# ENTROPY OF A CHOQUET CAPACITY (FUZZY MEASURE)

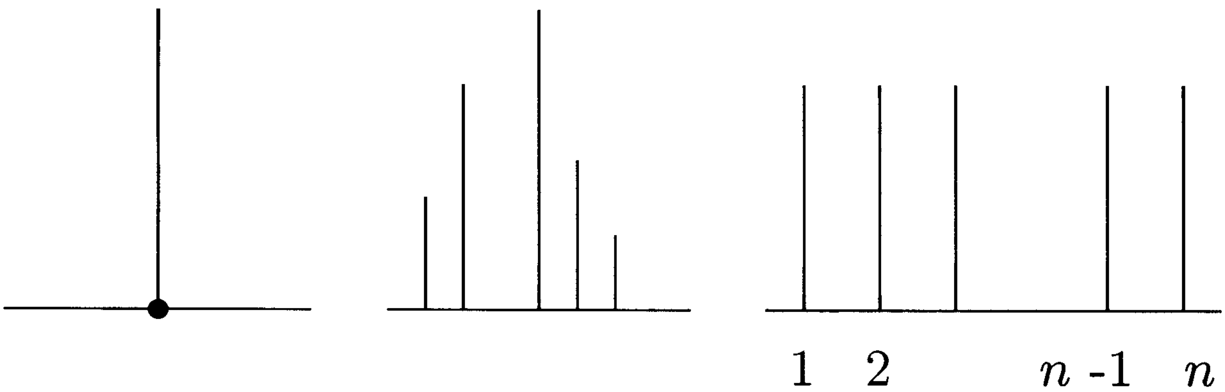
J.-L. Marichal
M. Roubens

University of Liège
Belgium

EUSFLAT, PALMA, Sept. 1999

MOTIVATION :

Classical Shannon entropy can be used as a measure of dispersion of probabilistic weights




$H_n(p) = 0$ $\quad 0 < H_n(p) < 1$ $\quad H_n(p) = 1$

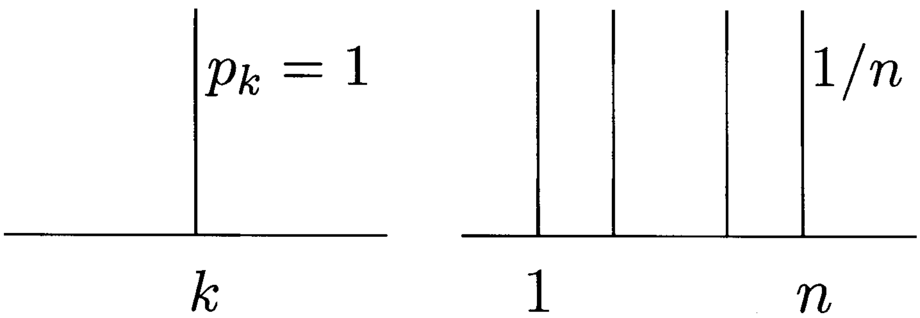
$$H_n(p) = -\sum_{i=1}^{n} p_i \log_n p_i$$

$H_n$ is characterized (see Yaglom and Yaglom (1959)) by

- $H_n$ is continuous.
- $H_n(p) = H_n(\pi p)$ : symmetric function (does not depend on the labels)
- $H_n(p_1 q_1, \dots, p_n q_n) = H_n(p_1, p_2, \dots, p_n) + H_n(q_1 \dots q_n)$ : additive function.
- Monotonic function.

$H_n$ can be used to determine how the weights of an aggregator of the type "weighted mean" are distributed over the scores :

$$WM(x_1, \ldots, x_n) = \sum_i p_i x_i$$




median ($n$ odd) arithmetic mean

max

min

We want to define a measure of dispersion of the capacities (Sugeno measure)

$$v : \text{ set function } : 2^N \rightarrow [0, 1]$$

$$\begin{cases} v(\emptyset) = 0 \\ v(R) \leq v(S) \text{ if } v \subset S \\ v(N) = 1 \end{cases}$$

particular case :

$v(T) = \sum_{i \in T} p(i)$, additive Sugeno measure.

Many proposals of extension of the notion of entropy have been proposed : (Klir and Folger, 1988, for a review)

As a measure of uncertainty and information over capacities

$$\left.\begin{array}{l}\text{Measure of dissonance}\\ \text{Measure of confusion}\\ \text{Measure of nonspecificity}\end{array}\right\}\ \begin{array}{l}\text{math. theory}\\ \text{of evidence}\end{array}$$

$$v \longrightarrow \underset{\substack{|\\ \text{basic probability assignment}}}{m} \quad : v(S) = \sum_{T \subset S} m(T)$$

$m$ is the Möbius transform of $v$

$$P\ell(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad Cr(A) = \sum_{B \subseteq A} m(B)$$

(plausibility) (credibility, belief)

$$E(m) = -\sum_{A \subseteq X} m(A) \ell n P\ell(A) \text{ (dissonance)}$$

$$C(n) = -\sum_{A} m(A) \ell n Cr(A) \text{ (confusion).}$$

We are concerned with an extension of the Shannon entropy to measure the amount of information in the arguments used in an aggregation operator.
This measure of information is called dispersion (DISP).

For any given capacity (fuzzy measure), we would like that

- If $v$ is a probability distribution, i.e. $v(S) = \sum_{i \in S} p_i$

$$disp(v) = H_n(p)$$

- $0 \leq disp(v) \leq 1$
- If $v(S)$ is a binary-valued capacity, $v(S) \in \{0, 1\}$

$$disp(v) = 0$$

(extension of $H(p) = 0$ iff $p_k = 1$ for some $k$).
- If $\delta_i v(T \cup i) = v(T \cup i) - v(T) = \frac{1}{n}, \forall i, \forall T \subset N$

$$disp(v) = 1$$

(extension of $H(p) = 1$ iff $p_k = \frac{1}{n}$, for all $k$).

Why are we concerned with $\delta_i v(T \cup i)$ ?

1) A classical aggregator used in ordinal expected utility with interactive criteria is the Choquet integral

$$
\begin{aligned}
C(x_1, \dots, x_n) &= \sum_i x_{(i)} [v((i), \dots, (n)) - v((i+1), \dots, (n))] \\
&= \sum_i x_{(i)} \delta_i v((i), \dots, (n))
\end{aligned}
$$

where $x_{(i)}$ are such that

$$x_{(1)} \leq \cdots \leq x_{(i)} \leq \cdots \leq x_{(n)}$$

$C$ is a weighted sum of ordered scores $x_{(i)}$. It extends the classical weighted mean

$$WM(x_1, \dots, x_n) = \sum_i x_i p_i$$

2) In classical game theory,

| | |
|---|---|
| $v$ : | characteristic function that defines a cooperative game |
| $N$ : | set of players ($i \in N$) |

the real power of a player $i$ is given by the Shapley value

$$Sh(i) = \sum_{T \subset N \backslash i} w_{|T|} \delta_i v(T \cup i)$$

$\delta_i v(T \cup i) = v(T \cup i) - v(T)$ : contribution of $i$ when joining the coalition of players $T$.

$w_{|T|}$ are the Shapley coefficients

$$w_{|T|} = w_t = \frac{(n-t-1)!t!}{n!} \qquad \left( \sum_{T \subset N \backslash i} w_t = 1 \right)$$

Important property : $Sh(i)$ share the value $v(N) = 1$

$$\sum_{i=1}^{n} Sh(i) = v(N) = 1.$$

Extension of Shannon entropy for Choquet capacities (Sugeno measures)

$$H(p) = -\sum_i p_i \log_n p_i \quad \left(\sum_i p_i = 1\right)$$

$$\downarrow \quad v(S) \;\#\; \begin{cases} v(\emptyset) = 0 \\ v(S) \le v(T), \ \forall S \subset T \\ v(N) = 1 \end{cases}$$

$$disp(v) = -\sum_i \sum_{T \subset N \setminus i} w_t \delta_i v(T \cup i) \log_n \delta_i v(T \cup i)$$

with $\sum_{T \subset N \setminus i} w_t = 1, \forall i, \ \forall T, \ |T| = t$

Properties

- $disp(v)$ is continuous
- $disp(v)$ is symmetric
- $disp(v) = 0$ if and only if $v$ is a $(1-0)$ fuzzy measure
- $disp(v) = 1$ if and only $v(T \cup i) - v(i) = \frac{1}{n}$ with $w_t$ being the Shapley coefficients

$$w_t = \frac{(n-t-1)!t!}{n!}$$

*disp* for *OWA*

$$OWA(x_1, \dots, x_n) = \sum_i x_{(i)} p_i$$

$OWA(x_1, \dots, x_n)$ is a Choquet integral such that

$$v(T \cup i) - v(T) = p_{n-t},$$

$\delta_i v(T)$ depends only on the cardinality of $T$, not on $i$ (symmetry)

$$\begin{aligned} C(x_1, \dots, x_n) \\ &= \sum_i x_{(i)} [v((i), \dots, (n)) - v((i+1), \dots, (n))] \\ &= \sum_i x_{(i)} [p_i] \end{aligned}$$

It implies that

$$\begin{aligned} disp_{OWA}(v) \\ &= -\sum_i \left( \overbrace{\left( \sum_{T \subset N \setminus i} w_t \right)}^{=1} \overbrace{\delta_i v(T \cup i)}^{p_i} \log_n \overbrace{\delta_i v(T \cup i)}^{p_i} \right) \\ &= -\sum_i p_i \log_n p_i \end{aligned}$$

This is the way Yager (1988) introduced the concept of "dispersion" in the framework of OWA and claims "the more disperse the $p$'s, the more of the information about the individual criteria is being used in the aggregation of the aggregate value".

## Disp. is useful in MCDM

Consider the following MC problem :

| | $M$ | $\varphi$ | $L$ |
|---|---|---|---|
| $a$ | 18 | 16 | 10 |
| $b$ | 10 | 12 | 18 |
| $c$ | 14 | 15 | 15 |

- Students good in $M$ and $L$ should be favoured or $\varphi$ and $L$

- $M$ and $\varphi$ give the same information about the profile of a student

- $M$ and $\varphi$ are more important than $L$

$$\Rightarrow \begin{cases} p(M) = p(\varphi) > p(L) \\ c > a > b \end{cases}$$

If weighted mean is used :

$$W(x) = x(M)p(M) + x(\varphi)p(\varphi) + x(L)p(L)$$

We end with a contradiction

$c > a$ and $(p(M) = p(\varphi))$

$$\Rightarrow \quad (14 + 15)p(M) + 15p(L) > (18 + 16)p(M) + 10p(L)$$

$$\Rightarrow \quad p(L) > p(M) \text{ !!!}$$

Use of Choquet integral can help

If

$$\begin{cases} v(M,\varphi) = .5 < v(M) + v(\varphi) = .9 & \text{: redondancy} \\ v(M,L) = v(\varphi,L) = .9 > v(M) + v(L) = .75 : & \text{: sinergy} \\ v(M) = v(\varphi) = .45 \\ v(L) = .3 \end{cases}$$

| | $M$ | $\varphi$ | $L$ | Choquet | Weighted mean |
|---|---|---|---|---|---|
| $a$ | 18 | 16 | 10 | 13.9 | 15.25 |
| $b$ | 10 | 12 | 18 | 13.6 | 12.75 |
| $c$ | 14 | 15 | 15 | 14.6 | 14.625 |
| | | | | $c > a > b$ | $a > c > b$ |

(with $p(M) = p(\varphi) = 3/8$, $p(L) = 2/8$)

However :

$$\frac{v(M)}{v(L)} = \frac{p(M)}{p(L)} = \frac{3}{2} \quad !$$

$disp\,(v) = .82$ $v$ rather well distributed over the total capacity.

*disp* can be used as a (non-linear) objective to determine the $v$'s.

$v$'s ?

$$\max disp(v)$$

under given constraints

$$\begin{cases} v(M,\varphi) > v(M) + v(\varphi) \\ v(M,L) < v(M) + v(L) \\ \quad \vdots \end{cases}$$

Re. paper by Marichal and Roubens

"Determination of weights of interacting criteria from a reference set"

EJOR, 99, to appear.