On Generalized Entropy and Entropic Metrics

Dan Simovici [*]

Abstract

Starting from an axiomatization of a generalization of Shannon entropy we introduce a set of axioms for a parametric family of distances over sets of partitions of finite sets. This family includes some well-known metrics used in data mining and in the study of finite functions.

1 Introduction

The notion of entropy is as a probabilistic concept that lies at the foundation of information theory. Our goal is to define entropy in an algebraic setting, namely, introduce the notion of entropy of a partition taking advantage of the partial order that is naturally defined on the set of partitions of a set. Actually, we will introduce a generalization of the notion of entropy that has the Gini index and Shannon entropy as special cases and further extends some of our previous results.

Another goal of this paper is the study an axiomatization of a parameterized family of metrics on sets of partitions of finite sets that generalizes the entropic metric introduced by R. López de Mántaras [6], as well as the Mirkin metric introduced in [10]. This unifies the separate axiom systems for these metrics introduced in [9] and illuminates the relationship of the axiomatization of these metrics with our previous axiomatization of generalized entropy [16, 14].

Metrics on sets of partitions of finite sets are useful because they allow us to study properties of finite functions related to their kernel partitions. In a different direction, these metrics are interesting for data mining because the attributes of a table induce partitions on the sets of tuples of the table. Thus, metrics on partitions allow us to determine interesting relationships between attributes and to use these relationship for classification, data summarization and other applications. Also, exclusive clusterings can be regarded as partitions of the set of clustered objects and partition metrics can be used for evaluating clusterings, a point of view presented in [9].

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 $\mathsf{PART}(S)$.

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

A natural link exists between random variables and partitions of sets. Namely, if S is a finite set and $\pi = \{B_1, \ldots, B_n\}$ is a partition of S, then

$$\mathbf{p} = \left(\frac{|B_1|}{|S|}, \ldots, \frac{|B_n|}{|S|} \right)$$

is a discrete probability distribution. As we shall see, the Shannon entropy of **p** equals the Shannon entropy of π, as follows from our axiomatization. This link allows the transfer of certain probabilistic and information-theoretical notions to partitions of sets, where we can take advantage of the partial order between partitions.

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 of two partitions π and π' is the partition that consists of non-empty intersections of blocks of π and π'. 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 $\pi \prec \sigma$. If $\pi \leq \pi'$, then there exists a sequence of partitions $\sigma_0, \sigma_1, \ldots, \sigma_r$ such that $\pi = \sigma_0 \prec \sigma_1 \prec \cdots \prec \sigma_r = \pi'$.

If S, T are two disjoint and nonempty sets, $\pi \in \mathsf{PART}(S)$, $\sigma \in \mathsf{PART}(T)$, where $\pi = \{B_1, \ldots, B_m\}$, $\sigma = \{C_1, \ldots, C_n\}$, then the partition $\pi + \sigma$ is the partition of $S \cup T$ given by

$$\pi + \sigma = \{B_1, \ldots, B_m, C_1, \ldots, C_n\}.$$

Whenever the "$+$" operation is defined, then it is easily seen to be associative. In other words, if S, T, U are pairwise disjoint and nonempty sets, and $\pi \in \mathsf{PART}(S)$, $\sigma \in \mathsf{PART}(T)$, $\tau \in \mathsf{PART}(U)$, then $\pi + (\sigma + \tau) = (\pi + \sigma) + \tau$. Observe that if S, T are disjoint, then $\alpha_S + \alpha_T = \alpha_{S \cup T}$. Also, $\omega_S + \omega_T$ is the partition $\{S, T\}$ of the set $S \cup T$.

If $\pi = \{B_1, \ldots, B_m\}$, $\sigma = \{C_1, \ldots, C_n\}$ are partitions of two arbitrary sets S, T, then we denote the partition $\{B_i \times C_j \mid 1 \leq i \leq m, 1 \leq j \leq n\}$ of $S \times T$ by $\pi \times \sigma$. Note that there is a natural bijection betwen $\alpha_S \times \alpha_T$ and $\alpha_{S \times T}$ and also, between $\omega_S \times \omega_T$ and $\omega_{S \times T}$.

2 An Axiomatization of Generalized Entropy

We present a new system of axioms for generalized entropies that have as special cases Shannon entropy and the generalized entropy introduced in [4, 5]). We further show that, under certain hypotheses, the second type of entropies in the single generalization possible.

Definition 2.1 A function $h : \mathbb{N} \longrightarrow \mathbb{R}_{\geq 0}$ is *multiplicative* if $h(pq) = h(p)h(q)$ for $p, q \in \mathbb{N}$. □

This notion of multiplicative function is stronger than the usual notion used in number theory, when the multiplicative equality holds if p and q are relatively prime. An example of such a function is $h(r) = r^\beta$ for $r \in \mathbb{N}$ and $\beta \in \mathbb{R}_{\geq 0}$; this function will play an important role in our axiomatization. Another trivial example is the constant function $h(p) = 0$ for $p \in \mathbb{N}$, which will be referred to as the *zero function*.

Note that if h a multiplicative function distinct from the zero function, then $h(1) = 1$.

We introduce below a system of four axioms:

Definition 2.2 Let $\beta \in \mathbb{R}$, $\beta \geq 1$, $\Phi : \mathbb{R}^2_{\geq 0} \longrightarrow \mathbb{R}_{\geq 0}$ be a continuous function such that $\Phi(x, y) = \Phi(y, x)$, and $\Phi(x, 0) = x$ for $x, y \in \mathbb{R}_{\geq 0}$, and $h : \mathbb{N} \longrightarrow \mathbb{R}_{\geq 0}$ a non-zero multiplicative function such that $h(n) = 0$ implies $n = 0$.

A (Φ, h)-*system of axioms for a partition entropy* $\mathcal{H}_h : \mathsf{PART}(S) \longrightarrow \mathbb{R}_{\geq 0}$ consists of the following axioms:

(P1) If $\pi, \pi' \in \mathsf{PART}(S)$ are such that $\pi \leq \pi'$, then $\mathcal{H}_h(\pi) \geq \mathcal{H}_h(\pi')$.

(P2) If S, T are two finite sets such that $|S| \leq |T|$, then $\mathcal{H}_h(\alpha_S) \leq \mathcal{H}_h(\alpha_T)$.

(P3) For every disjoint sets S, T and partitions $\pi \in \mathsf{PART}(S)$, and $\sigma \in \mathsf{PART}(T)$ we have:

$$\mathcal{H}_h(\pi + \sigma) = \frac{h(|S|)}{h(|S| + |T|)}\mathcal{H}_h(\pi) + \frac{h(|T|)}{h(|S| + |T|)}\mathcal{H}_h(\sigma) + \mathcal{H}_h(\{S, T\}).$$

(P4) We have:

$$\mathcal{H}_h(\pi \times \sigma) = \Phi(\mathcal{H}_h(\pi), \mathcal{H}_h(\sigma))$$

for $\pi \in \mathsf{PART}(S)$ and $\sigma \in \mathsf{PART}(T)$.

$\square$

Observe that we postulate that $\mathcal{H}_h(\pi) \geq 0$ for any partition π since the range of every function $\mathcal{H}_h$ is $\mathbb{R}_{\geq 0}$.

Lemma 2.3 *For every (Φ, h)-entropy $\mathcal{H}_h$ and set S we have $\mathcal{H}_h(\omega_S) = 0$.*

Proof. Let S, T be two non-empty disjoint sets that have the same cardinality, $|S| = |T|$. Since $\omega_S + \omega_T$ is the partition $\{S, T\}$ of the set $S \cup T$, by Axiom **(P3)** we have

$$\mathcal{H}_h(\omega_S + \omega_T) = \frac{h(|S|)}{h(2|S|)}(\mathcal{H}_h(\omega_S) + \mathcal{H}_h(\omega_T)) + \mathcal{H}_h(\{S, T\}),$$

which implies $\mathcal{H}_h(\omega_S) + \mathcal{H}_h(\omega_T) = 0$. Since $\mathcal{H}_h(\omega_S) \geq 0$ and $\mathcal{H}_h(\omega_T)) \geq 0$ it follows that $\mathcal{H}_h(\omega_S) = \mathcal{H}_h(\omega_T) = 0$. ∎

Lemma 2.4 *Let U, V be two disjoint sets and let $\pi, \pi' \in \mathsf{PART}(U \cup V)$ be defined by $\pi = \sigma + \alpha_V$ and $\pi' = \sigma + \omega_V$, where $\sigma \in \mathsf{PART}(U)$. Then,*

$$\mathcal{H}_h(\pi) = \mathcal{H}_h(\pi') + \frac{h(|V|)}{h(|U| + |V|)}\mathcal{H}_h(\alpha_V).$$

Proof. By Axiom **(P3)** we can write:

$$\begin{aligned}\mathcal{H}_h(\pi) &= \frac{h(|U|)}{h(|U|+|V|)}\mathcal{H}_h(\sigma)\\ &\quad+\frac{h(|V|)}{h(|U|+|V|)}\mathcal{H}_h(\alpha_T)+\mathcal{H}_h(\{U,V\}),\end{aligned}$$

and

$$\begin{aligned}\mathcal{H}_h(\pi') &= \frac{h(|U|)}{h(|U|+|V|)}\mathcal{H}_h(\sigma)\\ &\quad+\frac{h(|V|)}{h(|U|+|V|)}\mathcal{H}_h(\omega_T)+\mathcal{H}_h(\{U,V\})\\ &= \frac{h(|U|)}{h(|U|+|V|)}\mathcal{H}_h(\sigma)+\mathcal{H}_h(\{U,V\})\\ &\quad\text{(by Lemma 2.3).}\end{aligned}$$

The above equalities imply immediately the equality of the lemma. ∎

Theorem 2.5 *For every (Φ, h)-entropy and partition $\pi = \{B_1, \ldots, B_m\} \in \mathsf{PART}(S)$ we have:*

$$\mathcal{H}_h(\pi) = \mathcal{H}_h(\alpha_S) - \sum_{i=1}^{m} \frac{h(|B_i|)}{h(|S|)}\mathcal{H}_h(\alpha_{B_i}).$$

Proof. Starting from the partition π consider the following sequence of partitions in $\mathsf{PART}(S)$:

$$\begin{aligned}\pi_0 &= \omega_{B_1}+\omega_{B_2}+\omega_{B_3}+\cdots+\omega_{B_m}\\ \pi_1 &= \alpha_{B_1}+\omega_{B_2}+\omega_{B_3}+\cdots+\omega_{B_m}\\ \pi_2 &= \alpha_{B_1}+\alpha_{B_2}+\omega_{B_3}+\cdots+\omega_{B_m}\\ &\vdots\\ \pi_n &= \alpha_{B_1}+\alpha_{B_2}+\alpha_{B_3}+\cdots+\alpha_{B_m}.\end{aligned}$$

Let $\sigma_j = \alpha_{B_1} + \cdots + \alpha_{B_j} + \omega_{B_{i+2}} + \cdots + \omega_{B_m}$. Then, $\pi_i = \sigma_i + \omega_{B_{i+1}}$ and $\pi_{i+1} = \sigma_i + \alpha_{B_{i+1}}$; therefore, by Lemma 2.4, we have:

$$\mathcal{H}_h(\pi_{i+1}) = \mathcal{H}_h(\pi_i) + \frac{h(|B_{i+1}|)}{h(|S|)}\mathcal{H}_h(\alpha_{B_{i+1}})$$

for $0 \le i \le m-1$.

A repeated application of this equality yields:

$$\mathcal{H}_h(\pi_m) = \mathcal{H}_h(\pi_0) + \sum_{i=0}^{m-1} \frac{h(|B_{i+1}|)}{h(|S|)}\mathcal{H}_h(\alpha_{B_{i+1}}).$$

Observe that $\pi_0 = \pi$ and $\pi_m = \alpha_S$. Consequently,

$$\mathcal{H}_h(\pi) = \mathcal{H}_h(\alpha_S) - \sum_{i=1}^{m} \frac{h(|B_i|)}{h(|S|)} \mathcal{H}_h(\alpha_{B_i}).$$

∎

Note that if S, T are two sets such that $|S| = |T| > 0$, then, by Axiom **(P2)**, we have $\mathcal{H}_h(\alpha_S) = \mathcal{H}_h(\alpha_T)$. Therefore, the value of $\mathcal{H}_h(\alpha_S)$ depends only on the cardinality of S, and there exists a function $\mu_h : \mathbb{N}_1 \longrightarrow \mathbb{R}_{\geq 0}$ such that $\mathcal{H}_h(\alpha_S) = \mu_h(|S|)$ for every nonempty set S. Axiom **(P2)** also implies that μ_h is an increasing function. We will refer to μ_h as the *core* of the (Φ, h)-system of axioms.

Corollary 2.6 *Let $\mathcal{H}_h$ be a (Φ, h)-entropy. For the core μ_h defined in accordance to Axiom* **(P2)** *and every partition $\pi = \{B_1, \ldots, B_m\} \in \mathsf{PART}(S)$ we have:*

$$\mathcal{H}_h(\pi) = \mu_h(|S|) - \sum_{i=1}^{m} \frac{h(|B_i|)}{h(|S|)} \mu_h(|B_i|). \tag{1}$$

Proof. The statement is an immediate consequence of Theorem 2.5. ∎

Theorem 2.7 *Let $\pi = \{B_1, \ldots, B_m\}$ be a partition of the set S. Define the partition π' obtained by fusing the blocks B_1 and B_2 of π as $\pi' = \{B_1 \cup B_2, B_3, \ldots, B_m\}$ of the same set. Then*

$$\mathcal{H}_h(\pi) = \mathcal{H}_h(\pi') + \frac{h(|B_1 \cup B_2|)}{h(|S|)} \mathcal{H}_h(\{B_1, B_2\}).$$

Proof. A double application of Corollary 2.6 yields:

$$\begin{aligned} \mathcal{H}_h(\pi') &= \mu_h(|S|) - \frac{h(|B_1 \cup B_2|)}{h(|S|)} \mu_h(|B_1 \cup B_2|) \\ &\quad - \sum_{i>2}^{m} \frac{h(|B_i|)}{h(|S|)} \mu_h(|B_i|) \end{aligned}$$

and

$$\begin{aligned} \mathcal{H}_h(\{B_1, B_2\}) &= \mu(|B_1 \cup B_2|) - \frac{h(|B_1|)}{h(|B_1 \cup B_2|)} \mu(|B_1|) \\ &\quad - \frac{h(|B_2|)}{h(|B_1 \cup B_2|)} \mu(|B_2|). \end{aligned}$$

Substituting the above expressions in

$$\mathcal{H}_h(\pi') + \frac{h(|B_1 \cup B_2|)}{h(|S|)} \mathcal{H}_h(\{B_1, B_2\})$$

we obtain $\mathcal{H}_h(\pi)$. ∎

Theorem 2.7 allows us to extend Axiom **(P3)**:

Corollary 2.8 *Let $B_1, \ldots, B_m$ be m nonempty, disjoint sets and let $\pi_i \in \mathsf{PART}(B_i)$ for $1 \leq i \leq m$. We have:*

$$\mathcal{H}_h(\pi_1 + \cdots + \pi_m) = \sum_{i=1}^{m} \frac{h(|B_i|)}{h(|S|)} \mathcal{H}_h(\pi_i) + \mathcal{H}_h(\{B_1, \ldots, B_m\}),$$

where $S = B_1 \cup \cdots \cup B_m$.

Proof. The argument is by induction on $m \geq 2$. The basis step, $m = 2$, is Axiom **(P3)**. Suppose that the statement holds for m and let $B_1, \ldots, B_m, B_{m+1}$ be $m+1$ disjoint sets. Further, suppose that $\pi_1, \ldots, \pi_m, \pi_{m+1}$ are partitions of these sets, respectively. Then, $\pi_m + \pi_{m+1}$ is a partition of the set $B_m \cup B_{m+1}$. By the inductive hypothesis we have

$$\begin{aligned}
&\mathcal{H}_h(\pi_1 + \cdots + (\pi_m + \pi_{m+1})) \\
&= \sum_{i=1}^{m-1} \frac{h(|B_i|)}{h(|S|)} \mathcal{H}_h(\pi_i) + \frac{h(|B_m| + |B_{m+1}|)}{h(|S|)} \mathcal{H}_h(\pi_m + \pi_{m+1}) \\
&\quad + \mathcal{H}_h(\{B_1, \ldots, (B_m \cup B_{m+1})\}),
\end{aligned}$$

where $S = B_1 \cup \cdots \cup B_m \cup B_{m+1}$.

Axiom **(P3)** implies:

$$\begin{aligned}
&\mathcal{H}_h(\pi_1 + \cdots + (\pi_m + \pi_{m+1})) \\
&= \sum_{i=1}^{m-1} \frac{h(|B_i|)}{h(|S|)} \mathcal{H}_h(\pi_i) + \frac{h(|B_m|)}{h(|S|)} \mathcal{H}_h(\pi_m) \\
&\quad + \frac{h(|B_{m+1}|)}{h(|S|)} \mathcal{H}_h(\pi_{m+1}) + \frac{h(|B_m| + |B_{m+1}|)}{h(|S|)} \mathcal{H}_h\{B_m, B_{m+1}\} \\
&\quad + \mathcal{H}_h(\{B_1, \ldots, (B_m \cup B_{m+1})\}).
\end{aligned}$$

Finally, an application of Theorem 2.7 gives the desired equality. ∎

Theorem 2.9 *Let μ_h be the core of a (Φ, h)-system such that the function h is not the identity function $h(n) = n$ for $n \in \mathbb{N}$. There exists a number $k \in \mathbb{R}$ such that*

$$\mu(a) = k \left(1 - \frac{a}{h(a)}\right)$$

for $a \in \mathbb{N}_1$.

Proof. Let $A = \{x_1, \ldots, x_a\}$ and $B = \{y_1, \ldots, y_b\}$ be two nonempty sets, where $a, b \in \mathbb{N}_1$. The partition $\omega_A \times \alpha_B$ consists of b blocks of size a: $A \times \{y_1\}, \ldots, A \times \{y_b\}$. By Axiom **(P4)**,

$$\begin{aligned}
&\mathcal{H}_h(\omega_A \times \alpha_B) \\
&= \Phi(\mathcal{H}_h(\omega_A), \mathcal{H}_h(\alpha_B)) = \Phi(0, \mathcal{H}_h(\alpha_B)) = \mathcal{H}_h(\alpha_B) = \mu_h(b).
\end{aligned}$$

On the other hand, by Theorem 2.5 we have

$$\begin{aligned}\mathcal{H}_h(\omega_A \times \alpha_B) &= \mathcal{H}_h(\alpha_{A\times B}) - \sum_{i=1}^{b} \frac{h(a)}{h(ab)} \mathcal{H}_h(\alpha_{A\times\{y_i\}})\\ &= \mu_h(ab) - b\frac{h(a)}{h(ab)} \cdot \mu_h(a),\end{aligned}$$

which implies

$$\mu_h(ab) = \mu_h(b) + b\frac{h(a)}{h(ab)}\mu_h(a) = \mu_h(b) + b\frac{1}{h(b)}\mu_h(a), \tag{2}$$

for $a, b \in \mathbb{N}_1$, since h is a multiplicative function. Inverting the roles of a and b we obtain also

$$\mu_h(ab) = \mu_h(a) + a\frac{1}{h(a)}\mu_h(b),$$

which implies

$$\frac{\mu_h(a)}{1 - \frac{a}{h(a)}} = \frac{\mu_h(b)}{1 - \frac{b}{h(b)}},$$

for every $a, b \in \mathbb{N}_1$, which yields the desired equality for μ_h. ∎

An entropy is said to be *non-Shannon* if it is defined by a (Φ, h)-system of axioms such that $h(n) \neq n$ for $n \in \mathbb{N}$. If $h(n) = n$ for $n \in \mathbb{N}$, then the entropy will be referred to as a *Shannon* entropy. As we shall see, the choice of the function h determines the form of the function Φ. Initially we focus on non-Shannon entropies.

The next Corollary shows that within the framework of our axiomatization the one obtains necessarily the non-Shannon generalized entropy introduced in [4, 5]).

Corollary 2.10 *If $h : \mathbb{N} \longrightarrow \mathbb{R}_{\geq 0}$ is a multiplicative function used in a (Φ, h) axiomatization of a non-Shannon entropy, then $h(n) = n^\beta$ for some $\beta > 1$ and $\mu_h(n) = k(1 - n^{1-\beta})$.*

Proof. We observed that Axiom **(P2)** implies that μ_h is an increasing function. Since, by Theorem 2.9, $h(n) = \frac{n}{1-\frac{\mu_h(n)}{k}}$, it follows that h is also a non-decreasing function. Applying a result of Moser and Lambeck [11], the function h has the form $h(n) = n^\beta$ for some $\beta \in \mathbb{R}^{\geq 0}$. The equality defining μ_h follows immediately and, since μ is a non-decreasing function we also have $\beta > 1$. ∎

Corollary 2.11 *If $\mathcal{H}_h$ is a non-Shannon entropy defined by a (Φ, h)-system of axioms and $\pi \in \mathsf{PART}(S)$, where $\pi = \{B_1, \ldots, B_m\}$, then there exists a constant $k \in \mathbb{R}$ such that*

$$\mathcal{H}_h(\pi) = k\left[1 - \sum_{i=1}^{n}\left(\frac{|B_i|}{|S|}\right)^\beta\right] \tag{3}$$

for some $\beta > 1$.

Proof. We saw that for non-Shannon entropy, the function h is necessarily of the form $h(n) = n^\beta$ for $n \in \mathbb{N}$ and $\beta > 1$. By Corollary 2.6 we have

$$\begin{aligned}
\mathcal{H}_h(\pi) &= \mu_h(|S|) - \sum_{i=1}^{m} \left(\frac{|B_i|}{|S|}\right)^\beta \mu_h(|B_i|) \\
&= k(1 - |S|^{1-\beta}) - k \sum_{i=1}^{m} \left(\frac{|B_i|}{|S|}\right)^\beta (1 - |B_i|^{1-\beta}) \\
&= k(1 - |S|^{1-\beta}) - k \sum_{i=1}^{m} \left(\frac{|B_i|}{|S|}\right)^\beta + k|S|^{1-\beta} \\
&= k\left[1 - \sum_{i=1}^{n} \left(\frac{|B_i|}{|S|}\right)^\beta\right].
\end{aligned}$$

∎

The next theorem shows that the function Φ introduced by Definition 2.2 and used in Axiom **(P4)** is essentially determined by the choice made for h.

Theorem 2.12 *Let $\mathcal{H}_h$ be the non-Shannon entropy defined by a (Φ, h)-system, where $h(n) = n^\beta$ for some $\beta > 1$.*

The function Φ of Axiom **(P4)** *is given by*

$$\Phi(x, y) = x + y - \frac{1}{k}xy$$

for $x, y \in \mathbb{R}_{\geq 0}$.

Proof. Let $\pi = \{B_1, \ldots, B_m\} \in \mathsf{PART}(S)$ and $\sigma = \{C_1, \ldots, C_n\} \in \mathsf{PART}(T)$ be two partitions. Since

$$\begin{aligned}
\sum_{i=1}^{m} \left(\frac{|B_i|}{|S|}\right)^\beta &= 1 - \frac{1}{k}\mathcal{H}_h(\pi) \\
\sum_{j=1}^{n} \left(\frac{|C_j|}{|T|}\right)^\beta &= 1 - \frac{1}{k}\mathcal{H}_h(\sigma)
\end{aligned}$$

we can write:

$$\begin{aligned}
\mathcal{H}_h(\pi \times \sigma) &= k\left(1 - \sum_{i=1}^{m}\sum_{j=1}^{n} \left(\frac{|B_i||C_j|}{|S||T|}\right)^\beta\right) \\
&= k\left(1 - \left(1 - \frac{1}{k}\mathcal{H}_h(\pi)\right)\left(1 - \frac{1}{k}\mathcal{H}_h(\sigma)\right)\right) \\
&= \mathcal{H}_h(\pi) + \mathcal{H}_h(\sigma) - \frac{1}{k}\mathcal{H}_h(\pi)\mathcal{H}_h(\sigma).
\end{aligned}$$

Since $\beta > 1$ the set of rational numbers of the form

$$1 - \sum_{l=1}^{n} r_l^\beta,$$

where $r_l \in \mathbb{Q}$, $0 \leq r_l \leq 1$ for $1 \leq l \leq n$ and $\sum_{l=1}^{n} r_l = 1$, for some $n \in \mathbb{N}_1$, is dense in the interval $[0,1]$. Thus, formula (3) shows that the set of entropy values is dense in the interval $[0,k]$ because the sets $B_1, \ldots, B_m$ are finite but of arbitrarily large cardinalities. Since the set of values of entropies is dense in the interval $[0,k]$, the continuity of Φ implies the desired form of Φ. ∎

Choosing $k = \frac{1}{1-2^{1-\beta}}$ in the equality (3) we obtain the Havrda-Charvat entropy (see [5]):

$$\mathcal{H}_h(\pi) = \frac{1}{1-2^{1-\beta}} \cdot \left(1 - \sum_{i=1}^{m} \left(\frac{|B_i|}{|S|}\right)^{\beta}\right).$$

If $\beta = 2$ we obtain $\mathcal{H}_2(\pi)$ which is twice the Gini index,

$$\mathcal{H}_h(\pi) = 2 \cdot \left(1 - \sum_{i=1}^{m} \left(\frac{|B_i|}{|S|}\right)^{2}\right).$$

The *Gini index*, $\mathsf{gini}(\pi) = 1 - \sum_{i=1}^{m} \left(\frac{|B_i|}{|S|}\right)^2$ is widely used in machine learning and data mining.

The limit case, $\lim_{\beta \to 1} \mathcal{H}_h(\pi)$ yields

$$\begin{aligned}
\lim_{\beta \to 1} \mathcal{H}_h(\pi) &= \lim_{\beta \to 1} \frac{1}{1-2^{1-\beta}} \cdot \left(1 - \sum_{i=1}^{m} \left(\frac{|B_i|}{|S|}\right)^{\beta}\right) \\
&= \lim_{\beta \to 1} \frac{1}{2^{1-\beta} \ln 2} \cdot \left(- \sum_{i=1}^{m} \left(\frac{|B_i|}{|S|}\right)^{\beta} \ln \frac{|B_i|}{|S|}\right) \\
&= - \sum_{i=1}^{m} \frac{|B_i|}{|S|} \log_2 \frac{|B_i|}{|S|},
\end{aligned}$$

which is the Shannon entropy of π.

When $\beta = 1$, by Theorem 2.9, we have

$$\mu(ab) = \mu(a) + \mu(b)$$

for $a, b \in \mathbb{N}_1$. If $\eta : \mathbb{N}_1 \longrightarrow \mathbb{R}$ is the function defined by $\eta(a) = a\mu(a)$ for $a \in \mathbb{N}_1$, then η is clearly an increasing function and we have

$$\eta(ab) = ab\mu(ab) = b\eta(a) + a\eta(b)$$

for $a, b \in \mathbb{N}_1$. By Theorem A.6 , there exists a constant $c \in \mathbb{R}$ such that $\eta(a) = ca \log_2 a$ for $a \in \mathbb{N}_1$, so $\mu(a) = c \log_2(a)$. Then, equation (1) implies:

$$\mathcal{H}_h(\pi) = c \cdot \sum_{i=1}^{m} \frac{a_i}{a} \log_2 \frac{a_i}{a},$$

for every partition $\pi = \{A_1, \ldots, A_m\}$ of a set A, where $|A_i| = a_i$ for $1 \leq i \leq m$, and $|A| = a$. This is exactly the expression of Shannon's entropy.

The continuous function Φ is determined, as in the previous case. Indeed, if A, B are two sets such that $|A| = a$ and $|B| = b$, then we must have

$$c \cdot \log_2 ab = \mathcal{H}_h(\alpha_A \times \alpha_B) = \Phi(c \cdot \log_2 a, c \cdot \log_2 b)$$

for any $a, b \in \mathbb{N}_1$ and any $c \in \mathbb{R}$. The continuity of Φ implies $\Phi(x, y) = x + y$.

If $h(n) = n^\beta$ for $n \in \mathbb{N}$ and $\beta > 1$, we shall refer to the $\mathcal{H}_h(\pi)$ as the *β-entropy* of π and will denote this entropy from now on by $\mathcal{H}_\beta(\pi)$.

3 Generalized Conditional Entropies

The entropies previously introduced generate corresponding conditional entropies.

Let $\pi \in \mathsf{PART}(S)$ and let $C \subseteq S$. Denote by π_C the "trace" of π on C given by

$$\pi_C = \{B \cap C | B \in \pi \text{ such that } B \cap C \neq \emptyset\}.$$

Clearly, $\pi_C \in \mathsf{PART}(C)$; also, if C is a block of π, then $\pi_C = \omega_C$.

Definition 3.1 Let $\pi, \sigma \in \mathsf{PART}(S)$ and let $\sigma = \{C_1, \ldots, C_n\}$. The *$\beta$-conditional entropy* of the partitions $\pi, \sigma \in \mathsf{PART}(S)$ (where $\beta > 1$) is the function $\mathcal{H}_\beta : \mathsf{PART}(S)^2 \longrightarrow \mathbb{R}_{\geq 0}$ defined by:

$$\mathcal{H}_\beta(\pi|\sigma) = \sum_{j=1}^{n} \left(\frac{|C_j|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{C_j})$$

$\square$

Observe that $\mathcal{H}_\beta(\pi|\omega_S) = \mathcal{H}_\beta(\pi)$ and that $\mathcal{H}_\beta(\omega_S|\pi) = \mathcal{H}_\beta(\pi|\alpha_S) = 0$ for every partition $\pi \in \mathsf{PART}(S)$.

For $\pi = \{B_1, \ldots, B_m\}$ and $\sigma = \{C_1, \ldots, C_n\}$ the conditional entropy can be written explicitly as:

$$\begin{aligned} \mathcal{H}_\beta(\pi|\sigma) &= \sum_{j=1}^{n} \left(\frac{|C_j|}{|S|}\right)^\beta \sum_{i=1}^{m} \frac{1}{1 - 2^{1-\beta}} \left[1 - \left(\frac{|B_i \cap C_j|}{|C_j|}\right)^\beta\right] \\ &= \frac{1}{1 - 2^{1-\beta}} \sum_{j=1}^{n} \left(\left(\frac{|C_j|}{|S|}\right)^\beta - \sum_{i=1}^{m} \left(\frac{|B_i \cap C_j|}{|S|}\right)^\beta\right). \end{aligned} \tag{4}$$

For the special case when $\pi = \alpha_S$ we can write

$$\mathcal{H}_\beta(\alpha_S|\sigma) = \sum_{j=1}^{n} \left(\frac{|C_j|}{|S|}\right)^\beta \mathcal{H}_\beta(\alpha_{C_j}) = \frac{1}{1 - 2^{1-\beta}} \left(\sum_{j=1}^{n} \left(\frac{|C_j|}{|S|}\right)^\beta - \frac{1}{|S|^{\beta-1}}\right). \tag{5}$$

Theorem 3.2 *Let π, σ be two partitions of a finite set S. We have $\mathcal{H}_\beta(\pi|\sigma) = 0$ if and only if $\sigma \leq \pi$.*

Proof. Suppose that $\sigma = \{C_1, \ldots, C_n\}$. If $\sigma \leq \pi$, then $\pi_{C_j} = \omega_{C_j}$ for $1 \leq j \leq n$ and, therefore,

$$\mathcal{H}_\beta(\pi|\sigma) = \sum_{j=1}^{n} \left(\frac{|C_j|}{|S|}\right)^\beta \mathcal{H}_\beta(\omega_{C_j}) = 0.$$

Conversely, suppose that

$$\mathcal{H}_\beta(\pi|\sigma) = \sum_{j=1}^{n} \left(\frac{|C_j|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{C_j}) = 0.$$

This implies $\mathcal{H}_\beta(\pi_{C_j}) = 0$ for $1 \leq j \leq n$, which means that $\pi_{C_j} = \omega_{C_j}$ for $1 \leq j \leq n$ by a previous remark. This means that every block C_j of σ is included in a block of π. so $\sigma \leq \pi$. ∎

The next statement is a generalization of a well-known property of Shannon's entropy.

Theorem 3.3 *Let π, σ be two partitions of a finite set S. We have:*

$$\mathcal{H}_\beta(\pi \wedge \sigma) = \mathcal{H}_\beta(\pi|\sigma) + \mathcal{H}_\beta(\sigma) = \mathcal{H}_\beta(\sigma|\pi) + \mathcal{H}_\beta(\pi),$$

Proof. Suppose that $\pi = \{B_1, \ldots, B_m\}$ and that $\sigma = \{C_1, \ldots, C_n\}$. Observe that

$$\pi \wedge \sigma = \pi_{C_1} + \cdots + \pi_{C_n} = \sigma_{B_1} + \cdots + \sigma_{B_m}.$$

Therefore, by Corollary 2.8 we have:

$$\mathcal{H}_\beta(\pi \wedge \sigma) = \sum_{j=1}^{n} \left(\frac{|C_j|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{C_j}) + \mathcal{H}_\beta(\sigma),$$

which implies

$$\mathcal{H}_\beta(\pi \wedge \sigma) = \mathcal{H}_\beta(\pi|\sigma) + \mathcal{H}_\beta(\sigma).$$

The second equality has a similar proof. ∎

Corollary 3.4 *If $\mathcal{H}_\beta(\pi \wedge \sigma) = \mathcal{H}_\beta(\pi)$, then $\pi \leq \sigma$.*

Proof. Since $\mathcal{H}_\beta(\pi \wedge \sigma) = \mathcal{H}_\beta(\pi)$, Theorem 3.3 implies $\mathcal{H}_\beta(\sigma|\pi) = 0$. By Theorem 3.2 we have $\pi \leq \sigma$. ∎

Lemma 3.5 *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:

$$x^\beta + (1-x)^\beta \leq 1 \tag{6}$$

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

Inequality (6) implies

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

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

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

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

Thus, we can write

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

which is desired inequality. ∎

Theorem 3.6 *Let S be a set, $\pi \in \mathsf{PART}(S)$ and let C, D be two disjoint subsets of S. For $\beta \geq 1$ we have:*

$$\left(\frac{|C \cup D|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{C\cup D}) \geq \left(\frac{|C|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_C) + \left(\frac{|D|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_D).$$

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 3.5 to the numbers $w_1, \ldots, w_m$ and $a_1, \ldots, a_m$ we obtain:

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

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:

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

which is equivalent to

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

which yields the inequality of the theorem. ∎

The next result shows that the β-conditional entropy is dually monotonic with respect to its first argument and is monotonic with respect to its second argument.

Theorem 3.7 *Let $\pi, \sigma, \sigma' \in \mathsf{PART}(S)$, where S is a finite set. If $\sigma \leq \sigma'$, then $\mathcal{H}_\beta(\sigma|\pi) \geq \mathcal{H}_\beta(\sigma'|\pi)$ and $\mathcal{H}_\beta(\pi|\sigma) \leq \mathcal{H}_\beta(\pi|\sigma')$.*

Proof. Since $\sigma \leq \sigma'$ we have $\pi \wedge \sigma \leq \pi \wedge \sigma'$, so $\mathcal{H}_\beta(\pi \wedge \sigma) \geq \mathcal{H}_\beta(\pi \wedge \sigma')$. Therefore, $\mathcal{H}_\beta(\sigma|\pi) + \mathcal{H}_\beta(\pi) \geq \mathcal{H}_\beta(\sigma'|\pi) + \mathcal{H}_\beta(\pi)$, which implies $\mathcal{H}_\beta(\sigma|\pi) \geq \mathcal{H}_\beta(\sigma'|\pi)$.

For the second part of the theorem it suffices to prove the inequality 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:

$$\begin{aligned}
&\mathcal{H}_\beta(\pi|\sigma') \\
&= \sum_{j=1}^{n-2} \left(\frac{|C_j|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{C_j}) + \left(\frac{|C_{n-1} \cup C_n|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{C_{n-1} \cup C_n}) \\
&\geq \left(\frac{|C_j|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{C_j}) + \left(\frac{|C_{n-1}|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{C_{n-1}}) + \left(\frac{|C_n|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{C_n}) \\
&\quad \text{(by Theorem 3.6)} \\
&= \mathcal{H}(\pi|\sigma).
\end{aligned}$$

∎

Corollary 3.8 *We have $\mathcal{H}_\beta(\pi) \geq \mathcal{H}_\beta(\pi|\sigma)$ for every $\pi, \sigma \in \mathsf{PART}(S)$.*

Proof. We observed that $\mathcal{H}_\beta(\pi) = \mathcal{H}_\beta(\pi|\omega_S)$. Since $\omega_S \geq \sigma$ the statement follows from the second part of Theorem 3.7. ∎

Corollary 3.9 *Let ξ, θ, θ' be three partitions of a finite set S. If $\theta \geq \theta'$, then*

$$\mathcal{H}_\beta(\xi \wedge \theta) - \mathcal{H}_\beta(\theta) \geq \mathcal{H}_\beta(\xi \wedge \theta') - \mathcal{H}_\beta(\theta').$$

Proof. By Theorem 3.3 we have:

$$\mathcal{H}_\beta(\xi \wedge \theta) - \mathcal{H}_\beta(\xi \wedge \theta') = \mathcal{H}_\beta(\xi|\theta) + \mathcal{H}_\beta(\theta) - \mathcal{H}_\beta(\xi|\theta') - \mathcal{H}_\beta(\theta').$$

The monotonicity of $\mathcal{H}_\beta(|)$ in its second argument means that: $\mathcal{H}_\beta(\xi|\theta) - \mathcal{H}_\beta(\xi|\theta') \geq 0$, so $\mathcal{H}_\beta(\xi \wedge \theta) - \mathcal{H}_\beta(\xi \wedge \theta') \geq \mathcal{H}_\beta(\theta) - \mathcal{H}_\beta(\theta')$, which implies the desired inequality. ∎

The behavior of β-conditional entropies with respect to the "addition" of partitions is discussed in the next statement.

Theorem 3.10 *Let S be a finite set, π, θ be two partitions of S, where $\theta = \{D_1, \ldots, D_h\}$. If $\sigma_i \in \mathsf{PART}(D_i)$ for $1 \leq i \leq h$, then*

$$\mathcal{H}_\beta(\pi|\sigma_1 + \cdots + \sigma_h) = \sum_{i=1}^{h} \left(\frac{|D_i|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{D_i}|\sigma_i).$$

If $\tau = \{F_1, \ldots, F_k\}$, $\sigma = \{C_1, \ldots, C_n\}$ be two partitions of S, and let $\pi_i \in \mathsf{PART}(F_i)$ for $1 \leq i \leq k$. Then,

$$\mathcal{H}_\beta(\pi_1 + \cdots + \pi_k|\sigma) = \sum_{i=1}^{k} \left(\frac{|F_i|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_i|\sigma_{F_i}) + \mathcal{H}_\beta(\tau|\sigma).$$

Proof. Suppose that $\sigma_i = \{E_i^\ell \mid 1 \leq \ell \leq p_i\}$. The blocks of the partition $\sigma_1 + \cdots + \sigma_h$ are the sets of the collection $\bigcup_{i=1}^h \{E_i^\ell \mid 1 \leq \ell \leq p_i\}$. Thus, we have:

$$\mathcal{H}_\beta(\pi|\sigma_1 + \cdots + \sigma_h) = \sum_{i=1}^h \sum_{\ell=1}^{p_i} \left(\frac{|E_i^\ell|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{E_i^\ell}).$$

On the other hand, since $(\pi_{D_i})_{E_i^\ell} = \pi_{E_i^\ell}$, we have:

$$\begin{aligned}
\sum_{i=1}^h \left(\frac{|D_i|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{D_i}|\sigma_i) &= \sum_{i=1}^h \left(\frac{|D_i|}{|S|}\right)^\beta \sum_{\ell=1}^{p_i} \left(\frac{|E_i^\ell|}{|D_i|}\right)^\beta \mathcal{H}_\beta(\pi_{E_i^\ell}) \\
&= \sum_{i=1}^h \sum_{\ell=1}^{p_i} \left(\frac{|E_i^\ell|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{E_i^\ell}),
\end{aligned}$$

which gives the first equality of the theorem.

To prove the second part observe that $(\pi_1 + \cdots + \pi_k)_{C_j} = (\pi_1)_{C_j} + \cdots + (\pi_k)_{C_j}$ for every block C_j of σ. Thus, we have:

$$\mathcal{H}_\beta((\pi_1 + \cdots + \pi_k|\sigma) = \sum_{j=1}^n \left(\frac{|C_j|}{|S|}\right)^\beta \mathcal{H}_\beta((\pi_1)_{C_j} + \cdots + (\pi_k)_{C_j}).$$

By applying Corollary 2.8 to partitions $(\pi_1)_{C_j}, \ldots, (\pi_k)_{C_j}$ of C_j we can write:

$$\mathcal{H}_\beta((\pi_1)_{C_j} + \cdots + (\pi_k)_{C_j}) = \sum_{i=1}^k \left(\frac{|F_i \cap C_j|}{|C_j|}\right)^\beta \mathcal{H}_\beta((\pi_i)_{C_j}) + \mathcal{H}_\beta(\tau_{C_j}).$$

Thus,

$$\begin{aligned}
&\mathcal{H}_\beta(\pi_1 + \cdots + \pi_k|\sigma) \\
&= \sum_{j=1}^n \sum_{i=1}^k \left(\frac{|F_i \cap C_j|}{|S|}\right)^\beta \mathcal{H}_\beta((\pi_i)_{C_j}) + \sum_{j=1}^n \left(\frac{|C_j|}{|S|}\right)^\beta \mathcal{H}_\beta(\tau_{C_j}) \\
&= \sum_{i=1}^k \left(\frac{|F_i|}{|S|}\right)^\beta \sum_{j=1}^n \left(\frac{|F_i \cap C_j|}{|F_i|}\right)^\beta \mathcal{H}_\beta((\pi_i)_{F_i \cap C_j}) + \mathcal{H}_\beta(\tau|\sigma) \\
&= \sum_{i=1}^k \left(\frac{|F_i|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_i|\sigma_{F_i}) + \mathcal{H}_\beta(\tau|\sigma),
\end{aligned}$$

which is the desired equality. ∎

Theorem 3.11 *Let π, σ, τ be three partitions of the finite set S. We have:*

$$\mathcal{H}_\beta(\pi|\sigma \wedge \tau) + \mathcal{H}_\beta(\sigma|\tau) = \mathcal{H}_\beta(\pi \wedge \sigma|\tau).$$

Proof. By Theorem 3.3 we can write

$$\begin{aligned} \mathcal{H}_\beta(\pi|\sigma \wedge \tau) &= \mathcal{H}_\beta(\pi \wedge \sigma \wedge \tau) - \mathcal{H}_\beta(\sigma \wedge \tau) \\ \mathcal{H}_\beta(\sigma|\tau) &= \mathcal{H}_\beta(\sigma \wedge \tau) - \mathcal{H}_\beta(\tau). \end{aligned}$$

By adding these equalities and applying again Theorem 3.3 we obtain the equality of the theorem. ∎

Corollary 3.12 *Let π, σ, τ be three partitions of the finite set S. Then, we have:*

$$\mathcal{H}_\beta(\pi|\sigma) + \mathcal{H}_\beta(\sigma|\tau) \geq \mathcal{H}_\beta(\pi|\tau).$$

Proof. By Theorem 3.11, the monotonicity of β-conditional entropy in its second argument and the anti-monotonicity of the same in its first argument we can write:

$$\begin{aligned} \mathcal{H}_\beta(\pi|\sigma) + \mathcal{H}_\beta(\sigma|\tau) &\geq \mathcal{H}_\beta(\pi|\sigma \wedge \tau) + \mathcal{H}_\beta(\sigma|\tau) \\ &= \mathcal{H}_\beta(\pi \wedge \sigma|\tau) \\ &\geq \mathcal{H}_\beta(\pi|\tau), \end{aligned}$$

which is the desired inequality. ∎

Corollary 3.13 *Let π, σ be two partitions of the finite set S. Then, we have:*

$$\mathcal{H}_\beta(\pi \vee \sigma) + \mathcal{H}_\beta(\pi \wedge \sigma) \leq \mathcal{H}_\beta(\pi) + \mathcal{H}_\beta(\sigma).$$

Proof. By Corollary 3.12 we have $\mathcal{H}_\beta(\pi|\sigma) \leq \mathcal{H}_\beta(\pi|\tau) + \mathcal{H}_\beta(\tau|\sigma)$. Then, by Theorem 3.3 we obtain

$$\mathcal{H}_\beta(\pi \wedge \sigma) - \mathcal{H}_\beta(\sigma) \leq \mathcal{H}_\beta(\pi \wedge \tau) - \mathcal{H}_\beta(\tau) + \mathcal{H}_\beta(\tau \wedge \sigma) - \mathcal{H}_\beta(\sigma),$$

hence

$$\mathcal{H}_\beta(\tau) + \mathcal{H}_\beta(\pi \wedge \sigma) \leq \mathcal{H}_\beta(\pi \wedge \tau) + \mathcal{H}_\beta(\tau \wedge \sigma).$$

Choosing $\tau = \pi \vee \sigma$ implies immediately the inequality of the Corollary. ∎

The property of $\mathcal{H}_\beta$ described in Corollary 3.13 is known as the *submodularity* of the generalized entropy. This result generalizes the modularity of the Gini index proven in [15] and gives an elementary proof of a result shown in [8] concerning Shannon's entropy.

4 Generalized Entropic Metrics and Their Axiomatization

In [6] L. de Mántaras proved that Shannon's entropy generates a metric $d : \mathsf{PART}(S)^2 \longrightarrow \mathbb{R}^2$ given by $d(\pi, \sigma) = \mathcal{H}(\pi|\sigma) + \mathcal{H}(\sigma|\pi)$, for $\pi, \sigma \in \mathsf{PART}(S)$. His result can be extended to a class of metrics that can be defined by β-entropies, thereby improving our earlier results [17].

We can show now a central result:

Theorem 4.1 *The mapping $d_\beta : \mathsf{PART}(S)^2 \longrightarrow \mathbb{R}_{\geq 0}$ defined by: $d_\beta(\pi,\sigma) = \mathcal{H}_\beta(\pi|\sigma) + \mathcal{H}_\beta(\sigma|\pi)$ for $\pi,\sigma \in \mathsf{PART}(S)$ is a metric on $\mathsf{PART}(S)$.*

Proof. A double application of Corollary 3.12 yields:

$$\begin{aligned}\mathcal{H}_\beta(\pi|\sigma) + \mathcal{H}_\beta(\sigma|\tau) &\geq \mathcal{H}_\beta(\pi|\tau),\\ \mathcal{H}_\beta(\sigma|\pi) + \mathcal{H}_\beta(\tau|\sigma) &\geq \mathcal{H}_\beta(\tau|\pi).\end{aligned}$$

Adding these inequality gives

$$d_\beta(\pi,\sigma) + d_\beta(\sigma,\tau) \geq d_\beta(\pi,\tau),$$

which is the triangular inequality for d_β.

The symmetry of d_β is obvious and it is clear that $d_\beta(\pi,\pi) = 0$ for every $\pi \in \mathsf{PART}(S)$.

Suppose now that $d_\beta(\pi,\sigma) = 0$. Since the values of β-conditional entropies are non-negative this implies $\mathcal{H}_\beta(\pi|\sigma) = \mathcal{H}_\beta(\sigma|\pi) = 0$. By Theorem 3.2 we have both $\sigma \leq \pi$ and $\pi \leq \sigma$, respectively, so $\pi = \sigma$. Thus, d_β is a metric on $\mathsf{PART}(S)$. ∎

It is clear that $d_\beta(\pi,\omega_S) = \mathcal{H}_\beta(\pi)$ and $d_\beta(\pi,\alpha_S) = \mathcal{H}_\beta(\alpha_S|\pi)$.

Another useful form of d_β can be obtained starting from the equalities Since $\mathcal{H}_\beta(\pi|\sigma) = \mathcal{H}_\beta(\pi \wedge \sigma) - \mathcal{H}_\beta(\sigma)$ and $\mathcal{H}_\beta(\sigma|\pi) = \mathcal{H}_\beta(\pi \wedge \sigma) - \mathcal{H}_\beta(\sigma)$. Thus, we have:

$$d_\beta(\pi,\sigma) = 2\mathcal{H}_\beta(\pi \wedge \sigma) - \mathcal{H}_\beta(\pi) - \mathcal{H}_\beta(\sigma), \tag{7}$$

for $\pi,\sigma \in \mathsf{PART}(S)$.

The behavior of the distance d_β with respect to partition addition is discussed in the next statement.

Theorem 4.2 *Let S be a finite set, π,θ be two partitions of S, where $\theta = \{D_1,\ldots,D_h\}$. If $\sigma_i \in \mathsf{PART}(D_i)$ for $1 \leq i \leq h$, then*

$$d_\beta(\pi,\sigma_1 + \cdots + \sigma_h) = \sum_{i=1}^{h} \left(\frac{|D_i|}{|S|}\right)^\beta d_\beta(\pi_{D_i},\sigma_i) + \mathcal{H}_\beta(\theta|\pi).$$

Proof. This statement follows directly from Theorem 3.10. ∎

The next statement is a generalization of the axiom system proposed in [9] for the Shannon entropic metric and for the Mirkin metric.

Theorem 4.3 *The following properties hold in the metric space $(\mathsf{PART}(S), d_\beta)$:*

1. *if $\sigma \leq \pi$, then $d_\beta(\pi,\sigma) = \mathcal{H}_\beta(\sigma) - \mathcal{H}_\beta(\pi)$;*
2. *$d_\beta(\alpha_S,\sigma) + d_\beta(\sigma,\omega_S) = d_\beta(\alpha_S,\omega_S)$;*
3. *$d_\beta(\pi,\pi\wedge\sigma) + d_\beta(\pi\wedge\sigma,\sigma) = d_\beta(\pi,\sigma)$,*

for every partitions $\pi,\sigma \in \mathsf{PART}(S)$.

Furthermore, we have $d(\omega_T,\alpha_T) = \frac{1-|T|^{1-\beta}}{1-2^{1-\beta}}$, for every subset T of S.

Proof. The first three statements of the theorem follow immediately from Equality 7; the last part is an application of the definition of d_β. ∎

A generalization of a result obtained in [9] is contained in the next statement, which gives an axiomatization of the metric d_β.

Theorem 4.4 *Let $d : \mathsf{PART}(S)^2 \longrightarrow \mathbb{R}_{\geq 0}$ be a function that satisfies the following conditions:*

(D1) *d is symmetric, that is, $d(\pi, \sigma) = d(\sigma, \pi)$;*

(D2) *$d(\alpha_S, \sigma) + d(\sigma, \omega_S) = d(\alpha_S, \omega_S)$;*

(D3) *$d(\pi, \sigma) = d(\pi, \pi \wedge \sigma) + d(\pi \wedge \sigma, \sigma)$;*

(D4) *if $\sigma, \theta \in \mathsf{PART}(S)$ such that $\theta = \{D_1, \ldots, D_h\}$ and $\sigma \leq \theta$ then we have:*

$$d(\theta, \sigma) = \sum_{i=1}^{h} \left(\frac{|D_i|}{|S|}\right)^\beta d(\omega_{D_i}, \sigma_{D_i});$$

(D5) *$d(\omega_T, \alpha_T) = \frac{1-|T|^{1-\beta}}{1-2^{1-\beta}}$, for every $T \subseteq S$.*

Then, $d = d_\beta$.

Proof. Choosing $\sigma = \alpha_S$ in axiom **(D4)** and using **(D5)** we can write:

$$\begin{aligned} d(\alpha_S, \theta) &= \sum_{i=1}^{h} \left(\frac{|D_i|}{|S|}\right)^\beta d(\omega_{D_i}, \alpha_{D_i}) \\ &= \sum_{i=1}^{h} \left(\frac{|D_i|}{|S|}\right)^\beta \frac{1-|D_i|^{1-\beta}}{1-2^{1-\beta}} \\ &= \frac{\sum_{i=1}^{h} |D_i|^\beta - |S|}{(1-2^{1-\beta})|S|^\beta}. \end{aligned}$$

From Axioms **(D2)** and **(D5)** it follows that

$$\begin{aligned} d(\theta, \omega_S) &= d(\alpha_S, \omega_S) - d(\alpha_S, \theta) \\ &= \frac{1-|S|^{1-\beta}}{1-2^{1-\beta}} - \frac{\sum_{i=1}^{h} |D_i|^\beta - |S|}{(1-2^{1-\beta})|S|^\beta} \\ &= \frac{|S|^\beta - \sum_{i=1}^{h} |D_i|^\beta}{(1-2^{1-\beta})|S|^\beta}. \end{aligned}$$

Let now $\pi, \sigma \in \mathsf{PART}(S)$, where $\pi = \{B_1, \ldots, B_m\}$ and $\sigma = \{C_1, \ldots, C_n\}$. Since $\pi \wedge \sigma \leq \pi$ and $\sigma_{B_i} = \{C_1 \cap B_i, \ldots, C_n \cap B_i\}$, an application of Axiom **(D4)** yields

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

because $(\pi \wedge \sigma)_{B_i} = \sigma_{B_i}$. ∎

By Axiom **(D1)** we obtain the similar equality:

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

which, by Axiom **(D3)**, implies:

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

that is $d(\pi, \sigma) = d_{\beta}(\pi, \sigma)$. ∎

In fact, the Mirkin metric [10] (up to a multiplicative constant) is obtained for $\beta = 2$:

$$\begin{aligned}
d_2(\pi, \sigma) &= \frac{2}{|S|^2} \left(\sum_{i=1}^{m} |B_i|^2 + \sum_{j=1}^{n} |C_j|^2 \right. \\
& \qquad \left. -2 \sum_{i=1}^{m} \sum_{j=1}^{n} |B_i \cap C_j|^2 \right).
\end{aligned}$$

The corresponding generalized entropy $\mathcal{H}_2(\pi)$ is double the Gini index of the partition $\pi = \{B_1, \ldots, B_m\}$:

$$\mathcal{H}_2(\pi) = 2 \left(\sum_{i=1}^{m} \left(\frac{|B_i|}{|S|} \right)^2 - 1 \right)$$

It is worth noting that one could axiomatize the entropy starting from the notion metric between partitions. Indeed, if the β-entropy of a partition $\pi \in \mathsf{PART}(S)$ is defined as:

$$\mathcal{H}_\beta(\pi) = d_\beta(\pi, \omega_S),$$

then we would retrieve the β-entropy:

$$\mathcal{H}_\beta(\pi) = \frac{1}{1 - 2^{1-\beta}} \left(1 - \sum_{i=1}^{m} \left(\frac{|B_i|}{|S|}\right)^\beta\right).$$

5 Partition Valuations and β-Entropy

Metrics generated by β-conditional entropies are closely related to lower valuations of the upper semi-modular lattices of partitions of finite sets. This connection was established in [3] and studied in [2, 1, 12].

A *lower valuation* on a lattice $(L, \vee, \wedge)$ is a mapping $v : L \longrightarrow \mathbb{R}$ such that $v(\pi \vee \sigma) + v(\pi \wedge \sigma) \geq v(\pi) + v(\sigma)$ for every $\pi, \sigma \in L$. If the reverse inequality is satisfied, that is, if $v(\pi \vee \sigma) + v(\pi \wedge \sigma) \leq v(\pi) + v(\sigma)$ for every $\pi, \sigma \in L$, then v is referred to as an *upper valuation*.

If $v \in L$ is both a lower and upper valuation, that is, if $v(\pi \vee \sigma) + v(\pi \wedge \sigma) = v(\pi) + v(\sigma)$ for every $\pi, \sigma \in L$, then v is a valuation on L.

We have the following result:

Theorem 5.1 *Let $\pi, \sigma \in \mathsf{PART}(S)$ be two partitions. We have:*

$$\begin{aligned} d_\beta(\pi, \sigma) &= 2 \cdot d_\beta(\pi \wedge \sigma, \omega_S) - d_\beta(\pi, \omega_S) - d_\beta(\sigma, \omega_S) \\ &= d_\beta(\alpha_S, \pi) + d_\beta(\alpha_S, \sigma) - 2 \cdot d_\beta(\alpha_S, \pi \wedge \sigma). \end{aligned}$$

Proof. The equalities of the theorem can be immediately verified by using the definition of d_β. ∎

Corollary 5.2 *Let θ, τ be two partitions from $\mathsf{PART}(S)$. If $\theta \leq \tau$ and we have either $d_\beta(\theta, \omega_S) = d_\beta(\tau, \omega_S)$ or $d_\beta(\alpha_S, \theta) = d_\beta(\alpha_S, \tau)$, then $\theta = \tau$.*

Proof. Observe that if $\theta \leq \tau$, then Theorem 5.1 implies

$$d_\beta(\theta, \tau) + d_\beta(\tau, \omega_S) = d_\beta(\theta, \omega_S),$$

and

$$d_\beta(\theta, \tau) = d_\beta(\alpha_S, \tau) - d_\beta(\alpha_S, \theta).$$

Suppose that $d_\beta(\theta, \omega_S) = d_\beta(\tau, \omega_S)$. Since $d_\beta(\tau, \omega_S) = d_\beta(\theta, \omega_S)$ it follows that $d_\beta(\theta, \tau) = 0$, so $\theta = \tau$.

If $d_\beta(\alpha_S, \theta) = d_\beta(\alpha_S, \tau)$ the same conclusion can be reached immediately. ∎

It is known [3] that if there exists a positive valuation v on L, then L must be a modular lattice. Since the partition lattice of a set is an upper-semimodular lattice that is not modular ([3]) it is clear that positive valuations do not exist on partition lattices. However, lower and upper valuations do exist, as shown next:

Theorem 5.3 *Let S be a finite set. Define the mappings $v_\beta : \mathsf{PART}(S) \longrightarrow \mathbb{R}$ and let $w_\beta : \mathsf{PART}(S) \longrightarrow \mathbb{R}$ be by $v_\beta(\pi) = d_\beta(\alpha_S, \pi)$ and $w_\beta(\pi) = d_\beta(\pi, \omega_S)$, respectively, for $\pi \in \mathsf{PART}(S)$. Then, v_β is a lower valuation and w_β is an upper valuation on the lattice $(\mathsf{PART}(S), \vee, \wedge)$.*

Proof. Theorem 5.1 allows us to write:

$$\begin{aligned} d_\beta(\pi,\sigma) &= v_\beta(\pi) + v_\beta(\sigma) - 2v_\beta(\pi \wedge \sigma) \\ &= 2w_\beta(\pi \wedge \sigma) - w_\beta(\pi) - w_\beta(\sigma), \end{aligned}$$

for every $\pi, \sigma \in \mathsf{PART}(S)$.

If we rewrite the triangular inequality $d_\beta(\pi,\tau) + d_\beta(\tau,\sigma) \geq d_\beta(\pi,\sigma)$ using the valuations v_β and w_β we obtain:

$$\begin{aligned} v_\beta(\tau) + v_\beta(\pi \wedge \sigma) &\geq v_\beta(\pi \wedge \tau) + v_\beta(\tau \wedge \sigma), \\ w_\beta(\pi \wedge \tau) + w_\beta(\tau \wedge \sigma) &\geq w_\beta(\tau) + w_\beta(\pi \wedge \sigma), \end{aligned}$$

for every $\pi, \tau, \sigma \in \mathsf{PART}(S)$. If we choose $\tau = \pi \vee \sigma$ the last inequalities yield:

$$\begin{aligned} v_\beta(\pi) + v_\beta(\sigma) &\leq v_\beta(\pi \vee \sigma) + v_\beta(\pi \wedge \sigma) \\ w_\beta(\pi) + w_\beta(\sigma) &\geq w_\beta(\pi \vee \sigma) + w_\beta(\pi \wedge \sigma), \end{aligned}$$

for every $\pi, \sigma \in \mathsf{PART}(S)$, which shows that v_β is a lower valuation and w_β is an upper valuation on the lattice $(\mathsf{PART}(S), \vee, \wedge)$. ∎

6 Conclusion and Future Work

We introduced an axiomatization of a generalization of entropy that has as special case Shannon entropy and Havrda-Charvat entropy. Moreover, we have shown, that under certain assumption the Havrda-Charvat entropy is the unique alternative to Shannon's entropy.

A general axiomatization of a family of metrics on the set of partitions of a finite set that is related to generalized entropies was also introduced. These metrics are used for a variety of data mining tasks ranging from clustering [9, 19] to classification [17, 18] and discretization [13].

The value of the parameter β that gives optimal results depends on the statistical properties of the data set that is analyzed. Developing algorithms that learn the values of β for a specific data set and mining task remains an open problem.

References

[1] J. Barthélemy. Remarques sur les propriétés metriques des ensembles ordonnés. *Math. Sci. hum.*, 61:39–60, 1978.

[2] J. Barthélemy and B. Leclerc. The median procedure for partitions. In *Partitioning Data Sets*, pages 3–34. American Mathematical Society, Providence, 1995.

[3] G. Birkhoff. *Lattice Theory*. American Mathematical Society, Providence, 1973.

[4] Z. Daróczy. Generalized information functions. *Information and Control*, 16:36–51, 1970.

[5] J. H. Havrda and F. Charvat. Quantification Methods of Classification Processes: Concepts of Structural α-Entropy. *Kybernetica*, 3:30–35, 1967.

[6] R. L. de Mántaras. A distance-based attribute selection measure for decision tree induction. *Machine Learning*, 6:81–92, 1991.

[7] I. C. Lerman. *Classification et analyse ordinale des données*. Dunod, Paris, 1981.

[8] E. H. Lieb and M. Ruskai. Proof of the strong subadditivity of quantum-mechanical entropy. *Journal of Mathematical Physics*, 14:1938–1941, 1973.

[9] M. Meilă. Comparing clusterings: an axiomatic view. In *Proc. of ICML*, pages 577–584, Bonn, Germany, 2005.

[10] B. Mirkin. *Mathematical Classification and Clustering*. Kluwer, Boston, 1996.

[11] L. Moser and J. Lambek. *On Monotone Multiplicative Functions*. *Proceedings of the American Mathematical Society*, 4:544–545, 1953.

[12] B. Monjardet. Metrics on partially ordered sets – a survey. *Discrete Mathematics*, 35:173–184, 1981.

[13] D. S. S. R. Butterworth, D. A. Simovici and L. Ohno-Machado. A greedy algorithm for supervised discretization. *Journal of Biomedical Informatics*, 37(4):285–292, 2004.

[14] I. Rosenberg and D. A. Simovici. An abstract axiomatization of the notion of entropy. In *Proceedings of ISMVL*, pages 8–13, Calgary, Canada, 2005.

[15] B. Sayrafi. *A Measure-theoretic Framework for Constraints and Bounds on Measurements of Data*. Ph.D. Thesis, Indiana University, 2005.

[16] D. A. Simovici and S. Jaroszewicz. An axiomatization of partition entropy. *IEEE Transactions on Information Theory*, 48:2138–2142, 2002.

[17] 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.

[18] D. A. Simovici and S. Jaroszewicz. A new metric splitting criterion for decision trees. *Journal of Parallel, Emerging and Distributed Computing*, 21:239–256, 2006.

[19] D. A. Simovici, N. Singla, and M. Kuperberg. Metric incremental clustering of nominal data. In *Proceedings of ICDM 2004*, pages 523–527, Brighton, UK, 2004.

[20] A. Zygmund. *Trigonometric Series*. Cambridge University Press, 2nd edition, 1974.

A A Useful Characterization of a Function

The goal of this section is to show that if $h : \mathbb{N} \longrightarrow \mathbb{R}$ is an increasing function such that $h(2) = 2$ and $h(mn) = mh(n) + nh(m)$ for $m, n \in \mathbb{N}$, then $h(n) = n \log_2 n$ for every $n \in \mathbb{N}_1$.

For a number $x \in \mathbb{R}$ we denote the largest integer that is less or equal to x by $\lfloor x \rfloor$; the *fractional part* of x will be denoted by $\langle x \rangle$, where $\langle x \rangle = x - \lfloor x \rfloor$.

The following technical Lemma is a special case of a result of Dirichlet (see[20], pp.235).

Lemma A.1 *Let α be a real number and let q be a positive integer. There exists m such that $1 \leq m < q$ and an integer n such that $|m\alpha - n| < \dfrac{1}{q} < \dfrac{1}{m}$.*

Proof. Divide the set $I = \{x \in \mathbb{R} | 0 \leq x < 1\}$ into q equal subintervals $\left[\frac{i-1}{q}, \frac{i}{q}\right)$ for $1 \leq i \leq q$. Of the $q + 1$ numbers $\langle n\alpha \rangle$, where $0 \leq n \leq q$, at least two, say $\langle n_1\alpha \rangle$ and $\langle n_2\alpha \rangle$ are in the same subinterval. This means $|\langle n_1\alpha \rangle - \langle n_2\alpha \rangle| < \frac{1}{q}$. Setting $\lfloor n_1\alpha \rfloor = n'$ and $\lfloor n_2\alpha \rfloor = n''$ we obtain $|n_1\alpha - \lfloor n_1\alpha \rfloor - n_2\alpha + \lfloor n_2\alpha \rfloor| < \frac{1}{q}$, or $|(n_1 - n_2)\alpha - (n' - n'')| < \frac{1}{q}$. We can take $m = n_1 - n_2 \geq 1$ and $n = n' - m''$ in order to obtain the desired inequality. ∎

Lemma A.2 *If $n \leq m\epsilon < n + \epsilon$, then there exist m', n' such that $n' - \epsilon < m'\epsilon < n'$.*

Similarly, if $n - \epsilon < m\epsilon < n$, there exist n'', m'' such that $n'' \leq m''\epsilon < n'' + \epsilon$.

Proof. Let $\theta, \sigma > 0$. We have $\theta n + \sigma\alpha \leq (\theta m + \sigma)\alpha < \theta n + \theta\epsilon + \sigma\alpha$. Choose $\theta > \max\{1, 1/\epsilon\}$. Under this choice, the interval $[\theta n + \sigma\alpha, \theta n + \theta\epsilon + \sigma\alpha)$ is of length greater than 1 and, therefore, there is an integer m' in this interval for any choice of σ. This allows us to choose θ and σ such that $\theta m + \sigma = m'$ and $\theta n + \theta\epsilon + \sigma\alpha = n'$. Note that if we make these choices, then $\theta n + \sigma\alpha = n' - \theta\epsilon > n' - \epsilon$ and, therefore, $n' - \epsilon \leq m\alpha < n'$.

The second part of the argument is similar and it is left to the reader. ∎

Lemma A.3 *If $h : \mathbb{N} \longrightarrow \mathbb{R}$ is a function such that*

$$h(mn) = mh(n) + nh(m),$$

for every $m, n \in \mathbb{N}$, then $h(p^k) = kp^{k-1}h(p)$ for every $p, k \in \mathbb{N}$ and $k \geq 1$.

Proof. The argument is by induction on k and it is left to the reader. ∎

Lemma A.4 *If $h : \mathbb{N} \longrightarrow \mathbb{R}$ is a function such that*

$$h(mn) = mh(n) + nh(m),$$

for every $m, n \in \mathbf{N}$, then

$$h(p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n}) = p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n} \sum_{1 \leq i \leq n} \frac{k_i h(p_i)}{p_i}.$$

Let $\ell : \mathbb{N} \longrightarrow \mathbb{R}$ be the function given by

$$\ell(n) = \begin{cases} 0 & \text{if } n = 0, \\ \frac{h(n)}{n} & \text{if } n > 0. \end{cases}$$

Note that $\ell(mn) = \ell(m) + \ell(n)$ and, therefore,

$$\ell(p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n}) = \ell(p_1^{k_1}) + \ell(p_2^{k_2}) + \cdots + \ell(p_n^{k_n}).$$

Since $\ell(p^k) = \dfrac{k}{p} h(p)$ (because of Lemma A.3 we obtain:

$$\ell(p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n}) = \sum_{1 \leq i \leq n} \frac{k_i h(p_i)}{p_i},$$

which gives immediately the equality of the Lemma. ∎

Theorem A.5 *Let $h : \mathbb{N} \longrightarrow \mathbb{R}$ be a function such that $h(p) = p \log p$ if $p = 1$ or if p is prime. If $h(mn) = mh(n) + nh(m)$ for every $m, n \in \mathbb{N}$, then $h(n) = n \log n$ for every $n \in \mathbb{N}$, $n \geq 1$.*

Proof. Since every positive integer n other than 1 can be written uniquely as a product of powers of primes $n = p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n}$, we have:

$$\begin{aligned} h(n) &= p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n} \sum_{1 \leq i \leq n} \frac{k_i h(p_i)}{p_i} \\ &= n \sum_{1 \leq i \leq n} k_i \log p_i \\ &= n \log n, \end{aligned}$$

for $n \geq 2$. ∎

Theorem A.6 *Let $h : \mathbb{N} \longrightarrow \mathbb{R}$ be an increasing function such that $h(mn) = mh(n) + nh(m)$ for every $m, n \in \mathbb{N}$. If $h(2) = 2$, then $h(n) = n \log_2 n$ for $n \in \mathbb{N}$.*

Proof. Define the function $b : \{n \in \mathbb{N} | n > 1\} \longrightarrow \mathbb{R}$ by $b(n) = h(n)/(n * \log n)$.

We shall prove initially that if $p > 2$ is a prime number, than $b(p) \geq 1$. Let $\epsilon > 0$ be a real number. Taking $q < 1/\epsilon$ in Lemma A.1 we obtain the existence of $m, n \in \mathbb{N}$ such that $|m\alpha - n| < \epsilon$. In other words, we have $n - \epsilon < m\alpha < n + \epsilon$. If $n < m\alpha < n + \epsilon$, then by Lemma A.2, there are m', n' such that $n' - \epsilon < m'\epsilon < n'$. If $n - \epsilon < m\alpha < n$, then the same lemma implies the existence of n'', m'' such that $n'' \leq m''\epsilon < n'' + \epsilon$.

If we choose $\alpha = \log p$, then we may assume that there are $m, n \in \mathbb{N}$, $m, n \geq 1$ such that $n \leq m \log p < n + \epsilon$. Equivalently, we have $2^n \leq p^m < 2^n\ 2^\epsilon$. Since h is an increasing function, we obtain $n2^n \leq h(p^m)$, or $n2^n \leq mp^{m-1} h(p)$. Because of the definition of b we have $n2^n \leq mp^m b(p) \log p$, or $n2^n \leq b(p) p^m \log p^m$. In view of the previous inequality, this implies

$$n2^n \leq b(p) 2^n 2^\epsilon (n + \epsilon),$$

or, equivalently,

$$b(p) \geq \frac{n}{2^{\epsilon}(n+\epsilon)}.$$

Taking $\epsilon \to 0$ we obtain $b(p) \geq 1$.

Similarly, there exists a number $m \in \mathbb{N}$ such that $n - \epsilon < m \log p \leq n$. A similar argument which makes use of Lemma A.2 shows that $b(p) \leq 1$, so $b(p) = 1$, which proves that $h(p) = p \log p$ for every prime p. ∎