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

Ivan Kojadinovic

Iremia, Université de la Réunion

15 avenue René Cassin - BP 7151, 97715 Saint-Denis messag cedex 9

Ile de la Réunion, France.

Jean-Luc Marichal

Department of Mathematics, Brigham Young University

TMCB, Provo UT 84602, U.S.A.

Marc Roubens

Institute of Mathematics, University of Liège

Grande Traverse 12 (B37), B-4000 Liège, Belgium.

Working paper, October 2, 2001

Abstract

An axiomatization of the concept of entropy of a discrete Choquet capacity is given. It is based on three axioms: the symmetry property, a boundary condition for which the entropy reduces to the classical Shannon entropy, and a generalized version of the well-known recursivity property. This entropy, recently introduced to extend the Shannon entropy to non-additive measures, fulfills several properties considered as requisites for defining an entropy-like measure of uncertainty. An interpretation of it in the framework of aggregation by the discrete Choquet integral is given as well.

Keywords: entropy, Choquet capacity, Choquet integral, information theory.

1 Introduction

In 1948, Shannon introduced a measure of uncertainty of a discrete stochastic system known as *entropy* [19]. For a probability distribution p defined on a finite set $[n] = \{1, \dots, n\}$, the Shannon entropy of p is defined by

$$H_S(p) := -\sum_{i=1}^{n} p(i) \ln p(i)$$

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

Although several other measures of uncertainty have been proposed as generalizations of the Shannon entropy (see e.g. [17] for an overview), the most widely used uncertainty remains that of Shannon mainly because of its attractive properties, its connections with the *Kullback-Leibler divergence* [12] and its role in the *maximum entropy principle* [9].

Several axiomatic characterizations of the Shannon entropy have been proposed in the literature (see e.g. [1, 3, 4, 10, 11]), amongst which the most famous is probably *Shannon's theorem* [19].

By replacing the additivity property of probability measures by that of monotonicity, one obtains *Choquet capacities* [2] also known as *fuzzy measures* [21] which are able to model other types of possibly uncertain information. More formally, a discrete Choquet capacity μ on $[n]$ is a monotone set function defined on the power set of $[n]$ that is zero at the empty set. Such a concept can be used to model the *importance* of a coalition of elements of $[n]$. The label set $[n]$ could correspond to criteria in a multicriteria decision problem [5, 6, 15], to players in a cooperative game [6, 8, 20], to attributes in a classification problem [7], to variables in a regression problem, voters in an opinion pooling problem, etc. In all these cases, for any subset $S \subseteq [n]$, $\mu(S)$ can be interpreted as the *degree of importance* or the *strength* of the coalition S of elements for the particular problem under consideration.

A discrete Choquet capacity being clearly a generalization of a discrete probability distribution, the following natural question arises : How could one appraise the "uncertainty" associated with a Choquet capacity in the spirit of the Shannon entropy? Recently, Marichal [13, 14, 16] proposed the notion of *entropy of a discrete Choquet capacity* as a generalization of the Shannon entropy and showed that it satisfies many properties that one would intuitively require from such a measure. This generalized entropy is defined as

$$H_M(\mu) := \sum_{i=1}^{n} \sum_{S \subseteq [n] \setminus i} \gamma_s(n)\, h[\mu(S \cup i) - \mu(S)] \tag{1}$$

where the functions

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

and

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

will be used throughout.

Note that this formulation is very close to that of the *Shapley value* [20] of a Choquet capacity μ on $[n]$, which is a fundamental concept in game theory. For an element $i \in [n]$, it 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

$$\sum_{i=1}^{n} \phi_i(\mu) = \mu([n]). \tag{2}$$

In this paper, after defining the notion of *uncertainty* in the setting of discrete Choquet capacities (Section 2), we propose a characterization of the generalized entropy (1) 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 Uncertainty contained in a discrete Choquet capacity

Although discrete Choquet capacities can be seen as generalizations of discrete probability distributions, it is not clear what "uncertainty" should mean for such non-additive measures. After introducing the notation, we propose an intuitive definition of the notion of *uncertainty* based on the lattice representation of a discrete Choquet capacity.

2.1 Notation and first definitions

Throughout this paper, $[n] = \{1, \dots, n\}$ is a finite label set, the power set of which is denoted $\mathcal{P}([n])$.

A *discrete Choquet capacity* [2] or *discrete fuzzy measure* [21] on $[n]$ is a set function $\mu : \mathcal{P}([n]) \to \mathbb{R}^+$ satisfying the following conditions :

1. $\mu(\emptyset) = 0$,
2. for all $S, T \subseteq [n]$, $S \subseteq T \Rightarrow \mu(S) \leq \mu(T)$.

The Choquet capacity is said to be *normalized* if $\mu([n]) = 1$. In the sequel, we restrict ourselves to the study of the notion of *uncertainty* only for normalized Choquet capacities. For any integer $n \geq 1$, we denote by $\mathcal{F}_{[n]}$ the set of all normalized Choquet capacity on $[n]$.

In order to avoid a heavy notation, we adopt that used in [15]. Thus, we will omit braces for singletons, e.g., by writing $\mu(i)$, $[n] \setminus i$ instead of $\mu(\{i\})$, $[n] \setminus \{i\}$. Furthermore, cardinalities of subsets $S, T, \dots$, will be denoted by the corresponding lower case letters $s, t, \dots$

A Choquet capacity $\mu \in \mathcal{F}_{[n]}$ 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 cardinal of T. In this case, there exist $\mu_1, \dots, \mu_n \in [0, 1]$ such that $\mu(T) = \mu_t$ for all $T \subseteq N$.

Figure 1: Hasse diagram $\mathcal{H}_{[n]}$ corresponding to the lattice of subsets of $[n] = \{1, 2, 3, 4\}$.

There is only one Choquet capacity on $[n]$ that is both additive and cardinality-based. We shall call it *the uniform Choquet capacity* on $[n]$ and denote it by μ^*. It is easy to check that μ^* is given by

$$\mu^*(T) = t/n \qquad \forall T \subseteq [n].$$

2.2 Choquet capacities and maximal chains

Consider the lattice $\mathcal{L}_{[n]}$ related to the power set of $[n]$ under the inclusion relation. The lattice $\mathcal{L}_{[n]}$ 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 subsets denoted

$$m = (\emptyset \subsetneq \{i_1\} \subsetneq \{i_1, i_2\} \subsetneq \cdots \subsetneq \{i_1, \ldots, i_n\} = [n]).$$

For instance, in Figure 1, the maximal chain $(\emptyset \subsetneq \{1\} \subsetneq \{1,4\} \subsetneq \{1,2,4\} \subsetneq \{1,2,3,4\})$ is emphasized.

We denote by $\mathcal{C}_{[n]}$ the set of maximal chains of $\mathcal{H}_{[n]}$. The cardinality of $\mathcal{C}_{[n]}$ is clearly $n!$.

Given a Choquet capacity μ on $[n]$, with each maximal chain $m \in \mathcal{C}_{[n]}$ can be associated a discrete probability distribution p_m^μ on $[n]$ defined by

$$p_m^\mu(i) := \mu(m_i) - \mu(m_{i-1}) \qquad \forall i \in [n]$$

where m_i denotes the subset of cardinal i of m.

Denote by $\Pi_{[n]}$ the set of permutations on $[n]$. With each permutation $\pi \in \Pi_{[n]}$ is associated a unique maximal chain $m^\pi \in \mathcal{C}_{[n]}$ defined by

$$m^\pi = (\emptyset \subsetneq \{\pi(n)\} \subsetneq \{\pi(n-1), \pi(n)\} \subsetneq \cdots \subsetneq \{\pi(1), \ldots, \pi(n)\} = [n]).$$

We then write

$$p_\pi^\mu(i) := p_{m^\pi}^\mu(i) = \mu(\{\pi(i), \ldots, \pi(n)\}) - \mu(\{\pi(i+1), \ldots, \pi(n)\}) \qquad \forall i \in [n].$$

The set $\{p_\pi^\mu\}_{\pi \in \Pi_{[n]}} = \{p_m^\mu\}_{m \in \mathcal{C}_{[n]}}$ of $n!$ probability distributions obtained from μ will be denoted P^μ as we continue.

Notice that, if μ 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 μ is additive then we simply have $p_\pi^\mu(i) = \mu(\pi(i))$ for all $i \in [n]$.

2.3 Uncertainty associated with a discrete Choquet capacity

For a discrete probability distribution, the notion of *uncertainty* has an intuitive meaning which is directly linked with that of *uniformity*. Indeed, the more "uniform" a discrete probability distribution, the higher the uncertainty of the underlying discrete stochastic system (for a discussion on the notion of uncertainty, see e.g. [17]).

Although discrete Choquet capacities can be clearly seen as generalizations of discrete probability distributions, it is not clear what "uncertainty" should mean for such non-additive measures. As we have seen in the previous subsection, a Choquet capacity on $[n]$ can be represented by the set $P^\mu = \{p_m^\mu\}_{m \in \mathcal{C}_{[n]}}$ of $n!$ probability distributions on $[n]$. We therefore propose to define the intuitive notion of *uncertainty associated with a discrete Choquet capacity* as a kind of *average* of the uncertainties contained in the probability distributions $\{p_m^\mu\}_{m \in \mathcal{C}_{[n]}}$. Hence, the more uniform on average the probability distributions p_m^μ, $m \in \mathcal{C}_{[n]}$, the higher the *uncertainty* contained in the discrete Choquet capacity μ. As no maximal chain should be privileged, the average uncertainty should be defined by means of a symmetric function over all the $n!$ maximal chains m of $H_S(p_m^\mu)$. It is worth mentioning that, in terms of maximal chains, the entropy H_M can be rewritten as

$$H_M(\mu) = \frac{1}{n!} \sum_{m \in \mathcal{C}_{[n]}} H_S(p_m^\mu) \tag{3}$$

or equivalently,

$$H_M(\mu) = \frac{1}{n!} \sum_{\pi \in \Pi_{[n]}} H_S(p_\pi^\mu). \tag{4}$$

This result immediately follows from the next proposition.

Proposition 2.1 *Let μ be any Choquet capacity on $[n]$ (normalized or not) and let f be any function. Then, we have*

$$\frac{1}{n!} \sum_{m \in \mathcal{C}_{[n]}} \sum_{j=1}^{n} f[\mu(m_j) - \mu(m_{j-1})] = \sum_{i=1}^{n} \sum_{S \subseteq [n] \setminus i} \gamma_s(n)\, f[\mu(S \cup i) - \mu(S)].$$

Proof. For all $i \in [n]$, for all $S \subseteq [n] \setminus i$, let us denote by $\mathcal{C}_{[n]}^{S,S\cup i}$ the subset of $\mathcal{C}_{[n]}$ composed of maximal chains whose subsets of cardinal s and $s+1$ are equal to S and $S \cup i$ respectively. It is easy to check that $|\mathcal{C}_{[n]}^{S,S\cup i}| = s!(n-s-1)!$.

It follows therefore that, for a given $i \in [n]$ and for a subset $S \subseteq [n] \setminus i$, when summing the term $\sum_{j=1}^{n} f[\mu(m_j) - \mu(m_{j-1})]$ over the set of maximal chains, the term $f[\mu(S \cup i) - \mu(S)]$ will appear $s!(n-s-1)!$ times. $\square$

If $\Pi_{[n]}$ is considered as a probability space, a straightforward probabilistic interpretation of H_M directly follows from Eq. (4): for any $\mu \in \mathcal{F}_{[n]}$, $H_M(\mu)$ is the mathematical expectation of $H_S(p_\pi^\mu)$ with respect to the uniform distribution on $\Pi_{[n]}$.

3 Axiomatization of the entropy H_M

Before stating the axioms that a measure of *uncertainty* or *entropy* of a discrete Choquet capacity should satisfy, we define some additional concepts that will be needed in the sequel.

3.1 Additional definitions

Let $\mu \in \mathcal{F}_{[n]}$ and let S and T be two disjoint subsets of $[n]$. The *Choquet capacity on S in the presence of T* [8] is denoted $\mu_{\cup T}^{S}$ and is defined by

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

Clearly, the Choquet capacity $\mu_{\cup T}^{S}$ is not normalized. The normalized version of the Choquet capacity $\mu_{\cup T}^{S}$ is denoted $\bar{\mu}_{\cup T}^{S}$ and is defined by

$$\bar{\mu}_{\cup T}^{S}(K) := \begin{cases} \dfrac{\mu_{\cup T}^{S}(K)}{\mu_{\cup T}^{S}(S)} & \text{for all } K \subseteq S \text{ if } \mu_{\cup T}^{S}(S) \neq 0, \\ 0 & \text{for all } K \subsetneq S \text{ if } \mu_{\cup T}^{S}(S) = 0, \\ 1 & \text{if } K = S \text{ and } \mu_{\cup T}^{S}(S) = 0. \end{cases}$$

Let μ be a Choquet capacity on $[n]$ and let $A_1, \ldots, A_k$ form a partition of $[n]$. The *reduced Choquet capacity with respect to $A_1, \ldots, A_k$* [8] is a Choquet capacity denoted $\mu^{[A_1]\ldots[A_k]}$ defined on a set of k elements noted as $[A_1]\ldots[A_k]$, where, for all $i \in [k]$, $[A_i]$ stands for an hypothetical element which is the union (or the representative) of the elements in A_i, that is,

$$\mu^{[A_1]\ldots[A_k]}\Big(\bigcup_{i\in S}[A_i]\Big) = \mu\Big(\bigcup_{i\in S} A_i\Big) \qquad \forall S \subseteq [k]. \tag{5}$$

For any Choquet capacity $\mu \in \mathcal{F}_{[n]}$ and any permutation $\pi \in \Pi_{[n]}$, we denote by $\pi\mu$ the Choquet capacity on $[n]$ defined by

$$\pi\mu(\pi(S)) = \mu(S) \qquad \forall S \subseteq [n],$$

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

3.2 Axioms

Consider a sequence of functions $H^{(n)} : \mathcal{F}_{[n]} \to \mathbb{R}$ $(n = 2, 3, \dots)$, where, for any $\mu \in \mathcal{F}_{[n]}$, $H^{(n)}(\mu)$ is a measure of entropy of μ. When $n = 1$ we simply set $H^{(1)}(\mu) = 0$.

Given $\mu \in \mathcal{F}_{[n]}$, it is easy to check that a permutation on $[n]$ leaves unchanged the set P^μ of probability distributions. Hence, naturally, the first axiom is:

- *Symmetry axiom (S)*: For any $n \geq 2$, any $\mu \in \mathcal{F}_{[n]}$, and any $\pi \in \Pi_{[n]}$, we have $H^{(n)}(\pi\mu) = H^{(n)}(\mu)$.

Recall from the previous section that, for a cardinality-based Choquet capacity $\mu \in \mathcal{F}_{[n]}$, there exits a probability distribution p^μ such that all elements of the set P^μ are equal to p^μ. This suggests measuring the uncertainty of a cardinality-based Choquet capacity μ as that of the probability distribution p^μ. The choice of the Shannon entropy being natural as a measure of uncertainty contained in a probability distribution, our second axiom is:

- *Shannon entropy axiom (SE)*: For any $n \geq 2$ and any $\mu \in \mathcal{F}_{[n]}$, if μ is cardinality-based, then $H^{(n)}(\mu) = H_S^{(n)}(p^\mu)$.

Note that the previous axiom implies that among all the cardinality-based capacities on $[n]$, μ^* is the one that has maximum uncertainty.

The Shannon entropy is known to satisfy the so-called *recursivity property* [1, 3, 4, 17, 18, 19], which basically states that the entropy of a discrete s-tochastic system can be calculated either directly or by dividing the system into subsystems[1]. Let p be a probability distribution on $[n]$ 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, we have

$$H_S^{(n)}(p) = H_S^{(2)}(p^{[A_1][A_2]}) + p(A_1)\, H_S^{(a_1)}(\bar{p}^{A_1}) + p(A_2)\, H_S^{(a_2)}(\bar{p}^{A_2}), \qquad (6)$$

where $p^{[A_1][A_2]}$ is a probability distribution on the set $\{[A_1], [A_2]\}$ defined by $p^{[A_1][A_2]}([A_i]) = p(A_i)$, $i = 1, 2$, 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,

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

As a generalization of the Shannon entropy, we require that our measure of uncertainty $H^{(n)}(\mu)$ satisfy a similar property that would thus reflect the

[1] Note that the recursivity property is important because it implies amongst other things the additivity of the Shannon entropy, i.e.,

$$H_S^{(n^2)}(p * q) = H_S^{(n)}(p) + H_S^{(n)}(q),$$

for all probability distributions p, q on $[n]$ where $p * q$ denotes the distribution

$$(p_1 q_1, \dots, p_1 q_n, \dots, p_n q_1, \dots, p_n q_n).$$

Figure 2: Decomposition resulting from a partition of $[n] = \{1, 2, 3, 4\}$ into the subsets $\{1\}$ and $\{2, 3, 4\}$.

possibility to decompose *in an additive way* the calculation of the uncertainty of the discrete Choquet capacity $\mu \in \mathcal{F}_{[n]}$. 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 Choquet capacity $\mu \in \mathcal{F}_{[n]}$ and a partition of $[n]$ into two subsets A_1 and A_2. The Choquet capacities that appear from such a decomposition are:

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

For instance, when $[n] = \{1, 2, 3, 4\}$, Figure 2 shows the decomposition resulting from considering $A_1 = \{1\}$ and $A_2 = \{2, 3, 4\}$.

Assume now that n is even so that the subsets A_1 and A_2 can be chosen to have the same cardinal $n/2$. Then, we have $\bar{\mu}^{A_1} = \bar{\mu}^{A_2}$, $\bar{\mu}^{A_1}_{\cup A_2} = \bar{\mu}^{A_2}_{\cup A_1}$, and the reduced Choquet capacity $\mu^{[A_1][A_2]}$ is cardinality-based. According to axiom SE and Eq. (6), the following functional equation holds

$$H^{(n)}(\mu) = H^{(2)}(\mu^{[A_1][A_2]}) + \mu^{A_1}(A_1)\, H^{(a_1)}(\bar{\mu}^{A_1}) + \mu^{A_2}_{\cup A_1}(A_2)\, H^{(a_2)}(\bar{\mu}^{A_2}_{\cup A_1}). \quad (7)$$

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 cardinal anymore? Indeed, for a general choice of n and of the subsets A_1 and A_2, the Choquet capacities $\bar{\mu}^{A_1}$ and $\bar{\mu}^{A_1}_{\cup A_2}$ as well as the capacities $\bar{\mu}^{A_2}$ and $\bar{\mu}^{A_2}_{\cup A_1}$ are not necessarily equal. Furthermore, the reduced Choquet capacity

$\mu^{[A_1][A_2]}$ is not cardinality-based anymore. As an extension of the previous case, we then require that the following functional equation, which is a generalization of (7) and is still in accordance with our intuitive additivity requirement, holds for any cardinality-based Choquet capacity $\mu \in \mathcal{F}_{[n]}$:

$$\begin{aligned} H^{(n)}(\mu) = H^{(2)}(\mu^{[A_1][A_2]}) & \\ + \alpha^1\, \mu^{A_1}(A_1)\, H^{(a_1)}(\bar{\mu}^{A_1}) + \alpha_2^1\, \mu^{A_1}_{\cup A_2}(A_1)\, H^{(a_1)}(\bar{\mu}^{A_1}_{\cup A_2}) & \\ + \alpha^2\, \mu^{A_2}(A_2)\, H^{(a_2)}(\bar{\mu}^{A_2}) + \alpha_1^2\, \mu^{A_2}_{\cup A_1}(A_2)\, H^{(a_2)}(\bar{\mu}^{A_2}_{\cup A_1}), & \end{aligned}$$

where α^1, α_2^1, α^2, and α_1^2 are real numbers left undetermined thus far.

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

- *Recursivity axiom (R)*: For any integers $n \geq k \geq 2$, there exists a family of real coefficients $\{\alpha^i_S(n,k) \mid i \in [k],\ S \subseteq [k] \setminus i\}$ such that, for any partition $\{A_1, \ldots, A_k\}$ of $[n]$ and any cardinality-based capacity $\mu \in \mathcal{F}_{[n]}$,

$$\begin{aligned} H^{(n)}(\mu) = H^{(k)}(\mu^{[A_1]\cdots[A_k]}) & \\ + \sum_{i=1}^{k} \sum_{S \subseteq [k] \setminus i} \alpha^i_S(n,k)\, \mu^{A_i}_{\cup_{j \in S} A_j}(A_i)\, H^{(a_i)}(\bar{\mu}^{A_i}_{\cup_{j \in S} A_j}). & \quad (8) \end{aligned}$$

3.3 Characterization

We can now state our main result.

Theorem 3.1 *The sequence $H^{(n)} : \mathcal{F}_{[n]} \to \mathbb{R}$ $(n \geq 2)$ fulfills axioms S, SE, and R, if and only if*

$$H^{(n)} = H^{(n)}_M \qquad (n \geq 2).$$

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 [13, 16]).

1. **Boundary property for additive measures.** For any additive Choquet capacity $\mu \in \mathcal{F}_{[n]}$, we have

$$H^{(n)}_M(\mu) = H^{(n)}_S(p)$$

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

2. **Expansibility.** Let $\mu \in \mathcal{F}_{[n]}$ 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

$$H_M^{(n)}(\mu) = H_M^{(n-1)}(\mu^{[n]\setminus i})$$

where $\mu^{[n]\setminus i}$ is the restriction of μ to $[n] \setminus i$.

3. **Decisivity.** We have

$$H_M^{(n)}(\mu) \geq 0 \qquad \forall \mu \in \mathcal{F}_{[n]}.$$

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

4. **Maximality.** For any $\mu \in \mathcal{F}_{[n]}$, we have

$$H_M^{(n)}(\mu) \leq H_S^{(n)}(\phi(\mu)) \leq \ln n,$$

where $\phi(\mu) = (\phi_1(\mu), \dots, \phi_n(\mu))$ is the probability distribution (see E-q. (2)) formed by the Shapley value of the elements of $[n]$ with respect to μ. Moreover,

   - $H_M^{(n)}(\mu) = H_S^{(n)}(\phi(\mu))$ if and only if μ is additive,
   - $H_S^{(n)}(\phi(\mu)) = \ln n$ if and only if $\phi(\mu) = (1/n, \dots, 1/n)$,
   - $H_M^{(n)}(\mu) = \ln n$ if and only if μ is the uniform Choquet capacity μ^* on $[n]$.

5. **Increasing monotonicity.** Let $\mu \in \mathcal{F}_{[n]} \setminus \{\mu^*\}$ and define $\mu_\lambda \in \mathcal{F}_{[n]}$ by

$$\mu_\lambda := \mu + \lambda(\mu^* - \mu) \qquad \forall \lambda \in [0, 1].$$

Then for any $0 \leq \lambda_1 < \lambda_2 \leq 1$ we have

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

We now state another very important property of $H_M^{(n)}$ which follows from Eq. (3) and the strict concavity of $H_S^{(n)}$.

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

$$H_M^{(n)}(\lambda\, \mu_1 + (1 - \lambda)\, \mu_2) > \lambda\, H_M^{(n)}(\mu_1) + (1 - \lambda)\, H_M^{(n)}(\mu_2).$$

An immediate consequence of the previous property is that maximizing $H_M^{(n)}$ over a convex subset of $\mathcal{F}_{[n]}$ 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 stated in 1957 by Jaynes [9] as follows: 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 states 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 $H_M^{(n)}$ suggests to extend such an inference principle to Choquet capacities. The main difference comes from the fact that the interpretation of the maximum entropy principle for Choquet capacities is less natural than for probability distributions because the notion of *uncertainty* associated with a discrete Choquet capacity (cf. subsection 2.3) is less intuitive than that associated with a probability distribution.

5 Entropy in the aggregation framework

Suppose that $[n]$ represents a set of n criteria in a multicriteria decision making problem and consider a Choquet capacity $\mu \in \mathcal{F}_{[n]}$. 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 [5, 15]). 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 commensurable, 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 is an appropriate extension of the classical weighted arithmetic mean for the aggregation of interacting criteria.

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

$$\mathcal{C}_\mu(x) := \sum_{i=1}^{n} x_{(i)} \left[\mu(\{(i), \dots, (n)\}) - \mu(\{(i+1), \dots, (n)\})\right],$$

where $(\cdot)$ is a permutation on $[n]$ such that $x_{(1)} \leq \cdots \leq x_{(n)}$. For more details, see e.g. [15] and the references therein.

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

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

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

It would be interesting to appraise the degree to which the argument $x \in [0,1]^n$ is really used when calculating $\mathcal{C}_\mu(x)$. In this final section we shall show

that the function $H_M^{(n)}(\mu)$ measures an average value over all $x \in [0,1]^n$ of this degree of utilization.

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

$$\mathcal{C}_\mu(x) = \sum_{i=1}^n x_i\,\mu(i) = \sum_{i=1}^n x_i\,p(i),$$

where p is the probability distribution defined by $p(i) := \mu(i)$ for all $i \in [n]$. In this case, the function $H_S^{(n)}(p)$ behaves like a *dispersion index* [13], which measures the evenness (uniformity) of the weights $p(i)$. For example,

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

In other terms, $H_S^{(n)}(p)$ measures the extent to which the information about the individual criteria is being used in the aggregation, or equivalently, the extent to which the argument x is being used in the calculation of the aggregated value $\mathcal{C}_\mu(x)$. In that sense, the more uniform the probability distribution p the more the argument x is being used in the aggregation process.

Consider now a general (non-additive) Choquet capacity $\mu \in \mathcal{F}_{[n]}$ and define the sets

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

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

$$\mathcal{C}_\mu(x) = \sum_{i=1}^n x_{\pi(i)}\,p_\pi^\mu(i).$$

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, $H_S^{(n)}(p_\pi^\mu)$ 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 of utilization[2] of argument x in the calculation of the aggregated value $\mathcal{C}_\mu(x)$.

Starting from Eq. (4) and using the fact that

$$\int_{x\in\mathcal{O}_\pi} \mathrm{d}x = \frac{1}{n!}$$

[2]Should $H_S^{(n)}(p_\pi^\mu)$ be close to $\ln n$, the distribution p_π^μ will be approximately uniform and all the partial evaluations x_i ($i \in [n]$) will be involved almost equally in the calculation of $\mathcal{C}_\mu(x)$, which will be close to the arithmetic mean of the x_i's. On the contrary, should $H_S^{(n)}(p_\pi^\mu)$ be close to zero, one $p_\pi^\mu(i)$ will be very close to one and $\mathcal{C}_\mu(x)$ will be very close to the corresponding partial evaluation.

the entropy $H_M^{(n)}(\mu)$ can be rewritten as

$$H_M^{(n)}(\mu) = \sum_{\pi \in \Pi_{[n]}} \int_{x \in \mathcal{O}_\pi} H_S^{(n)}(p_\pi^\mu)\,\mathrm{d}x.$$

We thus see that $H_M^{(n)}(\mu)$ measures the average value over all $x \in [0,1]^n$ of the degree of utilization of argument x in the calculation of $\mathcal{C}_\mu(x)$. From E-q. (3), it can also be interpreted as a measure of the average regularity of the increasingness of μ over all maximal chains $m \in \mathcal{C}_{[n]}$.

In probabilistic terms, it corresponds to the expectation over all $x \in [0,1]^n$, with uniform distribution, of the degree of utilization of argument x in the calculation of $\mathcal{C}_\mu(x)$, or equivalently, to the expectation over all maximal chains $m \in \mathcal{C}_{[n]}$, with uniform distribution, of the regularity of the increasingness of μ.

It should also be mentioned that the interpretation of $H_M^{(n)}(\mu)$ as an average degree of utilization of the argument is in full accordance with the properties listed in Section 4 (see [13, 16]). For example, the decisivity property, which states that $H_M^{(n)}(\mu) = 0$ if and only if μ is a binary-valued Choquet capacity, is quite relevant for our aggregation framework. Indeed, it can be shown that this latter condition holds if and only if

$$\mathcal{C}_\mu(x) \in \{x_1, \ldots, x_n\} \qquad (x \in [0,1]^n).$$

In other terms, $H_M^{(n)}(\mu)$ is minimum ($= 0$) if and only if only one partial evaluation is really used in the calculation of $\mathcal{C}_\mu(x)$.

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, linear constraints on the corresponding Choquet capacity μ. Then, among all the feasible (admissible) Choquet integrals, choosing the Choquet integral with respect to the maximum entropy Choquet capacity is equivalent to choosing the Choquet integral that will have the highest average degree of utilization of its argument during 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 the information contained in its argument.

6 Concluding remarks

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 as a quantification of the uncertainty contained in a Choquet capacity, as a measure of the average regularity of a Choquet capacity along all maximal chains, or as the average degree of utilization of an argument in the framework of aggregation by the Choquet integral.

We hope that this characterization will enable us to justify the use of this concept in different contexts, such as decision making and game theory.

Acknowledgements

The paper was partially prepared during the stay of Marc Roubens as visiting professor at University of Reunion Island, France, in May 2001. This opportunity is greatly acknowledged.

A Proof of Theorem 3.1

In order to prove Theorem 3.1, we shall go through three technical lemmas. Moreover, for any cardinality-based Choquet capacity $\mu \in \mathcal{F}_{[n]}$, we shall use the notation $\mu_t := \mu(T)$ for all $T \subseteq [n]$.

Lemma A.1 *If the sequence $H^{(n)} : \mathcal{F}_{[n]} \to \mathbb{R}$ $(n \geq 2)$ fulfills axiom SE then, for any integers $n \geq k \geq 2$, any partition $\{A_1, \dots, A_k\}$ of $[n]$, and any cardinality-based Choquet capacity $\mu \in \mathcal{F}_{[n]}$, we have, for all $i \in [k]$ and all $S \subseteq [k] \setminus i$,*

$$\begin{aligned}\mu^{A_i}_{\cup_{j\in S} A_j}(A_i)\, H^{(a_i)}(\bar{\mu}^{A_i}_{\cup_{j\in S} A_j}) \\ = \sum_{t=1}^{a_i} h[\mu_{\sum_{j\in S} a_j + t} - \mu_{\sum_{j\in S} a_j + t-1}] - h[\mu_{\sum_{j\in S} a_j + a_i} - \mu_{\sum_{j\in S} a_j}] \qquad (9)\end{aligned}$$

Proof. Since μ is cardinality-based, there exist $\mu_1, \dots, \mu_n \in [0,1]$ such that $\mu(T) = \mu_t$ for any $T \subseteq N$. Moreover, it is easy to check that the Choquet capacities $\bar{\mu}^{A_i}_{\cup_{j\in S} A_j}$ are cardinality-based. From axiom SE, for all $i \in [k]$, for all $S \subseteq [k] \setminus i$, we then have

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

which completes the proof. □

Lemma A.2 *If the sequence $H^{(n)} : \mathcal{F}_{[n]} \to \mathbb{R}$ $(n \geq 2)$ fulfills axioms S, SE, and R, then, for any integers $k \geq 2$ and $n = k(2^k - 1)$, there exists a family of constants $\alpha_0(n,k), \dots, \alpha_{k-1}(n,k)$ such that*

$$\alpha^i_S(n,k) = \alpha_s(n,k) \text{ for all } i \in [k], \text{ for all } S \subseteq [k] \setminus i.$$

Proof. Let $k \geq 2$ be an integer and set $n := k(2^k - 1)$. Let $\mu \in \mathcal{F}_{[n]}$ be cardinality-based and let $\{A_1, \ldots, A_k\}$ be a partition of $[n]$.

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

$$H^{(k)}(\mu^{[A_1]\ldots[A_k]}) = H^{(k)}(\mu^{[A_{\pi(1)}]\ldots[A_{\pi(k)}]}). \tag{10}$$

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

$$u(S) := \mu\Big(\bigcup_{i\in S} A_{\pi(i)}\Big), \qquad \forall S \in [k],$$

and

$$v(S) := \mu\Big(\bigcup_{i\in S} A_i\Big), \qquad \forall S \in [k],$$

we clearly see that $v(\pi(S)) = u(S)$ for all $S \subseteq [k]$ and hence $v = \pi u$. By the definition (5) and axiom S, we then have

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

¿From axiom R, we know that the functional equation (8) must hold for all partitions of $[n]$ into k subsets. Writing it for $\{A_1, \ldots, A_k\}$ and for $\{A_{\pi(1)}, \ldots, A_{\pi(k)}\}$ and making use of (10), we have

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

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

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

that is, by Lemma A.1,

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

which must hold for all cardinality-based Choquet capacities $\mu \in \mathcal{F}_{[n]}$ and for all partitions $\{A_1, \ldots, A_k\}$.

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

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

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 $\sum_{i\in S} b_i$ $(S\subseteq[k])$ are all 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 $\mu\in\mathcal{F}_{[n]}$ by

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

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

Then, we can show that

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

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

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

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 $\sum_{i\in S} a_i$ $(S\subseteq[k])$ are all different.

The proof is now complete. □

Lemma A.3 *If the sequence $H^{(n)}:\mathcal{F}_{[n]}\to\mathbb{R}$ $(n\ge 2)$ fulfills axioms S, SE, and R, then, for any integers $k\ge 2$ and $n = k(2^k-1)$, we have*

$$\alpha_s(n,k) = \gamma_s(k) \text{ for all } s = 0,\dots,k-1.$$

Proof. Let $k\ge 2$ be an integer and set $n := k(2^k-1)$. Let $\mu\in\mathcal{F}_{[n]}$ be cardinality-based and let $\{A_1,\dots,A_k\}$ be a partition of $[n]$ such that $a_1 = \cdots = a_k := a$, with $a = 2^k - 1 > 1$.

From axiom SE, we have that

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

which can be rewritten as

$$H^{(n)}(\mu) = \sum_{i=1}^{k} \sum_{t=1}^{a} h[\mu_{(i-1)a+t} - \mu_{(i-1)a+t-1}].$$

Using the fact that, for any $i \in [k]$, the coefficients $\{\gamma_s(k) \mid S \subseteq [k] \setminus i\}$, form a probability distribution, we obtain

$$H^{(n)}(\mu) = \sum_{i=1}^{k} \sum_{S \subseteq [k] \setminus i} \gamma_s(k) \sum_{t=1}^{a} h[\mu_{sa+t} - \mu_{sa+t-1}]. \tag{11}$$

The subsets A_i having the same cardinal, it is easy to check that $\mu^{[A_1]\ldots[A_k]}$ is a cardinality-based Choquet capacity on $\{[A_1], \ldots, [A_k]\}$. According to axiom SE, $H^{(k)}(\mu^{[A_1]\ldots[A_k]})$ can be written as

$$H^{(k)}(\mu^{[A_1]\ldots[A_k]}) = \sum_{i=1}^{k} h[p^{\mu^{[A_1]\ldots[A_k]}}(i)] = \sum_{i=1}^{k} h[\mu_{ia} - \mu_{(i-1)a}],$$

which, using again the fact that, for any $i \in [k]$, the coefficients $\{\gamma_s(k) \mid S \subseteq [k] \setminus i\}$, form a probability distribution, is equivalent to

$$H^{(k)}(\mu^{[A_1]\ldots[A_k]}) = \sum_{i=1}^{k} \sum_{S \subseteq [k] \setminus i} \gamma_s(k)\, h[\mu_{(s+1)a} - \mu_{sa}]. \tag{12}$$

Using Lemma A.1 in the case where the numbers a_i are all equal, for all $i \in [k]$, for all $S \subseteq [k] \setminus i$, we obtain

$$\mu^{A_i}_{\cup_{j \in S} A_j}(A_i)\, H^{(a)}(\bar{\mu}^{A_i}_{\cup_{j \in S} A_j}) = \sum_{t=1}^{a} h[\mu_{sa+t} - \mu_{sa+t-1}] - h[\mu_{(s+1)a} - \mu_{sa}] \tag{13}$$

Substituting in the functional equation (8) of axiom R the terms

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

by their expressions given in Eq. (11), (12), and (13), we obtain, using Lemma A.2,

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

that is,

$$\sum_{s=0}^{k-1}\left(\begin{array}{c}k-1\\s\end{array}\right)(\gamma_s(k)-\alpha_s(n,k))$$
$$\times\Big[\sum_{t=1}^{a}h[\mu_{sa+t}-\mu_{sa+t-1}]-h[\mu_{(s+1)a}-\mu_{sa}]\Big]=0 \quad (14)$$

which must hold for all cardinality-based Choquet capacities $\mu\in\mathcal{F}_{[n]}$. By choosing μ in an appropriate way, we will now show that

$$\alpha_s(n,k)=\gamma_s(k) \text{ for all } s=0,\ldots,k-1,$$

which will complete the proof of the lemma.

Let us fix $s^*\in\{0,\ldots,k-1\}$, and define $\mu\in\mathcal{F}_{[n]}$ by

$$\mu(T):=\begin{cases}0 & \text{if } t\leq s^*a\\[2ex] \dfrac{t-s^*a}{a} & \text{if } s^*a\leq t\leq(s^*+1)a\\[2ex] 1 & \text{if } t\geq(s^*+1)a\end{cases}$$

Using axiom SE, we can easily show that Eq. (14) becomes

$$\left(\begin{array}{c}k-1\\s^*\end{array}\right)(\gamma_{s^*}(k)-\alpha_{s^*}(n,k))\,\ln a=0,$$

where $a>1$, which is sufficient. □

We now prove Theorem 3.1.

Proof. (Necessity) Let $k\geq 2$ be an integer and let $\nu\in\mathcal{F}_{[k]}$. Let us show first that there exist an integer $n>k$, multiple of k, a partition $\{A_1,\ldots,A_k\}$ of $[n]$, and a cardinality-based Choquet capacity $\mu\in\mathcal{F}_{[n]}$ such that $\nu=\mu^{[A_1]\cdots[A_k]}$, that is,

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

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

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

Consider now a set function $v:\mathcal{P}([k])\to\mathbb{R}$ defined by

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

It is easy to check that v is a Choquet capacity on $[k]$.

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

$$\begin{aligned}
&\sum_{i=1}^{k} \sum_{S\subseteq[k]\setminus i} \gamma_s(k) \sum_{t=1}^{a_i} h[\mu_{t+\sum_{j\in S} a_j} - \mu_{t-1+\sum_{j\in S} a_j}] \\
&\quad = \sum_{i=1}^{k} \sum_{S\subseteq[k]\setminus i} \gamma_s(k)\,[v(S\cup i) - v(S)] \\
&\quad = v([k]) \qquad \text{(cf. Eq. (2))} \\
&\quad = \sum_{t=1}^{n} h[\mu_t - \mu_{t-1}] \\
&\quad = H_S^{(n)}(p^\mu).
\end{aligned}$$

Finally, by axiom SE, we obtain

$$\sum_{i=1}^{k} \sum_{S\subseteq[k]\setminus i} \gamma_s(k) \sum_{t=1}^{a_i} h[\mu_{t+\sum_{j\in S} a_j} - \mu_{t-1+\sum_{j\in S} a_j}] = H^{(n)}(\mu). \tag{15}$$

Therefore, combining (9) and (15), we have

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

Finally, using Lemmas A.2 and A.3, the functional equation stated in axiom R becomes

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

that is,

$$H^{(k)}(\nu) = H_M^{(k)}(\nu).$$

(Sufficiency) We know that the sequence $(H_M^{(n)})_{n\geq 2}$ fulfills axioms S and SE. Let us show that it fulfills axiom R.

Let $n \geq k \geq 2$ be integers. Let $\mu \in \mathcal{F}_{[n]}$ be cardinality-based and let $\{A_1, \dots, A_k\}$ be a partition of $[n]$. By Lemma A.1, we have

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

and hence (cf. (15))

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

We have therefore shown that the sequence $(H^{(n)}_M)_{n\geq 2}$ fulfills also axiom R which completes the proof. □

References

[1] J. Aczél and Z. Daróczy. *On measures of information and their characterizations.* Academic Press, New York–San Francisco–London, 1975.

[2] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953.

[3] B. Ebanks, P. Sahoo, and W. Sander. *Characterizations of information measures.* World Scientific, Singapore, 1997.

[4] D.K. Fadeev. Zum begriff der entropie einer endlichen wahrscheinlichkeitschemes. *Arbeit zur Informationstheorie, Deutscher Verlag der Wissenschaften*, 1, 1957.

[5] M. Grabisch. Alternative representations of discrete fuzzy measures for decison making. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 5(5):587–607, 1997.

[6] M. Grabisch, J.-L. Marichal, and M. Roubens. Equivalent representations of set functions. *Mathematics of Operations Research*, 25(2):157–178, 2000.

[7] M. Grabisch and J-M. Nicolas. Classification by fuzzy integrals: Performance and test. *Fuzzy Sets and Systems*, 65:255–271, 1994.

[8] M. Grabisch and M. Roubens. An axiomatic approach to the concept of interaction among players in cooperative games. *Int. J. Game Theory*, 124:547–565, 1999.

[9] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630, 1957.

[10] E. T. Jaynes. *Probability Theory: The Logic of Science*. Fragmentary Edition of March 1996, 1996.

[11] A.I. Khinchin. *Mathematical foundations of information theory*. Dover, 1957.

[12] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Stat.*, 22:79–86, 1951.

[13] J.-L. Marichal. Entropy of discrete choquet capacities. *European Journal of Operational Research*. to appear.

[14] J.-L. Marichal. *Aggregation operators for multicriteria decison aid*. PhD thesis, University of Liège, Liège, Belgium, 1998.

[15] J.-L. Marichal. An axiomatic approach of the discrete choquet integral as a tool to aggregate interacting criteria. *IEEE Transactions on Fuzzy Systems*, 8(6):800–807, 2000.

[16] J.-L. Marichal and M. Roubens. Entropy of discrete fuzzy measures. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 8(6):625–640, 2000.

[17] D. Morales, L. Pardo, and I. Vajda. Uncertainty of discrete stochastic systems: general theory and statistical theory. *IEEE Trans. on System, Man and Cybernetics*, 26(11):1–17, 1996.

[18] A. Rényi. On the foundations of information theory. *Review of the International Statistical Institute*, 33(1):1–14, 1965.

[19] C. E. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423,623–656, 1948.

[20] L.S. Shapley. A value for n-person game. In H.W. Kuhn and A.W. Tucker, editors, *Contributions to the Theory of Games*, volume 2, pages 307–317. Princeton University Press, 1953.

[21] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, Tokyo, Japan, 1974.