Entropy of discrete Choquet capacities

JEAN-LUC MARICHAL*

Institute of Mathematics, University of Liège

Grande Traverse 12 - B37, Sart Tilman, B-4000 Liège, Belgium.

Email: jl.marichal[at]ulg.ac.be

Revised version, February 23, 2001

Abstract

We introduce a measure of entropy for any discrete Choquet capacity and we interpret it in the setting of aggregation by the Choquet integral.

Keywords : entropy, discrete Choquet capacity, aggregation, Choquet integral.

1 Introduction

Given a probability distribution $p = (p_1, \ldots, p_n) \in [0, 1]^n$, with $\sum_i p_i = 1$, the expression

$$H(p) := \sum_{i=1}^{n} h(p_i),$$

where $h(x) := -x \ln x$ if $x > 0$ and 0 if $x = 0$, is called the *Shannon entropy* of p, see [16, 17]. This value, well-known in information theory, measures in some sense the *uncertainty* which prevailed before an experiment was accomplished, or the *information* expected from the experiment. Note that it was characterized axiomatically by many authors, see e.g. [1, 3].

Now, consider a Choquet capacity (or fuzzy measure) on $N := \{1, \ldots, n\}$, that is, a set function $v : 2^N \to [0, 1]$ such that $v(\emptyset) = 0$, $v(N) = 1$, and $v(S) \leq v(T)$ whenever $S \subseteq T$. The following question arises : What is the generalized counterpart of the Shannon entropy for such a capacity?

For particular capacities, such as belief and plausibility measures, some candidates were proposed in evidence theory in the early 1980s, see e.g. [2, 7, 8, 19]. However, it seems that no definition of entropy for a general Choquet capacity was yet proposed in the literature.

In this paper we present an entropy-like measure defined for all discrete Choquet capacities. This "entropy" was proposed very recently by the author [9] in the framework of aggregation. Although it has yet to be characterized, it satisfies properties considered as requisites for defining an entropy. In particular, it collapses into the Shannon entropy as soon as the capacity is additive.

The outline of the paper is as follows. In Section 2 we interpret the use of the Shannon entropy in the context of aggregation. In Sections 3 and 4 we introduce the concept of entropy of any discrete capacity and we point out several of its properties. In Section 5 we give an example demonstrating its use in practical situations. Finally, in Section 6 we give a probabilistic interpretation of it.

In order to avoid a heavy notation, cardinality of subsets $S, T, \ldots$ will often be denoted by the corresponding lower case letters $s, t, \ldots$. Moreover, we will often omit braces for singletons, e.g., writing $T \cup i$ instead of $T \cup \{i\}$.

2 Entropy and aggregation operators

Consider the weighted arithmetic mean (WAM) as an aggregation operator:

$$\mathrm{WAM}_\omega(x_1, \ldots, x_n) = \sum_{i=1}^n \omega_i \, x_i \qquad (x \in \mathbb{R}^n) \,,$$

with

$$\sum_{i=1}^n \omega_i = 1 \quad \text{and} \quad \omega_i \geq 0 \quad (i \in N) \,.$$

Clearly, the extent to which the arguments $x_1, \ldots, x_n$ are used in such an aggregation process strongly depends upon the weight vector ω. For example, consider two weighted arithmetic means with weight vectors of the form

$$(1, 0, \ldots, 0) \quad \text{and} \quad (1/n, \ldots, 1/n) \,,$$

respectively. We note that these operators are quite different in the sense that the first one focuses the total weight on only one argument (projection on the first argument) whereas the second one distributes the total weight among all the arguments evenly (arithmetic mean).

In order to capture this idea, we can define a *measure of dispersion* associated to the weight vector of the weighted arithmetic mean WAM_ω as the Shannon entropy of ω:

$$H(\omega) = \sum_{i=1}^n h(\omega_i) \,.$$

Such a function enables us to measure to what extent the arguments are really used by the weighted arithmetic mean. In a certain sense the more disperse the ω the more the arguments are being used in the aggregation process.

Now, consider the so-called ordered weighted averaging operator (OWA), proposed in 1988 by Yager [20]:

$$\mathrm{OWA}_\omega(x_1, \ldots, x_n) = \sum_{i=1}^n \omega_i \, x_{(i)} \qquad (x \in \mathbb{R}^n) \,,$$

with

$$\sum_{i=1}^n \omega_i = 1 \quad \text{and} \quad \omega_i \geq 0 \quad (i \in N) \,,$$

where $(\cdot)$ indicates a permutation of indices such that $x_{(1)} \leq \cdots \leq x_{(n)}$. For this aggregation operator, the measure of dispersion, which should not be influenced when changing the

assignment of weights to the arguments, should also be given by the Shannon entropy. In fact, Yager [20] proposed explicitly to use this concept as measure of dispersion for the OWA operators.

It is known that $H(\omega)$ is maximum only when ω corresponds to the weight vector of the arithmetic mean, see e.g. [19]:

$$H(\omega) = \ln n \ \text{ for } \ \omega = (1/n, \ldots, 1/n) \,,$$

and minimum only when ω is a binary vector (Dirac measure):

$$H(\omega) = 0 \ \text{ if } \ \omega_i = 1 \ \text{ for some } \ i \in N.$$

Thus, the measure of dispersion can be normalized into

$$\frac{1}{\ln n} \, H(\omega) = - \sum_{i=1}^{n} \omega_i \log_n \omega_i \,,$$

so that it ranges in $[0, 1]$.

3 Entropy of discrete Choquet capacities

Given a Choquet capacity v on N, the *Choquet integral* of $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ with respect to v is defined by

$$\mathcal{C}_v(x) = \sum_{i=1}^{n} x_{(i)} \left[v(A_{(i)}) - v(A_{(i+1)}) \right] \,,$$

with the convention that $x_{(1)} \leq \cdots \leq x_{(n)}$. Also $A_{(i)} := \{(i), \ldots, (n)\}$, and $A_{(n+1)} := \emptyset$. Note that, in this definition, the coefficients of the capacity are interpreted as weights (or importance) of subsets. For more details, see e.g. [5, 10] and the references therein.

It is easy to see that the WAM operators correspond to the Choquet integrals with respect to additive capacities (i.e., such that $v(S \cup T) = v(S) + v(T)$ whenever $S \cap T = \emptyset$). On the other hand, one can show [4] that the OWA operators are exactly those Choquet integrals which are symmetric, that is, independent of any permutation of the arguments. Thus, the Choquet integral is a simultaneous generalization of both WAM and OWA operators.

Starting from these facts, the author [9, §6.2.4] proposed to define the entropy of a capacity v as a measure of dispersion for the Choquet integral $\mathcal{C}_v$. This measure should identify with the Shannon entropy when the Choquet integral is either a WAM or an OWA.

On the one hand, comparing

$$\mathrm{OWA}_\omega(x) = \sum_{i=1}^{n} x_{(i)} \, \omega_i$$

and

$$\mathcal{C}_v(x) = \sum_{i=1}^{n} x_{(i)} \left[v(A_{(i)}) - v(A_{(i+1)}) \right]$$

suggests proposing as measure of dispersion for $\mathcal{C}_v$ a sum over $i \in N$ of an average value of

$$h[v(T \cup i) - v(T)] \qquad (T \subseteq N \setminus i) \,,$$

3

that is, an expression of the form

$$H_\mathrm{M}(v) = \sum_{i=1}^{n} \sum_{T \subseteq N \setminus i} p_t \, h[v(T \cup i) - v(T)] \,, \tag{1}$$

where the coefficients p_t $(t = |T|)$ are non-negative and such that $\sum_{T \subseteq N \setminus i} p_t = 1$.

On the other hand, imposing the condition

$$\mathcal{C}_v = \mathrm{OWA}_\omega \quad \Rightarrow \quad H_\mathrm{M}(v) = H(\omega) \tag{2}$$

determines uniquely the coefficients p_t, as the following result shows.

Proposition 3.1 *Condition (2) holds if and only if*

$$p_t = \frac{(n - t - 1)! \, t!}{n!} \qquad (t = 0, \dots, n - 1) \,.$$

Proof. One can easily show [11, §2.2] that if $\mathcal{C}_v = \mathrm{OWA}_\omega$ then

$$v(T \cup i) - v(T) = \omega_{n-t} \qquad (i \in N; \ T \subseteq N \setminus i) \,.$$

Thus, the measure of dispersion (1) becomes:

$$\begin{aligned}
H_\mathrm{M}(v) &= \sum_{i=1}^{n} \sum_{T \subseteq N \setminus i} p_t \, h(\omega_{n-t}) \\
&= \sum_{i=1}^{n} \sum_{t=0}^{n-1} \binom{n-1}{t} p_t \, h(\omega_{n-t}) \\
&= \sum_{t=0}^{n-1} n \binom{n-1}{t} p_t \, h(\omega_{n-t}) \\
&= \sum_{i=1}^{n} q_i \, h(\omega_i) \,,
\end{aligned}$$

where $q_i := n \binom{n-1}{n-i} p_{n-i}$ for all $i \in N$. Hence, condition (2) is equivalent to

$$\sum_{i=1}^{n} (q_i - 1) \, h(\omega_i) = 0 \quad \text{for all weight vectors } \omega. \tag{3}$$

Now, let $k, l \in N$, $k \neq l$, and consider the weight vector ω defined by

$$\omega_i = \begin{cases} \alpha \,, & \text{if } i = k, \\ 1 - \alpha \,, & \text{if } i = l, \\ 0 \,, & \text{otherwise,} \end{cases}$$

where $\alpha \in \,]0, 1[$. In this case condition (3) becomes

$$(q_k - 1)\alpha \ln \alpha + (q_l - 1)(1 - \alpha) \ln(1 - \alpha) = 0 \qquad (k, l \in N; \ \alpha \in \,]0, 1[) \,.$$

Replacing α first by $1/2$ and then by $1/4$ provides $q_i = 1$ for all $i \in N$. This completes the proof. ∎

We then propose the following definition.

Definition 3.1 *The entropy of a Choquet capacity v on N is defined by*

$$H_{\mathrm{M}}(v) := \sum_{i=1}^{n} \sum_{T \subseteq N \setminus i} \frac{(n-t-1)!\, t!}{n!}\, h[v(T \cup i) - v(T)]\,.$$

When the Choquet integral $\mathcal{C}_v$ is used as an aggregation operator, the normalized entropy

$$\frac{1}{\ln n}\, H_{\mathrm{M}}(v)$$

can be interpreted as the degree to which one uses the arguments $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ when calculating the aggregated value $\mathcal{C}_v(x)$.

Introducing the notation $\gamma_t(n) := (n-t-1)!\, t!/n!$, we can write

$$H_{\mathrm{M}}(v) = \sum_{i=1}^{n} \sum_{T \subseteq N \setminus i} \gamma_t(n)\, h[v(T \cup i) - v(T)]\,.$$

Interestingly enough, this expression is very similar to that of the Shapley value of elements in N, which is a fundamental concept in game theory [18] expressing a power index:

$$\phi_i(v) = \sum_{T \subseteq N \setminus i} \gamma_t(n)\, [v(T \cup i) - v(T)] \qquad (i \in N)\,.$$

4 Properties of H_{M}

To justify the use of this new definition, one can show that it fulfills several properties required for an entropy (see [1, 3]). In this section we present some of them.

4.1 Symmetry

We observe that $H_{\mathrm{M}}(v)$ is symmetric in the sense that permuting the elements of N has no effect on the dispersion. To prove this, we need the following notation. For any permutation π on N, we denote by πv the capacity on N defined by $\pi v(\pi(S)) = v(S)$ for all $S \subseteq N$, where $\pi(S) := \{\pi(i) \,|\, i \in S\}$.

One can easily show that

$$\mathcal{C}_{\pi v}(x_1, \ldots, x_n) = \mathcal{C}_v(x_{\pi(1)}, \ldots, x_{\pi(n)}) \qquad (x \in \mathbb{R}^n)\,.$$

Moreover, we have the following result.

Proposition 4.1 *For any capacity v on N and any permutation π on N, we have*

$$H_{\mathrm{M}}(\pi v) = H_{\mathrm{M}}(v)\,.$$

Proof. By definition, we have

$$
\begin{aligned}
H_{\mathrm{M}}(\pi v) &= \sum_{i=1}^{n} \sum_{T \subseteq N \setminus i} \gamma_t(n)\, h[v(\pi^{-1}(T \cup i)) - v(\pi^{-1}(T))] \\
&= \sum_{\pi^{-1}(i)=1}^{n} \sum_{\pi^{-1}(T) \subseteq N \setminus \pi^{-1}(i)} \gamma_t(n)\, h[v(\pi^{-1}(T \cup i)) - v(\pi^{-1}(T))]\,.
\end{aligned}
$$

Then, setting $j := \pi^{-1}(i)$ and $S := \pi^{-1}(T)$ (implying $s = t$), we have

$$H_{\mathrm{M}}(\pi v) = \sum_{j=1}^{n} \sum_{S \subseteq N \setminus j} \gamma_s(n)\, h[v(S \cup j) - v(S)] = H_{\mathrm{M}}(v)\,,$$

which completes the proof. ∎

4.2 Maximality

We now prove that $0 \le H_{\mathrm{M}}(v) \le \ln n$ and that $H_{\mathrm{M}}(v)$ is maximum ($= \ln n$) if and only if $\mathcal{C}_v$ is the arithmetic mean. The proof is a straightforward adaptation of that used for the Shannon entropy.

Lemma 4.1 *If the numbers $c_T^i > 0$ and $d_T^i \ge 0$ ($i \in N$; $T \subseteq N \setminus i$) are such that*

$$\sum_{i=1}^{n} \sum_{T \subseteq N \setminus i} \gamma_t(n)\, c_T^i \le 1 \quad and \quad \sum_{i=1}^{n} \sum_{T \subseteq N \setminus i} \gamma_t(n)\, d_T^i = 1$$

then

$$\sum_{i=1}^{n} \sum_{T \subseteq N \setminus i} \gamma_t(n)\, d_T^i \, \ln\Big(\frac{c_T^i}{d_T^i}\Big) \le 0\,.$$

Moreover, the equality holds if and only if

$$d_T^i \ne 0 \;\; \Rightarrow \;\; d_T^i = c_T^i \qquad (i \in N;\; T \subseteq N \setminus i)\,.$$

Proof. Since $\ln x \le x - 1$ for all $x > 0$, we simply have

$$
\begin{aligned}
\sum_{i=1}^{n} \sum_{\substack{T \subseteq N \setminus i \\ d_T^i \ne 0}} \gamma_t(n)\, d_T^i \, \ln\Big(\frac{c_T^i}{d_T^i}\Big) \;\; &\le \;\; \sum_{i=1}^{n} \sum_{\substack{T \subseteq N \setminus i \\ d_T^i \ne 0}} \gamma_t(n)\, d_T^i \, \Big(\frac{c_T^i}{d_T^i} - 1\Big) \\
&= \;\; \sum_{i=1}^{n} \sum_{\substack{T \subseteq N \setminus i \\ d_T^i \ne 0}} \gamma_t(n)\, c_T^i - 1 \\
&\le \;\; 1 - 1 = 0.
\end{aligned}
$$

The second part follows from the fact that $\ln x = x - 1$ if and only if $x = 1$. ∎

Proposition 4.2 *For any capacity v on N, we have $0 \le H_{\mathrm{M}}(v) \le \ln n$. Moreover, $H_{\mathrm{M}}(v) = \ln n$ if and only if $v(T) = t/n$ for all $T \subseteq N$.*

Proof. For all $i \in N$ and all $T \subseteq N \setminus i$, we have $v(T \cup i) - v(T) \in [0, 1]$ and hence $h[v(T \cup i) - v(T)] \ge 0$ and $H_{\mathrm{M}}(v) \ge 0$.

For the second inequality, we have, since $\sum_{i=1}^{n} \phi_i(v) = 1$,

$$
\begin{aligned}
H_{\mathrm{M}}(v) - \ln n \;\; &= \;\; \sum_{i=1}^{n} \sum_{T \subseteq N \setminus i} \gamma_t(n)\, [v(T \cup i) - v(T)] \Big(-\ln[v(T \cup i) - v(T)] - \ln n \Big) \\
&= \;\; \sum_{i=1}^{n} \sum_{T \subseteq N \setminus i} \gamma_t(n)\, [v(T \cup i) - v(T)] \ln \frac{1}{n\,[v(T \cup i) - v(T)]}\,.
\end{aligned}
$$

Applying Lemma 4.1 with $c_T^i = 1/n$ and $d_T^i = v(T \cup i) - v(T)$ leads to $H_{\mathrm{M}}(v) - \ln n \le 0$. The equality holds if and only if $v(T \cup i) - v(T) = 1/n$. ∎

4.3 Decisivity

We now prove that $H_{\mathrm{M}}(v)$ is minimum if and only if v is a binary-valued capacity. This extends the classical decisivity property.

Proposition 4.3 *For any capacity v on N, we have $H_{\mathrm{M}}(v) = 0$ if and only if v is a binary-valued capacity, that is, $v(S) \in \{0, 1\}$ for all $S \subseteq N$.*

Proof. We simply have

$$
\begin{aligned}
H_{\mathrm{M}}(v) = 0 \;\; &\Leftrightarrow \;\; h[v(T \cup i) - v(T)] = 0 \qquad (i \in N; \; T \subseteq N \setminus i) \\
&\Leftrightarrow \;\; v(T \cup i) - v(T) \in \{0, 1\} \qquad (i \in N; \; T \subseteq N \setminus i) \\
&\Leftrightarrow \;\; v(T) \in \{0, 1\} \qquad (T \subseteq N) \,,
\end{aligned}
$$

which proves the result. ∎

Proposition 4.3 is in accordance with the idea of a dispersion measure. Indeed, it can be easily shown (cf. [12, Theorem 5.1]) that v is a binary-valued capacity if and only if

$$
\mathcal{C}_v(x) \in \{x_1, \ldots, x_n\} \qquad (x \in \mathbb{R}^n) \,.
$$

In other terms, $H_{\mathrm{M}}(v)$ is minimum $(= 0)$ if and only if only one argument is used in the aggregation.

4.4 Boundary conditions

The next result is a straightforward consequence of the definition of H_{M}.

Proposition 4.4 *We have*

$$
\mathcal{C}_v = \mathrm{WAM}_\omega \;\; or \;\; \mathrm{OWA}_\omega \quad \Rightarrow \quad H_{\mathrm{M}}(v) = H(\omega) \,.
$$

Proof. Trivial. ∎

4.5 Expansibility

Let $k \in N$ be a *null* element for v, that is, $v(T \cup k) = v(T)$ for all $T \subseteq N \setminus k$. Since such an element does not contribute in the aggregation process, it can be omitted without changing the result. Indeed, if k is a null element then, denoting by v_{-k} the restriction of v to $N \setminus k$, one can easily show that

$$
\mathcal{C}_v(x_1, \ldots, x_n) = \mathcal{C}_{v_{-k}}(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n) \qquad (x \in \mathbb{R}^n) \,,
$$

(see [15]). We then have the following result, which is an extension of the classical expansibility property.

Proposition 4.5 *Let v be a capacity on N. If $k \in N$ is a null element then*

$$
H_{\mathrm{M}}(v) = H_{\mathrm{M}}(v_{-k}) \,.
$$

Proof. Assume that k is a null element. We then have

$$
\begin{aligned}
H_{\mathrm{M}}(v) &= \sum_{\substack{i=1 \\ i \neq k}}^{n} \sum_{T \subseteq N \setminus i} \gamma_t(n)\, h[v(T \cup i) - v(T)] \\
&= \sum_{\substack{i=1 \\ i \neq k}}^{n} \sum_{\substack{T \subseteq N \setminus i \\ T \ni k}} \gamma_t(n)\, h[v(T \cup i) - v(T)] + \sum_{\substack{i=1 \\ i \neq k}}^{n} \sum_{\substack{T \subseteq N \setminus i \\ T \not\ni k}} \gamma_t(n)\, h[v(T \cup i) - v(T)].
\end{aligned}
$$

However, if $T \ni k$ then

$$
v(T \cup i) - v(T) = v((T \setminus k) \cup i) - v(T \setminus k),
$$

and hence,

$$
\begin{aligned}
H_{\mathrm{M}}(v) &= \sum_{\substack{i=1 \\ i \neq k}}^{n} \sum_{\substack{T \subseteq N \setminus i \\ T \not\ni k}} [\gamma_{t+1}(n) + \gamma_t(n)]\, h[v(T \cup i) - v(T)] \\
&= \sum_{\substack{i=1 \\ i \neq k}}^{n} \sum_{\substack{T \subseteq N \setminus i \\ T \not\ni k}} \gamma_t(n-1)\, h[v(T \cup i) - v(T)] \\
&= H_{\mathrm{M}}(v_{-k}),
\end{aligned}
$$

which completes the proof. ∎

4.6   Strict increasing monotonicity

For any weight vector ω and any $\lambda \in [0,1]$, we define the weight vector ω_λ by

$$
\omega_\lambda := \omega + \lambda\,(\omega^* - \omega),
$$

where $\omega^* := (1/n, \ldots, 1/n)$. Then it is easy to prove that $H(\omega_\lambda)$ is strictly increasing on the parameter λ, that is,

$$
0 \leq \lambda_1 < \lambda_2 \leq 1 \quad \Rightarrow \quad H(\omega_{\lambda_1}) < H(\omega_{\lambda_2}).
$$

In other terms, the Shannon entropy $H(\omega)$ strictly increases whenever ω moves closer to ω^*.

We now prove that a similar property holds for the entropy H_{M}. Let v^* be the capacity on N defined by

$$
v^*(S) = s/n \qquad (S \subseteq N).
$$

For any capacity v on N, $v \neq v^*$, and any $\lambda \in [0,1]$, we define the capacity v_λ on N by

$$
v_\lambda := v + \lambda\,(v^* - v).
$$

Proposition 4.6 *Let v be a capacity on N, with $v \neq v^*$. Then*

$$
0 \leq \lambda_1 < \lambda_2 \leq 1 \quad \Rightarrow \quad H_{\mathrm{M}}(v_{\lambda_1}) < H_{\mathrm{M}}(v_{\lambda_2}).
$$

Proof. Let us prove that

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} H_{\mathrm{M}}(v_\lambda) > 0 \qquad (\lambda \in \,]0,1[).$$

For any $\lambda \in \,]0,1[$, we have

$$H_{\mathrm{M}}(v_\lambda) = \sum_{i=1}^{n} \sum_{T \subseteq N \setminus i} \gamma_t(n)\, h[\delta_T^i + \lambda\,(1/n - \delta_T^i)],$$

where $\delta_T^i := v(T \cup i) - v(T)$. Hence,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\lambda} H_{\mathrm{M}}(v_\lambda) &= \sum_{i=1}^{n} \sum_{T \subseteq N \setminus i} \gamma_t(n)\,(\delta_T^i - 1/n)(1 + \ln[\delta_T^i + \lambda\,(1/n - \delta_T^i)]) \\
&= \sum_{i=1}^{n} \sum_{T \subseteq N \setminus i} \gamma_t(n)\,(\delta_T^i - 1/n) \ln[\delta_T^i + \lambda\,(1/n - \delta_T^i)].
\end{aligned}$$

Now, for any fixed $i \in N$ and $T \subseteq N \setminus i$,

- if $\delta_T^i > 1/n$ then

$$\delta_T^i + \lambda\,(1/n - \delta_T^i) \in \,]1/n, \delta_T^i[$$

  and hence,

$$\ln[\delta_T^i + \lambda\,(1/n - \delta_T^i)] \in \,]\ln(1/n), \ln\delta_T^i[$$

- if $\delta_T^i < 1/n$ then

$$\delta_T^i + \lambda\,(1/n - \delta_T^i) \in \,]\delta_T^i, 1/n[$$

  and hence,

$$\ln[\delta_T^i + \lambda\,(1/n - \delta_T^i)] \in \,]\ln\delta_T^i, \ln(1/n)[$$

We then have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\lambda} H_{\mathrm{M}}(v_\lambda) &= \sum_{\substack{i \in N \\ T \subseteq N \setminus i \\ \delta_T^i > 1/n}} \gamma_t(n)\,(\delta_T^i - 1/n) \ln[\delta_T^i + \lambda\,(1/n - \delta_T^i)] \\
&\quad - \sum_{\substack{i \in N \\ T \subseteq N \setminus i \\ \delta_T^i < 1/n}} \gamma_t(n)\,(1/n - \delta_T^i) \ln[\delta_T^i + \lambda\,(1/n - \delta_T^i)] \\
&> \sum_{\substack{i \in N \\ T \subseteq N \setminus i \\ \delta_T^i > 1/n}} \gamma_t(n)\,(\delta_T^i - 1/n) \ln(1/n) - \sum_{\substack{i \in N \\ T \subseteq N \setminus i \\ \delta_T^i < 1/n}} \gamma_t(n)\,(1/n - \delta_T^i) \ln(1/n) \\
&= \ln(1/n) \sum_{i=1}^{n} \sum_{T \subseteq N \setminus i} \gamma_t(n)\,(\delta_T^i - 1/n) \\
&= 0,
\end{aligned}$$

which proves the result. ∎

5   An illustrative example

In this section we give an example, borrowed from Grabisch [5]. Let us consider the problem of evaluating students in a high school with respect to three subjects: mathematics (M), physics (P), and literature (L). Usually, this is done by a simple weighted arithmetic mean, whose weights are the coefficients of importance of the different subjects. Suppose that the school is more scientifically than literary oriented, so that weights could be for example $\omega_M = 3/8$, $\omega_P = 3/8$, and $\omega_L = 2/8$, respectively. Then the weighted arithmetic mean will give the following results for three students a, b, and c (marks are given on a scale from 0 to 20):

Student	M	P	L	Global evaluation (weighted arithmetic mean)
a	18	16	10	15.25
b	10	12	18	12.75
c	14	15	15	14.625

The total weight is well distributed over the three subjects since we have

$$\frac{1}{\ln n}\, H(\omega) = 0.985\,.$$

Now, if the school wants to favor well equilibrated students without weak points then student c should be considered better than student a, who has a severe weakness in literature. Unfortunately, no weight vector $(\omega_M, \omega_P, \omega_L)$ satisfying $\omega_M = \omega_P > \omega_L$ is able to favor student c. Indeed, we have:

$$\mathrm{WAM}_\omega(14, 15, 15) > \mathrm{WAM}_\omega(18, 16, 10) \quad \Leftrightarrow \quad \omega_L > \omega_M.$$

The reason of this problem is that too much importance is given to mathematics and physics, which present some overlap effect since, usually, students good at mathematics are also good at physics (and vice versa), so that the evaluation is overestimated (resp. underestimated) for students good (resp. bad) at mathematics and/or physics. This problem can be overcome by using a suitable capacity v and the Choquet integral as follows.

- Since scientific subjects are more important than literature, the following weights can be put on subjects taken individually: $v(M) = v(P) = 0.45$ and $v(L) = 0.3$. Note that the initial ratio of weights $(3, 3, 2)$ is kept unchanged.

- Since mathematics and physics overlap, the weight attributed to the pair $\{M, P\}$ should be less than the sum of the weights of mathematics and physics: $v(MP) = 0.5$.

- Since students equally good at scientific subjects and literature must be favored, the weight attributed to the pair $\{L, M\}$ should be greater than the sum of individual weights (the same for physics and literature): $v(ML) = v(PL) = 0.9$.

- $v(\emptyset) = 0$ and $v(MPL) = 1$ by definition.

Applying Choquet integral with the above capacity leads to the following new global evaluations:

Student	M	P	L	Global evaluation (Choquet integral)
a	18	16	10	13.9
b	10	12	18	13.6
c	14	15	15	14.6

The expected result is then obtained. Also remark that student b has still the lowest rank, as requested by the scientific tendency of this high school.

Finally, we have

$$\frac{1}{\ln n}\, H_{\mathrm{M}}(v) = 0.820\,,$$

which shows that the total weight is still rather well distributed.

Now, in the spirit of the maximum entropy techniques, we might search for a capacity that both satisfies some linear constraints and maximizes the dispersion. Since H_{M} is clearly a strictly concave function, the solution of such an optimization problem is unique. Thus, a possible formulation would be the following:

$$\text{maximize} \quad H_{\mathrm{M}}(v)$$

$$\text{subject to} \quad \begin{cases} v(\mathrm{M}) = v(\mathrm{P}) \geq v(\mathrm{L}) + \varepsilon \\ v(\mathrm{MP}) \leq v(\mathrm{M}) + v(\mathrm{P}) - \varepsilon \\ \quad\quad \vdots \end{cases}$$

where $\varepsilon > 0$ is a threshold fixed beforehand.

For a detailed development about the constraints of such a problem, see Marichal and Roubens [13].

6 Probabilistic interpretation

In this final section, we give a probabilistic interpretation of $H_{\mathrm{M}}(v)$. This interpretation is similar to that of the Shapley value $\phi_i(v)$ as a mathematical expectation of the marginal contribution $v(T \cup i) - v(T)$ of element i when joining the subset $T \subseteq N \setminus i$.

First of all, through the usual identification of subsets $S \subseteq N$ with elements of $\{0,1\}^n$, any Choquet capacity v on N can be viewed as a pseudo-Boolean function $f_v : \{0,1\}^n \to [0,1]$ that is increasing in each variable and such that

$$f_v(0,\ldots,0) = 0 \quad \text{and} \quad f_v(1,\ldots,1) = 1\,.$$

The correspondence is written

$$f_v(x) = \sum_{T \subseteq N} v(T) \prod_{i \in T} x_i \prod_{i \notin T} (1 - x_i) \qquad (x \in \{0,1\}^n)\,,$$

and

$$v(S) = f_v(e_S) \qquad (S \subseteq N)\,,$$

where e_S denotes the characteristic vector of S in $\{0,1\}^n$.

Define also the first derivative of f_v with respect to $i \in N$ by

$$\Delta_i\, f_v(x) := f_v(x \,|\, x_i = 1) - f_v(x \,|\, x_i = 0) \qquad (x \in \{0,1\}^n)$$

and the gradient of f_v by

$$(\operatorname{grad} f_v)(x) := (\Delta_1 f_v(x), \ldots, \Delta_n f_v(x)) \qquad (x \in \{0,1\}^n).$$

Now, consider $\{0,1\}^n$ as a probability space with the following distribution:

$$p(x) = \frac{1}{n+1} \binom{n}{\sum_i x_i}^{-1} \qquad (x \in \{0,1\}^n).$$

This is a well-defined distribution since we have $p(x) \geq 0$ for any $x \in \{0,1\}^n$ and

$$
\begin{aligned}
\sum_{x \in \{0,1\}^n} p(x) &= \frac{1}{n+1} \sum_{T \subseteq N} \binom{n}{t}^{-1} \\
&= \frac{1}{n+1} \sum_{t=0}^{n} \frac{1}{\binom{n}{t}} \sum_{\substack{T \subseteq N \\ |T|=t}} 1 \\
&= 1.
\end{aligned}
$$

We now prove that, for any capacity v on N, $H_{\mathrm{M}}(v)$ is the mathematical expectation of the entropy of $\operatorname{grad} f_v$ for the distribution above. Notice that this entropy is well-defined even if the components of $\operatorname{grad} f_v$ do not sum up to one.

Proposition 6.1 *For any capacity v on N, we have*

$$H_{\mathrm{M}}(v) = E[H(\operatorname{grad} f_v)].$$

Proof. We have successively

$$
\begin{aligned}
E[H(\operatorname{grad} f_v)] &= \frac{1}{n+1} \sum_{x \in \{0,1\}^n} \binom{n}{\sum_j x_j}^{-1} \sum_{i=1}^{n} h[\Delta_i f_v(x)] \\
&= \sum_{i=1}^{n} \frac{1}{n+1} \underbrace{\sum_{T \subseteq N} \binom{n}{t}^{-1} h[\Delta_i f_v(e_T)]}_{(*)} .
\end{aligned}
$$

Now, partitioning T into $K \subseteq N \setminus i$ and $L \subseteq \{i\}$, the expression $(*)$ becomes

$$
\begin{aligned}
\sum_{K \subseteq N \setminus i} \sum_{L \subseteq \{i\}} \binom{n}{k+l}^{-1} h[\underbrace{\Delta_i f_v(e_{K \cup L})}_{=\Delta_i f_v(e_K)}] &= \sum_{K \subseteq N \setminus i} \left[\binom{n}{k}^{-1} + \binom{n}{k+1}^{-1} \right] h[\Delta_i f_v(e_K)] \\
&= (n+1) \sum_{K \subseteq N \setminus i} \gamma_k(n) \, h[\Delta_i f_v(e_K)] \\
&= (n+1) \sum_{K \subseteq N \setminus i} \gamma_k(n) \, h[v(K \cup i) - v(K)],
\end{aligned}
$$

which completes the proof. ∎

Note that a proof similar to that of Proposition 6.1 gives the following interpretation of the Shapley value (see also [6]):

$$\phi_i(v) = E[\Delta_i f_v] \qquad (i \in N).$$

7   Conclusions

We have proposed a measure of entropy for discrete Choquet capacities. In the framework of aggregation of numerical values by the Choquet integral, this entropy is interpreted as the degree to which one uses the arguments to calculate the aggregated value.

This new entropy-like function generalizes the Shannon entropy in a rather natural way. However, there exist other frameworks and other ways of defining an entropy for Choquet capacities; see e.g. [21, 22]. On this issue, researches are in progress and a comparative study of two proposals is presented in [14].

References

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

[2] D. Dubois and A. Ramer, Extremal properties of belief measures in the theory of evidence, *Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems* **1** (1993) 57–68.

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

[4] M. Grabisch, On equivalence classes of fuzzy connectives : the case of fuzzy integrals, *IEEE Trans. Fuzzy Systems* **3** (1) (1995) 96–109.

[5] M. Grabisch, The application of fuzzy integrals in multicriteria decision making, *European Journal of Operational Research* **89** (1996) 445-456.

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

[7] G.J. Klir and A. Ramer, Uncertainty in the Dempster-Shafer theory: a critical re-examination, *Int. J. General Systems* **18** (1990) 155–166.

[8] G.J. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications* (Prentice Hall PTR, Upper Saddle River, NJ, 1995).

[9] J.-L. Marichal, *Aggregation operators for multicriteria decision aid*, Ph.D. thesis, Institute of Mathematics, University of Liège, Liège, Belgium, 1998.

[10] 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.

[11] J.-L. Marichal, Behavioral analysis of aggregation in multicriteria decision aid, in: J. Fodor, B. De Baets, and P. Perny (eds.), *Preferences and Decisions under Incomplete Knowledge*. Series: Studies in Fuzziness and Soft Computing Vol. 51 (Physica Verlag, Heidelberg, 2000) 153–178.

[12] J.-L. Marichal, An axiomatic approach of the discrete Sugeno integral as a tool to aggregate interacting criteria in a qualitative framework, *IEEE Transactions on Fuzzy Systems*, to appear.

[13] J.-L. Marichal and M. Roubens, Determination of weights of interacting criteria from a reference set, *European Journal of Operational Research* **124** (3) (2000) 641–650.

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

[15] T. Murofushi and M. Sugeno, A theory of fuzzy measures. Representation, the Choquet integral and null sets, *J. of Math. Anal. and Appl.* **159** (2) (1991) 532–549.

[16] C.E. Shannon, A mathematical theory of communication, *Bell System Tech. J.* **27** (1948) 379–423, 623–656.

[17] C.E. Shannon and W. Weaver, A mathematical theory of communication (University of Illinois Press, Urbana, 1949).

[18] L.S. Shapley, A value for n-person games, in: H.W. Kuhn and A.W. Tucker (eds.), *Contributions to the Theory of Games*, Vol. II, Annals of Mathematics Studies, 28, (Princeton University Press, Princeton, NJ, 1953) 307–317.

[19] R.R. Yager, Entropy and specificity in a mathematical theory of evidence, *Int. J. Gen. Systems* **9** (1983) 249–260.

[20] R.R. Yager, On ordered weighted averaging aggregation operators in multicriteria decision making, *IEEE Trans. on Systems, Man and Cybernetics* **18** (1988) 183–190.

[21] R.R. Yager, A class of fuzzy measures generated from a Dempster-Shafer belief structure, *Int. J. of Intelligent Systems* **14** (12) (1999) 1239–1247.

[22] R.R. Yager, On the entropy of fuzzy measures, *IEEE Transactions on Fuzzy Systems* **8** (4) (2000) 453–461.