Minimum variance capacity identification

Ivan Kojadinovic

LINA CNRS FRE 2729, Site école polytechnique de l'université de Nantes, Rue Christian Pauc, 44306 Nantes, France

Abstract

In the framework of multi-criteria decision making whose aggregation process is based on the Choquet integral, we present a maximum entropy like method enabling to determine, if it exists, the "least specific" capacity compatible with the initial preferences of the decision maker. The proposed approach consists in solving a strictly convex quadratic program whose objective function is equivalently either the opposite of a *generalized entropy measure* or the *variance* of the capacity. The application of the proposed approach is illustrated on two examples.

Key words: Multi-criteria decision making; Interacting criteria; Choquet integral.

1 Introduction

Let us consider a multi-criteria decision making problem characterized by a set $\mathcal{A} := \{a, b, c, \dots\}$ of *alternatives* described by a set $N := \{1, \dots, n\}$ of *criteria*. We further assume that with each alternative $a \in \mathcal{A}$ can be associated a *profile* $(a_1, \dots, a_n) \in \mathbb{R}^n$ where, for any $i \in N$, $a_i \in E_i \subseteq \mathbb{R}$ represents the *utility* of a related to criterion i. For each $a \in \mathcal{A}$, we additionally make the hypothesis that the values $a_1, \dots, a_n$ are given on the same interval scale, which implies that all the utilities are *commensurable* [1] . In particular, $E_i = E, \forall\, i \in \{1, \dots, n\}$.

In classical multi-attribute utility theory (MAUT) (see e.g. [3]), the aim is to model the preferences of the decision maker represented by a binary relation

Email address: `ivan.kojadinovic@polytech.univ-nantes.fr` (Ivan Kojadinovic).

[1] From a practical perspective, the obtention of the utilities on a common interval scale can be performed, in the considered context, using the extension of the MACBETH methodology [1] proposed in [2].

$\succeq_{\mathcal{A}}$, by means of a *utility function* $U : E^n \to \mathbb{R}$ such that

$$a \succ_{\mathcal{A}} b \iff U(a_1, \ldots, a_n) > U(b_1, \ldots, b_n), \qquad \forall\, a, b \in \mathcal{A}.$$

The form of the utility function U depends on the hypotheses on which the multi-criteria decision making problem is grounded. When it can be assumed that the criteria are *mutually preferentially independent* (see e.g. [4]), it is frequent to consider that the utility function is additive and takes the form of a weighted arithmetic mean, i.e.

$$U(a) = W_\omega(a) := \sum_{i=1}^{n} \omega_i a_i, \qquad \forall\, a \in \mathbb{R}^n,$$

where, for any $i \in \{1, \ldots, n\}$, ω_i is the *weight* of criterion i with $\omega_i \geq 0$ and $\sum_{i=1}^n \omega_i = 1$.

The assumption of independence among criteria is however rarely verified. In order to be able to take interaction phenomena among criteria into account, it has been recently proposed to substitute a monotone set function μ on $N := \{1, \ldots, n\}$, called *Choquet capacity* [5] or *fuzzy measure* [6], to the weight vector ω, thereby allowing to model not only the importance of each criterion but also the importance of each subset of criteria [7,8].

A suitable aggregation operator that generalizes the weighted arithmetic mean when the criteria interact is then the Choquet integral with respect to (w.r.t.) the capacity μ [7,8,2].

The use of a Choquet integral as a utility function clearly requires the prior identification of the underlying capacity μ. The learning data from which μ is to be determined consists of what Marchant [9] calls the *initial preferences* of the decision maker : usually, a partial weak order over the set of alternatives, a partial weak order over the set of criteria, intuitions about the importance of the criteria, about their interaction, etc. In such a context, Marichal and Roubens proposed a simple identification method based on linear programming [10] which is the starting point of our work. Note that the capacity identification problem was also addressed in [11–15] in a more general framework.

In this paper, we propose an identification method based on a maximum entropy like principle [8,16–18]. The learning data are the same as those used in Marichal's and Roubens' method [10]. The main difference comes from the fact that among all the admissible capacities (compatible with the initial preferences of the decision maker), we choose the "least specific one", i.e. the one such that the corresponding Choquet integral is the closest to the simple arithmetic mean in the sense of a natural distance. From a practical perspective, the approach consists in solving a strictly convex quadratic program whose objective function is equivalently either the opposite of the extended Havrda and Charvat entropy of order 2 or the *variance* of the capacity. The proposed methodology has been implemented within the `kappalab` package [19] for the GNU R statistical system [20] (cf. § 5.3).

The paper is organized as follows. The first section is devoted to the presentation of the Choquet integral as an aggregation operator and to numerical indices that can be used to understand its behavior. In the second section, extensions of probabilistic entropy measures to capacities are introduced and the notion of *variance* of a capacity is defined. Next, a generalization of the maximum entropy principle to capacities is discussed and an interpretation of it in the context of Choquet integral based aggregation is given. The last section deals with the practical implementation of this principle and ends with two examples.

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

2 Aggregation by the Choquet integral

2.1 Choquet capacities and integral

As mentioned in the introduction, interaction phenomena among criteria can be modeled by a *capacity* [5] also called *fuzzy measure* [6].

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

Definition 1 *A capacity on N is a set function $\mu : \mathcal{P}(N) \to [0, 1]$ satisfying the following conditions :*

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

Furthermore, a capacity μ on N is said to be

- *additive* if $\mu(S \cup T) = \mu(S) + \mu(T)$ for all disjoint subsets $S, T \subseteq N$,
- *cardinality-based* if, for any $T \subseteq N$, $\mu(T)$ depends only on the cardinality of T. Formally, there exist $\mu_0, \mu_1, \dots, \mu_n \in [0, 1]$ such that $\mu(T) = \mu_t$ for all $T \subseteq N$ such that $|T| = t$.

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

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

In the framework of aggregation, for each subset of criteria $S \subseteq N$, the number $\mu(S)$ can be interpreted as the *weight* or the *importance* of S. The monotonicity of μ means that the weight of a subset of criteria can only increase when new criteria are added to it.

When using a capacity to model the importance of the subsets of criteria, a suitable aggregation operator that generalizes the weighted arithmetic mean is the Choquet integral [7,8,2].

Definition 2 *The Choquet integral of a function $a : N \to \mathbb{R}$ represented by the profile $(a_1, \ldots, a_n)$ w.r.t. a capacity μ on N is defined by*

$$C_\mu(a) := \sum_{i=1}^{n} a_{(i)}[\mu(A_{(i)}) - \mu(A_{(i+1)})],$$

where the notation $_{(.)}$ indicates a permutation such that $a_{(1)} \leq \cdots \leq a_{(n)}$. Also, $A_{(i)} := \{(i), \ldots, (n)\}$, for all $i \in \{1, \ldots, n\}$, and $A_{(n+1)} := \emptyset$.

Seen as an aggregation operator, the Choquet integral w.r.t. μ can be regarded as taking into account interaction phenomena among criteria, that is, complementarity or redundancy effects among elements of N modeled by μ [8]. Indeed, complementarity (resp. redundancy) between two disjoints subsets of criteria A and B can be naturally modeled by the inequality $\mu(A \cup B) \geq \mu(A) + \mu(B)$ (resp. $\leq$).

The Choquet integral generalizes the weighted arithmetic mean in the sense that, as soon as the capacity is additive (which intuitively coincides with the independence of the criteria), it collapses into a weighted arithmetic mean (i.e. a Lebesgue integral), that is

$$C_\mu(a) := \sum_{i=1}^{n} a_i \mu(i), \qquad \forall a = (a_1, \ldots, a_n) \in \mathbb{R}^n.$$

An intuitive presentation of the Choquet integral is given in [21]. Note that an axiomatic characterization of the Choquet integral as an aggregation operator was proposed by Marichal in [8].

2.2 The Möbius representation of a capacity

Any set function $\nu : \mathcal{P}(N) \to \mathbb{R}$ can be uniquely expressed in terms of its *Möbius representation* [22] by

$$\nu(T) = \sum_{S \subseteq T} m_\nu(S), \qquad \forall\, T \subseteq N, \tag{1}$$

where the set function $m_\nu : \mathcal{P}(N) \to \mathbb{R}$ is called the *Möbius transform* or *Möbius representation* of ν and is given by

$$m_\nu(S) = \sum_{T \subseteq S} (-1)^{s-t} \nu(T), \qquad \forall\, S \subseteq N.$$

Of course, any set of 2^n coefficients $\{m(S)\}_{S \subseteq N}$ does not necessarily correspond to the Möbius transform of a capacity on N. The boundary and monotonicity conditions must be ensured [23], i.e., we must have

$$\begin{cases} m(\emptyset) = \emptyset, \qquad \sum_{T \subseteq N} m(T) = 1, \\ \\ \sum_{\substack{T \subseteq S \\ T \ni i}} m(T) \geq 0, \qquad \forall\, S \subseteq N,\ \forall\, i \in S. \end{cases} \tag{2}$$

As shown in [23], in terms of the Möbius representation of a capacity μ on N, for any $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$, the Choquet integral of a w.r.t. μ is given by

$$\mathcal{C}_{m_\mu}(a) = \sum_{T \subseteq N} m_\mu(T) \bigwedge_{i \in T} a_i, \tag{3}$$

where the symbol $\wedge$ denotes the minimum operator.

2.3 Behavioral analysis of the aggregation

The behavior of the Choquet integral as an aggregation operator is generally difficult to understand. For a better comprehension of the interaction phenomena modeled by the underlying capacity, several numerical indices can be computed. In the sequel, we present two of them in detail.

The global importance of a criterion $i \in N$ can be measured by means of its Shapley value [24], which is defined by

$$\phi_\mu(i) := \sum_{T \subseteq N \setminus i} \gamma_s(n)[\mu(T \cup i) - \mu(T)],$$

where

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

The Shapley value of i can be thought of as an average value of the *marginal contribution* $\mu(T \cup i) - \mu(T)$ of criterion i to a subset T not containing it. A fundamental property is that the numbers $\phi_\mu(1), \ldots, \phi_\mu(n)$ form a probability distribution on N. In terms of the Möbius representation of μ, the Shapley value of i can be rewritten as

$$\phi_{m_\mu}(i) = \sum_{T \subseteq N \setminus i} \frac{1}{t+1} m_\mu(T \cup i). \tag{4}$$

The average interaction between two criteria i and j can be measured by means of their Shapley interaction index [25,26] which is defined as

$$I_\mu(ij) := \sum_{T \subseteq N \setminus ij} \frac{(n-t-2)!t!}{(n-1)!} [\mu(T \cup ij) - \mu(T \cup i) - \mu(T \cup j) + \mu(T)].$$

Similarly to the Shapley value, the Shapley interaction index between two criteria i and j can be regarded as an average value of their *marginal interaction*

$$\mu(T \cup ij) - \mu(T \cup i) - \mu(T \cup j) + \mu(T)$$

in the presence of a subset T of criteria not containing them. An important property is that $I_\mu(ij) \in [-1, 1]$ for all $ij \subseteq N$ [25], the value 1 (resp. -1) corresponding to maximum complementarity (resp. redundancy) between i and j. In terms of the Möbius representation of μ, $I_\mu(ij)$ can be rewritten as

$$I_{m_\mu}(ij) = \sum_{T \subseteq N \setminus ij} \frac{1}{t+1} m_\mu(T \cup ij). \tag{5}$$

Other indices that can help to understand the behavior of a Choquet integral are *veto* and *favor* indices [27,17,28], *orness* and *andness* degrees [27,17,28], *k-tolerance* and *k-intolerance* indices [29] and *amount of interaction* indices [30].

2.4 The concept of k-additivity

From the previous sections, it follows that a capacity μ on N is completely defined by the knowledge of 2^n coefficients, for instance $\{\mu(S)\}_{S \subseteq N}$ or $\{m_\mu(S)\}_{S \subseteq N}$. The use of capacities in practical applications is clearly curbed by this exponential complexity. Usually, as soon as n becomes large, it is frequent to consider that the capacity is additive and thus completely defined by the n coefficients $\{\mu(i)\}_{i \in N}$. Such a drastic simplification however greatly weakens the modeling ability of the capacity. The fundamental notion of *k-additivity* proposed by Grabisch in [25] enables to find a trade-off between the complexity of representation and the richness of the modeling.

Definition 3 *Let $k \in \{1, \ldots, n\}$. A capacity μ on N is said to be k-*additive* if its Möbius representation satisfies $m_\mu(T) = 0$ for all $T \subseteq N$ such that $t > k$ and there exists at least one subset T of cardinality k such that $m_\mu(T) \neq 0$.*

As one can easily check, the notion of 1-additivity coincides with that of additivity.

Let $k \in \{1, \ldots, n\}$ and let μ be a k-additive capacity on N. From Eq. (1), we immediately have that

$$\mu(S) = \sum_{\substack{T \subseteq S \\ t \leq k}} m_\mu(T), \qquad \forall\, S \subseteq N,$$

which confirms that a k-additive capacity is completely defined by the knowledge of $\sum_{l=1}^{k} \binom{n}{l}$ coefficients.

3 Entropy and variance of a capacity

3.1 Probabilistic entropies

The fundamental concept of *entropy of a probability distribution* was initially proposed by Shannon [31]. The Shannon entropy of a probability distribution p defined on a nonempty finite set $N := \{1, \ldots, n\}$ is defined by

$$H_S(p) := -\sum_{i \in N} p(i) \ln p(i),$$

with the convention that $0 \ln 0 := 0$. The quantity $H_S(p)$ is always non negative and 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 ($\ln n$) if and only if p is uniform (*maximality* property).

In a general non probabilistic setting, $H_S(p)$ 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.

Several axiomatic characterizations of the Shannon entropy were proposed in the literature [32–35], among which the most famous is probably Shannon's theorem [31].

A well-known generalization of the Shannon entropy is the Havrda and Charvat entropy of order β [36] defined, for any strictly positive real β and for any probability distribution p on N, by

$$H_{HC}^{\beta}(p) := \begin{cases} \dfrac{1}{1-\beta}\left[\sum_{i\in N} p(i)^{\beta} - 1\right], & \beta \neq 1, \\ \\ H_S(p), & \beta = 1. \end{cases} \tag{6}$$

As the Shannon entropy, the Havrda and Charvat entropy of order β is a strictly concave function of the probability distribution and satisfies the decisivity and maximality properties (with the exception that its maximal value is $\ln n$ only if $\beta = 1$).

The axiomatic characterization of the entropy H_{HC}^{β} is very similar to that of the Shannon entropy proposed by Faddeev [37]. The only difference comes from the form of the recursivity axiom [36].

Note that many other generalizations of the Shannon entropy were proposed in the literature. For an overview, see e.g. [33,38].

3.2 Choquet capacities, maximal chains and uniformity

A capacity being clearly a generalization of a discrete probability distribution, the following natural question arises : how could one appraise the 'uniformity' or 'uncertainty' associated with a Choquet capacity in the spirit of the Shannon entropy? In order to present the concept of uniformity of a capacity as intuitively defined in [18], we first introduce new definitions and notations.

The lattice related to the power set of $N := \{1,\dots,n\}$ under the inclusion relation can be represented by a graph $\mathcal{H}_N$, called Hasse diagram, whose nodes correspond to subsets $S \subseteq N$ and whose edges represent adding an element to the bottom subset to get the top subset.

A *maximal chain* m of $\mathcal{H}_N$ is an ordered collection of $n+1$ nested distinct subsets denoted

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

Denote by $\mathcal{M}_N$ the set of maximal chains of $\mathcal{H}_N$ 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 $\pi \in \Pi_N$ corresponds a unique maximal chain $m^\pi \in \mathcal{M}_N$ defined by

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

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

$$p_\pi^\mu(i) := \mu(\{\pi(i), \ldots, \pi(n)\}) - \mu(\{\pi(i+1), \ldots, \pi(n)\}), \qquad \forall i \in N.$$

Equivalently, with the maximal chain $m^\pi \in \mathcal{M}_N$ is associated the probability distribution $p_{m^\pi}^\mu := p_\pi^\mu$.

We will denote by P_N^μ the set $\{p_\pi^\mu\}_{\pi \in \Pi_N} = \{p_m^\mu\}_{m \in \mathcal{M}_N}$ of $n!$ probability distributions obtained from μ.

Notice that, if μ is cardinality-based, then there exists a unique probability distribution p^μ on N such that $p_\pi^\mu = p^\mu$ for all $\pi \in \Pi_N$. If μ is additive then we simply have $p_\pi^\mu(i) = \mu(\pi(i))$ for all $i \in N$.

As we have seen above, a capacity μ on N can be represented by the set $P_N^\mu = \{p_m^\mu\}_{m \in \mathcal{M}_N}$ of $n!$ probability distributions on N. As discussed in [18], the intuitive notion of *uniformity of a capacity* μ on N can then be defined as an average of the uniformity values of the probability distributions p_m^μ ($m \in \mathcal{M}_N$). Hence, the more uniform on average the probability distributions p_m^μ ($m \in \mathcal{M}_N$), the higher the uniformity of the capacity μ.

3.3 Definitions of generalized entropy measures and their properties

Let μ be a capacity on N. The following entropy was proposed by Marichal [39,40] (see also [41]) as an extension of the Shannon entropy to capacities :

$$H_M(\mu) := -\sum_{i \in N} \sum_{S \subseteq N \setminus i} \gamma_s(n)\, [\mu(S \cup i) - \mu(S)] \ln[\mu(S \cup i) - \mu(S)].$$

Regarded as a uniformity measure, H_M has been recently axiomatized by means of three axioms [18] : the symmetry property, a boundary condition for which it reduces to the Shannon entropy, and a generalized version of the well-known recursivity property.

A fundamental property of H_M is that it can be rewritten in terms of the maximal chains of the Hasse diagram of N. Using the notations introduced in the previous subsection, we have

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

or, equivalently,

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

The quantity $H_M(\mu)$ can therefore simply be seen as an average of the uniformity values of the probability distributions p^μ_m ($m \in \mathcal{M}_N$) calculated by means of the Shannon entropy. To stress on the fact that H_M is an average of Shannon entropies, we shall equivalently denote it by $\overline{H}_S$.

It has also been shown that $H_M = \overline{H}_S$ satisfies many properties that one would intuitively require from an entropy measure [18,40]. 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

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

where p^μ is the probability distribution on N defined by $p^\mu(i) = \mu(\{i,\dots,n\}) - \mu(\{i+1,\dots,n\})$ for all $i \in N$ (cf. § 3.2).

(3) **Decisivity.** For any capacity μ on N,

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

Moreover, $\overline{H}_S(\mu) = 0$ if and only if μ is a binary-valued capacity, that is, such that $\mu(T) \in \{0,1\}$ for all $T \subseteq N$.

(4) **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.

(5) **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

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

(6) **Strict concavity.** For any two capacities μ_1, μ_2 on N and any $\lambda \in]0,1[$, we have

$$\overline{H}_S(\lambda\,\mu_1 + (1-\lambda)\,\mu_2) > \lambda\,\overline{H}_S(\mu_1) + (1-\lambda)\,\overline{H}_S(\mu_2).$$

From Properties 3 and 4, we see that it is straightforward to define a normalized version of H_M to $[0,1]$ simply by division by $\ln n$ [17]. In the sequel, we shall denote this quantity H_M^*. From Property 6, we immediately have that maximizing H_M over a convex subset of the set of capacities on N always leads to a unique global maximum.

One can similarly extend the Havrda and Charvat entropy defined by Eq. (6) to capacities. Proceeding as in Eq. (7) and using Proposition 1 in [18], for any capacity μ on N, we define

$$\overline{H}_{HC}^{\beta}(\mu) := \frac{1}{1-\beta}\left[\sum_{i\in N}\sum_{S\subseteq N\setminus i}\gamma_s(n)[\mu(S\cup i)-\mu(S)]^{\beta}-1\right], \qquad \beta>0,\, \beta\neq 1. \tag{8}$$

From the characterization and the properties of the probabilistic Havrda and Charvat entropy presented in [36], it is easy to verify that $\overline{H}_{HC}^{\beta}$ satisfies the six properties stated above for H_M (up to a modification of the maximum value in the maximality property).

3.4 Variance of a capacity

Another straightforward way to measure the uniformity of a probability distribution p on N is to compute its variance :

$$V(p) := \frac{1}{n}\sum_{i\in N}\left[p(i)-\frac{1}{n}\right]^2,$$

which can be immediately rewritten as

$$V(p) = \frac{1}{n}\sum_{i\in N}p(i)^2 - \frac{1}{n^2}.$$

It is easy to verify that $V(p)=0$ if and only if p is uniform on N and that $V(p)$ reaches its maximum value $(n-1)/n^2$ if and only if p is a Dirac mass. The quantity $V(p)$ is in fact nothing else than the square of a Euclidean distance in $\mathbb{R}^n$ between p seen as a vector of $\mathbb{R}^n$ and the vector $(1/n,\ldots,1/n)\in\mathbb{R}^n$.

Proceeding as in Eq. (7) and using Proposition 1 in [18], the *variance* of a capacity μ on N can be immediately defined as

$$\overline{V}(\mu) := \frac{1}{n}\sum_{i\in N}\sum_{S\subseteq N\setminus i}\gamma_s(n)\,\left(\mu(S\cup i)-\mu(S)-\frac{1}{n}\right)^2, \qquad (9)$$
$$= \frac{1}{n}\sum_{i\in N}\sum_{S\subseteq N\setminus i}\gamma_s(n)\,\left(\mu(S\cup i)-\mu(S)\right)^2-\frac{1}{n^2}.$$

It is easy to verify that, for any capacity μ on N, the Havrda and Charvat entropy of order 2 and the variance are linked by the following linear equation :

$$\overline{H}^2_{HC}(\mu) = \frac{n-1}{n} - n\overline{V}(\mu). \qquad (10)$$

From the above equality for instance, it can be verified that V satisfies the following properties :

(1) **Boundary property for additive measures.** For any additive capacity μ on N, we have
$$\overline{V}(\mu) = V(p),$$
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
$$\overline{V}(\mu) = V(p^\mu),$$
where p^μ is the probability distribution on N defined by $p^\mu(i) = \mu(\{i,\dots,n\}) - \mu(\{i+1,\dots,n\})$ for all $i \in N$ (cf. § 3.2).

(3) **Minimality.** For any capacity μ on N, we have
$$\overline{V}(\mu) \geq 0.$$
with equality if and only if μ is the uniform capacity μ^* on N.

(4) **Decisivity.** For any capacity μ on N,
$$\overline{V}(\mu) \leq \frac{n-1}{n^2}$$
with equality if and only if μ is a binary-valued capacity, that is, such that $\mu(T) \in \{0,1\}$ for all $T \subseteq N$.

(5) **Decreasing monotonicity toward $\boldsymbol{\mu^*}$.** Let μ be a capacity on N such that $\mu \neq \mu^*$ and, for any $\lambda \in [0,1]$, define a 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
$$\overline{V}(\mu_{\lambda_1}) > \overline{V}(\mu_{\lambda_2}).$$

(6) **Strict convexity.** For any two capacities μ_1, μ_2 on N and any $\lambda \in]0,1[$, we have

$$\overline{V}(\lambda\,\mu_1 + (1-\lambda)\,\mu_2) < \lambda\,\overline{V}(\mu_1) + (1-\lambda)\,\overline{V}(\mu_2).$$

Finally, from the monotonicity constraints given in (2), it is easy to verify that, in terms of the Möbius representation, the variance of a capacity μ on N is simply given by

$$\overline{V}(m_\mu) = \frac{1}{n}\sum_{i\in N}\sum_{S\subseteq N\setminus i}\gamma_s(n)\left(\sum_{T\subseteq S} m_\mu(T\cup i) - \frac{1}{n}\right)^2. \qquad (11)$$

4 A maximum entropy principle for capacities

For probability distributions, the strict concavity of the Shannon entropy and its naturalness as a measure of uncertainty gave rise to the *maximum entropy principle*, which was pointed out in 1957 by Jaynes [16]. This principle states that, when one has only partial information about the possible outcomes of a random variable, one should choose its probability distribution so as to maximize the uncertainty about the missing information. In other words, all the available information should be used, but one should be as uncommitted as possible about missing information. In more mathematical terms, this principle means that among all the probability distributions that are in accordance with the available prior knowledge (i.e. a set of constraints), one should choose the one that has maximum uncertainty.

The strict concavity of H_M and $\overline{H}^\beta_{HC}$ suggests to extend such an inference principle to capacities as discussed in [8,17,18]. Let us give an interpretation of it in the framework of aggregation by a Choquet integral w.r.t. to a capacity μ on N in the presence of linear constraints. In such a context, as mentioned in [18], $H_M(\mu)$ can be interpreted as a measure of the average value over all $a \in E^n$ of the degree to which the arguments $a_1, \ldots, a_n$ of a profile contribute to the calculation of the aggregated value $C_\mu(a)$. This interpretation still holds for $\overline{H}^\beta_{HC}(\mu)$. In such a framework, the maximum entropy principle could therefore be stated as follows : Assume that we are given a set of linear constraints on the behavior of a Choquet integral C_μ, that is, constraints that are linear w.r.t. the corresponding capacity μ. Then, among all the feasible (admissible) Choquet integrals, choosing the Choquet integral w.r.t. the maximum entropy capacity amounts to choosing the Choquet integral that will have the highest average degree of contribution of its arguments in the aggregation phase. In other words, we could say the Choquet integral w.r.t. the maximum entropy capacity is the one that will exploit the most on average its arguments.

The constraints that we shall consider are naturally based on the indices presented in Subsection 2.3, which are all linear w.r.t. to the underlying capacity.

The application of the maximum entropy principle in the considered framework thus requires the maximization of a strictly concave function subject to linear constraints. The best studied such situation is probably that when the objective function is a quadratic form. It follows that the most interesting objective function from a practical perspective is $\overline{H}^2_{HC}$. This choice becomes even more natural from the point of view of the implementation of the approach since routines from solving quadratic programs are easily available.

From Eq. (10), we see that maximizing $\overline{H}^2_{HC}$ is clearly equivalent to minimizing $\overline{V}(\mu)$. This latter equivalence leads to another interpretation of the maximum $\overline{H}^2_{HC}$ entropy principle : among all the feasible Choquet integrals, choosing the Choquet integral w.r.t. the minimum variance capacity amounts to choosing the Choquet integral that will be the closest to the simple arithmetic mean in the sense of Eq. (9) subject to the set of considered linear constraints.

5 Application to Choquet integral based MAUT

5.1 Formulation of the optimization problem

Consider a multi-criteria decision making problem as described in the introduction and assume that a decision maker has some *initial preferences* about it. As discussed in [10], these initial preferences can take the form of :

- a partial weak order $\succeq_{\mathcal{A}}$ over $\mathcal{A}$ (ranking of the alternatives),
- a partial weak order $\succeq_N$ over N (ranking of the importance of the criteria),
- quantitative intuitions about the relative importance of some criteria,
- a partial weak order $\succeq_P$ over the set of pairs of criteria (ranking of interactions),
- intuitions about the type and the magnitude of the interaction between some criteria,
- the behavior of some criteria as *veto* or *favor*,
- etc.

In the context of aggregation by the Choquet integral, it seems natural to translate some of the above prior information as follows :

- $a \succ_{\mathcal{A}} b$ can be translated as $C_\mu(a) - C_\mu(b) \geq \delta_C$,
- $a \sim_{\mathcal{A}} b$ can be translated as $-\delta_C \leq C_\mu(a) - C_\mu(b) \leq \delta_C$,
- $i \succ_N j$ can be translated as $\phi_\mu(i) - \phi_\mu(j) \geq \delta_\phi$,

- $i \sim_N j$ can be translated as $-\delta_\phi \le \phi_\mu(i) - \phi_\mu(j) \le \delta_\phi$,
- $ij \succ_P kl$ can be translated as $I_\mu(ij) - I_\mu(kl) \ge \delta_I$,
- $ij \sim_P kl$ can be translated as $-\delta_I \le I_\mu(ij) - I_\mu(kl) \le \delta_I$,

where δ_C, δ_ϕ and δ_I are fixed preference thresholds to be defined by the decision maker.

In other terms, all the partial weak orders $\succeq_{\mathcal{A}}$, $\succeq_N$, $\succeq_P$ previously mentioned are translated into partial semiorders with fixed preference thresholds (for simplicity reasons).

The remaining more quantitative prior information could be translated as follows (although this is much more questionable) :

- intuitions about the relative importance $r \in [0,1]$ of a criterion i could be translated as $\phi_\mu(i) = r$ or as $\phi_\mu(i) \ge r$,
- intuitions about the type and the magnitude $m \in [0,1]$ of the interaction between two criteria i and j could be translated as $I_\mu(ij) = m$ in case of complementary interaction and as $-I_\mu(ij) = m$ in case of redundant interaction.

Finally, the veto (resp. favor) effect of a criterion i can be directly translated as $\mu(T) = 0$ for all $T \subseteq N$ such that $T \not\ni k$ (resp. $\mu(T) = 1$ for all $T \subseteq N$ such that $T \ni k$) [17, Propositions 3 and 4].

The maximum $\overline{H}^2_{HC}$ entropy principle can then be used to identify a capacity that is compatible with the initial preferences of the decision maker. The resulting Choquet integral can therefore be regarded as modeling the reasoning of the decision maker. The suggested approach can be stated as the following optimization problem :

$$
\min \overline{V}(\nu)
$$

$$
\text{subject to} \begin{cases} \nu(S \cup i) - \nu(S) \geq 0, \forall i \in N, \forall S \subseteq N \setminus i, \\ \nu(N) = 1, \\ C_\nu(a) - C_\nu(b) \geq \delta_C, \\ -\delta_C \leq C_\nu(c) - C_\nu(d) \leq \delta_C, \\ \vdots \\ \phi_\nu(i) - \phi_\nu(j) \geq \delta_\phi, \\ -\delta_\phi \leq \phi_\nu(k) - \phi_\nu(l) \leq \delta_\phi, \\ \vdots \\ I_\nu(ij) - I_\nu(kl) \geq \delta_I, \\ \vdots \\ I_\nu(ij) = \dots, \\ \vdots \end{cases}
$$

where ν is a *game* on N, i.e. a set function $\nu : \mathcal{P}(N) \to \mathbb{R}$ such that $\nu(\emptyset) = 0$.

Of course, the above problem may be infeasible if the constraints are inconsistent. Such a situation may arise when the considered constraints violate some of the axioms underlying the Choquet integral model (see e.g. [42]). In such a case, the Choquet integral cannot be considered as a sufficiently rich aggregation function for modeling the initial preferences of the decision maker.

A solution to the above problem is a full capacity defined by $2^n - 1$ coefficients. The number of variables involved in it increases exponentially with n and so will the computational time. For large problems, both for computational and simplicity reasons, it may be preferable to restrict the set of possible solutions to k-additive capacities, $k \in \{1, \dots, n\}$. The idea is here simply to rewrite the above optimization problem in terms of the Möbius transform of a k-additive game using Eqs. (11), (3), (4) and (5), which will decrease the number of variables from $2^n - 1$ to $\sum_{l=1}^{k} \binom{n}{l}$. We obtain

$$\min \overline{V}(m_\nu)$$
$$\text{subject to} \begin{cases} \sum_{\substack{T\subseteq S \\ t\leq k-1}} m_\nu(T\cup i) \geq 0,\ \forall i\in N,\ \forall S\subseteq N\setminus i, \\ \sum_{\substack{T\subseteq N \\ 0<t\leq k}} m_\nu(T) = 1, \\ C_{m_\nu}(a) - C_{m_\nu}(b) \geq \delta_C, \\ \vdots \\ \phi_{m_\nu}(i) - \phi_{m_\nu}(j) \geq \delta_\phi, \\ \vdots \end{cases} \tag{12}$$

where m_ν is the Möbius representation of a k-additive game ν on N.

5.2 Matrix formulation

Most solvers for solving quadratic programming problems on $\mathbb{R}^p$ use the following matrix formulation :

$$\min_{x\in\mathbb{R}^p} -d^t x + \tfrac{1}{2}x^t D x$$
$$\text{subject to} \begin{cases} A_e x = a_e \\ A_i x \geq a_i \end{cases} \tag{13}$$

where ^{t} denotes the matrix transpose, $d \in \mathbb{R}^p$, D is $p\times p$ real symmetric matrix, the matrix equality $A_e x = a_e$ represents the set of equality constraints and the matrix inequality $A_i x \geq a_i$ represents the set of inequality constraints.

In order to put the optimization problem (12) under the matrix form (13), let us first determine the form of the matrix D (d being clearly zero in our case).

To do so, in the rest of this subsection, we shall equivalently regard the Möbius transform m_ν of a k-additive game ν on N as a vector of $\mathbb{R}^{\sum_{l=1}^{k}\binom{n}{l}}$.

Consider first a matrix M of order $n2^{n-1} \times \sum_{l=1}^{k}\binom{n}{l}$ whose coefficients are either 0 or 1 and such that Mm_ν is a vector of $\mathbb{R}^{n2^{n-1}}$ whose coefficients are distinct elements from the set

$$\left\{ \sum_{\substack{T\subseteq S \\ t\leq k-1}} m_\nu(T\cup i) \right\}_{i\in N, S\subseteq N\setminus i}.$$

Next, let D_{Sh} be a diagonal matrix of order $n2^{n-1}$ whose elements are taken from the set of Shapley coefficients $\{\gamma_s(n)\}_{s\in\{0,\dots,n-1\}}$ multiplied by $1/n$ such that $D_{Sh}Mm_\nu$ is a vector of $\mathbb{R}^{n2^{n-1}}$ whose coefficients are distinct elements from the set

$$\left\{\frac{1}{n}\gamma_s(n)\sum_{\substack{T\subseteq S\\ t\leq k-1}} m_\nu(T\cup i)\right\}_{i\in N, S\subseteq N\setminus i}.$$

With these notations, the variance of a k-additive game ν can be written as

$$\begin{aligned}\overline{V}(\nu)=\overline{V}(m_\nu)&=(M(m_\nu-m_{\mu^*}))^tD_{Sh}(M(m_\nu-m_{\mu^*}))\\ &=(m_\nu-m_{\mu^*})^tM^tD_{Sh}M(m_\nu-m_{\mu^*}),\end{aligned}$$

where m_{μ^*} is a vector of $\mathbb{R}^{\sum_{l=1}^k\binom{n}{l}}$ such that Mm_{μ^*} is a vector of $\mathbb{R}^{n2^{n-1}}$ whose elements are all equal to $1/n$.

The matrix D in the matrix formulation (13) is thus defined by

$$D:=2M^tD_{Sh}M.$$

It is clearly symmetric and positive since, for any $m_\nu\in\mathbb{R}^{\sum_{l=1}^k\binom{n}{l}}$, $m_\nu^tDm_\nu$ is a weighted sum of squares with positive weights. Furthermore, the definiteness of D follows directly from the minimality property satisfied by $\overline{V}$ (see Subsection 3.4). It can be alternatively directly checked by noting that $m_\nu^tDm_\nu=0$ is equivalent to

$$\frac{2}{n}\sum_{i\in N}\left[\gamma_s(n)\sum_{S\subseteq N\setminus i}\left(\sum_{\substack{T\subseteq S\\ t\leq k-1}} m_\nu(T\cup i)\right)^2\right]=0,$$

which implies that $m_\nu(S)=0$ for all $S\subseteq N$, $0<s\leq k$.

We end this subsection by a few words about the constraints of the quadratic program.

The set of equality constraints should at least contain the constraint

$$\sum_{\substack{S\subseteq N\\ 0<s\leq k}} m_\nu(S)=1.$$

The monotonicity constraints straightforwardly correspond to the matrix inequality $Mm_\nu\geq 0$. The other constraints can be easily coded using the expressions of the Choquet integral, the Shapley value and the interaction index between two criteria in terms of the Möbius representation.

Table 1
Partial and global evaluations of the four cooks.

Cook	FL	ST	SC	Mean	Lin.Prog.	Min.Var
a	18	15	19	17.33	18.5	17.83
b	15	18	19	17.33	17	16.83
c	15	18	11	14.67	14.5	15.17
d	18	15	11	14.67	13	14.17

5.3 Practical implementation

The proposed approach was implemented within the `kappalab` package [19] for the `GNU R` statistical system [20]. The package is distributed as free software and can be downloaded from the `Comprehensive R Archive Network` (`http://cran.r-project.org`). The quadratic program is solved using the R `quadprog` package [43] which implements the dual method of Goldfarb and Idnani [44,45] for solving strictly convex quadratic programming problems.

5.4 A first simple example

We first consider the simple example presented in [10, Example 5.1] by Marichal and Roubens.

Four cooks a, b, c, d are evaluated according to their ability to prepare three dishes : frogs' legs (FL), steak tartare (ST) and stuffed clams (SC). Their evaluations on a $[0, 20]$ scale are given in Table 1.

The decision maker adopts the following reasoning : when a cook is renowned for his stuffed clams, it is preferable that he/she is also better in cooking frogs' legs than steak tartare, which implies that $a \succ_{\mathcal{A}} b$. However, when a cook badly prepares stuffed clams, it is more important that he/she is better in preparing steak tartare than frogs' legs, which leads to $c \succ_{\mathcal{A}} d$. Of course, we also immediately have $a \succ_{\mathcal{A}} d$ and $b \succ_{\mathcal{A}} c$ but these preferences do not contribute to anything since they naturally follow from the monotonicity of the Choquet integral [8].

Marichal and Roubens showed that there are no additive model that can lead to this partial ranking [10].

The minimum variance 2-additive game ν given in terms of its Möbius transform and satisfying, besides the monotonicity and normalization constraints,

Table 2
Coefficients, variance and normalized entropy of the Möbius representations of the 2-additive capacities obtained by linear programming (Lin.Prog.) and the minimum variance principle (Min.Var.).

	FL	ST	SC	FL, ST	FL, SC	ST, SC	$\overline{V}$	H_M^*
Lin.Prog.	0	0.5	0.5	0	0.5	-0.5	0.25	0.52
Min.Var.	0.17	0.5	0.33	0	0.33	-0.33	0.14	0.85

Table 3
Partial and global evaluations of the five students.

Student	S	P	E	M	En	Mean	Min.Var.1	Min.Var.2	Min.Var.3
a	18	11	18	11	11	13.8	15.04	14.83	14.88
b	18	18	11	11	11	13.8	13.04	12.83	12.88
c	11	11	18	18	11	13.8	12.04	11.83	11.88
d	18	11	11	11	18	13.8	16.04	15.83	15.88
e	11	11	18	11	18	13.8	14.04	13.83	13.88

the constraints $C_{m_\nu}(a) > C_{m_\nu}(b)$ and $C_{m_\nu}(c) > C_{m_\nu}(d)$ with the same preference threshold as in [10] (i.e. $\delta_C = 1$) is written in the second line of Table 2, the first line corresponding to the solution obtained using Marichal's and Roubens' approach.

As expected the minimum variance capacity has a lower variance than that obtained by linear programming. This is confirmed by the values of the normalized Marichal entropy. The global evaluations are given in Table 1. Again, as expected, the behavior of the minimum variance Choquet integral is closer to that of the simple arithmetic mean than that of the Choquet integral obtained by linear programming.

5.5 A larger example

We consider now the problem of the evaluation of students in an institute training econometricians. The students are evaluated with respect to five subjects : statistics (S), probability (P), economics (E), management (M) and English (En). The evaluations of five students a, b, c, d, e on a $[0, 20]$ scale are given in Table 3.

Assume that the institute is slightly more oriented towards statistics and probability and suppose that the decision maker considers that there are 3 groups

Table 4
Shapley value, variance and normalized Marichal entropy of the minimum variance k-additive capacities.

	k	S	P	E	M	En	$\overline{V}$	H_M^*
Min.Var.1	2	0.28	0.15	0.14	0.07	0.36	0.08	0.85
Min.Var.2	3	0.22	0.21	0.15	0.14	0.27	0.05	0.91
Min.Var.3	3	0.23	0.22	0.14	0.13	0.27	0.05	0.91

Table 5
Shapley interaction indices between subjects w.r.t. the minimum variance k-additive capacities.

	S,P	S,E	S,M	S,En	P,E	P,M	P,En	E,M	E,En	M,En
Min.Var.1	-0.18	0.13	0.0	0.03	0.0	0.1	0.0	-0.04	-0.11	0.0
Min.Var.2	-0.15	0.05	-0.02	0.00	0.06	0.13	-0.05	-0.04	-0.05	-0.02
Min.Var.3	-0.1	0.06	-0.02	-0.01	0.07	0.13	-0.07	-0.09	-0.06	0.0

of subjects : statistics and probability, economics and management, and English. Furthermore, he/she considers that within the two first groups, subjects have a redundant interaction, i.e. they overlap to a certain extent. Finally, if a student is good in statistics or probability, it is better that he/she is good in English than in economics or management. This reasoning leads to the following ranking :

$$d \succ_{\mathcal{A}} a \succ_{\mathcal{A}} e \succ_{\mathcal{A}} b \succ_{\mathcal{A}} c.$$

The Shapley value of the 2-additive capacity compatible with the Choquet integral constraints ($\delta_C = 1$) corresponding to the above ranking is given in the first line of Table 4. The interaction indices between subjects are given in the first line of Table 5.

Suppose now that by considering Table 4, the decision maker becomes aware of the fact that he/she did not mention that for him/her statistics and probability should have the same global importance as well as E and M, i.e. $S \sim_N P$ and $E \sim_N M$.

There are no 2-additive capacities compatible with these additional constraints. The Shapley value of the 3-additive capacity additionally compatible with the constraints $-\delta_\phi \leq \phi_{m_\nu}(S) - \phi_{m_\nu}(P) \leq \delta_\phi$ and $-\delta_\phi \leq \phi_{m_\nu}(E) - \phi_{m_\nu}(M) \leq \delta_\phi$ with $\delta_\phi = 0.01$ is given in the second line of Table 4. The interaction indices between subjects are given in the second line of Table 5.

After considering Table 5, the decision maker finally becomes aware that he would like statistics and probability, and economics and management, to overlap to the same extent, i.e. $\{S, P\} \sim_P \{E, M\}$.

The Shapley value of the 3-additive capacity additionally compatible with the constraint $-\delta_I \leq I_{m_\nu}(S, P) - I_{m_\nu}(E, M) \leq \delta_I$ with $\delta_I = 0.01$ is given in the third line of Table 4. The interaction indices between subjects are given in the third line of Table 5.

Note that the fact that the two last (more constrained) capacities have a lower variance than the first one is due to the fact that they are 3-additive whereas the first one is only 2-additive.

6 Conclusion

A maximum entropy like method for capacity identification was proposed and practically implemented. The objective function of the underlying quadratic program is the opposite of the extended Havrda and Charvat entropy of order 2 or equivalently the variance of the capacity. As in [10], the initial preferences of the decision maker are translated into linear constraints using the expressions of the Choquet integral and of some of the numerical indices that can be used to understand its behavior. The obtained solution, if any, can be regarded as leading to the "least specific" Choquet integral compatible with the prior reasoning of the decision maker.

Further research concerning the translation of the initial preferences of the decision maker into linear constraints would be required since some translations are clearly debatable. Moreover, the use of the extended Havrda and Charvat entropy of order 2 or equivalently of the variance as objective function would need to be more rigorously justified and compared with the use of other candidate objective functions based for instance on distances among capacities.

References

[1] C. Bana e Costa, J. Vansnick, Preference relations in MCDM, in: T. Gal, T. H. T. Steward (Eds.), MultiCriteria Decision Making : advances in MCDM models, algorithms, theory and applications, Kluwer, 1999.

[2] 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.

[3] R. L. Keeney, H. Raiffa, Decision with multiple objectives, Wiley, New-York, 1976.

[4] P. Vincke, Multicriteria Decision-aid, Wiley, 1992.

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

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

[7] M. Grabisch, The application of fuzzy integrals in multicriteria decision making, European Journal of Operational Research 89 (1992) 445–456.

[8] 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.

[9] T. Marchant, Towards a theory of MCDM : stepping away from social choice theory, Mathematical Social Sciences 45 (2003) 343–363.

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

[11] D. Filev, R. Yager, On the issue of obtaining OWA operator weights, Fuzzy Sets and Systems 94 (1998) 157–169.

[12] M. Grabisch, H. Nguyen, E. Walker, Fundamentals of uncertainty calculi with applications to fuzzy inference, Kluwer Academic, Dordrecht, 1995.

[13] M. Grabisch, M. Roubens, Application of the Choquet intergral in multicriteria decision making, in: M. Grabisch, T. Murofushi, M. Sugeno (Eds.), Fuzzy Measures and Integrals, Physica-Verlag, 2000, pp. 349–374.

[14] I. Kojadinovic, Estimation of the weights of interacting criteria from the set of profiles by means of information-theoretic functionals, European Journal of Operational Research 155 (2004) 741–751.

[15] A. Tanaka, T. Murofushi, A learning model using fuzzy measures and the Choquet integral, in: 5th Fuzzy System Symposium, Kobe, Japan, 1989, pp. 213–218.

[16] E. T. Jaynes, Information theory and statistical mechanics, Phys. Rev. 106 (1957) 620–630.

[17] 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.

[18] 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.

[19] M. Grabisch, I. Kojadinovic, `kappalab`: Non additive measure and integral manipulation functions, R package version 0.2 (2005).
URL `http://www.polytech.univ-nantes.fr/kappalab`

[20] R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, iSBN 3-900051-00-3 (2005).
URL `http://www.R-project.org`

[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] G.-C. Rota, On the foundations of combinatorial theory. I. Theory of Möbius functions, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 2 (1964) 340–368 (1964).

[23] A. Chateauneuf, J.-Y. Jaffray, Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion, Math. Social Sci. 17 (3) (1989) 263–283.

[24] L. S. Shapley, A value for n-person games, in: Contributions to the theory of games, vol. 2, Annals of Mathematics Studies, no. 28, Princeton University Press, Princeton, N. J., 1953, pp. 307–317.

[25] M. Grabisch, k-order additive discrete fuzzy measures and their representation, Fuzzy Sets and Systems 92 (2) (1997) 167–189.

[26] T. Murofushi, S. Soneda, Techniques for reading fuzzy measures (iii): interaction index, in: 9th Fuzzy System Symposium, Saporo, Japan, 1993, pp. 693–696.

[27] M. Grabisch, Alternative representations of discrete fuzzy measures for decision making, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 5 (5) (1997) 587–607.

[28] J.-L. Marichal, Tolerant or intolerant character of interacting criteria in aggregation by the Choquet integral, European Journal of Operational Research 155 (3) (2004) 771–791.

[29] J.-L. Marichal, k-intolerant capacities and Choquet integrals, in: 10th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2004), Perugia, Italia, 2004, pp. 601–608.

[30] I. Kojadinovic, An axiomatic approach to the measurment of the amount of interaction among criteria or players, Fuzzy Sets and Systems 152 (2005) 417–435.

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

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

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

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

[35] A. Khinchin, Mathematical foundations of information theory, Dover, 1957.

[36] J. Havrda, F. Charvat, Quantification method in classification processes: concept of structural α-entropy, Kybernetika 3 (1967) 30–35.

[37] D. Faddeev, Zum begriff der entropie einer endlichen wahrscheinlichkeitschemes, Arbeit zur Informationstheorie, Deutscher Verlag der Wissenschaften 1.

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

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

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

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

[42] P. Wakker, Additive Representations of Preferences : A New Foundation of Decision Analysis, Kluwer Academic Publishers, Dordrecht, 1989.

[43] B. Turlach, A. Weingessel, `quadprog`: Functions to solve quadratic programming problems., R package version 1.4-7 (2004).

[44] D. Goldfarb, A. Idnani, Dual and primal-dual methods for solving strictly convex quadratic programs, in: J. Hennart (Ed.), Numerical Analysis, Springer-Verlag, Berlin, 1982, pp. 226–239.

[45] D. Goldfarb, A. Idnani, A numerically stable dual method for solving strictly convex quadratic programs, Mathematical Programming 27 (1983) 1–33.