Cohomological aspects of information

Juan Pablo Vigneaux

Olga Taussky and John Todd Instructor in Mathematics
Department of Mathematics
California Institute of Technology
vigneaux@caltech.edu

Caltech

Topos Institute Colloquium, January 25th, 2023

Table of Contents

Outline

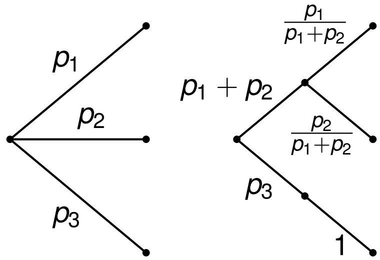
1. Foundations of information cohomology
2. Some recent results and open problems

Two perspectives on information

Shannon "justifies" the entropy $S_1(p_1, \ldots, p_s) = -\sum_{i=1}^{s} p_i \ln p_i$ in two different ways.

1. **Algebraic:** up to a factor, only continuous function that satisfies a certain recursive property, the *chain rule*.

$$S_1(p_1, p_2, p_3) =$$
$$S_1(p_1 + p_2, p_3) + (p_1 + p_2) S_1\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right)$$



2. **Probabilistically:** if words are generated by an i.i.d. process of discrete variables with law $(p_1, \ldots, p_s)$, when n large there are roughly $\exp(nH)$ "typical" words, all roughly equiprobable.

A new algebraic perspective

According to the *information cohomology* first introduced by Baudot and Bennequin in 2015 [1], Shannon entropy represents a cohomology class: an invariant associated with a category of discrete observables.

Information structure: pair $(\mathbf{S}, \mathcal{E})$ where

1. **S** conditional meet semilattice: poset + conditional existence of products (when there's a common lower bound). Objects represent *observables*.
2. $\mathcal{E} : \mathbf{S} \to \mathbf{Sets}$: functor of possible outcomes; $\mathcal{E}(X \to Y)$ surjection; $\mathcal{E}(X \wedge Y) \subset \mathcal{E}(X) \times \mathcal{E}(Y)$.

Π *covariant* functor of probabilities on $\mathcal{E}$, such that $\Pi(X \to Y)$ is marginalization.

Real-valued functions of probabilities define a contravariant one.

Information cohomology: Definition

For $X \in \mathrm{Ob}\,\mathbf{S}$, $\mathcal{S}_X := (\{\, Y \,:\, X \to Y \,\}, \wedge)$ monoid, multiplication $(Y, Z) \mapsto Y \wedge Z$ is 'joint variable'.

$X \mapsto \mathcal{S}_X$ presheaf, and also $X \mapsto \mathcal{A}_X := \mathbb{R}[\mathcal{S}_X]$.

Information cohomology: Definition

For $X \in \mathrm{Ob}\,\mathbf{S}$, $\mathcal{S}_X := (\{\, Y \,:\, X \to Y \,\}, \wedge)$ monoid, multiplication $(Y, Z) \mapsto Y \wedge Z$ is 'joint variable'.

$X \mapsto \mathcal{S}_X$ presheaf, and also $X \mapsto \mathcal{A}_X := \mathbb{R}[\mathcal{S}_X]$.

Category $\mathbf{Mod}(\mathcal{A})$ of $\mathcal{A}$-modules is abelian $\Rightarrow$ Derived functors $D^\bullet \operatorname{Hom}_{\mathcal{A}}(\mathbb{R}, -) =: \operatorname{Ext}^\bullet(\mathbb{R}, -)$.

Definition

The **information cohomology** with coefficients in an $\mathcal{A}$-module M is

$$H^\bullet(\mathbf{S}, M) := \operatorname{Ext}^\bullet(\mathbb{R}, M).$$

Bar resolution $B_\bullet \to \mathbb{R}$ gives differential complex $(\operatorname{Nat}(B_\bullet, M), \delta)$ whose cohomology is $H^\bullet(\mathbf{S}, M)$.

1-cocycle condition: $m_Z[XY] = X.m_Z[Y] + m_Z[X]$, an equation in $M(Z)$, for every $Z \in \mathrm{Ob}\,\mathbf{S}$ and $X, Y \in \mathcal{S}_Z$.

Probabilistic functionals

$(\mathbf{S}, \mathcal{E})$ information structure. We assume that *each* $\mathcal{E}(X)$ *is finite*.
$\mathcal{S}(X) := (\{\, Y \,:\, X \to Y \,\}, \wedge)$.

Define $\mathcal{F}(X) = \mathrm{Meas}(\Pi(X), \mathbb{R})$, a vector space. The monoid $\mathcal{S}(X)$ acts on it: for $Y \in \mathcal{S}(X)$ and $\phi \in \mathcal{F}(X)$,

$$(Y.\phi)(P) = \sum_{y \in \mathcal{E}(Y),\, P(Y=y) \neq 0} P(Y = y)\phi(P_X|_{Y=y}). \tag{1}$$

Proposition (Baudot-Bennequin, 2015, [1])

When $\mathbf{S}$ *is connected and nondegenerate, every* 1*-cocycle is given by a multiple of* $S_1[X] = -\sum_{x \in \mathcal{E}(X)} P(x) \log P(x)$.
In the nonconnected case $H^1(\mathbf{S}, \mathcal{F}) \cong \mathbb{R}^{|\pi_0(\mathbf{S} \setminus \{\top\})|}$.

By functoriality, for any 1-cochain, $\phi_Z[X](P) = \phi_X[X](P)$. Above $S_1[X] = (S_1)_X[X]$.

Probabilistic functionals II

More general action: for $Y \in \mathcal{S}(X)$ and $\phi \in \mathcal{F}(X)$,

$$(Y.\phi)(P) = \sum_{y \in \mathcal{E}(Y),\, P(Y=y) \neq 0} P(Y=y)^{\alpha} \phi(P_X|_{Y=y}). \tag{2}$$

Proposition (V. 2017; cf. [5])

When $\mathbf{S}$ *is connected and nondegenerate, every* 1*-cocycle is given by a multiple of* $S_1[X] = \sum_{x \in \mathcal{E}(X)} P(x)^{\alpha} - 1$.
In the nonconnected case $H^1(\mathbf{S}, \mathcal{F}) \cong \mathbb{R}^{|\pi_0(\mathbf{S} \setminus \{\top\})| - 1}$.

Under nondegeneracy assumption, the problem locally reduces to considering four possible partitions of a three point set, and the 1-cocycle can be reduced to solving

$$f(x) + (1-x)^{\alpha} f\left(\frac{y}{1-x}\right) = f(y) + (1-y)^{\alpha} f\left(\frac{x}{1-y}\right).$$

Combinatorial functionals

C : **S** $\to$ **Sets** given by $C(X) = \{\, \nu : \mathcal{E}(X) \to \mathbb{N} \ : \ \|\nu\| := \sum_{x \in \mathcal{E}(X)} \nu(x) > 0 \,\}$ (frequencies).
$\pi_* = C(X \to Y)$ marginalization: $\pi_* \nu(y) = \sum_{x \in \pi^{-1}(y)} \nu(x)$.

$G(X)$: multiplicative abelian group of measurable $(0, \infty)$-valued functions on $C(X)$. Induces contravariant functor G on **S**.

For each $Y \in \mathcal{S}(X)$ and $\phi \in G(X)$, define

$$(Y.\phi)(\nu) = \prod_{\substack{y \in \mathcal{E}(Y) \\ \nu(Y=y) \neq 0}} \phi(\nu|_{Y=y_i}). \tag{3}$$

where $\nu|_{Y=y_i}$ is a restriction.

Computing $H^\bullet(\mathbf{S}, G)$

Proposition (V. 2019; cf. [7])

1. *$H^0(\mathbf{S}, G)$ has dimension 1 and is generated by the exponential function.*
2. *The 1-cocycles are generalized (Fontené-Ward) multinomial coefficients:*

$$\phi[Y](\nu) = \frac{[\|\nu\|]_D!}{\prod_{y \in \mathcal{E}(Y)} [\nu(y)]_D!}$$

where $[0]_D! = 1$ and $[n]_D! = D_n D_{n-1} \cdots D_1$, for any sequence $\{D_i\}_{i \geq 1}$ such that $D_1 = 1$.

Computing $H^\bullet(\mathbf{S}, G)$

Proposition (V. 2019; cf. [7])

1. *$H^0(\mathbf{S}, G)$ has dimension 1 and is generated by the exponential function.*
2. *The 1-cocycles are generalized (Fontené-Ward) multinomial coefficients:*

$$\phi[Y](\nu) = \frac{[\|\nu\|]_D!}{\prod_{y\in\mathcal{E}(Y)}[\nu(y)]_D!}$$

where $[0]_D! = 1$ and $[n]_D! = D_n D_{n-1}\cdots D_1$, for any sequence $\{D_i\}_{i\geq 1}$ such that $D_1 = 1$.

The 0-cocycle condition reads: $\varphi(\|\nu\|) = \varphi(\nu_1)\varphi(\nu_2)\cdots\varphi(\nu_s)$.
The 1-cocycle confition reads: $\phi[XY] = (X.\phi[Y])\phi[X]$ e.g. $\binom{n}{k_1,k_2,k_3} = \binom{n}{k_1+k_2,k_3}\binom{k_1+k_2}{k_1,k_2}$

$D_n = n$: usual multinomial coefficients; $D_n = \frac{q^n-1}{q-1}$: the q-multinomial coefficients.

Coincidence?

Example:

- 0-cocycles: the exponential $\exp(k\,\|\nu\|)$ is a combinatorial 0-cocycle, the constant k is a probabilistic 0-cocycle.
- 1-cocycles:

$$\binom{n}{p_1 n, \ldots, p_s n} = \exp(nS_1(p_1, \ldots, p_s) + o(n))$$

and (cf. V., IEEE ToIT, 2019)

$$\begin{bmatrix} n \\ p_1 n, \ldots, p_s n \end{bmatrix}_q = \exp(n^2 \frac{\ln q}{2} S_2(p_1, \ldots, p_s) + o(n^2)).$$

Multiplicative relations imply additive relations

The combinatorial identity

$$\binom{n}{p_1 n, p_2 n, p_3 n} = \binom{n}{(p_1 + p_2)n, p_3 n}\binom{(p_1 + p_2)n}{p_1 n, p_2 n}$$

Multiplicative relations imply additive relations

The combinatorial identity

$$\binom{n}{p_1 n, p_2 n, p_3 n} = \binom{n}{(p_1+p_2)n, p_3 n}\binom{(p_1+p_2)n}{p_1 n, p_2 n}$$

becomes asymptotically

$$\begin{aligned}\exp(nS_1(p_1, p_2, p_3) + o(n)) = \\ \exp\left(n\left\{S_1(p_1+p_2, p_3) + (p_1+p_2)S_1\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)\right\} + o(n)\right).\end{aligned}$$

Asymptotic relation

Proposition (V. 2019; cf. [7])

Let ϕ be a combinatorial 1-cocycle. Suppose that, for every variable X, there exists a measurable function $\psi[X] : \Delta(X) \to \mathbb{R}$ with the following property: for every sequence of counting functions $\{\nu_n\}_{n\geq 1} \subset C(X)$ such that $\nu_n(x) \sim P(x)n$ the asymptotic formula

$$\phi[X](\nu_n) = \exp(n^\alpha \psi[X](P) + o(\|\nu_n\|^\alpha))$$

holds. Then ψ is a 1-cocycle of type α, i.e. $\psi \in Z^1(\mathbf{S}, \mathcal{F}_\alpha)$.

Vector-valued observables

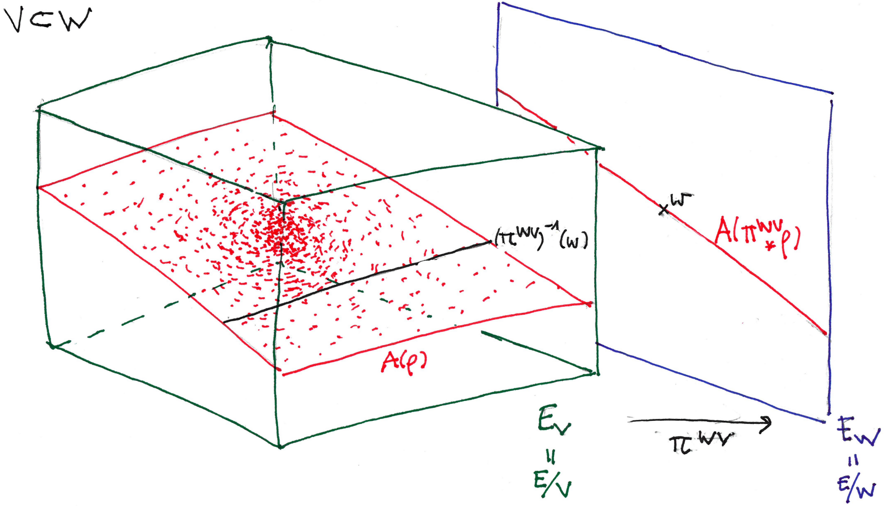
Let E be a f.d. euclidean space, and **S** a category of subspaces of E, with arrows corresponding to inclusions.
Conditionally closed under intersections: if Z, V, W are objects of **S** such that $Z \subset V$ and $Z \subset W$, then $V \cap W \in \mathrm{Ob}\,\mathbf{S}$.

Let $\mathcal{E}$ be the functor $V \mapsto E_V := V^\perp \cong E/V$, sending $V \subset W$ to the canonical projection $\pi^{WV} : E_V \to E_W$.

$\mathcal{P}(X)$: gaussian probability laws on E_X or some affine subspace of it ("supports" $\mathcal{N}$). Defines covariant functor $\mathcal{P}$.

$\mathcal{F}$: presheaf of functionals of probability laws, with polynomial growth in the mean, with Shannon's action: for $V \subset W$, and $\varphi \in \mathcal{F}_V$,

$$(W.\varphi)(\rho) := \int_{\pi^{WV}(A(\rho))} \varphi(\rho|_{X_W = w}) \,\mathrm{d}\pi_*^{WV}\rho(w), \tag{4}$$



$V \subset W$
$(\pi^{WV})^{-1}(w)$
w
$A(\pi^{WV}_* \rho)$
$A(\rho)$
$E_V = E/V$
π^{WV}
$E_W = E/W$

Dimension is a cocycle

If A is the support of $\rho \in \mathcal{P}_V$, then $\pi^{WV}(A)$ is the support of the marginal law $\pi_*^{WV}\rho \in \mathcal{P}_W$, and $(\pi^{WV})^{-1}(w)$ is the support of $\rho|_{X_W=w}$. One has the equality:

$$
\begin{aligned}
\dim(A) &= \dim(\pi^{WV}(A)) + \int_{\pi^{WV}(A)} \dim((\pi^{WV})^{-1}(w))\,\mathrm{d}\pi_*^{WV}\rho(w) \\
&= \dim(\operatorname{im}\pi^{WV}|_A) + \dim(\ker \pi^{WV}|_A).
\end{aligned}
$$

Cohomology of gaussian laws

Differential entropy $S(\rho) = -\int_{A(\rho)} \frac{\mathrm{d}\rho}{\mathrm{d}\lambda} \ln \frac{\mathrm{d}\rho}{\mathrm{d}\lambda}\,\mathrm{d}\lambda$ is not invariant under change of Lebesgue measure (changes in the ambient Euclidean metric). One introduces a extension $\mathcal{X}$ of $\mathcal{F}$ that takes this variations into account.

Theorem (V. 2019, [4])

Provided $(\mathbf{S}, \mathcal{E}, \mathcal{N})$ is sufficiently rich ("enough supports"), for every 1-cocycle φ with coefficients in $\mathcal{X}$, there are real constants a and c such that, for every $X \in \operatorname{Ob} \mathbf{S}$ and nondegenerate gaussian law ρ on E_X,

$$\varphi_X[X](\rho, \lambda) = a \det{}_\lambda(\operatorname{Cov}(\rho)) + c.\dim(\operatorname{supp}\rho). \tag{5}$$

φ is completely determined by its behavior on nondegenerate laws.

For gaussian probabilities, the 1-cocycle condition for differential entropy is Schur's determinantal formula $\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(A)\det(D - BA^{-1}C)$.

Cohomology of general laws

A similar result holds if we allow ourselves to mix discrete and continuous variables: $(\mathbf{S}_d, \mathcal{E}_d) \times (\mathbf{S}_c, \mathcal{E}_c)$.

The key tool is the approximation of any density by a convex combination of gaussian densities (cf. kernel estimates) in the L^1-norm. Equivalently: of the corresponding measures in the total variation norm.

Theorem (V. GSI 2021, [6])

Provided $(\mathbf{S}_c, \mathcal{E}_c, \mathcal{N})$ is sufficiently rich, for every 1-cocycle φ with coefficients in $\mathcal{X}_{\text{continuous}}$, there are real constants a and c such that, for every $X \in \operatorname{Ob} \mathbf{S}_c$ and "good" law ρ on E_X,

$$\varphi_X[X](\rho, \lambda) = aS_1(\rho, \lambda) + c.\dim(\operatorname{supp} \rho). \tag{6}$$

Outline

1. Foundations of information cohomology
2. Some recent results and open problems

Symmetry and entropy

The uniform distribution on a finite set E has maximum symmetry (under permutations of E) and maximum entropy. But this correlation breaks under any small perturbation.

We have to change perspective: thinks about long words with empirical law ρ. The group $\mathfrak{S}_n$ acts on E^n. If $\rho \approx \rho'$, then the stabilizers of the corresponding words are similar.

$$\binom{n}{p_1 n, \ldots, p_s n} = |\mathfrak{S}_n / (\mathfrak{S}_{p_1 n} \times \cdots \mathfrak{S}_{p_1 n})| = \exp(nS_1(p_1, \ldots, p_s) + o(n)).$$

New formulae

1. V., IEEE ToIT, 2019 [3]:

$$\begin{bmatrix} n \\ p_1 n, \dots, p_s n \end{bmatrix}_q = |GL_n(\mathbb{F}_q)/P| = \exp(n^2 S_2(p_1, \dots, p_s) + o(n^2)),$$

where P is a parabolic subgroup: it stabilizes a flag $V_1 \subset V_2 \subset \cdots \subset V_n = \mathbb{F}_q^n$ such that $\dim V_i = \sum_{j=1}^{i} p_i n$.

2. Leal and V., to appear: when $n \to \infty$,

$$\frac{1}{n} \ln |W/P| \sim H(p_1, \dots, p_s) + (1 - p_s) \ln 2, \quad \frac{1}{n^2} \log_q |G/P| \sim \frac{1}{2} H_2(p_1, \dots, p_s) + (1 - p_s^2),$$

where W is a reflection group of type B, C or D; $G = \mathrm{Sp}_{2n}(\mathbb{F}_q)$ (the $O_n(\mathbb{F}_q)$ case is similar), and P stands for a suitable parabolic subgroup, stabilizer of a flag of type $(p_1 n, \dots, p_s n)$.

Are these cocycles in information cohomology? For which module?

Dimension

Renyi's information dimension generalizes the vector-space dimension discussed above.

For a probability measure ρ on $\mathbb{R}^d$, ρ_n is a discretization obtained by partitioning $\mathbb{R}^d$ into cubes of size $1/n$. If

$$S_1(\rho_n) = D \ln n + h + o(1),$$

D is called the information dimension of ρ.

Conjecture: It is also a 1-cocycle, but for more general laws (not necessarily vector valued).

We have shown (IEEE ToIT, 2023, [8]) there is $W^{(n)} \subset (\mathbb{R}^d)^n$ such that $\rho^{\otimes n}(W^{(n)}) \geq 1 - \varepsilon$ and $W^{(n)} = \bigcup_{k \in [Dn - \sqrt{n} \ln n, DN + \sqrt{n} \ln n]} W_k$ with W_k rectifiable of Hausdorff dimension k.

Homogeneous spaces

S category of subgroups of a locally compact topological group G.

$\mathcal{E}(N) = G/N$: associated outcome space.
$N \subset N'$ implies $G/N \to G/N'$ surjection.

Introduce again the probabilistic functionals $\mathcal{F}$. They form an $\mathcal{A}$-module: the proof in the Euclidean case only depends on Weil's disintegration formula).

What are the cocycles in this case? They might capture more topological invariants than just the dimension. E.g. for an orientable fibration

$$\ln|\chi(G)| = \ln|\chi(G/N)| + \ln|\chi(N)|.$$

Entropy of categories

The magnitude is a categorical generalization cardinality. Entropy extends cardinality probabilistically. What is categorical entropy?

Chen and V., GSI 2023 [2]: akin to log-diversity of metric spaces,

$$\mathcal{H}(\mathbf{A}, p, \theta) = -\sum_{a \in \mathrm{Ob}\,\mathbf{A}} p(a) \ln \left(\sum_{b \in \mathrm{Ob}\,\mathbf{A}} \theta(a, b) p(b) \right).$$

Here $\mathbf{A}$ is a finite category, p a probability on $\mathrm{Ob}\,\mathbf{A}$ and $\theta(a, a')$ a function of pairs of objects that vanishes when $\mathrm{Hom}(a', a) = \emptyset$. Several good properties.

Algebraic characterization?
$F((\mathbf{A}, p, \theta) \to (\mathbf{B}, q, \phi)) = \mathcal{H}(\mathbf{A}, p, \theta) - \mathcal{H}(\mathbf{B}, q, \phi)$ defines a convex and continuous functor to the reals in the sense of Baez-Fritz-Leinster, but here this does not suffice to characterize. Problem: whereas every finite probability space is a convex combination of singletons, this does not hold for a categorical triple $(\mathbf{A}, p, \theta)$.
How about an homological characterization?

P. Baudot and D. Bennequin.
The homological nature of entropy.
Entropy, 17(5):3253–3318, 2015.

S. Chen and J. P. Vigneaux.
Categorical magnitude and entropy.
In F. Nielsen and F. Barbaresco, editors, *Geometric Science of Information: 6th International Conference, GSI 2023*, volume 14071 of *Lecture Notes in Computer Science*, pages 278–287, Cham, 2023. Springer Nature Switzerland.

J. P. Vigneaux.
Information theory with finite vector spaces.
IEEE Transactions on Information Theory, 65(9):5674–5687, Sep. 2019.

J. P. Vigneaux.
Topology of Statistical Systems: A Cohomological Approach to Information Theory.
PhD thesis, Université de Paris, 2019.
Available at
https://jpvigneaux.github.io/assets/pdf/manuscripts/thesis-v11022020.pdf.

J. P. Vigneaux.
Information structures and their cohomology.
Theory and Applications of Categories, 35(38):1476–1529, 2020.

J. P. Vigneaux.
Information cohomology of classical vector-valued observables.
In F. "Nielsen and F. Barbaresco, editors, *GSI 2021: Geometric Science of Information*, volume 12829 of *Lecture Notes in Computer Science*, pages 537–546. Springer, 2021.

J. P. Vigneaux.
A characterization of generalized multinomial coefficients related to the entropic chain rule.
Aequationes mathematicae, 97(2):231–255, 2023.

J. P. Vigneaux.
Typicality for stratified measures.
IEEE Transactions on Information Theory, 69(11):6922 – 6940, Nov. 2023.