



HAL
open science

A note on the Sobol' indices and interactive criteria

Michel Grabisch, Christophe Labreuche

► **To cite this version:**

Michel Grabisch, Christophe Labreuche. A note on the Sobol' indices and interactive criteria. *Fuzzy Sets and Systems*, 2017, 315, pp.99-108. 10.1016/j.fss.2016.05.009 . halshs-03220838

HAL Id: halshs-03220838

<https://shs.hal.science/halshs-03220838>

Submitted on 7 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A note on the Sobol' indices and interactive criteria

Michel GRABISCH^{1*} and Christophe LABREUCHE²

¹ Paris School of Economics, University of Paris I
106-112, Bd de l'Hôpital, 75013 Paris, France

`michel.grabisch@univ-paris1.fr`

² Thales Research and Technology
1, Avenue Augustin Fresnel, 91767 Palaiseau, France

`christophe.labreuche@thalesgroup.com`

Abstract

The Choquet integral and the Owen extension (or multilinear extension) are the most popular tools in multicriteria decision making to take into account the interaction between criteria. It is known that the interaction transform and the Banzhaf interaction transform arise as the average total variation of the Choquet integral and multilinear extension respectively. We consider in this note another approach to define interaction, by using the Sobol' indices which are related to the analysis of variance of a multivariate model. We prove that the Sobol' indices of the multilinear extension gives the square of the Fourier transform, a well-known concept in computer sciences. We also relate the latter to the Banzhaf interaction transform and compute the Sobol' indices for the 2-additive Choquet integral.

Keywords: capacity, interaction index, Sobol' index, multilinear extension, Fourier transform

1 Introduction

In multicriteria decision making (MCDM), the Choquet integral with respect to a capacity has become a popular tool to model situations where some interaction exists between criteria [7]. This often happens in practice, as the evaluation of an alternative under several criteria is a complex process, where the level of importance of criteria generally depends on which criteria are satisfied or not. Let us give an example.

Example 1. *The design of a complex system or software requires making trade-offs among multiple and conflicting concerns, such as safety, security, power consumption, communication overhead, components lifetime and costs for a Secure Radio Architecture [12]. A multi-criteria decision making approach is proposed in [12] in order to help finding the best compromise among these concerns seen as criteria. These criteria interact one another. For instance, safety and security can be considered as veto since the customer cannot accept an alternative not fulfilling these criteria. Criteria power consumption and communication overhead are complementary (positive interaction) as satisfying only one of them decreases drastically the quality of the alternative, according to the Secure Radio architect.*

The basic ingredient to model interaction is the capacity (or fuzzy measure) defined on the set N of criteria, through its interaction transform [5], which is a generalization of the Shapley value. Another type of interaction has been introduced by Roubens [15] under the name *Banzhaf*

*Corresponding author.

interaction transform, and extends the Banzhaf value. So far, emphasis has been put on the former one in theoretical developments and applications.

It is well known that the Choquet integral is an extension of a capacity, seen as a pseudo-Boolean function, and called the Lovász extension. Another popular extension of pseudo-Boolean functions is the Owen extension or multilinear extension, known in Multi-Attribute Utility Theory from a long time ago [10]. Both can be considered as aggregation functions on a bounded closed domain (say, $[0, 1]^n$).

In applications, interpretation of models and their sensitivity analysis is important for the user. The following example illustrates that several sensitivity analyses can be necessary.

Example 2 (Ex. 1 cont.). *Two problems arise in the design of a Secure Radio Architecture [12]. First of all, the engineering process is an incremental process in which an initial architectural alternative is assessed, and given its evaluation, the architect has to decide on improvement means to apply to this alternative. She needs thus to know on which criteria, an improvement is (on average) the most influential on the overall score.*

Another problem is that the values of the attributes are obtained from dedicated analysis or simulation tools that can be time consuming. For instance, safety can be assessed by fault tree analysis techniques. It can therefore happen that the assessments of some alternatives on some criteria are missing, because of time constraints. The architect needs to identify the criteria for which a missing value has the largest impact on the uncertainty of the global evaluation.

So far, in MCDM, the two issues raised in Ex. 2 are handled by the same concept of mean importance and interaction indices. The interaction index for $S \subseteq N$ w.r.t. an aggregation function F on $[0, 1]^n$ can be defined as the total variation of F w.r.t. the coordinates in S – the importance of a criterion being a particular case when S is a singleton. Grabisch et al. [8] showed that the interaction transform corresponds to the interaction index w.r.t. the Choquet integral, which generalizes the Shapley value, while the Banzhaf interaction transform is the interaction index w.r.t. the multilinear extension. We stress in this note that the interaction index is, by its definition, a relevant tool for sensitivity analysis in the sense of the first issue raised in Ex. 2. However it is not suited for the second one.

To this end, we need to adopt another view of interaction, namely the statistical view. We start by considering a well-known transform used in computer sciences but so far ignored in the field of decision making, which is the Fourier transform. ANOVA (analysis of variance) is an interesting approach to decompose a multivariate function into a sum of terms, each depending on a subset of the set of variables. The decomposition is orthogonal in the sense that each term has zero mean, except for the empty set. Then the variance of the function is decomposed as the variance of each term, which yields the Sobol indices. It is thus a natural way to deal with the second issue in Ex. 2. We show that considering the aggregation model as a multivariate function and defining the interaction index of $S \subseteq N$ through the Sobol' index of S , we come up with the square of the Fourier transform when the aggregation function is the multilinear model (Theorem 1). We also show the close relation between the Fourier transform and the Banzhaf interaction transform, and compute the Sobol' indices for the 2-additive Choquet integral (Theorem 2). The result for the Choquet integral is limited to 2-additive capacities as there does not seem to be a simple and compact expression for a general capacity. This case is of particular importance in practice, since it constitutes a good compromise between versatility and complexity. Experimental studies in multicriteria evaluation have shown that 2-additive capacities have almost the same approximation ability than general capacities (see, e.g., [6]). Most of the applications of the Choquet integral are restricted to a 2-additive capacity [12, 1, 2].

Apart from the distinction between the two issues depicted in Ex. 2, the Sobol' indices have the advantage to be comparable. One can say for instance that the influence of the interaction

between criteria 1 and 2 is larger than that of the importance of criterion 1 taken alone. Such a comparison is not relevant with the interaction indices.

Throughout the paper, cardinalities of sets S, T, \dots are denoted by the corresponding small letters s, t, \dots

2 Basic notions

We consider throughout a finite set $N = \{1, \dots, n\}$.

A *set function* is a mapping $\xi : 2^N \rightarrow \mathbb{R}$. A *game* v is a set function vanishing on the empty set: $v(\emptyset) = 0$. A *capacity* [3] or *fuzzy measure* [19] μ is a game satisfying monotonicity: for every $S, T \in 2^N$ such that $S \subseteq T$, we have $\mu(S) \leq \mu(T)$.

Clearly, the set \mathbb{R}^{2^N} of set functions on N is a 2^n -dimensional vector space. We introduce on \mathbb{R}^{2^N} the following inner product:

$$\langle \xi, \xi' \rangle = \frac{1}{2^n} \sum_{S \subseteq N} \xi(S) \xi'(S).$$

A classical basis of \mathbb{R}^{2^N} is the basis of the unanimity games. For any nonempty subset $S \subseteq N$, the *unanimity game centered on S* is the game defined by

$$u_S(T) = \begin{cases} 1, & \text{if } T \supseteq S \\ 0, & \text{otherwise} \end{cases}.$$

Defining the set function $u_\emptyset(S) = 1$ for every $S \subseteq N$, we obtain a basis of set functions. A drawback of this basis is that it is not orthogonal w.r.t. the above inner product. We will introduce later an orthonormal basis.

It is well known that the coordinates of ξ in this basis are the *Möbius transform* coefficients:

$$\xi = \sum_{S \in 2^N} m^\xi(S) u_S$$

with

$$m^\xi(S) = \sum_{T \subseteq S} (-1)^{|S \setminus T|} \xi(T).$$

The Möbius transform (or *Möbius inverse*) is a fundamental notion in combinatorics (see, e.g., [14]).

There is another vision of set functions, namely the pseudo-Boolean functions [9], noting that any subset A of N can be encoded by its characteristic function 1_A . Formally, a *pseudo-Boolean function* is a mapping $f : \{0, 1\}^n \rightarrow \mathbb{R}$. It follows that the set of pseudo-Boolean functions of n variables is a 2^n -dimensional vector space, with inner product

$$\langle f, f' \rangle = \frac{1}{2^n} \sum_{x \in \{0, 1\}^n} f(x) f'(x).$$

The standard polynomial expression of a pseudo-Boolean function f is

$$f(x) = \sum_{A \subseteq N} f(1_A) \prod_{i \in A} x_i \prod_{i \in A^c} (1 - x_i),$$

for every $x \in \{0, 1\}^n$, and with the convention $\prod_{i \in \emptyset} x_i = 1$. Rearranging terms, we get a sum of monomials:

$$f(x) = \sum_{T \subseteq N} a_T \prod_{i \in T} x_i, \quad (1)$$

for every $x \in \{0, 1\}^n$, where the coefficients a_T form the Möbius transform of ξ_f , the set function associated with f . Indeed, observe that unanimity games u_S correspond to monomials $\prod_{i \in S} x_i$.

Starting from (1), the *Owen extension* [13] or *multilinear extension* is obtained by letting x vary in $[0, 1]^N$:

$$f^{\text{Ow}}(x) = \sum_{T \subseteq N} a_T \prod_{i \in T} x_i \quad (x \in [0, 1]^N).$$

Another possible extension is obtained by remarking that in (1), the product can be replaced by the minimum without changing the value of the function. Letting again x to vary in $[0, 1]^N$, we obtain the *Lovász extension* [11]:

$$f^{\text{Lo}}(x) = \sum_{T \subseteq N} a_T \bigwedge_{i \in T} x_i \quad (x \in [0, 1]^N). \quad (2)$$

The Lovász extension coincides in fact with the Choquet integral [3].

3 Transforms of set functions

The Möbius transform presented above is an example of a linear and invertible transform on the set of set functions, in the sense that with each set function ξ is associated another set function m^ξ , called the Möbius transform of ξ , being linear because $m^{\xi + \alpha \xi'} = m^\xi + \alpha m^{\xi'}$ for every ξ, ξ' and $\alpha \in \mathbb{R}$, and ξ can be recovered from m^ξ by the inverse transform:

$$\xi(S) = \sum_{T \subseteq S} m^\xi(T) \quad (S \in 2^N).$$

3.1 Interaction transforms

Other such transforms exist, and three of them are of importance in this paper. The first one is the *interaction* transform [5], defined by

$$I^\xi(S) = \sum_{T \subseteq N \setminus S} \frac{(n-t-s)!t!}{(n-s+1)!} \sum_{L \subseteq S} (-1)^{|S \setminus L|} \xi(T \cup L),$$

and the inverse relation is given by

$$\xi(S) = \sum_{K \subseteq N} \beta_{|S \cap K|}^{|K|} I^\xi(K),$$

where

$$\beta_k^l = \sum_{j=0}^k \binom{k}{j} B_{l-j} \quad (k \leq l),$$

and B_0, B_1, \dots are the Bernoulli numbers. Also, the interaction transform has a simple expression in terms of the Möbius transform:

$$I^\xi(S) = \sum_{T \supseteq S} \frac{1}{t-s+1} m^\xi(T). \quad (3)$$

The interaction transform is of primary importance in multicriteria decision making, as it permits to model interaction between criteria [7]. Note that when S is a singleton $I^\xi(S)$ corresponds exactly to the Shapley index [16]. When S is a pair $\{i, j\}$, the interaction index $I^\xi(\{i, j\})$ can be interpreted as a kind of average value of the *added value* given by putting i and j together, all coalitions being considered. It is positive when i, j act in a cooperative or complementary way (i.e. the importance of the pair (i, j) is significantly greater than importances of i and j alone), and it is negative when i, j act in a redundant or substitutive way (i.e. the importance of the pair (i, j) is almost the same as importances of i and j alone). It is zero if the criteria i and j are independent. It is the case with a weighted sum: $I_{ij} = 0$ for any pair i, j of criteria. Lastly, $I^\xi(S)$ is harder to interpret when S has more than two elements.

A similar transform is the *Banzhaf interaction transform* [15] defined by

$$I_B^\xi(S) = \left(\frac{1}{2}\right)^{n-s} \sum_{K \subseteq N} (-1)^{|S \setminus K|} \xi(K),$$

with inverse relation

$$(I_B^{-1})^\xi(S) = \sum_{K \subseteq N} \left(\frac{1}{2}\right)^k (-1)^{|K \setminus S|} \xi(K).$$

Its expression in terms of the Möbius transform is

$$I_B^\xi(S) = \sum_{K \supseteq S} \left(\frac{1}{2}\right)^{|K \setminus S|} m^\xi(K), \quad (4)$$

and the converse relation is

$$m^\xi(S) = \sum_{K \supseteq S} \left(-\frac{1}{2}\right)^{|K \setminus S|} I_B^\xi(K). \quad (5)$$

In [8], the following close relation between the two extensions of a pseudo-Boolean function and the two interaction indices are shown:

$$I_B^\xi(S) = \int_{[0,1]^n} \frac{\partial^s f^{\text{Ow}}}{\partial x_S}(x) dx \quad (6)$$

$$I^\xi(S) = \int_{[0,1]^n} \Delta_S f^{\text{Lo}}(x) dx \quad (7)$$

where f is the pseudo-Boolean function corresponding to ξ , x_S is the restriction of x to coordinates in S , and

$$\Delta_S f^{\text{Lo}}(x) = \sum_{T \supseteq S} m^\xi(T) \bigwedge_{i \in T \setminus S} x_i,$$

which plays the role of a partial derivative.

3.2 Fourier transform

Another transform is the *Fourier transform*, well known in computer sciences. It is defined as the coordinates of a set function in the basis of the parity functions. For any subset $S \subseteq N$, the *parity function* associated with S is the function

$$\chi_S(x) = (-1)^{1_S \cdot x} = (-1)^{\sum_{i \in S} x_i} \quad (x \in \{0, 1\}^n),$$

where $1_S \cdot x$ is the inner product between the two vectors $1_S, x$. The parity function outputs 1 if the number of variables in S having value 1 is even, and -1 if it is odd. In terms of set functions, the parity function reads

$$\chi_S(T) = (-1)^{|S \cap T|} \quad (T \in 2^N).$$

It can be checked that the set of parity functions forms an orthonormal basis of \mathbb{R}^{2^N} .

Let us denote by $\widehat{f}(S)$, $S \subseteq N$, the coordinates of a pseudo-Boolean function f in the basis of parity functions:

$$f = \sum_{S \subseteq N} \widehat{f}(S) \chi_S. \quad (8)$$

The basis being orthonormal, it follows from (8) that $\widehat{f}(S)$ is simply given by

$$\widehat{f}(S) = \langle f, \chi_S \rangle = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} (-1)^{1_S \cdot x} f(x) \quad (S \subseteq N),$$

or, in terms of set functions,

$$\widehat{\xi}(S) = \frac{1}{2^n} \sum_{T \subseteq N} (-1)^{|S \cap T|} \xi(T) \quad (S \subseteq N).$$

The set of coordinates $\{\widehat{f}(S)\}_{S \subseteq N}$ is the *Fourier transform*.

We establish the relation between the Fourier, Möbius and Banzhaf transforms. Taking any set function ξ , we have

$$\begin{aligned} \widehat{\xi}(S) &= \frac{1}{2^n} \sum_{T \subseteq N} (-1)^{|S \cap T|} \xi(T) \\ &= \frac{1}{2^n} \sum_{T \subseteq N} (-1)^{|S \cap T|} \sum_{K \subseteq T} m^\xi(K) \\ &= \frac{1}{2^n} \sum_{K \subseteq N} m^\xi(K) \sum_{T \supseteq K} (-1)^{|S \cap T|}. \end{aligned}$$

Now,

$$\begin{aligned} \sum_{T \supseteq K} (-1)^{|S \cap T|} &= (-1)^{|K \cap S|} 2^{n-|K \cup S|} + (-1)^{|K \cap S|+1} 2^{n-|K \cup S|} \binom{|S \setminus K|}{1} + \\ &\quad (-1)^{|K \cap S|+2} 2^{n-|K \cup S|} \binom{|S \setminus K|}{2} + \dots + (-1)^{|S|} 2^{n-|K \cup S|} \\ &= (-1)^{|K \cap S|} 2^{n-|K \cup S|} \left(1 - \binom{|S \setminus K|}{1} + \binom{|S \setminus K|}{2} + \dots + (-1)^{|S \setminus K|} \right). \end{aligned}$$

Observe that

$$1 - \binom{|S \setminus K|}{1} + \binom{|S \setminus K|}{2} + \dots + (-1)^{|S \setminus K|} = 0$$

except if $|S \setminus K| = 0$. It follows that

$$\begin{aligned} \widehat{\xi}(S) &= \frac{1}{2^n} \sum_{K \supseteq S} m^\xi(K) (-1)^{|K \cap S|} 2^{n-|K \cup S|} \\ &= (-1)^{|S|} \sum_{K \supseteq S} \frac{1}{2^k} m^\xi(K). \end{aligned} \quad (9)$$

Now, using (4), we obtain

$$\widehat{\xi}(S) = \left(\frac{-1}{2}\right)^s I_B^\xi(S). \quad (10)$$

Lastly, we obtain from (10) and (5)

$$m^\xi(S) = (-2)^s \sum_{T \supseteq S} \widehat{\xi}(T).$$

4 The Sobol' indices

In statistics, the analysis of variance (ANOVA) (see, e.g., Fischer and Mackenzie [4]) is a well-known tool to model interaction between variables in a multivariate model. Consider n independent random variables Z_1, \dots, Z_n , with uniform distribution on $[0, 1]$, and a multivariate model $Y = F(Z)$, where $Z = (Z_1, \dots, Z_n)$. Let us denote for simplicity groups of variables $(Z_i)_{i \in S}$ by Z_S , and Z_{-S} denotes $(Z_i)_{i \notin S}$. Hence, we may write $Z = (Z_S, Z_{-S})$. Moreover, we denote by $\mathbb{E}[Y]$ the expected value of Y taken over all variables Z_1, \dots, Z_n . The expected value of Y can be taken on a subset Z_S of variables, with the corresponding notation $\mathbb{E}_{Z_S}[Y]$.

Any multivariate function can be decomposed in the following way (ANOVA decomposition):

$$Y = F(Z) = F_\emptyset + \sum_{i=1}^n F_i(Z_i) + \sum_{i < j} F_{ij}(Z_i, Z_j) + \dots + F_N(Z) = \sum_{S \subseteq N} F_S(Z_S),$$

with

$$\begin{aligned} F_\emptyset &= \mathbb{E}[Y] \\ F_i(Z_i) &= \mathbb{E}[Y|Z_i] - F_\emptyset \\ F_{ij}(Z_i, Z_j) &= \mathbb{E}[Y|Z_i, Z_j] - F_i(Z_i) - F_j(Z_j) - F_\emptyset \\ &= \mathbb{E}[Y|Z_i, Z_j] - \mathbb{E}[Y|Z_i] - \mathbb{E}[Y|Z_j] + \mathbb{E}[Y] \\ &\vdots \\ F_S(Z_S) &= \mathbb{E}_{Z_{-S}}[Y|Z_S] - \sum_{T \subset S} F_T(Z_T) = \sum_{T \subseteq S} (-1)^{|S \setminus T|} \mathbb{E}_{Z_{-T}}[Y|Z_T] \\ &\vdots \\ F_N(Z) &= \sum_{T \subseteq N} (-1)^{|N \setminus T|} \mathbb{E}_{Z_{-T}}[Y|Z_T]. \end{aligned}$$

The property of this decomposition is that each term has zero mean, except the first one, F_\emptyset . It follows that the variance of Y can be decomposed as follows:

$$\text{Var}[Y] = \sum_{\emptyset \neq S \subseteq N} \text{Var}[F_S(Z_S)].$$

The *first-order Sobol' indices* [17, 18] are the quantities $\frac{\text{Var}[F_S(Z_S)]}{\text{Var}[Y]}$, although one can omit the normalization factor. The next theorem establishes the close link between Sobol' indices and the Fourier transform (and consequently the Banzhaf transform) for the multilinear model.

Theorem 1. Consider the multilinear extension f_μ^{Ow} of a capacity μ . Then the (nonnormalized) Sobol' index for a subset $\emptyset \neq S \subseteq N$ is given by

$$\text{Var}[(f_\mu^{\text{Ow}})_S] = \frac{1}{3^s} (\widehat{\mu}(S))^2,$$

where $\widehat{\mu}$ is the Fourier transform of μ . Moreover, the ANOVA decomposition takes the following form

$$f_{\mu}^{\text{Ow}}(x) = \sum_{S \subseteq N} (-1)^s \prod_{i \in S} (2x_i - 1) \times \widehat{\mu}(S).$$

Proof. We set for simplicity $f = f_{\mu}^{\text{Ow}}$. We compute

$$f_S = \sum_{K \subseteq S} (-1)^k \mathbb{E}(f \mid Z_{S \setminus K}) \quad (S \subseteq N, |S| > 0).$$

We have for any such S :

$$\begin{aligned} \mathbb{E}(f \mid Z_{S \setminus K}) &= \int_{[0,1]^{N \setminus (S \setminus K)}} f \, dz_{N \setminus (S \setminus K)} = \sum_{T \subseteq N} m^{\mu}(T) \int_{[0,1]^{N \setminus (S \setminus K)}} \prod_{i \in T} z_i \, dz_{N \setminus (S \setminus K)} \\ &= \sum_{T \subseteq N} m^{\mu}(T) \frac{1}{2^{|T \setminus (S \setminus K)|}} \prod_{i \in T \cap (S \setminus K)} z_i \\ &= \sum_{L \subseteq N \setminus S} \sum_{T \subseteq S} m^{\mu}(L \cup T) \frac{1}{2^{|L \cup (T \cap K)|}} \prod_{i \in T \setminus K} z_i. \end{aligned} \quad (11)$$

It follows that

$$\begin{aligned} f_S &= \sum_{K \subseteq S} (-1)^k \sum_{L \subseteq N \setminus S} \sum_{T \subseteq S} m^{\mu}(L \cup T) \frac{1}{2^{|L \cup (T \cap K)|}} \prod_{i \in T \setminus K} z_i \\ &= \sum_{L \subseteq N \setminus S} \frac{1}{2^{|L|}} \sum_{T \subseteq S} m^{\mu}(L \cup T) \sum_{K \subseteq S} (-1)^k \frac{1}{2^{|T \cap K|}} \prod_{i \in T \setminus K} z_i. \end{aligned}$$

Letting $T' = T \cap K$, we have

$$\sum_{K \subseteq S} (-1)^k \frac{1}{2^{|T \cap K|}} \prod_{i \in T \setminus K} z_i = \sum_{T' \subseteq T} (-1)^{t'} \frac{1}{2^{t'}} \prod_{i \in T \setminus T'} z_i \sum_{K' \subseteq S \setminus T} (-1)^{k'}.$$

Observing that $\sum_{K' \subseteq S \setminus T} (-1)^{k'} = 0$ except if $S \setminus T = \emptyset$, it follows that

$$f_S = \sum_{L \subseteq N \setminus S} \frac{1}{2^{|L|}} m^{\mu}(L \cup S) \sum_{T \subseteq S} (-1)^t \frac{1}{2^t} \prod_{i \in S \setminus T} z_i.$$

Observe that

$$\sum_{T \subseteq S} (-1)^t \frac{1}{2^t} \prod_{i \in S \setminus T} z_i = \frac{1}{2^s} \prod_{i \in S} (2z_i - 1),$$

hence we finally get by (9):

$$f_S = (-1)^s \prod_{i \in S} (2z_i - 1) \widehat{\mu}(S).$$

Note that this expression is also true for $S = \emptyset$ as

$$f_{\emptyset} = \int_{[0,1]^N} f(z) \, dz = \sum_{T \subseteq N} m^{\mu}(T) \int_{[0,1]^N} \prod_{i \in T} z_i \, dz = \sum_{T \subseteq N} \frac{m^{\mu}(T)}{2^{|T|}} = \widehat{\mu}(\emptyset).$$

We obtain finally

$$\mathbb{E}[f_S^2] = \int_{[0,1]^S} \left(\prod_{i \in S} (2z_i - 1) \widehat{\mu}(S) \right)^2 \, dz_S = \frac{1}{3^s} (\widehat{\mu}(S))^2.$$

□

This result is not very surprising as, historically, Sobol' generalized the Fourier base to obtain the decomposition underlying the Sobol' indices [18].

Therefore, up to a multiplicative constant depending on the cardinality of the subset, the Sobol' indices are the coefficients of the square of the Fourier transform, or, due to (10), of the square of the Banzhaf transform (compare with (6)).

The computation of the Sobol' indices for a general Choquet integral is quite complex due to the presence of the minimum operator in (2), and there does not seem to be a compact and appealing expression. We restrict ourself to a sub-class of capacities – called 2 additive – where all Möbius coefficient of cardinality strictly greater than 2 are zero [5]. The Choquet integral of such capacities therefore becomes (see (2))

$$f^{\text{Lo}}(z) = \sum_{i \in N} m(i) z_i + \sum_{\{i,j\} \subseteq N} m(i,j) z_i \wedge z_j$$

with the notation $m(i) = m(\{i\})$ and $m(i,j) = m(\{i,j\})$, the Möbius transform of the capacity.

Theorem 2. Consider the Lovász extension f_μ^{Lo} of a 2-additive capacity μ with Möbius transform m . Then f_μ^{Lo} is decomposed in the following terms according to ANOVA

$$f_\emptyset^{\text{Lo}} = \sum_{i \in N} \frac{m(i)}{2} + \sum_{\{i,j\} \subseteq N} \frac{m(i,j)}{3} \quad (12)$$

$$f_k^{\text{Lo}}(z_k) = -\frac{m(k,\cdot)}{2} z_k^2 + (m(k) + m(k,\cdot)) z_k - \left(\frac{m(k)}{2} + \frac{m(k,\cdot)}{3} \right) \quad (13)$$

$$f_{p,q}^{\text{Lo}}(z_p, z_q) = m(p,q) \left(-z_p \vee z_q + \frac{z_p^2}{2} + \frac{z_q^2}{2} + \frac{1}{3} \right) \quad (14)$$

where $m(k,\cdot) := \sum_{i \in N \setminus \{k\}} m(k,i)$. Moreover, the (nonnormalized) Sobol' index are given by

$$\text{Var}[f_k^{\text{Lo}}] = \frac{m(k)^2}{12} + \frac{m(k,\cdot)^2}{45} + \frac{m(k)m(k,\cdot)}{12} \quad (15)$$

$$\text{Var}[f_{p,q}^{\text{Lo}}] = \frac{m(p,q)^2}{90} \quad (16)$$

Proof.

$$f_\emptyset^{\text{Lo}} = \sum_{i \in N} m(i) \mathbb{E}[Z_i] + \sum_{\{i,j\} \subseteq N} m(i,j) \mathbb{E}[Z_i \wedge Z_j]$$

where $\mathbb{E}[Z_i \wedge Z_j] = 2 \int_0^1 \int_0^{z_i} z_i \wedge z_j dz_i dz_j = 2 \int_0^1 \int_0^{z_i} z_j dz_i dz_j = 2 \int_0^1 \frac{z_i^2}{2} dz_i = \frac{1}{3}$. Hence (12) is proved.

For $k \in N$,

$$f_k^{\text{Lo}}(z_k) = \sum_{i \in N} m(i) \left(\mathbb{E}[Z_i | Z_k = z_k] - \frac{1}{2} \right) + \sum_{\{i,j\} \subseteq N} m(i,j) \left(\mathbb{E}[Z_i \wedge Z_j | Z_k = z_k] - \frac{1}{3} \right)$$

We observe that $\mathbb{E}[Z_k \wedge Z_i | Z_k = z_k] = \int_0^1 z_k \wedge z_i dz_i = \int_0^{z_k} z_i dz_i + \int_{z_k}^1 z_k dz_i = z_k - \frac{z_k^2}{2}$. Hence

$$\begin{aligned} f_k^{\text{Lo}}(z_k) &= m(k) \left(z_k - \frac{1}{2} \right) + m(k,\cdot) \left(z_k - \frac{z_k^2}{2} - \frac{1}{3} \right) \\ &= -\frac{m(k,\cdot)}{2} z_k^2 + (m(k) + m(k,\cdot)) z_k - \left(\frac{m(k)}{2} + \frac{m(k,\cdot)}{3} \right). \end{aligned}$$

Hence (13) is proved. Then

$$\begin{aligned}\mathbb{E} \left[(f_k^{\text{Lo}})^2 \right] &= \int_0^1 \left[m(k)^2 \left(z_k^2 - z_k + \frac{1}{4} \right) + m(k, \cdot)^2 \left(z_k^2 + \frac{z_k^4}{4} + \frac{1}{9} - z_k^3 - \frac{2z_k}{3} + \frac{z_k^2}{3} \right) \right. \\ &\quad \left. + 2m(k)m(k, \cdot) \left(z_k^2 - \frac{z_k^3}{2} - \frac{z_k}{3} - \frac{z_k}{2} + \frac{z_k^2}{4} + \frac{1}{6} \right) \right] dz_k \\ &= \frac{m(k)^2}{12} + \frac{m(k, \cdot)^2}{45} + \frac{m(k)m(k, \cdot)}{12}\end{aligned}$$

Hence (15) holds.

Consider now

$$\begin{aligned}f_{p,q}^{\text{Lo}}(z_p, z_q) &= m(p, q) (z_p \wedge z_q - \mathbb{E}[Z_p \wedge Z_q | Z_p = z_p] - \mathbb{E}[Z_p \wedge Z_q | Z_q = z_q] + \mathbb{E}[Z_p \wedge Z_q]) \\ &= m(p, q) \left(z_p \wedge z_q - \left(z_p - \frac{z_p^2}{2} \right) - \left(z_q - \frac{z_q^2}{2} \right) + \frac{1}{3} \right) \\ &= m(p, q) \left(-z_p \vee z_q + \frac{z_p^2}{2} + \frac{z_q^2}{2} + \frac{1}{3} \right)\end{aligned}$$

as $z_p \wedge z_q + z_p \vee z_q = z_p + z_q$. Hence (14) is proved. Then

$$\begin{aligned}\mathbb{E} \left[(f_{p,q}^{\text{Lo}})^2 \right] &= m(p, q)^2 \int_0^1 \int_0^1 \left(-z_p \vee z_q + \frac{z_p^2}{2} + \frac{z_q^2}{2} + \frac{1}{3} \right)^2 dz_p dz_q \\ &= m(p, q)^2 \int_0^1 \int_0^1 \left((z_p \vee z_q)^2 - (z_p \vee z_q) \left(z_p^2 + z_q^2 + \frac{2}{3} \right) \right) dz_p dz_q \\ &\quad + m(p, q)^2 \int_0^1 \int_0^1 \left(\frac{z_p^4}{4} + \frac{z_q^4}{4} + \frac{z_p^2 z_q^2}{2} + \frac{1}{3} (z_p^2 + z_q^2) + \frac{1}{9} \right) dz_p dz_q \\ &= 2 m(p, q)^2 \int_0^1 \int_0^{z_p} \left(z_p^2 - z_p \left(z_p^2 + z_q^2 + \frac{2}{3} \right) \right) dz_p dz_q \\ &\quad + m(p, q)^2 \left(\frac{1}{20} + \frac{1}{20} + \frac{1}{18} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} \right) \\ &= 2 m(p, q)^2 \left(\frac{1}{4} - \frac{1}{5} - \frac{1}{15} - \frac{2}{9} \right) + m(p, q)^2 \left(\frac{1}{10} + \frac{1}{18} + \frac{1}{3} \right) = \frac{m(p, q)^2}{90}\end{aligned}$$

Hence (16) is proved. □

The interaction indices for the two-additive model are, using (3)

$$\begin{aligned}I^\mu(k) &= m(k) + \frac{m(k, \cdot)}{2} \\ I^\mu(p, q) &= m(p, q)\end{aligned}$$

There is a clear difference between the Sobol' and the interaction indices, due to the presence of the square in the definition of the Sobol' indices. We observe that $I^\mu(p, q)$ and $\text{Var}[f_{p,q}^{\text{Lo}}]$ are both proportional to $m(p, q)$ or its square. By contrast, $\text{Var}[f_k^{\text{Lo}}]$ is not proportional to the square of $I^\mu(k)$. Term $m(k)$ takes more importance in $\text{Var}[f_k^{\text{Lo}}]$.

The Sobol' and interaction indices are both based on sensitivity analysis, but performed in a different manner. The interaction indices consider the average value of the partial derivative of f^{Lo} w.r.t. its components (see (7)), and is thus the mean value of a local sensitivity analysis. By contrast, the Sobol' indices comes from a sensitivity analysis based on variance. They are

used for instance to identify which factors shall be fixed (the other variables being unknown and governed by uniform distribution) in order to reduce as much as possible the variance on the output variable.

It is not clear how to compare the interaction indices for singletons and pairs. For instance, if $I^\mu(1) = 2 I^\mu(1, 2)$, does it mean that the importance of criterion 1 is twice more important than the interaction between variables 1 and 2? Such a comparison is possible with the Sobol' indices as the variance of the output variable is decomposed into the variance of each variable individually, each pair of variables, and so on. This allows to compare the Sobol' indices for different terms. Then if $\text{Var}[f_1^{\text{Lo}}] = 2 \text{Var}[f_{1,2}^{\text{Lo}}]$, then one can say that variable 1 alone is twice more influential than the interaction between variables 1 and 2.

References

- [1] S. Angilella, S. Greco, F. Lamantia, and B. Matarazzo. Assessing non-additive utility for multicriteria decision aid. *European Journal of Operational Research*, 158:734–744, 2004.
- [2] L. Berrah, G. Mauris, and J. Montmain. Monitoring the improvement of an overall industrial performance based on a Choquet integral aggregation. *Omega*, 36(3):340 – 351, 2008.
- [3] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953.
- [4] R. A. Fisher and W. A. Mackenzie. The manurial response of different potato varieties. *J. of Agricultural Science*, 13:311–320, 1923.
- [5] M. Grabisch. k -order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92:167–189, 1997.
- [6] M. Grabisch, J. Duchêne, F. Lino, and P. Perny. Subjective evaluation of discomfort in sitting position. *Fuzzy Optimization and Decision Making*, 1(3):287–312, 2002.
- [7] M. Grabisch and Ch. Labreuche. A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. *Annals of Operations Research*, 175:247–286, 2010. doi: 10.1007/s10479-009-0655-8.
- [8] M. Grabisch, J.-L. Marichal, and M. Roubens. Equivalent representations of set functions. *Mathematics of Operations Research*, 25(2):157–178, 2000.
- [9] P. L. Hammer and S. Rudeanu. *Boolean Methods in Operations Research and Related Areas*. Springer, 1968.
- [10] R. L. Keeney and H. Raiffa. *Decision with Multiple Objectives*. Wiley, New York, 1976.
- [11] L. Lovász. Submodular functions and convexity. In A. Bachem, M. Grötschel, and B. Korte, editors, *Mathematical programming. The state of the art*, pages 235–257. Springer Verlag, 1983.
- [12] S. Madélnat, Ch. Labreuche, J. Le Noir, and G. Gailliard. Comparing several candidate architectures variants: An industrial case study. In *8th European Congress on Embedded Real Time Software and Systems (ERTS2)*, Toulouse, France, January 2016.
- [13] G. Owen. Multilinear extensions of games. *Management Sci.*, 18:64–79, 1972.

- [14] G. C. Rota. On the foundations of combinatorial theory I. Theory of Möbius functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 2:340–368, 1964.
- [15] M. Roubens. Interaction between criteria and definition of weights in MCDA problems. In *44th Meeting of the European Working Group "Multicriteria Aid for Decisions"*, Brussels, Belgium, October 1996.
- [16] L. S. Shapley. A value for n -person games. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games, Vol. II*, number 28 in Annals of Mathematics Studies, pages 307–317. Princeton University Press, 1953.
- [17] I. M. Sobol'. On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe Modelirovanie*, 2:112–118, 1990. (in russian).
- [18] I. M. Sobol'. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, 1:407–414, 1993.
- [19] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.