

# Entropy of Discrete Fuzzy Measures

JEAN-LUC MARICHAL\*

MARC ROUBENS

Institute of Mathematics, University of Liège  
Grande Traverse 12 - B37, Sart Tilman, B-4000 Liège, Belgium.

{j1.marichal,m.roubens}@ulg.ac.be

Revised version, September 4, 2000

## Abstract

The concept of entropy of a discrete fuzzy measure has been recently introduced in two different ways. A first definition was proposed by Marichal in the aggregation framework, and a second one by Yager in the framework of uncertain variables. We present a comparative study between these two proposals and point out their properties. We also propose a definition for the entropy of an ordinal fuzzy measure, that is, a fuzzy measure taking its values in an ordinal scale in the sense of measurement theory.

**Keywords:** entropy, fuzzy measure, aggregation, uncertain variable, Choquet and Sugeno integrals, ordinal fuzzy measure.

## 1 Introduction

Consider a random experiment having  $n$  possible outcomes with probabilities  $p_1, \dots, p_n$ . In order to appraise the average uncertainty associated with the prediction of these outcomes, or equivalently, the amount of information received from the knowledge of which of these outcomes occurred, several information measures have been introduced. Among them, the best known is probably the Shannon entropy (see [17, 18]), defined by

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

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

It was proved in numerous ways, from several well-justified axiomatic characterizations, that this function is the only sensible measure of uncertainty in probability theory, see e.g. [1, 3].

It is known that probability is not the only type of uncertainty used to represent partial knowledge. A large class of types of partial information can be expressed by means of

---

\*Scientific Research Worker of FNRS (National Scientific Research Fund, Belgium).

fuzzy measures, which were introduced by Sugeno [20, 21] to model the subjective aspect of uncertainty. Formally a (discrete) fuzzy measure on a finite space  $N = \{\theta_1, \dots, \theta_n\}$  is a set function  $\mu : 2^N \rightarrow [0, 1]$  such that  $\mu(\emptyset) = 0$ ,  $\mu(N) = 1$ , and  $\mu(S) \leq \mu(T)$  whenever  $S \subseteq T \subseteq N$ . Since this concept is clearly a generalization of that of probability measure, the following natural question arises: how can we appraise the average uncertainty associated with a fuzzy measure in the spirit of Shannon entropy?

For particular fuzzy measures, such as belief and plausibility measures, two entropy-like measures were proposed in evidence theory in the early 1980s, namely the measure of dissonance and the measure of confusion, see e.g. [2, 8, 9, 22]. For general fuzzy measures it seems that no definition of entropy was given until 1999 when two proposals were introduced successively by Marichal [10, 13, 15] and Yager [25, 26]. Although they have yet to be axiomatically characterized, these proposals satisfy properties considered as requisites for defining an entropy. Particularly, they collapse into the Shannon entropy as soon as the fuzzy measure is additive.

In this paper we present a comparative analysis of these new definitions (see Section 3). This analysis is mainly done on the basis of extensions of some classical properties fulfilled by the Shannon entropy. We also propose a definition for the entropy of discrete fuzzy measures that range in a finite ordinal scale (see Section 4).

Throughout the paper, we will use the notation

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

and the function

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

## 2 Entropy of discrete fuzzy measures

Two definitions of entropy for discrete fuzzy measures have been recently proposed by Marichal [10] and Yager [25]. These definitions are very similar but have been introduced independently and within completely different frameworks. The first one gives the degree to which numerical values are really used when aggregated by a Choquet integral. The second one measures the amount of information in a fuzzy measure when it is being used to represent the knowledge about an uncertain variable. In this section we briefly present these two frameworks and the corresponding proposed entropies. In the next section we discuss the relevance of each entropy to its own framework.

### 2.1 Entropy in the aggregation framework

Suppose that  $N = \{\theta_1, \dots, \theta_n\}$  represents a set of criteria (or attributes) in a multicriteria decision making problem and consider a fuzzy measure  $\mu$  on  $N$ . For any  $S \subseteq N$ ,  $\mu(S)$  can be interpreted as the weight or the degree of importance of the combination  $S$  of criteria, or better, its power to make the decision alone (without the remaining criteria). Thus, in addition to the usual weights on criteria taken separately, weights on any combination of criteria are also defined. Monotonicity of  $\mu$  then means that adding a new element to a combination cannot decrease its importance. Obviously  $\mu(N)$  has the maximal value, being one by convention.

Now, suppose that  $x_1, \dots, x_n \in \mathbb{R}$  represent quantitative evaluations of an object with respect to criteria  $\theta_1, \dots, \theta_n$ , respectively. We assume that these evaluations are comparable, that is, defined on the same measurement scale. A global evaluation (average value) of this object can be calculated by means of the *Choquet integral* with respect to  $\mu$ .

Formally the Choquet integral of  $x \in \mathbb{R}^n$  with respect to a fuzzy measure  $\mu$  on  $N$  is defined by

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

where  $(\cdot)$  is a permutation on indices such that  $x_{(1)} \leq \dots \leq x_{(n)}$ . Also  $A_{(i)} := \{\theta_{(i)}, \dots, \theta_{(n)}\}$ , and  $A_{(n+1)} := \emptyset$ . For more details, see e.g. [5, 11] and the references therein.

It would be interesting to appraise the extent to which the arguments  $x_1, \dots, x_n$  are really used in the aggregation by the Choquet integral. This extent, which depends only on the importance  $\mu$  of criteria, can be measured by the following entropy-like function, proposed and justified by Marichal [10, 13, 15]. We call it “lower entropy”. This name will be justified in Section 3.7.

**Definition 2.1** *The lower entropy of a fuzzy measure  $\mu$  on  $N$  is defined by*

$$H_l(\mu) := \sum_{i=1}^n \sum_{T \subseteq N \setminus \{\theta_i\}} \gamma_{|T|}(n) h[\mu(T \cup \{\theta_i\}) - \mu(T)].$$

## 2.2 Entropy in the framework of uncertain variables

Consider a variable  $V$  whose the exact value, which lies in the space  $N = \{\theta_1, \dots, \theta_n\}$ , is not completely known. In many situations, the best we can do is to formulate our knowledge about  $V$  by means of a fuzzy measure on  $N$ . For each subset  $S \subseteq N$  of values,  $\mu(S)$  represents a measure associated with our belief (or the confidence we have) that the value of  $V$  is contained in the subset  $S$ . Here monotonicity of  $\mu$  means that we cannot be less confident that  $V$  lies in a smaller set than a larger one.

Now, consider the *Shapley value* [19] of  $\mu$ , i.e., the vector

$$\phi(\mu) = (\phi_1(\mu), \dots, \phi_n(\mu)) \in [0, 1]^n$$

whose  $i$ -th component, called Shapley index of  $\theta_i$ , is defined by:

$$\phi_i(\mu) := \sum_{T \subseteq N \setminus \{\theta_i\}} \gamma_{|T|}(n) [\mu(T \cup \{\theta_i\}) - \mu(T)].$$

This index, which is a fundamental concept in game theory, can be interpreted as the average marginal contribution of element  $\theta_i$  to a combination not containing it. In the present framework, it clearly measures the average increment in confidence obtained by adding  $\theta_i$  to a subset not containing it. Thus,  $\phi_i(\mu)$  somehow reflects our overall belief that the value of  $V$  is  $\theta_i$ .

It can be easily proved that the indices  $\phi_i(\mu)$  always sum up to one, so that the Shapley value of any fuzzy measure on  $N$  is a probability measure on  $N$ . From this observation, Yager [25] proposed to evaluate the uncertainty associated with the variable  $V$  by taking the Shannon entropy of the Shapley value of  $\mu$ . This leads to the “upper entropy” whose name will be justified in Section 3.7.

**Definition 2.2** *The upper entropy of a fuzzy measure  $\mu$  on  $N$  is defined by*

$$H_u(\mu) := H(\phi(\mu)),$$

that is,

$$H_u(\mu) := \sum_{i=1}^n h \left[ \sum_{T \subseteq N \setminus \{\theta_i\}} \gamma_{|T|}(n) [\mu(T \cup \{\theta_i\}) - \mu(T)] \right].$$

### 3 Comparative study of the entropies

Thus defined, the entropies  $H_l$  and  $H_u$  seem very similar. However, although they fulfill several properties required for an entropy, they also present very different behaviors.

This section is devoted to a comparative analysis of these two entropies. This analysis is mainly based on extensions of some well-known properties of the Shannon entropy such as symmetry, decisivity, expansibility (see e.g. [1, 3]). It will lead to the conclusion that each entropy is highly relevant for the framework in which it has been defined and not for the other.

Before going on we briefly recall some particular types of fuzzy measure, which will enable us to point out the behavior of each entropy. A fuzzy measure  $\mu$  on  $N$  is

- *additive* if, for any  $S, T \subseteq N$ , we have

$$S \cap T = \emptyset \quad \Rightarrow \quad \mu(S \cup T) = \mu(S) + \mu(T),$$

- *cardinality-based* if, for any  $S, T \subseteq N$ , we have

$$|S| = |T| \quad \Rightarrow \quad \mu(S) = \mu(T),$$

- *binary-valued* if, for any  $T \subseteq N$ , we have  $\mu(T) \in \{0, 1\}$ ,
- *the Dirac measure associated with  $\theta_k \in N$*  if, for any  $T \subseteq N$ , we have

$$\mu(T) = \begin{cases} 1, & \text{if } T \ni \theta_k, \\ 0, & \text{otherwise.} \end{cases}$$

#### 3.1 Connections with the Shannon entropy

When the fuzzy measure  $\mu$  is additive, we clearly have

$$\mu(T \cup \{\theta_i\}) - \mu(T) = \mu(\{\theta_i\})$$

for all  $\theta_i \in N$  and all  $T \subseteq N \setminus \{\theta_i\}$ . This leads to the following immediate result, which shows that the entropies  $H_l$  and  $H_u$  are generalizations of the Shannon entropy.

**Proposition 3.1** *For any additive measure  $\mu$  on  $N$ , we have*

$$H_l(\mu) = H_u(\mu) = H(\omega),$$

where  $\omega = (\omega_1, \dots, \omega_n)$  is the probability distribution defined by

$$\omega_i := \mu(\{\theta_i\}) \quad (i = 1, \dots, n).$$

Let us interpret this result in the aggregation framework. When the fuzzy measure  $\mu$  is additive, the Choquet integral  $\mathcal{C}_\mu$  reduces to the weighted arithmetic mean

$$\text{WAM}_\omega(x) = \sum_{i=1}^n \omega_i x_i \quad (x \in \mathbb{R}^n),$$

where  $\omega_i = \mu(\{\theta_i\})$  for all  $i = 1, \dots, n$ . In such an aggregation process, it is natural to evaluate the degree of use of the arguments  $x_1, \dots, x_n$  by means of the Shannon entropy  $H(\omega)$ , which represents a measure of dispersion associated with the weight vector  $\omega$ .

Let us consider the second framework. When the fuzzy measure  $\mu$  is additive, our knowledge about the unknown variable  $V$  is in the form of probabilistic uncertainty. For any  $S \subseteq N$ ,  $\mu(S)$  is referred to as the measure of probability of  $S$ . In this context the Shannon entropy of this probability distribution naturally provides a measure of uncertainty associated with the variable  $V$ .

### 3.2 The symmetry property

It is obvious that the Shannon entropy  $H$  is symmetric in the sense that permuting the elements of  $N$  has no effect on the entropy. This symmetry property is actually also fulfilled by  $H_l$  and  $H_u$ .

Formally, for any permutation  $\pi$  on  $\{1, \dots, n\}$ , we denote by  $\pi\mu$  the fuzzy measure on  $N$  defined by  $\pi\mu(\pi(S)) = \mu(S)$  for all  $S \subseteq N$ , where  $\pi(S) := \{\theta_{\pi(i)} \mid \theta_i \in S\}$ . We then have the following result.

**Proposition 3.2** *For any fuzzy measure  $\mu$  on  $N$ , we have*

$$H_l(\pi\mu) = H_l(\mu) \quad \text{and} \quad H_u(\pi\mu) = H_u(\mu).$$

**Proof.** The first equality was proved in [13]. The second one follows from the fact that  $\phi_i(\pi\mu) = \phi_{\pi^{-1}(i)}(\mu)$  for all  $i = 1, \dots, n$ , which is very easy to verify. ■

This accords with the interpretation of the entropy in the aggregation framework. Indeed, one can easily show that

$$\mathcal{C}_\mu(x_{\pi(1)}, \dots, x_{\pi(n)}) = \mathcal{C}_{\pi\mu}(x_1, \dots, x_n).$$

Thus, permuting the arguments of the Choquet integral has no effect on the degree to which one uses these arguments.

We also have an immediate interpretation in the second framework. The uncertainty associated with the variable  $V$  is independent of any permutation of elements of  $N$ .

### 3.3 The expansibility property

The classical expansibility property for the Shannon entropy says that suppressing an outcome with zero probability does not change the uncertainty of the outcome of an experiment (see e.g. [1]):

$$H(p_1, \dots, p_{n-1}, 0) = H(p_1, \dots, p_{n-1}).$$

This property can be extended to the framework of fuzzy measures in the following way. Let  $\mu$  be a fuzzy measure on  $N$  and let  $\theta_k \in N$  be a *null* element for  $\mu$ , that is such that  $\mu(T \cup \{\theta_k\}) = \mu(T)$  for all  $T \subseteq N \setminus \{\theta_k\}$ . Denote also by  $\mu_{-\theta_k}$  the restriction of  $\mu$  to  $N \setminus \{\theta_k\}$ . We then have the following result.

**Proposition 3.3** *Let  $\mu$  be a fuzzy measure on  $N$ . If  $\theta_k \in N$  is a null element for  $\mu$  then*

$$H_l(\mu) = H_l(\mu_{-\theta_k}) \quad \text{and} \quad H_u(\mu) = H_u(\mu_{-\theta_k}).$$

**Proof.** The first equality was proved in [13]. The second one follows from the expansibility property of the Shannon entropy and the fact that  $\phi_k(\mu) = 0$ . ■

This is a very natural property in the aggregation framework. Since element  $\theta_k$  does not contribute in the decision problem, it can be omitted without changing the result. Indeed, one can easily show [16] that

$$\mathcal{C}_\mu(x_1, \dots, x_n) = \mathcal{C}_{\mu_{-\theta_k}}(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n)$$

whenever  $\theta_k$  is a null element for  $\mu$ .

In the second framework, this property clearly shows that suppressing a value  $\theta_k$  from  $N$  for which we have no confidence that  $V = \theta_k$  does not change the uncertainty associated to  $V$ .

### 3.4 The decisivity property

For any probability measure  $p$  on  $N$ , we clearly have  $H(p) \geq 0$ . Now, the decisivity property for the Shannon entropy says that there is no uncertainty in an experiment in which one outcome has probability one (see e.g. [1]):

$$H(1, 0, \dots, 0) = \dots = H(0, \dots, 0, 1) = 0.$$

More precisely,  $H(p)$  reaches its minimal value ( $= 0$ ) if and only if  $p$  is a Dirac measure on  $N$ .

Concerning the entropies  $H_l$  and  $H_u$ , we observe two different behaviors, as the following result shows.

**Proposition 3.4** *Let  $\mu$  be a fuzzy measure on  $N$ . Then*

$$H_l(\mu) \geq 0 \quad \text{and} \quad H_u(\mu) \geq 0.$$

Moreover,

- i)  $H_l(\mu) = 0$  if and only if  $\mu$  is a binary-valued fuzzy measure,*
- ii)  $H_u(\mu) = 0$  if and only if  $\mu$  is a Dirac measure.*

**Proof.** The first inequality and property *i)* were proved in [10, 13]. The second inequality and property *ii)* follow from the above property of the Shannon entropy, that is,  $H_u(\mu) = 0$  if and only if the Shapley value  $\phi(\mu)$  is a Dirac measure, or equivalently, if and only if  $\mu$  itself is a Dirac measure. ■

Property *i)* is quite relevant for the aggregation framework. Indeed, it can be shown (cf. [12, Theorem 5.1]) that  $\mu$  is a binary-valued fuzzy measure if and only if

$$\mathcal{C}_\mu(x) \in \{x_1, \dots, x_n\} \quad (x \in \mathbb{R}^n).$$

In other terms,  $H_l(\mu)$  is minimum ( $= 0$ ) if and only if only one argument is really used in the aggregation. This is a fundamental condition.

In the second framework, property *ii*) corresponds to the case where the variable  $V$  is known exactly. Indeed, suppose that we know that  $V = \theta_k$ . In this case, the measure of uncertainty associated with  $V$  is given by

$$\mu(S) = \begin{cases} 1 & \text{if } \theta_k \in S, \\ 0 & \text{if } \theta_k \notin S, \end{cases}$$

which is the Dirac measure associated with  $\theta_k$ . Thus, the complete certainty of the value of  $V$  is characterized by a Dirac measure.

We thus see that each of the entropies  $H_l$  and  $H_u$  is suitable for the framework in which it has been defined and not for the other.

### 3.5 The maximality property

The maximality property for the Shannon entropy says that the uncertainty of the outcome of an experiment is maximum when all outcomes have equal probabilities. Formally, for any probability measure  $p$  on  $N$ ,

$$H(p) \leq H(1/n, \dots, 1/n) = \ln n.$$

More precisely,  $H(p)$  reaches its maximal value ( $= \ln n$ ) if and only if  $p$  is the uniform distribution on  $N$ .

Concerning the entropies  $H_l$  and  $H_u$ , we have the following result.

**Proposition 3.5** *Let  $\mu$  be a fuzzy measure on  $N$ . Then*

$$H_l(\mu) \leq \ln n \quad \text{and} \quad H_u(\mu) \leq \ln n.$$

Moreover,

- i)*  $H_l(\mu) = \ln n$  if and only if  $\mu(S) = |S|/n$  for all  $S \subseteq N$ ,
- ii)*  $H_u(\mu) = \ln n$  if and only if  $\phi_i(\mu) = 1/n$  for all  $i = 1, \dots, n$ .

**Proof.** The first inequality and property *i*) were proved in [10, 13]. The second inequality and property *ii*) follow from the above property of the Shannon entropy. ■

The interpretation of these properties is very simple. On the one hand,  $H_l(\mu)$  is maximum if and only if  $\mathcal{C}_\mu$  identifies with the arithmetic mean, which is the weighted arithmetic mean whose dispersion associated with the weight vector is maximum.

On the other hand, the uncertainty associated with  $V$  is maximum if and only if the overall belief that  $V$  has the value  $\theta_i$  is  $1/n$  for all  $i = 1, \dots, n$ .

As stated by Yager [25], the problem then arises of identifying all the fuzzy measures  $\mu$  such that  $\phi_i(\mu) = 1/n$  for all  $i = 1, \dots, n$ . It is easy to show that all the cardinality-based fuzzy measures belong to this class. However, as Yager [25] pointed out, there exist others. We give below a complete description of this class. For this purpose, we consider the well-known sequence  $(B_n)_{n \in \mathbb{N}}$  of Bernoulli numbers defined recursively by

$$\begin{cases} B_0 = 1, \\ \sum_{k=0}^n \binom{n+1}{k} B_k = 0, \quad (n \in \mathbb{N} \setminus \{0\}). \end{cases}$$

We then have the following result.

**Proposition 3.6** *All the fuzzy measures  $\mu$  on  $N$  such that  $\phi_i(\mu) = 1/n$  ( $i = 1, \dots, n$ ), are of the form*

$$\mu(S) = \frac{|S|}{n} + \sum_{\substack{T \subseteq N \\ |T| \geq 2}} \left[ \sum_{j=1}^{|T \cap S|} \binom{|T \cap S|}{j} B_{|T|-j} \right] c(T) \quad (S \subseteq N; 0 < |S| < n),$$

where the coefficients  $c(T) \in \mathbb{R}$ , with  $|T| \geq 2$ , are free up to the constraints

$$\frac{1}{n} + \sum_{\substack{T \ni \theta_i \\ |T| \geq 2}} \left[ \sum_{j=0}^{|T \cap S|} \binom{|T \cap S|}{j} B_{|T|-j-1} \right] c(T) \geq 0 \quad (\theta_i \in N; S \subseteq N \setminus \{\theta_i\}).$$

**Proof.** The result is a particular case of Theorem 5 in [6], which was revisited in [10, Theorem 6.3.4]. ■

Before closing this subsection, we present results on the monotonicity of the entropies  $H_l$  and  $H_u$ .

For any probability measure  $p$  on  $N$  and any  $\lambda \in [0, 1]$ , we define the probability measure  $p_\lambda$  on  $N$  by

$$p_\lambda := p + \lambda(p^* - p),$$

where  $p^* := (1/n, \dots, 1/n)$  denotes the uniform distribution on  $N$ . Then it is easy to prove that  $H(p_\lambda)$  is strictly increasing on the parameter  $\lambda$ , that is,

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

Thus, the Shannon entropy  $H(p)$  strictly increases whenever  $p$  moves closer to  $p^*$ . Note that this property does not characterize the entropy. Indeed, the function  $G$  defined over the set of probability measures on  $N$  by

$$G(p) := \sum_{i=1}^n p_i(1 - p_i)$$

also fulfills this property.

We now prove that similar properties hold for the entropies  $H_l$  and  $H_u$ . Let  $\mu^*$  be the fuzzy measure on  $N$  defined by

$$\mu^*(S) = |S|/n \quad (S \subseteq N).$$

For any fuzzy measure  $\mu$  on  $N$ ,  $\mu \neq \mu^*$ , and any  $\lambda \in [0, 1]$ , we define the fuzzy measure  $\mu_\lambda$  on  $N$  by

$$\mu_\lambda := \mu + \lambda(\mu^* - \mu).$$

We then have the following result.

**Proposition 3.7** *Let  $\mu$  be a fuzzy measure on  $N$ , with  $\mu \neq \mu^*$ , and let  $0 \leq \lambda_1 < \lambda_2 \leq 1$ .*

*Then*

- i)  $H_l(\mu_{\lambda_1}) < H_l(\mu_{\lambda_2})$ ,*
- ii)  $H_u(\mu_{\lambda_1}) \leq H_u(\mu_{\lambda_2})$ ,*
- iii)  $H_u(\mu_{\lambda_1}) = H_u(\mu_{\lambda_2})$  if and only if  $\phi_i(\mu) = 1/n$  for all  $i = 1, \dots, n$ .*

**Proof.** Property *i*) was proved in [13]. Let us prove *ii*) and *iii*).

For any  $\lambda \in ]0, 1[$  and any  $i = 1, \dots, n$ , we have, since  $\phi_i$  is a linear function,

$$\phi_i(\mu_\lambda) = \phi_i(\mu + \lambda(\mu^* - \mu)) = \phi_i(\mu) + \lambda(1/n - \phi_i(\mu)).$$

Hence we have

$$\begin{aligned} \frac{d}{d\lambda} H_u(\mu_\lambda) &= \sum_{i=1}^n \frac{d}{d\lambda} h[\phi_i(\mu_\lambda)] \\ &= \sum_{i=1}^n (\phi_i(\mu) - 1/n)(1 + \ln[\phi_i(\mu_\lambda)]) \\ &= \sum_{i=1}^n (\phi_i(\mu) - 1/n) \ln[\phi_i(\mu_\lambda)]. \end{aligned}$$

If  $\phi_i(\mu) = 1/n$  for all  $i = 1, \dots, n$  then the latter expression is identically zero. Otherwise, for any fixed  $i$  such that  $\phi_i(\mu) \neq 1/n$ ,

- if  $\phi_i(\mu) > 1/n$  then  $\phi_i(\mu_\lambda) \in ]1/n, \phi_i(\mu)[$ , hence  $\ln[\phi_i(\mu_\lambda)] \in ]\ln(1/n), \ln \phi_i(\mu)[$ ,
- if  $\phi_i(\mu) < 1/n$  then  $\phi_i(\mu_\lambda) \in ]\phi_i(\mu), 1/n[$ , hence  $\ln[\phi_i(\mu_\lambda)] \in ]\ln \phi_i(\mu), \ln(1/n)[$ .

We then have

$$\begin{aligned} \frac{d}{d\lambda} H_u(\mu_\lambda) &= \sum_{\substack{i=1 \\ \phi_i(\mu) > 1/n}}^n (\phi_i(\mu) - 1/n) \ln[\phi_i(\mu_\lambda)] - \sum_{\substack{i=1 \\ \phi_i(\mu) < 1/n}}^n (1/n - \phi_i(\mu)) \ln[\phi_i(\mu_\lambda)] \\ &> \sum_{\substack{i=1 \\ \phi_i(\mu) > 1/n}}^n (\phi_i(\mu) - 1/n) \ln(1/n) - \sum_{\substack{i=1 \\ \phi_i(\mu) < 1/n}}^n (1/n - \phi_i(\mu)) \ln(1/n) \\ &= \ln(1/n) \sum_{i=1}^n (\phi_i(\mu) - 1/n) \\ &= 0, \end{aligned}$$

which enables us to conclude. ■

### 3.6 Case of cardinality-based fuzzy measures

When the fuzzy measure is cardinality-based, there exist coefficients

$$0 = c_0 \leq \dots \leq c_n = 1$$

such that

$$\mu(S) = c_{|S|} \quad (S \subseteq N).$$

We then have

$$\mu(T \cup \{\theta_i\}) - \mu(T) = c_{|T|+1} - c_{|T|}$$

for all  $\theta_i \in N$  and all  $T \subseteq N \setminus \{\theta_i\}$ . This leads to the following result (see [10, 13, 25]), which shows that the entropies  $H_l$  and  $H_u$  present very different behaviors.

**Proposition 3.8** For any cardinality-based fuzzy measure  $\mu$  on  $N$ , we have

$$H_l(\mu) = H(\omega) \quad \text{and} \quad H_u(\mu) = \ln n,$$

where  $\omega = (\omega_1, \dots, \omega_n)$  is the probability distribution defined from the coefficients  $c_i$  above by

$$\omega_{n-i} := c_{i+1} - c_i \quad (i = 0, \dots, n-1).$$

The second equality has already been interpreted in Section 3.5. Let us comment on the first one. It can be shown [4] that, when the fuzzy measure  $\mu$  is cardinality-based, the Choquet integrals  $\mathcal{C}_\mu$  reduces to the ordered weighted averaging (OWA) operator [23]

$$\text{OWA}_\omega(x) = \sum_{i=1}^n \omega_i x_{(i)} \quad (x \in \mathbb{R}^n),$$

where  $(\cdot)$  is a permutation on indices such that  $x_{(1)} \leq \dots \leq x_{(n)}$ , and  $\omega_{n-i} := c_{i+1} - c_i$  for all  $i = 0, \dots, n-1$ .

Consequently, we see that the degree to which one uses the values  $x_1, \dots, x_n \in \mathbb{R}$  is the same when using either the weighted arithmetic mean (cf. Section 3.1) or the ordered weighted averaging defined from the same weight vector  $\omega$ . This is actually a very natural condition since this degree should not be influenced when changing the assignment of weights to the arguments. Furthermore Yager [23] proposed explicitly to use the Shannon entropy of  $\omega$  as measure of dispersion for the OWA operators.

### 3.7 Numerical comparison between the entropies

The strict concavity of function  $h$  can be used to establish a numerical comparison between the two entropies. This comparison actually justifies the names “lower entropy” and “upper entropy”. Before presenting the result, recall the *Jensen inequality* for the strictly concave functions.

Let  $I$  be a real interval. If a function  $f : I \rightarrow \mathbb{R}$  is strictly concave on  $I$  then, for all  $t_1, \dots, t_n \in I$  and all  $\lambda_1, \dots, \lambda_n \in ]0, 1[$  such that  $\sum_{i=1}^n \lambda_i = 1$ , one has

$$\sum_{i=1}^n \lambda_i f(t_i) \leq f\left(\sum_{i=1}^n \lambda_i t_i\right).$$

The equality occurs only if  $t_1 = \dots = t_n$ .

**Proposition 3.9** For any fuzzy measure  $\mu$  on  $N$ , we have

- i)  $H_l(\mu) \leq H_u(\mu)$ ,
- ii)  $H_l(\mu) = H_u(\mu)$  if and only if  $\mu$  is additive.

**Proof.** The first part follows directly from the Jensen inequality. Let us turn to the second part. If  $\mu$  is additive then, by Proposition 3.1, we have  $H_l(\mu) = H_u(\mu)$ .

Now suppose that  $H_l(\mu) = H_u(\mu)$  for a given fuzzy measure  $\mu$  on  $N$ . Then, by the Jensen inequality, we have

$$\sum_{T \subseteq N \setminus \{\theta_i\}} \gamma_{|T|}(n) h[\mu(T \cup \{\theta_i\}) - \mu(T)] = h\left[\sum_{T \subseteq N \setminus \{\theta_i\}} \gamma_{|T|}(n) [\mu(T \cup \{\theta_i\}) - \mu(T)]\right],$$

for all  $i = 1, \dots, n$ . This implies

$$\mu(T \cup \{\theta_i\}) - \mu(T) = \mu(S \cup \{\theta_i\}) - \mu(S)$$

for all  $i = 1, \dots, n$ , and all  $S, T \subseteq N \setminus \{\theta_i\}$ . Taking  $S = \emptyset$ , we observe that  $\mu$  is additive. ■

Rather interestingly, Proposition 3.9 means that the expression

$$D(\mu) := H_u(\mu) - H_l(\mu)$$

is non-negative and vanishes if and only if  $\mu$  is additive.

### 3.8 Probabilistic interpretations

We now give a probabilistic interpretation of the entropies  $H_l$  and  $H_u$ . This interpretation shows that these two entropies have very comparable forms.

First, through the usual identification of subsets  $S \subseteq N$  with elements of  $\{0, 1\}^n$ , any fuzzy measure  $\mu$  on  $N$  can be viewed as a pseudo-Boolean function  $f_\mu : \{0, 1\}^n \rightarrow [0, 1]$  that is increasing in each variable and such that

$$f_\mu(0, \dots, 0) = 0 \quad \text{and} \quad f_\mu(1, \dots, 1) = 1.$$

The correspondence is written

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

and

$$\mu(S) = f_\mu(e_S) \quad (S \subseteq N),$$

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

Define also the first derivative of  $f_\mu$  with respect to  $i \in \{1, \dots, n\}$  by

$$\Delta_i f_\mu(x) := f_\mu(x | x_i = 1) - f_\mu(x | x_i = 0) \quad (x \in \{0, 1\}^n)$$

and the gradient of  $f_\mu$  by

$$(\text{grad } f_\mu)(x) := (\Delta_1 f_\mu(x), \dots, \Delta_n f_\mu(x)) \quad (x \in \{0, 1\}^n).$$

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

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

It was proved [13] that, for any fuzzy measure  $\mu$  on  $N$ ,  $H_l(\mu)$  is the mathematical expectation of the entropy of  $\text{grad } f_\mu$  for the distribution above, that is,

$$H_l(\mu) = \frac{1}{n+1} \sum_{x \in \{0, 1\}^n} \binom{n}{\sum_i x_i}^{-1} H[(\text{grad } f_\mu)(x)],$$

or equivalently,

$$H_l(\mu) = E[H(\text{grad } f_\mu)].$$

Notice that this entropy is well-defined even if the components of  $\text{grad } f_\mu$  do not sum up to one.

It was proved in [7] that, for any fuzzy measure  $\mu$  on  $N$ , we have

$$\phi_j(\mu) = \frac{1}{n+1} \sum_{x \in \{0,1\}^n} \binom{n}{\sum_i x_i}^{-1} \Delta_j f_\mu(x) = E[\Delta_j f_\mu]$$

for all  $j = 1, \dots, n$ . From this we deduce immediately that

$$H_u(\mu) = H\left[\frac{1}{n+1} \sum_{x \in \{0,1\}^n} \binom{n}{\sum_i x_i}^{-1} (\text{grad } f_\mu)(x)\right],$$

that is,

$$H_u(\mu) = H[E(\text{grad } f_\mu)],$$

where  $E(\text{grad } f_\mu) = (E[\Delta_1 f_\mu], \dots, E[\Delta_n f_\mu])$ .

## 4 Entropy of ordinal fuzzy measures

In this final section we propose and investigate a definition for the entropy of ordinal fuzzy measures. This definition is actually a generalization of a measure of entropy proposed by Yager [24, §4].

Formally an ordinal fuzzy measure is a monotonic set function  $\mu : 2^N \rightarrow L$  fulfilling  $\mu(\emptyset) = \inf L$  and  $\mu(N) = \sup L$ , where  $L$  is a given finite ordinal scale

$$L := \{l_1 < \dots < l_m\}.$$

For such fuzzy measures we propose the following measure of entropy.

**Definition 4.1** *The entropy of an ordinal fuzzy measure*

$$\mu : 2^N \rightarrow L = \{l_1 < \dots < l_m\}$$

is defined by

$$H_L(\mu) = l_{|R|-1},$$

where  $R := \{\mu(S) \mid S \subseteq N\}$ .

We observe that this entropy is actually a measure of the diversity of the coefficients of the fuzzy measure. The more diversified the coefficients of the fuzzy measure the greater its entropy.

Now, let us examine the properties of this entropy in the framework of ordinal aggregation. Consider an ordinal fuzzy measure  $\mu : 2^N \rightarrow L$  modeling the importance of subsets of criteria. Suppose also that  $x_1, \dots, x_n \in L$  represent qualitative evaluations of an object with respect to criteria  $\theta_1, \dots, \theta_n$ , respectively. A global evaluation of this object can be calculated by means of the *Sugeno integral* [20, 21], which is defined from the fuzzy measure  $\mu$  by

$$\mathcal{S}_\mu(x) := \bigvee_{i=1}^n [x_{(i)} \wedge \mu(A_{(i)})],$$

where  $(\cdot)$  is a permutation on indices such that  $x_{(1)} \leq \dots \leq x_{(n)}$ . Also,  $\wedge := \min$ ,  $\vee := \max$ , and  $A_{(i)} := \{\theta_{(i)}, \dots, \theta_{(n)}\}$ . For a detailed analysis of this aggregation framework, see [12].

In some sense, the entropy  $H_L$  measures on the scale  $L$  the extent to which the arguments  $x_1, \dots, x_n$  are really used in the aggregation by the Sugeno integral. This observation is in accordance with the following properties fulfilled by  $H_L$ :

- **The symmetry property.** Entropy  $H_L$  clearly satisfies the symmetry property. With the notation of Section 3.2, we have

$$H_L(\pi\mu) = H_L(\mu).$$

This accords with the following property of the Sugeno integral:

$$\mathcal{S}_\mu(x_{\pi(1)}, \dots, x_{\pi(n)}) = \mathcal{S}_{\pi\mu}(x_1, \dots, x_n).$$

Thus, permuting the arguments of the Sugeno integral has no effect on the degree to which one uses these arguments.

- **The expansibility property.** With the notation of Section 3.3, we have

$$H_L(\mu) = H_L(\mu_{-\theta_k})$$

whenever  $\theta_k$  is a null element for  $\mu$ . This means that this element can be ignored in the aggregation process, as the following equality shows:

$$\mathcal{S}_\mu(x_1, \dots, x_n) = \mathcal{S}_{\mu_{-\theta_k}}(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n)$$

- **Lower and upper bounds.** We clearly have the following inequalities

$$l_1 \leq H_L(\mu) \leq l_k,$$

where  $k = \min(2^n, m) - 1$ .

- **The decisivity property.** We have

$$H_L(\mu) = l_1 \quad \Leftrightarrow \quad \mu(S) \in \{l_1, l_m\} \quad (S \subseteq N).$$

This is relevant for the ordinal aggregation framework. Indeed, it can be shown (cf. [12, Theorem 5.1]) that  $\mu(S) \in \{l_1, l_m\}$  for all  $S \subseteq N$  if and only if

$$\mathcal{S}_\mu(x) \in \{x_1, \dots, x_n\} \quad (x \in L^n).$$

- **The maximality property.** If  $m \leq 2^n$ , we have

$$H_L(\mu) = l_{m-1} \quad \Leftrightarrow \quad \{\mu(S) \mid S \subseteq N\} = L.$$

If  $m \geq 2^n$ , we have

$$H_L(\mu) = l_{2^n-1} \quad \Leftrightarrow \quad \text{all } \mu(S)\text{'s are distinct.}$$

- **Particular Sugeno integrals.** Let  $\omega = (\omega_1, \dots, \omega_n) \in L^n$ . If the Sugeno integral reduces to either a “weighted maximum”

$$\text{wmax}_\omega(x) = \bigvee_{i=1}^n (\omega_i \wedge x_i) \quad \text{with} \quad \bigvee_{i=1}^n \omega_i = l_m,$$

or a “weighted minimum”

$$\text{wmin}_\omega(x) = \bigwedge_{i=1}^n (\omega_i \vee x_i) \quad \text{with} \quad \bigwedge_{i=1}^n \omega_i = l_1,$$

or an “ordered weighted maximum”

$$\text{owmax}_\omega(x) = \bigvee_{i=1}^n (\omega_i \wedge x_{(i)}) \quad \text{with} \quad l_m = \omega_1 \geq \dots \geq \omega_n,$$

or an “ordered weighted minimum”

$$\text{owmin}_\omega(x) = \bigwedge_{i=1}^n (\omega_i \vee x_{(i)}) \quad \text{with} \quad \omega_1 \geq \dots \geq \omega_n = l_1,$$

(see [14] for details) then the entropy of the corresponding fuzzy measure  $\mu : 2^N \rightarrow L$  is given by

$$H_L(\mu) = l_{|\{\omega_1, \dots, \omega_n\}|-1}.$$

## 5 Conclusions

We have analyzed two recent proposals for the entropy of fuzzy measures and we have proposed an entropy for ordinal fuzzy measures. These concepts seem promising but need to be investigated more deeply and applied in practical situations. Particularly, axiomatic characterizations of these entropies would be welcome.

## 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, “ $k$ -order additive discrete fuzzy measures and their representation”, *Fuzzy Sets and Systems* **92** (1997) 167–189.
- [7] M. Grabisch, J.-L. Marichal, and M. Roubens, “Equivalent representations of set functions”, *Mathematics of Operations Research* **25** (2000) 157–178.
- [8] G.J. Klir and A. Ramer, “Uncertainty in the Dempster-Shafer theory: a critical re-examination”, *Int. J. General Systems* **18** (1990) 155–166.
- [9] G.J. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications* (Prentice Hall PTR, Upper Saddle River, NJ, 1995).
- [10] J.-L. Marichal, “Aggregation operators for multicriteria decision aid”, Ph.D. thesis, Institute of Mathematics, University of Liège, Liège, Belgium, 1998.
- [11] J.-L. Marichal, “An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria”, *IEEE Transactions on Fuzzy Systems*, to appear.
- [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*, submitted for revision.
- [13] J.-L. Marichal, “Entropy of discrete Choquet capacities”, *Eur. J. of Oper. Res.*, submitted for revision.
- [14] J.-L. Marichal, “On Sugeno integral as an aggregation function”, *Fuzzy Sets and Systems* **114** (3) (2000) 347–365.
- [15] J.-L. Marichal and M. Roubens, “Entropy of a Choquet capacity”, *Proc. of 1999 Eusflat-Estylf Joint Conference*, Palma de Mallorca, Spain, September 22–25, 1999, 383–385.
- [16] 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.
- [17] C.E. Shannon, “A mathematical theory of communication”, *Bell System Tech. J.* **27** (1948) 379–423, 623–656.
- [18] C.E. Shannon and W. Weaver, *A mathematical theory of communication* (University of Illinois Press, Urbana, 1949).
- [19] 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.
- [20] M. Sugeno, “Theory of fuzzy integrals and its applications”, Ph.D. Thesis, Tokyo Institute of Technology, Tokyo, 1974.
- [21] M. Sugeno, “Fuzzy measures and fuzzy integrals: a survey”, in: M.M. Gupta, G.N. Saridis, and B.R. Gaines (eds.), *Fuzzy Automata and Decision Processes*, (North-Holland, Amsterdam, 1977) 89–102.

- [22] R.R. Yager, “Entropy and specificity in a mathematical theory of evidence”, *Int. J. Gen. Systems* **9** (1983) 249–260.
- [23] R.R. Yager, “On ordered weighted averaging aggregation operators in multicriteria decision making”, *IEEE Trans. on Systems, Man and Cybernetics* **18** (1988) 183–190.
- [24] R.R. Yager, “An approach ordinal decision making”. Technical Report #MII-1410, Machine Intelligence Institute, Iona College, New Rochelle, NY, 1994.
- [25] R.R. Yager, “On the entropy of fuzzy measures”. Technical Report #MII-1917R, Machine Intelligence Institute, Iona College, New Rochelle, NY, 1999.
- [26] 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.