Topology of statistical systems

A cohomological approach to information theory

Juan Pablo Vigneaux Ariztía
IMJ-PRG
Université Paris Diderot

Paris, June 14, 2019

Outline

1. Introduction
   - Entropy and recurrence
   - Combinatorics
2. Generalized information theory
3. Information structures and their cohomology
   - Foundations
   - Cohomology of discrete variables
   - Cohomology of continuous (gaussian) variables

Outline

Shannon entropy

Shannon (1948) characterized the functions $S_1 := S_1^{(n)} : \Delta^n \to \mathbb{R}$, given by

$$S_1(p_0, \ldots, p_n) := -\sum_{k=0}^{n} p_i \ln p_i, \tag{1}$$

as the only continuous functions (up to a multiplicative constant), such that $S_1(1/n, \ldots, 1/n)$ is monotonic in n and

Shannon entropy

Shannon (1948) characterized the functions $S_1 := S_1^{(n)} : \Delta^n \to \mathbb{R}$, given by

$$S_1(p_0, ..., p_n) := -\sum_{k=0}^{n} p_i \ln p_i, \tag{1}$$

as the only continuous functions (up to a multiplicative constant), such that $S_1(1/n, ..., 1/n)$ is monotonic in n and

"If a choice be broken down into two successive choices, the original S_1 should be the weighted sum of the individual values of S_1."

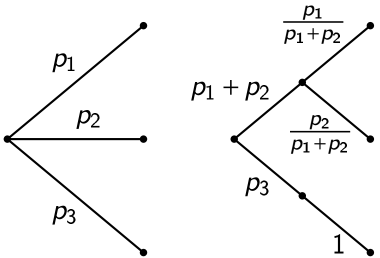
Shannon entropy

Shannon (1948) characterized the functions $S_1 := S_1^{(n)} : \Delta^n \to \mathbb{R}$, given by

$$S_1(p_0, \dots, p_n) := -\sum_{k=0}^{n} p_i \ln p_i, \tag{1}$$

as the only continuous functions (up to a multiplicative constant), such that $S_1(1/n, \dots, 1/n)$ is monotonic in n and

"If a choice be broken down into two successive choices, the original S_1 should be the weighted sum of the individual values of S_1."



$$S_1(p_1, p_2, p_3) =$$

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

Structural α-entropy (Tsallis α-entropy)

Havrda and Charvát introduced a generalized entropy, for each $\alpha > 0$, $\alpha \neq 1$:

$$S_\alpha(p_0, \ldots, p_n) = K_\alpha \left(1 - \sum_{i=0}^{n} p_i^\alpha\right). \tag{2}$$

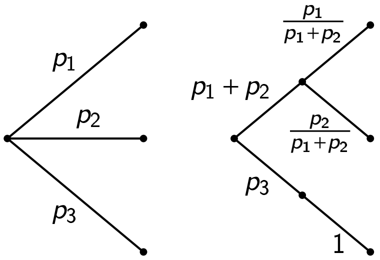
It satisfies a deformed version of the axioms.

Structural α-entropy (Tsallis α-entropy)

Havrda and Charvát introduced a generalized entropy, for each $\alpha > 0$, $\alpha \neq 1$:

$$S_\alpha(p_0, \dots, p_n) = K_\alpha \left(1 - \sum_{i=0}^{n} p_i^\alpha \right). \tag{2}$$

It satisfies a deformed version of the axioms.



$$S_1(p_1, p_2, p_3) =$$

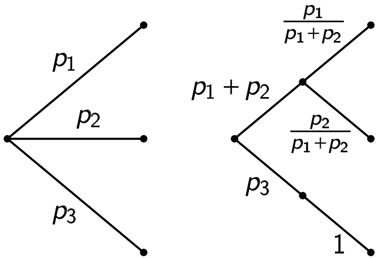
$$S_1(p_1 + p_2, p_3) + (p_1 + p_2)^\alpha S_1\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2} \right)$$

Structural α-entropy (Tsallis α-entropy)

Havrda and Charvát introduced a generalized entropy, for each $\alpha > 0$, $\alpha \neq 1$:

$$S_\alpha(p_0, \dots, p_n) = K_\alpha \left(1 - \sum_{i=0}^{n} p_i^\alpha\right). \tag{2}$$

It satisfies a deformed version of the axioms.



$$S_1(p_1, p_2, p_3) =$$

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

Why these axioms? What is their role in information theory?

Outline

1. Introduction
   - Entropy and recurrence
   - Combinatorics
2. Generalized information theory
3. Information structures and their cohomology
   - Foundations
   - Cohomology of discrete variables
   - Cohomology of continuous (gaussian) variables

Multinomial coefficients

The multinomial coefficient

$$\binom{n}{k_1,\ldots,k_s} := \frac{n!}{k_1!\cdots k_s!} = \frac{\Gamma(n+1)}{\Gamma(k_1+1)\cdots\Gamma(k_s+1)}$$

counts the number of words $w \in \Sigma^n$, with $\Sigma = \{\sigma_1,\ldots,\sigma_s\}$, such that σ_i appears k_i times.

Multinomial coefficients

The multinomial coefficient

$$\binom{n}{k_1,\dots,k_s} := \frac{n!}{k_1!\cdots k_s!} = \frac{\Gamma(n+1)}{\Gamma(k_1+1)\cdots\Gamma(k_s+1)}$$

counts the number of words $w \in \Sigma^n$, with $\Sigma = \{\sigma_1,\dots,\sigma_s\}$, such that σ_i appears k_i times.

From the point of view of probability and combinatorics, Shannon entropy $S_1(p_1,\dots,p_s) = -\sum_{i=1}^{s} p_i \ln p_i$ appears naturally in the asymptotic formula

$$\binom{n}{p_1 n,\dots,p_s n} = \exp(nS_1(p_1,\dots,p_s) + O(\ln n)) \tag{3}$$

q-multinomials

Let q be an indeterminate. Define

1. q-integers $[n]_q = \frac{q^n-1}{q-1}$,
2. q-factorials: $[n]_q! := [n]_q[n-1]_q \cdots [1]_q$.
3. q-multinomial coefficients by

$$\begin{bmatrix} n \\ k_1, \ldots, k_s \end{bmatrix}_q := \frac{[n]_q!}{[k_1]_q! \cdots [k_s]_q!}, \tag{4}$$

where $k_1, \ldots, k_s$ are such that $\sum_{i=1}^{s} k_i = n$.

q-multinomials

Let q be an indeterminate. Define

1. q-integers $[n]_q = \frac{q^n-1}{q-1}$,
2. q-factorials: $[n]_q! := [n]_q[n-1]_q \cdots [1]_q$.
3. q-multinomial coefficients by

$$\begin{bmatrix} n \\ k_1, \ldots, k_s \end{bmatrix}_q := \frac{[n]_q!}{[k_1]_q! \cdots [k_s]_q!}, \tag{4}$$

where $k_1, \ldots, k_s$ are such that $\sum_{i=1}^{s} k_i = n$.

Proposition

When q is a prime power, $\left[\begin{smallmatrix} n \\ k_1, \ldots, k_s \end{smallmatrix}\right]_q$ counts the number of flags of vector spaces $V_1 \subset V_2 \subset \ldots \subset V_s = \mathbb{F}_q^n$ such that $\dim V_i = \sum_{j=1}^{i} k_j$.

Tsallis 2-entropy in combinatorics

Proposition (V., 2018)

Let $(p_1, \dots, p_s)$ *be a probability. Then,*

$$\begin{bmatrix} n \\ p_1 n, \dots, p_s n \end{bmatrix}_q = q^{n^2 S_2(p_1,\dots,p_s)/2 + o(n^2)}. \tag{5}$$

Here

$$S_2(p_1, \dots, p_s) := 1 - \sum_{i=1}^{s} p_i^2 \tag{6}$$

is Tsallis 2-entropy (with the appropriate leading constant).

Recurrence (Shannon entropy)

The combinatorial identity

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

Recurrence (Shannon entropy)

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

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

Recurrence (α-entropy)

Since

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

Recurrence (α-entropy)

Since

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

the multiplicative relation

$$\begin{bmatrix} n \\ p_1 n, p_2 n, p_3 n \end{bmatrix}_q = \begin{bmatrix} n \\ (p_1+p_2)n, p_3 n \end{bmatrix}_q \begin{bmatrix} (p_1+p_2)n \\ p_1 n, p_2 n \end{bmatrix}_q,$$

Recurrence (α-entropy)

Since

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

the multiplicative relation

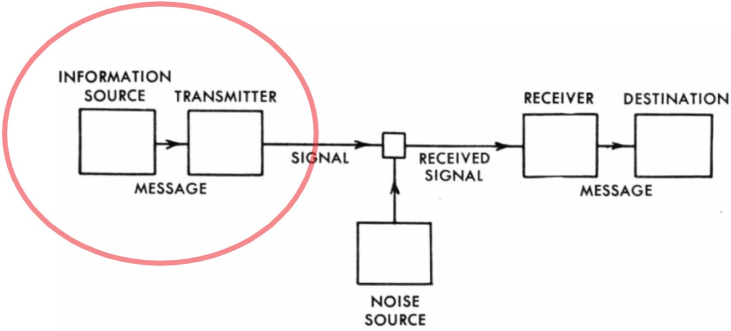
$$\begin{bmatrix} n \\ p_1 n, p_2 n, p_3 n \end{bmatrix}_q = \begin{bmatrix} n \\ (p_1+p_2)n, p_3 n \end{bmatrix}_q \begin{bmatrix} (p_1+p_2)n \\ p_1 n, p_2 n \end{bmatrix}_q,$$

implies

$$S_2(p_1,p_2,p_3) = S_2(p_1+p_2,p_3) + (p_1+p_2)^2 S_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right).$$

Outline

1. Introduction
   - Entropy and recurrence
   - Combinatorics
2. Generalized information theory
3. Information structures and their cohomology
   - Foundations
   - Cohomology of discrete variables
   - Cohomology of continuous (gaussian) variables



"[T]he actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design." (Shannon)

The source is described by a probabilistic model that quantifies the likelihood of any possible message.

A q-deformation of Shannon's theory

Concept	**Shannon case**	**q-case**
Message at time n (n-message)	Word $w \in \{0,1\}^n$	Vector subspace $v \subset \mathbb{F}_q^n$
Type	Number of ones	Dimension
Number of n-messages of type k	$\binom{n}{k}$	$\begin{bmatrix} n \\ k \end{bmatrix}_q$
Probability of a n-message of type k	$\xi^k (1-\xi)^{n-k}$	$\dfrac{\theta^k q^{k(k-1)/2}}{(-\theta;q)_n}$

Where q is a prime power, and $\theta > 0$, $\xi \in [0,1]$ are arbitrary parameters.

The q-binomial distribution

The binomial theorem implies that, for any $\xi \in [0,1]$,

$$1 = \sum_{k=0}^{n} \binom{n}{k} \xi^k (1-\xi)^{n-k}, \tag{7}$$

and also that $Y \sim \mathrm{Bin}(n,\xi)$ appears as the sum $Z_1 + \cdots Z_n$, with $Z_i \sim \mathrm{Ber}(\xi)$.

In turn, Gauss' binomial formula says that

$$\underbrace{(1+\theta)(1+\theta q)\cdots(1+\theta q^{n-1})}_{=:(-\theta;q)_n} = \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_q \theta^k q^{k(k-1)/2}. \tag{8}$$

The q-**binomial distribution** has probability mass function $k \mapsto \left[{n \atop k}\right]_q \frac{\theta^k q^{k(k-1)/2}}{(-\theta;q)_n}$. A variable $Y \sim \mathrm{Bin}_q(n,\theta)$ equals (in law) $X_1 + \cdots + X_n$, where $X_i \sim \mathrm{Ber}\left(\frac{\theta q^{i-1}}{1+\theta q^{i-1}}\right)$.

Grassmannian process I

Fix a sequence of embeddings $\mathbb{F}_q^1 \hookrightarrow \mathbb{F}_q^2 \hookrightarrow \ldots$.

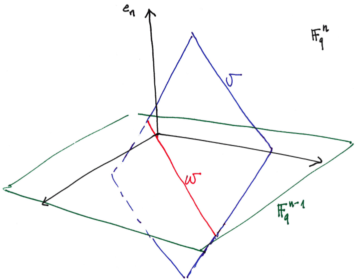
We introduce a stochastic process that generates at time n a generalized message $V_n \subset \mathbb{F}_q^n$ such that

1. V_{n-1} can be recovered from V_n (as $V_n \cap \mathbb{F}_q^{n-1}$);
2. the probability of $V_n = v$, when $\dim v = k$, equals

$$\frac{\theta^k q^{k(k-1)/2}}{(-\theta; q)_n}.$$

Grassmannian process II: How?

For $w \subset \mathbb{F}_q^{n-1}$,



$$\mathrm{Dil}_n(w) := \{ v \subset \mathbb{F}_q^n \mid \dim v - \dim w = 1, \\ w \subset v \text{ and } v \not\subset \mathbb{F}_q^{n-1} \}.$$

$\{X_i\}_{i\in\mathbb{N}}$ are independent, $X_i \sim \mathrm{Ber}\left(\frac{\theta q^{i-1}}{1+\theta q^{i-1}}\right)$. Set $V_0 = 0$ and, at time n,

1. if $X_n = 0$, do nothing $V_n = V_{n-1}$;
2. if $X_n = 1$, increase dimension: pick V_n at random, uniformly, from $\mathrm{Dil}_n(V_{n-1})$.

Theorem (Generalized AEP)

For every $\delta > 0$ *and almost every* $\varepsilon > 0$ *(except a countable set), there exist* $n_0 \in \mathbb{N}$ *and sets* $A_n = \bigcup_{k=0}^{\Delta(\varepsilon)} \mathrm{Gr}(n-k, n)$, *for all* $n \geq n_0$, *such that*

Theorem (Generalized AEP)

For every $\delta > 0$ and almost every $\varepsilon > 0$ (except a countable set), there exist $n_0 \in \mathbb{N}$ and sets $A_n = \bigcup_{k=0}^{\Delta(\varepsilon)} \mathrm{Gr}(n-k, n)$, for all $n \geq n_0$, such that

1. *$\Delta(\varepsilon)$ is an integer that just depends on ε;*

Theorem (Generalized AEP)

For every $\delta > 0$ and almost every $\varepsilon > 0$ (except a countable set), there exist $n_0 \in \mathbb{N}$ and sets $A_n = \bigcup_{k=0}^{\Delta(\varepsilon)} \mathrm{Gr}(n-k, n)$, for all $n \geq n_0$, such that

1. $\Delta(\varepsilon)$ *is an integer that just depends on* ε*;*
2. $\mathbb{P}\left(V_n \in A_n^c\right) \leq \varepsilon$*;*

Theorem (Generalized AEP)

For every $\delta > 0$ and almost every $\varepsilon > 0$ (except a countable set), there exist $n_0 \in \mathbb{N}$ and sets $A_n = \bigcup_{k=0}^{\Delta(\varepsilon)} \mathrm{Gr}(n-k, n)$, for all $n \geq n_0$, such that

1. $\Delta(\varepsilon)$ *is an integer that just depends on* ε;
2. $\mathbb{P}\left(V_n \in A_n^c\right) \leq \varepsilon$;
3. *for any* $v \in A_n$ *such that* $\dim v = k$,

$$\left| \frac{\log_q(\mathbb{P}\left(V_n = v\right)^{-1})}{n} - \frac{n}{2} S_2(k/n) \right| \leq \delta. \tag{9}$$

Theorem (Generalized AEP)

For every $\delta > 0$ and almost every $\varepsilon > 0$ (except a countable set), there exist $n_0 \in \mathbb{N}$ and sets $A_n = \bigcup_{k=0}^{\Delta(\varepsilon)} \mathrm{Gr}(n-k,n)$, for all $n \geq n_0$, such that

1. *$\Delta(\varepsilon)$ is an integer that just depends on ε;*
2. *$\mathbb{P}(V_n \in A_n^c) \leq \varepsilon$;*
3. *for any $v \in A_n$ such that $\dim v = k$,*

$$\left| \frac{\log_q(\mathbb{P}(V_n = v)^{-1})}{n} - \frac{n}{2} S_2(k/n) \right| \leq \delta. \tag{9}$$

Moreover, the size of A_n is optimal, up to the first order in the exponential: if $s(n,\varepsilon)$ is the cardinality of the smallest set of subspaces of $\mathbb{F}_q^n$ that accumulate probability $1-\varepsilon$, then

$$\lim_n \frac{1}{n} \log_q |A_n| = \lim_n \frac{1}{n} \log_q s(n,\varepsilon) = \lim_n \frac{n}{2} S_2(\Delta(\varepsilon)/n) = \Delta(\varepsilon). \tag{10}$$

Outline

Outline

1. Introduction
   - Entropy and recurrence
   - Combinatorics
2. Generalized information theory
3. Information structures and their cohomology
   - Foundations
   - Cohomology of discrete variables
   - Cohomology of continuous (gaussian) variables

Definition

An **information structure** is a couple $(\mathbf{S}, M)$, where $\mathbf{S}$ (observables, experiments, variables...) is a small category such that

1. $\mathbf{S}$ has a terminal object, denoted $\mathbf{1}$;
2. $\mathbf{S}$ is a skeletal partially ordered set (poset);
3. for objects $X, Y, Z \in \mathrm{Ob}\, S$, if $Z \to X$ and $Z \to Y$, then the categorical product $X \wedge Y$ exists;

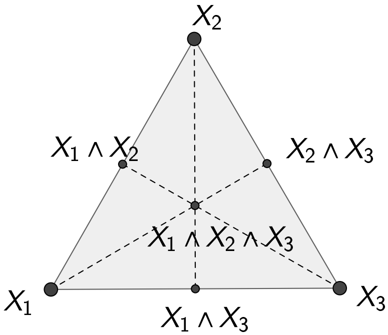
and M is a conservative covariant functor (the possible outputs) from $\mathbf{S}$ into the category $\mathbf{MeasurableSpace}_{surj}$, $X \mapsto M(X) = (E(X), B(X))$, that satisfies

4. $E(\mathbf{1}) \cong \{*\}$;
5. for every $X \in \mathrm{Ob}\,\mathbf{S}$ and any $x \in E(X)$, the σ-algebra $B(X)$ contains the singleton $\{x\}$;
6. for every diagram $X \xleftarrow{\pi} X \wedge Y \xrightarrow{\sigma} Y$ the measurable map $E(X \wedge Y) \hookrightarrow E(X) \times E(Y), z \mapsto (x(z), y(z)) := (\pi_*(z), \sigma_*(z))$ is an injection.

$$
\begin{tikzcd}
 & X \wedge P \arrow[r] \arrow[dr] & P \arrow[dr] & \\
P \wedge V \arrow[ur] \arrow[rr] \arrow[dr] & & X \arrow[r] & \mathbf{1} \\
 & X \wedge V \arrow[r] \arrow[ur] & V \arrow[ur] &
\end{tikzcd}
$$

$$
\begin{tikzcd}
 & (E_X \times E_P, 2^{E_X} \otimes \mathfrak{P}) \arrow[r] \arrow[dr] & (E_P, \mathfrak{P}) \arrow[dr] & \\
(M, \mathfrak{M}) \arrow[ur] \arrow[rr] \arrow[dr] & & (E_X, 2^{E_X}) \arrow[r] & \{*\} \\
 & (E_X \times E_V, 2^{E_X} \otimes \mathfrak{V}) \arrow[r] \arrow[ur] & (E_V, \mathfrak{V}) \arrow[ur] &
\end{tikzcd}
$$

where $M \hookrightarrow E_P \times E_V$ and $E_X := \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$.



with $E(X_i)$ arbitrary, $E(X_i \wedge X_j) = E(X_i) \times E(X_j)$, and $E(X_1 \wedge X_2 \wedge X_3) = E(X_1) \times E(X_2) \times E(X_2)$.

Information topos

The category of covariant functors $[\mathbf{S}, \mathbf{Sets}]$ as well as that of contravariant functors $[\mathbf{S}^{op}, \mathbf{Sets}]$ are important in applications. (For example, probabilities define a covariant functor; the probabilitic functionals—like entropy—, a contravariant one.)

Information topos

The category of covariant functors $[\mathbf{S}, \mathbf{Sets}]$ as well as that of contravariant functors $[\mathbf{S}^{op}, \mathbf{Sets}]$ are important in applications. (For example, probabilities define a covariant functor; the probabilitic functionals—like entropy—, a contravariant one.)

Both categories are Grothendieck topoi (related to geometric intuitions, logical interpretations, general algebraic constructions...).

Information topos

The category of covariant functors $[\mathbf{S}, \mathbf{Sets}]$ as well as that of contravariant functors $[\mathbf{S}^{op}, \mathbf{Sets}]$ are important in applications. (For example, probabilities define a covariant functor; the probabilitic functionals—like entropy—, a contravariant one.)

Both categories are Grothendieck topoi (related to geometric intuitions, logical interpretations, general algebraic constructions...).

There is an appropriate notion of localization: marginalizations induced by the surjections of M.

Information topos

The category of covariant functors $[\mathbf{S}, \mathbf{Sets}]$ as well as that of contravariant functors $[\mathbf{S}^{op}, \mathbf{Sets}]$ are important in applications. (For example, probabilities define a covariant functor; the probabilitic functionals—like entropy—, a contravariant one.)

Both categories are Grothendieck topoi (related to geometric intuitions, logical interpretations, general algebraic constructions...).

There is an appropriate notion of localization: marginalizations induced by the surjections of M.

We want to study the geometrical invariants attached to these functors (that are algebraic analogs of vector bundles).

Information cohomology: Definition

For each $X \in \mathrm{Ob}\,\mathbf{S}$, the set $\mathscr{S}_X := \{\, Y \,|\, X \to Y \,\}$ is a monoid under the multiplication $(Y, Z) \mapsto Y \wedge Z$ ('joint variable').

Each arrow $X \to Y$ induces an inclusion $\mathscr{S}_Y \to \mathscr{S}_X$, which defines a particular presheaf (contravariant functor). Let $\mathscr{A}$ denote the presheaf of algebras $X \mapsto \mathbb{R}[\mathscr{S}_X]$ (finite linear combinations).

Information cohomology: Definition

For each $X \in \mathrm{Ob}\,\mathbf{S}$, the set $\mathscr{S}_X := \{Y \mid X \to Y\}$ is a monoid under the multiplication $(Y,Z) \mapsto Y \wedge Z$ ('joint variable').

Each arrow $X \to Y$ induces an inclusion $\mathscr{S}_Y \to \mathscr{S}_X$, which defines a particular presheaf (contravariant functor). Let $\mathscr{A}$ denote the presheaf of algebras $X \mapsto \mathbb{R}[\mathscr{S}_X]$ (finite linear combinations).

The category $\mathbf{Mod}(\mathscr{A})$ of $\mathscr{A}$-modules is an abelian category (i.e. it behaves like the category of modules over a ring). We can introduce the derived functors of $\mathrm{Hom}_{\mathscr{A}}(\mathbb{R}, -)$, denoted $\mathrm{Ext}^\bullet(\mathbb{R}, -)$.

Definition

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

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

Resolution

There is a bar resolution $B_\bullet \to \mathbb{R}$ that allows us to compute $H^\bullet(\mathbf{S}, M)$ as the homology of a differential complex $(\mathrm{Nat}(B_\bullet, M), \delta)$.

Resolution

There is a bar resolution $B_\bullet \to \mathbb{R}$ that allows us to compute $H^\bullet(\mathbf{S}, M)$ as the homology of a differential complex $(\mathrm{Nat}(B_\bullet, M), \delta)$.
We get cochains, cocycles and coboundaries.

Important to retain: n-cochains $\varphi \in \mathrm{Nat}(B_n, M)$ are characterized by coherent/functorial collections of elements in M indexed by n-tuples of elements in $\mathrm{Ob}\,\mathbf{S}$ (since each $B_n(X)$ is a free module); they are n-cocyles when they satisfy $\delta\phi = 0$; an n-coboundary is an n-cochain ϕ that comes from an $(n-1)$-cochain ψ, such that $\varphi = \delta\psi$.

Resolution

There is a bar resolution $B_\bullet \to \mathbb{R}$ that allows us to compute $H^\bullet(\mathbf{S}, M)$ as the homology of a differential complex $(\mathrm{Nat}(B_\bullet, M), \delta)$.
We get cochains, cocycles and coboundaries.

Important to retain: n-cochains $\varphi \in \mathrm{Nat}(B_n, M)$ are characterized by coherent/functorial collections of elements in M indexed by n-tuples of elements in $\mathrm{Ob}\,\mathbf{S}$ (since each $B_n(X)$ is a free module); they are n-cocyles when they satisfy $\delta\phi = 0$; an n-coboundary is an n-cochain ϕ that comes from an $(n-1)$-cochain ψ, such that $\varphi = \delta\psi$.

We will see that the 1-cocycle condition encodes the recurrence properties of entropies and multinomial coefficients.

Outline

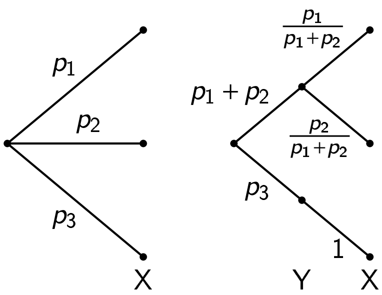
1. Introduction
   - Entropy and recurrence
   - Combinatorics
2. Generalized information theory
3. Information structures and their cohomology
   - Foundations
   - Cohomology of discrete variables
   - Cohomology of continuous (gaussian) variables

Probabilistic case

Discrete variables: *we suppose that each E_X is finite.*

Probabilistic case

Discrete variables: *we suppose that each E_X is finite.*

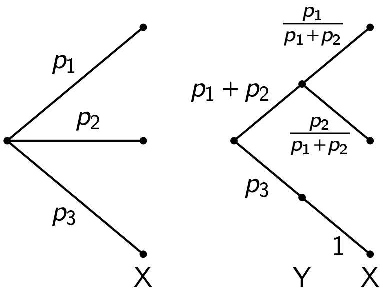


Let $Q:\mathbf{S}\to\mathbf{Sets}$ be a functor that associates to each X a set $Q(X)$ of probabilities on E_X (i.e. functions $p:E_X\to[0,1]$ such that $\sum_{x\in E_X}p(x)=1$) closed under conditioning by variables in $\mathbf{S}$. Every arrow $X\to Y$ in $\mathbf{S}$ translates into a surjection $\pi:E_X\to E_Y$ that induces a *marginalization* $\pi_*:=Q(\pi):Q(X)\to Q(Y)$ given by

$$\pi_*p(y)=\sum_{x\in\pi^{-1}(y)}p(x).$$

Probabilistic case

Discrete variables: *we suppose that each E_X is finite.*



Let $Q : \mathbf{S} \to \mathbf{Sets}$ be a functor that associates to each X a set $Q(X)$ of probabilities on E_X (i.e. functions $p : E_X \to [0,1]$ such that $\sum_{x \in E_X} p(x) = 1$) closed under conditioning by variables in $\mathbf{S}$. Every arrow $X \to Y$ in $\mathbf{S}$ translates into a surjection $\pi : E_X \to E_Y$ that induces a *marginalization* $\pi_* := Q(\pi) : Q(X) \to Q(Y)$ given by

$$\pi_* p(y) = \sum_{x \in \pi^{-1}(y)} p(x).$$

NB: Whenever $\mathbf{S}$ has no initial object, a section $q \in \Gamma(Q) := \mathrm{Hom}_{[\mathbf{S},\mathbf{Sets}]}(*, Q)$ is just a coherent collection of probabilities that does not necessarily come from a global law, called "pseudo-marginal" in the literature.

Probabilistic functionals

Let $F(X)$ be the additive abelian group of measurable real-valued functions on $Q(X)$, and $F(\pi) : F(Y) \to F(X)$ (contravariant) such that $F(\pi)(\phi) = \phi \circ \pi_*$.

In the previous example: $\pi^* f(p_1, p_2, p_3) = f(p_1 + p_2, p_3)$.

Let $F(X)$ be the additive abelian group of measurable real-valued functions on $Q(X)$, and $F(\pi): F(Y) \to F(X)$ (contravariant) such that $F(\pi)(\phi) = \phi \circ \pi_*$.

In the previous example: $\pi^* f(p_1, p_2, p_3) = f(p_1 + p_2, p_3)$.

For each $Y \in \mathscr{S}_X$ and $\phi \in F(X)$, define

$$(Y.\phi)(P) = \sum_{\substack{y \in E_Y \\ P_X(Y=y) \neq 0}} P(Y = y)^\alpha \phi(P_X|_{Y=y}). \tag{11}$$

This turns F into an $\mathscr{A}$-module that we denote F_α.

Probabilistic information cohomology: $H^\bullet(\mathbf{S}, F_\alpha)$

The 1-cocycles are characterized collections of functionals $\{\phi[X] : Q(X) \to \mathbb{R}\}_{X \in \mathrm{Ob}\, S}$ such that

$$0 = X.\phi[Y] - \phi[XY] + \phi[X] \tag{12}$$

Proposition (Baudot-Bennequin, 2015; V. 2017)

The only 1-cocycles are given by multiples of

$$S_\alpha[X] = \begin{cases} -\sum_{x \in E_X} P(x) \log P(x) & \text{when } \alpha = 1 \\ \sum_{x \in E_X} P(x)^\alpha - 1 & \text{when } \alpha \neq 1 \end{cases}.$$

Probabilistic information cohomology: $H^\bullet(\mathbf{S}, F_\alpha)$

The 1-cocycles are characterized collections of functionals $\{\phi[X] : Q(X) \to \mathbb{R}\}_{X \in \mathrm{Ob}\, S}$ such that

$$0 = X.\phi[Y] - \phi[XY] + \phi[X] \quad (12)$$

Proposition (Baudot-Bennequin, 2015; V. 2017)

The only 1-cocycles are given by multiples of

$$S_\alpha[X] = \begin{cases} -\sum_{x\in E_X} P(x)\log P(x) & \textit{when } \alpha = 1 \\ \sum_{x \in E_X} P(x)^\alpha - 1 & \textit{when } \alpha \neq 1 \end{cases}.$$

Globally, the number of free constants depends on the number β_0 of connected components of $\mathbf{S} \setminus \{\mathbf{1}\}$,

$$H^1(\mathbf{S}, F_1) \cong \mathbb{R}^{\beta_0}; \quad H^1(\mathbf{S}, F_\alpha) \cong \mathbb{R}^{\beta_0 - 1} \text{ when } \alpha \neq 1.$$

Information cohomology: Combinatorial case

Let $C : \mathbf{S} \to \mathbf{Sets}$ be a functor that associates to each X the set $C(X)$ of functions $\nu : E_X \to \mathbb{N}$ such that $\|\nu\| := \sum_{x \in E_X} \nu(x) > 0$ (counting functions, histograms...).

Information cohomology: Combinatorial case

Let $C : \mathbf{S} \to \mathbf{Sets}$ be a functor that associates to each X the set $C(X)$ of functions $\nu : E_X \to \mathbb{N}$ such that $\|\nu\| := \sum_{x \in E_X} \nu(x) > 0$ (counting functions, histograms...).
Given $\pi : X \to Y$, the arrow $\pi_* := C(\pi) : C(X) \to C(Y)$ is given by $\pi_* \nu(y) = \sum_{x \in \pi^{-1}(y)} \nu(x)$.

Information cohomology: Combinatorial case

Let $C : \mathbf{S} \to \mathbf{Sets}$ be a functor that associates to each X the set $C(X)$ of functions $\nu : E_X \to \mathbb{N}$ such that $\|\nu\| := \sum_{x \in E_X} \nu(x) > 0$ (counting functions, histograms...).
Given $\pi : X \to Y$, the arrow $\pi_* := C(\pi) : C(X) \to C(Y)$ is given by $\pi_* \nu(y) = \sum_{x \in \pi^{-1}(y)} \nu(x)$.

Let $G(X)$ be the multiplicative abelian group of measurable $(0,\infty)$-valued functions on $C(X)$, and $G(\pi) : G(Y) \to G(X)$ (contravariant) such that $G(\pi)(\phi) = \phi \circ \pi_*$.

Information cohomology: Combinatorial case

Let $C : \mathbf{S} \to \mathbf{Sets}$ be a functor that associates to each X the set $C(X)$ of functions $\nu : E_X \to \mathbb{N}$ such that $\|\nu\| := \sum_{x \in E_X} \nu(x) > 0$ (counting functions, histograms...).
Given $\pi : X \to Y$, the arrow $\pi_* := C(\pi) : C(X) \to C(Y)$ is given by $\pi_* \nu(y) = \sum_{x \in \pi^{-1}(y)} \nu(x)$.

Let $G(X)$ be the multiplicative abelian group of measurable $(0, \infty)$-valued functions on $C(X)$, and $G(\pi) : G(Y) \to G(X)$ (contravariant) such that $G(\pi)(\phi) = \phi \circ \pi_*$.

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

$$(Y.\phi)(\nu) = \prod_{\substack{y \in E_Y \\ \nu(Y=y) \neq 0}} \phi(\nu|_{Y=y_i}). \tag{13}$$

where $\nu|_{Y=y_i}$ is a restriction. This turns G into an $\mathscr{A}$-module.

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

Proposition (V. 2019)

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:*

$$\phi[Y](v) = \frac{[\|v\|]_D!}{\prod_{y \in E_Y} [v(y)]_D!}$$

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)

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:*

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

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.

Asymptotic relation

Example:

- 0-cocycles: the exponential $\exp(k\|v\|)$ 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

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

Recurrence (Shannon entropy)

The combinatorial identity

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

Recurrence (Shannon entropy)

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

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

Recurrence (α-entropy)

Since

$$\begin{bmatrix} n \\ p_1 n, \dots, p_s n \end{bmatrix}_q = q^{n^2 S_2(p_1,\dots,p_s)/2 + o(n^2)},$$

Recurrence (α-entropy)

Since

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

the multiplicative relation

$$\begin{bmatrix} n \\ p_1 n, p_2 n, p_3 n \end{bmatrix}_q = \begin{bmatrix} n \\ (p_1+p_2)n, p_3 n \end{bmatrix}_q \begin{bmatrix} (p_1+p_2)n \\ p_1 n, p_2 n \end{bmatrix}_q,$$

Recurrence (α-entropy)

Since

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

the multiplicative relation

$$\begin{bmatrix} n \\ p_1 n, p_2 n, p_3 n \end{bmatrix}_q = \begin{bmatrix} n \\ (p_1 + p_2)n, p_3 n \end{bmatrix}_q \begin{bmatrix} (p_1 + p_2)n \\ p_1 n, p_2 n \end{bmatrix}_q,$$

implies

$$S_2(p_1, p_2, p_3) = S_2(p_1 + p_2, p_3) + (p_1 + p_2)^2 S_2\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right).$$

Asymptotic relation I

Proposition (V. 2019)

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 $\{v_n\}_{n \geq 1} \subset C_X$ such that

1. $\|v_n\| \to \infty$, *and*
2. *for every $x \in E_X$, $v_n(x)/\|v_n\| \to p(x)$ as $n \to \infty$*

the asymptotic formula

$$\phi[X](v_n) = \exp(\|v_n\|^\alpha \psi[X](p) + o(\|v_n\|^\alpha))$$

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

Outline

Let E be a vector space, and $\mathbf{S}$ a category of subspaces of E, with arrows corresponding to inclusions. We suppose it is conditionally closed under intersections: if Z, V, W are objects of $\mathbf{S}$ such that $Z \subset V$ and $Z \subset W$, then $V \cap W \in \mathrm{Ob}\,\mathbf{S}$.

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

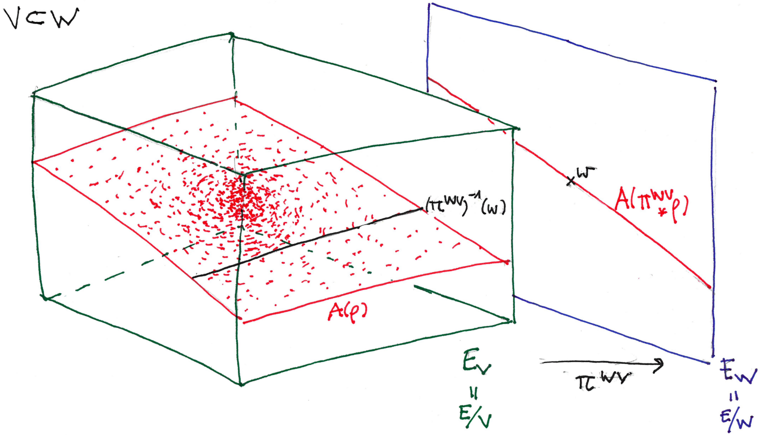
We introduce a functor $\mathscr{N}$ of supports, such that $\mathscr{N}_V$ contains affine subspaces of E and $\mathscr{N}(\pi^{WV})$ is the projection of subspaces under π^{WV}. We suppose $\mathscr{N}$ to be closed under certain operations.

Continuous variables

We then introduce a sheaf $\mathscr{P}$ of gaussian probability laws: $\rho \in \mathscr{P}_V$ if it is supported on $A \in \mathscr{N}_V$, and it is absolutely continuous and with Gaussian density with respect to a given Lebesgue measure on A. The mean and covariance of such ρ can be characterized without fixing coordinates or a Lebesgue measure.

As before, we introduce a sheaf $\mathscr{F}$ of functionals of probability laws with Shannon's action: for $\varphi \in \mathscr{F}_V$,

$$(W.\varphi)(\rho) := \int_{\pi^{WV}(A(\rho))} \varphi(\rho|_{X_W = w})\, \mathrm{d}\pi_*^{WV}\rho(w), \tag{14}$$


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

Dimension is a cocycle

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

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

Entropy is a cocycle

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.

We introduce a sheaf $\mathscr{X}^a$ that encodes variations of the Lebesgue measure. A section correspond to a collection of functions $\{\phi_V\}_{V\in \mathrm{Ob}\,\mathbf{S}}$, such that ϕ_V depends on a probability law ρ on E_V and a reference measure λ on its support (with $\rho \ll \lambda$), and

$$\forall C > 0, \quad \phi_V(\rho, C\lambda) = \phi_V(\rho, \lambda) + a \ln C. \tag{15}$$

The entropy S defines a section of $\mathscr{X}^{-1}$.

Theorem (V., 2019)

For every $a \in \mathbb{R}$, the cohomology $H^1(\mathbf{S}, \mathscr{X}^a)$ over a sufficiently rich grassmannian information structure is the affine space of dimension one made by the functions

$$\Phi_V(\rho) = -aS(\rho) + c.\dim(A(\rho)), \tag{16}$$

where c can be any real constant.

For gaussian probabilities, the fact that differential entropy is a 1-cocycle is equivalent to Schur's determinantal formula

$$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(A)\det(D - BA^{-1}C); \tag{17}$$

Extensions and open problems

There is a quantum version of information cohomology, where Von Neumann entropy appears as a 0-cochain, whose coboundary is related to Shannon entropy. What is the role of other quantum entropies? Relations with entanglement?

Functorial relation between classical and quantum information e.g. through geometric quantization.

The same formalism gives other derived functors. For example, the derived functors of the global sections functor $\Gamma_{\mathbf{S}}(-)$. What is the link between information cohomology and this cohomology of contextuality?

Extensions and open problems

There is a quantum version of information cohomology, where Von Neumann entropy appears as a 0-cochain, whose coboundary is related to Shannon entropy. What is the role of other quantum entropies? Relations with entanglement?

Functorial relation between classical and quantum information e.g. through geometric quantization.

The same formalism gives other derived functors. For example, the derived functors of the global sections functor $\Gamma_{\mathbf{S}}(-)$. What is the link between information cohomology and this cohomology of contextuality? Information cohomology in higher degrees? $\mathrm{Ext}(M, N)$? Products in cohomology?

q-deformed information theory for flags?

Are there combinatorial models for other α-entropies?

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

J. P. Vigneaux, *A combinatorial interpretation for tsallis 2-entropy*, arXiv preprint arXiv:1807.05152, (2018).

J. P. Vigneaux, *Information theory with finite vector spaces*, IEEE Transactions on Information Theory, (2019).
To appear.

J. P. Vigneaux, *Topology of Statistical Systems: A Cohomological Approach to Information Theory*, PhD thesis, Université Paris Diderot, 2019.