Metric-Entropy Pairs on Lattices

Dan Simovici
University of Massachusetts Boston,
Department of Computer Science, Boston,
Massachusetts 02125, USA,
e-mail: dsim@cs.umb.edu

Abstract: We introduce the notion of $\wedge$- and $\vee$-pairs of functions on lattices as an abstraction of the notions of metric and its related entropy for probability distributions. This approach allows us to highlight the relationships that exist between various properties of metrics and entropies and opens the possibility of extending these concepts to other algebraic structures.
Key Words: entropy, modular lattice, metric
Category: H.1.1, H.2.8

1 Introduction

The notion of entropy plays an important role in statistical physics and is a corner stone of information theory. More recently, several applications of entropy in data mining [10, 12], study of biodiversity [4], circuit design [2, 6, 3, 8], etc. have been investigated. A variety of axiomatizations of the notion of entropy have been developed, including axiomatizations that have an algebraic flavor [7, 9, 11].

In this paper we undertake a study of the relationships that exists between entropy and entropy-like concepts, their associated metrics and conditional entropies. The chosen framework is lattice theory, where we show that these notions can be naturally placed.

A lattice is defined as a partially ordered set $(P, \leq)$ such that $\sup\{x, y\}$ and $\inf\{x, y\}$ exist for all $x, y \in P$. It is well known that lattices can be regarded as algebras of the form $(P, \wedge, \vee)$, where "$\wedge$" and "$\vee$" are commutative, associative and idempotent operations linked by the absorption laws

$$x \vee (x \wedge y) = x \text{ and } x \wedge (x \vee y) = x,$$

for $x, y \in P$. The partial order relation "$\leq$" consists of those pairs $(x, y) \in P^2$ such that $x = x \wedge y$ or, equivalently, $y = x \vee y$.

If a least element of the partial ordered set $(P, \leq)$ exists we denoted it by 0; the largest element of $(P, \leq)$ is denoted by 1. If a lattice $(P, \leq)$ has both a least and a largest element we denote it as an algebra by $(P, \wedge, \vee, 0, 1)$, where we regard 0 and 1 as zero-ary operations.

A *partition of a set* S is a non-empty collection of non-empty subsets of S, $\pi = \{B_i \mid i \in I\}$ such that $\bigcup \pi = S$ and $B_i \cup B_j = \emptyset$ when $i \neq j$ for $i, j \in I$. The sets B_i are the *blocks* of π. The set of partitions of S is denoted by $\mathsf{PART}(S)$.

A partial order relation on $\mathsf{PART}(S)$ is defined by $\pi \leq \sigma$ for $\pi, \sigma \in \mathsf{PART}(S)$ if every block of B is included in a block of σ. This is easily seen to be equivalent to requiring that each block of σ is a union of blocks of π.

The partially ordered set $(\mathsf{PART}(S), \leq)$ is actually a bounded lattice. The infimum $\pi \wedge \pi'$ of two partitions π and π' is the partition that consists of non-empty intersections of blocks of π and π'. For a description of the supremum $\pi \vee \pi'$ of the partitions π, π' see [5], p. 251. The least element of this lattice is the partition $\alpha_S = \{\{s\} \mid s \in S\}$; the largest is the partition $\omega_S = \{S\}$.

The partition σ *covers* the partition π if σ is obtained from π by fusing two blocks of this partition. This is denoted by $\pi \prec \sigma$. We have $\pi \leq \pi'$, if and only if there exists a sequence of partitions $\sigma_0, \sigma_1, \ldots, \sigma_r$ such that $\pi = \sigma_0 \prec \sigma_1 \prec \cdots \prec \sigma_r = \pi'$.

Let C be a subset of the set S and let $\pi = \{B_i \mid i \in I\} \in \mathsf{PART}(S)$ be a partition. The *trace* of π on C is the partition $\pi_C = \{B_i \cap C \mid B_i \cap C \neq \emptyset \text{ and } i \in I\}$.

We introduce and study properties of pairs of functions (d, η) defined on lattices that formalize, in a general background, metrics and entropy-like functions defined on sets of partitions, which we investigated previously [11, 10]. This study illuminates the links that exist between various metric properties (non-negativity, definedness, triangular inequality) and monotonicity or modularity properties of entropy or conditional entropy.

2 Function Pairs on Lattices

Let $\mathcal{L} = (L, \vee, \wedge, 0, 1)$ be a lattice that has the least element 0 and the largest element 1.

Definition 1. Let $d : L^2 \longrightarrow \mathbb{R}$, $\eta : L \longrightarrow \mathbb{R}$ be two functions defined on the lattice $\mathcal{L}$.

The pair (d, η) is a $\wedge$-pair on $\mathcal{L}$ if

$$d(x, y) = 2\eta(x \wedge y) - \eta(x) - \eta(y) \tag{1}$$

for every $x, y \in L$.

The pair (d, η) is a $\vee$-*pair* if

$$d(x, y) = \eta(x) + \eta(y) - 2\eta(x \vee y) \tag{2}$$

for every $x, y \in L$.

If (d, η) is both a $\wedge$-pair and a $\vee$-pair, then we refer to (d, η) as a *double pair*.

Note that for any $\wedge$-pair or $\vee$-pair the function d is symmetric.

If (d, η) is a $\wedge$-pair (a $\vee$-pair) on the lattice L, then $(d, k + \eta)$ is also an $\wedge$-pair (a $\vee$-pair) on the same lattice for any number $k \in \mathbb{R}$. This shows that the first component does not determine the second component of an $\wedge$-pair or a $\vee$-pair.

Also, observe that for an $\wedge$-pair or a $\vee$-pair we have $\eta(x) = \eta(1) + d(x,1)$ for every $x \in L$; thus, the first component and the value $\eta(1)$ determines the second component of an $\wedge$-pair or a $\vee$-pair.

If $\eta(1) = 0$, then we say that the pair (d, η) is *regular*. In a regular $\wedge$-pair each component determines the other.

The function $\eta : L \longrightarrow \mathbb{R}$ is said to be *strictly anti-monotonic* if $u < v$ implies $\eta(u) > \eta(v)$.

The collection of pairs introduced here formalize in the realm of lattices the notions of partition entropy (η) and metric generated by an entropy (d).

Theorem 2. *Let (d, η) be an $\wedge$-pair or a $\vee$-pair on the lattice $\mathcal{L} = (L, \vee, \wedge, 0, 1)$. We have:*

1. *If $x \leq y$, then $d(x, y) = \eta(x) - \eta(y)$.*
2. *If $x \leq t \leq y$, then $d(x, t) + d(t, y) = d(x, y)$.*
3. *If (d, η) is an $\wedge$-pair, then*

$$d(x, x \wedge y) + d(x \wedge y, y) = d(x, y)$$

and

$$\begin{aligned} d(x, y) &= 2 \cdot d(x \wedge y, 1) - d(x, 1) - d(y, 1) \\ &= d(x, 0) + d(y, 0) - 2 \cdot d(x \wedge y, 0). \end{aligned}$$

4. *If (d, η) is a $\vee$-pair, then*

$$d(x, x \vee y) + d(x \vee y, y) = d(x, y)$$

and

$$\begin{aligned} d(x, y) &= 2 \cdot d(x \vee y, 0) - d(x, 0) - d(y, 0) \\ &= d(x, 1) + d(y, 1) - 2 \cdot d(x \vee y, 1). \end{aligned}$$

Proof. The properties mentioned in the theorem are direct consequences of the definition of $\wedge$-pairs. □

The purpose of the next two theorems is to show that defining properties of a metric can be expressed in terms of properties of the function η.

Theorem 3. *Let (d, η) be a $\wedge$-pair or a $\vee$-pair on the lattice $\mathcal{L} = (L, \vee, \wedge, 0, 1)$. We have $d(x, y) \geq 0$ for $x, y \in L$ if and only if η is an anti-monotonic function. Furthermore, $d(x, y) = 0$ implies $x = y$ if and only if η is a strictly anti-monotonic function.*

Proof. Suppose that $d(x, y) \geq 0$ for every $x, y \in L$. By the first part of Theorem 2, $x \leq y$ implies $\eta(x) \geq \eta(y)$.

Conversely, suppose that η is anti-monotonic. If (d, η) is a $\wedge$-pair, then $\eta(x \wedge y) \geq \eta(x), \eta(y)$, so $d(x, y) \geq 0$. If d is a $\vee$-pair the same conclusion can be reached, by observing that $\eta(x), \eta(y) \geq \eta(x \vee y)$.

Suppose that $d(x, y) = 0$, where (d, η) is a $\wedge$-pair, where η is an anti-monotonic function. Then, $2\eta(x \wedge y) - \eta(x) - \eta(y) = 0$, so $\eta(x) = \eta(y) = \eta(x \wedge y)$ because $\eta(x) \leq \eta(x \wedge y)$ and $\eta(y) \leq \eta(x \wedge y)$. Suppose that $x \neq y$. Then, at least one of the strict inequalities $x \wedge y < x$ or $x \wedge y < y$ holds. Since this yields a contradiction it follows that $x = y$.

The argument for $\vee$-pairs is similar. □

Example 1. Let S be a finite set and let $(\mathsf{PART}(S), \leq)$ be the partition lattice having α_S as its least element and ω_S as its largest element. For $\beta \in \mathbb{R}$ and $\beta > 1$ define the mapping $\eta_\beta : \mathsf{PART}(S) \longrightarrow \mathbb{R}$ as:

$$\eta_\beta(\pi) = \frac{1}{1 - 2^{1-\beta}} \left(1 - \sum_{i=1}^{m} \left(\frac{|B_i|}{|S|} \right)^\beta \right),$$

where π is the partition $\pi = \{B_1, \ldots, B_m\}$. Observe that $\eta_\beta(\omega_S) = 0$ and $\eta_\beta(\pi) > 0$ for $\pi \in \mathsf{PART}(S) - \{\omega_S\}$.

The function η_β is strictly anti-monotonic. To prove this property it suffices to consider two partitions π, π' such that $\pi \prec \pi'$. Without loss of generality we can assume that $\pi = \{B_1, \ldots, B_{n-2}, B_{n-1}, B_n\}$ and $\pi' = \{B_1, \ldots, B_{n-2}, B_{n-1} \cup B_n\}$.

Note that for $x, y > 0$ and $\beta > 1$ we have $x^\beta + y^\beta < (x + y)^\beta$. Therefore,

$$\left(\frac{|B_{n-1}|}{|S|} \right)^\beta + \left(\frac{|B_n|}{|S|} \right)^\beta < \left(\frac{|B_{n-1} \cup B_n|}{|S|} \right)^\beta,$$

which implies $\eta_\beta(\pi) < \eta_\beta(\pi')$, that is, the strict anti-monotonicity property. By Theorem 3, the function d_β of the $\wedge$-pair (d, η_β) given by $d_\beta(\pi, \sigma) = 2\eta_\beta(\pi \wedge \sigma) - \eta_\beta(\pi) - \eta_\beta(\sigma)$ for $\pi, \sigma \in \mathsf{PART}(S)$ is non-negative and $d_\beta(\pi, \sigma) = 0$ implies $\pi = \sigma$.

Example 2. Define the function $\eta_1 : \mathsf{PART}(S) \longrightarrow \mathbb{R}$ by

$$\eta_1(\pi) = -\sum_{i=1}^{m} \frac{|B_i|}{|S|} \log \frac{|B_i|}{|S|},$$

where $\pi = \{B_1, \ldots, B_m\}$ and the logarithm is in base 2.

This is the Shannon entropy of the probability distribution

$$\begin{pmatrix} \frac{|B_1|}{|S|} & \cdots & \frac{|B_m|}{|S|} \end{pmatrix}$$

defined by the partition $\pi \in \mathsf{PART}(S)$. It is easy to verify that $\lim_{\beta \to 1} \eta_\beta(\pi) = \eta_1(\pi)$, which implies that η_1 is anti-monotonic. An elementary argument can be used to verify that η_1 is, in fact, strictly anti-monotonic, so the function $d_1 : \mathsf{PART}(S)^2 \longrightarrow \mathbb{R}$ given by

$$d_1(\pi,\sigma) = \sum_{i=1}^{m} \frac{|B_i|}{|S|} \log \frac{|B_i|}{|S|} + \sum_{j=1}^{n} \frac{|C_j|}{|S|} \log \frac{|C_j|}{|S|} - \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{|B_i \cap C_j|}{|S|} \log \frac{|B_i \cap C_j|}{|S|},$$

where $\pi = \{B_1, \ldots, B_m\}$ and $\sigma = \{C_1, \ldots, C_n\}$ is non-negative and $d_1(\pi,\sigma) = 0$ implies $\pi = \sigma$.

Theorem 4. *The function d of an $\wedge$-pair (d,η) satisfies the triangular axiom, $d(x,y) \leq d(x,z) + d(z,y)$ if and only if*

$$\eta(z) + \eta(x \wedge y) \leq \eta(x \wedge z) + \eta(y \wedge z) \tag{3}$$

for $x, y, z \in L$.

If (d,η) is a $\vee$-pair, then d satisfies the triangular inequality if and only if

$$\eta(z) + \eta(x \vee y) \geq \eta(x \vee z) + \eta(y \vee z) \tag{4}$$

for $x, y, z \in L$.

Proof. Let d be a function that satisfies the triangular inequality. This implies

$$2\eta(x \wedge y) - \eta(x) - \eta(y) \leq 2\eta(x \wedge z) - \eta(x) - \eta(z) + 2\eta(y \wedge z) - \eta(y) - \eta(z),$$

which is easily seen to be equivalent to the Inequality (3). The reverse implication is as straightforward as the direct implication.

A similar straightforward argument can be made for $\vee$-pairs. □

Theorem 5. *Let (d,η) be an $\wedge$-pair on a lattice $\mathcal{L} = (L, \vee, \wedge, 0, 1)$, where d is a non-negative function. Then, d satisfies the triangular inequality if and only if η is an anti-monotonic and sub-modular, that is, $\eta(x \vee y) + \eta(x \wedge y) \leq \eta(x) + \eta(y)$ for every $x, y \in L$.*

If (d,η) be a $\vee$-pair on $\mathcal{L}$, where d is a non-negative function. The function d satisfies the triangular inequality if and only if η is an anti-monotonic and supra-modular, that is, $\eta(x \vee y) + \eta(x \wedge y) \geq \eta(x) + \eta(y)$ for every $x, y \in L$.

Proof. Suppose that d is a non-negative function of a $\wedge$-pair (d,η) that satisfies the triangular inequality. Then, by Theorems 3 and 4, η is an anti-monotonic function, and $\eta(z) + \eta(x \wedge y) \leq \eta(x \wedge z) + \eta(y \wedge z)$ for every $x, y, z \in L$. By replacing z by $x \vee y$ and using the absorption properties of $\mathcal{L}$ we obtain the sub-modular inequality $\eta(x \vee y) + \eta(x \wedge y) \leq \eta(x) + \eta(y)$.

If d is a non-negative function of a $\vee$-pair (d,η) that satisfies the triangular inequality, then η is an anti-monotonic function and $\eta(z) + \eta(x \vee y) \geq \eta(x \vee z) + \eta(y \vee z)$

for $x, y, z \in L$. Substituting $x \wedge y$ for z and applying the absorption properties we have the supra-modular inequality $\eta(x \vee y) + \eta(x \wedge y) \geq \eta(x) + \eta(y)$.

Conversely, suppose that η is an anti-monotonic, sub-modular function of a $\wedge$-pair (d, η). The anti-monotonicity of η implies the non-negativity of d. We need to show that the sub-modular inequality implies Inequality (3).

Observe that in every lattice $\mathcal{L} = (L, \vee, \wedge, 0, 1)$ we have the sub-distributive inequality

$$(x \vee y) \wedge z \geq (x \wedge z) \vee (y \wedge z), \tag{5}$$

for every $x, y, z \in L$. By substituting $x \wedge z$ for x and $y \wedge z$ for y in the sub-modular inequality we obtain:

$$\eta((x \wedge z) \vee (y \wedge z)) + \eta(x \wedge y \wedge z) \leq \eta(x \wedge z) + \eta(y \wedge z).$$

In view of Inequality (5) and of the anti-monotonicity of η we can write

$$\eta((x \vee y) \wedge z) \leq \eta((x \wedge z) \vee (y \wedge z)),$$

and, since $z \geq (x \vee y) \wedge z$ we have

$$\eta(z) \leq \eta((x \wedge z) \vee (y \wedge z)).$$

Since $\eta(x \wedge y) \leq \eta(x \wedge y \wedge z)$, we obtain the Inequality (3).

Let now η be an anti-monotonic, supra-modular function of a $\vee$-pair (d, η).

By replacing $x \vee z$ for x and $y \vee z$ for y in the supra-modular inequality we have:

$$\eta(x \vee y \vee z) + \eta((x \vee z) \wedge (y \vee z)) \geq \eta(x \vee z) + \eta(y \vee z)$$

for every $x, y, z \in L$. Starting from the inequality

$$(x \wedge y) \vee z \leq (x \vee z) \wedge (y \vee z), \tag{6}$$

that holds in every lattice we obtain

$$\eta(x \vee y \vee z) + \eta((x \wedge y) \vee z) \geq \eta(x \vee z) + \eta(y \vee z)$$

for every $x, y, z \in L$. Finally, since $z \leq (x \wedge y) \vee z$ and $x \vee y \vee z \geq x \vee y$ we get the Inequality 4. □

We retrieve a well-known property of modular lattices (cf. [1]):

Corollary 6. *If there exists a double pair (d, η) on a lattice $\mathcal{L} = (L, \vee, \wedge, 0, 1)$ such that d is a metric, then $\mathcal{L}$ is a metric lattice and $d(x, y) = \eta(x \wedge y) - \eta(x \vee y)$ for $x, y \in L$.*

Proof. Since (d, η) is a double pair and d is a metric the strictly anti-monotonic function η satisfies both the sub-modular and the supra-modular inequalities and therefore we have $\eta(x \wedge y) + \eta(x \vee y) = \eta(x) + \eta(y)$, so η is a modular valuation on L. This implies

$$d(x, y) = \eta(x \wedge y) - 2\eta(x \vee y),$$

for $x, y \in L$. □

3 Conditional Function of a Pair

Starting from an $\wedge$-pair (d, η), define the *conditional function* $\kappa : L^2 \longrightarrow \mathbb{R}$ of the pair (d, η) by $\kappa(x, y) = \eta(x \wedge y) - \eta(y)$ for $x, y \in L$. It is immediate that $d(x, y) = \kappa(x, y) + \kappa(y, x)$ and that $x \geq y$ implies $\kappa(x, y) = 0$ for $x, y \in L$. If the pair (d, η) is regular, then $\eta(x) = \kappa(x, 1)$.

The conditional function of a pair (d, η) formalizes the notion of conditional entropy corresponding to the entropy η.

Example 3. The conditional function of the $\wedge$-pair (d_β, η_β) introduced in Example 1 is given by

$$\begin{aligned}\kappa_\beta(\pi, \sigma) &= \eta_\beta(\pi \wedge \sigma) - \eta_\beta(\sigma) \\ &= \frac{1}{1 - 2^{1-\beta}} \cdot \left(\sum_{j=1}^{n} \left(\frac{|C_j|}{|S|} \right)^\beta - \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\frac{|B_i \cup C_j|}{|S|} \right)^\beta \right),\end{aligned}$$

where $\pi = \{B_1, \ldots, B_m\}$ and $\sigma = \{C_1, \ldots, C_n\}$ are two partitions of $\mathsf{PART}(S)$.

This function can be written alternatively as

$$\begin{aligned}\kappa_\beta(\pi, \sigma) &= \frac{1}{1 - 2^{1-\beta}} \cdot \sum_{j=1}^{n} \left(\frac{|C_j|}{|S|} \right)^\beta \left(1 - \sum_{i=1}^{m} \left(\frac{|B_i \cup C_j|}{|C_j|} \right)^\beta \right) \\ &= \frac{1}{1 - 2^{1-\beta}} \cdot \sum_{j=1}^{n} \left(\frac{|C_j|}{|S|} \right)^\beta \eta_\beta(\pi_{C_j}),\end{aligned}$$

where π_{C_j} is the trace of π on the block C_j of σ.

Theorem 7. *Let (d, η) be an $\wedge$-pair on $\mathcal{L}$. The non-negative function d satisfies the triangular inequality if and only if the conditional function κ is anti-monotonic in its first argument and monotonic in its second argument.*

Proof. Suppose that d satisfies the triangular inequality. The anti-monotonicity of κ in its first argument follows from the first part of Theorem 2. Let $y, y_1 \in L$ be such that $y \leq y_1$. It is clear that $(x \wedge y_1) \vee y \leq y_1$, so

$$\eta(y_1) \leq \eta((x \wedge y_1) \vee y). \tag{7}$$

By Theorem 5 we have the sub-modular inequality $\eta(x \vee y) + \eta(x \wedge y) \leq \eta(x) + \eta(y)$ for every $x, y \in L$. Taking into account Inequality 7 and replacing x by $x \wedge y_1$ in the sub-modular inequality yields

$$\begin{aligned}\eta(x \wedge y) + \eta(y_1) &\leq \eta(x \wedge y) + \eta((x \wedge y_1) \vee y) \\ &= \eta((x \wedge y_1) \wedge y) + \eta((x \wedge y_1) \vee y) \\ &\leq \eta(x \wedge y_1) + \eta(y) \\ &\quad \text{(by the sub-modular inequality).}\end{aligned}$$

The last equality implies $\kappa(x, y) \leq \kappa(x, y_1)$, that is, the monotonicity of κ in its second argument.

Conversely, suppose that κ is anti-monotonic in its first argument and monotonic in its second argument. Since $\kappa(1) = 0$ it follows that $\kappa(x, y) = \eta(x \wedge y) - \eta(y) \geq 0$. Similarly, $\eta(x \wedge y) - \eta(x) \geq 0$, so $d(x, y) \geq 0$.

If $y \leq y_1$, we have $\eta(x \wedge y) - \eta(y) \leq \eta(x \wedge y_1) - \eta(y_1)$. Choosing $y_1 = x \vee y$ we obtain the sub-modular inequality for η, which shows that d satisfies the triangular inequality by Theorem 5. □

In a similar manner one can define the conditional function of a $\vee$-pair by $\kappa(x, y) = \eta(x) - \eta(x \vee y)$. This time, we can prove the following statement:

Theorem 8. *Let (d, η) be an $\vee$-pair on $\mathcal{L}$. The non-negative function d satisfies the triangular inequality if and only if the conditional function κ is monotonic in its first argument and anti-monotonic in its second argument.*

Proof. The proof is analogous to the argument of Theorem 7. □

4 $\wedge$-Pairs on Partition Lattices

For partition lattices of finite sets the $\wedge$-pairs play a special role because they allow us to formalize the notion of entropy for a partition of a finite set and to introduce simultaneously a notion of metric on the partition lattice that has many applications in data mining and in other areas.

Lemma 9. *Let S be a finite set, $\pi \in \mathsf{PART}(S)$ and let C, D be two disjoint subsets of S. For $\beta \geq 1$ we have:*

$$\left(\frac{|C \cup D|}{|S|}\right)^\beta \eta_\beta(\pi_{C \cup D}) \geq \left(\frac{|C|}{|S|}\right)^\beta \eta_\beta(\pi_C) + \left(\frac{|D|}{|S|}\right)^\beta \eta_\beta(\pi_D),$$

where $\eta_\beta : \mathsf{PART}(S) \longrightarrow \mathbb{R}$ is the function introduced in Example 1.

Proof. Suppose that $\pi = \{B_1, \ldots, B_m\}$ is a partition of S. Define the numbers

$$w_i = \frac{|B_i \cap (C \cup D)|}{|C \cup D|}$$

for $1 \leq i \leq m$. It is clear that $\sum_{i=1}^m w_i = 1$. Let

$$a_i = \frac{|B_i \cap C|}{|B_i \cap (C \cup D)|},$$

for $1 \leq i \leq$. It is immediate that $1 - a_i = \frac{|B_i \cap D|}{|B_i \cap (C \cup D)|}$.

Applying Lemma 12 to the numbers $w_1, \ldots, w_m$ and $a_1, \ldots, a_m$ we obtain:

$$1 - \left(\sum_{i=1}^{n} \frac{|B_i \cap C|}{|C \cup D|}\right)^{\beta} - \left(\sum_{i=1}^{n} \frac{|B_i \cap D|}{|C \cup D|}\right)^{\beta}$$
$$\geq \sum_{i=1}^{n} \left(\frac{|B_i \cap (C \cup D)|}{|C \cup D|}\right)^{\beta} \left(1 - \left(\frac{|B_i \cap C|}{|B_i \cap (C \cup D)|}\right)^{\beta} - \left(\frac{|B_i \cap D|}{|B_i \cap (C \cup D)|}\right)^{\beta}\right).$$

Since

$$\sum_{i=1}^{n} \frac{|B_i \cap C|}{|C \cup D|} = \frac{|C|}{|C \cup D|} \text{ and } \sum_{i=1}^{n} \frac{|B_i \cap D|}{|C \cup D|} = \frac{|D|}{|C \cup D|},$$

the last inequality can be written:

$$1 - \left(\frac{|C|}{|C \cup D|}\right)^{\beta} - \left(\frac{|D|}{|C \cup D|}\right)^{\beta}$$
$$\geq \sum_{i=1}^{n} \left(\frac{|B_i \cap (C \cup D)|}{|C \cup D|}\right)^{\beta} - \sum_{i=1}^{n} \left(\frac{|B_i \cap C|}{|C \cup D|}\right)^{\beta} - \sum_{i=1}^{n} \left(\frac{|B_i \cap D|}{|C \cup D|}\right)^{\beta},$$

which is equivalent to

$$1 - \sum_{i=1}^{n} \left(\frac{|B_i \cap (C \cup D)|}{|C \cup D|}\right)^{\beta} \geq \left(\frac{|C|}{|C \cup D|}\right)^{\beta} \left(1 - \sum_{i=1}^{n} \left(\frac{|B_i \cap C|}{|C|}\right)^{\beta}\right)$$
$$+ \left(\frac{|D|}{|C \cup D|}\right)^{\beta} \left(1 - \sum_{i=1}^{n} \left(\frac{|B_i \cap D|}{|D|}\right)^{\beta}\right),$$

which yields the inequality of the lemma. □

The next result shows that $\kappa_\beta(\pi, \sigma)$, the conditional function of the $\wedge$-pair (d_β, η_β) is anti-monotonic with respect to its first argument and is monotonic with respect to its second argument.

Theorem 10. *Let $\pi, \sigma, \sigma' \in \mathsf{PART}(S)$, where S is a finite set. If $\sigma \leq \sigma'$, then $\kappa_\beta(\sigma, \pi) \geq \kappa_\beta(\sigma', \pi)$ and $\kappa_\beta(\pi, \sigma) \leq \kappa_\beta(\pi, \sigma')$.*

Proof. Since $\sigma \leq \sigma'$ we have $\pi \wedge \sigma \leq \pi \wedge \sigma'$, so $\eta_\beta(\pi \wedge \sigma) \geq \eta_\beta(\pi \wedge \sigma')$. Therefore, $\kappa_\beta(\sigma, \pi) \geq \kappa_\beta(\sigma', \pi)$.

For the monotonicity of κ_β in its second argument it suffices to prove the monotonicity for partitions σ, σ' such that $\sigma \prec \sigma'$. Without restricting the generality we may assume that $\sigma = \{C_1, \ldots, C_{n-2}, C_{n-1}, C_n\}$ and $\sigma' = \{C_1, \ldots, C_{n-2}, C_{n-1} \cup C_n\}$.

Thus, we can write:

$$
\begin{aligned}
&\kappa_\beta(\pi, \sigma') \\
&= \sum_{j=1}^{n-2} \left(\frac{|C_j|}{|S|}\right)^\beta \eta_\beta(\pi_{C_j}) + \left(\frac{|C_{n-1} \cup C_n|}{|S|}\right)^\beta \kappa_\beta(\pi_{C_{n-1} \cup C_n}) \\
&\geq \left(\frac{|C_j|}{|S|}\right)^\beta \eta_\beta(\pi_{C_j}) + \left(\frac{|C_{n-1}|}{|S|}\right)^\beta \eta_\beta(\pi_{C_{n-1}}) + \left(\frac{|C_n|}{|S|}\right)^\beta \eta_\beta(\pi_{C_n}) \\
&\quad \text{(by Lemma 9)} \\
&= \kappa_\beta(\pi, \sigma).
\end{aligned}
$$

□

Corollary 11. *Let* (d, η_β) *be the* $\wedge$*-pair on the lattice* $(\mathsf{PART}(S), \leq)$*, where* η_β *is the function introduced in Example 1. Then* d_β *is a metric on the lattice of partitions* $(\mathsf{PART}(S), \leq)$.

Proof. This statement follows from Theorems 3, 7, and 10. □

The function η_β is actually the entropy $\mathcal{H}_\beta$ that we axiomatized in [11] and d_β is its associated distance.

5 Function Pairs on Graded Lattices

A graded poset (cf. [1]) is a triple $(P, \leq, g)$, where $(P, \leq)$ is a partially ordered set, and $g : P \longrightarrow \mathbb{Z}$ is a function defined on P such that for $x, y \in L$ we have

(i) $x < y$ implies $g(x) < g(y)$ (strict monotonicity);

(ii) if y covers x, then $g(y) = g(x) + 1$.

If $(P, \leq)$ is a lattice, then we refer to $(P, \leq, g)$ as a *graded lattice*.

In a graded poset all maximal chains between the same elements have the same finite length (the Jordan-Dedekind condition).

Let $(P, \leq)$ be a poset that has the least element 0. The supremum of the lengths of all chains that join 0 to an element x is the *height* of x denoted by $\mathsf{height}(x)$. If $(P, \leq)$ has the largest element 1, then the height of $(P, \leq)$ is defined as $\mathsf{height}(1)$. A poset $(P, \leq)$ satisfies the Jordan-Dedekind condition if and only if it is graded by the function height.

It is known that a graded lattice of finite height is upper semimodular if height satisfies the sub-modular inequality and is lower semimodular if height satisfies the supra-modular inequality (cf. Theorem II.15, p. 40 of [1]).

The function $\eta : P \longrightarrow \mathbb{R}$ defined by $\eta(x) = \mathsf{height}(P) - \mathsf{height}(x)$ satisfies the supra-modular inequality and the associated function d in the $\wedge$-pair (d, η) satisfies the triangular inequality and, therefore, it is a pseudometric on the lattice L given by $d(x, y) = h(x) + h(y) - 2h(x \wedge y)$ for $x, y \in L$.

6 Conclusions

We present an lattice-theoretical framework for the study of entropy and entropy-like functions and the metrics and conditional entropies that can be associated to these entropies. This approach clarifies the dependencies that exist between properties of these concepts and opens the possibility of extending this study to broader classes of lattices, Boolean algebras, and partially ordered sets.

References

1. G. Birkhoff. *Lattice Theory*. American Mathematical Society, Providence, R. I., third edition, 1973.
2. K.T. Cheng and V. Agrawal. An entropy measure of the complexity of multi-output boolean function. In *Proc. of the 27th Design Automation Conference*, pages 302–305, 1990.
3. V. Cheushev, V. Shmerko, D. A. Simovici, and S. Yanushkevich. Functional entropy and decision trees. In *Proceedings of ISMVL 1998*, pages 257–262, Fukuoka, Japan, 1998.
4. K. S. Fritsch and J.C.Hsu. Multiple comparison of entropies with application to dinosaur biodiversity. *Biometrics*, 55:1300–1305, 1999.
5. G. Grätzer. *General Lattice Theory*. Birkhäuser, Basel, second edition, 2003.
6. C. H. Hwang and A. C. H. Wu. An entropy measure for power estimation of boolean function. In *Proc. of Asia and South Pacific Design Automation Conference (ASP-DAC)*, pages 101–106, 1997.
7. R. S. Ingarden and K. Urbanik. Information without probability. *Coll. Math.*, 1:281–304, 1962.
8. A. Lloris-Ruiz, J. F. Gomez-Lopera, and R. Roman-Roldan. Entropic minimization of multiple-valued logic functions. In *Proceedings of the 23rd ISMVL*, pages 24–28, 1993.
9. A. M. Mathai and P. N. Rathie. *Basic Concepts in Information Theory and Statistics — Axiomatic Foundations and Applications*. Halsted Press, John Wiley & Sons, New York, 1975.
10. D. A. Simovici, D. Cristofor, and L. Cristofor. Generalized entropy and projection clustering of categorial data. In *Proceedings of PKDD (Principles of Data Mining and Knowledge Discovery) the 4th European Conference, PKDD 2000*, pages 619–625, Lyon, France, 2000.
11. D. A. Simovici and S. Jaroszewicz. An axiomatization of partition entropy. *Transactions on Information Theory*, 48:2138–2142, 2002.
12. D. A. Simovici and S. Jaroszewicz. Generalized entropy and decision trees. In *EGC 2003 - Journees francophones d'Extraction et de Gestion de Connaissances*, pages 369–380, Lyon, France, 2003.

A A Technical Result

Lemma 12. *Let $\beta \geq 1$. If $w_1, \ldots, w_n$ are n positive numbers such that $\sum_{k=1}^n w_k = 1$, and $a_1, \ldots, a_n \in [0,1]$, then*

$$1 - \left(\sum_{i=1}^n w_i a_i\right)^\beta - \left(\sum_{i=1}^n w_i (1-a_i)\right)^\beta \geq \sum_{i=1}^n w_i^\beta \left(1 - a_i^\beta - (1-a_i)^\beta\right).$$

Proof. Let $\phi : [0,1] \longrightarrow \mathbb{R}$ be the function given by: $\phi(x) = x^\beta + (1-x)^\beta$ for $x \in [0,1]$. It is easy to see that $\phi(0) = \phi(1) = 1$ and that ϕ has a minimum for $x = 1/2$, $\phi(1/2) = 1/2^{1-\beta}$. Thus, we have:

$$x^\beta + (1-x)^\beta \le 1 \tag{8}$$

for $x \in [0,1]$.

Inequality (8) implies

$$w_i(1 - a_i^\beta - (1-a_i)^\beta) \ge w_i^\beta(1 - a_i^\beta - (1-a_i)^\beta),$$

because $w_i \in [0,1]$ and $\beta \ge 1$.

By applying Jensen's inequality for the convex function $f(x) = x^\beta$ we obtain the inequalities:

$$\left(\sum_{i=1}^n w_i a_i\right)^\beta \le \sum_{i=1}^n w_i a_i^\beta,$$
$$\left(\sum_{i=1}^n w_i(1-a_i)\right)^\beta \le \sum_{i=1}^n w_i(1-a_i)^\beta.$$

Thus, we can write

$$\begin{aligned}
&1 - \left(\sum_{i=1}^n w_i a_i\right)^\beta - \left(\sum_{i=1}^n w_i(1-a_i)\right)^\beta \\
&= \sum_{i=1}^n w_i - \left(\sum_{i=1}^n w_i a_i\right)^\beta - \left(\sum_{i=1}^n w_i(1-a_i)\right)^\beta \\
&\ge \sum_{i=1}^n w_i - \sum_{i=1}^n w_i a_i^\beta - \sum_{i=1}^n w_i(1-a_i)^\beta \\
&= \sum_{i=1}^n w_i \left(1 - a_i^\beta - (1-a_i)^\beta\right) \\
&\ge \sum_{i=1}^n w_i^\beta \left(1 - a_i^\beta - (1-a_i)^\beta\right),
\end{aligned}$$

which is the desired inequality. □