A maximum quadratic entropy principle for capacity identification

Ivan Kojadinovic
LINA CNRS FRE 2729
Site école polytechnique de l'université de Nantes
Rue Christian Pauc, 44306 Nantes, France
ivan.kojadinovic@univ-nantes.fr

Abstract

In the framework of multicriteria decision making based on Choquet integrals, 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.

Keywords: Multicriteria decision making; interacting criteria; Choquet integral.

1 Introduction

Consider a multicriteria decision making problem characterized by a set of alternatives $\mathcal{A} := \{a, b, c, \dots\}$ described by a set of criteria $N := \{1, \dots, n\}$. Each alternative $a \in \mathcal{A}$ is identified with its *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 alternative $a \in \mathcal{A}$, we shall further assume that the values $a_1, \dots, a_n$ are given on the same interval scale, which implies that all the utilities are commensurate ($E_i = E, \forall\, i \in \{1, \dots, n\}$).

In classical multiattribute utility theory (MAUT) [10, 18], the aim is to model the preferences of the decision maker represented by a binary relation $\succeq_{\mathcal{A}}$, by means of a *utility function* $U : E^n \to \mathbb{R}$ such that, for any $a, b \in \mathcal{A}$,

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

The form of the utility function U depends on the hypotheses on which the multicriteria decision making problem is grounded. When it can be assumed that the criteria are *mutually preferentially independent*, it is frequent to consider that the utility function is additive and takes the form of a weighted arithmetic mean. 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 $N := \{1, \dots, n\}$ 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 [15]. Such a set function μ, called *(discrete) Choquet capacity* [1] or *fuzzy measure* [22], satisfies $\mu(\emptyset) = 0$, $\mu(N) = 1$ and $\mu(S) \leq \mu(T)$ whenever $S \subseteq T \subseteq N$.

A suitable aggregation operator that generalizes the weighted arithmetic mean when the criteria interact is then the discrete Choquet integral with respect to (w.r.t) the capacity μ [15].

The use of a Choquet integral as utility function clearly requires the prior identification of the underlying capacity μ. The learning data from which μ is to be determined consists of the *initial preferences* of the decision maker : usually, a partial preorder over the set of alternatives, a partial preorder 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 [18] which is the starting point of our work. Note that the capacity identification problem was also addressed in [6, 7, 11, 23] in a more general framework.

In this paper, we propose an identification method based on a maximum entropy like principle [9, 13, 15, 16]. The learning data are the same as those used in Marichal's and Roubens' method [18].

The main difference comes from the fact that among all the admissible capacities (compatible with the initial preferences, if any), 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 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 [5] for the GNU R statistical system [19] (cf. § 5.2).

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 $\mu(\{i\})$, $N \setminus \{i\}$. Furthermore, cardinalities of subsets $S, T, \dots$, will often be denoted by the corresponding lower case letters $s, t, \dots$

2 Aggregation by the discrete Choquet integral

As mentioned in the introduction, interaction phenomena among criteria can be modeled by a *(discrete) Choquet capacity* [1] also called *fuzzy measure* [22].

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

Definition 2.1 *A discrete Choquet 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 $S, T \subseteq N$, $S \subseteq T \Rightarrow \mu(S) \leq \mu(T)$.

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 discrete Choquet integral [15].

Definition 2.2 *The Choquet integral of a function $a : N \to \mathbb{R}$ represented by the profile $(a_1, \dots, a_n)$ 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 $a_{(1)} \leq \cdots \leq a_{(n)}$, $A_{(i)} := \{(i), \dots, (n)\}$, for all $i \in \{1, \dots, n\}$, and $A_{(n+1)} := \emptyset$.

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, numerous numerical indices can be computed. The best-known ones are the Shapley importance indices [21], the interaction indices [4] and the *orness* and *andness* degrees [16].

3 Entropy and variance of a discrete Choquet capacity

3.1 Probabilistic entropies

The fundamental concept of *entropy of a probability distribution* was initially proposed by Shannon [20]. The Shannon entropy of a probability distribution p defined on a nonempty finite set $N := \{1, \dots, n\}$ 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). Note that in a general non probabilistic setting, $H_S(p)$ is merely a measure of the uniformity (evenness) of p.

A well-known generalization of the Shannon entropy is the Havrda and Charvat entropy of order β [8] defined, for any strictly positive real β, for any probability distribution p on N, by

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

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 $\beta = 1$).

3.2 Definitions of generalized entropy measures and their properties

Let μ be a discrete Choquet capacity on N. The following entropy was proposed by Marichal [14, 17] as an extension of the Shannon entropy to discrete Choquet 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 [13].

A fundamental property of H_M is that it can be rewritten in terms of the maximal chains of the Hasse diagram of N, i.e.

$$H_M(\mu) = \frac{1}{n!}\sum_{\pi\in\Pi_N} H_S(p_\pi^\mu), \quad (2)$$

where Π_N denotes the set of permutations on N and $p_\pi^\mu(i) := \mu(\{\pi(i),\ldots,\pi(n)\}) - \mu(\{\pi(i+1),\ldots,\pi(n)\})$ for all $\pi \in \Pi_N$ and all $i \in N$. More details can be found in [13].

The quantity $H_M(\mu)$ can therefore simply be seen as an average of the uniformity values of the probability distributions p_π^μ. To stress 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 [17]. The most important ones are the *decisivity, maximality, increasing monotonicity toward* μ^* and *strict concavity* properties [13].

One can similarly extend the Havrda and Charvat entropy defined by Eq. (1) to discrete Choquet capacities. Proceeding as in Eq. (2) and using Proposition 1 in [13], for any capacity μ on N, any $\beta > 0$, $\beta \neq 1$, we obtain

$$\overline{H}_{HC}^{\beta}(\mu) := \frac{1}{1-\beta}\left[\sum_{i\in N}\sum_{S\subseteq N\setminus i} \gamma_s(n)[\mu(S\cup i) - \mu(S)]^{\beta} - 1\right]. \quad (3)$$

From the characterization and the properties of the probabilistic Havrda and Charvat entropy presented in [8], it is easy to verify that $\overline{H}_{HC}^{\beta}$ satisfies the *decisivity, maximality, increasing monotonicity toward* μ^* and *strict concavity* properties presented in [13]. See [12] for more details.

3.3 Variance of a discrete Choquet capacity

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

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

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 on $\mathbb{R}^n$ 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. (2) and using Proposition 1 in [13], the *variance* of a capacity μ on N can be immediately defined as

$$\overline{V}(\mu) := \frac{1}{n}\sum_{i\in N}\sum_{S\subseteq N\setminus i}\gamma_s(n)\left(\mu(S\cup i)-\mu(S)-\frac{1}{n}\right)^2, \tag{4}$$
$$= \frac{1}{n}\sum_{i\in N}\sum_{S\subseteq N\setminus i}\gamma_s(n)\,(\mu(S\cup i)-\mu(S))^2-\frac{1}{n^2}.$$

The above quantity can be clearly interpreted as the square of a Euclidian distance between μ and the uniform capacity μ^*. See [12] for more details.

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 :

$$\overline{H}^2_{HC}(\mu) = \frac{n-1}{n} - n\overline{V}(\mu). \tag{5}$$

4 A maximum entropy principle for capacities

The strict concavity of H_M and $\overline{H}^{\beta}_{HC}$ suggests to extend the maximum entropy principle [9] to discrete Choquet capacities as discussed in [13, 15, 16]. 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 [13], $H_M(\mu)$ can be interpreted as a measure of the average value over all $a \in \mathbb{R}^n$ of the degree to which the arguments $a_1, \dots, a_n$ 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, if any, 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 based on the indices briefly mentioned in Subsection 2, which are all linear w.r.t to the underlying capacity. See [12] for more details.

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. (5), 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. (4) subject to the set of considered linear constraints.

5 Application to Choquet integral based MAUT

5.1 Formulation of the optimization problem

Consider a multicriteria decision making problem as described in the introduction and assume that a decision maker has some *initial preferences* about it. As discussed in [12, 18], these initial preferences can take the form of a partial preorder $\succeq_{\mathcal{A}}$ on $\mathcal{A}$ (ranking of the alternatives), a partial preorder $\succeq_N$ on N (ranking of the importance of the criteria), quantitative intuitions about the relative importance of some criteria, a partial preorder $\succeq_P$ on the set of pairs of criteria (ranking of interactions), intuitions about the type and the magnitude of the interaction between some criteria, etc.

In the context of aggregation by the discrete Choquet integral, it seems natural to translate the above partial preorders into partial semiorders with fixed preference thresholds (for simplicity reasons). For instance, $a \succ_{\mathcal{A}} b$ can be translated as $C_\mu(a) - C_\mu(b) \geq \delta_C$ and $a \sim_{\mathcal{A}} b$ can be translated as $-\delta_C \leq C_\mu(a) - C_\mu(b) \leq \delta_C$ where δ_C is a fixed preference threshold to be defined by the decision maker. For more details, see [12].

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 :

$$\min \overline{V}(\nu)$$
$$\text{subject to} \begin{cases} \nu(S \cup i) - \nu(S) \geq 0, \forall i \in N, \\ \qquad \forall S \subseteq N \setminus i, \\ \nu(N) = 1, \\ C_\nu(a) - C_\nu(b) \geq \delta_C, \\ \vdots \end{cases}$$

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 [25]. In such a case, the Choquet integral can not be considered as an 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, $k \in \{1, \dots, n\}$ [4]. The idea is here simply to rewrite the above optimization problem in terms of the Möbius transform of a k-additive game which will decrease the number of variables from $2^n - 1$ to $\sum_{l=1}^k \binom{n}{l} - 1$. See [12] for more details.

5.2 Practical implementation

The proposed approach was implemented within the `kappalab` package [5] for the `GNU R` statistical system [19]. The package is distributed as free software and should be soon downloadable from the `Comprehensive R Archive Network` (`http://cran.r-project.org`). The quadratic program is solved using the R `quadprog` package [24] which implements the dual method of Goldfarb and Idnani [2, 3] for solving strictly convex quadratic programming problems.

Table 1: Partial and global evaluations of the four cooks. LP stands for Linear Programming and MV for Minimum Variance.

Cook	FL	ST	SC	Mean	PL	MV
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 A simple example

We consider the example presented in [18, 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 [15].

Marichal and Roubens showed that there are no additive model that can lead to this partial ranking [18].

The minimum variance 2-additive game $\nu^{(2)}$ given in terms of its Möbius transform and satisfying, besides the monotonicity and normalization constraints, the constraints $C_{\nu^{(2)}}(a) > C_{\nu^{(2)}}(b)$

Table 2: Coefficients, variance and normalized entropy of the Möbius representations of the 2-additive capacities obtained by linear programming (1st line) and the minimum variance principle (2nd line).

FL	ST	SC	FL, ST	FL, SC	ST, SC	$\overline{V}$
0	0.5	0.5	0	0.5	-0.5	0.25
0.17	0.5	0.33	0	0.33	-0.33	0.14

and $C_{\nu^{(2)}}(c) > C_{\nu^{(2)}}(d)$ with the same preference threshold as in [18] (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.

References

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

[2] D. Goldfarb and A. Idnani. Dual and primal-dual methods for solving strictly convex quadratic programs. In J. Hennart, editor, *Numerical Analysis*, pages 226–239. Springer-Verlag, Berlin, 1982.

[3] D. Goldfarb and A. Idnani. A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming*, 27:1–33, 1983.

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

[5] M. Grabisch and I. Kojadinovic. `kappalab`*: Non additive measure and integral manipulation functions*, 2005. R package version 0.1-1.

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

[7] M. Grabisch and M. Roubens. Application of the Choquet intergral in multicriteria decision making. In M. Grabisch, T. Murofushi, and M. Sugeno, editors, *Fuzzy Measures and Integrals*, pages 349–374. Physica-Verlag, 2000.

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

[9] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630, 1957.

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

[11] 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*, 155:741–751, 2004.

[12] I. Kojadinovic. Unsupervised aggregation of commensurate correlated attributes by means of the choquet integral and entropy functionals. *International Journal of Intelligent Systems*, 2005. In Press.

[13] I. Kojadinovic, J.-L. Marichal, and M. Roubens. An axiomatic approach to the definition of the entropy of a discrete Choquet capacity. *Information Sciences*, 172:131–153, 2005.

[14] J.-L. Marichal. *Aggregation operators for multicriteria decison aid*. PhD thesis, University of Liège, Liège, Belgium, 1998.

[15] 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):800–807, 2000.

[16] J.-L. Marichal. Behavioral analysis of aggregation in multicriteria decision aid. In J. Fodor, B. D. Baets, and P. Perny, editors, *Preferences and Decisions under Incomplete Knowledge*, pages 153–178. Physica-Verlag, 2000.

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

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

[19] R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2004. ISBN 3-900051-00-3.

[20] C. E. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–623, 1948.

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

[22] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, Tokyo, Japan, 1974.

[23] A. Tanaka and T. Murofushi. A learning model using fuzzy measures and the Choquet integral. In *5th Fuzzy System Symposium*, pages 213–218, Kobe, Japan, 1989.

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

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