Information cohomology.

Juan Pablo Vigneaux
IMJ-PRG - Université Paris 7

November 15, 2016

Outline

Information theory

Shannon (1948) [5] defined the information content of a random variable $X : \Omega \to \{x_1, ..., x_n\}$ as

$$H_1(X) = -\sum_{k=0}^{n} \mathbb{P}(X = x_i) \log_2 \mathbb{P}(X = x_i), \tag{1}$$

where $\mathbb{P}$ denotes a probability law on the space Ω. The function H is called entropy.

Shannon (1948) [5] defined the information content of a random variable $X : \Omega \to \{x_1, ..., x_n\}$ as

$$H_1(X) = -\sum_{k=0}^{n} \mathbb{P}(X = x_i) \log_2 \mathbb{P}(X = x_i), \quad (1)$$

where $\mathbb{P}$ denotes a probability law on the space Ω. The function H is called entropy.

Information is related to **uncertainty**.

1. If $P(X = x_i) = 1$ for certain i, then $H_1(X) = 0$.
2. Uniform distribution on $\{x_1, ..., x_n\}$ implies $H_1(X)$ maximal.

Information theory

Shannon (1948) [5] defined the information content of a random variable $X : \Omega \to \{x_1, ..., x_n\}$ as

$$H_1(X) = -\sum_{k=0}^{n} \mathbb{P}(X = x_i) \log_2 \mathbb{P}(X = x_i), \tag{1}$$

where $\mathbb{P}$ denotes a probability law on the space Ω. The function H is called entropy.

Information is related to **uncertainty**.

1. If $P(X = x_i) = 1$ for certain i, then $H_1(X) = 0$.
2. Uniform distribution on $\{x_1, ..., x_n\}$ implies $H_1(X)$ maximal.

From the formula above, Shannon recognized an important relation,

$$H_1(X, Y) = H_1(X) + H_1(Y|X).$$

Generalized functions: α-entropies

Other example is given by the α-entropy, introduced in 1967 by Jan Havrda and František Charvát [4]: for $\alpha > 0$, $\alpha \neq 1$ define

$$H_\alpha[X](P) = c_\alpha \left(\sum_{k=1}^{n} P(X = x_k)^\alpha - 1 \right),$$

where c_α is some constant, positive if $0 < \alpha < 1$ and negative otherwise. Typical choices satisfy

$$\lim_{\alpha \to 1} (1 - \alpha) c_\alpha = 1;$$

in this case, $H_\alpha[X] \to H_1[X]$ when $\alpha \to 1$ (cf. [2]). The use of these functions in statistical mechanics has been popularized by Constantino Tsallis.

Outline

Observables

Fix a probability space Ω. Observables are random variables $1, X_1, X_2, X_3, \ldots$ (where 1 corresponds to certitude/a constant variable). We are just interested in the algebras of events defined by each variables (we consider $X \cong Y$ if $\sigma(X) = \sigma(Y)$).

We can write an arrow $X \to Y$ if $\sigma(Y) \subset \sigma(X)$ (if "X refines Y").

$\mathcal{O} = \mathcal{O}(\Omega)$ category of “finite observables”, whose objects are all the possible finite partitions of Ω and arrows are refinements.

- $\mathcal{O}$ has a terminal element, the trivial partition $\mathbf{1} := \{\Omega\}$. (If Ω is finite, $\mathcal{O}$ has also an initial element, the partition by points, that we denote by $\mathbf{0}$.)
- $(X,Y) : \Omega \to E_X \times E_Y, \omega \mapsto (X(\omega), Y(\omega))$ is the coarser partition that refines X and Y: it is the categorical product. We write simply $XY := (X,Y)$ (this product is commutative, associative, $\mathbf{1}$ is a unit, all elements are idempotents).

Definition (Information structure)

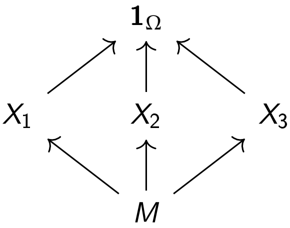
A ***concrete* classical information structure**, denoted by $\mathcal{S}$, is a *full* subcategory of $\mathcal{O}(\Omega)$ that satisfies the following properties:

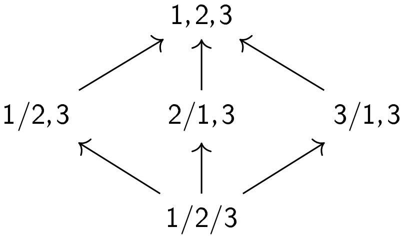
1. *Terminal object:*
   The partition $\mathbf{1}$ is in $\mathrm{Ob}(\mathcal{S})$.
2. *Conditional existence of products:*
   Given any three objects X, Y and Z in $\mathrm{Ob}(\mathcal{S})$, such that X refines Y and Z, then the partition YZ also belongs to $\mathrm{Ob}(\mathcal{S})$.

The properties above imply that, over each variable $X \in \mathrm{Ob}(\mathcal{S})$, the variables $\mathcal{S}_X := \{Y \in \mathrm{Ob}(\mathcal{S}) \mid X \to Y\}$ form a monoid (for the multiplication of partitions in $\mathcal{O}(\Omega)$).

Information structures: examples

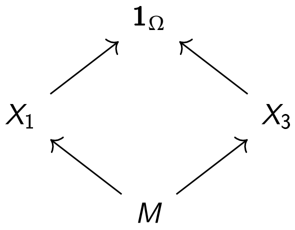
Example 1. Set $\Omega = \{1,2,3\}$ and define $X_i := \{\{i\}, \Omega \setminus \{i\}\}$. M is the atomic partition.





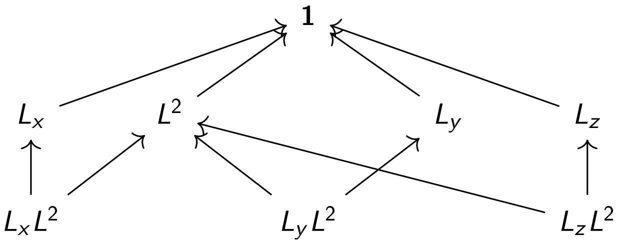
Information structures: examples

Example 2. As before, but the observable X_2 is not available.



Information structures: examples

Example 3. From quantum physics. Here, L_x, L_y, L_z are the quantum observables that correspond to angular momentum and $L^2 = L_x^2 + L_y^2 + L_z^2$.



We cannot measure simultaneously two components of the angular momentum since the operators do not commute.

Generalized structures

As Gromov in [3], we want to approach measurements from a categorical point of view and consider Ω as an emerging object.

Definition

A **classical information structure** $\mathcal{S}$ is a category whose objects are finite sets and whose morphisms are *surjections* between these sets, that satisfies:

1. $\mathcal{S}$ is a small category: both $\mathrm{Ob}(\mathcal{S})$ and $\mathrm{Hom}(S)$ are sets.
2. $\mathcal{S}$ has a terminal object, denoted $\mathbf{1}$, which is the one-point set.
3. $\mathcal{S}$ is a partially ordered set (poset): if A and B are different objects, $\mathrm{Hom}(A,B)$ has at most one element; if $A \neq B$ and $A \to B$, then $B \not\to A$.
4. Conditional existence of infima: for objects $X, Y, Z \in \mathrm{Ob}(\mathcal{S})$, if $Z \to X$ and $Z \to Y$ then $X \wedge Y$ exists ($X \wedge Y$ is the categorical product in the poset $\mathcal{S}$).

Definition (continued)

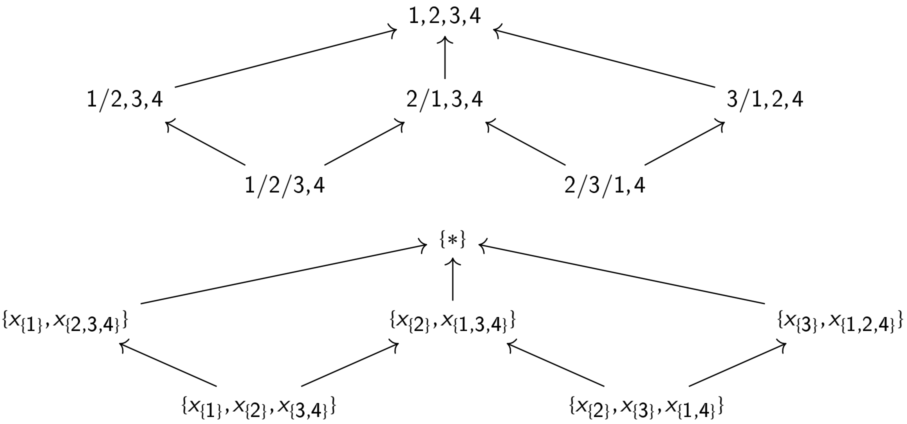
5. For every diagram

$$X \xleftarrow{\pi} X \wedge Y \xrightarrow{\sigma} Y$$

there is a bijection of sets $X \wedge Y \simeq \{(x,y) \in X \times Y \,|\, \pi^{-1}(x) \cap \sigma^{-1}(y) \neq \emptyset\}$. (Equivalently, we can impose that $|\pi^{-1}(x) \cap \sigma^{-1}(y)| \leq 1$ for every $x \in X$ and $y \in Y$.)

1. We call $X \in \mathrm{Ob}(\mathcal{S})$ a partition (sometimes: variable, observable) and its points, parts or elements.
2. The set-notation $\{X = x\}$ simply means "the part x contained in X".
3. $\{X = x, Y = y\}$ should be interpreted as *the* part z of $X \wedge Y$ mapped to x by $X \wedge Y \to X$ and to y by $X \wedge Y \to Y$ (if such z does not exist, write $\{X = x, Y = y\} = \emptyset$).
4. write $XY := X \wedge Y$ and define the monoid $\mathcal{S}_X$

Example:



where $x_A \mapsto x_B$ for $A \subset B$.

The category of information structures: $\mathscr{InfoStr}$

Given two structures $\mathscr{S}, \mathscr{S}'$, a morphism ϕ between them is a map $\phi : \mathrm{Ob}(\mathscr{S}) \to \mathrm{Ob}(\mathscr{S}')$ such that:

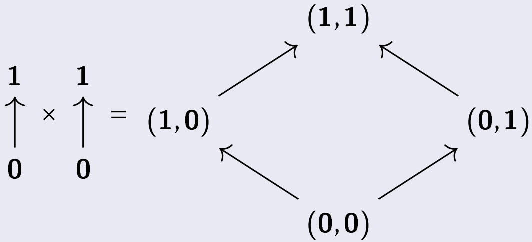
1. $\phi(\mathbf{1}) = \mathbf{1}$,
2. $X \to Y$ in $\mathscr{S}$ implies $\phi(X) \to \phi(Y)$,
3. if $X \wedge Y$ exists, then $\phi(X \wedge Y) = \phi(X) \wedge \phi(Y)$.

Note that, if $X \wedge Y$ exists, property (2) above implies $\phi(X \wedge Y) \to \phi(X)$ and $\phi(X \wedge Y) \to \phi(X)$, and thus the product $\phi(X) \wedge \phi(Y)$ exists too, in virtue of Definition 3-(4).

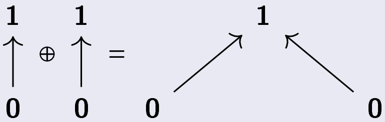
Proposition

The category $\mathscr{InfoStr}$ has finite products and coproducts.

Example



and



Probabilities

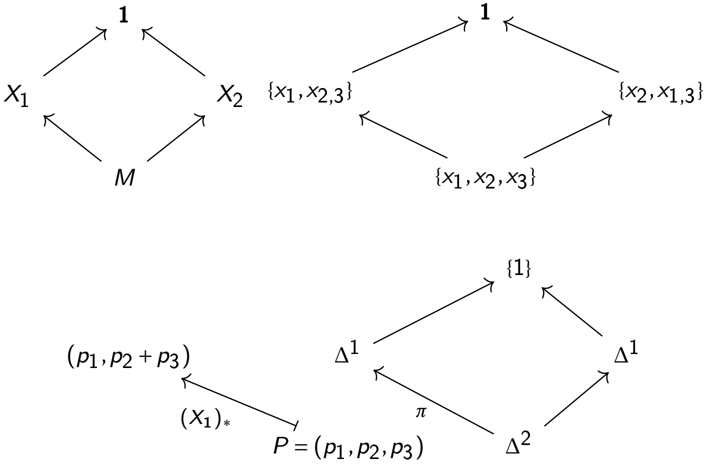
A probability functor Q on a classical information structure is a functor $Q : \mathcal{S} \to \mathcal{S}ets$ such that, for every $X \in \mathrm{Ob}(\mathcal{S})$, the set Q_X is a subcomplex of $\Delta^{|X|-1}$ (all possible laws on a set of cardinality $|X|$) and the morphisms $Q(\pi) : Q_X \to Q_Y$ are surjective simplicial mappings induced by the natural "collapsing map" from $\Delta^{|X|-1}$ onto $\Delta^{|Y|-1}$; this means that they satisfy the equations:

$$\forall f : X \to Y, \forall y \in Y, \forall P \in Q_X : \quad Q(\pi)(P)(y) = \sum_{x \in \pi^{-1}(y)} P(x).$$

The operation implemented by $Q(\pi)$ is called **marginalization** and denoted Y_* if π is clear from the context.

Example.

$\Delta^k := \{(x_0,\ldots,x_k) \in \mathbb{R}^2_{\geq 0} : x_0 + \ldots + x_k = 1\}$, the k-simplex.



Conditional probabilities

For $\pi : X \to Y$, $Y = \{y_1, ..., y_n\}$ variable/partition/observable, P a law of Q_X and $P(Y = y_i) \neq 0$ for certain $i \in [n]$, define a new law in Q_X:

$$P|_{Y=y_i}(A) = \frac{P(A \cap \{Y = y_i\})}{P(Y = y_i)} = \frac{P(A \cap \pi^{-1}(y_i))}{P(\pi^{-1}(y_i))}.$$

The functor is called adapted if it is stable by conditioning.

Functional module

For each observable X, consider the real vector space

$$F_X = \{f : Q_X \to \mathbb{R} \mid f \text{ measurable}\}.$$

The entropy $H[X]$ lives here.

If $\pi : X \to Y$, a function $f \in F_Y$ can be mapped naturally to F_X: just set $(F(\pi)f)(P) = f(Y_* P)$. With this conventions, F is a contravariant functor from $\mathcal{S}$ to $\mathcal{V}ect_{\mathbb{R}}$.

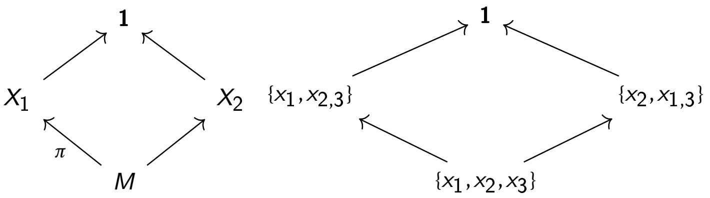
For each $\alpha > 0$, we define an action of $\mathcal{S}_X$ on F_X by the rule:

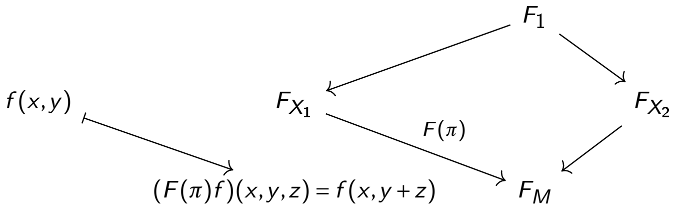
$$(Y.f)(P) = \sum_{i=1}^{k} P(Y = y_i)^{\alpha} f(P|_{Y=y_i}).$$

This gives a morphism of monoids $\Lambda_\alpha(X) : \mathcal{S}_X \to \mathrm{End}(F_X)$, natural in X.

Example.

Δ^k the k-simplex. $F_M = \{f : \Delta^2 \to \mathbb{R}\}$, etc.





Outline

(Some) General definitions

Let $\mathcal{C}$ be a category.

1. A presheaf of sets is any (contra-variant) functor $F : \mathcal{C}^{op} \to \mathcal{S}ets$ (here $\mathcal{S}ets =$ the category of sets). A morphism of presheaves $\phi : F \to G$ is a natural transformation of functors.
2. (We put on $\mathcal{C}$ the trivial Grothendieck topology, such that every presheaf is a sheaf.)
3. If $\mathcal{C}$ is a site and O is a sheaf of rings on $\mathcal{C}$, the couple $(\mathcal{C}, O)$ is called a **ringed site** and O, the structure ring. The couple $(\mathcal{S}h(\mathcal{C}), O)$ is called a **ringed topos**.

(Some) General definitions

Given a ringed site $(\mathcal{C}, O)$.

1. A **sheaf of O-modules** is given by an abelian sheaf F together with a map of presheaves of sets $O \times F \to F$, such that for every $X \in \mathrm{Ob}(\mathcal{C})$, the map $O(X) \times F(X) \to F(X)$ defines a structure of $O(X)$-module on the abelian group $F(X)$.
2. A morphism $\phi : F \to G$ between sheaves of O-modules is a morphism of abelian presheaves $\phi : F \to G$ such that

$$\begin{array}{ccc} O \times F & \longrightarrow & F \\ \downarrow{\scriptstyle 1\times\phi} & & \downarrow{\scriptstyle \phi} \\ O \times G & \longrightarrow & G \end{array}$$

   The set of O-module morphisms from F to G is denoted by $\mathrm{Hom}_O(F, G)$.

Sheaves of O-modules and its morphisms form the category $\mathcal{M}od(O)$.

Information topos

Proposition

Let $(\mathcal{Sh}(\mathcal{C}), O)$ be a ringed topos. The category $\mathcal{Mod}(O)$ is abelian. Moreover, it has enough injective objects.

- Over each $X \in \mathcal{S}$ there is monoid $\mathcal{S}_X$ of variables coarser than X. Denote by $\mathcal{A}_X$ the monoid ring $\mathbb{R}[\mathcal{S}_X]$.
- Put the trivial Grothendieck topology on $\mathcal{S}$. The couple $(\mathcal{S}, \mathcal{A})$ is a ringed site. We work in the category $\mathcal{Mod}(\mathcal{A})$: sheaves of abelian groups with an action of $\mathcal{A}$.
  $\mathcal{Mod}(\mathcal{A})$ **is the topos of information.**
- Define the information cohomology as (cf. Bennequin-Baudot, 2015 [1]):

$$H^n(\mathcal{S}, F_\alpha) := \operatorname{Ext}^n(\mathbb{R}_{\mathcal{S}}, F_\alpha).$$

An alternative definition

Objective: build a projective resolution of the trivial sheaf of $\mathcal{A}$-modules $\mathbb{R}_{\mathcal{S}}$.

$$0 \leftarrow \mathbb{R}_{\mathcal{S}} \leftarrow B_0 \leftarrow B_1 \leftarrow B_2 \leftarrow \dots$$

- Let $B_n(X)$ de be the free $\mathcal{A}_X$ module generated by the symbols $[X_1|\dots|X_n]$, where $\{X_1,\dots,X_n\} \subset \mathcal{S}_X$. Remark that $B_0(X)$ is the free module on one generator $[\,]$, in consequence isomorphic to $\mathcal{A}_X$. The B_i are sheaves in $\mathcal{Mod}(\mathcal{A})$.
- We introduce now $\mathcal{A}_X$-module morphisms $\varepsilon_X : B_0(X) \to \mathbb{R}_{S_X}$ given by the equation $\varepsilon_X([\,]) = 1$, and boundary morphisms $\partial : B_n(X) \to B_{n-1}(X)$, given by

$$\begin{aligned} \partial[X_1|\dots|X_n] =& X_1[X_2|\dots|X_n] \\ &+ \sum_{k=1}^{n-1} (-1)^k [X_1|\dots|X_k X_{k+1}|\dots|X_n] + (-1)^n [X_1|\dots|X_{n-1}]. \end{aligned}$$

These morphisms are natural in X.

Resolution

Proposition

The complex

$$0 \longleftarrow \mathbb{R}_{\mathcal{S}} \xleftarrow[\varepsilon]{} B_0 \xleftarrow[\partial_1]{} B_1 \xleftarrow[\partial_2]{} B_2 \xleftarrow[\partial_3]{} \dots \tag{2}$$

is a resolution of the sheaf $\mathbb{R}_{\mathcal{S}}$.

In fact, the construction correspond to a *relatively free resolution*, cf. MacLane, *Homology*.

Proposition

For each $n \geq 0$, *the sheaf* B_n *is a projective object in* $\mathcal{Mod}(\mathcal{A})$.

We have shown the existence of a projective resolution of $\mathbb{R}_{\mathcal{S}}$ in $\mathcal{Mod}(\mathcal{A})$,

$$0 \leftarrow \mathbb{R}_{\mathcal{S}} \leftarrow B_0 \leftarrow B_1 \leftarrow B_2 \leftarrow \ldots$$

Define:

$$H^0(\mathbb{R}_{\mathcal{S}}, F_\alpha) := \ker\{\delta : \mathrm{Hom}_{\mathcal{A}}(B_0, F_\alpha) \to \mathrm{Hom}_{\mathcal{A}}(B_1, F_\alpha)\}$$

and

$$H^i(\mathbb{R}_{\mathcal{S}}, F_\alpha) := \frac{\ker\{\delta : \mathrm{Hom}_{\mathcal{A}}(B_i, F_\alpha) \to \mathrm{Hom}_{\mathcal{A}}(B_{i+1}, F_\alpha)\}}{\mathrm{im}\{\delta : \mathrm{Hom}_{\mathcal{A}}(B_{i-1}, F_\alpha) \to \mathrm{Hom}_{\mathcal{A}}(B_i, F_\alpha)\}},$$

for $i \geq 1$.

Note: these correspond to the right derived functors $\underline{\mathrm{Ext}}^n(\mathbb{R}_{\mathcal{S}}, -)$ of $\mathrm{Hom}(\mathbb{R}_{\mathcal{S}}, -)$. By the **balancing theorem**:

$$\underline{\mathrm{Ext}}^n(\mathbb{R}_{\mathcal{S}}, F) \cong \mathrm{Ext}^n(\mathbb{R}_{\mathcal{S}}, F).$$

Computing cohomology

We study the cohomology of the complex
$\{C^n := \mathrm{Hom}_{\mathcal{A}}(B_n(\mathbb{R}_{\mathcal{S}}), F(\mathcal{Q}))\}_{n\geq 0}$.

1. $f \in C^n$ is called n-cochain.
2. The coboundary of $f \in C^n$ is the $(n+1)$-cochain $\delta f = f\partial : B^{n+1} \to F(\mathcal{Q})$. More explicitly,

$$\delta f[X_1|...|X_{n+1}] = X_1.f[X_2|...|X_{n+1}] + \sum_{k=1}^{n}(-1)^k f[X_1|...|X_kX_{k+1}|...|X_n] + (-1)^{n+1} f[X_1|...|X_n]$$

3. $Z_n = \{f \in C^n \mid \delta f = 0\}$: n-cocycles.
4. δC^{n-1}: n-coboundaries. $(\delta C^{-1} := 0)$
5. $H^n(\mathbb{R}_{\mathcal{S}}, F(\mathcal{Q})) = Z^n/\delta C^{n-1}$, for every $n \geq 0$.

A n-cochain f consist of a collection of morphism $f_X \in \mathrm{Hom}_{\mathcal{A}_X}(B_n(X), F(\mathcal{Q}_X))$ that satisfy the following conditions:

1. f is a morphism of (pre)sheaves (a natural transformation): given $X \to Y$,

$$\begin{array}{ccc} B_n(Y) & \xrightarrow{f_Y} & F(\mathcal{Q}_Y) \\ \downarrow & \ulcorner & \downarrow \\ B_n(X) & \xrightarrow{f_X} & F(\mathcal{Q}_X) \end{array}$$

2. f is compatible with the action of $\mathcal{A}$: given $X \to Y$, the diagram

$$\begin{array}{ccc} O \times B_n(X) & \longrightarrow & B_n(X) \\ \downarrow{\scriptstyle 1 \times f_X} & & \downarrow{\scriptstyle f_X} \\ O \times F(\mathcal{Q}_X) & \longrightarrow & F(\mathcal{Q}_X) \end{array}$$

commutes, i.e. f_X is equivariant, $f_X(Y[Z]) = Y.f_X([Z])$.

☞Notate bene

Remark

Given E in $\mathcal{PSh}(\mathcal{S})$, it is possible to define a presheaf of $\mathcal{A}$-modules (denoted by $\mathcal{A}[E]$) that associates to each $X \in \mathrm{Ob}(\mathcal{S})$ the free $\mathcal{A}(X)$-module on generators $E(X)$. There is an adjunction

$$\mathrm{Hom}_{\mathcal{Mod}(\mathcal{A})}(\mathcal{A}[E], F) \cong \mathrm{Hom}_{\mathcal{PSh}(\mathcal{S})}(E, \square F). \tag{3}$$

Here, $\square$ denotes the forgetful functor, and will be omitted in the sequel.

Since $B_n(X)$ is a free module, f_X is completely determined by the values on the generators $[X_1|...|X_n]$. Just to simplify notation, we shall write $f_X([X_1|...|X_n])$ as $f_X[X_1|...|X_n]$. Note that $f_X[X_1|...|X_n]$ is itself a function (on $\mathcal{Q}_X$).

Locality

Requirement (1) is equivalent to the following condition, that we call **joint locality**: given observables X and Y such that $X \to Y$, and a set of observables $\{X_1, \ldots, X_n\}$ in $\mathcal{S}_Y$, a morphism $f \in \mathrm{Hom}_{\mathcal{A}}(B_n(\mathbb{R}_{\mathcal{S}}), F(\mathcal{Q}))$ must satisfy

$$f_X[X_1|\ldots|X_n](P_X) = f_Y[X_1|\ldots|X_n](Y_* P_X). \tag{4}$$

Some consequences:

1. 0-cochains satisfy $f_Y(Y_* P_X) = f_X(P_X)$ for each arrow $X \to Y$. As we assume that $\mathbf{1} \in \mathrm{Ob}(\mathcal{S})$, this implies that $f_\cdot$ is constant $(= f_{\mathbf{1}}(1))$. Every constant function is closed and $H^0(\mathcal{S}, \mathcal{Q}) \cong \mathbb{R}$.
2. 1-cochains: it is sufficient to know $f_Y[Y](Y_* P)$ to recover $f_X[Y](P)$ (for $X \to Y$); in this sense, we usually omit the subindex and just write $f[Y]$.

Proposition

Let f be a 1-cocycle. Then

1. $f[\mathbf{1}] \equiv 0$
2. *For any $X \in \mathrm{Ob}(\mathcal{S})$, any value x_i of X, and $P \in \mathcal{Q}_X$ with $P(X = x_i) > 0$, one has $f[X](P|_{X=x_i}) = 0$.*

Proof.

1. Note that $f[\mathbf{1}] = f[\mathbf{11}] = f[\mathbf{1}] + \mathbf{1}.f[\mathbf{1}]$. Then, $\mathbf{1}.f[\mathbf{1}] = 0$. But $\mathbf{1}$ takes only one value, say c, and $\mathbf{1}.f[\mathbf{1}](P) = f[\mathbf{1}](P|_{\mathbf{1}=c}) = f[1](P)$.
2. From $f[XX] = f[X] + X.f[X]$, we conclude

$$X.f[X] = \sum_{k=0}^{n} P(X = x_k) f[X](P|_{X=x_k}) = 0.$$

In each term, $P(X = x_k)^{\alpha} \geq 0$. Given an index i with $P(X = x_i) > 0$, one obtains $f[X](P|_{X=x_i}) = 0$.

□

À la poursuite de H^1 (I)

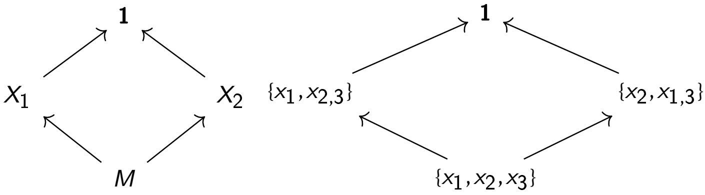
The bad guy: Take $\mathcal{S}$ as $\mathbf{0} = \{a, b\} \to \mathbf{1}$. If f is a 1-cocyle, the only effective condition on it is $f[\mathbf{0}] = f[\mathbf{0}] + \mathbf{0}.f[\mathbf{0}]$. In consequence,

$$P(\{a\})^{\alpha} f[\mathbf{0}](1,0) + P(\{b\})^{\alpha} f[\mathbf{0}](0,1) = 0. \tag{5}$$

The presence in $\mathcal{Q}_{\mathbf{0}}$ of any laws P, P' with $P(\{a\}) > 0$ and $P'(\{b\}) > 0$ implies $f[\mathbf{0}](1,0) = f[\mathbf{0}](0,1) = 0$. Then, 1-cocycles are in correspondence with measurable functions f on arguments (p_a, p_b) such that $f(1,0) = f(0,1) = 0$. (All the others cocycle equations become tautological.) We conclude that $H^1(\mathcal{S}, F_{\alpha})$ has infinite dimension.

À la poursuite de H^1 (II)

The good guy:



$\mathcal{Q}_\bullet$ full set of probabilities. The general construction says that a 1-cocycle is defined by 3 functions $f[X_1] : \mathcal{Q}_{X_1} \to \mathbb{R}$, $f[X_2] : \mathcal{Q}_{X_2} \to \mathbb{R}$, $f[M] : \mathcal{Q}_M \to \mathbb{R}$ subject to the cocycle equations.

Cocycle equations:

$$
\begin{aligned}
0 &= X_1.f[X_2] - f[M] + f[X_1] \\
0 &= X_2.f[X_1] - f[M] + f[X_2] \\
&\dots
\end{aligned}
$$

These are functional equations. They imply $X_2.f[X_1] + f[X_2] = X_1.f[X_2] + f[X_1]$, i.e., for every $(p_0, p_1, p_2) \in \Delta^2$:

$$
\begin{aligned}
&(1-p_2)^\alpha f[X_1]\left(\frac{p_0}{1-p_2}, \frac{p_1}{1-p_2}\right) - f[X_1](1-p_1, p_1) \\
&\qquad = (1-p_1)^\alpha f[X_2]\left(\frac{p_0}{1-p_1}, \frac{p_2}{1-p_1}\right) - f[X_2](1-p_2, p_2).
\end{aligned}
$$

Fundamental technology: functional equations

We introduce the functions:

$$S_1(q) = -q\log(q) - (1-q)\log(1-q) \quad \text{and} \quad S_\alpha(q) = \frac{1}{2^{\alpha-1}-1}(q^\alpha + (1-q)^\alpha - 1).$$

Proposition (Tverberg, Lee, Ng...)

Let $f_1, f_2 : \Delta^2 \to \mathbb{R}$ be two unknown functions satisfying

1. $f_i(0,1) = f_i(1,0) = 0$ *for* $i = 1,2$.
2. *for all* $(p_0, p_1, p_2) \in \Delta^2$,

$$(1-p_2)^\alpha f_1\left(\frac{p_0}{1-p_2}, \frac{p_1}{1-p_2}\right) - f_1(1-p_1, p_1) = (1-p_1)^\alpha f_2\left(\frac{p_0}{1-p_1}, \frac{p_2}{1-p_1}\right) - f_2(1-p_2, p_2). \qquad (6)$$

Then, $f_1 = f_2$ *and there exists* $\lambda \in \mathbb{R}$ *such that* $f_1(x, 1-x) = \lambda S_\alpha(x)$.

Two sources of problems

1. If an element is irreducible (cannot be factored), like in the example $\mathbf{0} \to \mathbf{1}$, then $\dim H^1(\mathcal{S}, F_\alpha) = \infty$.
2. If an element is reducible but it does not have enough probabilities, then $\dim H^1(\mathcal{S}, F_\alpha) = \infty$ (for example, if $Q_M \neq \Delta^2$ in the example above). We call such elements/products *degenerate*.

Universality of the entropy

Proposition

Let $\mathcal{S}$ be a bounded information structure and $\mathcal{Q}$ an adapted probability family. Suppose that every minimal object can be factored as a non-degenerate product. Then,

$$H^1(\mathcal{S}, F_\alpha) = \prod_{C \in H_0^{CW}(\mathcal{S}_0)} \mathbb{R} H_\alpha^C$$

where C represents a connected component of $\mathcal{S}_0 = \mathcal{S} \setminus \mathbf{1}$ and

$$H_\alpha^C[X] = \begin{cases} H_\alpha[X] & \text{if } X \in C \\ 0 & \text{if } X \notin C \end{cases}$$

Universality of the entropy

Proposition (continued)

If $(\mathcal{S}_1,\mathcal{Q}_1),\dots,(\mathcal{S}_n,\mathcal{Q}_n)$ *satisfy the hypothesis above, then*

$$H^1\left(\bigoplus_{i=1}^n \mathcal{S}, F_\alpha\right) = \bigoplus_{i=1}^n \prod_{C\in H_0^{CW}(\mathcal{S}_{i0})} \mathbb{R}H_\alpha^C.$$

Proof.

If every minimal element $M\in\mathbf{M}$ can be factored as a non-degenerate product, then $f[M]=\lambda_M H[M]$, where λ_M is a constant that depends *a priori* on M. Given $M,N\in\mathbf{M}$ such that $M\vee N = Z\neq\mathbf{1}$, every cocycle f satisfies

$$f[Z]=f[M]-Z.f[M]=\lambda_M H[M]-Z.(\lambda_M H[M])=\lambda_M H[Z]$$

and

$$f[Z]=f[N]-Z.f[N]=\lambda_N H[N]-Z.(\lambda_N H[N])=\lambda_N H[Z].$$

Therefore, $\lambda_M=\lambda_N$. Conversely, if $M\vee N=\mathbf{1}$, no cocycle equation relates λ_M and λ_N. □

Outline

Suppose we have a set of possible events whose probabilities of occurrence are $p_1, p_2, \cdots, p_n$. These probabilities are known but that is all we know concerning which event will occur. Can we find a measure of how much "choice" is involved in the selection of the event or of how uncertain we are of the outcome?

If there is such a measure, say $H(p_1, p_2, \cdots, p_n)$, it is reasonable to require of it the following properties:

1. H should be continuous in the p_i.
2. If all the p_i are equal, $p_i = \frac{1}{n}$, then H should be a monotonic increasing

3. If a choice be broken down into two successive choices, the original H should be the weighted sum of the individual values of H. The meaning of this is illustrated in Fig. 6. At the left we have three possibilities $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{6}$. On the right we first choose between two possibilities each with probability $\frac{1}{2}$, and if the second occurs make another choice with probabilities $\frac{2}{3}$, $\frac{1}{3}$. The final results have the same probabilities as before. We require, in this special case, that

$$H(\tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{6}) = H(\tfrac{1}{2}, \tfrac{1}{2}) + \tfrac{1}{2}H(\tfrac{2}{3}, \tfrac{1}{3})$$

The coefficient $\frac{1}{2}$ is because this second choice only occurs half the time.

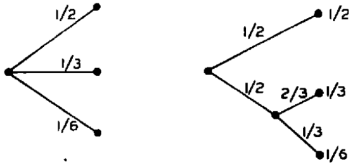


Fig. 6—Decomposition of a choice from three possibilities.

Reinterpretation

Inspired by how Shannon *uses* H, I propose to read the left tree as the three possible results of a variable XY, and the right tree as the iterated choices given by Y and X.

The *weighted sum* should be read

$$H[XY](P) = H[Y](Y_* P) + \sum_{y_i} P(Y = y_i) H[X](X_*(P|_{Y=y_i})).$$

Example

If $\Omega = [0,1]$, $Y = \chi_{[0,1/2]}$, $X = \chi_{[0,1/6]}$, you obtain the example of Shannon for $P =$ the Lebesgue measure.

Cohomology of algebras

Let Λ be a presheaf of Rr-algebras. An extension of Λ is an epimorphism $\sigma : \Gamma \to \Lambda$.

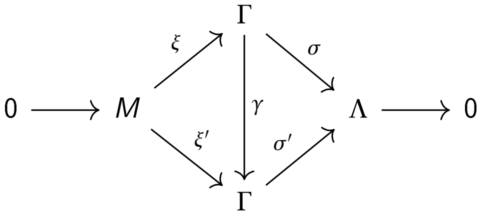
1. The extension is called **singular** (or square zero) if $\ker(\sigma)^2 = 0$ (in this case, $\ker(\sigma)$ can be regarded as a Λ-bimodule).
2. An extension is called **cleft** if there exists a morphism $\phi : \Lambda \to \Gamma$, ϕ_X morphism of algebras, such that $\sigma \circ \phi = 1_\Lambda$.
3. Given a bimodule M, a singular extension of Λ by M is a short exact sequence

$$0 \longrightarrow M \xrightarrow{\xi} \Gamma \xrightarrow{\sigma} \Lambda \longrightarrow 0$$

where ξ is a morphism of Γ-bimodules (M is Γ-bimodule by $\gamma.m = \sigma(\gamma).m$, etc.). commute.

Cohomology of algebras

4. Two extensions are called **congruent** if there is an algebra morphism $\gamma : \Gamma \to \Gamma'$ making



commute.

A particular singular cleft extension of Λ by M is given by the **semidirect sum**, defined to be the (pre)sheaf of vector spaces $M \oplus \Lambda$ with product defined by $(m_1, \lambda_1) \bullet (m_2, \lambda_2) = (m_1\lambda_2 + \lambda_1 m_2, \lambda_1\lambda_2)$; with $\xi(m) = (m, 0)$ and $\sigma(m, \lambda) = \lambda$.

Proposition

Any singular cleft extension is congruent to $A \rtimes \Lambda$.

In our case, $\Lambda = \mathcal{A}$ and $M = F_\alpha$ (turned into a (pre)sheaf of $\mathcal{A} - \mathcal{A}$-bimodules with trivial right action). If Γ is a singular cleft extension of $\mathcal{A}$ by F_α, it is isomorphic to $F_\alpha \rtimes \mathcal{A}$.

What are the possible morphisms $\phi : \mathcal{A} \to F_\alpha \rtimes \mathcal{A}$ that implement this splitting? Set $\phi(X) = (X, d(X))$; since ϕ_X morphism of algebras,

$$(Y, d(Y)) \bullet (X, d(X)) = (YX, d(YX)) \Leftrightarrow (YX, d(Y) + Y.d(X)) = (YX, d(YX)).$$

Thus d must be a *derivation*: a 1-cocycle! There is no choice, we must take the entropy.

Final remarks

The extensions that are singular and $\mathbb{R}$-split (instead of cleft) are classified by $H^2(\mathcal{S}, F)$:
The morphism $\phi_X : \mathbb{A}_{\mathbb{X}} \to \Gamma_X$ gives a natural vector space decomposition $\Gamma_X \simeq \mathbb{A}_{\mathbb{X}} \oplus F_X$, whith product given by $(X, f) \bullet (Y, g) = (X, f.Y + X.g + f(X, Y))$. The function $f : \mathcal{A} \otimes_{\mathbb{R}} \mathcal{A} \to \mathbb{R}$ is called the factor set of the extension and the associativity of the product in Γ entails that f is a 2-cocycle.

Higher cohomology groups correspond to non-singular extensions?

P. Baudot and D. Bennequin, *The homological nature of entropy*, Entropy, 17 (2015), pp. 3253–3318.

I. Csiszár, *Axiomatic characterizations of information measures*, Entropy, 10 (2008), pp. 261–273.

M. Gromov, *In a search for a structure, part 1: On entropy*, Preprint available at http://www.ihes.fr/gromov, (2012).

J. Havrda and F. Charvát, *Quantification method of classification processes. concept of structural a-entropy*, Kybernetika, 3 (1967), pp. 30–35.

C. Shannon, *A mathematical theory of communication*, Bell System Technical Journal, 27 (1948), pp. 379–423, 623–656.