An Axiomatization of Partition Entropy

Dan A. Simovici, Member, IEEE, Szymon Jaroszewicz

Abstract— **The aim of this paper is to present an axiomatization of a generalization of Shannon's entropy starting from partitions of finite sets. The proposed axiomatization yields as special cases the Havrda-Charvat entropy, and thus, provides axiomatizations for the Shannon entropy, the Gini index, and for other types of entropy used in classification and data mining.**

Keywords— **Shannon entropy, Gini index, Havrda-Charvat entropy, non-Shannon entropy.**

I. Introduction and Basic Notations

The notion of partition of a finite set is naturally linked to the notion of probability distribution. Namely, if A is a finite set and $\pi = \{B_1, \ldots, B_n\}$ is a partition of A, then the probability distribution attached to π is $(p_1, \ldots, p_n)$, where $p_i = \frac{|B_i|}{|A|}$ for $1 \leq i \leq n$. Thus, it is natural to consider the notion of entropy of a partition via the entropy of the corresponding probability distribution. Axiomatizations for entropy and entropy-like characteristics of probability distributions represent a problem with a rich history in information theory. Previous relevant work include the results of A.I. Khinchin [11], D.K. Faddeev [5], R.S. Ingarden and K. Urbanik [8], A. Rényi [14] who investigated various axiomatizations of entropy, and Z. Daróczy who presented in [4] an unified treatment of entropy-like characteristics of probability distributions using the notion of information function.

In our previous work (see [15], [9]) we introduced an axiomatization for the notion of functional entropy. This numerical characteristic of functions is related to the complexity of circuits that realize functions (cf.[1]) and serves as an estimate for power dissipation of a circuit realizing a function (cf.[7]) and is linked to the notion of entropy for partitions, since every function $f : A \longrightarrow B$ between the finite sets A, B defines a partition on its definition domain A whose blocks are $\{f^{-1}(b) \mid b \in \mathrm{Ran}(f)\}$.

Information measures, especially conditional entropy of a logic function and its variables, have been used for minimization of logic functions (See [12] and [2]).

In a different direction, starting from the notion of impurity of a set relative to a partition, we found a common generalization of Shannon entropy and of Gini index and we used this generalization in clustering of non-categorial data (see [16]). P. A. Devijer used the Gini index in pattern recognition in [3].

D. A. Simovici and S. Jaroszewicz are with the Univ. of Massachusetts Boston, Dept. of Computer Science, Boston, 02125 USA.
A preliminary form of this paper appeared in the Proceedings of the 31 International Symposium for Multiple-Valued Logic, Warsaw, 2001, pp. 259–268

Our main result is an axiomatization of this generalization that illuminates the common nature of several known ways of evaluating concentrations of values of functions.

All sets considered in the following discussion are nonempty and finite unless stated explicitly otherwise. The sets $\mathbb{R}, \mathbb{R}_{\geq 0}, \mathbb{Q}, \mathbb{N}, \mathbb{N}_1$ denote the set of reals, the set of nonnegative reals, the set of rational numbers, the set of natural numbers, and the set $\{n \in \mathbb{N} \mid n \geq 1\}$, respectively. The domain and range of a function f are denoted by $\mathrm{Dom}(f)$ and $\mathrm{Ran}(f)$ respectively.

Let $\mathsf{PART}(A)$ be the set of partitions of the nonempty set A. The class of all partitions of finite sets is denoted by PART. The one-block partition of A is denoted by ω_A. The partition $\{\{a\} \mid a \in A\}$ is denoted by ι_A.

If $\pi, \pi' \in \mathsf{PART}(A)$, then $\pi \leq \pi'$ if every block of π is included in a block of π'. Clearly, for every $\pi \in \mathsf{PART}(A)$ we have $\iota_A \leq \pi \leq \omega_A$.

If A, B are two disjoint and nonempty sets, $\pi \in \mathsf{PART}(A)$, $\sigma \in \mathsf{PART}(B)$, where $\pi = \{A_1, \ldots, A_m\}$, $\sigma = \{B_1, \ldots, B_n\}$, then the partition $\pi + \sigma$ is the partition of $A \cup B$ given by

$$\pi + \sigma = \{A_1, \ldots, A_m, B_1, \ldots, B_n\}.$$

Whenever the "+" operation is defined, then it is easily seen to be associative. In other words, if A, B, C are pairwise disjoint and nonempty sets, and $\pi \in \mathsf{PART}(A)$, $\sigma \in \mathsf{PART}(B)$, $\tau \in \mathsf{PART}(C)$, then $\pi+(\sigma+\tau) = (\pi+\sigma)+\tau$. Note that if A, B are disjoint, then $\iota_A + \iota_B = \iota_{A\cup B}$. Also, $\omega_A + \omega_B$ is the partition $\{A, B\}$ of the set $A \cup B$.

If $\pi = \{A_1, \ldots, A_m\}$, $\sigma = \{B_1, \ldots, B_n\}$ are partitions of two arbitrary sets, then we denote the partition $\{A_i \times B_j \mid 1 \leq i \leq m, 1 \leq j \leq n\}$ of $A \times B$ by $\pi \times \sigma$. Note that $\iota_A \times \iota_B = \iota_{A\times B}$ and $\omega_A \times \omega_B = \omega_{A\times B}$.

II. An Axiomatization of Generalized Entropy

We introduce below a system of four axioms satisfied by several types of entropy-like characteristics of probability distributions.

Definition II.1: Let $\beta \in \mathbb{R}$, $\beta > 0$, and let $\Phi : \mathbb{R}^2_{\geq 0} \longrightarrow \mathbb{R}_{\geq 0}$ be a continuous function such that $\Phi(x, y) = \Phi(y, x)$, $\Phi(x, 0) = x$ for $x, y \in \mathbb{R}$.

A (Φ, β)*-system of axioms for a partition entropy* $\mathcal{H} : \mathsf{PART}(A) \longrightarrow \{x \in \mathbb{R} \mid x \geq 0\}$ consists of the following axioms:

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

(P2) If A, B are two finite sets such that $|A| \leq |B|$, then $\mathcal{H}(\iota_A) \leq \mathcal{H}(\iota_B)$.

(P3) For every disjoint sets A, B and partitions $\pi \in$

$\mathsf{PART}(A)$, and $\sigma \in \mathsf{PART}(B)$ we have:

$$\begin{aligned}&\mathcal{H}(\pi+\sigma)\\ &= \left(\frac{|A|}{|A|+|B|}\right)^{\beta}\mathcal{H}(\pi)+\left(\frac{|B|}{|A|+|B|}\right)^{\beta}\mathcal{H}(\sigma)\\ &\quad+\mathcal{H}(\{A,B\}).\end{aligned}$$

(P4) If $\pi \in \mathsf{PART}(A)$ and $\sigma \in \mathsf{PART}(B)$, then

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

□

Lemma II.2: For every (Φ,β)-entropy $\mathcal{H}$ and set A we have $\mathcal{H}(\omega_A)=0$.

Proof: Let A,B be two sets that have the same cardinality, $|A|=|B|$. Since $\omega_A+\omega_B$ is the partition $\{A,B\}$ of the set $A\cup B$, by Axiom **(P3)** we have

$$\mathcal{H}(\omega_A+\omega_B)=\left(\frac{1}{2}\right)^{\beta}(\mathcal{H}(\omega_A)+\mathcal{H}(\omega_B))+\mathcal{H}(\{A,B\}),$$

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

Lemma II.3: Let A,B be two disjoint sets and let $\pi,\pi'\in\mathsf{PART}(A\cup B)$ be defined by $\pi=\sigma+\iota_B$ and $\pi'=\sigma+\omega_B$, where $\sigma\in\mathsf{PART}(A)$. Then,

$$\mathcal{H}(\pi)=\mathcal{H}(\pi')+\left(\frac{|B|}{|A|+|B|}\right)^{\beta}\mathcal{H}(\iota_B).$$

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

$$\begin{aligned}\mathcal{H}(\pi) &= \left(\frac{|A|}{|A|+|B|}\right)^{\beta}\mathcal{H}(\sigma)\\ &\quad+\left(\frac{|B|}{|A|+|B|}\right)^{\beta}\mathcal{H}(\iota_B)+\mathcal{H}(\{A,B\}),\end{aligned}$$

and

$$\begin{aligned}\mathcal{H}(\pi') &= \left(\frac{|A|}{|A|+|B|}\right)^{\beta}\mathcal{H}(\sigma)\\ &\quad+\left(\frac{|B|}{|A|+|B|}\right)^{\beta}\mathcal{H}(\omega_B)+\mathcal{H}(\{A,B\})\\ &= \left(\frac{|A|}{|A|+|B|}\right)^{\beta}\mathcal{H}(\sigma)+\mathcal{H}(\{A,B\})\\ &\quad\text{(by Lemma II.2).}\end{aligned}$$

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

Theorem II.4: For every (Φ,β)-entropy and partition $\pi=\{A_1,\ldots,A_n\}\in\mathsf{PART}(A)$ we have:

$$\mathcal{H}(\pi)=\mathcal{H}(\iota_A)-\sum_{j=1}^{n}\left(\frac{|A_j|}{|A|}\right)^{\beta}\mathcal{H}(\iota_{A_j}).$$

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

$$\begin{aligned}\pi_0 &= \omega_{A_1}+\omega_{A_2}+\omega_{A_3}+\cdots+\omega_{A_n}\\ \pi_1 &= \iota_{A_1}+\omega_{A_2}+\omega_{A_3}+\cdots+\omega_{A_n}\\ \pi_2 &= \iota_{A_1}+\iota_{A_2}+\omega_{A_3}+\cdots+\omega_{A_n}\\ &\vdots\\ \pi_n &= \iota_{A_1}+\iota_{A_2}+\iota_{A_3}+\cdots+\iota_{A_n}.\end{aligned}$$

Let $\sigma_j=\iota_{A_1}+\cdots+\iota_{A_j}+\omega_{A_{j+2}}+\cdots+\omega_{A_n}$. Then, $\pi_j=\sigma_j+\omega_{A_{j+1}}$ and $\pi_{j+1}=\sigma_j+\iota_{A_{j+1}}$; therefore, by Lemma II.3, we have

$$\mathcal{H}(\pi_{j+1})=\mathcal{H}(\pi_j)+\left(\frac{|A_{j+1}|}{|A|}\right)^{\beta}\mathcal{H}(\iota_{A_{j+1}})$$

for $0\leq j\leq n-1$.

A repeated application of this equality yields:

$$\mathcal{H}(\pi_n)=\mathcal{H}(\pi_0)+\sum_{j=0}^{n-1}\left(\frac{|A_{j+1}|}{|A|}\right)^{\beta}\mathcal{H}(\iota_{A_{j+1}}).$$

Observe that $\pi_0=\pi$ and $\pi_n=\iota_A$. Consequently,

$$\mathcal{H}(\pi)=\mathcal{H}(\iota_A)-\sum_{j=1}^{n}\left(\frac{|A_j|}{|A|}\right)^{\beta}\mathcal{H}(\iota_{A_j}).$$

∎

Note that if A,B are two sets such that $|A|=|B|>0$, then, by Axiom **(P2)**, we have $\mathcal{H}(\iota_A)=\mathcal{H}(\iota_B)$. Therefore, the value of $\mathcal{H}(\iota_A)$ depends only on the cardinality of A, and there exists a function $\mu:\mathbb{N}_1\longrightarrow\mathbb{R}$ such that $\mathcal{H}(\iota_A)=\mu(|A|)$ for every nonempty set A. Axiom **(P2)** also implies that μ is an increasing function. We will refer to μ as the *core* of the (Φ,β)-system of axioms.

Corollary II.5: Let $\mathcal{H}$ be a (Φ,β)-entropy. For the function μ defined in Axiom **(P2)** and every partition $\pi=\{A_1,\ldots,A_n\}\in\mathsf{PART}(A)$ we have:

$$\mathcal{H}(\pi)=\mu(|A|)-\sum_{j=1}^{n}\left(\frac{|A_j|}{|A|}\right)^{\beta}\mu(|A_j|). \tag{1}$$

Proof: The statement is an immediate consequence of Theorems II.4. ∎

Theorem II.6: Let $\pi=\{B_1,\ldots,B_n\}$ be a partition of the set A. Define the partition π' obtained by fusing the blocks B_1 and B_2 of π as $\pi'=\{B_1\cup B_2,B_3,\ldots,B_n\}$ of the same set. Then

$$\mathcal{H}(\pi)=\mathcal{H}(\pi')+\left(\frac{|B_1\cup B_2|}{|A|}\right)^{\beta}\mathcal{H}(\{B_1,B_2\}).$$

Proof: A double application of Corollary II.5 yields:

$$\begin{aligned}\mathcal{H}(\pi') &= \mu(|A|)-\left(\frac{|B_1\cup B_2|}{|A|}\right)^{\beta}\mu(|B_1\cup B_2|)\\ &\quad-\sum_{i>2}^{n}\left(\frac{|B_i|}{|A|}\right)\mu(|B_i|)\end{aligned}$$

and

$$\mathcal{H}(\{B_1, B_2\}) = \mu(|B_1 \cup B_2|) - \left(\frac{|B_1|}{|B_1 \cup B_2|}\right)^\beta \mu(|B_1|) - \left(\frac{|B_2|}{|B_1 \cup B_2|}\right)^\beta \mu(|B_2|).$$

Substituting the above expressions in

$$\mathcal{H}(\pi') + \left(\frac{|B_1 \cup B_2|}{|A|}\right)^\beta \mathcal{H}(\{B_1, B_2\})$$

we obtain $\mathcal{H}(\pi)$. ∎

Theorem II.6 allows us to extend Axiom **(P3)**:

Corollary II.7: Let $A_1, \ldots, A_m$ be n nonempty, disjoint sets and let $\pi_i \in \mathsf{PART}(A)$ for $1 \leq i \leq m$. We have

$$\mathcal{H}(\pi_1 + \cdots + \pi_m) = \sum_{i=1}^{m} \left(\frac{|A_i|}{|A|}\right)^\beta \mathcal{H}(\pi_i) + \mathcal{H}(\{A_1, \ldots, A_m\}),$$

where $A = A_1 \cup \cdots \cup A_m$.

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 $A_1, \ldots, A_m, A_{m+1}$ 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 $A_m \cup A_{m+1}$. By the inductive hypothesis we have

$$\begin{aligned}&\mathcal{H}(\pi_1 + \cdots + (\pi_m + \pi_{m+1})) = \textstyle\sum_{i=1}^{m-1} \left(\frac{|A_i|}{|A|}\right)^\beta \mathcal{H}(\pi_i)\\ &+ \left(\frac{|A_m|+|A_{m+1}|}{|A|}\right)^\beta \mathcal{H}(\pi_m + \pi_{m+1})\\ &+\mathcal{H}(\{A_1, \ldots, (A_m \cup A_{m+1})\}),\end{aligned}$$

where $A = A_1 \cup \cdots \cup A_m \cup A_{m+1}$.

Axiom **(P3)** implies:

$$\begin{aligned}&\mathcal{H}(\pi_1 + \cdots + (\pi_m + \pi_{m+1})) = \textstyle\sum_{i=1}^{m-1} \left(\frac{|A_i|}{|A|}\right)^\beta \mathcal{H}(\pi_i)\\ &+ \left(\frac{|A_m|}{|A|}\right)^\beta \mathcal{H}(\pi_m) + \left(\frac{|A_{m+1}|}{|A|}\right)^\beta \mathcal{H}(\pi_{m+1})\\ &+ \left(\frac{|A_m|+|A_{m+1}|}{|A|}\right)^\beta \mathcal{H}\{A_m, A_{m+1}\}\\ &+\mathcal{H}(\{A_1, \ldots, (A_m \cup A_{m+1})\}).\end{aligned}$$

Finally, an application of Theorem II.6 gives the desired equality. ∎

Theorem II.8: Let μ be the core of a (Φ, β)-system. If $a, b \in \mathbb{N}_1$, then

$$\mu(ab) - \frac{\mu(a)}{b^{\beta-1}} = \mu(b).$$

Proof: Let $A = \{x_1, \ldots, x_a\}$ and $B = \{y_1, \ldots, y_b\}$ be two nonempty sets. Note that $\omega_A \times \iota_B$ consists of b blocks of size a: $A \times \{y_1\}, \ldots, A \times \{y_b\}$. By Axiom **(P4)**,

$$\mathcal{H}(\omega_A \times \iota_B) = \Phi(\mathcal{H}(\omega_A), \mathcal{H}(\iota_B)) = \Phi(0, \mathcal{H}(\iota_B)) = \mathcal{H}(\iota_B).$$

On the other hand,

$$\begin{aligned}\mathcal{H}(\omega_A \times \iota_B) &= \mathcal{H}(\iota_{A \times B}) - \sum_{i=1}^{b} \left(\frac{1}{b}\right)^\beta \mathcal{H}(\iota_{A \times \{y_i\}})\\ &= \mu(ab) - \frac{1}{b^\beta} b \cdot \mu(a),\end{aligned}$$

which gives the needed equality. ∎

An entropy is said to be *non-Shannon* if it defined by a (Φ, β)-system of axioms such that $\beta \neq 1$; otherwise, that is if $\beta = 1$, the entropy will be referred to as a *Shannon* entropy.

III. Axiomatization of non-Shannon Entropies

Theorem III.1: Let $\mathcal{H}$ be a non-Shannon entropy defined by a (Φ, β)-system of axioms and let μ be the core of this system of axioms.

There is a constant $k \in \mathbb{R}$ such that $k \cdot (\beta - 1) \geq 0$ and

$$\mu(a) = k \cdot \left(1 - \frac{1}{a^{\beta-1}}\right)$$

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

Proof: Theorem II.8 implies that

$$\mu(ab) = \frac{\mu(a)}{b^{\beta-1}} + \mu(b) = \frac{\mu(b)}{a^{\beta-1}} + \mu(a),$$

for every $a, b \in \mathbb{N}_1$. Consequently,

$$\frac{\mu(a)}{1 - \frac{1}{a^{\beta-1}}} = \frac{\mu(b)}{1 - \frac{1}{b^{\beta-1}}},$$

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

Note that for $\beta \neq 1$ we have:

$$k = \begin{cases} \lim_{a \to \infty} \mu(a) \text{ if } \beta > 1 \\ \lim_{a \to 0} \mu(a) \text{ if } \beta < 1. \end{cases} \tag{2}$$

Corollary III.2: If $\mathcal{H}$ is a non-Shannon entropy defined by a (Φ, β)-system of axioms and $\pi \in \mathsf{PART}(A)$, where $\pi = \{A_1, \ldots, A_n\}$, then there exists a constant $k \in \mathbb{R}$ such that

$$\mathcal{H}(\pi) = k \cdot \left(1 - \sum_{j=1}^{n} \left(\frac{|A_j|}{|A|}\right)^\beta\right). \tag{3}$$

Proof: By Corollary II.5 and by Theorem III.1 we have

$$\begin{aligned}&\mathcal{H}(\pi)\\ &= \mu(|A|) - \sum_{j=1}^{n} \left(\frac{|A_j|}{|A|}\right)^\beta \mu(|A_j|)\\ &= k \cdot \left(1 - \frac{1}{|A|^{\beta-1}}\right) - k \cdot \sum_{j=1}^{n} \left(\frac{|A_j|}{|A|}\right)^\beta \cdot \left(1 - \frac{1}{|A_j|^{\beta-1}}\right)\\ &= k \cdot \left(1 - \frac{1}{|A|^{\beta-1}}\right) - k \cdot \sum_{j=1}^{n} \left(\frac{|A_j|}{|A|}\right)^\beta + k \cdot \sum_{j=1}^{n} \frac{|A_j|}{|A|^\beta}\\ &= k \cdot \left(1 - \sum_{j=1}^{n} \left(\frac{|A_j|}{|A|}\right)^\beta\right).\end{aligned}$$

■

Theorem III.3: Let $\mathcal{H}$ be the non-Shannon entropy defined by a (Φ, β)-system and let k be as defined by Equality (2), where μ is the core of the (Φ, β)-system of axioms. The function Φ introduced by Axiom **(P4)** is given by $\Phi(x, y) = x + y - \frac{1}{k}xy$ for $x, y \in \mathbb{R}_{\geq 0}$.

Proof: Let $\pi = \{A_1, \ldots, A_n\} \in \mathsf{PART}(A)$ and $\sigma = \{B_1, \ldots, B_m\} \in \mathsf{PART}(B)$ be two partitions. Since

$$\begin{aligned}\sum_{j=1}^{n} \frac{|A_j|}{|A|} &= 1 - \frac{1}{k}\mathcal{H}(\pi)\\ \sum_{k=1}^{m} \frac{|B_k|}{|B|} &= 1 - \frac{1}{k}\mathcal{H}(\sigma)\end{aligned}$$

we can write:

$$\begin{aligned}\mathcal{H}(\pi \times \sigma) &= k\left(1 - \sum_{j=1}^{n}\sum_{k=1}^{m}\left(\frac{|A_j||B_k|}{|A||B|}\right)^{\beta}\right)\\ &= k\left(1 - \left(1 - \frac{1}{k}\mathcal{H}(\pi)\right)\left(1 - \frac{1}{k}\mathcal{H}(\sigma)\right)\right)\\ &= \mathcal{H}(\pi) + \mathcal{H}(\sigma) - \frac{1}{k}\mathcal{H}(\pi)\mathcal{H}(\sigma).\end{aligned}$$

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 $k = \frac{1}{\beta - 1}$ in the equality (3) we obtain the Havrda-Charvat entropy (see [13]):

$$\mathcal{H}(\pi) = \frac{1}{\beta - 1} \cdot \left(1 - \sum_{j=1}^{n}\left(\frac{|A_j|}{|A|}\right)^{\beta}\right).$$

The limit case, $\lim_{\beta \to 1} \mathcal{H}(\pi)$ yields the Shannon entropy. The case $\beta = 1$ is considered independently in the next section.

If $\beta = 2$ we obtain the Gini index,

$$\mathcal{H}(\pi) = 1 - \sum_{j=1}^{n}\left(\frac{|A_j|}{|A|}\right)^{2},$$

which is widely used in machine learning and data mining.

IV. Axiomatization of Shannon Entropy

When $\beta = 1$, by Theorem II.8, 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 from [15], 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:

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

for every partition $\pi = \{A_1, \ldots, A_n\}$ of a set A, where $|A_i| = a_i$ for $1 \leq i \leq n$, 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

$$c \cdot \log_2 ab = \mathcal{H}(\iota_A \times \iota_B) = \Phi(c \cdot \log_2 a, c \cdot \log_2 b)$$

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

V. Conditional Entropy

The entropies previously introduced generate corresponding conditional entropies.

Let $\pi \in \mathsf{PART}(A)$ and let $C \subseteq A$. 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 V.1: The *conditional entropy* defined by the (Φ, β)-entropy $\mathcal{H}$ is the function $\mathcal{C} : \mathsf{PART}^2 \longrightarrow \mathbb{R}_{\geq 0}$ given by:

$$\mathcal{C}(\pi, \sigma) = \sum_{j=1}^{n} \frac{|C_j|}{|A|} \cdot \mathcal{H}(\pi_{C_j}),$$

where $\pi, \sigma \in \mathsf{PART}(A)$ and $\sigma = \{C_1, \ldots, C_n\}$. □

We denote the value of $\mathcal{C}(\pi, \sigma)$ by $\mathcal{H}(\pi|\sigma)$. Note that $\mathcal{H}(\pi|\omega_A) = \mathcal{H}(\pi)$.

The partition $\pi \wedge \sigma$ whose blocks consist of the nonempty intersections of the blocks of π and σ can be written as

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

Therefore, by Corollary II.7, we have:

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

For those entropies with $\beta > 1$ we have

$$\mathcal{H}(\pi \wedge \sigma) \geq \mathcal{H}(\pi|\sigma) + \mathcal{H}(\sigma),$$

while for those having $\beta < 1$, the reverse inequality holds. In the case of Shannon entropy, $\beta = 1$ and

$$\begin{aligned}\mathcal{H}(\pi \wedge \sigma) &= \mathcal{H}(\pi|\sigma) + \mathcal{H}(\sigma)\\ &= \mathcal{H}(\sigma|\pi) + \mathcal{H}(\pi).\end{aligned}$$

If $\mathcal{H}$ is a (Φ, β)-entropy, $\pi, \sigma \in \mathsf{PART}(A)$ are such that $\pi = \{B_1, \ldots, B_m\}$ and $\sigma = \{C_1, \ldots, C_n\}$, then the conditional entropy $\mathcal{H}(\pi|\sigma)$ is given by:

$$\begin{aligned}\mathcal{H}(\pi|\sigma) &= \sum_{j=1}^{n}\frac{|C_j|}{|A|}\mu(|C_j|)\\ &= \sum_{j=1}^{n}\frac{|C_j|}{|A|}\mu(|C_j|)\\ &\quad - \sum_{j=1}^{n}\sum_{i=1}^{m}\frac{|B_i \cap C_j|^{\beta}}{|A| \cdot |C_j|^{\beta-1}}\mu(|B_i \cap C_j|).\end{aligned}$$

This equality follows immediately from Corollary II.5.

In the case of Shannon entropy, taking $\beta = 1$ and $\mu(n) = \log_2 n$ we obtain the well-known expression of conditional entropy:

$$\mathcal{H}(\pi|\sigma) = -\sum_{i=1}^{m}\sum_{j=1}^{n} \frac{|B_i \cap C_j|}{|A|} \log_2 \frac{|B_i \cap C_j|}{|C_j|}.$$

In the case of the Gini index we have $\beta = 2$ and $\mu(a) = k\left(1 - \frac{1}{a}\right)$ for $a \in \mathbb{N}_1$. Consequently, after some elementary transformations, the conditional Gini index is:

$$\mathcal{H}(\pi|\sigma) = 1 - \sum_{j=1}^{n}\sum_{i=1}^{m} \frac{|B_i \cap C_j|^2}{|A| \cdot |C_j|}.$$

For $\beta = 0.5$, we obtain the "square-root entropy" used by us in clustering of categorial data (see [16]).

VI. Conclusions

The main result of the paper is a common axiomatization of several numerical characteristics of partitions (or, equivalently, of functions) that measure the "concentration" of values. Some of these characteristics (the Shannon entropy, Gini index, etc.) are widely used in data mining, machine learning, and, in the area of multiple-valued logic, in constructing decision diagrams, minimization, etc.

The axiomatization constructed opens the possibility that some of these measures can be used in new areas of application, and some entirely new characteristics can be used for the same purpose.

References

[1] Cheng K.T. and Agrawal V., *An Entropy Measure of the Complexity of Multi-Output Boolean Function*, Proc. of the 27th Design Automation Conference, pp. 302–305, 1990.

[2] Cheushev V., Shmerko V., Simovici D.A., and Yanushkevich S., *Functional Entropy and Decision Trees*, Proc. ISMVL'98, pp. 257–262.

[3] Devijer P. A., *Entropie quadratique et reconnaisance des formes*, Computer Oriented Learning Processes, Proceedings of the NATO Advanced Study Institute, Château de Bonas, France, 1974, J, C. Simon, editor, pp. 257–278.

[4] Daróczy, Z., *Generalized Information Functions*, Information and Control, vol. 16, pp. 36–51, 1970.

[5] Faddeev, D. K., *On the notion of entropy of finite probability distributions*, in Usp. Mat. Nauk (in Russian), 11(1956), pp. 227–231.

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

[7] Hwang C.H. and Wu A.C.H., *An Entropy Measure for Power Estimation of Boolean Function*, Proc. of Asia and South Pacific Design Automation Conference (ASP-DAC), pp. 101–106, 1997.

[8] Ingarden R. S. and Urbanik, K., *Information without Probability*, Coll. Math., 1(1962), pp. 281–304.

[9] Jaroszewicz S. and Simovici, D. A., *On Axiomatization of Conditional Entropy of Functions between Finite Sets*, Proc.of the 29th ISMVL, Freiburg, Germany, pp. 24–31.

[10] Kapur J.N. and Kesavan H.K., *Entropy Optimization Principles with Applications*, Academic Press, San Diego, 1992.

[11] Khinchin, A. Ia., *Mathematical Foundations of Information Theory*, Dover Publ., New York, 1957.

[12] Lloris-Ruiz A., Gomez-Lopera J.F. and Roman-Roldan R., *Entropic Minimization of Multiple-Valued Logic Functions*, Proc. ISMVL'93, pp. 24–28.

[13] Mathai A.M. and Rathie P.N., *Basic Concepts in Information Theory and Statistics — Axiomatic Foundations and Applications*, Halsted Press, John Wiley & Sons, 1975.

[14] Rényi, A., *On the Dimension and Entropy of Probability Distributions*, Acta Math. Acad. Sci. Hung., 10(1959), pp. 193–215.

[15] Simovici D.A. and Reischer C., *On Functional Entropy*, Proc. ISMVL'93, pp. 100–104.

[16] D. A. Simovici, D. Cristofor, L. Cristofor, *Generalized Entropy and Projection Clustering of Categorial Data*, in Principles of Data Mining and Knowledge Discovery, the 4th European Conference, PKDD 2000, Lyon, Lecture Notes in Artificial Intelligence, 1910, Springer-Verlag, pp. 619–625.