

# Betweenness, Metrics and Entropies in Lattices

Dan A. Simovici  
Univ. of Massachusetts Boston,  
Dept. of Comp. Science,  
100 Morrissey Blvd.  
Boston, Massachusetts  
02125 USA  
dsim@cs.umb.edu

## Abstract

We investigate a class of metrics on lattices that are compatible with the partial order defined by the lattice using the ternary relation of betweenness that can be naturally defined on a metric space. The relationships between entropy-like functions and metrics defined on lattices are studied and we show the links that exists between various properties of entropies and properties of metrics. Applications to metrics defined on the lattice of partitions of finite sets are examined.

## 1 Introduction

The notion of entropy that is fundamental for information theory and statistical mechanics has been applied in a variety of disciplines ranging from circuit design [2, 6, 3, 8], to data mining [10, 12] and biodiversity [4].

A variety of axiomatizations of this notion have been developed, including axiomatizations that have an algebraic flavor [7, 9, 11].

In this paper we examine the interaction between metrics on lattices, their associated betweenness relations, and a family of generalizations of entropy. Lattice theory is the natural framework for this study due to the relationship that exists between partitions of sets and finite discrete probabilistic distributions.

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 “ $\leq$ ” consists of those pairs  $(x, y) \in P^2$  such that  $x = x \wedge y$  or, equivalently,  $y = x \vee y$ .

If the least element of the partially 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 metric on a set  $S$  is a mapping  $d : S \times S \longrightarrow \mathbb{R}$  that satisfies the following conditions:

1.  $d(x, y) = 0$  if and only if  $x = y$ ,
2.  $d(x, y) = d(y, x)$ , and
3.  $d(x, y) \leq d(x, z) + d(z, y)$  (the *triangular inequality*),

for every  $x, y, z \in S$ . The pair  $(S, d)$  will be referred to as a *metric space*.

An element  $u$  of a metric space  $(S, d)$  is *between*  $x$  and  $y$  if  $d(x, y) = d(x, u) + d(u, y)$ . The betweenness ternary relation  $B_d \subseteq S \times S \times S$  is defined by

$$B_d = \{(x, t, y) \in S^3 \mid d(x, y) = d(x, t) + d(t, y)\}.$$

We use this relation to introduce a compatibility requirement for metrics defined on lattices. Then, we show that compatible metrics define entropy-like functions on lattices and we examine also the reverse process that starts from entropy-like functions and generates compatible metrics on lattices. We clarify the connections that exist between various metric properties (non-negativity, definedness, triangular inequality) and monotonicity or modularity properties of entropy or conditional entropy.

Finally, we exemplify our results by a study on a family of entropies on the lattice of partitions of a finite set.

## 2 Compatible Metrics on Lattices

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

We say that a metric  $d : L^2 \longrightarrow \mathbb{R}_{\geq 0}$  is *compatible* with the lattice if  $x \leq u \leq y$  implies  $B_d(x, u, y)$ .

Let  $d$  be a compatible metric on the lattice  $\mathcal{L} = (L, \vee, \wedge, 0, 1)$ . If  $u \leq v$ , we have  $d(u, 1) = d(u, v) + d(v, 1)$  because of  $B_d(u, v, 1)$ . Similarly,  $B(0, u, v)$  implies  $d(0, v) = d(0, u) + d(u, v)$ , so

$$d(u, v) = d(0, v) - d(0, u) = d(u, 1) - d(v, 1)$$

for  $u, v \in L$ .

Define the functions  $\eta_d : L \longrightarrow \mathbb{R}$  and  $\zeta_d : L \longrightarrow \mathbb{R}$  by  $\eta_d(u) = d(u, 1)$  and  $\zeta_d(u) = d(0, u)$ . It is clear that  $\eta_d$  is an anti-monotonic function, while  $\zeta_d$  is a monotonic function on  $L$  and that  $\eta_d(x) + \zeta_d(x) = d(0, 1)$  because we have  $B(0, x, 1)$  for every  $x \in L$ .

Note that the triangular inequality implies

$$\begin{aligned} d(x, y) &\leq d(x, x \wedge y) + d(x \wedge y, y), \\ d(x, y) &\leq d(x, x \vee y) + d(x \vee y, y), \end{aligned}$$

for  $x, y \in L$ , hence  $2d(x, y) \leq d(x, x \wedge y) + d(x \wedge y, y) + d(x, x \vee y) + d(x \vee y, y) = 2d(x \wedge y, x \vee y)$ . Thus, since we have both  $B_d(x \wedge y, x, x \vee y)$  and  $B_d(x \wedge y, y, x \vee y)$  it follows that  $d(x, y) \leq d(x \wedge y, x \vee y)$  for  $x, y \in L$ .

We are especially interested in two types of metrics introduced in the next.

**Definition 2.1** Let  $\mathcal{L} = (L, \vee, \wedge, 0, 1)$  be a lattice and let  $d : L \times L \longrightarrow \mathbb{R}_{\geq 0}$  be a metric defined on  $L$ . Then,  $d$  is a  $\wedge$ -metric if  $\mathbb{B}_d(x, x \wedge y, y)$  and is a  $\vee$ -metric if  $\mathbb{B}_d(x, x \vee y, y)$ , respectively, for every  $x, y \in L$ .  $\square$

**Lemma 2.2** If  $d$  is a  $\wedge$ -metric on the lattice  $\mathcal{L} = (L, \vee, \wedge, 0, 1)$ , then

$$d(x, y) = 2\eta_d(x \wedge y) - \eta_d(x) - \eta_d(y) \quad (1)$$

for every  $x, y \in L$ . If  $d$  is a  $\vee$ -metric then

$$d(x, y) = \zeta_d(x) + \zeta_d(y) - 2\zeta_d(x \vee y) \quad (2)$$

for every  $x, y \in L$ .

**Proof.** Suppose that  $d$  is a  $\wedge$ -metric, so  $d(x, y) = d(x, x \wedge y) + d(x \wedge y, y)$  for  $x, y \in L$ . Observe that  $d(x \wedge y, x) + d(x, 1) = d(x \wedge y, 1)$ , since  $d$  is a compatible metric and, similarly,  $d(x \wedge y, y) + d(y, 1) = d(x \wedge y, 1)$ . Therefore,

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

because  $d$  is a  $\wedge$ -metric. The argument for  $\vee$ -metrics is similar.  $\blacksquare$

For reasons that will become here later, if  $d$  is a  $\wedge$ -metric, then we refer to  $\eta_d$  as a  $\wedge$ -entropy; similarly, if  $d$  is a  $\vee$ -metric we refer to  $\zeta_d$  as a  $\vee$ -entropy.

### 3 Metrics Derived from Entropies

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

Let  $f : L \longrightarrow \mathbb{R}$  be a function. Define the mappings  $d_f : L^2 \longrightarrow \mathbb{R}$  and  $d^f : L^2 \longrightarrow \mathbb{R}$  by

$$d_f(x, y) = 2f(x \wedge y) - f(x) - f(y),$$

and

$$d^f(x, y) = f(x) + f(y) - 2f(x \vee y)$$

for  $x, y \in L$ .

**Theorem 3.1** Let  $\mathcal{L} = (L, \vee, \wedge, 0, 1)$  be a lattice and let  $f : L \longrightarrow \mathbb{R}$  be a function. The following statements hold:

- (i)  $d_f(x, y) \geq 0$  for every  $x, y \in L$  if and only if  $f$  is anti-monotonic;
- (ii)  $d_f(x, y) > 0$  for every  $x, y \in L$  such that  $x \neq y$  if and only if  $f$  is strictly anti-monotonic;
- (iii)  $d^f(x, y) \geq 0$  for every  $x, y \in L$  if and only if  $f$  is anti-monotonic;
- (iv)  $d^f(x, y) > 0$  for every  $x, y \in L$  such that  $x \neq y$  if and only if  $f$  is strictly anti-monotonic;

**Proof.** Suppose that  $d_f(x, y) \geq 0$  for every  $x, y \in L$ . If  $x \leq y$ , we have

$$\begin{aligned} d_f(x, y) &= 2f(x \wedge y) - f(x) - f(y) \\ &= f(x) - f(y) \geq 0, \end{aligned}$$

so  $f$  is anti-monotonic. Conversely, suppose that  $f$  is anti-monotonic. Then,  $f(x \wedge y) \geq f(x), f(y)$ , so  $d(x, y) \geq 0$ .

For  $d^f$  the same conclusion can be reached, by observing that  $f(x), f(y) \geq f(x \vee y)$ .

Suppose that  $d_f(x, y) = 0$ , where  $(d, \eta)$ , where  $f$  is an anti-monotonic function. Then,  $2f(x \wedge y) - f(x) - f(y) = 0$ , so  $f(x) = f(y) = f(x \wedge y)$  because  $f(x) \leq f(x \wedge y)$  and  $f(y) \leq f(x \wedge y)$ . 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  $d^f$  is similar. ■

**Lemma 3.2** *The function  $d_f$  satisfies the triangular inequality,  $d(x, y) \leq d(x, z) + d(z, y)$  if and only if*

$$f(z) + f(x \wedge y) \leq f(x \wedge z) + f(y \wedge z) \quad (3)$$

for  $x, y, z \in L$ . Furthermore, the function  $d^f$  satisfies the triangular inequality if and only if

$$f(z) + f(x \vee y) \geq f(x \vee z) + f(y \vee z) \quad (4)$$

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

**Proof.** Let  $d_f$  be a function that satisfies the triangular inequality. This implies

$$\begin{aligned} &2f(x \wedge y) - f(x) - f(y) \\ &\leq 2f(x \wedge z) - f(x) - f(z) \\ &\quad + 2f(y \wedge z) - f(y) - f(z), \end{aligned}$$

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

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

**Theorem 3.3** *Let  $\mathcal{L} = (L, \vee, \wedge, 0, 1)$ ,  $f : L \rightarrow \mathbb{R}$  be a function such that  $d_f$  is a non-negative. Then,  $d_f$  satisfies the triangular inequality if and only if  $f$  is an anti-monotonic and sub-modular, that is,  $f(x \vee y) + f(x \wedge y) \leq f(x) + f(y)$  for every  $x, y \in L$ .*

*The function  $d^f$  satisfies the triangular inequality if and only if  $f$  is an anti-monotonic and supramodular, that is,  $f(x \vee y) + f(x \wedge y) \geq f(x) + f(y)$  for every  $x, y \in L$ .*

**Proof.** Suppose that  $d_f$  is a non-negative function that satisfies the triangular inequality. Then, by Theorem 3.1 and Lemma 3.2,  $f$  is an anti-monotonic function, and  $f(z) + f(x \wedge y) \leq f(x \wedge z) + f(y \wedge z)$  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 submodular inequality  $f(x \vee y) + f(x \wedge y) \leq f(x) + f(y)$ .

If  $d^f$  is a non-negative function that satisfies the triangular inequality, then  $f$  is an anti-monotonic function and  $f(z) + f(x \vee y) \geq f(x \vee z) + f(y \vee z)$  for  $x, y, z \in L$ . Substituting  $x \wedge y$  for  $z$  and applying the absorption properties we have the supramodular inequality  $f(x \vee y) + f(x \wedge y) \geq f(x) + f(y)$ .

Conversely, suppose that  $f$  is an anti-monotonic, submodular function. The anti-monotonicity of  $f$  implies the non-negativity of  $d^f$ . We need to show that the submodular inequality implies (3). The definition of suprema and infima implies that

$$z \geq (x \wedge z) \vee (y \wedge z),$$

for  $x, y, z \in L$ . From the anti-monotonicity of  $f$  we have

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

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

$$(x \vee y) \wedge z \geq (x \wedge z) \vee (y \wedge z), \quad (5)$$

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

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

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

When  $f$  be an anti-monotonic, supramodular function, the Dualization Principle gives Inequality (4). ■

**Corollary 3.4** *Let  $\mathcal{L} = (L, \vee, \wedge, 0, 1)$  be a lattice and let  $f : L \rightarrow \mathbb{R}$  be a strictly anti-monotonic function. If  $f$  is submodular, then  $d_f$  is a compatible  $\wedge$ -metric and for the  $\wedge$ -entropy we have  $\eta_{d_f}(x) = f(x) - f(1)$ .*

*If  $f$  is supramodular, then  $d^f$  is a compatible  $\vee$ -metric and for the  $\vee$ -entropy we have  $\zeta^{d^f}(x) = f(0) - f(x)$ .*

**Proof.** The corollary is an immediate consequence of Theorem 3.3. ■

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

**Corollary 3.5** *If  $f$  is a function such that  $d_f = d^f$  is a metric on a lattice  $\mathcal{L} = (L, \vee, \wedge, 0, 1)$ , then  $\mathcal{L}$  is a metric lattice and  $d_f(x, y) = f(x \wedge y) - f(x \vee y)$  for  $x, y \in L$ .*

**Proof.** Since  $d_f = d^f$  and  $d_f$  is a metric the strictly anti-monotonic function  $f$  satisfies both the submodular and the supramodular inequalities and therefore we have  $f(x \wedge y) + f(x \vee y) = f(x) + f(y)$ , so  $f$  is a modular valuation on  $L$ . This implies

$$d_f(x, y) = f(x \wedge y) - f(x \vee y),$$

for  $x, y \in L$ . ■

Starting from a function  $f : L \rightarrow \mathbb{R}$ , define the conditional function  $\kappa_f : L^2 \rightarrow \mathbb{R}$  by  $\kappa_f(x, y) = f(x \wedge y) - f(y)$  for  $x, y \in L$ . It is immediate that  $d_f(x, y) = \kappa_f(x, y) + \kappa_f(y, x)$  and that  $x \geq y$  implies  $\kappa_f(x, y) = 0$  for  $x, y \in L$ .

As we shall see, the conditional function  $\kappa_f$  formalizes the notion of conditional entropy corresponding to an entropy.

**Theorem 3.6** *The non-negative function  $d_f$  satisfies the triangular inequality if and only if the conditional function  $\kappa_f$  is anti-monotonic in its first argument and monotonic in its second argument.*

**Proof.** Suppose that  $d_f$  satisfies the triangular inequality. The anti-monotonicity of  $\kappa_f$  in its first argument follows from the anti-monotonicity of  $f$ .

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

$$f(y_1) \leq f((x \wedge y_1) \vee y). \quad (6)$$

By Theorem 3.3 we have the submodular inequality  $f(x \vee y) + f(x \wedge y) \leq f(x) + f(y)$  for every  $x, y \in L$ . Taking Inequality (6) into account and replacing  $x$  by  $x \wedge y_1$  in the submodular inequality yields

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

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

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

If  $y \leq y_1$ , we have  $f(x \wedge y) - f(y) \leq f(x \wedge y_1) - f(y_1)$ . Choosing  $y_1 = x \vee y$  we obtain the submodular inequality for  $f$ , which shows that  $d$  satisfies the triangular inequality by Theorem 3.3. ■

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

**Theorem 3.7** *The non-negative function  $d^f$  satisfies the triangular inequality if and only if the conditional function  $\kappa^f$  is monotonic in its first argument and anti-monotonic in its second argument.*

**Proof.** The proof is analogous to the argument of Theorem 3.6. ■

## 4 Applications to Partition Lattices

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  $B_i \cap B_j = \emptyset$  when  $i \neq j$  for  $i, j \in I$ . The sets  $B_i$  are the *blocks* of  $\pi$ . The set of partitions of  $S$  is denoted by  $\text{PART}(S)$ .

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

The partially ordered set  $(\text{PART}(S), \leq)$  is actually a bounded lattice. The infimum  $\pi \wedge \pi'$  of two partitions  $\pi$  and  $\pi'$  is the partition that consists of non-empty intersections of blocks of  $\pi$  and  $\pi'$ . For a description of the supremum  $\pi \vee \pi'$  of the partitions  $\pi, \pi'$

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  $\sigma$  covers the partition  $\pi$  if  $\sigma$  is obtained from  $\pi$  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  $\sigma_0, \sigma_1, \dots, \sigma_r$  such that  $\pi = \sigma_0 \prec \sigma_1 \prec \dots \prec \sigma_r = \pi'$ .

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

For partition lattices of finite sets the metrics  $d_f$  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.

Let  $S$  be a finite set and let  $(\mathbf{PART}(S), \leq)$  be the partition lattice having  $\alpha_S$  as its least element and  $\omega_S$  as its largest element. For  $\beta \in \mathbb{R}$  and  $\beta > 1$  define the mapping  $f_\beta : \mathbf{PART}(S) \rightarrow \mathbb{R}$  as:

$$f_\beta(\pi) = \frac{1}{1 - 2^{1-\beta}} \left( 1 - \sum_{i=1}^m \left( \frac{|B_i|}{|S|} \right)^\beta \right), \quad (7)$$

where  $\pi = \{B_1, \dots, B_m\}$ . Observe that  $f_\beta(\omega_S) = 0$  and  $f_\beta(\pi) > 0$  for  $\pi \in \mathbf{PART}(S) - \{\omega_S\}$ .

**Lemma 4.1** *Let  $S$  be a finite set,  $\pi \in \mathbf{PART}(S)$  and let  $C, D$  be two disjoint subsets of  $S$ . For  $\beta \geq 1$  we have:*

$$\begin{aligned} & \left( \frac{|C \cup D|}{|S|} \right)^\beta f_\beta(\pi_{C \cup D}) \\ & \geq \left( \frac{|C|}{|S|} \right)^\beta f_\beta(\pi_C) + \left( \frac{|D|}{|S|} \right)^\beta f_\beta(\pi_D), \end{aligned}$$

where  $f_\beta : \mathbf{PART}(S) \rightarrow \mathbb{R}$  is the function introduced in Equality (7).

**Proof.** The proof is elementary and is omitted. ■

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

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  $f_\beta(\pi) < f_\beta(\pi')$ , that is, the strict anti-monotonicity property. By Theorem 3.1, the function  $d_{f_\beta}$  given by  $d_{f_\beta}(\pi, \sigma) = 2f_\beta(\pi \wedge \sigma) - f_\beta(\pi) - f_\beta(\sigma)$  for  $\pi, \sigma \in \mathbf{PART}(S)$  is non-negative and  $d_{f_\beta}(\pi, \sigma) = 0$  implies  $\pi = \sigma$ .

**Example 4.2** Define the function  $f_1 : \mathbf{PART}(S) \rightarrow \mathbb{R}$  by

$$f_1(\pi) = - \sum_{i=1}^m \frac{|B_i|}{|S|} \log \frac{|B_i|}{|S|},$$

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

This is the Shannon entropy of the probability distribution

$$\left( \frac{|B_1|}{|S|} \quad \dots \quad \frac{|B_m|}{|S|} \right)$$

defined by the partition  $\pi \in \text{PART}(S)$ . It is easy to verify that  $\lim_{\beta \rightarrow 1} f_\beta(\pi) = f_1(\pi)$ , which implies that  $f_1$  is anti-monotonic. An elementary argument can be used to verify that  $f_1$  is, in fact, strictly anti-monotonic, so the function  $d_{f_1} : \text{PART}(S)^2 \rightarrow \mathbb{R}$  given by

$$\begin{aligned} d_{f_1}(\pi, \sigma) &= \sum_{i=1}^m \frac{|B_i|}{|S|} \log \frac{|B_i|}{|S|} + \sum_{j=1}^n \frac{|C_j|}{|S|} \log \frac{|C_j|}{|S|} \\ &\quad - \sum_{i=1}^m \sum_{j=1}^n \frac{|B_i \cap C_j|}{|S|} \log \frac{|B_i \cap C_j|}{|S|}, \end{aligned}$$

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

The conditional function of  $f_\beta$  introduced in Equality (7) is given by

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

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

This function can be written alternatively as

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

where  $\pi_{C_j}$  is the trace of  $\pi$  on the block  $C_j$  of  $\sigma$ .

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

**Theorem 4.3** *Let  $\pi, \sigma, \sigma' \in \text{PART}(S)$ , where  $S$  is a finite set. If  $\sigma \leq \sigma'$ , then  $\kappa_{f_\beta}(\sigma, \pi) \geq \kappa_{f_\beta}(\sigma', \pi)$  and  $\kappa_{f_\beta}(\pi, \sigma) \leq \kappa_{f_\beta}(\pi, \sigma')$ .*

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

For the monotonicity of  $\kappa_{f_\beta}$  in its second argument it suffices to prove the monotonicity for partitions  $\sigma, \sigma'$  such that  $\sigma \prec \sigma'$ . Without restricting the generality we may



assume that  $\sigma = \{C_1, \dots, C_{n-2}, C_{n-1}, C_n\}$  and  $\sigma' = \{C_1, \dots, C_{n-2}, C_{n-1} \cup C_n\}$ . Thus, we can write:

$$\begin{aligned}
& \kappa_{f_\beta}(\pi, \sigma') \\
&= \sum_{j=1}^{n-2} \left( \frac{|C_j|}{|S|} \right)^\beta f_\beta(\pi_{C_j}) + \left( \frac{|C_{n-1} \cup C_n|}{|S|} \right)^\beta f_\beta(\pi_{C_{n-1} \cup C_n}) \\
&\geq \sum_{j=1}^{n-2} \left( \frac{|C_j|}{|S|} \right)^\beta f_\beta(\pi_{C_j}) + \left( \frac{|C_{n-1}|}{|S|} \right)^\beta f_\beta(\pi_{C_{n-1}}) + \left( \frac{|C_n|}{|S|} \right)^\beta f_\beta(\pi_{C_n}) \\
&\quad \text{(by Lemma 4.1)} \\
&= \kappa_{f_\beta}(\pi, \sigma).
\end{aligned}$$

■

**Corollary 4.4** *Let  $f_\beta$  be the function introduced in Equality (7). Then  $d_{f_\beta}$  is a metric on the lattice of partitions  $(\text{PART}(S), \leq)$ .*

**Proof.** This statement follows from Theorems 3.1, 3.6, and 4.3. ■

By Corollary 3.4, since  $f_\beta(\omega_S) = 0$ , we see that  $f_\beta = \eta_\beta$ , that is,  $f_\beta$  is actually the  $\wedge$ -entropy associated to the  $\wedge$ -metric  $d_{f_\beta}$ . An independent axiomatization of this entropy was presented in [11].

## 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  $g : P \rightarrow \mathbb{Z}$  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  $\text{height}(x)$ . If  $(P, \leq)$  has the largest element 1, then the height of  $(P, \leq)$  is defined as  $\text{height}(1)$ . 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 submodular inequality and is lower semimodular if height satisfies the supramodular inequality (cf. Theorem II.15, p. 40 of [1]).

The function  $f : P \rightarrow \mathbb{R}$  defined by  $f(x) = \text{height}(P) - \text{height}(x)$  satisfies the supramodular inequality and the associated function  $d_f$  satisfies the triangular inequality and, therefore, it is a pseudometric on the lattice  $L$  given by  $d_f(x, y) = h(x) + h(y) - 2h(x \wedge y)$  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.

## 7 Acknowledgement

The author wishes to thank an anonymous reviewer whose careful reading and judicious observations contributed to the improvement of this paper.

## 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 - Journées francophones d'Extraction et de Gestion de Connaissances*, pages 369–380, Lyon, France, 2003.