ENTROPY OF A CHOQUET CAPACITY*

JEAN-LUC MARICHAL
University of Liège - Faculty of Economics
Sart Tilman - B31 - 4000 Liège, Belgium

MARC ROUBENS
University of Liège - Institute of Mathematics
Sart Tilman - B37 - 4000 Liège, Belgium

1 Introduction

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

$$H_n(p) = -\sum_{i=1}^{n} p_i \log_2 p_i,$$

with the convention that $0 \log_2 0 := 0$, is called the *Shannon entropy* of p, see [8, 9]. 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 also 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, 5, 7, 11]. However, it seems that no definition of entropy for a general Choquet capacity was yet proposed in literature.

In this paper we present an entropy-like measure defined for all Choquet capacities. This "entropy" was proposed very recently by Marichal [6] 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.

*To be presented at 1999 Eusflat-Estylf Joint Conference, Palma de Mallorca, Spain, September 22–25, 1999.

2 Entropy and aggregation operators

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

$$\text{WAM}_\omega(x_1, \ldots, x_n) = \sum_{i=1}^{n} \omega_i \, x_i$$

with

$$\sum_{i=1}^{n} \omega_i = 1 \quad \text{and} \quad \omega_i \geq 0 \;\; \forall i = 1, \ldots, n.$$

It is clear that, in such an aggregation process, the use of the information contained in the arguments $x_1, \ldots, x_n$ strongly depends upon the weight vector ω. For example, consider two weighted arithmetic means with weights 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, one can define a *measure of dispersion* associated to the weight vector of the weighted arithmetic mean WAM_ω as the Shannon entropy of ω:

$$H_n(\omega) = -\sum_{i=1}^{n} \omega_i \log_2 \omega_i.$$

Such a function enables to measure how much of the information in the arguments is really used. In a certain sense the more disperse the ω the more the information contained in the arguments is being used in the aggregation process.

Now, consider the so-called ordered weighted averaging operator (OWA), proposed in 1988 by Yager [12]:

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

with

$$\sum_{i=1}^n \omega_i = 1 \quad \text{and} \quad \omega_i \geq 0 \;\; \forall i=1,\ldots,n,$$

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 depend on a reordering of the arguments, should also be given by the Shannon entropy. In fact, Yager [12] proposed explicitely to use this concept as measure of dispersion for the OWA operators.

It is known that $H_n(\omega)$ is maximum only when ω corresponds to the weight vector of the arithmetic mean, see e.g. [11]:

$$H_n(\omega)=\log_2 n \quad \text{for } \omega=(1/n,\ldots,1/n),$$

and minimum only when ω is a binary vector:

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

Thus, the measure of dispersion can be normalized into

$$\mathrm{disp}(\omega)=\frac{1}{\log_2 n}\,H_n(\omega)=-\sum_{i=1}^n \omega_i \log_n \omega_i,$$

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

3 Entropy of a capacity

Given a Choquet capacity v on N, the *(discrete) Choquet integral* of $x=(x_1,\ldots,x_n)$ w.r.t. v is defined by

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

with the usual convention that $x_{(1)} \leq \cdots \leq x_{(n)}$. Also $A_{(i)} := \{(i),\ldots,(n)\}$, and $A_{(n+1)}=\emptyset$. For more details, see e.g. [4] and the references there.

It is easy to see that the WAM operators correspond to the Choquet integrals w.r.t. additive capacities (i.e., such that $v(S\cup T)=v(S)+v(T)$ whenever $S\cap T=\emptyset$). Moreover, one can show 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, Marichal [6, §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

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

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

$$[v(T\cup\{i\})-v(T)]\log_n[v(T\cup\{i\})-v(T)],\; T\subseteq N\setminus\{i\},$$

that is an expression of the form

$$\begin{aligned}\mathrm{disp}(v) \;=\; & -\sum_{i=1}^n \sum_{T\subseteq N\setminus\{i\}} p_{|T|}\,[v(T\cup\{i\})-v(T)] \\ & \times \log_n[v(T\cup\{i\})-v(T)],\end{aligned}$$

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

On the other hand, imposing the condition

$$\mathcal{C}_v=\mathrm{OWA}_\omega \quad\Rightarrow\quad \mathrm{disp}(v)=\mathrm{disp}(\omega)$$

determines uniquely the coefficients $p_{|T|}$, so that the definition proposed is the following.

Definition 1 *The entropy of a Choquet capacity v on N is defined by*

$$\begin{aligned}\mathrm{disp}(v) := & -\sum_{i=1}^n \sum_{T\subseteq N\setminus\{i\}} \frac{(n-t-1)!\,t!}{n!} \\ & \times[v(T\cup\{i\})-v(T)]\log_n[v(T\cup\{i\})-v(T)].\end{aligned}$$

When the Choquet integral $\mathcal{C}_v$ is used as an aggregation operator, this entropy can be interpreted as the degree to which one uses all the information contained in the arguments $x=(x_1,\ldots,x_n)$ when calculating the aggregated value $\mathcal{C}_v(x)$.

Interestingly enough, its expression is very similar to that of the Shapley value of elements in N, which is a fundamental concept in game theory [10] expressing a power index:

$$\phi_i(v)=\sum_{T\subseteq N\setminus\{i\}} \frac{(n-t-1)!\,t!}{n!}\,[v(T\cup\{i\})-v(T)],$$

$i \in N$.

Notice also that this new definition has yet to be axiomatically characterized. However, to justify its use, one can show that it fulfils several properties required for an entropy [6], namely:

- $\text{disp}(v)$ is *continuous* w.r.t. v.
- disp is *symmetric*, that is
$$\text{disp}(\pi v) = \text{disp}(v)$$
for any permutation π of N. Here, πv is the capacity on N defined by $\pi v(\pi(S)) = v(S)$ for all $S \subseteq N$, where $\pi(S) = \{\pi(i) \,|\, i \in S\}$.
- We have
$$0 \leq \text{disp}(v) \leq 1.$$
Moreover, $\text{disp}(v)$ is maximum ($= 1$) only when $\mathcal{C}_v$ is the arithmetic mean, and minimum ($= 0$) only when v is a binary-valued capacity: $v(S) \in \{0, 1\}$ for all $S \subseteq N$. Note that this latter case occurs if and only if $\mathcal{C}_v(x) \in \{x_1, \ldots, x_n\}$ (only one piece of information is used in the aggregation).
- We have
$$\mathcal{C}_v = \text{WAM}_\omega \text{ or } \text{OWA}_\omega$$
$$\Downarrow$$
$$\text{disp}(v) = \text{disp}(\omega).$$
- Let $k \in N$ be a *null* element for v, that is, $v(T \cup \{k\}) = v(T)$ for all $T \subseteq N \setminus \{k\}$. Then
$$\text{disp}(v) = \text{disp}(v^{N \setminus \{k\}})$$
where $v^{N \setminus \{k\}}$ is the restriction of v to $N \setminus \{k\}$.

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

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

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

[7] A. Ramer, Uniqueness of information measure in the theory of evidence, *Fuzzy Sets and Systems* **24** (1987) 183–196.

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

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

[10] 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.

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

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