AN ALGEBRAIC APPROACH TO GENERALIZED MEASURES OF INFORMATION

Daniel Halpern-Leistner
6/20/08

Abstract. I propose an algebraic framework in which to study measures of information. One immediate consequence is a new proof of the classification of branching entropies on a Boolean algebra. Common constructions in the theory of entropy such as the conditional entropy and the mutual information arise naturally in this framework. A generalized theory of information can be interpreted in a category-theoretic context as a natural transformation between two functors.

In 1948, Claude E. Shannon established the statistical theory of entropy (Shannon 1948), which offered deep insights into coding theory and compression. Since then researchers have been generalizing the axiomatic structure of the Shannon entropy so that they might bring it to bear on other mathematical questions.

One such generalization is the notion of a branching inset entropy (Kannappan & Sander 2004). In 1983, J. Aczel named the problem of classifying inset information measures on a list of unsolved problems in the theory of functional equations at the Twenty-First Symposium on Functional Equations. The problem was solved (Ebanks 1986, Ḃ.R. Ebanks & Ng 1990), but the proof is 20 pages long and involves checking over a dozen special cases.

Below we introduce an algebraic object, the partition algebra, and we identify the space of branching entropies as the dual of this object. In addition to producing a new and simple algebraic proof of the classification of branching entropies, this construction endows the space of branching entropies with a rich algebraic structure with information theoretic interpretations and leads to a notion of a generalized information theory.

1. Definition of the partition algebra

We fix an arbitrary base ring k, but for most applications we can consider $k = \mathbb{R}$. For any boolean algebra $\mathcal{F}$, we consider the space of formal linear combinations of elements of $\mathcal{F}$ with coefficients in k, where we take the element $\emptyset \in \mathcal{F}$ to be zero. We can extend the intersection operation " $\cap$ " bilinearly to define a multiplication of formal sums. This k-module, along with the multiplication, is a k-algebra denoted $k_0\mathcal{F}$.

Define the k-algebra $\mathfrak{A}_k\mathcal{F}$ to be the quotient of $k_0\mathcal{F}$ by the ideal generated by expressions of the form $a \cup b + a \cap b - a - b$ for all $a, b \in \mathcal{F}$. In other words $\mathfrak{A}_k\mathcal{F}$ is the space of formal linear combinations of elements of $\mathcal{F}$, but where we consider $a \cup b = a + b$ whenever $a, b \in \mathcal{F}$ are disjoint.

We refer to $\mathfrak{A}_k\mathcal{F}$ as the "algebra of simple functions on $\mathcal{F}$", because if $\mathcal{F}$ is a boolean algebra of subsets of some set X, then $\mathfrak{A}_k\mathcal{F}$ is just the algebra of k-valued functions on X whose range is a finite set in k and whose level sets are elements of $\mathcal{F}$. We see this by identifying the formal element $E \in \mathcal{F} \subseteq \mathfrak{A}_k\mathcal{F}$ with the indicator function χ_E for all subsets $E \in \mathcal{F}$.

Denote the surjection onto the quotient algebra $\pi_{\mathcal{F}} : k_0\mathcal{F} \to \mathfrak{A}_k\mathcal{F}$. This surjection is natural in the sense that if $\phi : \mathcal{F} \to \mathcal{G}$ is a map of boolean algebras, then there is a

corresponding map of short exact sequences

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \ker \pi_{\mathcal{F}} & \longrightarrow & k_0\mathcal{F} & \xrightarrow{\pi_{\mathcal{F}}} & \mathfrak{A}_k\mathcal{F} & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \phi_*} & & \downarrow{\scriptstyle \phi_*} & & \downarrow{\scriptstyle \phi_*} & & \\ 0 & \longrightarrow & \ker \pi_{\mathcal{G}} & \longrightarrow & k_0\mathcal{G} & \xrightarrow{\pi_{\mathcal{G}}} & \mathfrak{A}_k\mathcal{G} & \longrightarrow & 0 \end{array}$$

Finally we define the "partition algebra" as the smallest subalgebra of $k_0\mathcal{F}$ containing $\ker \pi_{\mathcal{F}}$. So we have

$$\mathfrak{P}_k\mathcal{F} := \ker \pi_{\mathcal{F}} \oplus k \cdot \mathbf{1} \subseteq k_0\mathcal{F}$$

The projection $\pi_{\mathcal{F}}$ induces an augmentation map $\pi_{\mathcal{F}} : \mathfrak{P}_k\mathcal{F} \to k$, and a map of boolean algebras $\mathcal{F} \to \mathcal{G}$ naturally induces a map of augmented k-algebras $\mathfrak{P}_k\mathcal{F} \to \mathfrak{P}_k\mathcal{G}$. The space $\ker \pi_{\mathcal{F}}$ is a $k_0\mathcal{F}$-module and hence a $\mathfrak{P}_k\mathcal{F}$-module.

1.1. Calculation in $\mathfrak{P}_k\mathcal{F}$ and $\mathfrak{A}_k\mathcal{F}$

The main tool for calculation in the algebra of simple functions $\mathfrak{A}_k\mathcal{F}$ is the existence of a canonical form

Proposition 1..1. *For any element $f \in \mathfrak{A}_k\mathcal{F}$ there is a unique minimal (i.e. fewest terms) expression $f = \lambda_1 a_1 + \cdots + \lambda_n a_n$ such that $\lambda_i \in k$ are non-zero and the elements $a_i \in \mathcal{F}$ are pairwise disjoint.*

This proposition makes rigorous the analogy between elements of $\mathfrak{A}_k\mathcal{F}$ and "functions", and many arguments can be made using this analogy.

Remark 1..2. Another way to make this analogy rigorous when k is a field is to consider the spectrum $\operatorname{Spec} \mathfrak{A}_k\mathcal{F}$, then the boolean algebra $\mathcal{F}$ is canonically isomorphic to the lattice of open-and-closed subsets of this space. When $k = \mathbb{Z}/2$ this is just Stone's theorem as presented in (M.F. Atiyah 1969)

Next we introduce a notation for calculation in the algebra $\mathfrak{P}_k\mathcal{F}$. If $\omega = \{\omega_1, \ldots, \omega_n\}$ is an unordered n-tuple of pairwise disjoint elements of $\mathcal{F}$ (we ignore the zero element $\emptyset \in \mathcal{F}$), i.e. $\omega_i \cap \omega_j = \emptyset$ for $i \neq j$, then we define

$$\omega^* := \omega_1 + \cdots + \omega_n \in k_0\mathcal{F}$$
$$\omega^\sharp := (\omega_1 \cup \cdots \cup \omega_n) - \omega^* \in \mathfrak{P}_k\mathcal{F}$$

We call such an n-tuple a partition of $\mathcal{F}$ and say that ω is full if $\omega_1 \cup \cdots \cup \omega_n = \mathbf{1}$. If ω is full then $\omega^* \in \mathfrak{P}_k\mathcal{F} \subseteq k_0\mathcal{F}$ and $\omega^\sharp = 1 - \omega^*$ by definition. Furthermore in this case $\pi_{\mathcal{F}}(\omega^*) = 1$.

Next if ω and ξ are two partitions, we can define their join to be the tuple of pairwise intersections

$$\omega \vee \xi := \{\omega_1 \cap \xi_1, \omega_1 \cap \xi_2, \ldots, \ldots, \omega_m \cap \xi_n\}$$

For any two partitions we have $\omega^*\xi^* = (\omega \vee \xi)^*$, and $\omega^\sharp$ and ω^* are orthogonal idempotents, i.e.

$$(\omega^*)^2 = \omega^*, \quad (\omega^\sharp)^2 = \omega^\sharp, \quad \omega^*\omega^\sharp = 0$$

Finally if both ω and ξ are full partitions then $\omega^\sharp \xi^\sharp = \omega^\sharp + \xi^\sharp - (\omega \vee \xi)^\sharp$.

There is a correspondence between full partitions of $\mathcal{F}$ and finite boolean subalgebras of $\mathcal{F}$. If $\mathcal{G} \subseteq \mathcal{F}$ is a finite boolean subalgebra, then it has a set of irreducible generators (sometimes called "atoms"). The set of atoms is the unique collection $a_1, \ldots, a_n \in \mathcal{G}$ which are distinct, nonzero generators and such that

$$a_i \cap x = a_i \text{ or } \emptyset \text{ for all } x \in \mathcal{G}$$

In particular $a_1 \cup \cdots \cup a_n = \mathbf{1}$ and the a_i are pairwise disjoint. We can associate to $\mathcal{G}$ the full partition $\{a_1, \ldots, a_n\}$, and this establishes a bijection between finite subalgebras and full partitions. Sometimes it is convenient to confuse ω with the subalgebra it generates and vice versa. In the context of subalgebras the join $\omega \vee \xi$ is the smallest boolean subalgebra of $\mathcal{F}$ containing both ω and ξ, i.e. the subalgebra generated by the two subalgebras ω and ξ.

One important fact about the partition algebra is the following

Proposition 1..3. *Let k be a ring and $\mathcal{F}$ a boolean algebra. The space $\ker \pi_{\mathcal{F}}$ is generated as a k-module by the set of $\omega^\sharp$, where $\omega = \{\omega_1, \omega_2\}$ ranges over all partitions of $\mathcal{F}$ with just two elements.*

This is an immediate consequence of the Proposition 2..2 below, so we will postpone the proof until then.

Corollary 1..4. *The space $\ker \pi_{\mathcal{F}}$ is generated as a k-module by the set of $\omega^\sharp$, where $\omega = \{\omega_1, \omega_2, \omega_3\}$ ranges over all* full *partitions of $\mathcal{F}$ with three elements. The space $\mathfrak{P}_k \mathcal{F}$ is generated as a k-modules by the set of ω^*, where ω ranges over all full partitions of $\mathcal{F}$ with three elements.*

2. The relation to entropy

The algebraic objects defined in the previous section capture many important structures in information theory and measure theory. Dualizing the relationship between the partition algebra and the algebra of simple functions will yield a relationship between finitely additive measures and so-called "branching entropies" on a boolean algebra.

2.1. Additive measures

Given a boolean algebra $\mathcal{F}$, the space $\mathrm{M}_k(\mathcal{F})$ of additive measures on $\mathcal{F}$ is defined to be the k-module of functions $\mu : \mathcal{F} \to k$ such that $\mu(\emptyset) = 0$ and

$$\mu(a \cup b) = \mu(a) + \mu(b) \text{ whenever } a \cap b = \emptyset$$

Note that these are finitely additive measures and need not be countably additive even when $\mathcal{F}$ is a σ-algebra. One can see that the space $\mathrm{M}_k(\mathcal{F})$ is naturally identified with the k dual of the space $\mathfrak{A}_k \mathcal{F}$, i.e. the space $(\mathfrak{A}_k \mathcal{F})^*$ of k-linear maps $\mathfrak{A}_k \mathcal{F} \to k$.

In the case $k = \mathbb{R}$ we can relate this definition to the usual notion of a measure on a measure space. First of all, a "measure" is usually considered to take nonnegative values, i.e. $\mu(a) \geq 0$ for all $a \in \mathcal{F}$. We can express this condition in algebraic terms by saying that the symmetric bilinear form on $\mathfrak{A}_{\mathbb{R}} \mathcal{F}$ given by

$$(f, g)_\mu := \mu(f \cdot g)$$

is positive semi-definite. One can use Proposition 1..1 to show that μ nonnegative is equivalent to $(\bullet, \bullet)_\mu$ positive semi-definite. We denote by $\mathrm{M}^+_{\mathbb{R}}(\mathcal{F})$ the space of nonnegative measures on $\mathcal{F}$ – it is closed under addition and multiplication by positive scalars (i.e. it is a semigroup under addition with a compatible $\mathbb{R}^+$ group action).

Classically $(\bullet, \bullet)_\mu$ is just the L^2 inner product on the algebra of simple functions. From this algebraic perspective, the classical Jordan decomposition theorem – which states that a signed measure μ can be written uniquely as the difference of two orthogonal nonnegative measures $\mu = \mu_1 - \mu_2$ – can be seen as a decomposition of a symmetric bilinear form as the difference of two orthogonal nonnegative forms (i.e. such that the null space of μ is precisely the intersection of the null spaces of μ_1 and μ_2).

The inner product perspective gives us yet another way to think about measures on the space $\mathfrak{A}_k\mathcal{F}$. It can be shown that $\mu \mapsto (\bullet, \bullet)_\mu$ gives a one-to-one onto correspondence between measures on $\mathcal{F}$ and inner products on $\mathfrak{A}_k\mathcal{F}$ under which the action of $k_0\mathcal{F}$ on $\mathfrak{A}_k\mathcal{F}$ is self-adjoint.

Remark 2..1. In order to discuss countably additive measures on a σ-algebra, we must introduce a suitable topology on the vector space $\mathfrak{A}_{\mathbb{R}}\mathcal{F}$. Then the space of countably additive measures can be constructed as the space of continuous linear functionals on $\mathfrak{A}_{\mathbb{R}}\mathcal{F}$.

2.2. Branching entropies

The notion of a (symmetric) branching entropy is a common generalization of the Shannon entropy and several other information measures. For a particular boolean algebra $\mathcal{F}$ and base ring k, a branching entropy h is a k-valued function on the set of all partitions of $\mathcal{F}$ which satisfies the following axioms

- $h(\omega_1, \ldots, \omega_n) = h(\emptyset, \omega_1, \ldots, \omega_n)$
- $h(\omega_1, \ldots, \omega_{n+1}) = h(\omega_1, \omega_2) + h(\omega_1 \cup \omega_2, \omega_3, \ldots, \omega_{n+1})$

In particular these axioms imply that for a one-element partition we have $h(\omega_1) = 0$. The space of all such branching entropies is a k-module which we will denote as $\mathrm{Br}_k(\mathcal{F})$. A map of boolean algebras $\phi : \mathcal{F} \to \mathcal{G}$ induces a map of k-modules $\phi^* : \mathrm{Br}_k(\mathcal{G}) \to \mathrm{Br}_k(\mathcal{F})$ and $(\phi \circ \varphi)^* = \varphi^* \circ \phi^*$.

As with additive measures we can consider the k-module

$$\mathrm{Rel}_k(\mathcal{F}) := \left\{\text{formal sums of partitions } \lambda_1\omega^{(1)} + \cdots \lambda_m\omega^{(m)}\right\} / \sim$$
$$\text{where } \begin{array}{l} \{\emptyset, \omega_1, \ldots, \omega_n\} \sim \{\omega_1, \ldots, \omega_n\} \\ \{\omega_1, \ldots, \omega_{n+1}\} \sim \{\omega_1, \omega_2\} + \{\omega_1 \cup \omega_2, \omega_3, \ldots, \omega_{n+1}\} \end{array}$$

and one can see that $\mathrm{Br}_k(\mathcal{F})$ is naturally identified with the k dual of $\mathrm{Rel}_k(\mathcal{F})$, i.e. $\mathrm{Br}_k(\mathcal{F}) \cong (\mathrm{Rel}_k(\mathcal{F}))^*$ naturally. We have the following theorem

Proposition 2..2. *The assignment $\omega \mapsto \omega^\sharp$ induces a natural isomorphism of k-modules $\mathrm{Rel}_k(\mathcal{F}) \xrightarrow{\cong} \ker \pi_{\mathcal{F}}$.*

Corollary 2..3. *There is a natural isomorphism of k-modules $\mathrm{Br}_k(\mathcal{F}) \cong (\ker \pi_{\mathcal{F}})^*$.*

Proof. $k_0\mathcal{F}$ is the union of the subalgebras $k_0\xi \subseteq k_0\mathcal{F}$ where ξ ranges over all Boolean subalgebras of $\mathcal{F}$. The map $\omega \to \omega^\sharp$ is natural with respect to these inclusions and the corresponding inclusions $\mathrm{Rel}_k(\xi) \subseteq \mathrm{Rel}_k(\mathcal{F})$, so it suffices to show that the map is an isomorphism for $\mathcal{F} = \xi$ finite.

We equip $\mathrm{Rel}_k(\mathcal{F})$ with a natural $k_0\mathcal{F}$-module structure (hence a $\mathfrak{P}_k\mathcal{F}$-module structure) via the action

$$a \cdot \{\omega_1, \ldots, \omega_n\} := \{a \cap \omega_1, \ldots, a \cap \omega_n\}$$

And for finite $\mathcal{F} = \xi$ the space $\ker \pi_\xi$ is generated as a $\mathfrak{P}_k\xi$-module by $\xi^\sharp$ because $\mathfrak{A}_k\xi$ can be naturally identified with $\xi^* \cdot k_0\xi$ and the projection $k_0\xi \to \mathfrak{A}_k\xi$ is just multiplication by ξ^*. It follows that for finite ξ we can define an inverse explicitly by $u \cdot \xi^\sharp \mapsto u \cdot \xi$ for $u \in \mathfrak{P}_k\xi$. □

It can be shown that the spaces $\mathfrak{A}_k\mathcal{F}$ and $\mathfrak{P}_k\mathcal{F}$ are both projective as k-modules A, and thus taking the dual spaces gives a short exact sequence

$$0 \to \mathrm{M}_k(\mathcal{F}) \to (k_0\mathcal{F})^* \to \mathrm{Br}_k(\mathcal{F}) \to 0$$

This short exact sequence classifies the space of branching entropies. Explicitly it states that any branching entropy is of the form

$$h(\omega_1, \ldots, \omega_n) = f(\omega^\sharp) = f(\omega_1 \cup \cdots \cup \omega_n) - f(\omega_1) - \cdots - f(\omega_n)$$

for some function $f : \mathcal{F} \to k$ with $f(\emptyset) = 0$. Furthermore, the short exact sequence implies that two functions f, f' induce the same branching entropy if and only if their difference $f - f'$ is an additive measure on $\mathcal{F}$.

This gives a simple algebraic proof of the classification of branching entropies, and this proof works over a general base ring k and not just $\mathbb{R}$. However, the isomorphism 2..3 does more than just classify the k-module of branching entropies. As we show in the next section, this isomorphism endows $\mathrm{Br}_k(\mathcal{F})$ with a rich algebraic structure.

2.3. The algebraic structure of branching entropies

From the proposition 2..2, the space $\mathrm{Rel}_k \mathcal{F}$ is naturally a $\mathfrak{P}_k(\mathcal{F})$-module. This induces a $\mathfrak{P}_k\mathcal{F}$-module structure on $\mathrm{Br}_k(\mathcal{F})$, given explicitly by

$$\begin{aligned}(\omega^* \cdot h)(\xi) &= h(\omega^*\xi^\sharp) = h((\omega \vee \xi)^\sharp - \omega^\sharp) \\ &= h(\omega \vee \xi) - h(\omega)\end{aligned}$$

Here we have assumed that ω and ξ are both full partitions. In classical information theory the quantity $H(\omega \vee \xi) - H(\omega)$ is *defined* to be the conditional entropy $H(\xi|\omega)$, although here it arises as the natural action of the partition algebra on the space of branching entropies. In the classical notation we write this action as $(\omega^* \cdot h)(\bullet) = h(\bullet|\omega)$.

Standard formulas involving conditional entropy, such as the fact that $h(\zeta|\omega \vee \xi) = h(\zeta \vee \xi|\omega) - h(\xi|\omega)$ follow immediately from this interpretation of conditional entropy.

Note: The space $\ker \pi_\mathcal{F}$ is a $k_0\mathcal{F}$-module in addition to being a $\mathfrak{P}_k\mathcal{F}$-module. Thus studying the action of all of $k_0\mathcal{F}$ on $\mathrm{Br}_k(\mathcal{F})$ provides a new more general notion of conditional entropy which does not seem to appear in the classical theory.

Given a branching entropy $h \in (\ker \pi_{\mathcal{F}})^*$ we define a symmetric bilinear form $(u, v)_h := h(u \cdot v)$ for $u, v \in \ker \pi_{\mathcal{F}}$. As usual we can think of $|u|^2 = (u, u)$ as a kind of "pseudo-norm" on $\ker \pi_{\mathcal{F}}$, although one should exercise caution as to when this "norm" is non-negative (see below). This symmetric bilinear form gives rise to many of the classical quantities studied in information theory

$$\begin{aligned} (\omega^\sharp, \xi^\sharp) &= h(\omega) + h(\xi) - h(\omega \vee \xi) \\ |\omega^* \xi^\sharp|^2 &= h(\xi|\omega) \\ |\omega^\sharp - \xi^\sharp|^2 &= 2h(\omega \vee \xi) - h(\omega) - h(\xi) \\ &= h(\xi|\omega) + h(\omega|\xi) \end{aligned} \tag{1}$$

The quantity $(\xi^\sharp, \omega^\sharp)$ is usually called the mutual information and denoted $I(\omega; \xi)$. The quantity $\left|\xi^\sharp - \omega^\sharp\right|^2$ is called the Rohklin metric and is sometimes denoted $\rho(\omega, \xi)$.

The short exact sequence 2.2. allows us to define the notion of the complement of a branching entropy. We will need the following lemma

Lemma 1. *Fix a base ring k and let $\mathcal{F}$ be an arbitrary boolean algebra. There exists a finitely additive measure $\mu \in \mathrm{M}_k(\mathcal{F})$ with $\mu(\mathbf{1}) = 1$. If $k = \mathbb{R}$, the measure can be chosen to be nonnegative.*

This is just a rephrasing of the existence of ultrafilters (). The ultrafilter lemma implies that there are boolean algebra homomorphisms $\psi : \mathcal{F} \to \{\emptyset, \mathbf{1}\}$. One can verify that if we consider $\mathbf{1} = 1 \in k$ and $\emptyset = 0 \in k$ then the boolean algebra homomorphism *is* precisely a finitely additive measure.

We summarize the definition and properties of the complement map in the following proposition (we use $[f(x)] \in \mathrm{Br}_k(\mathcal{F})$ to denote the branching entropy represented by the function $f : \mathcal{F} \to k$)

Proposition 2..4. *Let $\mu \in \mathrm{M}_k(\mathcal{F})$ with $\mu(\mathbf{1}) = 1$, then the k-linear map*

$$\begin{aligned} (\bullet)^c : \mathrm{Br}_k(\mathcal{F}) &\to \mathrm{Br}_k(\mathcal{F}) \\ : [f(x)] &\mapsto [f(x^c) - f(\mathbf{1})\mu(x^c)] \end{aligned}$$

is well-defined and independent of the choice of μ. Also $(h^c)^c = h$ for any $h \in \mathrm{Br}_k(\mathcal{F})$.

If we think of Br_k as a functor $\mathrm{Br}_k : \mathbf{Bool}^{op} \to \mathbf{Mod_k}$ from the (opposite) category of boolean algebras to the category of k-modules, then the complement is a natural transformation of functors $(\bullet)^c : \mathrm{Br}_k \Rightarrow \mathrm{Br}_k$.

All of the above structure on $\mathrm{Br}_k(\mathcal{F})$ leads one to interpret a branching entropy h as a sort of "state of knowledge" in a whole configuration space $\mathrm{Br}_k(\mathcal{F})$. Imagine h represents my knowledge of the outcome of a chemistry experiment. The relation $(\omega^* \cdot h)(\bullet) = h(\bullet|\omega)$ allows us to interpret the action of the partition algebra on h as appending facts (i.e. additional data and measurements) to my current state of knowledge h. In the same vein, $\left|\omega^\sharp\right| = h(\omega^\sharp)$ represents the knowledge I stand to gain once I am given the data encoded by ω. Finally the complement h^c represents "complementary" knowledge in the sense that if h were to encode which of several possible events occured in my experiment, the h^c encodes a weighted average of the knowledge of whether each event *failed* to occur.

Example 2..5. For any nonnegative measure $\mu \in \mathrm{M}^+_{\mathbb{R}}(\mathcal{F})$ we can assign the Shannon entropy $H_\mu \in \mathrm{Br}_{\mathbb{R}}(\mathcal{F})$ induced by the function $a \mapsto \eta(\mu(a)) := \mu(a)\log\mu(a)$. Explicitly the Shannon entropy assigns

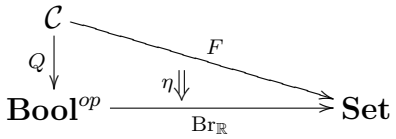
$$H_\mu(\omega) = \eta(\mu(\omega_1 \cup \cdots \cup \omega_n)) - \eta(\mu(\omega_1)) - \cdots - \eta(\mu(\omega_n))$$

This assignment rule $\mu \mapsto H_\mu \in \mathrm{Br}_{\mathbb{R}}(\mathcal{F})$ is the entire content of the classical statistical definition of information. If we consider $\mathrm{Br}_{\mathbb{R}}$ and $\mathrm{M}_{\mathbb{R}}$ as contravariant functors $\mathrm{M}^+_{\mathbb{R}}, \mathrm{Br}_{\mathbb{R}} : \mathbf{Bool}^{op} \to \mathbf{Set}$, then the Shannon entropy is just a natural transformation of functors $S : \mathrm{M}^+_{\mathbb{R}} \to \mathrm{Br}_{\mathbb{R}}$.

3. Generalized measures of information

Motivated by the Shannon entropy we define

Definition 3..1. A **generalized information theory** for the category $\mathcal{C}$ is a triple(F, Q, η), where $F : \mathcal{C} \to \mathbf{Set}$ is a "structure" functor, $Q : \mathcal{C} \to \mathbf{Bool}^{op}$ associates a Boolean algebra to each object of $\mathcal{C}$, and $\eta : F \to \mathrm{Br}_{\mathbb{R}} \circ Q$ is a natural transformation of functors.



The Shannon entropy is purely statistical, $\mathcal{C} = \mathbf{Bool}^{op}$ and $F = \mathrm{M}_{\mathbb{R}}$ so the entropy only depends on a measure on a Boolean algebra. However, we can imagine theories of information that depend on additional structures such as that of a metric space or a group action, etc.. This definition of a generalized information theory is a first step towards extending classical information theory uniformly to other branches of mathematics.

3.1. Nonnegativity of branching entropies

We know from the classical theory that one of the most important properties of the Shannon entropy is the quantities $H_\mu(\omega)$, $I_\mu(\omega;\xi)$, $\rho_\mu(\omega,\xi)$, and $H_\mu(\omega|\xi)$ are nonnegative and even obey some subadditivity laws.

To generalize this call a branching entropy *nonnegative* if the form $(\bullet,\bullet)_h$ is positive semi-definite, and denote by $\mathrm{Br}^+_{\mathbb{R}}(\mathcal{F})$ the space of nonnegative entropies on $\mathcal{F}$. As before, $\mathrm{Br}^+_{\mathbb{R}}(\mathcal{F})$ is a semigroup under addition and has an $\mathbb{R}^+$ group action under scalar multiplication. Similarly we call an information theory nonnegative if the map $\eta : F \to \mathrm{Br}_{\mathbb{R}}\, Q$ lands in the subfunctor $\mathrm{Br}^+_{\mathbb{R}}\, Q$.

Unfortunately the Shannon entropy is *not* nonnegative in this sense, but the complement of the Shannon entropy is.

Define for any finite Boolean algebra ξ the special element $e_\xi := \sum_{a\in\mathcal{F}}(-1)^{|a|}a^c$, where $|a|$ denotes the number of atoms in a. We have the lemma

Lemma 2. *Let $\eta : F \to \mathrm{Br}_{\mathbb{R}}$ be an information theory on $\mathcal{C} = \mathbf{Bool}^{op}$. Then η is nonnegative iff $\eta_{\mathcal{F}}[x](e_{\mathcal{F}}) \geq 0$ for all finite Boolean algebras $\mathcal{F}$ and every $x \in F(\mathcal{F})$.*

Using this lemma

Proposition 3..2. *The complement of the Shannon entropy $H^c : \mu \to H^c_\mu$ is nonnegative.*

These ideas are still in their infancy, but the great range of information-theoretic constructions captured by these algebraic structures suggests that they are a step in the right direction towards a general theory of mathematical complexity. In addition to developing the idea of nonnegativity further, we are studying specific examples of information theories on metric spaces and dynamical systems, and we are studying the relationship between entropy and coding in this general context.

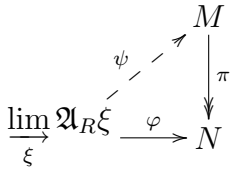
A Proof that $\mathfrak{A}_k\mathcal{F}$ and $\mathfrak{P}_k\mathcal{F}$ are projective k modules

Proposition A.1. *For any boolean algebra $\mathcal{F}$ and any ring R, the R-module $\mathfrak{A}_R\mathcal{F}$ is projective.*

Proof. Any element of $\mathfrak{A}_R\mathcal{F}$ lies in the image of some finite subalgebra $\mathfrak{A}_R\xi \hookrightarrow \mathfrak{A}_R\mathcal{F}$, we have the isomorphism

$$\mathfrak{A}_R\mathcal{F} \cong \varinjlim_{\substack{\xi\subseteq\mathcal{F}\\ |\xi|<\infty}} \mathfrak{A}_R\xi$$

To show $\mathfrak{A}_R\mathcal{F}$ is projective, we must show that any diagram of the form



has a lift ψ. Any such φ is simply a family of homomorphisms $\varphi_\xi : \mathfrak{A}_R\xi \to N$ that are compatible with the inclusion homomorphisms. $\mathfrak{A}_R\xi$ is free on the set of atoms $\{\xi_1, \ldots, \xi_n\}$, so we can think of φ as a compatible family $\varphi_\xi : \{\xi_1, \ldots, \xi_n\} \to N$ of maps of sets. The compatibility condition means that for all $\xi \subseteq \omega$ we have

$$\varphi_\xi(\xi_i) = \sum_{\omega_j \subseteq \xi_i} \varphi_\omega(\omega_j) \in N \tag{2}$$

for all atoms $\xi_1, \ldots, \xi_n$ of ξ.

We seek a family of lifts ψ_ξ that satisfy the relation 2. Define the set $S(\xi)$ for each finite ξ by

$$S(\xi) = \{f : \{\xi_1, \ldots, \xi_n\} \to M \text{ so that } \pi \circ f(\xi_i) = \varphi_\xi(\xi_i)\}$$

and for any $\xi \subseteq \omega \subseteq \mathcal{F}$ we have the maps of sets

$$\begin{aligned} S(\omega) &\to S(\xi) \\ f &\mapsto f'(\xi_i) = \sum_{\omega_j \subseteq \xi_i} f(\omega_j) \end{aligned}$$

where $f' \in S(\xi)$ because π is linear, and thus

$$\pi \circ f'(\xi_i) = \sum_{\omega_j \subseteq \xi_i} \pi \circ f(\omega_j) = \sum_{\omega_j \subseteq \xi_i} \varphi_\omega(\omega_j) = \varphi_\xi(\xi_i)$$

Furthermore the maps $S(\omega) \to S(\xi)$ are surjective. To see this, take a particular $f' \in S(\xi)$, then we construct a lift $f \in S(\omega)$ as follows: For each ξ_i, take the set of $\omega_j \subseteq \xi_i$. We call them $\omega_1, \ldots, \omega_k$ for simplicity. Now assign the values of $f(\omega_2), \ldots, f(\omega_k) \in M$ arbitrarily so that $\pi(f(\omega_j)) = \phi_\omega(\omega_j) \in N$ for $j = 2, \ldots, k$. Then we finally assign the value of

$$f(\omega_1) = f'(\xi_i) - f(\omega_2) - \cdots f(\omega_k)$$

And it follows that

$$\pi \circ f(\omega_1) = \pi \circ f'(\xi_i) - \sum_{j=2}^{k} \pi \circ f(\omega_j) = \varphi_\xi(\xi_i) - \sum_{j=2}^{k} \varphi_\omega(\omega_j) = \varphi_\omega(\omega_1)$$

And repeating this procedure for each ξ_i gives a lift $f \in S(\omega)$, so in fact we have $S(\omega) \twoheadrightarrow S(\xi)$

Now a lift ψ of φ corresponds to a compatible choice of elements of $S(\xi)$ for all ξ, hence an element of $\varprojlim_\xi S(\xi)$. But the limit over a direct system of non-empty sets and surjections is always non-empty, and hence there is a lift $\psi : \mathfrak{A}_R\mathcal{F} \to M$ so that $\varphi = \pi \circ \psi$. □

Corollary A.2. *The k-module $\ker \pi_\mathcal{F}$ is projective*

Proof. This follows because $k_0\mathcal{F}$ is a free k-module. Because $\mathfrak{A}_k\mathcal{F}$ is projective, it follows that $\ker \pi_\mathcal{F}$ is a direct summand of $k_0\mathcal{F}$, hence projective. □

B Proof that the Shannon entropy is nonnegative

For $n \geq 0$ and a continuous map $\eta : [0, \infty) \to \mathbb{R}$, let

$$F_n[\eta](x_1, \ldots, x_n) = \sum_{k=0}^{n} (-1)^k \sum_{i_1 < \cdots < i_k} \eta(x_{i_1} + \cdots + x_{i_k})$$

where the $k = 0$ contribution is just $\eta(0)$. It will be convenient to introduce the notation $t \cdot \eta(z) := \eta(z + t)$ for the action by translation of positive numbers on the space of functions $[0, \infty) \to \mathbb{R}$.

Observe that if η is differentiable then $F_n[\eta]$ is as well, and we can calculate

$$\frac{\partial F_n[\eta]}{\partial x_i}(x_1, \ldots, x_n) = -F_{n-1}[x_i \cdot \eta'](x_1, \ldots, \hat{x_i}, \ldots, x_n)$$

$$\frac{\partial}{\partial t} F_n[t \cdot \eta](x_1, \ldots, x_n) = F_n[t \cdot \eta'](x_1, \ldots, x_n)$$

We also have that $F_n[\eta](0, x_2, \ldots, x_n) = 0$, because each term containing x_i will cancel the corresponding term not containing x_1.

Proposition B.1. *Let $\eta : [0, \infty) \to \mathbb{R}$ be a smooth function such that*

$$\frac{d^{2k}\eta(z)}{dz^{2k}} \geq 0$$

for all $z > 0$ and $k \geq n$. Then $F_n[\eta](x_1, \ldots, x_n)$ is nonnegative for $x_1, \ldots, x_n \geq 0$.

Proof. We prove this by induction on n. For the case $n = 0$, we have $F_0[\eta]() = \eta(0) \leq 0$ by continuity and the hypothesis that $\frac{d^0}{dz^0}\eta = \eta \geq 0$ for $z > 0$.

Now fix a choice of $p_1, \ldots, p_n \geq 0$. Then we will show that

$$F_n[\eta](p_1 + t, p_2 - t, p_3, \ldots, p_n) \tag{3}$$

is concave as a function of $t \in [-p_1, p_2]$. Computing the second derivative gives

$$\begin{aligned}
&\frac{\partial^2}{\partial t^2} F_n[\eta](p_1 + t, p_2 - t, p_3, \ldots, p_n) \\
&\qquad = -\frac{\partial}{\partial t} F_{n-1}[(t + p_1) \cdot \eta'](p_2 - t, p_3, \ldots, p_n) \\
&\qquad\qquad + \frac{\partial}{\partial t} F_{n-1}[(p_2 - t) \cdot \eta'](p_1 + t, p_3, \ldots, p_n) \\
&\qquad = -F_{n-1}[(p_1 + t) \cdot \eta''](p_2 - t, p_3, \ldots, p_n) \\
&\qquad\qquad - F_{n-1}[(p_2 - t) \cdot \eta''](p_1 + t, p_3, \ldots, p_4) \\
&\qquad\qquad - 2F_{n-2}[(p_1 + p_2) \cdot \eta''](p_3, \ldots, p_n)
\end{aligned}$$

And the new functions $\tilde{\eta} = (p_1 + t) \cdot \eta'', (p_2 - t) \cdot \eta''$, and $(p_1 + p_2) \cdot \eta''$ all have the property that $\frac{d^{2k}}{dz^{2k}}\tilde{\eta} \geq 0$ for $k \geq n - 1$. The inductive hypothesis implies that each term above is non-positive and so the whole expression is ≤ 0.

Thus we have verified that the function (3) is concave for $t \in [-p_1, p_2]$. In addition, the function is 0 when $t = -p_1$ or $t = p_2$ because $F_n[\eta]$ vanishes when any one of its arguments vanishes. Concavity implies the function is nonnegative when $t = 0 \in [-p_1, p_2]$. The proposition now follows by induction on n. □

Remark B.2. Under the hypotheses of proposition B.1, the second derivative of

$$F_n[\eta](tp_1, (1 - t)p_2, p_3, \ldots, p_n)$$

is nonnegative. It follows that $F_n[\eta]$ is a concave function.

References

P. K. B́.R. Ebanks & C. Ng (1990). 'Recursive Inset Entropies of Multiplicative Type on Open Domains'. *Aequationes Math.* **3p.**(pp. 100).

B. Ebanks (1986). 'Branching Inset Entropies on Open Domains'. *Aequationes Math.* **30.**(pp. 187).

P. Kannappan & W. Sander (2004). 'Inset Information Measures on Open Domains'. *Aequationes Math.* **68.**(pp. 289).

I. M. M.F. Atiyah (1969). *Introduction to Commutative Algebra*. Westview Press, Boulder,CO.

C. Shannon (1948). 'A Mathematical Theory of Communication'. *Bell System Technical Journal* **27.**(pp. 379, 623).