Measuring the Interactions among Variables of Functions over the Unit Hypercube

Jean-Luc Marichal and Pierre Mathonet

Mathematics Research Unit, FSTC, University of Luxembourg,
6, rue Coudenhove-Kalergi, L-1359 Luxembourg,
Grand Duchy of Luxembourg
jean-luc.marichal@uni.lu, pierre.mathonet@uni.lu

Abstract. By considering a least squares approximation of a given square integrable function $f\colon [0,1]^n \to \mathbb{R}$ by a multilinear polynomial of a specified degree, we define an index which measures the overall interaction among variables of f. This definition extends the concept of Banzhaf interaction index introduced in cooperative game theory. Our approach is partly inspired from multilinear regression analysis, where interactions among the independent variables are taken into consideration. We show that this interaction index has appealing properties which naturally generalize the properties of the Banzhaf interaction index. In particular, we interpret this index as an expected value of the difference quotients of f or, under certain natural conditions on f, as an expected value of the derivatives of f. These interpretations show a strong analogy between the introduced interaction index and the overall importance index defined by Grabisch and Labreuche [7]. Finally, we discuss a few applications of the interaction index.

1 Introduction

Sophisticated mathematical models are extensively used in a variety of areas of mathematics and physics, and especially in applied fields such as engineering, life sciences, economics, finance, and many others. Here we consider the simple situation where the model aims at explaining a single dependent variable, call it y, in terms of n independent variables $x_1, \ldots, x_n$. Such a model is usually described through an equation of the form

$$y = f(x_1, \ldots, x_n),$$

where f is a real function of n variables.

Now, suppose that the function f describing the model is given and that we want to investigate its behavior through simple terms. For instance, suppose we want to measure the overall contribution (importance or influence) of each independent variable to the model. A natural approach to this problem consists in defining the overall importance of each variable as the coefficient of this variable

V. Torra, Y. Narukawa, and M. Daumas (Eds.): MDAI 2010, LNAI 6408, pp. 19–30, 2010.

in the least squares linear approximation of f. This approach was considered by Hammer and Holzman [11] for pseudo-Boolean functions and cooperative games $f\colon \{0,1\}^n \to \mathbb{R}$. Interestingly enough, they observed that the coefficient of each variable in the linear approximation is exactly the Banzhaf power index [2,5] of the corresponding player in the game f.

In many practical situations, the information provided by the overall importance degree of each variable may be far insufficient due to the possible interactions among the variables. Then, a more flexible approach to investigate the behavior of f consists in measuring an overall importance degree for each combination (subset) of variables. Such a concept was first introduced in [13] for Boolean functions $f\colon \{0,1\}^n \to \{0,1\}$ (see also [3,4]), then in [14] for pseudo-Boolean functions and games $f\colon \{0,1\}^n \to \mathbb{R}$ (see also [15]), and in [7] for square integrable functions $f\colon [0,1]^n \to \mathbb{R}$.

In addition to these importance indexes, we can also measure directly the interaction degree among the variables by defining an overall interaction index for each combination of variables. This concept was introduced axiomatically in [10] (see also [6]) for games $f\colon \{0,1\}^n \to \mathbb{R}$. However, it has not yet been extended to real functions defined on $[0,1]^n$. In this paper we intend to fill this gap by defining and investigating an appropriate index to measure the interaction degree among variables of a given square integrable function $f\colon [0,1]^n \to \mathbb{R}$. Our sources of inspiration to define such an index are actually threefold:

In cooperative game theory. Interaction indexes were introduced axiomatically a decade ago [10] for games $f\colon \{0,1\}^n \to \mathbb{R}$ (see also [6]). The best known interaction indexes are the Banzhaf and Shapley interaction indexes, which extend the Banzhaf and Shapley power indexes. Following Hammer and Holzman's approach [11], it was shown in [9] that the Banzhaf interaction index can be obtained from least squares approximations of the game under consideration by games whose multilinear representations are of lower degrees.

In analysis. Considering a sufficiently differentiable real function f of several variables, the *local* interaction among certain variables at a given point $\mathbf{a}$ can be obtained through the coefficients of the Taylor expansion of f at $\mathbf{a}$, that is, through the coefficients of the *local* polynomial approximation of f at $\mathbf{a}$. By contrast, if we want to define an *overall* interaction index, we naturally have to consider a *global* approximation of f by a polynomial function.

In statistics. Multilinear statistical models have been proposed to take into account the interaction among the independent variables (see for instance [1]): two-way interactions appear as the coefficients of leading terms in quadratic models, three-way interactions appear as the coefficients of leading terms in cubic models, and so forth.

On the basis of these observations, we naturally consider the least squares approximation problem of a given square integrable function $f\colon [0,1]^n \to \mathbb{R}$ by a polynomial of a given degree. As multiple occurrences in combinations of variables are not relevant, we will only consider multilinear polynomial functions. Then, given a subset $S \subseteq \{1,\ldots,n\}$, an index $\mathcal{I}(f,S)$ measuring the interaction among the variables $\{x_i : i \in S\}$ of f is defined as the coefficient of the monomial $\prod_{i\in S} x_i$ in the best approximation of f by a multilinear polynomial of degree at most $|S|$. This definition is given and discussed in Section 2.

In Section 3 we show that this new index has many appealing properties, such as linearity, continuity, and symmetry. In particular, we show that, similarly to the Banzhaf interaction index introduced for games, the index $\mathcal{I}(f, S)$ can be interpreted in a sense as an expected value of the discrete derivative of f in the direction of S (Theorem 2) or, equivalently, as an expected value of the difference quotient of f in the direction of S (Corollary 1). Under certain natural conditions on f, the index can also be interpreted as an expected value of the derivative of f in the direction of S (Proposition 4). These latter results reveal a strong analogy between the interaction index and the overall importance index introduced by Grabisch and Labreuche [7].

In Section 4 we discuss the computation of explicit expressions of the interaction index for certain classes of functions, namely pseudo-multilinear polynomials and discrete Choquet integrals.

We employ the following notation throughout the paper. Let $\mathbb{I}^n$ denote the n-dimensional unit cube $[0,1]^n$. We denote by $F(\mathbb{I}^n)$ the class of all functions $f\colon \mathbb{I}^n \to \mathbb{R}$ and by $L^2(\mathbb{I}^n)$ the subclass of square integrable functions $f\colon \mathbb{I}^n \to \mathbb{R}$ modulo equality almost everywhere. For any $S \subseteq N = \{1, \ldots, n\}$, we denote by $\mathbf{1}_S$ the characteristic vector of S in $\{0,1\}^n$.

2 Interaction Indexes

In this section we first recall the concepts of power and interaction indexes introduced in cooperative game theory and how the Banzhaf index can be obtained from the solution of a least squares approximation problem. Then we show how this approximation problem can be extended to functions in $L^2(\mathbb{I}^n)$ and, from this extension, we introduce an interaction index for such functions.

Recall that a (*cooperative*) *game* on a finite set of players $N = \{1, \ldots, n\}$ is a set function $v\colon 2^N \to \mathbb{R}$ which assigns to each coalition S of players a real number $v(S)$ representing the *worth* of S.[1] Through the usual identification of the subsets of N with the elements of $\{0,1\}^n$, a game $v\colon 2^N \to \mathbb{R}$ can be equivalently described by a pseudo-Boolean function $f\colon \{0,1\}^n \to \mathbb{R}$. The correspondence is given by $v(S) = f(\mathbf{1}_S)$ and

$$f(\mathbf{x}) = \sum_{S\subseteq N} v(S) \prod_{i\in S} x_i \prod_{i\in N\setminus S} (1 - x_i). \tag{1}$$

Equation (1) shows that any pseudo-Boolean function $f\colon \{0,1\}^n \to \mathbb{R}$ can always be represented by a multilinear polynomial of degree at most n (see [12]), which can be further simplified into

$$f(\mathbf{x}) = \sum_{S\subseteq N} a(S) \prod_{i\in S} x_i\,, \tag{2}$$

[1] Usually, the condition $v(\varnothing) = 0$ is required for v to define a game. However, we do not need this restriction in the present paper.

where the set function $a\colon 2^N \to \mathbb{R}$, called the *Möbius transform* of v, is defined by

$$a(S) = \sum_{T \subseteq S} (-1)^{|S|-|T|}\, v(T).$$

Let $\mathcal{G}^N$ denote the set of games on N. A *power index* [17] on N is a function $\phi\colon \mathcal{G}^N \times N \to \mathbb{R}$ that assigns to every player $i \in N$ in a game $f \in \mathcal{G}^N$ his/her prospect $\phi(f,i)$ from playing the game. An *interaction index* [10] on N is a function $I\colon \mathcal{G}^N \times 2^N \to \mathbb{R}$ that measures in a game $f \in \mathcal{G}^N$ the interaction degree among the players of a coalition $S \subseteq N$.

For instance, the *Banzhaf interaction index* [10] of a coalition $S \subseteq N$ in a game $f \in \mathcal{G}^N$ can be defined (in terms of the Möbius transformation of f) by

$$I_{\mathrm{B}}(f,S) = \sum_{T \supseteq S} \Big(\frac{1}{2}\Big)^{|T|-|S|} a(T), \tag{3}$$

and the *Banzhaf power index* [5] of a player $i \in N$ in a game $f \in \mathcal{G}^N$ is defined by $\phi_{\mathrm{B}}(f,i) = I_{\mathrm{B}}(f,\{i\})$.

It is noteworthy that $I_{\mathrm{B}}(f,S)$ can be interpreted as an average of the S-*difference* (or *discrete S-derivative*) $\Delta^S f$ of f. Indeed, it also writes (see [9, §2])

$$I_{\mathrm{B}}(f,S) = \frac{1}{2^n} \sum_{\mathbf{x} \in \{0,1\}^n} (\Delta^S f)(\mathbf{x}), \tag{4}$$

where $\Delta^S f$ is defined inductively by $\Delta^{\varnothing} f = f$ and $\Delta^S f = \Delta^{\{i\}} \Delta^{S\setminus\{i\}} f$ for $i \in S$, with $\Delta^{\{i\}} f(\mathbf{x}) = f(\mathbf{x} \mid x_i = 1) - f(\mathbf{x} \mid x_i = 0)$.

We now recall how the Banzhaf interaction index can be obtained from a least squares approximation problem. For $k \in \{0,\ldots,n\}$, denote by V_k the set of all multilinear polynomials $g\colon \{0,1\}^n \to \mathbb{R}$ of degree at most k, that is of the form

$$g(\mathbf{x}) = \sum_{\substack{S \subseteq N \\ |S| \leqslant k}} c(S) \prod_{i \in S} x_i\,, \tag{5}$$

where the coefficients $c(S)$ are real numbers. For a given pseudo-Boolean function $f\colon \{0,1\}^n \to \mathbb{R}$, the best kth approximation of f is the unique multilinear polynomial $f_k \in V_k$ that minimizes the distance $\sum_{\mathbf{x} \in \{0,1\}^n} (f(\mathbf{x}) - g(\mathbf{x}))^2$ among all $g \in V_k$. A closed-form expression of f_k was given in [11] for $k = 1$ and $k = 2$ and in [9] for arbitrary $k \leqslant n$. In fact, when f is given in its multilinear form (2) we obtain

$$f_k(\mathbf{x}) = \sum_{\substack{S \subseteq N \\ |S| \leqslant k}} a_k(S) \prod_{i \in S} x_i,$$

where

$$a_k(S) = a(S) + (-1)^{k-|S|} \sum_{\substack{T \supseteq S \\ |T| > k}} \binom{|T|-|S|-1}{k-|S|} \Big(\frac{1}{2}\Big)^{|T|-|S|} a(T).$$

It is then easy to see that

$$I_{\mathrm{B}}(f,S) = a_{|S|}(S). \tag{6}$$

Thus, $I_{\mathrm{B}}(f,S)$ is exactly the coefficient of the monomial $\prod_{i\in S} x_i$ in the best approximation of f by a multilinear polynomial of degree at most $|S|$.

Taking into account this approximation problem, we now define an interaction index for functions in $L^2(\mathbb{I}^n)$ as follows. Denote by W_k the set of all multilinear polynomials $g\colon \mathbb{I}^n \to \mathbb{R}$ of degree at most k. Clearly, these functions are also of the form (5). For a given function $f \in L^2(\mathbb{I}^n)$, we define the *best kth (multilinear) approximation of f* as the multilinear polynomial $f_k \in W_k$ that minimizes the distance

$$\int_{\mathbb{I}^n} \big(f(\mathbf{x}) - g(\mathbf{x})\big)^2\, d\mathbf{x} \tag{7}$$

among all $g \in W_k$.

It is easy to see that W_k is a linear subspace of $L^2(\mathbb{I}^n)$ of dimension $\sum_{s=0}^{k}\binom{n}{s}$. Indeed, W_k is the linear span of the basis $B_k = \{v_S : S \subseteq N,\ |S| \leqslant k\}$, where the functions $v_S\colon \mathbb{I}^n \to \mathbb{R}$ are defined by $v_S(\mathbf{x}) = \prod_{i\in S} x_i$. Note that formula (7) also writes $\|f-g\|^2$ where $\|\cdot\|$ is the standard norm of $L^2(\mathbb{I}^n)$ associated with the inner product $\langle f,g\rangle = \int_{\mathbb{I}^n} f(\mathbf{x})g(\mathbf{x})\,d\mathbf{x}$. Therefore, using the general theory of Hilbert spaces, the solution of this approximation problem exists and is uniquely determined by the orthogonal projection of f onto W_k. This projection can be easily expressed in any orthonormal basis of W_k. But here it is very easy to see that the set $B'_k = \{w_S : S \subseteq N,\ |S| \leqslant k\}$, where $w_S\colon \mathbb{I}^n \to \mathbb{R}$ is given by

$$w_S(\mathbf{x}) = 12^{|S|/2} \prod_{i\in S}\Big(x_i - \frac{1}{2}\Big) = 12^{|S|/2} \sum_{T\subseteq S}\Big(-\frac{1}{2}\Big)^{|S|-|T|} v_T(\mathbf{x}),$$

forms such an orthonormal basis for W_k (actually, this basis can be obtained from B_k via Gram Schmidt orthogonalization).

The following immediate theorem gives the components of the best kth approximation of a function $f \in L^2(\mathbb{I}^n)$ in the bases B_k and B'_k.

Theorem 1. *For every $k \in \{0,\ldots,n\}$, the best kth approximation of $f \in L^2(\mathbb{I}^n)$ is the function*

$$f_k = \sum_{\substack{T\subseteq N\\ |T|\leqslant k}} \langle f, w_T\rangle\, w_T \;=\; \sum_{\substack{S\subseteq N\\ |S|\leqslant k}} a_k(S)\, v_S\,, \tag{8}$$

where

$$a_k(S) = \sum_{\substack{T\supseteq S\\ |T|\leqslant k}} \Big(-\frac{1}{2}\Big)^{|T|-|S|} 12^{|T|/2}\, \langle f, w_T\rangle. \tag{9}$$

By analogy with (6), to measure the interaction degree among variables of an arbitrary function $f \in L^2(\mathbb{I}^n)$, we naturally define an index $\mathcal{I}\colon L^2(\mathbb{I}^n)\times 2^N \to \mathbb{R}$ as $\mathcal{I}(f,S) = a_{|S|}(S)$, where $a_{|S|}(S)$ is obtained from f by (9). We will see in the next section that this index indeed measures an importance degree when $|S| = 1$ and an interaction degree when $|S| \geqslant 2$.

Definition 1. *Let $\mathcal{I}\colon L^2(\mathbb{I}^n) \times 2^N \to \mathbb{R}$ be defined as $\mathcal{I}(f,S) = 12^{|S|/2}\langle f, w_S\rangle$, that is,*

$$\mathcal{I}(f,S) = 12^{|S|} \int_{\mathbb{I}^n} f(\mathbf{x}) \prod_{i\in S}\Big(x_i - \frac{1}{2}\Big)\, d\mathbf{x}. \tag{10}$$

Thus we have defined an interaction index from an approximation (projection) problem. Conversely, this index characterizes this approximation problem. Indeed, as the following result shows, the best kth approximation of $f \in L^2(\mathbb{I}^n)$ is the unique function of W_k that preserves the interaction index for all the s-subsets such that $s \leqslant k$. The discrete analogue of this result was established in [9] for the Banzhaf interaction index (3).

Proposition 1. *A function $f_k \in W_k$ is the best kth approximation of $f \in L^2(\mathbb{I}^n)$ if and only if $\mathcal{I}(f,S) = \mathcal{I}(f_k,S)$ for all $S \subseteq N$ such that $|S| \leqslant k$.*

3 Properties and Interpretations

Most of the interaction indexes defined for games, including the Banzhaf interaction index, share a set of fundamental properties such as linearity, symmetry, and k-monotonicity (see [6]). Many of them can also be expressed as expected values of the discrete derivatives (differences) of their arguments (see for instance (4)). In this section we show that the index $\mathcal{I}$ fulfills direct generalizations of these properties to the framework of functions of $L^2(\mathbb{I}^n)$. In particular, we show that $\mathcal{I}(f,S)$ can be interpreted as an expected value of the difference quotient of f in the direction of S or, under certain natural conditions on f, as an expected value of the derivative of f in the direction of S.

The first result follows from the very definition of the index.

Proposition 2. *For every $S \subseteq N$, the mapping $f \mapsto \mathcal{I}(f,S)$ is linear and continuous.*

Recall that if π is a permutation on N, then, for every function $f \in F(\mathbb{I}^n)$, the permutation π acts on f by $\pi(f)(x_1,\ldots,x_n) = f(x_{\pi(1)},\ldots,x_{\pi(n)})$. The following result is then an easy consequence of the change of variables theorem.

Proposition 3. *The index $\mathcal{I}$ is symmetric. That is, for every permutation π on N, every $f \in L^2(\mathbb{I}^n)$, and every $S \subseteq N$, we have $\mathcal{I}(\pi(f),\pi(S)) = \mathcal{I}(f,S)$.*

We now provide an interpretation of $\mathcal{I}(f,S)$ as an expected value of the S-derivative $D^S f$ of f. The proof immediately follows from repeated integrations by parts of (10) and thus is omitted.

For $S \subseteq N$, denote by h_S the probability density function of independent beta distributions on $\mathbb{I}^n$ with parameters $\alpha = \beta = 2$, that is, $h_S(\mathbf{x}) = 6^{|S|}\prod_{i\in S} x_i$ $(1-x_i)$.

Proposition 4. *For every $S \subseteq N$ and every $f \in L^2(\mathbb{I}^n)$ such that $D^T f$ is continuous and integrable on $]0,1[^n$ for all $T \subseteq S$, we have*

$$\mathcal{I}(f,S) = \int_{\mathbb{I}^n} h_S(\mathbf{x})\, D^S f(\mathbf{x})\, d\mathbf{x}. \tag{11}$$

Remark 1. (a) Formulas (4) and (11) show a strong analogy between the indexes I_{B} and $\mathcal{I}$. Indeed, $I_{\mathrm{B}}(f,S)$ is the expected value of the S-difference of f with respect to the discrete uniform distribution whereas $\mathcal{I}(f,S)$ is the expected value of the S-derivative of f with respect to a beta distribution. We will see in Theorem 2 a similar interpretation of $\mathcal{I}(f,S)$ which does not require all the assumptions of Proposition 4.

(b) Propositions 1 and 4 reveal an analogy between least squares approximations and Taylor expansion formula. Indeed, while the k-degree Taylor expansion of f at a given point $\mathbf{a}$ can be seen as the unique polynomial of degree at most k whose derivatives at $\mathbf{a}$ coincide with the derivatives of f at the same point, the best kth approximation of f is the unique multilinear polynomial of degree at most k that agrees with f in all average S-derivatives for $|S| \leqslant k$.

We now give an alternative interpretation of $\mathcal{I}(f,S)$ as an expected value, which does not require the additional assumptions of Proposition 4. In this more general framework, we naturally replace the derivative with a difference quotient. To this extent, we introduce some further notation. As usual, we denote by $\mathbf{e}_i$ the ith vector of the standard basis for $\mathbb{R}^n$. For every $S \subseteq N$ and every $\mathbf{h} \in \mathbb{I}^n$, we define the S-*shift* operator $E_{\mathbf{h}}^S$ on $F(\mathbb{I}^n)$ by

$$E_{\mathbf{h}}^S f(\mathbf{x}) = f\Bigg(\mathbf{x} + \sum_{j\in S} h_j \mathbf{e}_j\Bigg)$$

for every $\mathbf{x} \in \mathbb{I}^n$ such that $\mathbf{x} + \mathbf{h} \in \mathbb{I}^n$.

We also define the S-*difference* (or *discrete S-derivative*) operator $\Delta_{\mathbf{h}}^S$ on $F(\mathbb{I}^n)$ inductively by $\Delta_{\mathbf{h}}^{\varnothing} f = f$ and $\Delta_{\mathbf{h}}^S f = \Delta_{\mathbf{h}}^{\{i\}} \Delta_{\mathbf{h}}^{S\setminus\{i\}} f$ for $i \in S$, with $\Delta_{\mathbf{h}}^{\{i\}} f(\mathbf{x}) = E_{\mathbf{h}}^{\{i\}} f(\mathbf{x}) - f(\mathbf{x})$. Similarly, we define the S-*difference quotient* operator $Q_{\mathbf{h}}^S$ on $F(\mathbb{I}^n)$ by $Q_{\mathbf{h}}^{\varnothing} f = f$ and $Q_{\mathbf{h}}^S f = Q_{\mathbf{h}}^{\{i\}} Q_{\mathbf{h}}^{S\setminus\{i\}} f$ for $i \in S$, with $Q_{\mathbf{h}}^{\{i\}} f(\mathbf{x}) = \frac{1}{h_i} \Delta_{\mathbf{h}}^{\{i\}} f(\mathbf{x})$.

The next straightforward lemma provides a direct link between the difference operators and the shift operators. It actually shows that, for every fixed $\mathbf{h} \in \mathbb{I}^n$, the map $S \mapsto \Delta_{\mathbf{h}}^S$ is nothing other than the Möbius transform of the map $S \mapsto E_{\mathbf{h}}^S$.

Lemma 1. *For every $f \in F(\mathbb{I}^n)$ and every $S \subseteq N$, we have*

$$\Delta_{\mathbf{h}}^S f(\mathbf{x}) = \sum_{T\subseteq S} (-1)^{|S|-|T|} \, E_{\mathbf{h}}^T f(\mathbf{x}). \tag{12}$$

Let us interpret the S-difference operator through a simple example. For $n = 3$ and $S = \{1,2\}$, we have

$$\Delta_{\mathbf{h}}^S f(\mathbf{x}) = f(x_1+h_1, x_2+h_2, x_3) - f(x_1+h_1, x_2, x_3) - f(x_1, x_2+h_2, x_3) + f(x_1, x_2, x_3).$$

In complete analogy with the discrete concept of marginal interaction among players in a coalition $S \subseteq N$ (see [9, §2]), the value $\Delta_{\mathbf{h}}^S f(\mathbf{x})$ can be interpreted as the *marginal interaction* among variables x_i $(i \in S)$ at $\mathbf{x}$ with respect to the increases h_i for $i \in S$.

Setting $\mathbf{h} = \mathbf{y} - \mathbf{x}$ in the example above, we obtain

$$\Delta^S_{\mathbf{y}-\mathbf{x}} f(\mathbf{x}) = f(y_1, y_2, x_3) - f(y_1, x_2, x_3) - f(x_1, y_2, x_3) + f(x_1, x_2, x_3).$$

If $x_i \leqslant y_i$ for every $i \in S$, then $\Delta^S_{\mathbf{y}-\mathbf{x}} f(\mathbf{x})$ is naturally called the *f-volume* of the box $\prod_{i\in S}[x_i, y_i]$. The following straightforward lemma shows that, when $f = v_S$, $\Delta^S_{\mathbf{y}-\mathbf{x}} f(\mathbf{x})$ is exactly the volume of the box $\prod_{i\in S}[x_i, y_i]$.

Lemma 2. *For every $S \subseteq N$, we have $\Delta^S_{\mathbf{y}-\mathbf{x}} v_S(\mathbf{x}) = \prod_{i\in S}(y_i - x_i)$.*

In the remaining part of this paper, the notation $\mathbf{y}_S \in [\mathbf{x}_S, \mathbf{1}]$ means that $y_i \in [x_i, 1]$ for every $i \in S$.

Theorem 2. *For every $f \in L^2(\mathbb{I}^n)$ and every $S \subseteq N$, we have*

$$\mathcal{I}(f, S) = \frac{1}{\mu(S)} \int_{\mathbf{x}\in\mathbb{I}^n} \int_{\mathbf{y}_S\in[\mathbf{x}_S,\mathbf{1}]} \Delta^S_{\mathbf{y}-\mathbf{x}} f(\mathbf{x})\, d\mathbf{y}_S\, d\mathbf{x}, \tag{13}$$

where

$$\mu(S) = \int_{\mathbf{x}\in\mathbb{I}^n} \int_{\mathbf{y}_S\in[\mathbf{x}_S,\mathbf{1}]} \Delta^S_{\mathbf{y}-\mathbf{x}} v_S(\mathbf{x})\, d\mathbf{y}_S\, d\mathbf{x} = 6^{-|S|}.$$

Remark 2. (a) By Lemma 2, we see that $\mathcal{I}(f, S)$ can be interpreted as the average f-volume of the box $\prod_{i\in S}[x_i, y_i]$ divided by its average volume, when $\mathbf{x}$ and $\mathbf{y}_S$ are chosen at random with the uniform distribution.

(b) As already mentioned in Remark 1(a), Theorem 2 appears as a natural generalization of formula (4) (similarly to Proposition 4) in the sense that the marginal interaction $\Delta^S_{\mathbf{h}} f(\mathbf{x})$ at $\mathbf{x}$ is averaged over the whole domain $\mathbb{I}^n$ (instead of its vertices).

(c) We note an analogy between formula (13) and the importance index defined by Grabisch and Labreuche in [7, Theorem 1]. Indeed, up to the normalization constant, this importance index is obtained by replacing in formula (13) the operator $\Delta^S_{\mathbf{y}-\mathbf{x}}$ by $E^S_{\mathbf{y}-\mathbf{x}} - I$. Moreover, when S is a singleton, both operators coincide and so do the normalization constants.

As an immediate consequence of Theorem 2, we have the following interpretation of the index $\mathcal{I}$ as an expected value of the difference quotients of its argument with respect to some probability distribution.

Corollary 1. *For every $f \in L^2(\mathbb{I}^n)$ and every $S \subseteq N$, we have*

$$\mathcal{I}(f, S) = \int_{\mathbf{x}\in\mathbb{I}^n} \int_{\mathbf{y}_S\in[\mathbf{x}_S,\mathbf{1}]} p_S(\mathbf{x}, \mathbf{y}_S)\, Q^S_{\mathbf{y}-\mathbf{x}} f(\mathbf{x})\, d\mathbf{y}_S\, d\mathbf{x},$$

where the function $p_S(\mathbf{x}, \mathbf{y}_S) = 6^{|S|} \prod_{i\in S}(y_i - x_i)$ defines a probability density function on the set $\{(\mathbf{x}, \mathbf{y}_S) : \mathbf{x} \in \mathbb{I}^n, \mathbf{y}_S \in [\mathbf{x}_S, \mathbf{1}]\}$.

Let us now analyze the behavior of the interaction index $\mathcal{I}$ on some special classes of functions. The following properties generalize in a very natural way to our setting the behavior of the Banzhaf interaction index I_B with respect to the presence of null players and dummy coalitions.

Recall that a null player in a game (or a set function) $v \in \mathcal{G}^N$ is a player $i \in N$ such that $v(T \cup \{i\}) = v(T)$ for every $T \subseteq N \setminus \{i\}$. Equivalently, the corresponding pseudo-Boolean function $f\colon \{0,1\}^n \to \mathbb{R}$, given by (1), is independent of x_i. The notion of null player for games is then naturally extended through the notion of ineffective variables for functions in $F(\mathbb{I}^n)$ as follows. A variable x_i ($i \in N$) is said to be *ineffective* for a function f in $F(\mathbb{I}^n)$ if $f(\mathbf{x}) = E^{\{i\}}_{-\mathbf{x}} f(\mathbf{x})$ for every $\mathbf{x} \in \mathbb{I}^n$, or equivalently, if $\Delta^{\{i\}}_{\mathbf{y}-\mathbf{x}} f(\mathbf{x}) = 0$ for every $\mathbf{x}, \mathbf{y} \in \mathbb{I}^n$.

Define $I_f = \{i \in N : x_i \text{ ineffective for } f\}$. From either (10) or (13), we immediately derive the following result, which states that any combination of variables containing at least one ineffective variable for a function $f \in L^2(\mathbb{I}^n)$ has necessarily a zero interaction.

Proposition 5. *For every $f \in L^2(\mathbb{I}^n)$ and every $S \subseteq N$ such that $S \cap I_f \neq \varnothing$, we have $\mathcal{I}(f, S) = 0$.*

We say that a coalition $S \subseteq N$ is *dummy* in a game (or a set function) $v \in \mathcal{G}^N$ if $v(R \cup T) = v(R) + v(T) - v(\varnothing)$ for every $R \subseteq S$ and every $T \subseteq N \setminus S$. This means that $\{S, N \setminus S\}$ forms a partition of N such that, for every coalition $K \subseteq N$, the relative worth $v(K) - v(\varnothing)$ is the sum of the relative worths of its intersections with S and $N \setminus S$. It follows that a coalition S and its complement $N \setminus S$ are simultaneously dummy in any game $v \in \mathcal{G}^N$.

We propose the following extension of this concept.

Definition 2. *We say that a subset $S \subseteq N$ is* dummy *for a function $f \in F(\mathbb{I}^n)$ if $f(\mathbf{x}) = E^S_{-\mathbf{x}} f(\mathbf{x}) + E^{N \setminus S}_{-\mathbf{x}} f(\mathbf{x}) - f(\mathbf{0})$ for every $\mathbf{x} \in \mathbb{I}^n$.*

The following proposition gives an immediate interpretation of this definition.

Proposition 6. *A subset $S \subseteq N$ is* dummy *for a function $f \in F(\mathbb{I}^n)$ if and only if there exist functions $f_S, f_{N \setminus S} \in F(\mathbb{I}^n)$ such that $I_{f_S} \supseteq N \setminus S$, $I_{f_{N \setminus S}} \supseteq S$ and $f = f_S + f_{N \setminus S}$.*

The following result expresses the natural idea that interaction index for subsets that are properly partitioned by a dummy subset must be zero. It is an immediate consequence of Propositions 2, 5, and 6.

Proposition 7. *For every $f \in L^2(\mathbb{I}^n)$, every nonempty subset $S \subseteq N$ that is dummy for f, and every subset $K \subseteq N$ such that $K \cap S \neq \varnothing$ and $K \setminus S \neq \varnothing$, we have $\mathcal{I}(f, K) = 0$.*

4 Applications

We now calculate explicit expressions of the interaction index for two classes of functions, namely pseudo-multilinear polynomials and discrete Choquet integrals.

4.1 Pseudo-multilinear polynomials

As a first application, we derive an explicit expression of the index $\mathcal{I}$ for the class of pseudo-multilinear polynomials, that is, the class of multilinear polynomials with transformed variables.

Definition 3. *We say that a function $f \in L^2(\mathbb{I}^n)$ is a* pseudo-multilinear polynomial *if there exists a multilinear polynomial $g \in F(\mathbb{R}^n)$ and n unary functions $\varphi_1, \ldots, \varphi_n \in L^2(\mathbb{I})$ such that $f(\mathbf{x}) = g(\varphi_1(x_1), \ldots, \varphi_n(x_n))$ for every $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{I}^n$.*

Using expression (5) of multilinear polynomials, we immediately see that any pseudo-multilinear polynomial $f \in L^2(\mathbb{I}^n)$ can be written in the form

$$f(\mathbf{x}) = \sum_{T \subseteq N} a(T) \prod_{i \in T} \varphi_i(x_i).$$

The following result yields an explicit expression of the interaction index for this function in terms of the interaction indexes for the unary functions $\varphi_1, \ldots, \varphi_n$.

Proposition 8. *For every pseudo-multilinear polynomial $f \in L^2(\mathbb{I}^n)$ and every $S \subseteq N$, we have*

$$\mathcal{I}(f, S) = \sum_{T \supseteq S} a(T) \prod_{i \in T \setminus S} \mathcal{I}(\varphi_i, \varnothing) \prod_{i \in S} \mathcal{I}(\varphi_i, \{i\}).$$

Remark 3. Proposition 8 can actually be easily extended to functions of the form

$$f(\mathbf{x}) = \sum_{T \subseteq N} a(T) \prod_{i \in T} \varphi_i^T(x_i),$$

where $\varphi_i^T \in L^2(\mathbb{I})$ for $i = 1, \ldots, n$ and $T \subseteq N$.

An interesting subclass of pseudo-multilinear polynomials is the class of multiplicative functions, that is, functions of the form $f(\mathbf{x}) = \prod_{i=1}^n \varphi_i(x_i)$, where $\varphi_1, \ldots, \varphi_n \in L^2(\mathbb{I})$. For every multiplicative function $f \in L^2(\mathbb{I}^n)$ and every $S \subseteq N$, assuming $\mathcal{I}(f, \varnothing) \neq 0$, the ratio $\mathcal{I}(f, S)/\mathcal{I}(f, \varnothing)$ is also multiplicative in the sense that

$$\frac{\mathcal{I}(f, S)}{\mathcal{I}(f, \varnothing)} = \prod_{i \in S} \frac{\mathcal{I}(\varphi_i, \{i\})}{\mathcal{I}(\varphi_i, \varnothing)}. \tag{14}$$

4.2 The Discrete Choquet Integrals

A *discrete Choquet integral* is a function $f \in F(\mathbb{I}^n)$ of the form

$$f(\mathbf{x}) = \sum_{T \subseteq N} a(T) \min_{i \in T} x_i, \tag{15}$$

where the set function $a\colon 2^N \to \mathbb{R}$ is nondecreasing with respect to set inclusion and such that $a(\varnothing) = 0$ and $\sum_{S\subseteq N} a(S) = 1$.[2] These functions are mainly used in aggregation function theory and decision making. For general background, see for instance [8, Section 5.4].

The following proposition yields an explicit expression of the interaction index for the class of discrete Choquet integrals. We first consider a lemma and recall that the *beta function* is defined, for any integers $p, q > 0$, by

$$B(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1}\,dt = \frac{(p-1)!(q-1)!}{(p+q-1)!}\,.$$

Proposition 9. *If $f \in F(\mathbb{I}^n)$ is of the form (15), then we have*

$$\mathcal{I}(f,S) = 6^{|S|}\sum_{T\supseteq S} a(T)\,B(|S|+1,|T|+1).$$

Remark 4. The map $a \mapsto \mathcal{I}(f,S) = 6^{|S|}\sum_{T\supseteq S} a(T)\,B(|S|+1,|T|+1)$ defines an interaction index, in the sense of [6], that is not a probabilistic index (see [6, Section 3.3]). However, if we normalize this interaction index (with respect to $|S|$) to get a probabilistic index, we actually divide $\mathcal{I}(f,S)$ by $6^{|S|}B(|S|+1,|S|+1)$ and retrieve the index I_M defined in [16].

Acknowledgments

The authors wish to thank Michel Beine, Miguel Couceiro, Paul Gérard, and Samuel Nicolay for fruitful discussions. This research is supported by the internal research project F1R-MTH-PUL-09MRDO of the University of Luxembourg.

References

1. Aiken, L.S., West, S.G.: Multiple Regression: Testing and Interpreting Interactions. Sage Publications, Newbury Park (1991)
2. Banzhaf, J.F.: Weighted voting doesn't work: A mathematical analysis. Rutgers Law Review 19, 317–343 (1965)
3. Ben-Or, M., Linial, N.: Collective coin flipping. In: Randomness and Computation, pp. 91–115. Academic Press, New York (1990); Earlier version: Collective coin flipping, robust voting games and minima of Banzhaf values. In: Proc. 26th IEEE Symposium on the Foundation of Computer Sciences, Portland, pp. 408–416 (1985)
4. Bourgain, J., Kahn, J., Kalai, G., Katznelson, Y., Linial, N.: The influence of variables in product spaces. Isr. J. Math. 77(1-2), 55–64 (1992)
5. Dubey, P., Shapley, L.S.: Mathematical properties of the Banzhaf power index. Math. Oper. Res. 4, 99–131 (1979)
6. Fujimoto, K., Kojadinovic, I., Marichal, J.-L.: Axiomatic characterizations of probabilistic and cardinal-probabilistic interaction indices. Games Econom. Behav. 55(1), 72–99 (2006)

[2] Whether the conditions on the set function a are assumed or not, the function given in (15) is also called the *Lovász extension* of the pseudo-Boolean function $f|_{\{0,1\}^n}$.

7. Grabisch, M., Labreuche, C.: How to improve acts: An alternative representation of the importance of criteria in MCDM. Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 9(2), 145–157 (2001)
8. Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E.: Aggregation functions. Encyclopedia of Mathematics and its Applications, vol. 127. Cambridge University Press, Cambridge (2009)
9. Grabisch, M., Marichal, J.-L., Roubens, M.: Equivalent representations of set functions. Math. Oper. Res. 25(2), 157–178 (2000)
10. Grabisch, M., Roubens, M.: An axiomatic approach to the concept of interaction among players in cooperative games. Int. J. Game Theory 28(4), 547–565 (1999)
11. Hammer, P., Holzman, R.: Approximations of pseudo-Boolean functions; applications to game theory. Z. Oper. Res. 36(1), 3–21 (1992)
12. Hammer, P., Rudeanu, S.: Boolean methods in operations research and related areas. Springer, Heidelberg (1968)
13. Kahn, J., Kalai, G., Linial, N.: The influence of variables on Boolean functions. In: Proc. 29th Annual Symposium on Foundations of Computational Science, pp. 68–80. Computer Society Press (1988)
14. Marichal, J.-L.: The influence of variables on pseudo-Boolean functions with applications to game theory and multicriteria decision making. Discrete Appl. Math. 107(1-3), 139–164 (2000)
15. Marichal, J.-L., Kojadinovic, I., Fujimoto, K.: Axiomatic characterizations of generalized values. Discrete Applied Mathematics 155(1), 26–43 (2007)
16. Marichal, J.-L., Mathonet, P.: Approximations of Lovász extensions and their induced interaction index. Discrete Appl. Math. 156(1), 11–24 (2008)
17. Shapley, L.: A value for n-person games. In: Contributions to the Theory of Games II. Annals of Mathematics Studies, vol. 28, pp. 307–317. Princeton University Press, Princeton (1953)