

November 7, 2018

Symmetries among multivariate information measures explored using Möbius operators*

David J. Galas^{1,2} and Nikita A. Sakhnenko¹

¹Pacific Northwest Research Institute
720 Broadway
Seattle, Washington 98122
USA

² correspondence to: David J. Galas, dgalas@pnri.org

* *This paper is a revision of a previous manuscript, arXiv: 1601.06780*

November 7, 2018

Abstract

Information-related measures are useful tools for multi-variable data analysis, as measures of dependence among variables, and as descriptions of order and disorder in biological and physical systems. Measures, like marginal entropies, mutual / interaction / multi-information, have long been used in a number of fields including descriptions of systems complexity and biological data analysis. The mathematical relationships among these measures are therefore of significant inherent interest. Relations between common information measures include the duality relations based on Möbius inversion on lattices. These are the direct consequence of the symmetries of the lattices of the sets of variables (subsets ordered by inclusion). While these relationships are of significant interest there has been, to our knowledge, no systematic examination of the full range of relationships of this diverse range of functions into a unifying formalism as we do here. In this paper we define operators on functions on these lattices based on the Möbius inversions that map functions into one another (Möbius operators). We show that these operators form a simple group isomorphic to the symmetric group S_3 . Relations among the set of functions on the lattice are transparently expressed in terms of the operator algebra, and, applied to the information measures, can be used to derive a wide range of relationships among diverse information measures. The Möbius operator algebra is naturally generalized which yields extensive new relationships. This formalism now provides a fundamental unification of information-related measures, and the isomorphism of all distributive lattices with the subset lattice implies an even broader application of these results.

Keywords: information, entropy, interaction-information, multi-information, Möbius inversion, lattices, multivariable dependence, symmetric group, MaxEnt, networks

November 7, 2018

1. Introduction

The description of order and disorder in systems of all kinds is fundamental. In the physics and chemistry of condensed matter it plays a central role, but for systems with biological levels of complexity, including interactions of genes, macromolecules, cells and of networks of neurons, it is also central, and certainly not well understood. Mathematical descriptions of the underlying order, and transitions between states of order, are still far from satisfactory and a subject of much current research (for example [16,17]). The difficulty arises in several ways, but the dominant contributors are, in our view, the number and high degree of effective interactions among components, and their non-linearity. There have been many efforts to define information-based measures as a language for describing the order and disorder of systems and the transfer of information. Negative entropy, joint entropies, multi-information and various manifestations of Kullback-Leibler (K-L) divergence are among the key concepts. Interaction information is one of these. It is an entropy-based measure for multiple variables introduced by McGill in 1954 [1]. It has been used effectively in a number of developments and applications of information-based analysis [2-5], and has several interesting properties, including symmetry under permutation of variables, like joint entropies and multi-information, though its interpretation as a form of information in the usual sense is ambiguous as it can have negative values. In previous work we have proposed complexity and dependence measures related to this quantity [6,9].

Here we focus on elucidating the character and source of some of the mathematical properties that relate these measures, and on extending both the definitions and spectrum of relations among all these quantities. The formalism presented here can thus be viewed as a unification of a wide range of information-related measures in the sense that the relations between them are elucidated. This paper is structured as follows. We briefly review a number of definitions and review preliminaries relevant to information measures and Möbius inversion. In the next section we define the operators that map the functions on the lattice into one another, expressing the Möbius inversions as operator equations. We then

November 7, 2018

determine the products of the operators and, completing the set of operators with a lattice complement operator, we show that they form a group that is isomorphic to the symmetric group, S_3 . In the next section we express previous results in defining dependency and complexity measures in terms of the operator formalism, and illustrate relationships between many commonly used information measures, like interaction information and multi-information. We derive a number of new relations using the formalism, and point out the relationship between multi-information and certain maximum entropy limits. This suggests a wide range of maximum entropy criteria in the relationships inherent in the operator algebra. The next section focuses on the relations between these functions and the probability distributions underlying the symmetries. We then illustrate an operator equation expressing our dependence measure in terms of conditional log likelihood functions. Finally, we define a generalized form of the inversion relation, which also has S_3 symmetry, and show how these operators on functions can be additively decomposed in a variety of interesting ways.

2. Preliminaries

We review the elements of information theory and lattices that are relevant to this paper here, and clarify some notational conventions used.

2.1 Information Theory

Consider a set of n discrete variables $v_n = \{X_1, X_2, \dots, X_n\}$ sometimes denoted as v if there is no ambiguity. We use v_{n-1} to denote the set v_n without variable X_n . We write $\Pr(v_n)$ to denote a joint probability density function over v_n , and $\Pr(X_n|v_{n-1})$ to denote a conditional probability density function.

Marginal entropy of a single variable X_i is defined as $H(X_i) = -\sum_{x_i} \Pr(x_i) \log(\Pr(x_i))$. Similarly given a set of variables v_n , *joint entropy* is defined as $H(v_n) = -\sum_s \Pr(s) \log(\Pr(s))$, where s traverses all possible states of v_n . We write $H(X_n|v_{n-1})$ to denote *conditional entropy* of X_n on the rest of the variables

November 7, 2018

v_{n-1} . The difference in joint entropy of sets of variables with and without X_n is called *differential entropy* $\delta H(v_n)$:

$$\delta H(v_n) \equiv H(v_n) - H(v_{n-1}) \quad (1)$$

The *mutual information* $I(X_i, X_j)$ measuring the mutual dependence between two variables X_i and X_j is defined as

$$I(X_i, X_j) \equiv -\sum_{x_i x_j} \Pr(x_i, x_j) \log \left(\frac{\Pr(x_i, x_j)}{\Pr(x_i)\Pr(x_j)} \right) \quad (2)$$

Equivalently, the mutual information can be expressed via marginal and joint entropies:

$$I(X_i, X_j) \equiv H(X_i) + H(X_j) - H(X_i, X_j) \quad (3)$$

Similar to Equation 3, given three variables X_i, X_j , and X_k , the conditional mutual information can be defined as

$$I(X_i, X_j | X_k) = H(X_i | X_k) + H(X_j | X_k) - H(X_i, X_j | X_k) \quad (4)$$

A generalization of mutual information to more than two variables is called *interaction information*. For three variables it is defined as the difference between mutual information with and without knowledge of the third variable

$$I(X_i, X_j, X_k) = I(X_i, X_j | X_k) - I(X_i, X_j) \quad (5)$$

When expressed entirely in terms of entropies we have

$$\begin{aligned} I(X_i, X_j, X_k) = & H(X_i) + H(X_j) + H(X_k) \\ & - H(X_i, X_j) - H(X_i, X_k) - H(X_j, X_k) \\ & + H(X_i, X_j, X_k) \end{aligned} \quad (6)$$

Consider the interaction information for a set of n variables v_n

$$I(v_n) = -\sum_{\tau \subseteq v_n} (-1)^{|\tau|} H(\tau) \quad (7)$$

Given Equation 7, we define the *differential interaction information*, Δ , as the difference between values of successive interaction informations arising from adding variables

November 7, 2018

$$\Delta(\nu_{n-1}; X_n) \equiv I(\nu_n) - I(\nu_{n-1}) = -I(\nu_{n-1}|X_n) \quad (8)$$

The last equality in Equation 8 comes from the recursive relation for the interaction information, Equation 5. The differential interaction information is based on providing the target variable X_n to be added to the set of $n - 1$ variables, and is therefore asymmetric. If we multiply differential interaction informations with all possible choices of the target variable, the resulting measure will be symmetric and called *symmetric delta*, $\bar{\Delta}$

$$\bar{\Delta}(\nu_n) = (-1)^n \prod_{X \in \nu_n} \Delta(\nu_n - \{X\}, X) \quad (9)$$

There is another measure for multivariable dependence called *multi-information*, or *total correlation* [10]. It is defined as the difference between the sum of single entropies for each variable of a set and the joint entropy for the entire set

$$\Omega(\nu_n) \equiv \sum_{X_i} H(X_i) - H(\nu_n) \quad (10)$$

Multi-information is frequently used because it is always positive and goes to zero when all the variables are independent. We can think of it as a kind of conglomerate of dependencies among members of the set ν_n .

At the two-variable level multi-information, Kullback-Leibler divergence and interaction information are all identical, and equal to mutual information. The interaction information $I(\nu_n)$ for a set of n variables, obeys a recursion relation that parallels that for the joint entropy of sets of variables, $H(\nu_n)$, which is derived in turn directly from the probability chain rule:

$$\begin{aligned} H(\nu_n) &= H(\nu_{n-1}) + H(X_n|\nu_{n-1}) \\ I(\nu_n) &= I(\nu_{n-1}) - I(\nu_{n-1}|X_n) \end{aligned} \quad (11)$$

where the second terms on the right are conditionals. These two information functions are known to be related by Möbius inversion [2-5]. There is an inherent duality between the marginal entropy functions and the interaction information functions based on Möbius inversion, which we will show in detail in Section 2. Bell described an elegantly symmetric form of the inversion and identified the source of

November 7, 2018

this duality in the lattice associated with the variables [2]. The duality is based on the inclusion lattice of the set of variables. We start with this symmetric inversion relation and extend it to an algebra of operators on these lattices. We will first define the lattice and other relevant concepts from lattice theory before discussing Möbius inversion.

2.2 Lattice theory

We now review some definitions from the lattice theory [13]. We say that a set P is a *poset* (a *partially ordered set*) if there is a partial order defined on it, $\langle P, \leq \rangle$. A partial order (\leq) is a binary relation that is reflexive, antisymmetric, and transitive. Note that we would write $x \leq y$ to denote the partial order between elements x and y of a poset. Note also that an inverse of a partial order is a partial order. A *chain* of a poset $\langle P, \leq \rangle$ is a subset $C \subseteq P$ such that for any two elements $x, y \in C$ either $x \leq y$ or $y \leq x$. Similarly, a *path* of length k is a subset $C \subseteq P$ such that $C = [x_1, x_2, \dots, x_k]$ for any $1 \leq i < k$ either $x_i \leq x_{i+1}$ or $x_{i+1} \leq x_i$. Note that any chain is a path, but not other way around, since x_i and x_j of a path need not be ordered if $|i - j| > 1$.

Let X be a subset of a poset $\langle P, \leq \rangle$. The *minimum* of X , if exists, is $\min(X)$ such that $\min(X) \in X$ and for any $x \in X: \min(X) \leq x$. Similarly, the *maximum* of X , if exists, is $\max(X)$ such that $\max(X) \in X$ and for any $x \in X: x \leq \max(X)$. A poset $\langle P, \leq \rangle$ has a *top element* (a greatest element) T iff $T \in P$ and for any $x \in P: x \leq T$. Similarly, a poset $\langle P, \leq \rangle$ has a *bottom element* (a least element) \perp iff $\perp \in P$ and for any $x \in P: \perp \leq x$.

The *dual* of a poset $\langle P, \leq \rangle$ is $\langle P, \geq \rangle$, where \geq is the inverse partial order of \leq . For any statement based on the partial order \leq and true about all posets, the dual statement (based on the inverse partial order \geq) is also true about all posets.

For a poset $\langle P, \leq \rangle$ we call $D \subseteq P$ a *down-set* (or an *ideal*) iff, for any $d \in D: \forall p \in P: (p \leq d) \Rightarrow (p \in D)$. Dually, we call $U \subseteq P$ an *up-set* (or a *filter*) iff for any $u \in U: \forall p \in P: (p \geq u) \Rightarrow (p \in U)$. Note that a set S is a down-set of $\langle P, \leq \rangle$ iff its set complement $P \setminus S$ is an up-set of $\langle P, \leq \rangle$.

November 7, 2018

Given a subset S of a poset $\langle P, \leq \rangle$, M is an *upper bound* of S iff for any $x \in S: x \leq M$. And dually, m is a *lower bound* of S iff for any $x \in S: m \leq x$. The *join* of S , if exists, is called an upper bound of S , which is the least of the upper bounds of S . And dually, the *meet* of S is the greatest lower bound of S .

A poset where for every two elements there exist the unique join and meet is called a *lattice*. A lattice that contains a top element T and a bottom element \perp , such that for every element x of the lattice, $\perp \leq x \leq T$, is called a *bounded lattice*. An *inclusion lattice* (also called a *subset lattice*) is a typical example of a lattice defined on all subsets of a given set S ordered by a subset inclusion \subseteq . If a set S is finite, then its corresponding inclusion lattice is bounded, where the top element is S itself and the bottom element is the empty set.

3. Möbius Dualities

Many applications make use of the relations among information theoretic quantities like joint entropies and interaction information that are formed by what can be called Möbius duality [2]. Restricting ourselves to functions on subset lattices, we note that a function on a lattice is a mapping of each of the lattice element (subset of variables) to the reals. The Möbius function for this lattice is $\mu(\nu, \tau) = (-1)^{|\nu| - |\tau|}$ where τ is a subset of ν , $|\tau|$ is the cardinality of the subset.

2a. Möbius Inversion

Consider a set of n variables ν and define g , the dual of f for the set of variables

$$g(\eta) = \sum_{\tau \subseteq \eta} \mu(\nu, \tau) f(\tau) = \sum_{\tau \subseteq \eta} (-1)^{|\nu| - |\tau|} f(\tau); \eta \subseteq \nu \quad (12a)$$

Note that function g is the interaction information if f were the entropy function H , adopting the sign convention of [2]. It can easily be shown that the symmetric relation

November 7, 2018

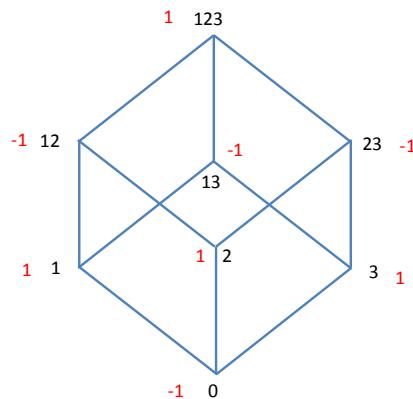
holds,

$$f(\eta) = \sum_{\tau \subseteq \eta} (-1)^{|\nu| - |\tau|} g(\tau); \eta \subseteq \nu \quad (12b)$$

The relations defined in Equation 12(a,b) represent a symmetric form of Möbius inversion, and the functions f and g can be called Möbius duals.

Now consider an inclusion lattice. The Möbius inversion is a convolution of the Möbius function with any function defined on the lattice over all its elements (subsets) between the argument subset, τ , of the function and the empty set. The summation in the inversion is over all the elements on all chains between τ and the empty set, counting the elements only once, which is called a down-set of the inclusion lattice (see section 2). The empty set, at the limit of the range of the convolution, can be considered as the “reference element”. We use the idea of a reference element in section 6 in generalizing the inversion relations. The range of the convolution can of course be limited at the top element (largest subset) and the bottom element of the lattice. In defining the Möbius operators below we need to carefully define how the range is determined.

To illustrate the relations concretely the nodes and the Möbius function are shown graphically for three variables in Figure 1. When the functions in Equation 12 are mapped onto the lattice for three variables, these equations represent the convolution of the lattice functions and the Möbius function over the lattice.



November 7, 2018

Figure 1. The Hasse diagram of the subset lattice for three variables. The numbers in black are the variable subsets, while the Möbius function $\mu(\nu, \tau)$ on this lattice (1 or -1) is indicated in red.

3b. Möbius operators

The convolutions with the Möbius function over the lattice in Equation 12 define mappings that can be expressed as operators. The operators can be thought of as mapping of one function on the lattice into another. A function on the lattice, in turn, is a map of the subsets of variables at each node into the real numbers.

Definition: *Möbius down-set operator.* Given a set of variables, τ , which is the element in the inclusion lattice, we define the Möbius *down-set operator*, \hat{m} , that operates on a function on this lattice.

$$\hat{m}(f(\tau)) \equiv \sum_{\eta \subseteq \tau} (-1)^{|\eta|-1} f(\eta) = g(\tau), \tau \subseteq \nu \quad (13a)$$

The down-set operator is defined as an operator form of the convolution with the Möbius function: the sum over the lattice of subsets of τ , of product of the values of the function times the Möbius function. The upper bound of this convolution is the entire set, τ , the lower bound is the empty set.

Likewise, we can define a Möbius *up-set operator*. The definition is significantly different in that the lower limit needs to be specified, whereas the downset operator uses the empty set unless otherwise specified.

Definition: *Möbius up-set operator.* Given a set of variables, ν , the operator, \hat{M} , is defined as the convolution operator on a function on the inclusion lattice which is the sum is over the lattice of supersets of τ .

$$\hat{M}(f(\tau)) \equiv \sum_{\eta \supseteq \tau} (-1)^{|\eta|+1} f(\eta) = h(\tau), \eta, \tau \subseteq \nu \quad (13b)$$

November 7, 2018

The lower bound of this convolution is τ and the upper bound is the complete set ν .

Given a function, f , Equations 13(a,b) define the functions g and h , respectively: the *down-set* and *up-set* inverses, or duals, of f . The sum in the expression of Eqn. 13a is the same as the symmetric form of the Möbius inversion [2]: f and g in Eqn. 13a are interchangeable, dual with respect to the down set operator (see Eqn. 12a and 12b). Given a set to begin with, the up-set operator is referenced to the full set, and the down-set operator to the empty set.

From Equation 13a Möbius inversion implies that applying the down-set operator twice yields the identity, $\hat{m}^2 = \hat{I}$. Similarly, using Equation 13b we can show that $\hat{M}^2 = \hat{I}$. This is an expression of the duality: this idempotent property of the Möbius operators is equivalent to the symmetry in Equation 12; in other words, the exchangability in these equations, or duality of the functions is exactly the same property as the idempotency of the operators. The relationships between pairs of the dual functions, generated by the operators are shown in the diagram in Figure 2. The range of the convolution operator is clear here, but this is not always true, and where it is ambiguous we use a subscript on the operator to identify the reference set. We will need this in section 4.

$$h(\tau) \xleftrightarrow{\hat{M}} f(\tau) \xleftrightarrow{\hat{m}} g(\tau)$$

Figure 2. The Möbius operators define the duality relationships between the functions on the subset lattice.

To advance this formalism further we need to define another operator on the inclusion lattice. The inversion, or complementation, operator \hat{X} has the effect of mapping function values of all elements of the lattice (subsets) into the function values of the corresponding set complement elements. For example, node 1 maps into node 23 in Figure 1. Viewed as a 3D geometric space, as shown in Figure 1, the complementation corresponds to an inversion of the lattice, all such 3-D coordinates

November 7, 2018

mapping into their opposites through the origin at the geometric center of the cube. We define the operator \hat{X} , acting on functions whose arguments are subsets τ of the set ν

$$\hat{X}f(\tau) = (-1)^{|\nu|}f(\tilde{\tau}): \tau \subseteq \nu, \tau \cap \tilde{\tau} = \emptyset, \tau \cup \tilde{\tau} = \nu \quad (14)$$

The sign change factor is added since inversion of the lattice also has the effect of shifting the Möbius function by a sign for odd numbers of total variables on the lattice.

If we define the composite operators, \hat{P} and \hat{R} , as

$$\hat{P} = \hat{X}\hat{M}, \hat{R} = \hat{X}\hat{m} \quad (15)$$

the pairwise relations among the functions and the operators shown in Figure 3 then follow. The three- and four-variable case for the relationships in Figure 3 can easily be confirmed by direct calculation, and as it happens the general case is also easy to prove. The proofs are direct and follow from the Möbius inversion sums, by keeping track of the effects of each of the inversion and convolution operators, and are not presented here.

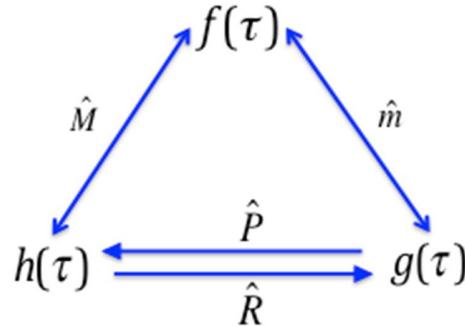


Figure 3. Diagram of the mappings of the functions on the subset lattice into one another by the operators. The operators \hat{P} and \hat{R} are: $\hat{P} = \hat{X}\hat{M}, \hat{R} = \hat{X}\hat{m}$.

Let us collect the operators of Figure 3, add the identity operator and the composite operators \hat{P} and \hat{R} , and calculate the full product table of the set of operators.

The full product table of the operators is shown in Table 1.

November 7, 2018

		right					
		\hat{I}	\hat{m}	\hat{X}	\hat{M}	\hat{P}	\hat{R}
		\hat{I}	\hat{I}	\hat{m}	\hat{X}	\hat{M}	\hat{P}
left		\hat{m}	\hat{m}	\hat{I}	\hat{P}	\hat{R}	\hat{X}
		\hat{X}	\hat{X}	\hat{R}	\hat{I}	\hat{P}	\hat{M}
		\hat{M}	\hat{M}	\hat{P}	\hat{R}	\hat{I}	\hat{m}
		\hat{P}	\hat{P}	\hat{M}	\hat{m}	\hat{X}	\hat{R}
		\hat{R}	\hat{R}	\hat{X}	\hat{M}	\hat{m}	\hat{I}
						\hat{I}	\hat{P}

Table 1. The product table for the 6 operators above. The operators \hat{P} and \hat{R} are defined as $\hat{P} = \hat{X}\hat{M}$, $\hat{R} = \hat{X}\hat{m}$. The convention is that the top row is on the right and the left column on the left in the products indicated; e.g. $\hat{M}\hat{X} = \hat{R}$, $\hat{X}\hat{M} = \hat{P}$.

It is immediately clear that this set of 6 operators forms a group: the set is closed, it contains an identity element, all its elements have an inverse included, and they demonstrate associativity. Furthermore, examination of the table immediately shows that it is isomorphic to the symmetric group S_3 , the group of permutations of three objects. Table 2 shows the 3x3 matrix representation of the group S_3 , with the one line notation of the operator effect, and the correspondence between the Möbius operators and the S_3 representation.

<u>One line Notation:</u> <u>(Permutation)</u>	<u>Matrix Representation</u> (left action convention)	<u>Möbius Operator</u>
123	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	\hat{I}
213	$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	\hat{m}

November 7, 2018

132	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$	\hat{M}
321	$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$	\hat{X}
231	$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$	\hat{P}
312	$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$	\hat{R}

Table 2. The 3x3 matrix representation of symmetric group S_3 and the corresponding Möbius operators. The one-line notation on the left shows the permutations.

Note that while the operators themselves, which act on functions, depend on the number of variables since they define convolutions, their relationships do not. Thus, the group structure is independent of the number of variables in the lattice. For any number of variables the structure is simply the permutation group, S_3 .

4. Connections to the Deltas

The differential interaction information and the symmetric deltas were defined in [4] as overall measures of both dependence and complexity (see definitions in Equations 8 and 9). We will now show the connection between these deltas and our operator algebra. We will use the three-variable case to illustrate the connection. If the marginal entropies are identified with the function f in Equation 12, and the interaction informations with g , then the differential interaction information is identified with h . For the three-variable case these examples are shown using simplified notation,

November 7, 2018

$$h(1) = \Delta(23; 1), \quad h(2) = \Delta(13; 2), \quad h(3) = \Delta(12; 3) \quad (16b)$$

Simplifying the notation we can express the relations between these functions using the Möbius operator as

$$\Delta(\tau; X) = \widehat{M}H(X) = -I(\tau|X) \quad (16c)$$

The full set of the lattice is $\tau \cup \{X\}$ and the variable X is singled out as in Equations 11 and 16a. Furthermore, the convolution takes place over the set $\tau \cup \{X\}$. Equation 16c, if interpreted properly, provides a simple connection between the deltas and the Möbius operator algebra, and expresses a key relation (Theorem 1). We have proved the following theorem.

Theorem 1. The Möbius *up-set* operator acting on the *join-irreducible elements* of the lattice of marginal entropies generates the conditional interaction informations, the deltas, for the full set of variables of the lattice.

Join-irreducible lattice elements are all those that cannot be expressed as the join, or union, of other elements. In this case they are all the single variables. Since the deltas are differentials of the interaction information at the top of the lattice (the argument of the function is the full set), their expression in terms of the join-irreducible elements is the most fundamental form. To illustrate the relation more concretely, Figure 4 shows the specific connection between the join-irreducible elements and deltas for the four-variable lattice.

November 7, 2018

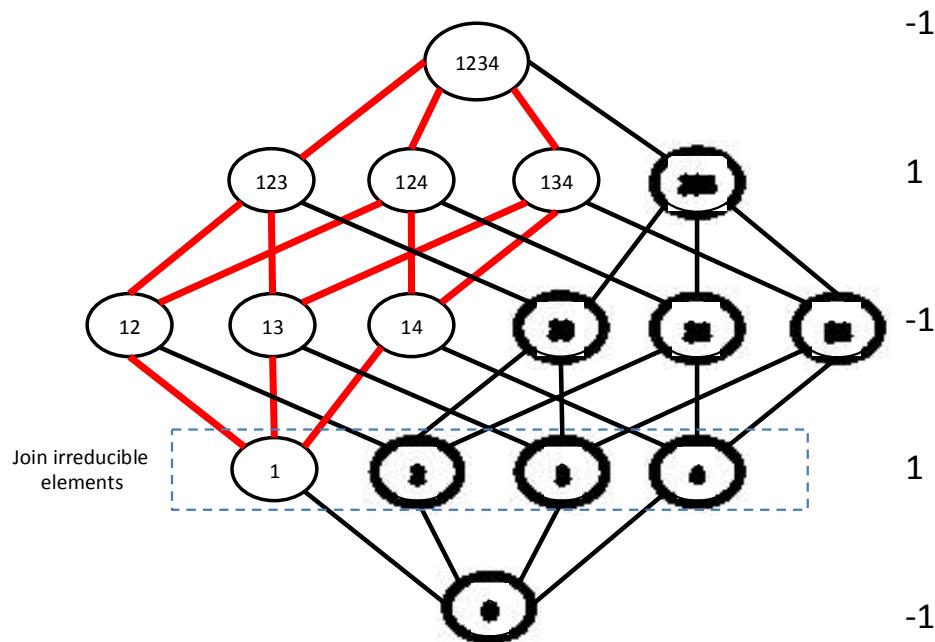


Figure 4. The four-variable lattice showing the 4 join-irreducible elements that generate the symmetric deltas as in Equation 16c. Möbius function values are shown on the right, and the red lines connect the elements of the delta function, $\Delta(234; 1)$, which form a 3D-cube.

A general statement of this connection emerging from this geometric picture is a general property of the algebraic structure of the subset lattice.

Corollary: The differential of one function on the lattice corresponds to the up-set operator on another function of the join-irreducible elements.

Written in terms of the functions related by the inversions, and using the same set notation as above, X indicating a join-irreducible element, we can state this general result as follows.

If $g(\tau) = \hat{M}f(\tau)$ and X is a join-irreducible element of lattice, then

$$\hat{M}f(X) = h(\tau; X) = g(\tau|X) \quad (17)$$

where the final term is a conditional form of the g function in which X is instantiated. This is defined as function over all τ for which $X \in \tau$. These deltas, and

November 7, 2018

delta-like functions more generally, are represented as convolutions over a lattice that is one dimension less than the full variable set lattice.

We have previously proposed the symmetric delta (the product of all variable permutations of the delta function, h) as a measure of complexity, and of collective variable dependence [6]. The symmetric delta expression simply the product of the individual deltas is seen to be the product of the results of the up-set operator acting on the functions of *all of the join-irreducible elements of the entropy lattice*. Note that by Equation 11 both the conditional entropies and conditional interaction informations, since they correspond to the differentials, imply a path independent chain rule. Note that these kinds of differential functions include more than just those keyed on the join-irreducible elements as shown in the next section.

5. Symmetries reveal a wide range of new relations

The system of functions and operators defined in the previous section reveals a wide range of relationships. Examination of Equation 11 and comparision with 16c shows that delta is also related to the differential entropy (defined by Equation 1) measuring the change in the entropy of a set when we consider an additional variable. Applying the down-set operator to Equation 1, and using sets v_n and v_{n-1} as the upper bounds, gives us

Theorem 2. Given the definition of the differential entropy (Equation 1), and the definitions of the up-set and down-set operators:

$$\begin{aligned}\widehat{m}\delta H(v_n) &= \widehat{m}(H(v_n) - H(v_{n-1})) = I(v_n) - I(v_{n-1}) = -I(v_n|X_n) \\ \widehat{m}\delta H(v_n) &= \widehat{M}(H(X_n))\end{aligned}\quad (18)$$

where X_n is the element that is the difference between the sets v_n and v_{n-1} .

We can consider δ as an operator, but note that it does not define a convolution over elements of the lattice as do the Möbius operators. Considering δ as an operator we note that δ and \widehat{m} commute. The duality between H and I implies a dual version of Equation 18 as well, which we will not derive. If we apply other

November 7, 2018

operators to the expression in equation 18 we find another set of relations among these marginal entropy functions. For example, another remarkable symmetry emerges.

$$\begin{aligned}\delta H(v_n) &= \hat{m}\hat{M}H(X_n) = \hat{X}\hat{m}H(X_n) = \hat{R}H(X_n) \\ H(X_n) &= \hat{P}\delta H(v_n)\end{aligned}\quad (19)$$

Equations 18 and 19 relate functions of the lattice elements to functions of join irreducible elements.

There are further symmetries in this set of information functions. Consider the mapping diagram of Figure 3. If we define a function which is simply the delta function with each lattice element mapped into its set complement, that is, acted on by the lattice inversion operator, from Equation 16c we have

$$\Phi \equiv \hat{X}\Delta, \quad \hat{X}\hat{m}\Phi = \hat{X}\hat{m}\hat{X}\Delta = H \quad (20)$$

Then these functions occupy different positions in the mapping diagram as seen in Figure 5. Several other such modifications can be generated by similar operations.

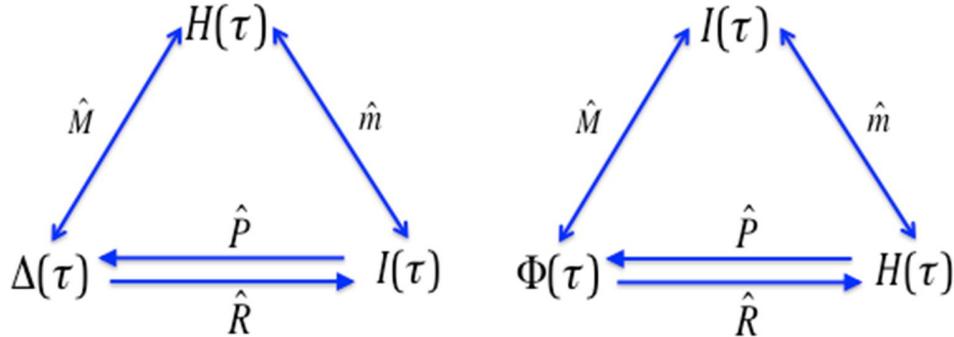


Figure 5. A simple modification of one of the functions by lattice inversion modifies the position of functions in the mapping diagram. The original diagram is on the left, the result of Δ modified by inversion is on the right. Note that the idempotent property of \hat{M} is applied to the Φ function relation.

There are a large number of similar relations that can be generated by such considerations.

There are other information-based measures that we can express using the operator algebra. Because it is a widely used measure for multi-variable

November 7, 2018

dependence we will now examine the example of multi-information Ω defined by Equation 10. In terms of entropy functions on the lattice elements, Ω , as expressed in this equation, can be thought of as the sum of the join-irreducible elements, minus the top element or the join of the inclusion lattice. To apply the down-set operator to the terms in Equation 10 we must carefully define the bounds of the convolutions. If we calculate the convolution over the Ω function, we have

$$\widehat{m}\Omega(v_n) = \sum_{X_i} \widehat{m}H(X_i) - I(v_n) \quad (21)$$

Since the upper bound of the down-set operator is defined as the argument set of the function, the down-set of a single variable function is the function itself (since $H_i - 0 = H_i$). Note that we are using the distributive property of the operator here. The application of the up-set operator to the multi-information function on the lattice, on the other hand, gives us

$$\widehat{M}\Omega(v_n) = \sum_{X_i} \Delta(v_{n-1}; X_i) - H(v_n) \quad (22)$$

Note that the multi-information is a composite function and therefore the results of the action of the (distributive) Möbius operators are also composite functions.

6. Relation to probability densities

6a. Conditional log Likelihoods and Deltas

Writing the differential entropy in terms of the probability distributions, using the definitions of the joint entropies and the probability chain rule, gives

$$\delta H(v_n) = -\langle \ln \frac{\Pr(v_n)}{\Pr(v_{n-1})} \rangle = -\langle \ln \Pr(X_n|v_{n-1}) \rangle = H(X_n|v_{n-1}) \quad (23a)$$

November 7, 2018

For simplicity of notation we define π as the expectation value on the right. We have

$$\pi(X_n|\nu_{n-1}) \equiv -\langle \ln \frac{\Pr(\nu_n)}{\Pr(\nu_{n-1})} \rangle = \langle \ln \Pr(\nu_{n-1}) - \ln \Pr(\nu_n) \rangle \quad (23b)$$

From Equation 23 we see that π is a conditional log likelihood function. By applying the down-set operator, \hat{m} , to π we generate some interesting relations. As seen in Equations 18, the result of this operation is the delta, the conditional interaction information,

$$\hat{m}\pi(X_n|\nu_{n-1}) = \hat{m}\delta H(\nu_n) = \hat{M}H(X_n) = -I(\nu_{n-1}|X_n) = \Delta(\nu_{n-1}; X_n) \quad (24)$$

Expressing this in another way, using the group table, we have the expressions from Equation 19, and therefore

$$\pi(X_n|\nu_{n-1}) = -\langle \ln \Pr(X_n|\nu_{n-1}) \rangle = \delta H(\nu_n) = \hat{R}H(X_n) \quad (25a)$$

The expected value of the log of the probability of a given, single variable, conditioned on the other variables in the subset, can therefore be expressed simply in terms of Möbius operators acting on the entropy functions of a lattice. This is the result of this section, embodied in Theorem 3.

Theorem 3. The symmetric delta is the product of all conditional log likelihood functions acted on by the down-set operator:

$$\bar{\Delta}(\nu_n) = \prod_{\text{all choices of } X_n} \Delta(\nu_{n-1}; X_n) = \prod_{\text{all choices of } X_n} \hat{m}\pi(X_n|\nu_{n-1}) \quad (25b)$$

The relation of the π 's to the deltas is clear here, and the subsets of the variables under consideration can then generate a series of conditional log likelihoods (CLL's) for $|\nu_m| = m$, $\{\pi(X_m|\nu_{m-1})\}$ for $m \geq 2$. The simplest approximation for dependencies among variables is realized in the case $m = 2$, where CLL's are

November 7, 2018

approximated by those with a single conditional variable. In this case (using simplified notation)

$$\begin{aligned}\pi(2|1) &= H_{12} - H_1 \\ \pi(3|1) &= H_{13} - H_1\end{aligned}\quad (26a)$$

and we have for the three-variable case

$$\begin{aligned}\Delta(23; 1) &= H_1 - H_{12} - H_{13} + H_{123} = -\pi(2|1) + \pi(2|13) \\ \Delta(23; 1) &= -\pi(3|1) + \pi(3|12)\end{aligned}\quad (26b)$$

There are two different ways to express deltas as sums of the π 's. Several conclusions follow from these considerations. Since the group table for the Möbius operators exhibits several different, equivalent operators, $\hat{R} = \hat{m}\hat{M} = \hat{X}\hat{m} = \hat{M}\hat{X} = \hat{P}^2$, we can express the correspondence between Δ and the CLL's in several equivalent ways. These expressions then provide direct links with other information functions.

6b. Towards prediction

An approach to extracting relations predictive of a variable from the information in a data set is suggested by the above considerations. The general problem can be defined as how to determine the “best” prediction formula for the value of one variable in the set, say X_1 , from analysis of a data set of all variables. We sketch the suggested approach here. Step one in a high level description of the process, is to define the maximum degree of dependence to consider (the number of variables involved.) Step two is to calculate the symmetric deltas to determine the set of variables that are dependent on one another [9]. Step three is to find the maximum expected CLL, from the set $\{\pi(X_1|X_i), \pi(X_1|X_i, X_j), \pi(X_1|X_i, X_j, X_k) \dots\}$ by calculating the expectations of the entropy differentials. Note that the specific, expected entropy differences tend to zero as the dependence of the single variable, X_1 , on the other variables increases. Finally, once the “best” likelihood function is found, a predictive function is estimated based on the data: an estimate of the probabilities of X_1 conditioned on all the other variables of the set. The general framework for

November 7, 2018

inference is clear. This procedure is reminiscent of the Chow-Liu algorithm [12] which is entirely pairwise and based on mutual information. Our approach is a direct way towards generating predictive rules from large, multivariable data sets. We will develop this approach further in a future paper.

7. Generalizing the Möbius operators

The up-set and down-set operators, \widehat{M} and \widehat{m} , defined above, generate convolutions over chains from each element of the inclusion lattice to the top element (full set) or to the bottom element (empty set) respectively. The convolutions are either “down”, towards subset elements, or “up” toward supersets. The chains over which the convolutions, sums of the product of function and Möbius function, are taken are clear and are defined by the subset lattice for these two operators. No element is included more than once in the sum. Moreover, the sign of the Möbius function is the same across all elements at the same distance from the extreme elements.

We can generalize the Möbius operators by defining the range of the convolution, the end elements of the paths, to be any pair of elements of the lattice, an upper and lower element, rather than the bounds of the lattice. Two elements are required: the starting element, and an ending element. The starting element may be determined by the argument of the function being operated on. We can call the ending element a *reference element* and associate it with the operator. The specification of both the upper and lower element is essential. Instead of the up-set operator, for example, with the full set ν as its reference element, we could designate an arbitrary subset element like $\{1,2\}$ as the reference and thereby define another operator. Consider now a lattice of the full set ν , where η designates a reference element.

Definition: The generalized Möbius operator F_η , acting on a function of a subset, $f(\tau), \tau \subseteq \nu$, is defined by Equation 27, where the subsets of variables, ς , ranges over

November 7, 2018

all of the shortest paths between τ and η , and the functions $f(\zeta)$ only occur once in the sum, even if they are on more than one path.

$$F_\eta f(\tau) = \sum_{\zeta \text{ on all shortest paths between } \tau \text{ and } \eta} (-1)^{|\nu| - |\zeta|} f(\zeta) \quad (27)$$

There are often multiple shortest paths between any two elements in the lattice, since the subset lattice is a hypercube. In this case, we are specifying the upper and lower elements by the reference and the element specified by the function. The two extreme reference elements, the empty set and the full set, then yield the down-set and up-set operators respectively

$$\begin{aligned} F_0 f &= \widehat{m} f \\ F_\nu f &= \widehat{M} f \end{aligned} \quad (28)$$

The reference element η establishes a relation between the lattice sums and the Möbius function. It is the juxtaposition of the lattice, anchored at η , to the Möbius function that defines the symmetries of the generalized Möbius operator algebra. Note that we now have the possibility of including elements that are not ordered along the paths by inclusion since the reference element can be chosen from any lattice element. For example, the convolution between $\{1\}$ and $\{2,3\}$ for the 3D-cube lattice, shows this clearly (see Figure 1) as it includes $\{1,2\}$, $\{2\}$ and the empty set.

Definition: Given $\mu, \eta \subseteq \nu$ we define the complement generalized operator Möbius operator as $\tilde{F}_\mu \equiv \widehat{X} F_\mu \widehat{X}$.

The products of the generalized operators can easily be calculated for the 3- and 4- element sets. We can identify some similarities of these general operators to the operators \widehat{M} and \widehat{m} . First, we note that the operators, F_μ , are all idempotent. This is easy to calculate for the 3D and 4D case, and to derive using the relations indicated in Equation 27. The idempotent property implies that there are pairs of functions that are related by each general Möbius operator – a generalized Möbius inversion

November 7, 2018

on the inclusion lattice, a generalized duality. Furthermore, the products exhibit other familiar symmetries. The notable relationships that involve a subset and its complement are summarized in the following theorem.

Theorem 4. For all $\mu, \eta \subseteq \nu$ the following properties of the generalized Möbius operator and its complement hold:

$$F_\mu F_\eta = \tilde{F}_\eta \tilde{F}_\mu \quad (29a)$$

$$F_\mu \tilde{F}_\eta = F_\eta \tilde{F}_\mu \quad (29b)$$

$$F_\mu = -\tilde{F}_{\tilde{\mu}} \quad (29c)$$

$$F_\mu f(\tilde{\mu}) = F_\eta f(\tilde{\eta}) \quad (29d)$$

where $\tilde{\mu}$ and $\tilde{\eta}$ are set complements of μ and η correspondingly.

Equation 29a is true since the products of the generalized Möbius operators involve the operator \hat{X} , namely $F_\mu F_\eta = \hat{X} F_\eta F_\mu \hat{X}$, which in the geometric metaphor is like a rotation of the hypercube (inclusion lattice). Applying Equation 29a to $F_\mu \tilde{F}_\eta$ results in Equation 29b. The property shown in Equation 29c follows directly from the definition of F_μ and its complement. The proof of the last property (Equation 29d) is direct. Since the limiting elements of the convolution are a subset and its complement, it encompasses the whole lattice. Thus $F_\gamma f(\tilde{\gamma})$ for any subset γ is seen to describe the convolution over all subsets of the entire lattice and therefore Equation 29d holds.

The full group structure of the general operator algebra is more complex than the group defined by the up-set and down-set operators as there are many more operators, defined by the full range of reference elements. (If N is the number of subsets on the lattice there are $N - 1$ down-set operators, while for the generalized case there are $(N - 1)^2$ operators.) The symmetry of the subgroups determined by pairs of complementary subsets are preserved, remaining isomorphic to S_3 (seen to be true for the 3D, and 4D case by direct calculation, and it appears to be generally true, though we do not yet have a proof of the general case.)

November 7, 2018

The relations between these pairs of functions on the lattice is described by the diagram in Figure 6. It appears that the sets of three functions, specific to a reference set η , with the operators that map one into the other exhibit the same overall symmetries reflected in the group S_3 . The pairs of operators identified with a subset and its complement are the key elements of the group. This is because this particular combination of operator and function defines a convolution over the entire set, ν . This identity therefore includes the specific up-set and down set relations, and is equal to the interaction information if f is the entropy function.

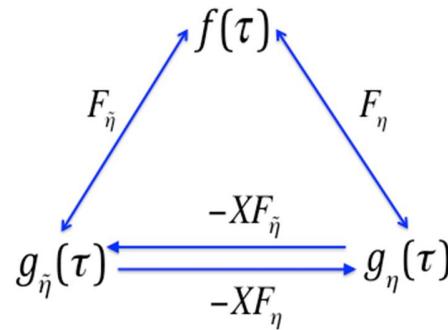


Figure 6. Generalized Möbius operator relations. A diagram of the relations among the functions as determined by the operators. The upper two arrows represent the generalized Möbius inversion relations. The function $g_\eta(\tau)$ is the designation of the function created by the operator F_η . The S_3 structure is reflected in the similarity with the diagram of Figure 3. Note that when $\eta = \emptyset$ the figure becomes identical to Figure 3.

We ask now if sums of such operator-function pairs can be used to decompose a convolution. This decomposition issue can be addressed by asking this specific question: are there sums of operators acting on functions that add up to a given specific operator on another function? If this is possible how do we decompose such convolutions and what do they mean? The simple decomposition of the hypercube into sub-lattices can be shown to be equivalent to the process of finding these convolutions, or operator decompositions. We will not deal with the decomposition relations in a general dimension here, but rather demonstrate them for $\{1,2,3\}$ and $\{1,2,3,4\}$. First, let's consider the 3D case. There are three possible ways to

November 7, 2018

decompose the 3D-cube Hasse diagram into two squares (2D hypercubes), which is done by passing a plane through the cube parallel to the faces (see Figure 7)

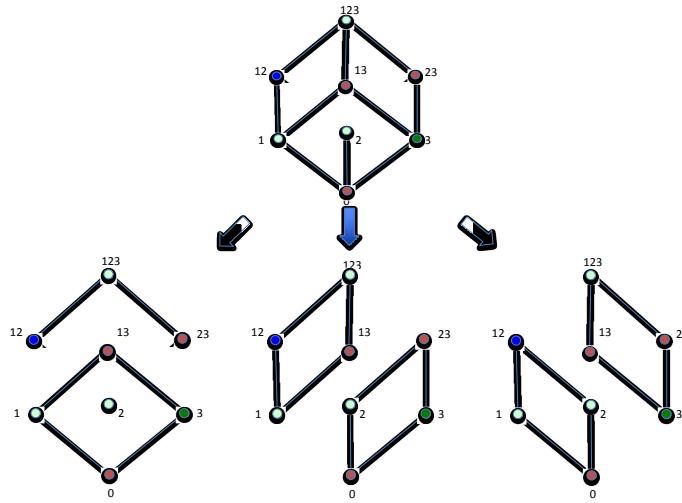


Figure 7. Decomposing the 3D-cube Hasse diagram into two squares (2D hypercubes), by passing a plane through the center of the cube three different ways.

Considering one of these decompositions (the leftmost decomposition in Figure 7) results in the following

$$F_0 f_{123} = F_2 f_{123} + F_0 f_{13} \quad (30)$$

Each of the two terms on the right-hand side could be expressed in operator terms in eight ways (each of the four elements of the sub-lattice being a reference element). There are thus a total of 192 decompositions of the full 3-set convolution, 64 per each of the three decompositions of the cube into two squares. Note that each decomposition leads to the same set of functions, but it is a distinct operator expression.

For the 4-set decomposition, there are four ways of decomposing the 4-hypercube into 2 cubes, so the total number of possible decompositions is $4 \times 192 \times 192 = 147456$.

November 7, 2018

8. Discussion

Many diverse measures have been used in descriptions of order and dependence in complex systems and as data analysis tools [1-9]. While the mathematical properties and relationships among these information-related measures are of significant interest in several fields, there has been, to our knowledge, no systematic examination of the full range of relationships and symmetries, and no unification of this diverse range of functions into a single formalism as we do here. Beginning with the duality relationships, based on Möbius inversions of functions on lattices, we define a set of operators on functions on subset inclusion lattices that map the functions into one another. We show here that they form a simple group, isomorphic to the symmetric group S_3 . A wide range of relationships among the set of functions on the lattice can be expressed simply in terms of this operator algebra formalism. When applied to the information-related measures they can express a wide range of relationships among various measures, providing a unified picture and allowing new ways to calculate one from the other using the subset lattice functions. For example, we can express the conditional mutual information in the 4D, $\{1,2,3,4\}$ lattice as sums of convolutions of entropy functions with few terms for multiple 3D and 2D lattices, or create new information functions with specific symmetries and desired practical properties for a given system. Much is left to explore in the full range of implications of this system, including algorithms for prediction from complex data sets, and other issues about how these functions may be used or computed.

We can make also connections with any other areas where lattices are useful. Since any distributive lattice is isomorphic to the lattice of sets ordered by inclusion, all the results presented here apply to any system of functions defined on a distributive lattice [13,14], so this unification extends beyond the information measure functions. Distributive lattices are widespread and include the following: every Boolean algebra is a distributive lattice; the Lindebaum algebra of most logics that support conjunction and disjunction is a distributive lattice; every Heyting algebra is a distributive lattice, every totally ordered set is a distributive lattice with

November 7, 2018

max as the join and *min* as the meet. The natural numbers also form a distributive lattice with the greatest common divisor as the meet and the least common multiple as the join (this infinite lattice, however, requires some extension of the equivalence proof.)

The relationships shown here unify, clarify, and can serve to guide the use of a range of measures in the development of the theoretical characterization of information and complexity, and in the algorithms and estimation methods needed for the computational analysis of multi-variable data. We have addressed the relationships between the interaction information, the deltas (conditional interaction information), and the underlying probability densities. We find that the deltas can be expressed as Möbius sums of conditional entropies, the multi-information is simply related by the operators to other information functions, and we made an initial connection to the maximum entropy method. We also note that Knuth has proposed generalizations of the zeta and Möbius functions that define *degrees of inclusion* on the lattices [11]. Knuth's formalism, integrated with ours, could lead to a more general set of relations, and add another dimension to this theory by incorporating uncertainty or variance in the information-related measures. This could be particularly useful in developing future methods for complexity descriptions and data analysis. Since the information-related functions have been directly linked to interpretations in algebraic topology [15] it will also be interesting to explore the topological interpretation of the Möbius operators.

From the simple symmetries of these functions and operators it is clear there is more to uncover in this complex of relationships. The information theory-based measures have a surprising richness and internal relatedness in addition to their practical value in data analysis. While we have described here a systematic structure of relationships and symmetries, the full range of possible relationships, insights and applications using Möbius pairs of functions remains to be explored.

November 7, 2018

Acknowledgements: We gratefully acknowledge support by the National Science Foundation, EAGER grant (IIS-1340619), by the Pacific Northwest Research Institute, and the Bill and Melinda Gates Foundation. In the early stages of this work the Luxembourg Centre for Systems Biomedicine (LCSB) of the University of Luxembourg also provided significant partial support.

References

1. McGill, W. J. "Multivariate information transmission". *Psychometrika*, 1954, 19(2):97–116, DOI:10.1109/TIT.1954.1057469.
2. Bell, A. J. "The co-information lattice". Proceedings of 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), Nara, Japan, April 2003.
3. Jakulin, A.; Bratko, I. "Quantifying and Visualizing Attribute Interactions: An Approach Based on Entropy". <http://arXiv:cs.AI/0308002>, v3, 2004.
4. Hu Kuo Tin. "On the Amount of Information". *Theory Prob. Appl.*, 1962, 7(4), 439-447, DOI:10.1137/1107041.
5. Han, T.S. "Multiple Mutual Information and multiple interactions in frequency data". *Information and Control*, 1980, 46:26-45, DOI:10.1016/S0019-9958(80)90478-7.
6. Galas, D. J.; Sakhnenko, N. A.; Ignac, T.; Skupin, A. "Describing the complexity of systems: multi-variable "set complexity" and the information basis of systems biology". *J Comput Biol.* 2014, Feb;21(2):118-40, DOI:10.1089/cmb.2013.0039. Epub 2013 Dec 30.
7. Galas, D.J.; M. Nykter; G. W. Carter; N. D. Price; I. Shmulevich. "Biological information as set based complexity," *IEEE Transactions on Information Theory*, 2010, 56(2):667 – 677. DOI:10.1109/TIT.2009.2037046.

November 7, 2018

8. Carter, G.W.; Galitski, T.; Galas, D.J. "Maximal Extraction of Biological Information from Genetic Interaction Data", *PLoS Comput. Biol.*, 2009, 5(4):e1000347, DOI: 10.1371/journal.pcbi.1000347. Epub 2009 Apr 3.
9. Sakhnenko, N.A.; Galas, D.J. "Biological data analysis as an information theory problem: Multivariable dependence measures and the Shadows algorithm". *J. Comput. Biol.* 2015, 22(11):1005–24, DOI:10.1089/cmb.2015.0051.
10. Watanabe, S. "Information theoretic analysis of multivariate correlation". *IBM Journal of Research and Development*, 1960, 4(1):66-82, DOI:10.1147/rd.41.0066.
11. Knuth, K. E. "Valuations on Lattices and their Application to Information Theory". Proceedings of 2006 IEEE World Congress on Computational Intelligence, Vancouver, Canada, 2006.
12. Chow, C. K.; Liu, C.N. "Approximating discrete probability distributions with dependence trees". *IEEE Transactions on Information Theory*, 1968, 14(3): 462–467, DOI:10.1109/TIT.1968.1054142.
13. Davey, B.A.; Priestley, H. A. *Introduction to Lattices and Order*, 3rd ed. Cambridge University Press, Cambridge, 2002; ISBN 978-0-521-78451-1.
14. Priestley, H. A. (1970). "Representation of distributive lattices by means of ordered Stone spaces". *Bull. London Math. Soc.*, 1970, 2(2):186-190. DOI:10.1112/blms/2.2.186.
15. Baudot, P.; Bennequin, D. "The Homological Nature of Entropy". *Entropy*, 2015, 17(5):3253-3318. DOI:10.3390/e17053253.
16. Singer, A. "Maximum entropy formulation of the Kirkwood superposition approximation". *J Chem Phys* 2004, 121(8), 3657-3666, doi: <http://dx.doi.org/10.1063/1.1776552>.
17. Cochran, R.V. and Lund, L.H.. "On the Kirkwood Superposition Approximation". *J Chem Phys* 1964, 41,3499-3504, doi: <http://dx.doi.org/10.1063/1.1725757>.