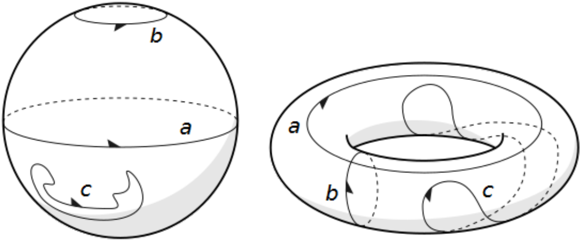
# Variations on information theory: categories, cohomology, entropy.

Juan Pablo Vigneaux
*IMJ-PRG - Université Paris 7*

May 17, 2016

# (CO)HOMOLOGY

In geometry, homology and cohomology are related to the notion of “shape”.




Define
$H^1 = \{$1-dimensional cycles$\}/\{$1-dimensional boundaries$\}$.
The fact that $\dim H^1(\text{sphere}) = 0$ and $\dim H^1(\text{torus}) = 2$ is stable under continuous deformations.

# INFORMATION THEORY

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

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

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

# INFORMATION THEORY

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

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

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

Information is related to **uncertainty**.

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

# Information theory

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

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

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

Information is related to **uncertainty**.

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

Shannon recognized an important relation,

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

# OBSERVABLES

Consider a set of observables $1, X_1, X_2, X_3, \ldots$ (where 1 corresponds to a 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 $Y$").

# Information structures: examples

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

$$
\begin{tikzcd}
 & \mathbf{1}_\Omega & \\
X_1 \arrow[ur] & X_2 \arrow[u] & X_3 \arrow[ul] \\
 & M \arrow[ul] \arrow[u] \arrow[ur] &
\end{tikzcd}
$$

# INFORMATION STRUCTURES: EXAMPLES

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

$$\begin{tikzcd}
 & \mathbf{1}_\Omega & \\
X_1 \arrow[ur] & & X_3 \arrow[ul] \\
 & M \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

# 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 $L^2 = L_x^2 + L_y^2 + L_z^2$.

$$\begin{tikzcd}
 & & \mathbf{1} & & \\
L_x \arrow[urr] & L^2 \arrow[ur] & & L_y \arrow[ul] & L_z \arrow[ull] \\
L_xL^2 \arrow[u] \arrow[ur] & & L_yL^2 \arrow[ul] \arrow[ur] & & L_zL^2 \arrow[u] \arrow[ulll]
\end{tikzcd}$$

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

# INFORMATION STRUCTURE: GENERAL DEFINITION

An information structure is a category, whose objects are observables (seen as partitions/$\sigma$-algebras) and whose arrows are refinements (they form a poset for this relation).

We suppose that:

- given any three observables $X$, $Y$ and $Z$ in $\mathcal{S}$, such that $X$ refines $Y$ and $Z$, then the joint observable

$$YZ := (Y, Z), \omega \mapsto (Y(\omega), Z(\omega))$$

  also belongs to $\mathcal{S}$.
- $\mathcal{S}$ has a final object (a constant variable/ a certitude).

# PROBABILITIES

Each observable $X$ defines an algebra of sets $\sigma(X)$. Fix a set $\mathcal{Q}_X$ of allowed laws on $(\Omega, \sigma(X))$, parametrized in some way.

To each arrow of refinement $X \to Y$, we want a surjective application $\mathcal{Q}_X \xrightarrow{Y_*} \mathcal{Q}_Y$, called marginalization.

# Example.

Set $\Omega = \{1,2,3\}$, $X_i := \{\{i\}, \Omega \setminus \{i\}\}$, $M$ atomic.
$\Delta^k := \{(x_0, \ldots, x_k) \in \mathbb{R}^2_{\geq 0} : x_0 + \ldots + x_k = 1\}$, the $k$-simplex.

$$\begin{tikzcd}
 & \mathbf{1}_\Omega & \\
X_1 \arrow[ur] & & X_3 \arrow[ul] \\
 & M \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

$$\begin{tikzcd}
 & & & \{1\} & \\
(p_1, p_2 + p_3) & & \Delta^1 \arrow[ur] & & \Delta^1 \arrow[ul] \\
 & (p_1, p_2, p_3) \arrow[ul, mapsto, "(X_1)_*"] & & \Delta^2 \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

# Functional module

Similarly, for each observable $X$, consider the real vector space $F_X$ of measurable functions on $\mathcal{Q}_X$ (the entropy $H[X]$ lives here!).

If $X \to Y$, a function $f \in \mathcal{Q}_X$ can be mapped naturally to $F_X$: just set $f^{X|Y}(P) = f(Y_* P)$.

The set $F_X$ accepts a natural action of $\mathcal{S}_X$ (these are the variables refined by $X$): for an observable $Y$ (call the possible values $\{y_1, ..., y_k\}$) in $\mathcal{S}_X$ and $f \in F(\mathcal{Q}_X)$, the new function $Y.f \in F_X$ is given by

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

# Example.

Set $\Omega = \{1,2,3\}$, $X_i := \{\{i\}, \Omega \setminus \{i\}\}$, $M$ atomic, $\Delta^k$ the $k$-simplex.
$F_M = \{f : \Delta^2 \to \mathbb{R}\}$, etc.

$$
\begin{tikzcd}
 & \mathbf{1}_\Omega & \\
X_1 \arrow[ur] & & X_3 \arrow[ul] \\
 & M \arrow[ul] \arrow[ur] &
\end{tikzcd}
$$

$$
\begin{tikzcd}
 & & & F_1 \arrow[dl] \arrow[dr] & \\
f(x,y) \arrow[dr, mapsto] & & F_{X_1} \arrow[dr] & & F_{X_2} \arrow[dl] \\
 & f^{X|Y}(x,y,z) = f(x,y+z) & & F_M &
\end{tikzcd}
$$

# FINITE QUANTUM CASE

- The role of $\Omega$ is played by a fixed finite dimensional, complex vector space $E$ with a distinguished basis (or a non-degenerate hermitian form).
- Observables are self-adjoint operators, they induce decompositions of $E$ as direct-sum of subspaces (Spectral theorem).
- We can measure simultaneously two quantities only if the corresponding observables commute as operators. In this case the joint $(X, Y)$ determines a refined decomposition.
- We obtain a category $\mathcal{S}$ of observables.
- Quantum laws are positive hermitian forms.
- Etc.

# DE RHAM COHOMOLOGY

**Question:** $U \subset \mathbb{R}^2$, functions $f_1, f_2 : U \to \mathbb{R}$. Is $\frac{\partial f_1}{\partial y} - \frac{\partial f_2}{\partial x} = 0$ a sufficient condition for the existence of $F$ such that $\nabla F = (f_1, f_2)$?

1. If $U$ is star-shaped (radially convex): yes!
2. if $U = \mathbb{R}^2 \setminus \{0\}$: no.

# DE RHAM COHOMOLOGY

**Question:** $U \subset \mathbb{R}^2$, functions $f_1, f_2 : U \to \mathbb{R}$. Is $\frac{\partial f_1}{\partial y} - \frac{\partial f_2}{\partial x} = 0$ a sufficient condition for the existence of $F$ such that $\nabla F = (f_1, f_2)$?

1. If $U$ is star-shaped (radially convex): yes!
2. if $U = \mathbb{R}^2 \setminus \{0\}$: no.

For example, for $(f_1, f_2) = \left(\frac{-x_2}{x_1^2+x_2^2}, \frac{x_1}{x_1^2+x_2^2}\right)$ such $F$ does not exist,

# DE RHAM COHOMOLOGY

**Question:** $U \subset \mathbb{R}^2$, functions $f_1, f_2 : U \to \mathbb{R}$. Is $\frac{\partial f_1}{\partial y} - \frac{\partial f_2}{\partial x} = 0$ a sufficient condition for the existence of $F$ such that $\nabla F = (f_1, f_2)$?

1. If $U$ is star-shaped (radially convex): yes!
2. if $U = \mathbb{R}^2 \setminus \{0\}$: no.

For example, for $(f_1, f_2) = \left(\frac{-x_2}{x_1^2+x_2^2}, \frac{x_1}{x_1^2+x_2^2}\right)$ such $F$ does not exist, since $\int_0^{2\pi} \frac{\mathrm{d}}{\mathrm{d}\theta} F(\cos\theta, \sin\theta) d\theta = F(1,0) - F(1,0) = 0$ but $\frac{\mathrm{d}}{\mathrm{d}\theta} F(\cos\theta, \sin\theta) = 1$ by the chain rule.

The answer depends on the "shape" (the topology) of $U$.

# Some algebra...

$$\begin{tikzcd}
C^\infty(U,\mathbb{R}) \arrow[d, equal] & \{1-\text{forms}\} \arrow[d, equal] & \{2-\text{forms}\} \arrow[d, equal] \\
\Omega^0(U) \arrow[r, "\delta_0=\nabla"] & \Omega^1(U) \arrow[r, "\delta_1=\text{curl}"] & \Omega^2(U)
\end{tikzcd}$$

$$f \longmapsto \frac{\partial f}{\partial x}\, \mathrm{d}\, x + \frac{\partial f}{\partial y}\, \mathrm{d}\, y$$

$$g(x,y)\, \mathrm{d}\, x + h(x,y)\, \mathrm{d}\, y \longmapsto \left(\frac{\partial g}{\partial y} - \frac{\partial h}{\partial x}\right) \mathrm{d}\, x \wedge \mathrm{d}\, y.$$

Remark that $\text{curl}(\nabla f) = 0$... this means that

$$\operatorname{im} \nabla \subset \ker(\text{curl}).$$

$$\Omega^0(U) \xrightarrow{\delta_0=\nabla} \Omega^1(U) \xrightarrow{\delta_1=\mathrm{curl}} \Omega^2(U)$$

Define,

$$H^1(U) = \ker \delta_1 / \operatorname{im} \delta_0 = \ker(\mathrm{curl}) / \operatorname{im} \nabla.$$

Then,

1. $H^1(U) \cong \{0\}$ if $U$ is star-shaped.

$$\Omega^0(U) \xrightarrow{\delta_0=\nabla} \Omega^1(U) \xrightarrow{\delta_1=\mathrm{curl}} \Omega^2(U)$$

Define,

$$H^1(U) = \ker \delta_1 / \operatorname{im} \delta_0 = \ker(\mathrm{curl}) / \operatorname{im} \nabla.$$

Then,

1. $H^1(U) \cong \{0\}$ if $U$ is star-shaped.
2. $H^1(\mathbb{R}^2 \setminus \{0\}) \neq \{0\}$.

$$\Omega^0(U) \xrightarrow{\delta_0=\nabla} \Omega^1(U) \xrightarrow{\delta_1=\mathrm{curl}} \Omega^2(U)$$

Define,

$$H^1(U) = \ker \delta_1 / \operatorname{im} \delta_0 = \ker(\mathrm{curl}) / \operatorname{im} \nabla.$$

Then,

1. $H^1(U) \cong \{0\}$ if $U$ is star-shaped.
2. $H^1(\mathbb{R}^2 \setminus \{0\}) \neq \{0\}$.
3. In general, $H^1(U) \cong \mathbb{R}^n$ if $U$ has $n$ holes.

# THE TRICKY TECHNICAL POINTS...

1. Consider your category $\mathcal{S}$. Over each $X \in \mathcal{S}$ there is monoid $\mathcal{S}_X$ of variables coarser than $X$. Denote by $\mathcal{A}_X$ the algebra generated over $\mathbb{R}$ by this monoid.
2. 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 groups with an action of $\mathcal{A}$ (the sheaf $F$ lives here!).
3. Define the information cohomology as (cf. Bennequin-Baudot, 2015 [1]):
$$H^n(\mathcal{S}, \mathcal{Q}) = \operatorname{Ext}^n(\mathbb{R}_{\mathcal{S}}, F).$$
4. The bar resolution construction allows us to construct a complex
$$0 \longrightarrow C^0 \xrightarrow{\delta_0} C^1 \xrightarrow{\delta_1} C^2 \xrightarrow{\delta_2} \dots$$
and compute $H^n(\mathcal{S}, \mathcal{Q}) \cong \ker \delta_n / \operatorname{im} \delta_{n-1}$.

# Back to the observables.

Set $\Omega = \{1,2,3\}$, $X_i := \{\{i\}, \Omega \setminus \{i\}\}$, $M$ atomic, $\Delta^k$ the $k$-simplex.

$$\begin{tikzcd}
& \mathbf{1}_\Omega & \\
X_1 \arrow[ur] & & X_2 \arrow[ul] \\
& M \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

The general construction says that a 1-cocycle is defined by 3 functions $f[X_1] : \underbrace{\mathcal{Q}_{X_1}}_{=\Delta^1} \to \mathbb{R}$, $f[X_2] : \mathcal{Q}_{X_2} \to \mathbb{R}$, $f[M] : \mathcal{Q}_M \to \mathbb{R}$ such that

... a 1-cocycle is defined by 3 functions $f[X_1] : \underbrace{\mathcal{Q}_{X_1}}_{=\Delta^1} \to \mathbb{R}$, $f[X_2] : \mathcal{Q}_{X_2} \to \mathbb{R}, f[M] : \mathcal{Q}_M \to \mathbb{R}$ such that

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

(The conditions for being in the kernel of $\delta_1$, like $\frac{\partial f_1}{\partial y} - \frac{\partial f_2}{\partial x} = 0$... but more complicated.)

These are functional equations (!), each term is a function. They imply $X_2.f[X_1] + f[X_2] = X_1.f[X_2] + f[X_1]$ and if you plug a particular probability $(p_0, p_1, p_2)$ here, you obtain

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

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

People (Tverberg, Lee, Ng, etc.) have proved that the only measurable solution to this equation are

$$f[X_1](x,1-x) = f[X_2](x,1-x) = \lambda(-x\log x - (1-x)\log(1-x))$$

where $\lambda$ is an arbitrary constant.

$$(1-p_2)f[X_1]\left(\frac{p_0}{1-p_2},\frac{p_1}{1-p_2}\right)-f[X_1](1-p_1,p_1)$$
$$=(1-p_1)f[X_2]\left(\frac{p_0}{1-p_1},\frac{p_2}{1-p_1}\right)-f[X_2](1-p_2,p_2).$$

People (Tverberg, Lee, Ng, etc.) have proved that the only measurable solution to this equation are

$$f[X_1](x,1-x)=f[X_2](x,1-x)=\lambda(-x\log x-(1-x)\log(1-x))$$

where $\lambda$ is an arbitrary constant.

This means that, in fairly general situations, the information cohomology $H^1(\mathcal{S},\mathcal{Q})$ is a 1-dimensional vector space, all cocycles being multiples of entropy function.

An interesting idea is to see the information category as a primary object and $\Omega$ as a derived one. In this view, the observables (the objects of $\mathcal{S}$) correspond to physical procedures and the arrows to particular ways of "attaching" one observable to another (given by certain protocol). A sample space corresponds to certain object that can be put "over" this category (see Gromov, 'On entropy' [2]).

Naïvely, we can start with certain category of (finite) observables and associate to it an initial object. This object is another set, whose elements correspond to combinations of compatible observations.

How many things can we see in this cohomology groups?
What are the higher cohomology groups?

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

M. Gromov, *In a search for a structure, part 1: On entropy.*, (2013).

**Thank you!**