Entropy of bi-capacities ⋆

Ivan Kojadinovic [a], Jean-Luc Marichal [b]

[a] *LINA CNRS FRE 2729, Site école polytechnique de l'Université de Nantes, Rue Christian Pauc, 44306 Nantes, France*

[b] *Applied Mathematics Unit, University of Luxembourg, 162A, avenue de la Faïencerie, L-1511 Luxembourg, Grand Duchy of Luxembourg*

Abstract

In the context of multicriteria decision making whose aggregation process is based on the Choquet integral, bi-capacities can be regarded as a natural extension of capacities when the underlying evaluation scale is bipolar. The notion of entropy, recently generalized to capacities to measure their uniformity, is now extended to bi-capacities. We show that the resulting entropy measure has a very natural interpretation in terms of the Choquet integral and satisfies many natural properties that one would expect from an entropy measure.

Key words: Multicriteria decision making, bi-capacity, Choquet integral, entropy.

1 Introduction

The well-known Shannon entropy [2–4] is a fundamental concept in probability theory and related fields. In a general non-probabilistic setting, it is merely a measure of the uniformity (evenness) of a discrete probability distribution. In a probabilistic context, it can be naturally interpreted as a measure of unpredictability or information [5].

By relaxing the additivity property of probability measures, requiring only that they be monotone, one obtains *capacities* [6], also known as *fuzzy measures* [7], for which an extension of the Shannon entropy was proposed by

⋆ This paper is a revised and extended version with proofs of the conference paper [1].

Email addresses: ivan.kojadinovic[at]polytech.univ-nantes.fr (Ivan Kojadinovic), jean-luc.marichal[at]uni.lu (Jean-Luc Marichal).

Marichal [8–10]. The resulting entropy measure was recently characterized as a measure of uniformity of a capacity in [11]. In the context of multicriteria decision making based on Choquet integrals, it was further given a natural interpretation which gave rise to a maximum entropy like principle in that framework [11,12].

The concept of capacity can be further generalized. In the context of multicriteria decision making, *bi-capacities* have been recently introduced by Grabisch and Labreuche [13–15] to model in a flexible way the preferences of a decision maker when the underlying evaluation scale is *bipolar* [16]. Since a bi-capacity can be regarded as a generalization of a capacity, the following natural question arises : how could one appraise the 'uniformity' or 'uncertainty' associated with a bi-capacity in the spirit of the Shannon entropy?

In this paper, we propose an extension of the Shannon entropy to bi-capacities that has a natural interpretation in the framework of multicriteria decision making whose aggregation process is based on Choquet integrals. To introduce that extended entropy measure, we shall consider a set $N := \{1, \dots, n\}$ of *criteria* and a set $\mathcal{A}$ of *alternatives* described according to these criteria, i.e., real-valued functions on N. Then, given an alternative $x \in \mathcal{A}$, for any $i \in N$, $x_i := x(i)$ is regarded as the *evaluation* of x with respect to criterion i. The evaluations are further considered to be *commensurate* and to lie either on a unipolar or on a *bipolar* scale. Compared to a unipolar scale, a bipolar scale is characterized by the additional presence of a neutral value (usually 0) such that values above this neutral reference point are considered to be good by the decision maker, and values below it are considered to be bad [16]. As in [13,14], for simplicity reasons, we shall assume that the scale used for all evaluations is $[0, 1]$ if the scale is unipolar, and $[-1, 1]$ with 0 as neutral value, if the scale is bipolar.

This paper is organized as follows. The second section is devoted to a presentation of the notions of capacity, bi-capacity, and Choquet integral in the framework of aggregation. The third section presents the existing extension of the Shannon entropy to capacities and recalls its main properties. In the last section, we propose a generalization of the Shannon entropy to bi-capacities. We also give an interpretation of it in the context of multicriteria decision making based on Choquet integrals and we study its main properties.

In order to avoid a heavy notation, we will omit braces for singletons and pairs, e.g., by writing $v(i)$, $N \setminus ij$ instead of $v(\{i\})$, $N \setminus \{ij\}$. Furthermore, cardinalities of subsets $S, T, \dots$, will be denoted by the corresponding lower case letters $s, t, \dots$

2 Capacities, bi-capacities, and Choquet integral

2.1 Capacities and bi-capacities

In the context of aggregation, *capacities* [6], also called *fuzzy measures* [7], can be regarded as generalizations of weighting vectors involved in the calculation of weighted arithmetic means [17].

Let $\mathcal{P}(N)$ denote the power set of N.

Definition 1 *A function $\mu : \mathcal{P}(N) \to [0,1]$ is a capacity on N if it satisfies :*

(i) $\mu(\emptyset) = 0$, $\mu(N) = 1$,
(ii) for any $S, T \subseteq N$, $S \subseteq T \Rightarrow \mu(S) \le \mu(T)$.

A capacity μ on N is said to be *additive* if $\mu(S \cup T) = \mu(S) + \mu(T)$ for all disjoint subsets $S, T \subseteq N$. It is said to be *cardinality-based* if, for any $T \subseteq N$, $\mu(T)$ depends only on its cardinality t. There is only one capacity on N, denote it by μ^*, that is both additive and cardinality-based. It is called the *uniform* capacity on N and it is defined by

$$\mu^*(T) = \frac{t}{n}, \qquad \forall T \subseteq N.$$

The *dual* (or *conjugate*) of a capacity μ on N is a capacity $\bar{\mu}$ on N defined by $\bar{\mu}(A) = \mu(N) - \mu(N \setminus A)$, for all $A \subseteq N$. An element $k \in N$ is said to be *null* for a capacity μ on N if $\mu(T \cup k) = \mu(T)$, for all $T \subseteq N \setminus k$.

As mentioned in the introduction, the concept of capacity can be further generalized. In the context of aggregation, *bi-capacities* arise as a natural extension of capacities when the underlying evaluation scale is bipolar [13].

Let $\mathcal{Q}(N) := \{(A,B) \in \mathcal{P}(N) \times \mathcal{P}(N) \mid A \cap B = \emptyset\}$.

Definition 2 *A function $v : \mathcal{Q}(N) \to \mathbb{R}$ is a bi-capacity on N if it satisfies :*

(i) $v(\emptyset, \emptyset) = 0$, $v(N, \emptyset) = 1$, $v(\emptyset, N) = -1$,
(ii) $A \subseteq B$ *implies* $v(A, \cdot) \le v(B, \cdot)$ *and* $v(\cdot, A) \ge v(\cdot, B)$.

Furthermore, a bi-capacity v is said to be :

- of the *Cumulative Prospect Theory (CPT) type* [13,14,18] if there exist two capacities μ_1, μ_2 such that

$$v(A,B) = \mu_1(A) - \mu_2(B), \qquad \forall (A,B) \in \mathcal{Q}(N).$$

When $\mu_2 = \mu_1$ (resp. $\mu_2 = \bar{\mu}_1$), the bi-capacity is further said to be *symmetric* (resp. *asymmetric*).

- *additive* if it is of the CPT type with μ_1, μ_2 additive, i.e., for any $(A, B) \in \mathcal{Q}(N)$

$$v(A, B) = \sum_{i \in A} \mu_1(i) - \sum_{i \in B} \mu_2(i).$$

Note that an additive bi-capacity with $\mu_1 = \mu_2$ is both symmetric and asymmetric since $\bar{\mu}_1 = \mu_1$.

As we continue, to indicate that a CPT type bi-capacity v is constructed from two capacities μ_1, μ_2, we shall denote it by v_{μ_1,μ_2}.

We shall further say that a bi-capacity v on N is *difference cardinality-based* if, for any $(A, B) \in \mathcal{Q}(N)$, $v(A, B)$ depends on the difference of cardinalities $a - b$.

It is easy to verify that there is only one bi-capacity on N that is both additive and difference cardinality-based. We shall call it the *uniform* bi-capacity and denote it by v^*. It is defined by

$$v^*(A, B) = \frac{a - b}{n}, \qquad \forall (A, B) \in \mathcal{Q}(N).$$

Finally, we say that an element $k \in N$ is *null* for a bi-capacity v on N if

$$v(A \cup k, B) = v(A, B) = v(A, B \cup k), \qquad \forall (A, B) \in \mathcal{Q}(N \setminus k).$$

2.2 Capacities and maximal chains

Before recalling the definition of the Choquet integral w.r.t. a capacity, we introduce some definitions [11].

A *maximal chain* m of the lattice $(\mathcal{P}(N), \subseteq)$ is an ordered collection of $n + 1$ nested distinct subsets denoted

$$m = (\varnothing \subsetneq \{i_1\} \subsetneq \{i_1, i_2\} \subsetneq \cdots \subsetneq \{i_1, \ldots, i_n\} = N).$$

Denote by $\mathcal{M}_N$ the set of maximal chains of $(\mathcal{P}(N), \subseteq)$ and by Π_N the set of permutations on N. We can readily see that the sets $\mathcal{M}_N$ and Π_N are equipollent. Indeed, to each permutation $\sigma \in \Pi_N$ corresponds a unique maximal chain $m^\sigma \in \mathcal{M}_N$ defined by

$$m^\sigma = (\varnothing \subsetneq \{\sigma(n)\} \subsetneq \{\sigma(n-1), \sigma(n)\} \subsetneq \cdots \subsetneq \{\sigma(1), \ldots, \sigma(n)\} = N).$$

Now, given a capacity μ on N, with each permutation $\sigma \in \Pi_N$ can be associated a discrete probability distribution p^μ_σ on N defined by

$$p^\mu_\sigma(i) := \mu\Big(\{\sigma(i), \ldots, \sigma(n)\}\Big) - \mu\Big(\{\sigma(i+1), \ldots, \sigma(n)\}\Big), \qquad \forall i \in N.$$

Equivalently, with the maximal chain $m^\sigma \in \mathcal{M}_N$ is associated the probability distribution $p^\mu_{m^\sigma} := p^\mu_\sigma$.

2.3 The Choquet integral w.r.t. a capacity

When evaluations are considered to lie on a unipolar scale, the importance of the subsets of (interacting) criteria can be modelled by a capacity. A suitable aggregation operator that generalizes the weighted arithmetic mean is then the Choquet integral [19,20].

Definition 3 *The Choquet integral of a function $x : N \to \mathbb{R}^+$, represented by the vector $(x_1, \ldots, x_n)$, w.r.t. the capacity μ on N is defined by*

$$C_\mu(x) := \sum_{i=1}^{n} p^\mu_\sigma(i)\, x_{\sigma(i)}, \tag{1}$$

where σ is a permutation on N such that $x_{\sigma(1)} \leq \cdots \leq x_{\sigma(n)}$.

An intuitive presentation of the Choquet integral can be found in [21]. Note also that an axiomatic characterization of the Choquet integral as an aggregation operator was proposed by Marichal in [17].

As one can see from the previous definition, the Choquet integral is defined for positive integrands only. For real integrand functions, two extensions of the Choquet integral exist [22,23,16]. For any $x : N \to \mathbb{R}$, let us denote by x^+ and x^- the positive and the negative parts of x, respectively defined by

$$x^+ := \max(x, 0) \quad \text{and} \quad x^- := \max(-x, 0).$$

The first generalization of the Choquet integral is known as the *symmetric Choquet integral*, also called the *Šipoš integral* [24]. It is defined, for any capacity μ on N, by

$$\check{S}_\mu(x) := C_\mu(x^+) - C_\mu(x^-), \qquad \forall x \in \mathbb{R}^n.$$

The second generalization, which coincides with the usual definition of the Choquet integral for real integrands, is known as the *asymmetric Choquet integral* and is defined by

$$C_\mu(x) := C_\mu(x^+) - C_{\bar{\mu}}(x^-), \qquad \forall x \in \mathbb{R}^n.$$

The following result follows from [16, Eqs. (11) and (12)].

Proposition 4 *For any capacity μ on N and any $x \in \mathbb{R}^n$, we have*

$$C_\mu(x) = \sum_{i=1}^{n} p_\sigma^\mu(i)\, x_{\sigma(i)}, \tag{2}$$

and

$$\check{S}_\mu(x) = \sum_{i=1}^{r} p_\sigma^{\bar{\mu}}(i)\, x_{\sigma(i)} + \sum_{i=r+1}^{n} p_\sigma^\mu(i)\, x_{\sigma(i)}, \tag{3}$$

where σ is a permutation on N such that $x_{\sigma(1)} \le \cdots \le x_{\sigma(r)} \le 0 \le x_{\sigma(r+1)} \le \cdots \le x_{\sigma(n)}$.

As discussed in [16], in the context of aggregation, the symmetric Choquet integral can be regarded as a first mean to deal with evaluations defined on a bipolar scale, whereas using the asymmetric Choquet integral in such a framework amounts to ignoring the neutral level of the scale.

2.4 The Choquet integral w.r.t. a bi-capacity

In a bipolar context, when bi-capacities are used to model the importance of the subsets of criteria, Grabisch and Labreuche [14] proposed the following natural extension of the Choquet integral, which, as we shall see, encompasses the symmetric and asymmetric Choquet integrals.

Definition 5 *The Choquet integral of a function $x : N \to \mathbb{R}$, represented by the vector $(x_1, \dots, x_n)$, w.r.t. the bi-capacity v on N is defined by*

$$C_v(x) := C_{\nu^v_{N^+}}(|x|),$$

where $\nu^v_{N^+}$ is a game on N (i.e., a set function on N vanishing at the empty set) defined by

$$\nu^v_{N^+}(C) := v(C \cap N^+, C \cap N^-), \qquad \forall C \subseteq N,$$

and $N^+ := \{i \in N | x_i \ge 0\}$, $N^- := N \setminus N^+$.

As shown in [14], an equivalent expression of $C_v(x)$ is :

$$C_v(x) = \sum_{i \in N} q^v_{\sigma,N^+}(i)\, |x_{\sigma(i)}|, \tag{4}$$

where σ is a permutation on N so that $|x_{\sigma(1)}| \le \cdots \le |x_{\sigma(n)}|$ and

$$q^v_{\sigma,N^+}(i) := \nu^v_{N^+}\Big(\{\sigma(i), \dots, \sigma(n)\}\Big) - \nu^v_{N^+}\Big(\{\sigma(i+1), \dots, \sigma(n)\}\Big), \qquad \forall i \in N.$$

From the monotonicity conditions of bi-capacities, it is easy to see that, for any $i \in N$, $x_{\sigma(i)}$ and $q^v_{\sigma,N^+}(i)$ have the same sign. It follows that Eq. (4) can be equivalently rewritten as

$$C_v(x) = \sum_{i \in N} |q^v_{\sigma,N^+}(i)|\, x_{\sigma(i)} \tag{5}$$

or

$$C_v(x) = \sum_{\substack{i \in N \\ \sigma(i) \in N^+}} q^v_{\sigma,N^+}(i)\, x_{\sigma(i)} - \sum_{\substack{i \in N \\ \sigma(i) \in N^-}} q^v_{\sigma,N^+}(i)\, x_{\sigma(i)}.$$

The Choquet integral w.r.t. a bi-capacity generalizes the symmetric and asymmetric Choquet integral. Indeed, as shown in [14, Proposition 1], for any symmetric bi-capacity $v_{\mu,\mu}$ on N, we have

$$C_{v_{\mu,\mu}}(x) = \check{S}_\mu(x), \qquad \forall x \in \mathbb{R}^n,$$

and, for any asymmetric bi-capacity $v_{\mu,\bar{\mu}}$ on N, we have

$$C_{v_{\mu,\bar{\mu}}}(x) = C_\mu(x), \qquad \forall x \in \mathbb{R}^n. \tag{6}$$

3 Entropy of a capacity

3.1 The concept of probabilistic entropy

The fundamental concept of *entropy of a probability distribution* was initially proposed by Shannon [25,4]. The Shannon entropy of a probability distribution p defined on the set $N := \{1, \ldots, n\}$ is given by

$$H_S(p) := \sum_{i \in N} h[p(i)],$$

where

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

The quantity $H_S(p)$ is always non negative and it is zero if and only if p is a Dirac mass (*decisivity* property). As a function of p, H_S is strictly concave. Furthermore, it reaches its maximum value (which is $\ln n$) if and only if p is uniform (*maximality* property).

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.

Note that many generalizations of the Shannon entropy were proposed in the literature. For an overview, see e.g. [26].

3.2 Extension of the Shannon entropy to capacities

The Shannon entropy has been extended to capacities by Marichal [8,9] (see also [10]), who proposed the following definition.

Definition 6 *The extension of the Shannon entropy to a capacity μ on N 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

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

Regarded as a uniformity measure, H_M has been recently axiomatized by means of three axioms [11] : the symmetry property, a boundary condition for which H_M reduces to the Shannon entropy, and a generalized version of the well-known recursivity property [2,3,27].

A fundamental property of H_M is that it can be rewritten in terms of the maximal chains of the $(\mathcal{P}(N), \subseteq)$ [11]. For any capacity μ on N, we have

$$H_M(\mu) = \frac{1}{n!} \sum_{m\in\mathcal{M}_N} H_S(p^\mu_m), \tag{7}$$

or equivalently,

$$H_M(\mu) = \frac{1}{n!} \sum_{\sigma\in\Pi_N} H_S(p^\mu_\sigma). \tag{8}$$

The quantity $H_M(\mu)$ can therefore simply be seen as an average over $\mathcal{M}_N$ of the uniformity values of the probability distributions p^μ_m calculated by means of the Shannon entropy. In other words, $H_M(\mu)$ can be interpreted as a measure of the average regularity of the monotonicity of μ over all maximal chains $m \in \mathcal{M}_N$.

As shown in [11], in the context of aggregation by a Choquet integral w.r.t. a capacity μ on N, $H_M(\mu)$ can also be interpreted as a measure of 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 the aggregated value $C_\mu(x)$.

To see this, define the sets

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

which cover the hypercube $[0,1]^n$. For a fixed $x \in [0,1]^n$, there exists $\sigma \in \Pi_N$ such that $x \in \mathcal{O}_\sigma$ and hence

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

The permutation σ corresponds to the maximal chain m^σ, along which the Choquet integral boils down to a weighted arithmetic mean whose weights are defined by the probability distribution p_σ^μ. In that case, $H_S(p_\sigma^\mu)$ measures the degree to which the arguments $x_1, \ldots, x_n$ contribute[1] to the calculation of the aggregated value $C_\mu(x)$.

Starting from Eq. (8) and using the fact that $\int_{x \in \mathcal{O}_\sigma} \mathrm{d}x = 1/n!$ the entropy $H_M(\mu)$ can be rewritten as

$$H_M(\mu) = \sum_{\sigma \in \Pi_N} \int_{x \in \mathcal{O}_\sigma} H_S(p_\sigma^\mu)\, \mathrm{d}x = \int_{[0,1]^n} H_S(p_{\sigma_x}^\mu)\, \mathrm{d}x,$$

where σ_x is a permutation of Π_N such that $x \in \mathcal{O}_{\sigma_x}$. 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 $x_1, \ldots, x_n$ contribute to the calculation of $C_\mu(x)$.

To stress on the fact that H_M is an average of Shannon entropies, we shall equivalently denote it by $\overline{H}_S$ as we go on.

It has also been shown that $\overline{H}_S$ satisfies many properties that one would intuitively require from an entropy measure [11,9]. The most important ones are :

(1) **Boundary property for additive measures.** For any additive capacity μ on N, we have

$$\overline{H}_S(\mu) = H_S(p),$$

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

(2) **Boundary property for cardinality-based measures.** For any cardinality-based capacity μ on N, we have

$$\overline{H}_S(\mu) = H_S(p^\mu),$$

[1] Should $H_S(p_\sigma^\mu)$ be close to $\ln n$, the distribution p_σ^μ will be approximately uniform and all the partial evaluations $x_1, \ldots, x_n$ 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 $H_S(p_\sigma^\mu)$ be close to zero, only one $p_\sigma^\mu(i)$ will be very close to one and $C_\mu(x)$ will be very close to the corresponding partial evaluation.

where p^μ is the probability distribution on N defined by $p^\mu := p^\mu_\sigma$ for all $\sigma \in \Pi_N$.

(3) **Symmetry.** For any capacity μ on N and any permutation $\pi \in \Pi_N$, we have
$$\overline{H}_S(\mu \circ \pi) = \overline{H}_S(\mu).$$

(4) **Expansibility.** Let μ be a capacity on N. If $k \in N$ is a null element for μ then
$$\overline{H}_S(\mu) = \overline{H}_S(\mu_{-k}),$$
where μ_{-k} denotes the restriction of μ to $N \setminus k$.

(5) **Decisivity.** For any capacity μ on N,
$$\overline{H}_S(\mu) \geq 0,$$
with equality if and only if μ is a binary-valued capacity, that is, such that $\mu(T) \in \{0,1\}$ for all $T \subseteq N$.

(6) **Maximality.** For any capacity μ on N, we have
$$\overline{H}_S(\mu) \leq \ln n,$$
with equality if and only if μ is the uniform capacity μ^* on N.

(7) **Increasing monotonicity toward $\boldsymbol{\mu^*}$.** Let μ be a capacity on N such that $\mu \neq \mu^*$ and, for any $\lambda \in [0,1]$, define the capacity μ_λ on N as $\mu_\lambda := \mu + \lambda(\mu^* - \mu)$. Then for any $0 \leq \lambda_1 < \lambda_2 \leq 1$ we have
$$\overline{H}_S(\mu_{\lambda_1}) < \overline{H}_S(\mu_{\lambda_2}).$$

(8) **Strict concavity.** For any two capacities μ_1, μ_2 on N and any $\lambda \in \,]0,1[$, we have
$$\overline{H}_S(\lambda\, \mu_1 + (1-\lambda)\, \mu_2) > \lambda\, \overline{H}_S(\mu_1) + (1-\lambda)\, \overline{H}_S(\mu_2).$$

4 Entropy of a bi-capacity

As discussed earlier, bi-capacities can be regarded as a natural generalization of capacities when the underlying evaluation scale is bipolar. It seems then natural to investigate the possibility of extending the Shannon entropy to bi-capacities, as was already done for capacities.

In the context of aggregation by the Choquet integral w.r.t. a capacity μ, we have seen in the previous section that one of the fundamental characteristics of $\overline{H}_S(\mu)$ is that it can be interpreted as a measure of the average degree to which the arguments of a profile contribute to the calculation of the aggregated value.

Let us now focus on the Choquet integral w.r.t. a bi-capacity v on N and see how we could measure the same behavior. To do so, consider an alternative $x : N \to [-1, 1]$, denote by $N^+ \subseteq N$ the subset of criteria for which $x \geq 0$, and let $N^- := N \setminus N^+$. Then, from Eq. (5), we see that the Choquet integral of x w.r.t. v is simply a weighted sum of $x_{\sigma(1)}, \ldots, x_{\sigma(n)}$, where each $x_{\sigma(i)}$ is weighted by $|q^v_{\sigma,N^+}(i)|$. From the boundary conditions that define a bi-capacity, it is easy to verify that these weights sum up to one if $N^+ \in \{\emptyset, N\}$. However, this is not true in general if $N^+ \in \mathcal{P}(N) \setminus \{\emptyset, N\}$.

For any bi-capacity v on N and any $N^+ \subseteq N$, let p^v_{σ,N^+} be the probability distribution on N defined by

$$p^v_{\sigma,N^+}(i) := \frac{|q^v_{\sigma,N^+}(i)|}{\sum_{j \in N} |q^v_{\sigma,N^+}(j)|}, \qquad \forall i \in N. \tag{9}$$

The Choquet integral of x w.r.t. v can then be rewritten as

$$C_v(x) = \Big(\sum_{i \in N} p^v_{\sigma,N^+}(i)\, x_{\sigma(i)} \Big) \times \Big(\sum_{i \in N} |q^v_{\sigma,N^+}(i)| \Big). \tag{10}$$

As we could have expected, the degree of contribution of the evaluations $x_1, \ldots, x_n$ in the calculation of $C_v(x)$ depends on the evenness of the distribution p^v_{σ,N^+}.

A straightforward way to measure the uniformity of p^v_{σ,N^+} consists in computing $H_S(p^v_{\sigma,N^+})$. Should $H_S(p^v_{\sigma,N^+})$ be close to $\ln n$, the distribution p^v_{σ,N^+} will be approximately uniform and all the evaluations $x_1, \ldots, x_n$ will be involved almost equally in the calculation of $C_v(x)$. On the contrary, should $H_S(p^v_{\sigma,N^+})$ be close to zero, only one $p^v_{\sigma,N^+}(i)$ will be very close to one and $C_v(x)$ will be almost proportional to the corresponding partial evaluation as can be seen from Eq. (10).

From the above remarks, it seems natural to adopt the following definition.

Definition 7 *The extension of the Shannon entropy to a bi-capacity v on N is defined by*

$$\overline{\overline{H}}_S(v) := \frac{1}{2^n} \sum_{N^+ \subseteq N} \frac{1}{n!} \sum_{\sigma \in \Pi_N} H_S(p^v_{\sigma,N^+}). \tag{11}$$

As in the case of capacities, the extended Shannon entropy $\overline{\overline{H}}_S(v)$ is nothing else than an average of the uniformity values of the probability distributions p^v_{σ,N^+} calculated by means of H_S. Interestingly enough, the definition given in Eq. (11) is in complete accordance with the vision of $\mathcal{Q}(N)$ recently adopted by Grabisch and Labreuche in [28,29]. Indeed, the probability distributions $\{p^v_{\sigma,N^+} \,|\, \sigma \in \Pi_N, N^+ \subseteq N\}$ can be regarded as obtained along the maximal

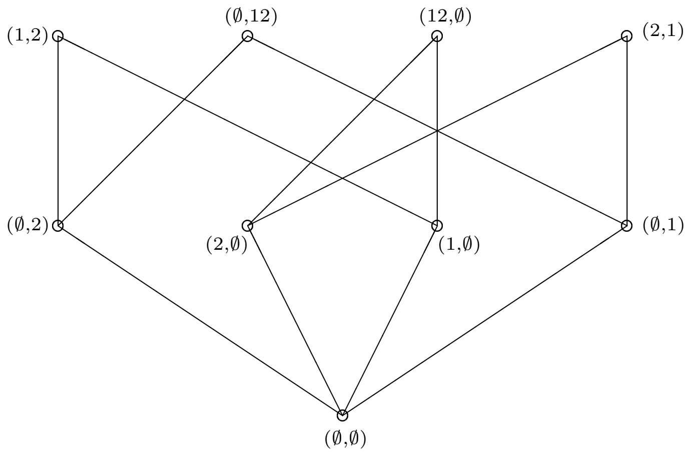

Fig. 1. $(\mathcal{Q}(N), \subseteq)$ with $n = 2$.

chains of the inf-semilattice $(\mathcal{Q}(N), \subseteq)$ where $\subseteq$ denotes the product order :

$$(A, A') \subseteq (B, B') \iff A \subseteq B \text{ and } A' \subseteq B'.$$

The inf-semilattice $(\mathcal{Q}(N), \subseteq)$ is represented in Figure 1 for $n = 2$. More details can be found in [28,29].

In the context of aggregation by the Choquet integral w.r.t. a bi-capacity v on N, let us now show, more formally, that $\overline{\overline{\overline{H}}}_S(v)$ is a measure of the average value over all $x \in [-1,1]^n$ of the degree to which the arguments $x_1, \ldots, x_n$ contribute to the calculation of the aggregated value $C_v(x)$.

To do so, define the sets

$$\begin{aligned}\mathcal{O}_{\sigma,N^+} := \{x \in [-1,1]^n \mid x_i \in [0,1]\, \forall i \in N^+, x_i \in [-1,0[\ \forall i \in N^-,\\ \text{and } |x_{\sigma(1)}| \leq \cdots \leq |x_{\sigma(n)}|\}, \quad (N^+ \subseteq N,\ \sigma \in \Pi_N),\end{aligned}$$

which cover the hypercube $[-1,1]^n$. For a fixed $x \in [-1,1]^n$, there exist $N^+ \subseteq N$ and $\sigma \in \Pi_N$ such that $x \in \mathcal{O}_{\sigma,N^+}$ and, as we can see from Eq. (10), $C_v(x)$ is proportional to $\sum_{i \in N} p^v_{\sigma,N^+}(i)\, x_{\sigma(i)}$.

Starting from Eq. (11) and using the fact that $\int_{x \in \mathcal{O}_{\sigma,N^+}} \mathrm{d}x = 1/n!$, the entropy $\overline{\overline{\overline{H}}}_S(v)$ can be rewritten as

$$\begin{aligned}\overline{\overline{\overline{H}}}_S(\mu) &= \frac{1}{2^n} \sum_{N^+ \subseteq N} \sum_{\sigma \in \Pi_N} \int_{x \in \mathcal{O}_{\sigma,N^+}} H_S(p^v_{\sigma,N^+})\, \mathrm{d}x \\ &= \frac{1}{2^n} \int_{[-1,1]^n} H_S(p^v_{\sigma_x,N^+_x})\, \mathrm{d}x,\end{aligned}$$

where $N^+_x \subseteq N$ and $\sigma_x \in \Pi_N$ are defined such that $x \in \mathcal{O}_{\sigma_x,N^+_x}$.

We thus observe that $\overline{\overline{\overline{H}}}_S(v)$ measures the average value over all $x \in [-1,1]^n$

of the degree to which the arguments $x_1,\ldots,x_n$ contribute to the calculation of $C_v(x)$. In probabilistic terms, it corresponds to the expectation over all $x \in [-1,1]^n$, with uniform distribution, of the degree of contribution of arguments $x_1,\ldots,x_n$ in the calculation of $C_v(x)$.

5 Properties of $\overline{\overline{H}}_S$

The entropy $\overline{\overline{H}}_S$ satisfies many properties that one would expect from an entropy measure. These properties are stated and proved hereafter.

5.1 Boundary conditions

We first present two lemmas giving the form of the probability distributions p^v_{σ,N^+} for CPT type bi-capacities.

For any subset $S \subseteq N$ and any capacity μ on N, we denote by $\mu^{\cap S}$ and $\mu^{\cup S}$ the set functions on $\mathcal{P}(N)$ defined respectively by $\mu^{\cap S}(T) := \mu(T \cap S)$ and $\mu^{\cup S}(T) := \mu(T \cup S)$ for all $T \subseteq N$.

Lemma 8 *For any CPT type bi-capacity v_{μ_1,μ_2} on N, any $N^+ \subseteq N$, and any $\sigma \in \Pi_N$, we have*

$$p^{v_{\mu_1,\mu_2}}_{\sigma,N^+}(i) = \frac{p^{\mu_1^{\cap N^+}}_\sigma(i) + p^{\mu_2^{\cap N^-}}_\sigma(i)}{\mu_1(N^+) + \mu_2(N^-)} \qquad \forall i \in N.$$

Proof. Let v_{μ_1,μ_2} be a bi-capacity of the CPT type on N. Then, for any $N^+ \subseteq N$, any $\sigma \in \Pi_N$, and any $i \in N$, we have, after some algebra,

$$|q^{v_{\mu_1,\mu_2}}_{\sigma,N^+}(i)| = |p^{\mu_1^{\cap N^+}}_\sigma(i) + p^{\mu_2^{\cap N^-}}_\sigma(i)| = p^{\mu_1^{\cap N^+}}_\sigma(i) + p^{\mu_2^{\cap N^-}}_\sigma(i).$$

We further immediately obtain that

$$\sum_{j\in N} |q^{v_{\mu_1,\mu_2}}_{\sigma,N^+}(j)| = \mu_1(N^+) + \mu_2(N^-).$$

□

Lemma 9 *For any asymmetric bi-capacity $v_{\mu,\bar{\mu}}$ on N, any $N^+ \subseteq N$, and any*

$\sigma \in \Pi_N$, *we have*

$$p^{v_{\mu,\bar{\mu}}}_{\sigma,N^+}(i) = \mu^{\cap N^+}(\{\sigma(i),\ldots,\sigma(n)\}) - \mu^{\cap N^+}(\{\sigma(i+1),\ldots,\sigma(n)\}) \\ + \mu^{\cup N^+}(\{\sigma(1),\ldots,\sigma(i)\}) - \mu^{\cup N^+}(\{\sigma(1),\ldots,\sigma(i-1)\}), \qquad \forall i \in N.$$

Proof. For any asymmetric bi-capacity $v_{\mu,\bar{\mu}}$ on N, any $N^+ \subseteq N$, we immediately have that $\mu(N^+) + \bar{\mu}(N^-) = \mu(N) = 1$.

The rest of the proof follows from the immediate identity

$$\bar{\mu}^{\cap N^-}(T) = \mu(N) - \mu^{\cup N^+}(N \setminus T) \qquad \forall T \subseteq N. \qquad \square$$

We now give the form of $\overline{\overline{H}}_S$ for asymmetric bi-capacities.

Proposition 5.1 *For any asymmetric bi-capacity* $v_{\mu,\bar{\mu}}$ *on* N, *we have*

$$\overline{\overline{H}}_S(v_{\mu,\bar{\mu}}) = \overline{H}_S(\mu).$$

Proof. Let $v_{\mu,\bar{\mu}}$ be an asymmetric bi-capacity on N and let $N^+ \subseteq N$. Let $\mathcal{M}_N^{N^+}$ denote the subset of $\mathcal{M}_N$ formed by the maximal chains of $(\mathcal{P}(N), \subseteq)$ containing N^+. From Lemma 9, we see that the $n!$ probability distributions $p^{v_{\mu,\bar{\mu}}}_{\sigma,N^+}$, $\sigma \in \Pi_N$, are defined along the maximal chains of $\mathcal{M}_N^{N^+}$. Now, among the $n!$ maximal chains of $\mathcal{M}_N$, it is easy to verify that there are $n^+!(n-n^+)!$ chains containing N^+. It follows that

$$\sum_{\sigma \in \Pi_N} H_S(p^{v_{\mu,\bar{\mu}}}_{\sigma,N^+}) = \binom{n}{n^+} \sum_{m \in \mathcal{M}_N^{N^+}} H_S(p^\mu_m),$$

where the probability distributions p^μ_m are defined as in Subsection 2.2.

We can then write

$$\begin{aligned} \overline{\overline{H}}_S(v_{\mu,\bar{\mu}}) &= \frac{1}{2^n} \sum_{N^+ \subseteq N} \frac{1}{n!} \binom{n}{n^+} \sum_{m \in \mathcal{M}_N^{N^+}} H_S(p^\mu_m) \\ &= \frac{1}{2^n} \sum_{n^+=0}^{n} \sum_{\substack{N^+ \subseteq N \\ |N^+|=n^+}} \frac{1}{n!} \binom{n}{n^+} \sum_{m \in \mathcal{M}_N^{N^+}} H_S(p^\mu_m), \end{aligned}$$

from which, using the fact that

$$\bigcup_{\substack{N^+ \subseteq N \\ |N^+|=n^+}} \mathcal{M}_N^{N^+} = \mathcal{M}_N$$

and using Eq. (7), we obtain

$$\overline{\overline{H}}_S(v_{\mu,\bar{\mu}}) \;=\; \frac{1}{2^n}\sum_{n^+=0}^{n}\binom{n}{n^+}\frac{1}{n!}\sum_{m\in\mathcal{M}_N} H_S(p_m^\mu) \;=\; \overline{H}_S(\mu). \qquad \square$$

Note that the above property is in full accordance with Eq. (6) and the expression of the asymmetric Choquet integral given in Eq. (2), which coincides with the original definition of the Choquet integral given in Eq. (1).

The following proposition gives the form of $\overline{\overline{H}}_S$ for additive bi-capacities.

Proposition 5.2 *For any additive bi-capacity v_{μ_1,μ_2} on N, we have*

$$\overline{\overline{H}}_S(v_{\mu_1,\mu_2}) = \frac{1}{2^n}\sum_{N^+\subseteq N}\sum_{i\in N} h\left[\frac{\mu_1^{\cap N^+}(i)+\mu_2^{\cap N^-}(i)}{\mu_1(N^+)+\mu_2(N^-)}\right]$$

Proof. Let v_{μ_1,μ_2} be an additive bi-capacity on N. Then, using Lemma 8, for any $N^+\subseteq N$, any $\sigma\in\Pi_N$, and any $i\in N$, we immediately have

$$p_{\sigma,N^+}^{v_{\mu_1,\mu_2}}(i) = \frac{\mu_1^{\cap N^+}(\sigma(i))+\mu_2^{\cap N^-}(\sigma(i))}{\mu_1(N^+)+\mu_2(N^-)}$$

and hence the result. $\square$

We end this subsection by a natural result giving the form of $\overline{\overline{H}}_S$ for additive asymmetric/symmetric bi-capacities.

Proposition 5.3 *For any additive asymmetric/symmetric bi-capacity $v_{\mu,\mu}$ on N, we have*

$$\overline{\overline{H}}_S(v_{\mu,\mu}) = H_S(p),$$

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

Proof. The result follows from Proposition 5.2, or equivalently from Property 5.1 and the properties of $\overline{H}_S$. $\square$

5.2 Symmetry

Proposition 5.4 *For any bi-capacity v on N and any permutation π on N, we have*

$$\overline{\overline{H}}_S(v\circ\pi) = \overline{\overline{H}}_S(v).$$

Proof. Let v be a bi-capacity on N, let $N^+ \subseteq N$, and let $\sigma, \pi \in \Pi_N$. It is very easy to check that $q^{v\circ\pi}_{\sigma,N^+} = q^v_{\pi\circ\sigma,\pi(N^+)}$, which implies $p^{v\circ\pi}_{\sigma,N^+} = p^v_{\pi\circ\sigma,\pi(N^+)}$.

Therefore,

$$\overline{\overline{H}}_S(v\circ\pi) = \frac{1}{2^n}\sum_{N^+\subseteq N}\frac{1}{n!}\sum_{\sigma\in\Pi_N} H_S(p^v_{\pi\circ\sigma,\pi(N^+)}).$$

Setting $\sigma' := \pi\circ\sigma$ and $S := \pi(N^+)$, we obtain

$$\overline{\overline{H}}_S(v) = \frac{1}{2^n}\sum_{S\subseteq N}\frac{1}{n!}\sum_{\sigma'\in\Pi_N} H_S(p^v_{\sigma',S}),$$

which completes the proof. □

5.3 Expansibility

Proposition 5.5 *Let v be a bi-capacity on N. If $k \in N$ is a null element for v, then*

$$\overline{\overline{H}}_S(v) = \overline{\overline{H}}_S(v_{-k}),$$

where v_{-k} denotes the restriction of v to $N \setminus k$.

Proof. Let v be a bi-capacity on N and let $k \in N$ be a null element for v. Since $\overline{\overline{H}}_S$ satisfies the symmetry property, we can assume w.l.o.g. that $k = 1$. We then have

$$\overline{\overline{H}}_S(v) = \frac{1}{2^n}\sum_{N^+\subseteq N\setminus 1}\frac{1}{n!}\sum_{\sigma\in\Pi_N}\left[H_S(p^v_{\sigma,N^+}) + H_S(p^v_{\sigma,N^+\cup 1})\right].$$

Since 1 is a null element, it is easy to verify that $p^v_{\sigma,N^+} = p^v_{\sigma,N^+\cup 1}$. Hence, we have

$$\overline{\overline{H}}_S(v) = \frac{1}{2^n}\sum_{N^+\subseteq N\setminus 1}\frac{1}{n!}\sum_{\sigma\in\Pi_N} 2H_S(p^v_{\sigma,N^+}).$$

Furthermore,

$$\begin{aligned}\sum_{\sigma\in\Pi_N} H_S(p^v_{\sigma,N^+}) &= \sum_{\sigma\in\Pi_N}\sum_{i\in N} h[p^v_{\sigma,N^+}(i)] = \sum_{\sigma\in\Pi_N}\sum_{\substack{i\in N\\ \sigma(i)\neq 1}} h[p^v_{\sigma,N^+}(i)]\\ &= \sum_{j\in N}\sum_{\substack{\sigma\in\Pi_N\\ \sigma(j)=1}}\sum_{i\in N\setminus j} h[p^v_{\sigma,N^+}(i)].\end{aligned}$$

However, we can write

$$\sum_{\substack{\sigma\in\Pi_N\\ \sigma(j)=1}}\sum_{i\in N\setminus j} h[p^v_{\sigma,N^+}(i)] = \sum_{\pi\in\Pi_{N\setminus 1}}\sum_{l\in N\setminus 1} h[p^{v_{-1}}_{\pi,N^+}(l)] = \sum_{\pi\in\Pi_{N\setminus 1}} H_S(p^{v_{-1}}_{\pi,N^+}).$$

Finally,

$$\begin{aligned}\overline{\overline{H}}_S(v) &= \frac{1}{2^{n-1}} \sum_{N^+ \subseteq N \setminus 1} \frac{1}{n!} \sum_{j \in N} \sum_{\pi \in \Pi_{N \setminus 1}} H_S(p^{v_{-1}}_{\pi,N^+}) \\ &= \frac{1}{2^{n-1}} \sum_{N^+ \subseteq N \setminus 1} \frac{1}{(n-1)!} \sum_{\pi \in \Pi_{N \setminus 1}} H_S(p^{v_{-1}}_{\pi,N^+}) \;=\; \overline{\overline{H}}_S(v_{-1}). \qquad \square\end{aligned}$$

5.4 Decisivity

Proposition 5.6 *For any bi-capacity v on N, we have*

$$\overline{\overline{H}}_S(v) \geq 0.$$

Moreover, $\overline{\overline{H}}_S(v) = 0$ if and only, for any $x \in [-1,1]^n$, there exists $\lambda \in]0,2]$ and $i \in N$ such that $C_v(x) = \lambda x_i$.

Proof. From the decisivity property satisfied by the Shannon entropy, we have that, for any probability distribution p on N, $H_S(p) \geq 0$ with equality if and only if p is a Dirac measure.

It follows that, for any given bi-capacity v on N, we have $\overline{\overline{H}}_S(v) \geq 0$, with equality if and only if, for any $N^+ \subseteq N$ and any $\sigma \in \Pi_N$, p^v_{σ,N^+} is a Dirac measure.

We can then write the following equivalences

$$\begin{aligned}&\forall N^+ \subseteq N, \forall \sigma \in \Pi_N,\ p^v_{\sigma,N^+} \text{ is a Dirac measure,}\\ &\quad\Leftrightarrow\ \forall N^+ \subseteq N,\ \forall \sigma \in \Pi_N, \exists!\, i \in N \text{ such that } q^v_{\sigma,N^+}(i) \neq 0,\\ &\quad\Leftrightarrow\ \forall x \in [-1,1]^n, \exists N^+ \subseteq N, \sigma \in \Pi_N, \text{ and } i \in N \text{ such that}\\ &\qquad\ C_v(x) = |q^v_{\sigma,N^+}(i)|\, x_{\sigma(i)},\\ &\quad\Leftrightarrow\ \forall x \in [-1,1]^n, \exists \lambda \in]0,2] \text{ and } i \in N \text{ such that } C_v(x) = \lambda x_i,\end{aligned}$$

since for any $N^+ \subseteq N$, and any $\sigma \in \Pi_N$, $|q^v_{\sigma,N^+}(i)| \in [0,2]$, for all $i \in N$. $\square$

5.5 Maximality

Proposition 5.7 *For any bi-capacity v on N, we have*

$$\overline{\overline{H}}_S(v) \leq \ln n,$$

with equality if and only if v is the uniform bi-capacity v^* on N.

Proof. From the maximality property satisfied by the Shannon entropy, we have that, for any probability distribution p on N, $H_S(p) \leq \ln n$, with equality if and only if p is uniform.

It follows that, given a bi-capacity v on N, we have $\overline{\overline{H}}_S(v) \leq \ln n$, with equality if and only if, for any $N^+ \subseteq N$ and any $\sigma \in \Pi_N$, p^v_{σ,N^+} is uniform.

It is easy to see that if $v = v^*$ then p^v_{σ,N^+} is uniform and hence $\overline{\overline{H}}_S(v) = \ln n$.

Let us now show that if $\overline{\overline{H}}_S(v) = \ln n$, that is, if p^v_{σ,N^+} is uniform for any $N^+ \subseteq N$ and any $\sigma \in \Pi_N$, then necessarily $v = v^*$, that is, for any $N^+ \subseteq N$ and any $\sigma \in \Pi_N$,

$$|q^v_{\sigma,N^+}(i)| = \frac{1}{n}, \qquad \forall i \in N.$$

To do so, consider first the case where $N^+ \in \{\emptyset, N\}$. From the normalization condition $v(N, \emptyset) = 1 = -v(\emptyset, N)$, it easy to verify that, for any $\sigma \in \Pi_N$,

$$\sum_{j \in N} |q^v_{\sigma,N^+}(j)| = 1.$$

It follows that, if, for any $\sigma \in \Pi_N$, p^v_{σ,N^+} is uniform, then,

$$|q^v_{\sigma,N^+}(i)| = \frac{1}{n}, \qquad \forall i \in N, \forall \sigma \in \Pi_N.$$

This implies that,

$$v(i, \emptyset) = \frac{1}{n} = -v(\emptyset, i), \qquad \forall i \in N. \tag{12}$$

Consider now the case where $N^+ \in \mathcal{P}(N) \setminus \{\emptyset, N\}$. From Eq. (12), we have that, for any $\sigma \in \Pi_N$,

$$|q^v_{\sigma,N^+}(n)| = \frac{1}{n}.$$

Since, for any $\sigma \in \Pi_N$, p^v_{σ,N^+} is uniform, we obtain that

$$|q^v_{\sigma,N^+}(i)| = \frac{1}{n}, \qquad \forall i \in N.$$

□

5.6 Increasing monotonicity toward v^*

Proposition 5.8 *Let v be a bi-capacity on N such that $v \neq v^*$ and, for any $\lambda \in [0, 1]$, define the bi-capacity v_λ on N as $v_\lambda := v + \lambda(v^* - v)$. Then for any*

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

$$\overline{\overline{H}}_S(v_{\lambda_1}) < \overline{\overline{H}}_S(v_{\lambda_2}).$$

Proof. We proceed mainly as in [9]. Let $\lambda \in]0,1[$ and let us prove that

$$\frac{d}{d\lambda}\overline{\overline{H}}_S(v_\lambda) > 0.$$

Let $N^+ \subseteq N$ and let $\sigma \in \Pi_N$. Since $v_\lambda = \lambda v^* + (1-\lambda)v$, for any $i \in N$, we have

$$\begin{aligned}|q^{v_\lambda}_{\sigma,N^+}(i)| &= |\lambda\, q^{v^*}_{\sigma,N^+}(i) + (1-\lambda)\, q^{v}_{\sigma,N^+}(i)| \\ &= \lambda|q^{v^*}_{\sigma,N^+}(i)| + (1-\lambda)|q^{v}_{\sigma,N^+}(i)|,\end{aligned}$$

since $q^{v^*}_{\sigma,N^+}(i)$ and $q^{v}_{\sigma,N^+}(i)$ have the same sign. It follows that

$$|q^{v_\lambda}_{\sigma,N^+}(i)| = \frac{\lambda}{n} + (1-\lambda)|q^{v}_{\sigma,N^+}(i)|.$$

Consequently,

$$p^{v_\lambda}_{\sigma,N^+}(i) = \frac{\frac{\lambda}{n} + (1-\lambda)|q^{v}_{\sigma,N^+}(i)|}{\lambda + (1-\lambda)\sum_{j\in N}|q^{v}_{\sigma,N^+}(j)|}, \qquad \forall i \in N.$$

Let us now compute

$$\frac{d}{d\lambda}H_S(p^{v_\lambda}_{\sigma,N^+}) = \sum_{i\in N}\frac{d}{d\lambda}h[p^{v_\lambda}_{\sigma,N^+}(i)].$$

For any $i \in N$, we have

$$\frac{d}{d\lambda}h[p^{v_\lambda}_{\sigma,N^+}(i)] = \frac{d}{d\lambda}p^{v_\lambda}_{\sigma,N^+}(i) \times [-1 - \ln p^{v_\lambda}_{\sigma,N^+}(i)],$$

with

$$\frac{d}{d\lambda}p^{v_\lambda}_{\sigma,N^+}(i) = \frac{-|q^{v}_{\sigma,N^+}(i)| + \frac{1}{n}\sum_{j\in N}|q^{v}_{\sigma,N^+}(j)|}{\left(\lambda + (1-\lambda)\sum_{j\in N}|q^{v}_{\sigma,N^+}(j)|\right)^2}.$$

Since $\sum_{i\in N}\frac{d}{d\lambda}p^{v_\lambda}_{\sigma,N^+}(i) = 0$, we obtain

$$\begin{aligned}\frac{d}{d\lambda}H_S(p^{v_\lambda}_{\sigma,N^+}) &= -\sum_{i\in N}\frac{d}{d\lambda}p^{v_\lambda}_{\sigma,N^+}(i) \times \ln p^{v_\lambda}_{\sigma,N^+}(i) \\ &= -\sum_{i\in N}\frac{d}{d\lambda}p^{v_\lambda}_{\sigma,N^+}(i) \times \ln\left(\frac{\lambda}{n} + (1-\lambda)|q^{v}_{\sigma,N^+}(i)|\right).\end{aligned}$$

Hence, $\frac{d}{d\lambda} H_S(p^{v_\lambda}_{\sigma,N^+})$ has the same sign as

$$\sum_{i\in N}\left(|q^v_{\sigma,N^+}(i)| - \frac{1}{n}\sum_{j\in N}|q^v_{\sigma,N^+}(j)|\right) \times \ln\left(\frac{\lambda}{n} + (1-\lambda)|q^v_{\sigma,N^+}(i)|\right). \tag{13}$$

Let

$$d^v_{\sigma,N^+}(i) := |q^v_{\sigma,N^+}(i)| - \frac{1}{n}\sum_{j\in N}|q^v_{\sigma,N^+}(j)|, \qquad \forall i \in N.$$

Now, for any $i \in N$, if $d^v_{\sigma,N^+}(i) > 0$ (resp < 0), then, clearly,

$$|q^v_{\sigma,N^+}(i)| > \frac{1}{n}\sum_{j\in N}|q^v_{\sigma,N^+}(j)| \text{ (resp. <)},$$

and hence

$$\ln\left(\frac{\lambda}{n} + (1-\lambda)|q^v_{\sigma,N^+}(i)|\right) > \ln\left(\frac{\lambda}{n} + (1-\lambda)\frac{1}{n}\sum_{j\in N}|q^v_{\sigma,N^+}(j)|\right) \text{ (resp. <)}. \tag{14}$$

Eq. (13) can then be rewritten as

$$\sum_{\substack{i\in N\\ d^v_{\sigma,N^+}(i)>0}} d^v_{\sigma,N^+}(i)\ln\left(\frac{\lambda}{n} + (1-\lambda)|q^v_{\sigma,N^+}(i)|\right) - \sum_{\substack{i\in N\\ d^v_{\sigma,N^+}(i)<0}} -d^v_{\sigma,N^+}(i)\ln\left(\frac{\lambda}{n} + (1-\lambda)|q^v_{\sigma,N^+}(i)|\right),$$

which, using Eq. (14), is strictly greater than

$$\begin{aligned}&\sum_{\substack{i\in N\\ d^v_{\sigma,N^+}(i)>0}} d^v_{\sigma,N^+}(i)\ln\left(\frac{\lambda}{n} + (1-\lambda)\frac{1}{n}\sum_{j\in N}|q^v_{\sigma,N^+}(j)|\right)\\ &\quad - \sum_{\substack{i\in N\\ d^v_{\sigma,N^+}(i)<0}} -d^v_{\sigma,N^+}(i)\ln\left(\frac{\lambda}{n} + (1-\lambda)\frac{1}{n}\sum_{j\in N}|q^v_{\sigma,N^+}(j)|\right)\\ &\quad = \ln\left(\frac{\lambda}{n} + (1-\lambda)\frac{1}{n}\sum_{j\in N}|q^v_{\sigma,N^+}(j)|\right)\underbrace{\sum_{i\in N} d^v_{\sigma,N^+}(i)}_{=0}\\ &\quad = 0,\end{aligned}$$

which implies that $\frac{d}{d\lambda} H_S(p^{v_\lambda}_{\sigma,N^+}) > 0$, and therefore the desired result. $\square$

Table 1
Partial evaluations of four students. The marks are given of a $[0, 20]$ bipolar scale with 10 as neutral level.

Student	M	S	E	Student	M	S	E
a	14	16	7	c	9	16	7
b	14	15	8	d	9	15	8

6 On the practical use of $\overline{\overline{H}}_S$

To illustrate the potential usefulness of the defined entropy in practical applications, we consider the following simple example taken from [30].

The dean of a scientific faculty aims at evaluating his students from their marks in Mathematics (M), Statistics (S), and English (E). Among others, he adopts the following rules of reasoning :

- If a student is good in M, then E is more important than S.
- If a student is bad in M, then S is more important than E.

Considering the four students whose partial evaluations are given in Table 1, his preferences are then naturally $b \succ a \succ c \succ d$. As shown in [30], the preferences of the dean cannot be modeled by a Choquet integral w.r.t. a capacity. However, they can be modelled by a Choquet integral w.r.t. a bi-capacity v on $\{M, S, E\}$ whose coefficients have to satisfy :

$$v(MS, \emptyset) - v(MS, L) > v(S, \emptyset) \qquad \text{and} \qquad v(S, L) > 0.$$

Assume now that based on additional preferential information, the dean hesitates between two Choquet integral models grounded respectively on two bi-capacities v_1 and v_2 on $\{M, S, E\}$, both of course satisfying the previous inequalities, and such that $\overline{\overline{H}}_S(v_1)/\ln n = 0.523$ and $\overline{\overline{H}}_S(v_2)/\ln n = 0.892$. From the previous section, in such a case, we know that C_{v_2} will exploit more its arguments on average than C_{v_1} which may be a sensible argument in favor of the aggregation model C_{v_2}.

More generally, extending the approaches adopted in [17,12,11,31], the entropy $\overline{\overline{H}}_S$ could be used in the framework of a maximum entropy like principle to determine, if it exists, the "least specific" bi-capacity compatible with the preferences of the decision maker, the expression "least specific" refering to a bi-capacity such that the corresponding Choquet integral is the closest to the simple arithmetic mean in the sense of $\overline{\overline{H}}_S$.

7 Conclusion

An extension of the Shannon entropy to bi-capacities has been proposed. In the framework of aggregation by the Choquet integral w.r.t. a bi-capacity, the main characteristic of this generalized entropy is that it can be interpreted as a measure of the average value of the degree to which the partial evaluations contribute in the calculation of the aggregated value. It can therefore be used, in combination with other behavioral indices, to better understand the aggregation process. In that respect, the generalized entropy, $\overline{\overline{H}}_S$, arises as a natural generalization of Marichal's extension of the Shannon entropy to capacities, $\overline{H}_S$, which has the same interpretation in the context of aggregation by the Choquet integral w.r.t. a capacity. Interestingly enough, $\overline{\overline{H}}_S$ satisfies most of the properties satisfied by $\overline{H}_S$. Notice however, that certain properties, such as the decisivity property, have a slightly weaker form in the case of $\overline{\overline{H}}_S$.

References

[1] I. Kojadinovic, J.-L. Marichal, An extension of the Shannon entropy to bi-capacities, in: 3rd International Summer School on Aggregation Operators (AGOP 2005), Lugano, Switzerland, 2005, pp. 67–72.

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

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

[4] C. E. Shannon, A mathematical theory of communication, Bell Systems Technical Journal 27 (1948) 379–623.

[5] T. Cover, J. Thomas, Elements of Information Theory, John Wiley and Sons, 1991.

[6] G. Choquet, Theory of capacities, Annales de l'Institut Fourier 5 (1953) 131–295.

[7] M. Sugeno, Theory of fuzzy integrals and its applications, Ph.D. thesis, Tokyo Institute of Technology, Tokyo, Japan (1974).

[8] J.-L. Marichal, Aggregation operators for multicriteria decison aid, Ph.D. thesis, University of Liège, Liège, Belgium (1998).

[9] J.-L. Marichal, Entropy of discrete Choquet capacities, European Journal of Operational Research 3 (137) (2002) 612–624.

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

[11] I. Kojadinovic, J.-L. Marichal, M. Roubens, An axiomatic approach to the definition of the entropy of a discrete Choquet capacity, Information Sciences 172 (2005) 131–153.

[12] J.-L. Marichal, Behavioral analysis of aggregation in multicriteria decision aid, in: J. Fodor, B. D. Baets, P. Perny (Eds.), Preferences and Decisions under Incomplete Knowledge, Physica-Verlag, 2000, pp. 153–178.

[13] M. Grabisch, C. Labreuche, Bi-capacities I : definition, Möbius transform and interaction, Fuzzy Sets and Systems 151 (2005) 211–236.

[14] M. Grabisch, C. Labreuche, Bi-capacities II : the Choquet integral, Fuzzy Sets and Systems 151 (2005) 237–259.

[15] C. Labreuche, M. Grabisch, Generalized Choquet-like aggregation functions for handling bipolar scales, European Journal of Operational Research (2005) in press.

[16] M. Grabisch, C. Labreuche, The symmetric and asymmetric choquet integrals on finite spaces for decision making, Statistical Papers 43 (2002) 37–52.

[17] 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) (2000) 800–807.

[18] A. Tversky, D. Kahneman, Advances in prospect theory : cumulative representation of uncertainty, Journal of Risk and Uncertainty 5 (1992) 297–323.

[19] M. Grabisch, The representation of importance and interaction of features by fuzzy measures, Pattern Recognition Letters 17 (6) (1996) 567–575.

[20] C. Labreuche, M. Grabisch, The Choquet integral for the aggregation of interval scales in multicriteria decision making, Fuzzy Sets and Systems 137 (2003) 11–16.

[21] T. Murofushi, M. Sugeno, Fuzzy measures and fuzzy integrals, in: M. Grabisch, T. Murofushi, M. Sugeno (Eds.), Fuzzy Measures and Integrals: Theory and Applications, Physica-Verlag, 2000, pp. 3–41.

[22] D. Denneberg, Non-additive Measure and Integral, Kluwer Academic, Dordrecht, 1994.

[23] M. Grabisch, C. Labreuche, To be symmetric or asymmetric? A dilemna in decision making, in: Preferences, D. under Incomplete Knowledge (Eds.), Preferences and Decisions under Incomplete Knowledge, Physica-Verlag, 2000, pp. 179–194.

[24] J. Šipoš, Integral with respect to a pre-measure, Math. Slovaca 29 (1979) 141–155.

[25] C. Shannon, W. Weaver, A mathematical theory of communication, University of Illinois, Urbana, 1949.

[26] M. Esteban, D. Morales, A summary on entropy statistics, Kybernetika 31 (4) (1995) 337–346.

[27] E. Jaynes, Probability Theory: The Logic of Science, Cambridge University Press, 2003.

[28] M. Grabisch, C. Labreuche, A general construction for unipolar and bipolar interpolative aggregation, in: Joint 4th EUSFLAT and 11th LFA Conference (EUSFLAT 2005), Barcelona, Spain, 2005, pp. 916–921.

[29] M. Grabisch, C. Labreuche, Bi-capacities: towards a generalization of cumulative prospect theory, working paper (2005).

[30] M. Grabisch, C. Labreuche, Fuzzy measures and integrals in MCDA, in: S. G. J. Figueira, M. Ehrgott (Eds.), Multiple Criteria Decision Analysis, Kluwer Academic Publishers, 2004, pp. 563–608.

[31] I. Kojadinovic, Minimum variance capacity identification, European Journal of Operational Research (2005) in press.