poly}$, the set $\Gamma(qy) \neq 0$ is nonempty since every (qy)-type has at least one term.
 ii. If $\dot{p}(1) = 0$ then $p \in \mathbf{Set}$ is constant, so $\dot{p}y = 0$ as well, and $\Gamma(0) = 1$ by (4).
iii. By assumption $^{\#}A = {}^{\#}\dot{p}(1) < \infty$. For B, note that there are only a finite number of $I \in p(1)$ for which $p[I]$ is nonempty, so by (4) and (6) the set $\Gamma(\dot{p}y)$ is finite. $\quad\square$

Remark 3.10 (Log aspect ratio). With the interpretation of an object $(A, B) \in \mathbf{Set} \times \mathbf{Set}^{\mathrm{op}}$ as a rectangle with length A and width $\sqrt[A]{B}$, as in Remark 3.6, we can think of $L(A, B) = \log A - \log \sqrt[A]{B}$ as its *log aspect ratio*, the log of its length divided by its width. This is a quantity that has come up in the study of vision [TGH11; Dic+17], though we're making no claim about whether this connection is meaningful. $\quad\Diamond$

Definition 3.11 (Empirical distribution). Let $p \neq 0$ be a nonzero polynomial and suppose that the cardinality of $\dot{p}(1) \in \mathbf{Set}$ is finite, $^{\#}\dot{p}(1) < \infty$. We define the *empirical distribution defined by p* to be the following function $P\colon p(1) \to [0, 1]$:

$$P(I) := \frac{^{\#}p[I]}{^{\#}\dot{p}(1)}$$

Note that $1 = \sum_{I \in p(1)} P(I)$, so P is indeed a probability distribution. $\quad\Diamond$

Remark 3.12. One may ask how to view **Poly**'s monoidal structures, especially $+$ and $\otimes$, under the correspondence from Definition 3.11. Suppose given polynomials $p, q \in \mathbf{Poly}$ with associated probability distributions P_p and P_q. For Dirichlet product we have

$$P_{p \otimes q} = P_p \otimes P_q$$

where the left-hand side is the probability distribution associated to $p \otimes q$ and the right-hand side is the usual tensor (independent) product of probability distributions. For sums we have

$$P_{p+q} = \frac{\dot{p}(1)}{\dot{p}(1) + \dot{q}(1)} P_p + \frac{\dot{q}(1)}{\dot{p}(1) + \dot{q}(1)} P_q$$

the convex combination of P_p and Q_q, weighted according to the relative number of draws $\dot{p}(1)$ and $\dot{q}(1)$ in each. $\diamond$

Recall that the Shannon entropy $H(P)$ of a probability distribution $P \colon X \to [0, 1]$ is given by

$$H(P) := -\sum_{x \in X} P(x) \log P(x).$$

The following theorem could be summarized as follows: "thinking of $p \in \textbf{Poly}^{\textbf{Cart}}$ as a statistical sample, the entropy $H(P)$ of the corresponding probability distribution P is equal to the log ratio of the rectangular aspect of p's total polynomial"; see Remarks 3.3, 3.6, and 3.10.

Theorem 3.13. *Let $p \neq 0$ be a nonzero polynomial with $^{\#}\dot{p}(1) < \infty$, and let P be the empirical distribution defined by p. Then the following equation holds*

$$H(P) = L(\hbar(p))$$

where H is the Shannon entropy and $L, \hbar$ are as defined in Eqs. (11) and (12).

Proof. We need to show that the following holds:

$$H(P) = \log \dot{p}(1) - \frac{\log \Gamma(\dot{p}y)}{\dot{p}(1)}.$$

With the fact $\Gamma(\dot{p}y) \cong \prod_{I \in p(1)} p[I]^{p[I]}$ from Example 3.4 in hand, this is a routine calculation:

$$\begin{aligned}
H(P) &:= -\sum_{I \in p(1)} \frac{^{\#}p[I]}{^{\#}\dot{p}(1)} \log \frac{^{\#}p[I]}{^{\#}\dot{p}(1)} \\
&= \frac{1}{^{\#}\dot{p}(1)} \sum_{I \in p(1)} {}^{\#}p[I] \left(\log {}^{\#}\dot{p}(1) - \log {}^{\#}p[I] \right) \\
&= \frac{1}{^{\#}\dot{p}(1)} \left({}^{\#}\dot{p}(1) \log {}^{\#}\dot{p}(1) - \log \prod_{I \in p(1)} {}^{\#}p[I]^{{}^{\#}p[I]} \right) \\
&= \log {}^{\#}\dot{p}(1) - \frac{\log \Gamma(\dot{p}y)}{\dot{p}(1)}
\end{aligned}$$

$\square$

Example 3.14 (Entropy of a uniform distribution). It is well-known and easy to calculate that if P is a uniform distribution on A elements, then $H(P) = \log(A)$. There are many samples that correspond to P; what differs are their sample sizes.

The sample in which AB-many observations are taken—each outcome occurring B-many times—corresponds to the polynomial Ay^B.

Our formula for entropy needs to agree, and it does. The rectangular aspect of the total polynomial is $\hbar(p) \cong (AB, B^{AB})$: length AB and width $B = \sqrt[AB]{B^{AB}}$, so its log aspect ratio is

$$L(\hbar(p)) = \log(AB) - \frac{\log(B^{AB})}{AB} = \log A. \qquad \diamond$$

References

[AAG03]   Michael Gordon Abbott, Thorsten Altenkirch, and Neil Ghani. "Categories of Containers". In: *FoSSaCS*. 2003 (cit. on p. 2).

[ACH19]   Jeremy Avigad, Mario Carneiro, and Simon Hudon. "Data types as quotients of polynomial functors". In: *10th International Conference on Interactive Theorem Proving (ITP 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 2019 (cit. on p. 2).

[AN18]   Steve Awodey and Clive Newstead. *Polynomial pseudomonads and dependent type theory*. 2018. arXiv: 1802.00997 [math.CT] (cit. on p. 2).

[BD96]   Richard Bird and Oege De Moor. "The algebra of programming". In: *NATO ASI DPD*. 1996, pp. 167–203 (cit. on p. 2).

[BF14]   John C. Baez and Tobias Fritz. "A Bayesian characterization of relative entropy". In: *Theory and Applications of Categories* 29 (2014), Paper No. 16, 421–456 (cit. on p. 4).

[BFL11]   John C. Baez, Tobias Fritz, and Tom Leinster. "A Characterization of Entropy in Terms of Information Loss". In: *Entropy* 13.11 (2011), pp. 1945–1957. URL: https://www.mdpi.com/1099-4300/13/11/1945 (cit. on p. 4).

[Dic+17]   J Edwin Dickinson, Sarah K Morgan, Matthew F Tang, and David R Badcock. "Separate banks of information channels encode size and aspect ratio". In: *Journal of Vision* 17.3 (2017), pp. 27–27 (cit. on p. 11).

[GK12]   Nicola Gambino and Joachim Kock. "Polynomial functors and polynomial monads". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 154.1 (Sept. 2012), pp. 153–192 (cit. on p. 4).

[Lei21]   Tom Leinster. *Entropy and Diversity: The Axiomatic Approach*. Cambridge University Press, 2021 (cit. on p. 4).

[Par22]   Arthur Parzygnat. "A functorial characterization of von Neumann entropy". In: *Cahiers* 63 (1 2022) (cit. on p. 4).

[SH21]   David I. Spivak and Timothy Hosgood. "Dirichlet Polynomials and Entropy". In: *Entropy* 23.8 (2021). URL: https://www.mdpi.com/1099-4300/23/8/1085 (cit. on p. 4).

[Sha21]   Brandon Shapiro. "Familial Monads as Higher Category Theories". In: *arXiv preprint arXiv:2111.14796* (2021) (cit. on p. 2).

[Sha48]   C. E. Shannon. "A mathematical theory of communication". In: *The Bell System Technical Journal* 27.3 (1948), pp. 379–423. DOI: `10.1002/j.1538-7305.1948.tb01338.x` (cit. on p. 2).

[SN22]   David I. Spivak and Nelson Niu. *Polynomial functors: a general theory of interaction. In preparation*. 2022 (cit. on pp. 2, 4).

[Spi20]   David I. Spivak. "Poly: An abundant categorical setting for mode-dependent dynamics". In: (2020). eprint: `arXiv:2005.01894` (cit. on p. 2).

[Spi21]   David I. Spivak. *Functorial aggregation*. 2021. arXiv: `2111.10968 [math.CT]` (cit. on p. 2).

[SW15]   David I. Spivak and Ryan Wisnesky. "Relational Foundations for Functorial Data Migration". In: *Proceedings of the 15th Symposium on Database Programming Languages*. DBPL. Pittsburgh, PA: ACM, 2015, pp. 21–28 (cit. on p. 2).

[TCM19]   Cédric Ho Thanh, Pierre-Louis Curien, and Samuel Mimram. "A sequent calculus for opetopes". In: *2019 34th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*. IEEE. 2019, pp. 1–12 (cit. on p. 2).

[TGH11]   Justin Talbot, John Gerth, and Pat Hanrahan. "Arc length-based aspect ratio selection". In: *IEEE transactions on visualization and computer graphics* 17.12 (2011), pp. 2276–2282 (cit. on p. 11).