

An axiomatic approach to the definition of the entropy of a discrete Choquet capacity [★]

Ivan Kojadinovic ^a, Jean-Luc Marichal ^b, Marc Roubens ^c

^a*École polytechnique de l'Université de Nantes, LINA CNRS FRE 2729, Rue Christian Pauc, 44306 Nantes, France*

^b*Faculty of Law, Economics, and Finance, University of Luxembourg, 162A, avenue de la Faïencerie, L-1511 Luxembourg, Grand Duchy of Luxembourg.*

^c*Institute of Mathematics, University of Liège, Grande Traverse 12 (B37), B-4000 Liège, Belgium*

Abstract

To extend the classical Shannon entropy to non-additive measures, Marichal [1–3] recently introduced the concept of generalized entropy for discrete Choquet capacities. We provide a first axiomatization of this new concept on the basis of three axioms: the symmetry property, a boundary condition for which the entropy reduces to the Shannon entropy, and a generalized version of the well-known recursivity property. We also show that this generalized entropy fulfills several properties considered as requisites for defining an entropy-like measure. Lastly, we provide an interpretation of it in the framework of aggregation by the discrete Choquet integral.

Key words: Entropy; Choquet capacity; Choquet integral; Information theory.

1 Introduction

The well-known Shannon entropy [5] of a probability distribution p defined on a nonempty finite set $N := \{1, \dots, n\}$ is given by

$$H_S(p) := - \sum_{i \in N} p(i) \ln p(i)$$

[★] This paper is a revised and extended version with proofs of the conference paper [4].

Email addresses: Ivan.Kojadinovic[at]polytech.univ-nantes.fr (Ivan Kojadinovic), jean-luc.marichal[at]uni.lu (Jean-Luc Marichal), m.roubens[at]ulg.ac.be (Marc Roubens).

with the convention that $0 \ln 0 := 0$.

In a general non probabilistic setting, it is merely a measure of the uniformity (evenness) of p . In a probabilistic context, when p is associated with an n -state discrete stochastic system, it is naturally interpreted as a measure of its unpredictability and thus reflects the uncertainty associated with a future state of the system.

Although several other measures of uniformity and uncertainty have been proposed in probability theory as generalizations of the Shannon entropy (see e.g. [6] for an overview), the most widely used measure remains that of Shannon mainly because of its attractive properties, its connections with the Kullback-Leibler divergence [7], and its role in the maximum entropy principle [8]. Also, several axiomatic characterizations of the Shannon entropy have been proposed in the literature (see e.g. [9–12]), among which the most famous is probably Shannon’s theorem [5].

By relaxing the additivity property of probability measures, requiring only that they be monotone, one obtains Choquet capacities [13], also known as fuzzy measures [14], which are able to model other types of possibly uncertain information. Formally, a discrete Choquet capacity μ on a nonempty finite set N is a monotone set function defined on the power set of N that is zero at the empty set.

Discrete Choquet capacities are at the root of many fields such as the Dempster-Shafer theory of evidence (see e.g. [15,16]), possibility theory (see e.g. [16,17]), cooperative game theory (see e.g. [18,19]), and multicriteria decision making (see e.g. [20,21]). In these fields, Choquet capacities are generally used to model either uncertainty (in evidence and possibility theories) or the importance of coalitions (in cooperative game theory and multicriteria decision making).

A discrete Choquet capacity being clearly a generalization of a discrete probability distribution, the following natural question arises : how could one appraise the ‘uniformity’ or ‘uncertainty’ associated with a Choquet capacity in the spirit of the Shannon entropy?

For particular Choquet capacities such as belief, plausibility, and possibility measures used in evidence and possibility theories, several definitions of entropy-like measures of uncertainty, known as measures of ‘conflict’, ‘confusion’, ‘dissonance’, ‘discord’, and ‘aggregate uncertainty’, have been proposed (see e.g. [15,16,22–24]). For general Choquet capacities, it seems that no definition of generalized entropy was available until recently when three proposals were introduced successively by Marichal [1–3], Yager [25,26], and Dukhovny [27]. All three proposals can be regarded as direct extensions of the Shannon entropy in the sense that they coincide with the latter when the capacity is additive.

The main purpose of this paper is to propose a first axiomatization, as a measure of ‘uniformity’, of the generalized entropy proposed by Marichal, which is defined by

$$H_M(\mu) := \sum_{i \in N} \sum_{S \subseteq N \setminus \{i\}} \gamma_{|S|}(|N|) h[\mu(S \cup \{i\}) - \mu(S)],$$

where μ is a Choquet capacity on $N = \{1, \dots, n\}$ and where the functions

$$\gamma_s(n) := \frac{(n-s-1)! s!}{n!} \quad (s = 0, 1, \dots, n-1),$$

and

$$h(x) := \begin{cases} -x \ln x, & \text{if } x > 0, \\ 0, & \text{if } x = 0, \end{cases}$$

will be used throughout.

As we will see in Section 3, this axiomatization is based on three properties: the symmetry property, a boundary condition for which the entropy reduces to the Shannon entropy, and a generalized version of the well-known recursivity property.

It has been shown that this generalized entropy H_M satisfies many properties that one would intuitively require from such a measure [2]. We will recall some of them in Section 4.

Note also that its formulation is very close to that of the Shapley value [18] of a Choquet capacity μ on N , which is a fundamental concept in cooperative game theory. It is an n -dimensional vector $\phi(\mu)$ whose i th component is defined by

$$\phi_i(\mu) := \sum_{S \subseteq N \setminus \{i\}} \gamma_{|S|}(|N|) [\mu(S \cup \{i\}) - \mu(S)]$$

and can be interpreted as the average marginal contribution of i to a coalition not containing it. It is worth noting that a basic property of the Shapley value is

$$\sum_{i \in N} \phi_i(\mu) = \mu(N). \quad (1)$$

Hence, for a normalized Choquet capacity μ on N (i.e., satisfying $\mu(N) = 1$), the Shapley value $\phi(\mu)$ forms a probability distribution on N .

This latter observation led Yager [25,26] to propose another definition for normalized Choquet capacities on N . His generalized entropy is defined as the Shannon entropy of the Shapley value of the capacity, i.e.

$$H_Y(\mu) := H_S(\phi(\mu)).$$

A comparative study of H_M and H_Y can be found in [3].

The last and most recent proposal, due to Dukhovny [27], is based on an interpretation of H_M in terms of maximal chains of the Hasse diagram of N and will be presented in Section 2.

In this paper, after defining the notion of uniformity in the setting of discrete Choquet capacities (Section 2), we propose a characterization of the generalized entropy H_M by means of three axioms (Section 3). We also list some of its properties (Section 4) and present an interpretation of it in the framework of aggregation by the discrete Choquet integral (Section 5).

2 Uniformity of a discrete Choquet capacity

The Shannon entropy $H_S(p)$ of a distribution $p = (p_1, \dots, p_n)$ of n positive real numbers summing up to one is a measure of the uniformity of these numbers. That is, the closer $H_S(p)$ is to $\ln n$, the more uniform is p . In a probabilistic context, when p is regarded as the probability distribution associated with an n -state discrete stochastic system, $H_S(p)$ becomes a natural measure of the uncertainty associated with the system. That is, the more uniform is p , the higher $H_S(p)$ and thus the more difficult it is to predict a future state of the system.

The approach to the definition of a generalized entropy measure adopted in this paper is similarly grounded on the notion of uniformity. In the present section, we propose an intuitive definition of this notion for discrete Choquet capacities based on their lattice representation.

2.1 Notation and first definitions

Throughout this paper we shall consider a countably infinite set U , the *universe of discourse*. A *discrete Choquet capacity* [13] or *discrete fuzzy measure* [14] on U is then a set function $\mu : 2^U \rightarrow \mathbb{R}^+$ ($\mathbb{R}^+ := [0, +\infty[$) satisfying the following conditions :

- (i) $\mu(\emptyset) = 0$,
- (ii) for any $S, T \subseteq U$, we have $S \subseteq T \Rightarrow \mu(S) \leq \mu(T)$.

A set $N \subseteq U$ is said to be a *carrier* (or *support*) of a Choquet capacity μ if $\mu(S) = \mu(N \cap S)$ for all $S \subseteq U$. Thus, a capacity μ with carrier $N \subseteq U$ is completely defined by the knowledge of the coefficients $\{\mu(S)\}_{S \subseteq N}$ and the elements outside N have no influence on the capacity since they do not contribute to any subset. Note that it follows that any superset of N is also

a carrier of μ . Clearly, a Choquet capacity on a nonempty finite set N can be identified with a Choquet capacity on U with carrier N .

A Choquet capacity with nonempty finite carrier $N \subset U$ is said to be *normalized* if $\mu(N) = 1$. In that case we also have $\mu(N') = 1$ for any other carrier N' of μ .

In this paper we restrict our study of the notion of uniformity to normalized Choquet capacities. We denote by \mathcal{C} the set of normalized capacities with nonempty finite carriers on U and by \mathcal{C}_N the set of normalized capacities having $N \subset U$ as nonempty finite carrier.

In order to avoid a heavy notation, we will often omit braces for singletons, e.g., by writing $\mu(i)$, $U \setminus i$ instead of $\mu(\{i\})$, $U \setminus \{i\}$. Furthermore, cardinalities of subsets S, T, \dots , will be denoted by the corresponding lower case letters s, t, \dots . Finally, for any integer $k \geq 1$, the set $\{1, \dots, k\}$ will simply be denoted by $[k]$.

A Choquet capacity $\mu \in \mathcal{C}_N$ is said to be

- *additive* if $\mu(S \cup T) = \mu(S) + \mu(T)$ for all disjoint subsets $S, T \subseteq N$,
- *cardinality-based* if, for all $T \subseteq N$, $\mu(T)$ depends only on the cardinality of T . Formally, there exist $\mu_0, \mu_1, \dots, \mu_n \in [0, 1]$ such that $\mu(T) = \mu_t$ for all $T \subseteq N$ such that $|T| = t$. Note that if $\mu \in \mathcal{C}_N$ is cardinality-based, μ seen as an element of $\mathcal{C}_{N'}$, with $N' \supsetneq N$, is no longer cardinality-based.

There is only one Choquet capacity of \mathcal{C}_N that is both additive and cardinality-based. We shall call it *the uniform Choquet capacity* and denote it by μ_N^* . It is easy to check that μ_N^* is given by

$$\mu_N^*(T) = t/n, \quad \forall T \subseteq N.$$

For the sake of convenience, we will henceforth assume that N is the n -element set $[n]$.

2.2 Choquet capacities and maximal chains

The lattice related to the power set of N under the inclusion relation can be represented by a graph \mathcal{H}_N , called Hasse diagram, whose nodes correspond to subsets $S \subseteq N$ and whose edges represent adding an element to the bottom subset to get the top subset. Figure 1 shows an example of such a graph for $N = \{1, 2, 3, 4\}$.

A *maximal chain* m of \mathcal{H}_N is an ordered collection of $n + 1$ nested distinct

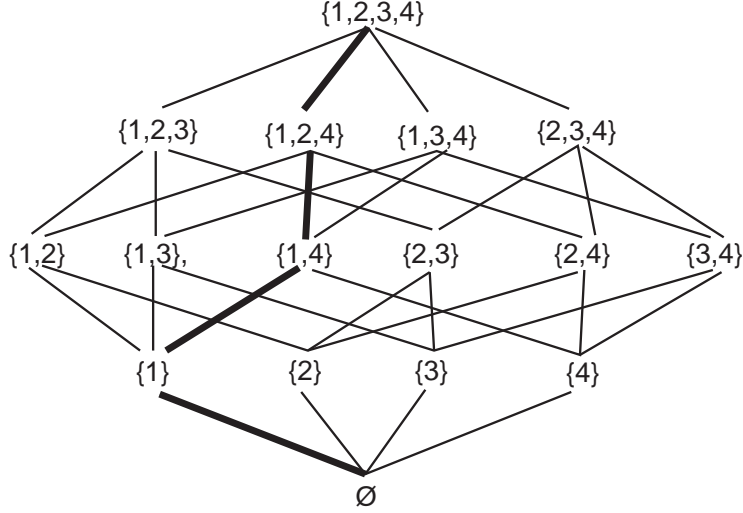


Fig. 1. Hasse diagram corresponding to the lattice of subsets of $N = \{1, 2, 3, 4\}$.
subsets denoted

$$m = (\emptyset \subsetneq \{i_1\} \subsetneq \{i_1, i_2\} \subsetneq \cdots \subsetneq \{i_1, \dots, i_n\} = N).^1$$

Denote by \mathcal{M}_N the set of maximal chains of \mathcal{H}_N and by Π_K the set of permutations on any set $K \subseteq U$. We can readily see that the sets \mathcal{M}_N and Π_N are equipollent. Indeed, to each permutation $\pi \in \Pi_N$ corresponds a unique maximal chain $m^\pi \in \mathcal{M}_N$ defined by

$$m^\pi = (\emptyset \subsetneq \{\pi(n)\} \subsetneq \{\pi(n-1), \pi(n)\} \subsetneq \cdots \subsetneq \{\pi(1), \dots, \pi(n)\} = N).$$

Now, given a capacity $\mu \in \mathcal{C}_N$, with each permutation $\pi \in \Pi_N$ can be associated a discrete probability distribution p_π^μ on N defined by

$$p_\pi^\mu(i) := \mu(\{\pi(i), \dots, \pi(n)\}) - \mu(\{\pi(i+1), \dots, \pi(n)\}), \quad \forall i \in N.$$

Equivalently, with the maximal chain $m^\pi \in \mathcal{M}_N$ is associated the probability distribution $p_{m^\pi}^\mu := p_\pi^\mu$.

We will denote by P_N^μ the set $\{p_\pi^\mu\}_{\pi \in \Pi_N} = \{p_m^\mu\}_{m \in \mathcal{M}_N}$ of $n!$ probability distributions obtained from $\mu \in \mathcal{C}_N$.

Notice that, if $\mu \in \mathcal{C}_N$ is cardinality-based, then there exists a unique probability distribution p^μ on N such that $p_\pi^\mu = p^\mu$ for all $\pi \in \Pi_N$. If $\mu \in \mathcal{C}_N$ is

¹ For instance, in Figure 1, the maximal chain

$$(\emptyset \subsetneq \{1\} \subsetneq \{1, 4\} \subsetneq \{1, 2, 4\} \subsetneq \{1, 2, 3, 4\} = N)$$

is emphasized.

additive then we simply have $p_\pi^\mu(i) = \mu(\pi(i))$ for all $i \in N$.

2.3 Uniformity of a discrete Choquet capacity

As discussed earlier, in a general non probabilistic context, the Shannon entropy of a discrete probability distribution can be regarded a measure of its uniformity (for a discussion, see e.g. [6]).

Although discrete Choquet capacities can be clearly seen as generalizations of discrete probability distributions, it is not clear what ‘uniformity’ should mean for such non-additive measures. As we have seen above, a capacity $\mu \in \mathcal{C}_N$ can be represented by the set $P_N^\mu = \{p_m^\mu\}_{m \in \mathcal{M}_N}$ of $n!$ probability distributions on N . We therefore propose to define the intuitive notion of *uniformity of a Choquet capacity* $\mu \in \mathcal{C}_N$ as an average of the uniformity values of the probability distributions p_m^μ ($m \in \mathcal{M}_N$). Hence, the more uniform on average the probability distributions p_m^μ ($m \in \mathcal{M}_N$), the higher the uniformity of the capacity μ . As no maximal chain should be privileged, a measure of the average uniformity should be defined by means of a symmetric function of the values $H_S(p_m^\mu)$ ($m \in \mathcal{M}_N$). In this respect, it can be shown that the entropy H_M is written in terms of maximal chains as

$$H_M(\mu) = \frac{1}{n!} \sum_{m \in \mathcal{M}_N} H_S(p_m^\mu), \quad \forall \mu \in \mathcal{C}_N, \quad (2)$$

or, equivalently,

$$H_M(\mu) = \frac{1}{n!} \sum_{\pi \in \Pi_N} H_S(p_\pi^\mu), \quad \forall \mu \in \mathcal{C}_N. \quad (3)$$

This result immediately follows from the next proposition [4] (see also [27]).

Proposition 1 *Let μ be any Choquet capacity with nonempty finite carrier N (normalized or not) and let f be any function. Then, we have*

$$\frac{1}{n!} \sum_{m \in \mathcal{M}_N} \sum_{j \in N} f[p_m^\mu(j)] = \sum_{i \in N} \sum_{S \subseteq N \setminus i} \gamma_s(n) f[\mu(S \cup i) - \mu(S)].$$

Proof. For any $i \in N$ and any $S \subseteq N \setminus i$, let us denote by $\mathcal{M}_N^{S, S \cup i}$ the subset of \mathcal{M}_N formed by the maximal chains whose subsets of cardinality s and $s+1$ are respectively S and $S \cup i$. It is easy to check that

$$|\mathcal{M}_N^{S, S \cup i}| = s!(n-s-1)!$$

It follows therefore that, for any fixed $i \in N$ and any fixed $S \subseteq N \setminus i$, when summing the term $\sum_{j \in N} f[p_m^\mu(j)]$ over the set of maximal chains, the expression $f[\mu(S \cup i) - \mu(S)]$ appears $s!(n - s - 1)!$ times. \square

If Π_N is considered as a probability space, a straightforward probabilistic interpretation of H_M directly follows from Eq. (3). For any $\mu \in \mathcal{C}_N$, $H_M(\mu)$ is the mathematical expectation of $H_S(p_\pi^\mu)$ with respect to the uniform distribution on Π_N .

Interestingly, Dukhovny [27] recently proposed the following alternative entropy-like measure

$$H_D(\mu) := \min_{m \in \mathcal{M}_N} H_S(p_m^\mu), \quad \forall \mu \in \mathcal{C}_N,$$

which is nothing but the lowest uniformity value contained in the probability distributions $\{p_m^\mu\}_{m \in \mathcal{M}_N}$. Its advantage resides in the fact that it can be defined even for capacities having no finite carriers.

3 Axiomatization of the entropy H_M

Before stating the axioms we use in our characterization, we define some additional concepts that will be needed as we proceed.

3.1 Additional definitions

Consider any nonempty finite set $N \subset U$ and let $\mu \in \mathcal{C}_N$. Let also S and T be two disjoint subsets of N , and let A_1, \dots, A_k form any partition of N .

We consider the following definitions (see e.g. [28]):

- The *restriction of μ to S* is a capacity of \mathcal{C}_S denoted μ^S and defined by

$$\mu^S(K) := \mu(K), \quad \forall K \subseteq S.$$

- The *restriction of μ to S in the presence of T* is a capacity of \mathcal{C}_S denoted $\mu_{\cup T}^S$ and defined by

$$\mu_{\cup T}^S(K) := \mu(K \cup T) - \mu(T), \quad \forall K \subseteq S.$$

- The *reduction of μ with respect to A_1, \dots, A_k* is a capacity of $\mathcal{C}_{\{[A_1], \dots, [A_k]\}}$ denoted $\mu^{[A_1] \dots [A_k]}$ and defined by

$$\mu^{[A_1] \dots [A_k]} \left(\bigcup_{i \in S} [A_i] \right) = \mu \left(\bigcup_{i \in S} A_i \right), \quad \forall S \subseteq [k], \quad (4)$$

where, for all $i \in [k]$, $[A_i]$ stands for an hypothetical macro element that is the union (or the representative) of the elements in A_i .

Clearly, these new capacities need not be normalized. The normalized version of any Choquet capacity μ on N , denoted $\bar{\mu}$, is defined by

$$\bar{\mu}(S) := \begin{cases} \mu(S)/\mu(N), & \text{for all } S \subseteq N \text{ if } \mu(N) \neq 0, \\ 0, & \text{for all } S \subsetneq N \text{ if } \mu(N) = 0, \\ 1, & \text{if } S = N \text{ and } \mu(N) = 0. \end{cases}$$

Finally, for any capacity $\mu \in \mathcal{C}_N$ and any permutation $\pi \in \Pi_U$, we denote by $\pi\mu$ the capacity of $\mathcal{C}_{\pi(N)}$ defined by

$$\pi\mu[\pi(S)] = \mu(S), \quad \forall S \subseteq U,$$

where $\pi(S) := \{\pi(i) \mid i \in S\}$.

3.2 Axioms

To present our axiomatization, we will regard a generalized entropy merely as a function

$$H : \mathcal{C} \rightarrow \mathbb{R}^+,$$

where, for any nonempty finite set $N \subset U$ and any $\mu \in \mathcal{C}_N$, the number $H(\mu)$ quantifies the uniformity of μ .

Given $\mu \in \mathcal{C}_N$, it is easy to verify that any permutation $\pi \in \Pi_U$ transforms the set P_N^μ of probability distributions into the set $P_{\pi(N)}^{\pi\mu}$, thus only affecting the labels of the elements of N . Hence, naturally, our first axiom is:

- *Symmetry axiom (S)*: For any $\mu \in \mathcal{C}$ and any $\pi \in \Pi_U$, we have $H(\pi\mu) = H(\mu)$.

Recall from the previous section that, for a cardinality-based capacity $\mu \in \mathcal{C}_N$, there exists a probability distribution p^μ on N such that the elements of the set P_N^μ all are equal to p^μ . This suggests measuring the uniformity of a cardinality-based capacity $\mu \in \mathcal{C}_N$ as that of the probability distribution p^μ . The choice of the Shannon entropy being natural as that of a measure of uniformity of a probability distribution, our second axiom is:

- *Shannon entropy axiom (SE)*: For any nonempty finite set $N \subset U$ and any cardinality-based capacity $\mu \in \mathcal{C}_N$, we have $H(\mu) = H_S(p^\mu)$.

Note that the previous axiom naturally implies that among all the cardinality-based capacities of \mathcal{C}_N , μ_N^* is the most uniform one.

The Shannon entropy is known to satisfy the so-called recursivity property [5,6,9,10,29], which, in a probabilistic context, basically states that the entropy of a discrete stochastic system can be calculated either directly or by dividing the system into subsystems.² Let p be a probability distribution on a nonempty finite set $N \subset U$ and assume that there exists a partition $\{A_1, A_2\}$ of N such that $p(A_1) \neq 0$ and $p(A_2) \neq 0$. Then, recursivity means that

$$H_S(p) = H_S(p^{[A_1][A_2]}) + p(A_1) H_S(\bar{p}^{A_1}) + p(A_2) H_S(\bar{p}^{A_2}), \quad (5)$$

where, according to our terminology, $p^{[A_1][A_2]}$ is a probability distribution on the pair $\{[A_1], [A_2]\}$ defined by

$$p^{[A_1][A_2]}([A_i]) = p(A_i), \quad \forall i \in \{1, 2\},$$

and where \bar{p}^{A_1} and \bar{p}^{A_2} are the normalized probability distributions on A_1 and A_2 , respectively, obtained from p , that is,

$$\bar{p}^{A_i}(j) = \frac{p(j)}{p(A_i)} \quad \forall j \in A_i, \forall i \in \{1, 2\}.$$

As a generalization of the Shannon entropy, we require that our measure of uniformity H satisfy a similar property that would thus reflect the possibility to decompose in an additive way the calculation of the uniformity of any capacity $\mu \in \mathcal{C}_N$. Under certain conditions and when μ is cardinality-based, such a decomposition already exists if axiom SE is satisfied. To demonstrate this, let us consider a cardinality-based capacity $\mu \in \mathcal{C}_N$ and a partition of N into two subsets A_1 and A_2 . The capacities that can be constructed from such a decomposition are :

- the reduced capacity $\mu^{[A_1][A_2]}$ on $\{[A_1], [A_2]\}$, which need not be cardinality-based,
- the capacities on A_1 : $\bar{\mu}^{A_1}$ and $\bar{\mu}_{\cup A_2}^{A_1}$, which are cardinality-based,
- and the capacities on A_2 : $\bar{\mu}^{A_2}$ and $\bar{\mu}_{\cup A_1}^{A_2}$, which are cardinality-based as well.

Assume now that n is even so that the subsets A_1 and A_2 can be chosen to have the same cardinality $n/2$. Then, there exist permutations $\pi, \sigma \in \Pi_N$ such that $\bar{\mu}^{A_1} = \pi \bar{\mu}^{A_2}$ and $\bar{\mu}_{\cup A_2}^{A_1} = \sigma \bar{\mu}_{\cup A_1}^{A_2}$, and the reduced capacity $\mu^{[A_1][A_2]}$ is cardinality-based. According to axiom SE and Eq. (5), the following identity

² Note that the recursivity property is important because it implies among other things the additivity of the Shannon entropy, i.e.,

$$H_S(p \otimes q) = H_S(p) + H_S(q),$$

for all probability distributions p, q on N , where $p \otimes q$ denotes the tensor product of p and q .

holds

$$H(\mu) = H(\mu^{[A_1][A_2]}) + \mu^{A_1}(A_1) H(\bar{\mu}^{A_1}) + \mu_{\cup A_1}^{A_2}(A_2) H(\bar{\mu}_{\cup A_1}^{A_2}), \quad (6)$$

which could be equivalently rewritten as

$$H(\mu) = H(\mu^{[A_1][A_2]}) + \mu^{A_2}(A_2) H(\bar{\mu}^{A_2}) + \mu_{\cup A_2}^{A_1}(A_1) H(\bar{\mu}_{\cup A_2}^{A_1}).$$

The following question then arises: how could we generalize the previous functional equation to situations when the subsets A_1 and A_2 do not have the same cardinality anymore? Indeed, for a general choice of N and of the subsets A_1 and A_2 , the capacities $\bar{\mu}^{A_1}$ and $\bar{\mu}^{A_2}$ need not have the same entropy. Neither do the capacities $\bar{\mu}_{\cup A_2}^{A_1}$ and $\bar{\mu}_{\cup A_1}^{A_2}$. Furthermore, the reduced capacity $\mu^{[A_1][A_2]}$ need not be cardinality-based. As an extension of the previous case, we then require that the following functional equation, which is a generalization of (6) and is still in accordance with our intuitive additivity requirement, holds for any cardinality-based capacity $\mu \in \mathcal{C}_N$.

$$\begin{aligned} H(\mu) = & H(\mu^{[A_1][A_2]}) \\ & + \alpha^1 \mu^{A_1}(A_1) H(\bar{\mu}^{A_1}) + \alpha_2^1 \mu_{\cup A_2}^{A_1}(A_1) H(\bar{\mu}_{\cup A_2}^{A_1}) \\ & + \alpha^2 \mu^{A_2}(A_2) H(\bar{\mu}^{A_2}) + \alpha_1^2 \mu_{\cup A_1}^{A_2}(A_2) H(\bar{\mu}_{\cup A_1}^{A_2}), \end{aligned}$$

where α^1 , α_2^1 , α^2 , and α_1^2 are as yet undetermined real numbers.

By generalizing the previous functional equation to any partition $\{A_1, \dots, A_k\}$ of N into k subsets, we obtain our third axiom:

- *Recursivity axiom (R)*: For any finite set $N \subset U$, with $n \geq 2$, and any integer k such that $2 \leq k \leq n$, there exists a family of real coefficients

$$\{\alpha_S^i(n, k) \mid i \in [k], S \subseteq [k] \setminus i\}$$

such that, for any partition $\{A_1, \dots, A_k\}$ of N and any cardinality-based capacity $\mu \in \mathcal{C}_N$, we have

$$H(\mu) = H(\mu^{[A_1] \dots [A_k]}) + \sum_{i=1}^k \sum_{S \subseteq [k] \setminus i} \alpha_S^i(n, k) \mu_{\cup_{j \in S} A_j}^{A_i}(A_i) H(\bar{\mu}_{\cup_{j \in S} A_j}^{A_i}). \quad (7)$$

The three axioms we consider here are natural enough to characterize the generalized entropy H_M as a measure of uniformity. Axiom S says that the labels of the elements of any carrier do not matter. Axiom SE means that in the limit case where the capacity is cardinality-based then H_M collapses into the Shannon entropy, which is a natural choice for measuring the uniformity along the maximal chains. Finally, axiom R makes it possible to calculate the uniformity of any capacity starting from this limit case and from certain of its restrictions and reductions.

3.3 Axiomatic characterization of H_M

We can now state our main result.

Theorem 2 $H : \mathcal{C} \rightarrow \mathbb{R}^+$ fulfills axioms S, SE, and R, if and only if $H = H_M$.

The proof of this theorem is given in the appendix.

4 Properties of the entropy H_M

In addition to axioms S, SE, and R, the entropy H_M fulfills several properties considered as natural for an entropy-like measure. In this section we list some of them (see [2,3,27]).

- (1) **Boundary property for additive measures.** For any additive capacity $\mu \in \mathcal{C}_N$, we have

$$H_M(\mu) = H_S(p),$$

where p is the probability distribution on N defined by $p(i) = \mu(i)$ for all $i \in N$.

- (2) **Expansibility.** Let $\mu \in \mathcal{C}_N$ and let $i \in N$ be a null element, that is, such that $\mu(S \cup i) = \mu(S)$ for all $S \subseteq N \setminus i$. Then

$$H_M(\mu) = H_M(\mu^{N \setminus i}).$$

- (3) **Decisivity.** We have

$$H_M(\mu) \geq 0, \quad \forall \mu \in \mathcal{C}_N.$$

Moreover, $H_M(\mu) = 0$ if and only if μ is a binary-valued capacity, that is, such that $\mu(T) \in \{0, 1\}$ for all $T \subseteq N$.

- (4) **Maximality.** For any $\mu \in \mathcal{C}_N$, we have

$$H_D(\mu) \leq H_M(\mu) \leq H_Y(\mu) \leq \ln n.$$

Moreover,

- $H_M(\mu) = H_D(\mu)$ if and only if μ is equientropic [27], that is if

$$H_S(p_m^\mu) = H_S(p_{m'}^\mu), \quad \forall m, m' \in \mathcal{M}_N,$$

- $H_M(\mu) = H_Y(\mu)$ if and only if μ is additive,
 - $H_M(\mu) = \ln n$ if and only if μ is the uniform capacity μ_N^* on N .
- (5) **Increasing monotonicity toward μ_N^* .** Let $\mu \in \mathcal{C}_N \setminus \{\mu_N^*\}$ and, for any $\lambda \in [0, 1]$, define $\mu_\lambda \in \mathcal{C}_N$ by $\mu_\lambda := \mu + \lambda(\mu_N^* - \mu)$. Then for any

$0 \leq \lambda_1 < \lambda_2 \leq 1$ we have

$$H_M(\mu_{\lambda_1}) < H_M(\mu_{\lambda_2}).$$

We now state another very important property of H_M which follows from Eq. (2) and the strict concavity of H_S .

(6) **Strict concavity.** For any $\mu_1, \mu_2 \in \mathcal{C}_N$ and any $\lambda \in]0, 1[$, we have

$$H_M(\lambda \mu_1 + (1 - \lambda) \mu_2) > \lambda H_M(\mu_1) + (1 - \lambda) H_M(\mu_2).$$

An immediate consequence of the previous property is that maximizing H_M over a convex subset of \mathcal{C}_N always leads to a unique global maximum.

For probability distributions, the strict concavity of the Shannon entropy and its naturalness as a measure of uncertainty gave rise to the *maximum entropy principle*, which was pointed out in 1957 by Jaynes [8]. This principle states that, when one has only partial information about the possible outcomes of a random variable, one should choose its probability distribution so as to maximize the uncertainty about the missing information. In other words, all the available information should be used, but one should be as uncommitted as possible about missing information. In more mathematical terms, this principle means that among all the probability distributions that are in accordance with the available prior knowledge (i.e. a set of constraints), one should choose the one that has maximum uncertainty.

The strict concavity of H_M suggests extending such an inference principle to discrete Choquet capacities.³ Note, however, that the interpretation of the maximum entropy principle for capacities is less natural than for probability distributions because the notion of uniformity in the context of capacities is not naturally linked with that of uncertainty as it is the case for probability distributions.

5 Interpretation of the entropy H_M in the aggregation framework

Suppose that the nonempty finite set $N = [n]$ represents a set of criteria in a multicriteria decision making problem and consider a capacity $\mu \in \mathcal{C}_N$. For any $S \subseteq N$, $\mu(S)$ can be interpreted as the weight or the degree of importance of the coalition S of criteria. Hence, in addition to the usual weights on criteria taken separately, weights on any coalition of criteria are also defined, thus allowing to model interaction phenomena among them (see e.g [20,21]).

³ See [30] for an application in cooperative game theory of the maximum entropy principle with H_M .

Monotonicity of μ then means that adding a new element to a coalition cannot decrease its importance. Obviously $\mu(N)$ has the maximal value, being one by convention.

Now, suppose that $x_1, \dots, x_n \in [0, 1]$ represent quantitative evaluations of an object with respect to criteria $1, \dots, n$, respectively. We further assume that these evaluations are commensurate, i.e., defined on the same measurement scale. A global evaluation for this object can then be calculated by means of the discrete Choquet integral with respect to μ , which has proved to be an appropriate extension of the classical weighted arithmetic mean for the aggregation of interacting criteria [21].

Formally, the Choquet integral of $x \in [0, 1]^n$ with respect to a Choquet capacity $\mu \in \mathcal{C}_N$ is defined by

$$C_\mu(x) := \sum_{i \in N} x_{\pi(i)} \left[\mu(\{\pi(i), \dots, \pi(n)\}) - \mu(\{\pi(i+1), \dots, \pi(n)\}) \right],$$

where π is a permutation of Π_N such that $x_{\pi(1)} \leq \dots \leq x_{\pi(n)}$. For more details, see e.g. [21] and the references therein.

For instance, if $x_3 \leq x_1 \leq x_2$, we have

$$\begin{aligned} C_\mu(x_1, x_2, x_3) &= x_3 [\mu(\{3, 1, 2\}) - \mu(\{1, 2\})] \\ &\quad + x_1 [\mu(\{1, 2\}) - \mu(\{2\})] \\ &\quad + x_2 [\mu(\{2\}) - \mu(\emptyset)]. \end{aligned}$$

Whether a partial evaluation x_i will have some influence on the calculation of $C_\mu(x)$ obviously depends upon its corresponding coefficient and hence upon the capacity μ .

It would then be interesting to appraise the degree to which the partial evaluations x_1, \dots, x_n contribute to the calculation of the aggregated value $C_\mu(x)$. In this final section we shall show that the function H_M measures the average value over all $x \in [0, 1]^n$ of the degree to which the arguments x_1, \dots, x_n contribute to the calculation of $C_\mu(x)$.

To demonstrate this, let us first consider the case when μ is additive, that is, when no interaction among criteria is allowed. The Choquet integral then reduces to the weighted arithmetic mean

$$C_\mu(x) = \sum_{i \in N} x_i \mu(i) = \sum_{i \in N} x_i p(i),$$

where p is the probability distribution defined by $p(i) := \mu(i)$ for all $i \in N$. In this case, the function $H_S(p)$ behaves like a dispersion index [2], which

measures the uniformity (evenness) of the weights $p(i)$. For example,

- if $H_S(p)$ is close to $\ln n$, then the weights are distributed among all criteria almost evenly,
- if $H_S(p)$ is close to zero, then the total weight is focused almost on only one criterion.

In other terms, $H_S(p)$ measures the extent to which all the individual criteria contribute to the aggregation process, or equivalently, the extent to which the arguments x_1, \dots, x_n contribute to the calculation of the aggregated value $C_\mu(x)$. In that sense, the more uniform the probability distribution p the more the arguments x_1, \dots, x_n contribute to the aggregation process.

Now consider a general (non-additive) capacity $\mu \in \mathcal{C}_N$ and define the sets

$$\mathcal{O}_\pi := \{x \in [0, 1]^n \mid x_{\pi(1)} \leq \dots \leq x_{\pi(n)}\} \quad (\pi \in \Pi_N)$$

which cover the hypercube $[0, 1]^n$.

Let $x \in [0, 1]^n$ be fixed. Then there exists $\pi \in \Pi_N$ such that $x \in \mathcal{O}_\pi$ and hence

$$C_\mu(x) = \sum_{i \in N} x_{\pi(i)} p_\pi^\mu(i).$$

The permutation π corresponds to the maximal chain m^π , along which the Choquet integral boils down to a weighted arithmetic mean whose weights are defined by the probability distribution p_π^μ . In that case, $H_S(p_\pi^\mu)$ measures the uniformity of the distribution p_π^μ , that is, the regularity of the increasingness of μ along the chain m^π . Equivalently, it measures the degree to which the arguments x_1, \dots, x_n contribute⁴ to the calculation of the aggregated value $C_\mu(x)$.

Starting from Eq. (3) and using the fact that $\int_{x \in \mathcal{O}_\pi} dx = 1/n!$ the entropy $H_M(\mu)$ can be rewritten as

$$H_M(\mu) = \sum_{\pi \in \Pi_N} \int_{x \in \mathcal{O}_\pi} H_S(p_\pi^\mu) dx = \int_{[0,1]^n} H_S(p_{\pi_x}^\mu) dx,$$

where π_x is a permutation of Π_N such that $x \in \mathcal{O}_{\pi_x}$.

We thus observe that $H_M(\mu)$ measures the average value over all $x \in [0, 1]^n$ of the degree to which the arguments x_1, \dots, x_n contribute to the calculation

⁴ Should $H_S(p_\pi^\mu)$ be close to $\ln n$, the distribution p_π^μ will be approximately uniform and all the partial evaluations x_1, \dots, x_n will be involved almost equally in the calculation of $C_\mu(x)$, which will be close to the arithmetic mean of the x_i 's. On the contrary, should $H_S(p_\pi^\mu)$ be close to zero, one $p_\pi^\mu(i)$ will be very close to one and $C_\mu(x)$ will be very close to the corresponding partial evaluation.

of $C_\mu(x)$. From Eq. (2), it can also be interpreted as a measure of the average regularity of the increasingness of μ over all maximal chains $m \in \mathcal{M}_N$.

In probabilistic terms, it corresponds to the expectation over all $x \in [0, 1]^n$, with uniform distribution, of the degree of contribution of arguments x_1, \dots, x_n in the calculation of $C_\mu(x)$. From Eq. (2), it also corresponds to the expectation over all maximal chains $m \in \mathcal{M}_N$, with uniform distribution, of the regularity of the increasingness of μ .

It should also be mentioned that the interpretation of $H_M(\mu)$ as an average degree of contribution of the arguments is in full accordance with the properties listed in Section 4 (see [2,3]). For example, the decisivity property, which states that $H_M(\mu) = 0$ if and only if μ is binary-valued, is quite relevant for our aggregation framework. Indeed, it can be shown that this latter condition holds if and only if

$$C_\mu(x) \in \{x_1, \dots, x_n\} \quad (x \in [0, 1]^n).$$

In other terms, $H_M(\mu)$ is minimum ($= 0$) if and only if only one partial evaluation contributes in the calculation of $C_\mu(x)$.

To conclude this section, we give an interpretation of the maximum entropy principle stated in the previous section in the framework of aggregation by the Choquet integral.

Assume that we are given a set of linear constraints on the behavior of a Choquet integral C_μ , that is, constraints that are linear with respect to the corresponding capacity μ . Then, among all the feasible (admissible) Choquet integrals, choosing the Choquet integral with respect to the maximum entropy Choquet capacity amounts to choosing the Choquet integral that will have the highest average degree of contribution of its arguments in the aggregation phase. In other words, we could say the Choquet integral with respect to the maximum entropy Choquet capacity is the one that will exploit the most on average its arguments.

6 Conclusion

We have proposed an axiomatic characterization of the concept of entropy of a discrete Choquet capacity, recently introduced on intuitive grounds in the framework of aggregation by the Choquet integral.

Depending upon the context in which this entropy is used, it can be interpreted either as a quantification of the uniformity of a Choquet capacity, that is, a

measure of the average regularity of a Choquet capacity along all maximal chains, or as the degree of average contribution of arguments in the framework of aggregation by the Choquet integral.

A Proof of Theorem 2

In order to prove Theorem 2, we shall go through three technical lemmas. Moreover, for any cardinality-based capacity $\mu \in \mathcal{C}_N$, we shall use the notation $\mu_t := \mu(T)$ for all $T \subseteq N$ such that $|T| = t$, with $t \in \{0, \dots, n\}$.

Lemma 3 *If the function $H : \mathcal{C} \rightarrow \mathbb{R}^+$ fulfills axiom SE then, for any finite set $N \subset U$, with $n \geq 2$, any integer k such that $2 \leq k \leq n$, any partition $\{A_1, \dots, A_k\}$ of N , and any cardinality-based capacity $\mu \in \mathcal{C}_N$, we have*

$$\begin{aligned} & \mu_{\bigcup_{j \in S} A_j}^{A_i}(A_i) H(\bar{\mu}_{\bigcup_{j \in S} A_j}^{A_i}) \\ &= \sum_{t=1}^{a_i} h \left[\mu_{\sum_{j \in S} a_j + t} - \mu_{\sum_{j \in S} a_j + t - 1} \right] - h \left[\mu_{\sum_{j \in S} a_j + a_i} - \mu_{\sum_{j \in S} a_j} \right] \quad (\text{A.1}) \end{aligned}$$

for all $i \in [k]$ and all $S \subseteq [k] \setminus i$.

Proof. Consider a finite set $N \subset U$, $n \geq 2$, a partition $\{A_1, \dots, A_k\}$ of N , with $2 \leq k \leq n$, and a cardinality-based capacity $\mu \in \mathcal{C}_N$. Then, there exist $\mu_0, \dots, \mu_n \in [0, 1]$ such that $\mu(T) = \mu_t$ for all $T \subseteq N$. Moreover, it is easy to check that the capacities $\bar{\mu}_{\bigcup_{j \in S} A_j}^{A_i}$ are cardinality-based. From axiom SE, for any $i \in [k]$ and any $S \subseteq [k] \setminus i$, we then have

$$\begin{aligned} & \mu_{\bigcup_{j \in S} A_j}^{A_i}(A_i) H(\bar{\mu}_{\bigcup_{j \in S} A_j}^{A_i}) \\ &= \left(\mu_{\sum_{j \in S} a_j + a_i} - \mu_{\sum_{j \in S} a_j} \right) \sum_{t=1}^{a_i} h \left[\frac{\mu_{\sum_{j \in S} a_j + t} - \mu_{\sum_{j \in S} a_j + t - 1}}{\mu_{\sum_{j \in S} a_j + a_i} - \mu_{\sum_{j \in S} a_j}} \right], \\ &= \sum_{t=1}^{a_i} h \left[\mu_{\sum_{j \in S} a_j + t} - \mu_{\sum_{j \in S} a_j + t - 1} \right] \\ & \quad + \ln \left(\mu_{\sum_{j \in S} a_j + a_i} - \mu_{\sum_{j \in S} a_j} \right) \sum_{t=1}^{a_i} \left(\mu_{\sum_{j \in S} a_j + t} - \mu_{\sum_{j \in S} a_j + t - 1} \right), \end{aligned}$$

which completes the proof. \square

Lemma 4 *If the function $H : \mathcal{C} \rightarrow \mathbb{R}^+$ fulfills axioms S, SE, and R, then, for any integer $k \geq 2$ and any finite set $N \subset U$, with $n = k(2^k - 1)$, there exists a family of constants $\alpha_0(n, k), \dots, \alpha_{k-1}(n, k)$ such that*

$$\alpha_S^i(n, k) = \alpha_s(n, k), \quad \forall i \in [k], \forall S \subseteq [k] \setminus i.$$

Proof. Let $k \geq 2$ be an integer and consider $N \subset U$ such that $n = k(2^k - 1)$. Let $\mu \in \mathcal{C}_N$ be cardinality-based and let $\{A_1, \dots, A_k\}$ be a partition of N .

Let $\pi \in \Pi_{[k]}$. It is easy to check that

$$H(\mu^{[A_1] \dots [A_k]}) = H(\mu^{[A_{\pi(1)}] \dots [A_{\pi(k)}]}). \quad (\text{A.2})$$

Indeed, defining $u, v \in \mathcal{C}_{[k]}$ by

$$u(S) := \mu\left(\bigcup_{i \in S} A_{\pi(i)}\right), \quad \forall S \subseteq [k],$$

and

$$v(S) := \mu\left(\bigcup_{i \in S} A_i\right), \quad \forall S \subseteq [k],$$

we clearly see that $v(\pi(S)) = u(S)$ for all $S \subseteq [k]$ and hence $v = \pi u$. By (4) and axiom S, we then have

$$H(\mu^{[A_1] \dots [A_k]}) = H(v) = H(u) = H(\mu^{[A_{\pi(1)}] \dots [A_{\pi(k)}]}).$$

From axiom R, we know that the functional equation (7) must hold for all partitions of N into k subsets. Writing it for $\{A_1, \dots, A_k\}$ and for $\{A_{\pi(1)}, \dots, A_{\pi(k)}\}$ and making use of (A.2), we have

$$\begin{aligned} \sum_{i=1}^k \sum_{S \subseteq [k] \setminus i} \alpha_S^i(n, k) \mu_{\bigcup_{j \in S} A_j}^{A_i} (A_i) H(\bar{\mu}_{\bigcup_{j \in S} A_j}^{A_i}) \\ = \sum_{i=1}^k \sum_{S \subseteq [k] \setminus i} \alpha_S^i(n, k) \mu_{\bigcup_{j \in S} A_{\pi(j)}}^{A_{\pi(i)}} (A_{\pi(i)}) H(\bar{\mu}_{\bigcup_{j \in S} A_{\pi(j)}}^{A_{\pi(i)}}). \end{aligned}$$

Rewriting the right-hand side of the previous equation, we obtain

$$\sum_{i=1}^k \sum_{S \subseteq [k] \setminus i} \left(\alpha_S^i(n, k) - \alpha_{\pi^{-1}(S)}^{\pi^{-1}(i)}(n, k) \right) \mu_{\bigcup_{j \in S} A_j}^{A_i} (A_i) H(\bar{\mu}_{\bigcup_{j \in S} A_j}^{A_i}) = 0,$$

that is, by Lemma 3,

$$\begin{aligned} \sum_{i=1}^k \sum_{S \subseteq [k] \setminus i} \left(\alpha_S^i(n, k) - \alpha_{\pi^{-1}(S)}^{\pi^{-1}(i)}(n, k) \right) \\ \times \left[\sum_{t=1}^{a_i} h[\mu_{\sum_{j \in S} a_j + t} - \mu_{\sum_{j \in S} a_j + t - 1}] - h[\mu_{\sum_{j \in S} a_j + a_i} - \mu_{\sum_{j \in S} a_j}] \right] = 0, \end{aligned}$$

which must hold for all cardinality-based capacities $\mu \in \mathcal{C}_N$ and all partitions $\{A_1, \dots, A_k\}$ of N .

By choosing μ and $\{A_1, \dots, A_k\}$ in an appropriate way, we will now show that

$$\alpha_S^i(n, k) = \alpha_{\pi^{-1}(S)}^{\pi^{-1}(i)}(n, k), \quad \forall \pi \in \Pi_{[k]}, \forall i \in [k], \forall S \subseteq [k] \setminus i,$$

which will complete the proof of the lemma.

Consider a partition $\{B_1, \dots, B_k\}$ of N such that, for all $i \in [k]$, $b_i = k 2^{i-1}$. This is possible since we have chosen $n = k(2^k - 1) = \sum_{i=1}^k k 2^{i-1}$. Such a choice ensures that the partial sums $\sum_{i \in S} b_i$ ($S \subseteq [k]$) all are different.

Let us fix $i^* \in [k]$, $S^* \subseteq [k] \setminus i^*$, $\sigma \in \Pi_{[k]}$ such that $\sigma(i^*) = k$, and choose the partition $\{A_1, \dots, A_k\}$ such that $A_i = B_{\sigma(i)}$ for all $i \in [k]$. Finally, define $\mu \in \mathcal{C}_N$ by

$$\mu(T) := \begin{cases} 0, & \text{if } t \leq \sum_{j \in S^*} a_j \\ 1/2, & \text{if } \sum_{j \in S^*} a_j < t < \sum_{j \in S^*} a_j + a_{i^*} \\ 1, & \text{if } t \geq \sum_{j \in S^*} a_j + a_{i^*} \end{cases}$$

with $a_{i^*} = b_k = k 2^{k-1} \geq 4$.

Then, we can show that

$$\begin{aligned} \sum_{t=1}^{a_i} h \left[\mu_{\sum_{j \in S} a_j + t} - \mu_{\sum_{j \in S} a_j + t - 1} \right] - h \left[\mu_{\sum_{j \in S} a_j + a_i} - \mu_{\sum_{j \in S} a_j} \right] \\ = \begin{cases} \ln 2, & \text{if } i = i^* \text{ and } S = S^*, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Indeed, by definition of μ , the left-hand side of the previous identity will be strictly positive (with value $\ln 2$) if and only if two elements in the sequence

$$\Delta_S^i := \left(\mu_{\sum_{j \in S} a_j + t} - \mu_{\sum_{j \in S} a_j + t - 1} \right)_{t=1, \dots, a_i}$$

are $1/2$.

The choice of the cardinality-based capacity μ ensures that only the first and the last elements of the sequence $\Delta_{S^*}^{i^*}$ are equal to $1/2$.

The choice of the partition $\{A_1, \dots, A_k\}$ ensures that there exists no pair $(i, S) \neq (i^*, S^*)$ such that $\Delta_{S^*}^{i^*}$ is a subsequence of Δ_S^i . Indeed, $a_{i^*} > a_i$ for all $i \in [k] \setminus i^*$ and the partial sums $\sum_{i \in S} a_i$ ($S \subseteq [k]$) all are different. \square

Lemma 5 *If the function $H : \mathcal{C} \rightarrow \mathbb{R}^+$ fulfills axioms S , SE , and R , then, for any integer $k \geq 2$ and any finite set $N \subset U$, with $n = k(2^k - 1)$, we have*

$$\alpha_s(n, k) = \gamma_s(k), \quad \forall s = 0, \dots, k - 1.$$

Proof. Let $k \geq 2$ be an integer and consider $N \subset U$ such that $n = k(2^k - 1)$. Let $\mu \in \mathcal{C}_N$ be cardinality-based and let $\{A_1, \dots, A_k\}$ be a partition of N such that $a_1 = \dots = a_k := a$, with $a = 2^k - 1 > 1$.

From axiom SE, we have that

$$H(\mu) = H_S(p^\mu) = \sum_{i=1}^n h[\mu_i - \mu_{i-1}],$$

which can be rewritten as

$$\begin{aligned} H(\mu) &= \sum_{s=0}^{k-1} \sum_{t=1}^a h[\mu_{sa+t} - \mu_{sa+t-1}] \\ &= k \sum_{s=0}^{k-1} \binom{k-1}{s} \gamma_s(k) \sum_{t=1}^a h[\mu_{sa+t} - \mu_{sa+t-1}], \end{aligned}$$

that is,

$$H(\mu) = \sum_{i=1}^k \sum_{S \subseteq [k] \setminus i} \gamma_s(k) \sum_{t=1}^a h[\mu_{sa+t} - \mu_{sa+t-1}]. \quad (\text{A.3})$$

Since the subsets A_i have the same cardinality, it is easy to check that the capacity $\mu^{[A_1] \dots [A_k]}$ is cardinality-based. According to axiom SE, $H(\mu^{[A_1] \dots [A_k]})$ can be written as

$$\begin{aligned} H(\mu^{[A_1] \dots [A_k]}) &= \sum_{i=1}^k h[p^{\mu^{[A_1] \dots [A_k]}}(i)] = \sum_{i=1}^k h[\mu_{ia} - \mu_{(i-1)a}] \\ &= \sum_{s=0}^{k-1} h[\mu_{(s+1)a} - \mu_{sa}] \\ &= k \sum_{s=0}^{k-1} \binom{k-1}{s} \gamma_s(k) h[\mu_{(s+1)a} - \mu_{sa}] \end{aligned}$$

that is,

$$H(\mu^{[A_1] \dots [A_k]}) = \sum_{i=1}^k \sum_{S \subseteq [k] \setminus i} \gamma_s(k) h[\mu_{(s+1)a} - \mu_{sa}]. \quad (\text{A.4})$$

Using Lemma 3 in the case where the numbers a_i all are equal, for all $i \in [k]$ and all $S \subseteq [k] \setminus i$, we obtain

$$\mu_{\bigcup_{j \in S} A_j}^{A_i}(A_i) H(\bar{\mu}_{\bigcup_{j \in S} A_j}^{A_i}) = \sum_{t=1}^a h[\mu_{sa+t} - \mu_{sa+t-1}] - h[\mu_{(s+1)a} - \mu_{sa}]. \quad (\text{A.5})$$

Substituting in the functional equation (7) of axiom R the terms

$$H(\mu), H(\mu^{[A_1] \dots [A_k]}), \text{ and } \mu_{\cup_{j \in S} A_j}^{A_i} H(\bar{\mu}_{\cup_{j \in S} A_j}^{A_i})$$

by their expressions given respectively in Eq. (A.3), (A.4), and (A.5), we obtain, using Lemma 4,

$$\sum_{i=1}^k \sum_{S \subseteq [k] \setminus i} (\gamma_s(k) - \alpha_s(n, k)) \left[\sum_{t=1}^a h[\mu_{sa+t} - \mu_{sa+t-1}] - h[\mu_{(s+1)a} - \mu_{sa}] \right] = 0,$$

that is,

$$\begin{aligned} \sum_{s=0}^{k-1} \binom{k-1}{s} (\gamma_s(k) - \alpha_s(n, k)) \\ \times \left[\sum_{t=1}^a h[\mu_{sa+t} - \mu_{sa+t-1}] - h[\mu_{(s+1)a} - \mu_{sa}] \right] = 0, \end{aligned} \quad (\text{A.6})$$

which must hold for all cardinality-based capacities $\mu \in \mathcal{C}_N$. By choosing μ in an appropriate way, we will now show that

$$\alpha_s(n, k) = \gamma_s(k), \quad \forall s = 0, \dots, k-1,$$

which will complete the proof of the lemma.

Let us fix $s^* \in \{0, \dots, k-1\}$ and define $\mu \in \mathcal{C}_N$ by

$$\mu(T) := \begin{cases} 0, & \text{if } t \leq s^*a, \\ \frac{t - s^*a}{a}, & \text{if } s^*a \leq t \leq (s^* + 1)a, \\ 1, & \text{if } t \geq (s^* + 1)a. \end{cases}$$

Using axiom SE, we can easily show that Eq. (A.6) becomes

$$\binom{k-1}{s^*} (\gamma_{s^*}(k) - \alpha_{s^*}(n, k)) \ln a = 0,$$

where $a > 1$, which is sufficient. \square

We now prove Theorem 2.

Proof. (Necessity) Let $H : \mathcal{C} \rightarrow \mathbb{R}^+$ fulfill axioms S, SE, and R. Let $k \geq 2$ be an integer and let $\nu \in \mathcal{C}_{[k]}$. We only need to prove that $H(\nu) = H_M(\nu)$.

Let us show first that, given a set $N \subset U$, with $n = k(2^k - 1)$, there exists a partition $\{A_1, \dots, A_k\}$ of N and a cardinality-based capacity $\mu \in \mathcal{C}_N$ such

that $\nu = \mu^{[A_1] \cdots [A_k]}$, that is,

$$\nu(S) = \mu^{[A_1] \cdots [A_k]} \left(\bigcup_{i \in S} [A_i] \right) = \mu_{\sum_{i \in S} a_i} \quad \forall S \subseteq [k].$$

To prove this, it suffices to consider sets A_1, \dots, A_k such that $a_i = k 2^{i-1}$ ($i \in [k]$). In that case, the partial sums $\sum_{i \in S} a_i$ ($S \subseteq [k]$) all are different and hence we can always define a cardinality-based capacity $\mu \in \mathcal{C}_N$, with $n = k(2^k - 1)$, such that

$$\mu_{\sum_{i \in S} a_i} = \nu(S) \quad \forall S \subseteq [k].$$

Now, consider a set function $v : 2^U \rightarrow \mathbb{R}$ with carrier $[k]$ defined by

$$v(R) := \begin{cases} 0, & \text{if } R = \emptyset, \\ \sum_{j \in R} a_j \\ \sum_{t=1} h[\mu_t - \mu_{t-1}], & \text{if } \emptyset \neq R \subseteq [k]. \end{cases}$$

It is easy to check that v is a capacity.

We then have

$$\sum_{t=1}^{a_i} h[\mu_{t+\sum_{j \in S} a_j} - \mu_{t-1+\sum_{j \in S} a_j}] = \sum_{t=1+\sum_{j \in S} a_j}^{a_i+\sum_{j \in S} a_j} h[\mu_t - \mu_{t-1}] = v(S \cup i) - v(S).$$

and hence

$$\begin{aligned} & \sum_{i=1}^k \sum_{S \subseteq [k] \setminus i} \gamma_s(k) \sum_{t=1}^{a_i} h[\mu_{t+\sum_{j \in S} a_j} - \mu_{t-1+\sum_{j \in S} a_j}] \\ &= \sum_{i=1}^k \sum_{S \subseteq [k] \setminus i} \gamma_s(k) [v(S \cup i) - v(S)] \\ &= v([k]) \quad (\text{cf. Eq. (1)}) \\ &= \sum_{t=1}^n h[\mu_t - \mu_{t-1}] \\ &= H_S(p^\mu). \end{aligned}$$

Finally, by axiom SE, we obtain

$$\sum_{i=1}^k \sum_{S \subseteq [k] \setminus i} \gamma_s(k) \sum_{t=1}^{a_i} h[\mu_{t+\sum_{j \in S} a_j} - \mu_{t-1+\sum_{j \in S} a_j}] = H(\mu). \quad (\text{A.7})$$

Therefore, combining (A.1) and (A.7), we have

$$\begin{aligned} \sum_{i=1}^k \sum_{S \subseteq [k] \setminus i} \gamma_s(k) \mu_{\cup_{j \in S} A_j}^{A_i}(A_i) H(\bar{\mu}_{\cup_{j \in S} A_j}^{A_i}) \\ = H(\mu) - \sum_{i=1}^k \sum_{S \subseteq [k] \setminus i} \gamma_s(k) h[\nu(S \cup i) - \nu(S)]. \end{aligned}$$

Finally, using Lemmas 4 and 5, Eq. (7) becomes

$$H(\mu^{[A_1] \cdots [A_k]}) = \sum_{i=1}^k \sum_{S \subseteq [k] \setminus i} \gamma_s(k) h[\nu(S \cup i) - \nu(S)],$$

that is,

$$H(\nu) = H_M(\nu).$$

(Sufficiency) We know that H_M fulfills axioms S and SE. Let us show that it fulfills axiom R.

Consider an integer $k \geq 2$ and a finite set $N \subset U$ such that $n \geq k$. Let $\mu \in \mathcal{C}_N$ be cardinality-based and let $\{A_1, \dots, A_k\}$ be a partition of N . By Lemma 3, we have

$$\begin{aligned} \mu_{\cup_{j \in S} A_j}^{A_i}(A_i) H_M(\bar{\mu}_{\cup_{j \in S} A_j}^{A_i}) \\ = \sum_{t=1}^{a_i} h[\mu_{\sum_{j \in S} a_j + t} - \mu_{\sum_{j \in S} a_j + t - 1}] - h[\mu_{\sum_{j \in S} a_j + a_i} - \mu_{\sum_{j \in S} a_j}] \end{aligned}$$

and hence (cf. (A.7))

$$\begin{aligned} \sum_{i=1}^k \sum_{S \subseteq [k] \setminus i} \gamma_s(k) \mu_{\cup_{j \in S} A_j}^{A_i}(A_i) H_M(\bar{\mu}_{\cup_{j \in S} A_j}^{A_i}) \\ = H_M(\mu) - \sum_{i=1}^k \sum_{S \subseteq [k] \setminus i} \gamma_s(k) h[\mu_{\sum_{j \in S} a_j + a_i} - \mu_{\sum_{j \in S} a_j}] \\ = H_M(\mu) - H_M(\mu^{[A_1] \cdots [A_k]}). \end{aligned}$$

We have therefore shown that H_M fulfills also axiom R which completes the proof. \square

References

- [1] J.-L. Marichal, Aggregation operators for multicriteria decision aid, Ph.D. thesis, Institute of Mathematics, University of Liège, Liège, Belgium (December

1998).

- [2] J.-L. Marichal, Entropy of discrete Choquet capacities, *European J. Oper. Res.* 137 (3) (2002) 612–624.
- [3] J.-L. Marichal, M. Roubens, Entropy of discrete fuzzy measures, *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems* 8 (6) (2000) 625–640.
- [4] I. Kojadinovic, J.-L. Marichal, M. Roubens, An axiomatic approach to the definition of the entropy of a discrete Choquet capacity, in: *Proc. 9th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2002)*, Annecy, France, 2002, pp. 763–768.
- [5] C. Shannon, A mathematical theory of communication, *Bell System Technical Journal* 27 (1948) 379–423, 623–656.
- [6] D. Morales, L. Pardo, I. Vajda, Uncertainty of discrete stochastic systems: general theory and statistical inference, *IEEE Transactions on Systems, Man and Cybernetics. Part A: Systems and Humans* 26 (6) (1996) 681–697.
- [7] S. Kullback, R. Leibler, On information and sufficiency, *The Annals of Mathematical Statistics* 22 (1951) 79–86.
- [8] E. Jaynes, Information theory and statistical mechanics, *Physical Review* 106 (4) (1957) 620–630.
- [9] J. Aczél, Z. Daróczy, On measures of information and their characterizations, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1975, *Mathematics in Science and Engineering*, Vol. 115.
- [10] B. R. Ebanks, P. Sahoo, W. Sander, *Characterization of information measures*, Singapore: World Scientific, 1997.
- [11] E. Jaynes, *Probability theory: the logic of science*, Cambridge University Press, 2003.
- [12] A. Khinchin, *Mathematical foundations of information theory*, New York: Dover Publications, 1957.
- [13] G. Choquet, Theory of capacities, *Ann. Inst. Fourier, Grenoble* 5 (1953–1954) 131–295 (1955).
- [14] M. Sugeno, *Theory of fuzzy integrals and its applications*, Ph.D. thesis, Tokyo Institute of Technology, Tokyo (1974).
- [15] G. J. Klir, B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications*, Upper Saddle River, NJ: Prentice Hall, 1995.
- [16] G. J. Klir, M. J. Wierman, *Uncertainty-based information. Elements of generalized information theory*. 2nd corr. ed., *Studies in Fuzziness and Soft Computing*. 15. Heidelberg: Physica-Verlag, 1999.
- [17] D. Dubois, H. Prade, *Possibility theory. An approach to computerized processing of uncertainty*, New York: Plenum Press, 1988.

- [18] L. Shapley, A value for n -person games., in: Contrib. Theory of Games, II, Ann. Math. Stud. No. 28, 307-317, 1953.
- [19] M. Grabisch, J.-L. Marichal, M. Roubens, Equivalent representations of set functions, Math. Oper. Res. 25 (2) (2000) 157–178.
- [20] M. Grabisch, Alternative representations of discrete fuzzy measures for decision making, Internat. J. Uncertain. Fuzziness Knowledge-Based Systems 5 (1997) 587–607.
- [21] J.-L. Marichal, An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria, IEEE Trans. Fuzzy Syst. 8 (6) (2000) 800–807.
- [22] D. Dubois, H. Prade, Properties of measures of information in evidence and possibility theories, Fuzzy Sets and Systems 24 (1987) 161–182.
- [23] D. Dubois, A. Ramer, Extremal properties of belief measures in the theory of evidence, Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems 1 (1) (1993) 57–68.
- [24] G. J. Klir, R. M. Smith, On measuring uncertainty and uncertainty-based information: recent developments, Annals of Mathematics and Artificial Intelligence 32 (1–4) (2001) 5–33.
- [25] R. R. Yager, A class of fuzzy measures generated from a Dempster-Shafer belief structure, Int. J. Intell. Syst. 14 (12) (1999) 1239–1247.
- [26] R. R. Yager, On the entropy of fuzzy measures, IEEE Transactions on Fuzzy Systems 8 (4) (2000) 453–461.
- [27] A. Dukhovny, General entropy of general measures, Internat. J. Uncertain. Fuzziness Knowledge-Based Systems 10 (3) (2002) 213–225.
- [28] M. Grabisch, M. Roubens, An axiomatic approach to the concept of interaction among players in cooperative games., Int. J. Game Theory 28 (4) (1999) 547–565.
- [29] A. Rényi, On the foundations of information theory, Rev. Inst. Int. Stat. 33 (1965) 1–14.
- [30] J.-L. Marichal, The influence of variables on pseudo-Boolean functions with applications to game theory and multicriteria decision making, Discrete Appl. Math. 107 (1-3) (2000) 139–164.