Information cohomology of classical vector-valued observables

Juan Pablo Vigneaux Ariztía
jpvigneaux@gmail.com

GSI 2021 - Statistics, Information and Topology

Paris, 23 July 2021

Outline

1. Introduction
2. Information cohomology
3. Mixing discrete and continuous variables
4. Main result: Characterization of 1-cocycles

Information cohomology

- Baudot and Bennequin (2015, GSI 2015 Plenary talk) introduced *information cohomology*, an invariant associated to sheaves of modules over a category of statistical observables...
- ...following the general constructions introduced by Grothendieck in the framework of *topos theory.*
  (The cohomology of sheaves over the category of open sets of a given topological space gives topological invariants of that space, etc.)

Information cohomology

- Baudot and Bennequin (2015, GSI 2015 Plenary talk) introduced *information cohomology*, an invariant associated to sheaves of modules over a category of statistical observables...
- ...following the general constructions introduced by Grothendieck in the framework of *topos theory.*
  (The cohomology of sheaves over the category of open sets of a given topological space gives topological invariants of that space, etc.)
- When the sheaf is made of *probabilistic functionals* on a category of *discrete* observables, Shannon's entropy $-\sum p \log p$ defines a non-trivial cohomology class in degree 1. It satisfies the 1-*cocycle condition*: the recurrence property

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

- In this sense, the chain rule is its defining property of the (discrete) Shannon entropy.

Does a similar result hold for the *differential entropy*?

Does a similar result hold for the *differential entropy*?

1. We consider here a category that mixes discrete and continuous observables.
2. The "coefficients" of the cohomology are again (a sheaf of) probabilistic functionals: continuous real-valued functions of probability laws. The cocycle equations are systems of *functional equations*.

Does a similar result hold for the *differential entropy*?

1. We consider here a category that mixes discrete and continuous observables.
2. The "coefficients" of the cohomology are again (a sheaf of) probabilistic functionals: continuous real-valued functions of probability laws. The cocycle equations are systems of *functional equations*.
3. Every 1-cocycle, when evaluated on probability measures absolutely continuous with respect to the Lebesgue measure, is a linear combination of the differential entropy and the dimension of the underlying space.

Outline

1 Introduction

2 Information cohomology

3 Mixing discrete and continuous variables

4 Main result: Characterization of 1-cocycles

Information structures

Let **S** be a partially ordered set (poset); we see it as a category, denoting the order relation by an arrow.

An object of X of **S** (i.e. $X \in \operatorname{Ob}\mathbf{S}$) is interpreted as an *observable*, with possible outcomes $\mathscr{E}(X) = (E_X, \mathfrak{B}_X)$. An arrow $X \to Y$ as Y being coarser than X, and $Y \wedge Z$ as the joint measurement of Y and Z.

Definition (Information structure)

A pair $(\mathbf{S}, \mathscr{E})$ made of a poset **S** and a functor $\mathscr{E} : \mathbf{S} \to \mathbf{Meas}_{surj}$ is an information structure if

1. **S** has a terminal object, denoted $\top$, and $E_\top \cong \{*\}$ (certainty);
2. Products exist conditionally in **S**: whenever $X, Y, Z \in \operatorname{Ob}\mathbf{S}$ are such that $X \to Y$ and $X \to Z$, the categorical product $Y \wedge Z$ exists in **S**; and
3. For every $Y, Z \in \operatorname{Ob}\mathbf{S}$, $\mathscr{E}(Y \wedge Z)$ is mapped *injectively* into $\mathscr{E}(Y) \times \mathscr{E}(Z)$ by $\mathscr{E}(Y \wedge Z \to Y) \times \mathscr{E}(Y \wedge Z \to Z)$.

Probabilities

We associate to each $X \in \operatorname{Ob} \mathbf{S}$ the set $\Pi(X)$ of probability measures on $\mathscr{E}(X)$, and to each arrow $\pi : X \to Y$ the *marginalization map* $\pi_* := \Pi(\pi)$ that maps ρ to the image measure $\mathscr{E}(\pi)_*(\rho)$.

Probabilities

We associate to each $X \in \operatorname{Ob} \mathbf{S}$ the set $\Pi(X)$ of probability measures on $\mathscr{E}(X)$, and to each arrow $\pi : X \to Y$ the *marginalization map* $\pi_* := \Pi(\pi)$ that maps ρ to the image measure $\mathscr{E}(\pi)_*(\rho)$.

More generally, we consider any subfunctor $\mathbb{Q}$ of Π that is stable under conditioning:

for all $X \in \operatorname{Ob} \mathbf{S}$, $\rho \in \mathbb{Q}(X)$, and $\pi : X \to Y$, $\rho|_{Y=y}$ belongs to $\mathbb{Q}(X)$ for $\pi_*\rho$-almost every $y \in E_Y$, where $\{\rho|_{Y=y}\}_{y \in E_Y}$ is the $(\mathscr{E}\pi, \pi_*\rho)$-disintegration of ρ.

We associate to each $X \in \operatorname{Ob}\mathbf{S}$ the set $\Pi(X)$ of probability measures on $\mathscr{E}(X)$, and to each arrow $\pi : X \to Y$ the *marginalization map* $\pi_* := \Pi(\pi)$ that maps ρ to the image measure $\mathscr{E}(\pi)_*(\rho)$.

More generally, we consider any subfunctor $\mathbb{Q}$ of Π that is stable under conditioning:

for all $X \in \operatorname{Ob}\mathbf{S}$, $\rho \in \mathbb{Q}(X)$, and $\pi : X \to Y$, $\rho|_{Y=y}$ belongs to $\mathbb{Q}(X)$ for $\pi_*\rho$-almost every $y \in E_Y$, where $\{\rho|_{Y=y}\}_{y\in E_Y}$ is the $(\mathscr{E}\pi, \pi_*\rho)$-disintegration of ρ.

Remark

Conditional probabilities are understood as **disintegrations**.
Let $(E, \mathfrak{B}, \nu)$ and $(E_T, \mathfrak{B}_T, \xi)$ be σ-finite measure spaces, $T : E \to E_T$ measurable. The measure ν has a *(T, ξ)-disintegration* $\{\nu_t\}_{t\in E_T}$ if each ν_t is a σ-finite measure on $\mathfrak{B}$ concentrated on $\{T = t\}$ and for each measurable function $f : E \to \mathbb{R}$,

$$\int_E f \,\mathrm{d}\nu = \int_{E_T} \left(\int_E f(x) \,\mathrm{d}\nu_t(x) \right) \mathrm{d}\xi(t).$$

Modules

1. We associate to $X \in \operatorname{Ob}\mathbf{S}$ the monoid $\mathcal{S}_X = \{\, Y \mid X \to Y \,\}$, equipped with the product $(Y, Z) \mapsto Y \wedge Z$ (joint variable). Set $\mathcal{A}_X = \mathbb{R}[\mathcal{S}_X]$ (induced algebra: finite linear combinations...).

Modules

1. We associate to $X \in \mathrm{Ob}\,\mathbf{S}$ the monoid $\mathcal{S}_X = \{\, Y \mid X \to Y \,\}$, equipped with the product $(Y, Z) \mapsto Y \wedge Z$ (joint variable). Set $\mathcal{A}_X = \mathbb{R}[\mathcal{S}_X]$ (induced algebra: finite linear combinations...).
2. $X \mapsto \mathcal{S}_X$ and $X \mapsto \mathcal{A}_X$ define contravariant functors (presheaves).

Modules

1. We associate to $X \in \operatorname{Ob} \mathbf{S}$ the monoid $\mathscr{S}_X = \{ Y \mid X \to Y \}$, equipped with the product $(Y, Z) \mapsto Y \wedge Z$ (joint variable). Set $\mathscr{A}_X = \mathbb{R}[\mathscr{S}_X]$ (induced algebra: finite linear combinations...).
2. $X \mapsto \mathscr{S}_X$ and $X \mapsto \mathscr{A}_X$ define contravariant functors (presheaves).
3. An $\mathscr{A}$-module is a collection of modules $\mathscr{M}_X$ over $\mathscr{A}_X$, for $X \in \operatorname{Ob} \mathbf{S}$, with an action that is "natural" in X.

Modules

1. We associate to $X \in \operatorname{Ob} \mathbf{S}$ the monoid $\mathscr{S}_X = \{\, Y \mid X \to Y \,\}$, equipped with the product $(Y, Z) \mapsto Y \wedge Z$ (joint variable). Set $\mathscr{A}_X = \mathbb{R}[\mathscr{S}_X]$ (induced algebra: finite linear combinations...).
2. $X \mapsto \mathscr{S}_X$ and $X \mapsto \mathscr{A}_X$ define contravariant functors (presheaves).
3. An $\mathscr{A}$-module is a collection of modules $\mathscr{M}_X$ over $\mathscr{A}_X$, for $X \in \operatorname{Ob} \mathbf{S}$, with an action that is "natural" in X.
4. Main example: given $\mathbb{Q}$ probability functor, introduce $\mathscr{F} = \mathscr{F}(\mathbb{Q})$ such that $\mathscr{F}(X)$ is the vector space of measurable functions on $\mathbb{Q}(X)$, and $\mathscr{F}(\pi)$ is precomposition with $\mathbb{Q}(\pi)$ for each morphism π in $\mathbf{S}$. The monoid $\mathscr{S}_X$ acts on $\phi \in \mathscr{F}(X)$ by the rule

$$\forall Y \in \mathscr{S}_X, \forall \rho \in \mathbb{Q}(X), \quad Y.\phi(\rho) = \int_{E_Y} \phi(\rho|_{Y=y}) \, \mathrm{d}\pi_*^{YX} \rho(y) \qquad (1)$$

where π_*^{YX} is the marginalization induced by $\pi^{YX} : X \to Y$. This action can be extended by linearity to $\mathscr{A}_X$.

Information cohomology

- $H^n(\mathbf{S}, \mathscr{F}) = \{n\text{-cocycles}\}/\{n\text{-coboundaries}\}$.
  The $\{n\text{-cocycles}\}$ and $\{n\text{-coboundaries}\}$ are vector subspaces of $\{n\text{-cochains}\}$.

Information cohomology

- $H^n(\mathbf{S}, \mathscr{F}) = \{n\text{-cocycles}\}/\{n\text{-coboundaries}\}$.
  The $\{n\text{-cocycles}\}$ and $\{n\text{-coboundaries}\}$ are vector subspaces of $\{n\text{-cochains}\}$.
- For $n = 1$: $\{1\text{-coboundaries}\} = \{0\}$, so we simply have to describe the 1-cocycles.

Information cohomology

- $H^n(\mathbf{S}, \mathcal{F}) = \{n\text{-cocycles}\}/\{n\text{-coboundaries}\}$.
  The $\{n\text{-cocycles}\}$ and $\{n\text{-coboundaries}\}$ are vector subspaces of $\{n\text{-cochains}\}$.
- For $n = 1$: $\{1\text{-coboundaries}\} = \{0\}$, so we simply have to describe the 1-cocycles.
  Let $\mathcal{B}_1(X)$ be the $\mathcal{A}_X$-module freely generated by a collection of bracketed symbols $\{[Y]\}_{Y \in \mathcal{S}_X}$; an arrow $\pi : X \to Y$ induces an inclusion $\mathcal{B}_1(Y) \hookrightarrow \mathcal{B}_1(X)$, so $\mathcal{B}_1$ is a presheaf. A 1-*cochain* is a natural transformations $\varphi : \mathcal{B}_1 \Rightarrow \mathcal{F}$, with components $\varphi_X : \mathcal{B}_1(X) \to \mathcal{F}(X)$; we use $\varphi_X[Y]$ as a shorthand for $\varphi_X([Y])$.

Information cohomology

- $H^n(\mathbf{S}, \mathscr{F}) = \{n\text{-cocycles}\}/\{n\text{-coboundaries}\}$.
  The $\{n\text{-cocycles}\}$ and $\{n\text{-coboundaries}\}$ are vector subspaces of $\{n\text{-cochains}\}$.
- For $n = 1$: $\{1\text{-coboundaries}\} = \{0\}$, so we simply have to describe the 1-cocycles.
  Let $\mathscr{B}_1(X)$ be the $\mathscr{A}_X$-module freely generated by a collection of bracketed symbols $\{[Y]\}_{Y \in \mathcal{S}_X}$; an arrow $\pi : X \to Y$ induces an inclusion $\mathscr{B}_1(Y) \hookrightarrow \mathscr{B}_1(X)$, so $\mathscr{B}_1$ is a presheaf. A 1-*cochain* is a natural transformations $\varphi : \mathscr{B}_1 \Rightarrow \mathscr{F}$, with components $\varphi_X : \mathscr{B}_1(X) \to \mathscr{F}(X)$; we use $\varphi_X[Y]$ as a shorthand for $\varphi_X([Y])$.
- The naturality implies that $\varphi_X[Z](\rho)$ equals $\varphi_Z[Z](\pi_*^{ZX}\rho)$ (*locality*).

Information cohomology

- $H^n(\mathbf{S}, \mathscr{F}) = \{n\text{-cocycles}\}/\{n\text{-coboundaries}\}$.
  The $\{n\text{-cocycles}\}$ and $\{n\text{-coboundaries}\}$ are vector subspaces of $\{n\text{-cochains}\}$.
- For $n = 1$: $\{1\text{-coboundaries}\} = \{0\}$, so we simply have to describe the 1-cocycles.
  Let $\mathscr{B}_1(X)$ be the $\mathscr{A}_X$-module freely generated by a collection of bracketed symbols $\{[Y]\}_{Y\in\mathcal{S}_X}$; an arrow $\pi : X \to Y$ induces an inclusion $\mathscr{B}_1(Y) \hookrightarrow \mathscr{B}_1(X)$, so $\mathscr{B}_1$ is a presheaf. A 1-*cochain* is a natural transformations $\varphi : \mathscr{B}_1 \Rightarrow \mathscr{F}$, with components $\varphi_X : \mathscr{B}_1(X) \to \mathscr{F}(X)$; we use $\varphi_X[Y]$ as a shorthand for $\varphi_X([Y])$.
- The naturality implies that $\varphi_X[Z](\rho)$ equals $\varphi_Z[Z](\pi_*^{ZX}\rho)$ (*locality*).
- A 1-cochain φ is a 1-cocycle (i.e. $\delta\varphi = 0$) iff

$$\forall X \in \operatorname{Ob}\mathbf{S}, \forall X_1, X_2 \in \mathcal{S}_X, \quad 0 = X_1.\varphi_X[X_2] - \varphi_X[X_1 \wedge X_2] + \varphi_X[X_1].$$

  Remark that this is an equality *of functions* in $\mathbb{Q}(X)$.

Known results: discrete case

An information structure is *finite* if for all $X \in \operatorname{Ob} S$, E_X is finite.

Theorem (Baudot & Bennequin 2015, V. 2020 (TAC))

Suppose $(\mathbf{S}, \mathscr{E})$ is finite. If X can be written as a non-degenerate product $Y \wedge Z$ (this means that E_X is sufficiently "close" to $E_Y \times E_Z$), there exists $b \in \mathbb{R}$ such that for all $W \in \mathcal{S}_X$ and ρ in $\mathbb{Q}(W)$

$$\varphi_W[W](\rho) = -b \sum_{w \in E_W} \rho(w) \log \rho(w). \tag{2}$$

Known results: continuous vector-valued case

- Let E be a vector with Euclidean metric, and $\mathbf{S}$ a poset of vector subspaces of E, ordered by inclusion, with E terminal and conditional products (intersections).

Known results: continuous vector-valued case

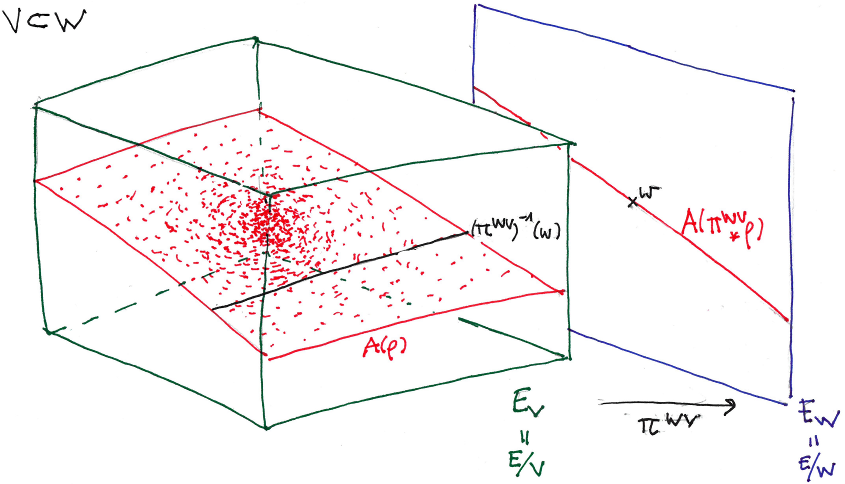
- Let E be a vector with Euclidean metric, and $\mathbf{S}$ a poset of vector subspaces of E, ordered by inclusion, with E terminal and conditional products (intersections).
- Introduce sheaf $\mathscr{E}$ by $V \in \operatorname{Ob}\mathbf{S} \mapsto E_V := V^\perp \cong E/V$ (using the metric).

Known results: continuous vector-valued case

- Let E be a vector with Euclidean metric, and $\mathbf{S}$ a poset of vector subspaces of E, ordered by inclusion, with E terminal and conditional products (intersections).
- Introduce sheaf $\mathscr{E}$ by $V \in \mathrm{Ob}\,\mathbf{S} \mapsto E_V := V^\perp \cong E/V$ (using the metric).
- Introduce sheaf $\mathscr{N}$ of *affine supports*: $\mathscr{N}(X)$ a set of affine subspaces of E_X + suitable hypotheses.
  Each $N \in \mathscr{N}(X)$ has a Lebesgue measure $\mu_{X,N}$ induced by the metric.

Known results: continuous vector-valued case

- Let E be a vector with Euclidean metric, and $\mathbf{S}$ a poset of vector subspaces of E, ordered by inclusion, with E terminal and conditional products (intersections).
- Introduce sheaf $\mathscr{E}$ by $V \in \mathrm{Ob}\,\mathbf{S} \mapsto E_V := V^\perp \cong E/V$ (using the metric).
- Introduce sheaf $\mathscr{N}$ of *affine supports*: $\mathscr{N}(X)$ a set of affine subspaces of E_X + suitable hypotheses.
  Each $N \in \mathscr{N}(X)$ has a Lebesgue measure $\mu_{X,N}$ induced by the metric.
- Introduce $\mathbb{Q}_{gauss}$ such that $\mathbb{Q}_{gauss}(X)$ are probabilities measures ρ with gaussian density w.r.t. $\mu_{X,N}$, for some $N \in \mathscr{N}(X)$.
  Consider then a subfunctor $\mathscr{F}'$ of $\mathscr{F}(\mathbb{Q}_{gauss})$ made of functions that *grow moderately* (i.e. at most polynomially) with respect to the mean, in such a way that the integral defining $Y.\phi(\rho)$ is always convergent.



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

Known results: continuous vector-valued case

A triple $(\mathbf{S}, \mathscr{E}, \mathscr{N})$ is *sufficiently rich* when there are "enough supports" (not all parallel to the spaces generated by a given basis).

Theorem (V. 2019 (Ph.D. thesis))

Provided $(\mathbf{S}, \mathscr{E}, \mathscr{N})$ is sufficiently rich, for every 1-cocycle φ, with coefficients in $\mathscr{F}'(\mathbb{Q})$, there are real constants a and c such that, for every $X \in \operatorname{Ob}\mathbf{S}$ and every gaussian law ρ with support E_X and variance Σ_ρ,

$$\varphi_X[X](\rho) = a\det(\Sigma_\rho) + c.\dim(E_X). \tag{3}$$

Moreover, φ is completely determined by its behavior on nondegenerate laws.

The variance is a nondegenerate, symmetric, positive bilinear form on E_X^*

Outline

1 Introduction

2 Information cohomology

3 Mixing discrete and continuous variables

4 Main result: Characterization of 1-cocycles

Product structure

- Let $(\mathbf{S}_d, \mathscr{E}_d)$ be a finite information structure and $(\mathbf{S}_c, \mathscr{E}_c, \mathscr{N}_c)$ associated to an Euclidean space E, such that the characterizations of 1-cocycle hold.

Product structure

- Let $(\mathbf{S}_d, \mathscr{E}_d)$ be a finite information structure and $(\mathbf{S}_c, \mathscr{E}_c, \mathscr{N}_c)$ associated to an Euclidean space E, such that the characterizations of 1-cocycle hold.
- Let $(\mathbf{S}, \mathscr{E} : \mathbf{S} \to \mathbf{Meas})$ be the product $(\mathbf{S}_c, \mathscr{E}_c) \times (\mathbf{S}_d, \mathscr{E}_d)$ in the category of information structures.
  By definition, every object $X \in \operatorname{Ob} \mathbf{S}$ has the form $\langle X_c, X_d \rangle$ for $X_c \in \operatorname{Ob} \mathbf{S}_c$ and $X_d \in \operatorname{Ob} \mathbf{S}_d$, and $\pi : \langle X_1, X_2 \rangle \to \langle Y_1, Y_2 \rangle$ in $\mathbf{S}$ if and only if there exist arrows $\pi_1 : X_1 \to Y_1$ in $\mathbf{S}_c$ and $\pi_2 : X_2 \to Y_2$ in $\mathbf{S}_d$.
  Outcomes: $\mathscr{E}(X) = \mathscr{E}_c(X_1) \times \mathscr{E}_d(X_2)$, etc.

Product structure

- Let $(\mathbf{S}_d, \mathscr{E}_d)$ be a finite information structure and $(\mathbf{S}_c, \mathscr{E}_c, \mathscr{N}_c)$ associated to an Euclidean space E, such that the characterizations of 1-cocycle hold.
- Let $(\mathbf{S}, \mathscr{E} : \mathbf{S} \to \mathbf{Meas})$ be the product $(\mathbf{S}_c, \mathscr{E}_c) \times (\mathbf{S}_d, \mathscr{E}_d)$ in the category of information structures.
  By definition, every object $X \in \operatorname{Ob} \mathbf{S}$ has the form $\langle X_c, X_d \rangle$ for $X_c \in \operatorname{Ob} \mathbf{S}_c$ and $X_d \in \operatorname{Ob} \mathbf{S}_d$, and $\pi : \langle X_1, X_2 \rangle \to \langle Y_1, Y_2 \rangle$ in $\mathbf{S}$ if and only if there exist arrows $\pi_1 : X_1 \to Y_1$ in $\mathbf{S}_c$ and $\pi_2 : X_2 \to Y_2$ in $\mathbf{S}_d$.
  Outcomes: $\mathscr{E}(X) = \mathscr{E}_c(X_1) \times \mathscr{E}_d(X_2)$, etc.
- There is an embedding in the sense of information structures $\mathbf{S}_c \hookrightarrow \mathbf{S}$, $X \to \langle X, \top \rangle$; we write X instead of $\langle X, \top \rangle$. Similarly for discrete part.

Supports

We extend the sheaf of supports $\mathcal{N}_c$ to the whole $\mathbf{S}$ setting $\mathcal{N}_d(Y) = 2^Y \setminus \{\emptyset\}$ when $Y \in \operatorname{Ob}\mathbf{S}_d$, and then $\mathcal{N}(Z) = \{\, A \times B \mid A \in N_c(X), B \in N_d(Y) \,\}$ for any $Z = \langle X, Y \rangle \in \operatorname{Ob}\mathbf{S}$.

For every $X \in \operatorname{Ob}\mathbf{S}$ and $N \in \mathcal{N}(X)$, there is a unique reference measure $\mu_{X,N}$ compatible with M: Lebesgue measure given by the metric on the affine subspaces of E, or the counting measure, or a mixture of both: for $A \times B \subset E_X \times E_Y$, with $X \in \operatorname{Ob}\mathbf{S}_c$ and $Y \in \operatorname{Ob}\mathbf{S}_d$, it is just $\sum_{y \in B} \mu_A^y$, where μ_A^y is the image of μ_A under the inclusion $A \hookrightarrow A \times B,\ a \mapsto (a, y)$. We write μ_X instead of μ_{X,E_X}.

Probability laws

Consider the subfunctor $\Pi(\mathcal{N})$ of Π that associates to each $X \in \operatorname{Ob} \mathbf{S}$ the set $\Pi(X;\mathcal{N})$ of probability measures on $\mathscr{E}(X)$ that are absolutely continuous with respect to the reference measure $\mu_{X,N}$ on some $N \in \mathcal{N}(X)$. We define the *(affine) support* or *carrier* of ρ, denoted $A(\rho)$, as the unique $A \in \mathcal{N}(X)$ such that $\rho \ll \mu_{X,A}$.

$\mathbb{Q} \subset \Pi(\mathcal{N})$ subfunctor such that:

1. $\mathbb{Q}$ is adapted (closed under conditioning);
2. for each $\rho \in \mathbb{Q}(X)$, the differential entropy $S_{\mu_{A(\rho)}}(\rho) := -\int \log \frac{\mathrm{d}\rho}{\mathrm{d}\mu_{A(\rho)}} \,\mathrm{d}\rho$ exists i.e. it is finite;
3. when restricted to probabilities in $\mathbb{Q}(X)$ with the same carrier A, the differential entropy is a continuous functional in the total variation norm;
4. for each $X \in \operatorname{Ob} \mathbf{S}_c$ and each $N \in \mathcal{N}(X)$, (enough) *gaussian mixtures* carried by N are contained in $\mathbb{Q}(X)$—cf. next section.

Information cohomology with coefficients in the module of probabilistic functionals with Shannon's action

Let $\mathbb{Q}$ be a probability functor satisfying conditions above.

For each $X \in \operatorname{Ob}\mathbf{S}$, let $\mathscr{F}(X)$ be the vector space of measurable functions of (ρ, μ_M), equivalently $(\mathrm{d}\rho/d\mu_{A(\rho)}, \mu_M)$, where ρ is an element of $\mathbb{Q}(X)$, μ_M is a global determination of reference measure on any affine subspace given by the metric M on E, and $\mu_{A(\rho)}$ is the corresponding reference measure on the carrier of ρ under this determination.

Let $\mathscr{G}$ be linear subfunctor of $\mathscr{F}$ such that the quantities

$$Y.\varphi(\rho) = \int_{E_Y} \varphi(\rho|_{Y=y})\,\mathrm{d}Y_*\rho$$

are always convergent. We want to compute $H^1(\mathbf{S}, \mathscr{G})$.

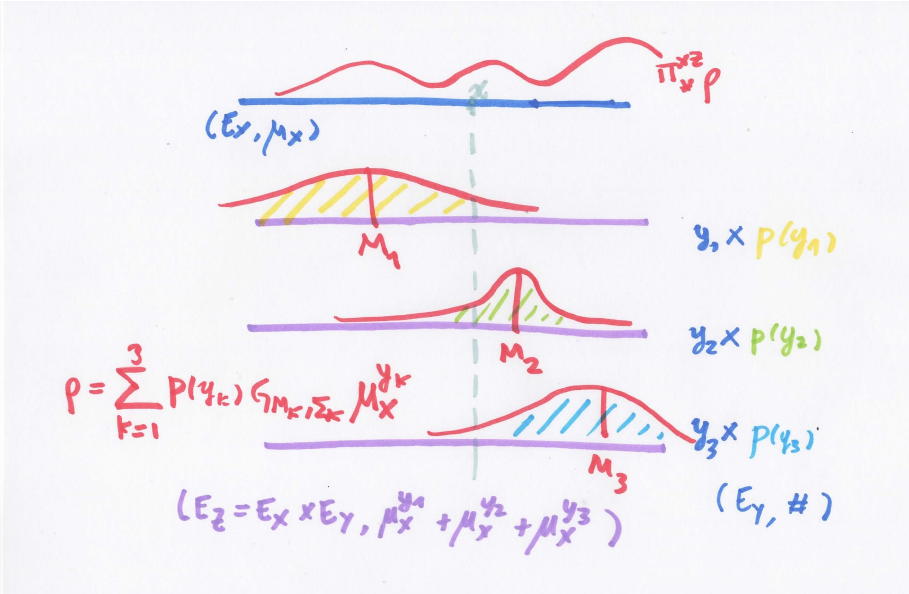
Outline

1 Introduction

2 Information cohomology

3 Mixing discrete and continuous variables

4 Main result: Characterization of 1-cocycles


$\pi^{xz}_{*}\rho$
(E_X, μ_X)
x
M_1
$y_1 \times p(y_1)$
M_2
$y_2 \times p(y_2)$
$\rho = \sum_{k=1}^{3} p(y_k)\, G_{M_k, \Sigma_k}\, \mu_X^{y_k}$
M_3
$y_3 \times p(y_3)$
$(E_Y, \#)$
$(E_Z = E_X \times E_Y, \mu_X^{y_1} + \mu_X^{y_2} + \mu_X^{y_3})$

Formula for gaussian mixtures

For $Z = \langle X, Y\rangle = \langle X, \top\rangle \wedge \langle \top, Y\rangle =: X \wedge Y$ and $\rho = \sum_{y\in E_Y} p(y) G_{M_y,\Sigma_y}\mu_X^y \in \mathbb{Q}(Z)$ as in the previous slide:

$$\varphi_Z[Z](\rho) = \varphi_X[X](\underbrace{\pi_*^{XZ}\rho}_{\text{gaussian mix.}}) + \int_{E_X} \varphi_Z[Y](\underbrace{\rho|_{X=x}}_{\text{discrete law}})\,\mathrm{d}\pi_*^{XZ}\rho$$

$$= \varphi_Y[Y](\underbrace{\pi_*^{YZ}\rho}_{\text{discrete law}}) + \sum_{y\in E_Y} \pi_*^{YZ}\rho(y)\varphi_Z[X](\underbrace{\rho|_{Y=y}}_{gaussian}).$$

An explicit computation yields that, for some real constants a, b, and c:

$$\varphi_X[X](\pi_*^{XZ}\rho) = \sum_{y\in E_Y} p(y)\left((a-\frac{b}{2})\log\det\Sigma_y + c\dim E_X - \frac{b\dim E_X}{2}\log(2\pi e)\right) + b\underbrace{S_{\mu_X}(\pi_*^{XZ}\rho)}_{\text{Diff. ent.}}. \tag{4}$$

Theorem (Characterization of 1-cocycles)

Let φ *be a* 1*-cocycle on* $\mathbf{S}$ *with coefficients in* $\mathcal{G}$*,* X *an object in* $\mathbf{S}_c$*, and* ρ *a probability law in* $\mathbb{Q}(X)$ *absolutely continuous with respect to* μ_X*. Then, there exist real constants* c_1, c_2 *such that*

$$\varphi_X[X](\rho) = c_1 S_{\mu_A}(\rho) + c_2 \dim E_X. \tag{5}$$

Theorem (Characterization of 1-cocycles)

Let φ be a 1-cocycle on $\mathbf{S}$ with coefficients in $\mathscr{G}$, X an object in $\mathbf{S}_c$, and ρ a probability law in $\mathbb{Q}(X)$ absolutely continuous with respect to μ_X. Then, there exist real constants c_1, c_2 such that

$$\varphi_X[X](\rho) = c_1 S_{\mu_A}(\rho) + c_2 \dim E_X. \tag{5}$$

Proof.

Every density can be approximated by a random mixture of gaussians in $L^1(E_X, \mu_X)$. Let f be any density of ρ with respect to μ_X, $(X_n)_{n\in\mathbb{N}}$ an i.i.d sequence of points of E_X with law ρ, and (h_n) any sequence such that $h_n \to 0$ and $nh_n^d \to \infty$. Each

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n} G_{X_i(\omega), h_n^2 I}(x)$$

is the density of a composite gaussian law ρ_n (kernel estimate); the $(X_n(\omega))_n$ is any realization of the process such that f_n tend to f in L^1. So $\rho_n \to \rho$ in total variation (cf. Scheffé's lemma)

In virtue of the hypotheses on $\mathbb{Q}$, $S_{\mu_A}(\rho)$ is finite and $S_{\mu_A}(\rho_n) \to S_{\mu_A}(\rho)$. Since $\varphi_X[X]$ is continuous when restricted to $\Pi(A, \mu_A)$ and $\varphi_X[X](f)$ is a real number, we conclude that necessarily $a = b/2$. The statement is then just a rewriting of (4). □

Conclusion

We get a characterization of the dimension and the differential entropy as information measures that depends solely on their chain rules.

This should be contrasted with the involved characterizations introduced e.g. by Ikeda (1959). The improvement is explained by the *naturality* encoded in the categorical constructions.

Moreover, the cocycle equations that express the chain rule come from a general algebro-geometric construction and suggest further connections with geometry/topology.