

Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Topological forms of information.

by

Pierre Baudot and Daniel Bennequin

Preprint no.: 112

2014



Topological forms of information

Pierre Baudot* and Daniel Bennequin†

**Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, 04103 Leipzig, Germany*

†*Universite Paris Diderot-Paris 7, UFR de Mathematiques, Equipe Geometrie et Dynamique, Batiment Sophie Germain, 5 rue Thomas Mann, 75205 Paris Cedex 13, France*

Abstract. We propose that entropy is a universal co-homological class in a theory associated to a family of observable quantities and a family of probability distributions. Three cases are presented: 1) classical probabilities and random variables; 2) quantum probabilities and observable operators; 3) dynamic probabilities and observation trees. This gives rise to a new kind of topology for information processes. We discuss briefly its application to complex data, in particular to the structures of information flows in biological systems. This short note summarizes results obtained during the last years by the authors. The proofs are not included, but the definitions and theorems are stated with precision.

Keywords: Shannon Information, Homology Theory, Entropy, Quantum Information, Homotopy of Links, Mutual Informations, Trees, Monads, Partitions

PACS: 02.40.Re, 03.67.-a, 05.20.-y

INTRODUCTION

"What is information ?" is a question that has received several answers according to the different problems investigated. The best known definition was given by Shannon [1], using random variables and a probability law, for the problem of optimal message compression. But the first definition was given by Fisher, as a metric associated to a smooth family of probability distributions, for optimal discrimination by statistical tests; it is a limit of the Kullback-Leibler divergence, which was introduced to estimate the accuracy of a statistical model of empirical data, and which can be also viewed as a quantity of information. More generally Kolmogorov considered that the concept of information must precede probability theory (cf. [2]). However, Evariste Galois saw the application of group theory for discriminating solutions of an algebraic equation as a first step toward a general theory of ambiguity, that was developed further by Riemann, Picard, Vessiot, Lie, Poincare and Cartan, for systems of differential equations; it is also a theory of information. In another direction Rene Thom claimed that information must have a topological content (see [3]); he gave the example of the unfolding of the coupling of two dynamical systems, but he had in mind the whole domain of algebraic or differential topology.

All these approaches have in common the definition of secondary objects, either functions, groups or homology cycles, for measuring in what sense a pair of objects departs from independency. For instance, in the case of Shannon, the mutual information is $I(X;Y) = H(X) + H(Y) - H(X,Y)$, where H denotes the usual Gibbs entropy ($H(X) = -\sum_x P(X=x) \ln_2 P(X=x)$), and for Galois it is the quotient set $IGal(L_1;L_2|K) = (Gal(L_1|K) \times Gal(L_2|K))/Gal(L|K)$, where L_1, L_2 are two fields

containing a field K in an algebraic closure Ω of K , where L is the field generated by L_1 and L_2 in Ω , and where $Gal(L_i|K)$ (for $i = 0, 1, 2$) denotes the group introduced by Galois, made by the field automorphisms of L_i fixing the elements of K .

We suggest that all information quantities are of co-homological nature, in a setting which depends on a pair of categories (cf.[4] [5]); one for the data on a system, like random variables or functions of solutions of an equation, and one for the parameters of this system, like probability laws or coefficients of equations; the first category generates an algebraic structure like a monoid, or more generally a monad (cf. [4]), and the second category generates a representation of this structure, as do for instance conditioning, or adding new numbers; then information quantities are co-cycles associated with this module.

We will see that, given a set of random variables on a finite set Ω and a simplicial subset of probabilities on Ω , the *entropy* appears as the only one universal co-homology class of degree one. The higher mutual information functions that were defined by Shannon are co-cycles (or twisted co-cycles for even orders), and they correspond to higher homotopical constructions. In fact this description is equivalent to the theorem of Hu Kuo Ting [6], that gave a set theoretical interpretation of the mutual information decomposition of the total entropy of a system. Then we can use information co-cycles to describe forms of the information distribution between a set of random data; figures like ordinary links, or chains or Borromean links appear in this context, giving rise to a new kind of topology.

INFORMATION HOMOLOGY

Here we call random variables (r.v) on a finite set Ω *congruent* when they define the same partition (remind that a partition of Ω is a family of disjoint non-empty subsets covering Ω and that the partition associated to a r.v X is the family of subsets Ω_x of Ω defined by the equations $X(\omega) = x$); the join r.v YZ , also denoted by (Y, Z) , corresponds to the less fine partition that is finer than Y and Z . This defines a monoid structure on the set $\Pi(\Omega)$ of partitions of Ω , with 1 as a unit, and where each element is idempotent, i.e. $\forall X, XX = X$. An *information category* is a set \mathcal{S} of r.v such that, for any $Y, Z \in \mathcal{S}$ less fine than $U \in \mathcal{S}$, the join YZ belongs to \mathcal{S} , cf. [7]. An ordering on \mathcal{S} is given by $Y \leq Z$ when Z refines Y , which also defines the morphisms $Y \rightarrow Z$ in \mathcal{S} . In what follows we always assume that 1 belongs to \mathcal{S} . The simplex $\Delta(\Omega)$ is defined as the set of families of numbers $\{p_\omega; \omega \in \Omega\}$, such that $\forall \omega, 0 \leq p_\omega \leq 1$ and $\sum_\omega p_\omega = 1$; it parameterizes all probability laws on Ω . We choose a simplicial sub-complex \mathcal{P} in $\Delta(\Omega)$, which is stable by all the conditioning operations by elements of \mathcal{S} . By definition, for $N \in \mathbb{N}$, an information N -cochain is a family of measurable functions of $P \in \mathcal{P}$, with values in \mathbb{R} or \mathbb{C} , indexed by the sequences $(S_1; \dots; S_N)$ in \mathcal{S} majored by an element of \mathcal{S} , whose values depend only of the image law $(S_1, \dots, S_N)_*P$. This condition is natural from a topos point of view, cf. [4]; we interpret it as a "locality" condition. Note that we write $(S_1; \dots; S_N)$ for a sequence, because (S_1, \dots, S_N) designates the joint variable. For $N = 0$ this gives only the constants. We denote by \mathcal{C}^N the vector space of N -cochains of information. The following formula corresponds to the *averaged conditioning* of

Shannon [1]:

$$S_0.F(S_1; \dots; S_N; \mathbb{P}) = \sum \mathbb{P}(S_0 = v_j) F(S_1; \dots; S_N; \mathbb{P} | S_0 = v_j), \quad (1)$$

where the sum is taken over all values of S_0 , and the vertical bar is ordinary conditioning.

It satisfies the associativity condition $(S'_0 S_0).F = S'_0.(S_0.F)$.

The coboundary operator δ is defined by

$$\begin{aligned} \delta F(S_0; \dots; S_N; \mathbb{P}) &= S_0.F(S_1; \dots; S_N; \mathbb{P}) \\ &+ \sum_0^{N-1} (-1)^{i+1} F(\dots; (S_i, S_{i+1}); \dots; S_N; \mathbb{P}) + (-1)^{N+1} F(S_0; \dots; S_{N-1}; \mathbb{P}), \end{aligned} \quad (2)$$

It corresponds to a standard non-homogeneous bar complex (cf. [5]). Another coboundary operator on \mathcal{C}^N is δ_t (t for twisted or trivial action or topological complex), that is defined by the above formula with the first term $S_0.F(S_1; \dots; S_N; \mathbb{P})$ replaced by $F(S_1; \dots; S_N; \mathbb{P})$. The corresponding co-cycles are defined by the equations $\delta F = 0$ or $\delta_t F = 0$, respectively. We easily verify that $\delta \circ \delta = 0$ and $\delta_t \circ \delta_t = 0$; then co-homology $H^*(\mathcal{S}; \mathcal{P})$ resp. $H_t^*(\mathcal{S}; \mathcal{P})$ is defined by taking co-cycles modulo the elements of the image of δ resp. δ_t , called co-boundaries. The fact that classical entropy $H(X; \mathbb{P}) = -\sum_i p_i \log_2 p_i$ is a 1-co-cycle is the fundamental equation $H(X, Y) = H(X) + X.H(Y)$.

Theorem 1 (cf. [7]): For the full simplex $\Delta(\Omega)$, and if \mathcal{S} is the monoid generated by a set of at least two variables, such that each pair takes at least four values, then the information co-homology space of degree one is one-dimensional and generated by the classical entropy.

Problem 1: Compute the homology of higher degrees.

We conjecture that for binary variables it is zero, but that in general non-trivial classes appear, deduced from polylogarithms. This could require us to connect with the works of Dupont, Bloch, Goncharov, Elbaz-Vincent, Gangl et al. on motives (cf. [8]), which started from the discovery of Cathelineau (1988) that entropy appears in the computation of the degree one homology of the discrete group SL_2 over \mathbb{C} with coefficients in the adjoint action (cf. [9]).

Suppose \mathcal{S} is the monoid generated by a finite family of partitions. The higher mutual informations were defined by Hu Kuo Ting [6] as alternating sums:

$$I_N(S_1; \dots; S_N; \mathbb{P}) = \sum_{k=1}^{k=N} (-1)^{k-1} \sum_{I \subset [N]; \text{card}(I)=k} H(S_I; \mathbb{P}), \quad (3)$$

where S_I denotes the join of the S_i such that $i \in I$. We have $I_1 = H$ and $I_2 = I$ is the usual mutual information: $I(S; T) = H(S) + H(T) - H(S, T)$.

Theorem 2 (cf. [7]): $I_{2m} = \delta_t \delta \delta_t \dots \delta \delta_t H$, $I_{2m+1} = -\delta \delta_t \delta \delta_t \dots \delta \delta_t H$, where there are $m-1$ δ and m δ_t factors for I_{2m} and m δ and m δ_t factors for I_{2m+1} .

Thus odd information quantities are information co-cycles, because they are in the image of δ , and even information quantities are twisted (or topological) co-cycles, because they are in the image of δ_t .

In [7] we show that this description is equivalent to the theorem of Hu Kuo Ting (1962)

[6], giving a set theoretical interpretation of the mutual information decomposition of the total entropy of a system: mutual information, join and averaged conditioning correspond respectively to intersection, union and difference $A \setminus B = A \cap B^c$. In special cases we can interpret I_N as homotopical algebraic invariants. For instance for $N = 3$, suppose that $I(X;Y) = I(Y;Z) = I(Z;X) = 0$, then $I_3(X;Y;Z) = -I((X,Y);Z)$ can be defined as a Milnor invariant for links, generalized by Massey, as they are presented in [10] (cf. page 284), through the 3-ary obstruction to associativity of products in a subcomplex of a differential algebra, cf. [7]. The absolute minima of I_3 correspond to Borromean links, interpreted as synergy, cf. [11], [12].

EXTENSION TO QUANTUM INFORMATION

Positive hermitian $n \times n$ -matrices ρ , normalized by $Tr(\rho) = 1$, are called density of states (or density operators) and are considered as quantum probabilities on $E = \mathbb{C}^n$. Real quantum observables are $n \times n$ -matrices hermitian matrices Z , and, by definition, the amplitude, or expectation, of the observable Z in the state ρ is given by the formula $\mathbb{E}(Z) = Tr(Z\rho)$ (see e.g. [13]). Two real observables Y, Z are said *congruent* if their eigenspaces are the same, thus orthogonal decomposition of E are quantum analogs of partitions. The join is well defined for commuting observables. An information structure \mathcal{S}_Q is given by a subset of observables, such that if Y, Z have common refined eigenspaces decomposition in \mathcal{S}_Q , their join (Y, Z) belongs to \mathcal{S}_Q . We assume that $\{E\}$ belongs to \mathcal{S}_Q . We define information N -cochains as for the classical case.

The image of a density ρ by an observable Y is $\rho_Y = \sum_A E_A^* \rho E_A$, where the E_A 's are the spectral projectors of the observable Y .

The action of a variable on the cochains space \mathcal{C}_Q^* is given by the *quantum averaged conditioning*:

$$Y.F(Y_0; \dots; Y_m; \rho) = \sum_A Tr(E_A^* \rho E_A) F(Y_0; \dots; Y_m; E_A^* \rho E_A) \quad (4)$$

From here we define coboundary operators δ_Q and δ_{Q_t} by the formula (2), then notions of co-cycles, co-boundaries and co-homology classes follow. We have $\delta_Q \circ \delta_Q = 0$ and $\delta_{Q_t} \circ \delta_{Q_t} = 0$; cf. [7].

The *Von-Neumann entropy* of ρ is $S(\rho) = \mathbb{E}_\rho(-\log_2(\rho)) = -Tr(\rho \log_2(\rho))$, the entropy of Y in state ρ is $S(Y; \rho) = S(\rho_Y)$, and the classical entropy is $H(Y; \rho) = -\sum_A Tr(E_A^* \rho E_A) \log_2(Tr(E_A^* \rho E_A))$. It is well known that $S((X, Y); \rho) = H(X; \rho) + X.S(Y; \rho)$ when X, Y commute, cf. [13]. In particular, by taking $Y = 1_E$ we see that classical entropy measures the default of equivariance of the quantum entropy, i.e. $H(X; \rho) = S(X; \rho) - (X.S)(\rho)$.

Then, if we define the *reduced quantum entropy* by $s(X; \rho) = S(X; \rho) - S(\rho)$, we get a 1-cocycle of quantum information. In fact H is also a 1-cocycle and it is co-homologous to s by the following lemma: $\delta_Q(S) = s - H$.

Theorem 3 (cf. [7]): when \mathcal{S}_Q is generated by at least two decompositions such each pair has at least four subspaces, then s or H generates the co-homology of δ_Q .

CONCAVITY AND CONVEXITY PROPERTIES OF INFORMATION QUANTITIES

The simplest classical information structure \mathcal{S} is the monoid generated by a family of "elementary" *binary* variables S_1, \dots, S_n . It is remarkable that in this case, the information functions $I_{N,J} = I_N(S_{j_1}; \dots; S_{j_N})$ over all the subsets $J = \{j_1, \dots, j_N\}$ of $[n] = \{1, \dots, n\}$, different from $[n]$ itself, give algebraically independent functions on the probability simplex $\Delta(\Omega)$ of dimension $2^n - 1$. They form coordinates on the quotient of $\Delta(\Omega)$ by a finite group.

Let \mathcal{L}_d denotes the Lie derivative with respect to $d = (1, \dots, 1)$ in the vector space \mathbb{R}^{2^n} , and Δ the Euclidian Laplace operator on \mathbb{R}^{2^n} , then $\Delta = \Delta - 2^{-n} \mathcal{L}_d \circ \mathcal{L}_d$ is the Laplace operator on the simplex $\Delta(\Omega)$ defined by equating the sum of coordinates to 1.

Theorem 4 (cf. [14]): On the affine simplex $\Delta(\Omega)$ the functions $I_{N,J}$ with N odd (resp. even) satisfies the inequality $\Delta I_N \geq 0$ (resp. $\Delta I_N \leq 0$).

In other terms, for N odd the $I_{N,J}$ are super-harmonic which is a kind of weak concavity and for N even they are sub-harmonic which is a kind of weak convexity. In particular, when N is even (resp. odd) $I_{N,J}$ has no local maximum (resp. minimum) in the interior of $\Delta(\Omega)$.

Problem 2: What can be said of the other critical points of $I_{N,J}$? What can be said of the restriction of one information function on the intersection of levels of other information functions? Information topology depends on the shape of these intersections and on the Morse theory for them.

MONADIC COHOMOLOGY OF INFORMATION

Now we consider the category \mathcal{S}_* of *ordered partitions* of Ω over \mathcal{S} , i.e. pairs (π, ω) where $\pi \in \mathcal{S}$ and ω is a bijection from $\{1, \dots, l(\omega)\}$ with the quotient set Ω/π , where $l(\omega)$ is the length of π , i.e. the number of pieces of Ω given by π . The indices of these pieces are the values of the r.v associated with (π, ω) . A *rooted tree decorated* by \mathcal{S}_* is an oriented finite tree Γ , with a marked initial vertex s_0 , named the *root* of Γ , where each vertex s is equipped with an element F_s of \mathcal{S}_* , such that edges issued from s correspond to the values of F_s .

The notation $\mu(m; n_1, \dots, n_m)$ denotes the operation which associates to an ordered partition (π, ω) of length m and m ordered partitions (π_i, ω_i) of respective lengths n_i , the ordered partition that is obtained by cutting the pieces of π using the π_i and respecting the order. An evident unit element for this operation is π_0 . The symbol μ_m denotes the collection of those operations for m fixed. Be aware that in general the result of $\mu(m; n_1, \dots, n_m)$ is not a partition of length $n_1 + \dots + n_m$, thus the μ_m do not define what is named an *operad*; cf. [10], [15]. However they allow the definition of a filtered version of operad, with unity, associativity and covariance for permutations, cf. [16]. See [15] and [10] for the definition of ordinary graded operads.

But the most important algebraic object which is associated to an operad is a *monad* (cf. [4],[15]), i.e. a functor \mathcal{V} from a category \mathcal{A} to itself, equipped with two natural

transformations $\mu : \mathcal{V} \circ \mathcal{V} \rightarrow \mathcal{V}$ and $\varepsilon : \mathbb{R} \rightarrow \mathcal{V}$, which satisfy to the following axioms:

$$\mu \circ (Id \circ \mu) = \mu \circ (\mu \circ Id), \quad \mu \circ (Id \circ \varepsilon) = Id = \varepsilon \circ (\eta \circ Id) \quad (5)$$

In our situation, we can apply the Schur construction (cf. [15]) to the μ_m to get a monad: take for V the real vector space freely generated by \mathcal{S}_* ; it is graded by the partition length as direct sum of spaces $V(m)$. As for ordinary operads we introduce $\mathcal{V} = \bigoplus_{m \geq 0} V(m) \otimes_{\mathfrak{S}_m} V^{\otimes m}$; Schur composition is defined by $\mathcal{V} \circ \mathcal{V} = \bigoplus_{m \geq 0} V(m) \otimes_{\mathfrak{S}_m} \mathcal{V}^{\otimes m}$. It is easy to verify that the collection $(\mu_m; m \in \mathbb{N})$ defines a linear map $\mu : \mathcal{V} \circ \mathcal{V} \rightarrow \mathcal{V}$, and the trivial partition π_0 defines a linear map $\varepsilon : \mathbb{R} \rightarrow \mathcal{V}$, that satisfied to the axioms of a monad.

Let \mathcal{F} be the vector space of real measurable functions on the set \mathcal{P} of probability laws, considered as a \mathcal{S}_* module of pure degree 1, in such a manner that $\mathcal{F} \circ \mathcal{V}^{\otimes m}$ coincides with $\mathcal{F} \otimes \mathcal{V}^{\otimes m}$. For a r.v S of length m and m decorated trees $(S_1^s; S_2^s; \dots; S_k^s)$; $1 \leq s \leq m$ of level k , we pose

$$F_S(S_1; S_2; \dots; S_k; \mathbb{P}) = \sum_s \mathbb{P}(S = s) F(S_1^s; S_2^s; \dots; S_k^s; \mathbb{P} | (S = s)); \quad (6)$$

this is a function of the decorated tree $(S; S_1; S_2; \dots; S_k)$ of level $k + 1$ that roots in S and S_i^s is placed at the end of the edge $S = v$. This formula extending (1) defines a map $\theta : \mathcal{F} \circ \mathcal{V} \rightarrow \mathcal{F}$, that is an action to the right in the sense of monads, i.e. $\theta \circ (Id \circ \mu) = \theta \circ (\theta \circ Id)$; $\theta \circ (Id \circ \varepsilon) = Id$.

We say that $F(S; S_1; S_2; \dots; S_{k-1}; \mathbb{P})$ is *local* if its value depends only of the images of \mathbb{P} by the join of the decorating variables of the corresponding tree. Then we copy the formalism of Beck (see [15]) with this locality condition, to get monadic information co-homology: a cochain of degree k is an element of $\mathcal{F} \circ \mathcal{V}^{\circ k}$ whose components are local; the operator δ comes from the simplicial structure associated to θ and μ :

$$\begin{aligned} \delta F(S; S_1; \dots; S_k; \mathbb{P}) &= F_S(S_1; \dots; S_k; \mathbb{P}) + (-1)^{k+1} F(S; \dots; S_{k-1}; \mathbb{P}) \\ &+ \sum_{i=1}^{i=k} (-1)^i F(S; \dots; \mu(S_{i-1} \circ S_i); S_{i+1}; \dots; S_k; \mathbb{P}) \quad (7) \end{aligned}$$

This gives co-homology groups $H_\tau^*(\mathcal{S}, \mathcal{P})$, τ for tree. The fact that entropy $H(S_*; \mathbb{P}) = H(S; \mathbb{P})$ defines a 1-cocycle is a result of an equation of Fadeev, generalized by Baez, Fritz and Leinster [17], who gave another interpretation, based on true operad structure over the set of all finite probability laws.

Theorem 5 (cf. [16]): If Ω has more than four points, $H_\tau^1(\Pi(\Omega), \Delta(\Omega))$ is the one dimensional vector space generated by the entropy.

Another right action of \mathcal{V} on \mathcal{F} is given by (6) where on the right side $\mathbb{P} | (S = s)$ is replaced by \mathbb{P} itself. From here and the simplicial structure associated to θ and μ , we define an operator δ_t , which gives a twisted version of information co-homology as we have done in the first paragraph. This allows us to define higher information quantities for strategies: for $N = 2M + 1$ odd $I_{\tau, N} = -(\delta \delta_t)^M H$ and for $N = 2M + 2$ even $I_{\tau, N} = \delta_t (\delta \delta_t)^M H$.

This gives for $N = 2$ a notion of mutual information between a variable S of length m

and a collection T of m variables T_1, \dots, T_m :

$$I_\tau(S; T; \mathbb{P}) = \sum_{i=1}^{i=m} \mathbb{P}(S = i)(H(T_i; \mathbb{P}) - H(T_i; \mathbb{P}|S = i)). \quad (8)$$

When all the T_i are equals we recover the ordinary mutual information of Shannon.

THE FORMS OF INFORMATION STRATEGIES

A rooted tree Γ decorated by \mathcal{S}_* can be seen as a strategy to discriminate between points in Ω . For each vertex s there is a minimal set of chained edges $\alpha_1, \dots, \alpha_k$ connecting s_0 to s ; the cardinal k is named the *level* of s ; this chain defines a sequence $(F_0, v_0; F_1, v_1; \dots; F_{k-1}, v_{k-1})$ of observables and values of them; then we can associate to s the subset Ω_s of Ω where each F_j takes the value v_j . At a given level k the sets Ω_s form a partition π_k of Ω ; the first one π_0 is the unit partition of length 1, and π_l is finer than π_{l-1} for any l . By recurrence over k it is easy to deduce from the orderings of the values of F_s an embedding in the Euclidian plane of the subtrees $\Gamma(k)$ at level k such that the values of the variables issued from each vertex are oriented in the direct trigonometric sense, thus π_k has a canonical ordering ω_k . Remark that many branches of the tree gives the empty set for Ω_s after some level; we name them dead branches. It is easy to prove that the set $\Pi(\mathcal{S})_*$ of ordered partitions that can be obtained as a (π_k, ω_k) for some tree Γ and some level k is closed by the natural ordered join operation, and, as $\Pi(\mathcal{S})_*$ contains π_0 , it forms a monoid, which contains the monoid $M(\mathcal{S}_*)$ generated by \mathcal{S}_* .

Complete discrimination of Ω by \mathcal{S}_* exists when the final partition of Ω by singletons is attainable as a π_k ; optimal discrimination correspond to minimal level k . When the set Ω is a subset of the set of words x_1, \dots, x_N with letters x_i belonging to given sets M_i of respective cardinalities m_i , the problem of optimal discrimination by observation strategies Γ decorated by \mathcal{S}_* is equivalent to a problem of minimal rewriting by words of type $(F_0, v_0), (F_1, v_1), \dots, (F_k, v_k)$; it is a variant of optimal coding, where the alphabet is given. The topology of the poset of discriminating strategies can be computed in terms of the free Lie algebra on Ω , cf. [15].

Probabilities \mathbb{P} in \mathcal{P} correspond to *a priori* knowledge on Ω . In many problems \mathcal{P} is reduced to one element, that is the uniform law. Let s be a vertex in a strategic tree Γ , and let \mathcal{P}_s be the set of probability laws that are obtained by conditioning through the equations $F_i = v_i; i = 0, \dots, k - 1$ for a minimal chain leading from s_0 to s . We can consider that the sets \mathcal{P}_s for different s along a branch measure the evolution of knowledge when applying the strategy. The entropy $H(F; \mathbb{P}_s)$ for F in \mathcal{S}_* and \mathbb{P}_s in \mathcal{P}_s gives a measure of information we hope when applying F at s in the state \mathbb{P}_s . The maximum entropy algorithm consists in choosing at each vertex s a variable that has the maximal conditioned entropy $H(F; \mathbb{P}_s)$.

Theorem 6 (cf. [16]): To find one false piece of different weight among N pieces for $N \geq 3$, when knowing the false piece is unique, by the minimal numbers of weighing, one can use the maximal entropy algorithm.

However we have another measure of information of the resting ambiguity at s , by

taking for the *Galois group* G_s the set of permutations of Ω_s which respect globally the set \mathcal{P}_s and the set of restrictions of elements of \mathcal{S}_* to Ω_s , and which preserve one by one the equations $F_i = v_i$. Along branches of Γ this gives a decreasing sequence of groups, whose successive quotients measure the evolution of acquired information in an algebraic sense.

Problem 3: Generalize Theorem 6. Can we use algorithms based on the Galoisian measure of information? Can we use higher information quantities associated to trees for optimal discrimination?

CONCLUSION AND PERSPECTIVE

Concepts of Algebraic topology were recently applied to Information theory by several researchers. In particular notions coming from category theory, homological algebra and differential geometry were used for revisiting the nature and scope of entropy, cf. for instance Baez et al. [17], Marcolli and Thorngren [18] and Gromov [19]. In the present note we interpreted entropy and Shannon information functions as co-cycles in a natural co-homology theory of information, based on categories of observable and complexes of probability. This allowed us to associate topological figures, like Borromean links, with particular configuration of mutual dependency of several observable quantities. Moreover we extended these results to a dynamical setting of system observation, and we connected probability evolutions with the measures of ambiguity given by Galois groups. All those results provide only the first steps toward a developed Information Topology. However, even at this preliminary stage, this theory can be applied to the study of distribution and evolution of Information in concrete physical and biological systems. This approach already proved its efficiency for detecting collective synergic dynamic in neural coding [12], in genetic expression [20], in cancer signature [21], or in signaling pathways [22]. In particular, information topology could provide the principles accounting for the structure of information flows in biological systems and notably in the central nervous system of animals.

ACKNOWLEDGMENTS

We thank MaxEnt14 for the opportunity to present these researches to the information science community. We thank Guillaume Marrelec for discussions and notably his participation to the research of the last part on optimal discrimination. We thank Frederic Barbaresco and Alain Chenciner for discussions and comments on the manuscript. We thank the "Institut des Systemes complexes" (ISC-PIF), and region Ile-de-France, for the financial support and hosting of P. Baudot.

REFERENCES

1. C. E. Shannon, *The Bell System Technical Journal* **27**, 379–423 (1948).

2. A. Kolmogorov, *Russ. Math. Surv.* **38** **29** (1983).
3. R. Thom, *Stabilité structurelle et morphogénèse.*, deuxième édition, InterEdition, Paris, 1977.
4. S. Mac Lane, *Categories for the Working Mathematician.*, Graduate Texts in Mathematics. Springer, 1998.
5. S. Mac Lane, *Homology.*, Classic in Mathematics, Springer, Reprint of the 1975 edition, 1975.
6. K. T. Hu, *Theory Probab. Appl.* **7**(4), 439–447 (1962).
7. P. Baudot, and D. Bennequin, *Preprint I* (2014).
8. P. Elbaz-Vincent, and H. Gangl, *Compositio Mathematica* **130**(2), 161–214 (2002).
9. J. Cathelineau, *Math. Scand.* **63**, 51–86 (1988).
10. J.-L. Loday, and B. Valette, *Algebraic operads.*, Springer, 2012.
11. H. Matsuda, *Physica A: Statistical Mechanics and its Applications.* **294** (1-2), 180–190 (2001).
12. N. Brenner, S. Strong, R. Koberle, and W. Bialek, *Neural Computation.* **12**, 1531–1552 (2000).
13. M. Nielsen, and I. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, 2000.
14. P. Baudot, and D. Bennequin, *Preprint II* (2014).
15. B. Fresse, *Contemp. Math. Amer. Math. Soc.* **346**, pp. 115–215 (2004).
16. P. Baudot, and D. Bennequin, *Preprint III* (2014).
17. J. Baez, T. Fritz, and T. Leinster, *Entropy* **13**, 1945–1957 (2011).
18. M. Marcolli, and R. Thorngren, *arXiv 10.4171/JNCG/159* **Vol. abs/1108.2874** (2011).
19. M. Gromov, *unpublished manuscript <http://www.ihes.fr/~gromov/PDF/structre-serch-entropy-july5-2012.pdf>* (2013).
20. J. Watkinson, K. Liang, X. Wang, T. Zheng, and D. Anastassiou, *The Challenges of Systems Biology: Ann. N.Y. Acad. Sci.* **1158**, 302–313 (2009).
21. H. Kim, J. Watkinson, V. Varadan, and D. Anastassiou, *BMC Medical Genomics* , **3:51** (2010).
22. S. Uda, T. H. Saito, T. Kudo, T. Kokaji, T. Tsuchiya, H. Kubota, Y. Komori, Y. ichi Ozaki, and S. Kuroda, *Science* **Vol. 341 no. 6145**, pp. 558–561 (2013).