Weighted Banzhaf power and interaction indexes through weighted approximations of games

Jean-Luc Marichal and Pierre Mathonet

Mathematics Research Unit
University of Luxembourg , Luxembourg-Kirchberg, Luxembourg
jean-luc.marichal@uni.lu, pierre.mathonet@uni.lu

Abstract. In cooperative game theory, various kinds of power indexes are used to measure the influence that a given player has on the outcome of the game or to define a way of sharing the benefits of the game among the players. The best known power indexes are due to Shapley [15, 16] and Banzhaf [1, 5] and there are many other examples of such indexes in the literature.
When one is concerned by the analysis of the behavior of players in a game, the information provided by power indexes might be far insufficient, for instance due to the lack of information on how the players interact within the game. The notion of *interaction index* was then introduced to measure an interaction degree among players in coalitions; see [13, 12, 7, 8, 14, 10, 6] for the definitions and axiomatic characterizations of the Shapley and Banzhaf interaction indexes as well as many others.
In addition to the axiomatic characterizations the Shapley power index and the Banzhaf power and interaction indexes were shown to be solutions of simple least squares approximation problems (see [2] for the Shapley index, [11] for the Banzhaf power index and [9] for the Banzhaf interaction index).
We generalize the non-weighted approach of [11, 9] by adding a weighted, probabilistic viewpoint: A weight $w(S)$ is assigned to every coalition S of players that represents the probability that coalition S forms. The solution of the weighted least squares problem associated with the probability distribution w was given in [3, 4] in the special case when the players behave independently of each other to form coalitions.
In this particular setting we introduce a weighted Banzhaf interaction index associated with w by considering, as in [11, 9], the leading coefficients of the approximations of the game by polynomials of specified degrees. We then study the most important properties of these weighted indexes and their relations with the classical Banzhaf and Shapley indexes.

A *cooperative game* on a finite set of players $N = \{1, \dots, n\}$ 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] Identifying the subsets of N with the elements of $\{0,1\}^n$, we see that a game $v\colon 2^N \to \mathbb{R}$ corresponds to a pseudo-Boolean function $f\colon \{0,1\}^n \to \mathbb{R}$ (the correspondence is given by $v(S) = f(\mathbf{1}_S)$, where $\mathbf{1}_S$ denotes the characteristic vector of

[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 work.

S in $\{0,1\}^n$). We will henceforth use the same symbol to denote both a given pseudo-Boolean function and its underlying set function (game).

Every pseudo-Boolean function $f\colon \{0,1\}^n \to \mathbb{R}$ can be represented by a multilinear polynomial of degree at most n of the form

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

where the set function $a\colon 2^N \to \mathbb{R}$ is the *Möbius transform* of f.

Let $\mathcal{G}^N$ denote the set of games on N. A *power index* [15] 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 $f \in \mathcal{G}^N$ is defined by

$$I_{\mathrm{B}}(f,S) = \sum_{T\supseteq S} \left(\frac{1}{2}\right)^{|T|-|S|} a(T) = \frac{1}{2^{n-|S|}} \sum_{T\subseteq N\setminus S} (\Delta^S f)(T), \tag{1}$$

where the *S-difference* $\Delta^S f$ is defined inductively by $\Delta^{\varnothing} f = f$ and $\Delta^S f = \Delta^{\{i\}}\Delta^{S\setminus\{i\}} f$ for $i \in S$, with $\Delta^{\{i\}} f(\mathbf{x}) = f(\mathbf{x} \mid x_i = 1) - f(\mathbf{x} \mid x_i = 0)$. The *Banzhaf power index* [5] of a player $i \in N$ in a game $f \in \mathcal{G}^N$ is then given by $\phi_{\mathrm{B}}(f,i) = I_{\mathrm{B}}(f,\{i\})$.

Let us now introduce a weighted least squares approximation problem which generalizes the one considered in [11, 9]. 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

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

We also consider a weight function $w\colon \{0,1\}^n \to\,]0,\infty[$. For every pseudo-Boolean function $f\colon \{0,1\}^n \to \mathbb{R}$, we define the *best kth approximation of f* as the unique multilinear polynomial $f_k \in V_k$ that minimizes the squared distance

$$\sum_{\mathbf{x}\in\{0,1\}^n} w(\mathbf{x})\big(f(\mathbf{x}) - g(\mathbf{x})\big)^2 = \sum_{S\subseteq N} w(S)\big(f(S) - g(S)\big)^2 \tag{2}$$

among all functions $g \in V_k$.

Clearly, we can assume without loss of generality that the weights $w(S)$ are (multiplicatively) normalized so that $\sum_{S\subseteq N} w(S) = 1$. We then immediately see that the weights define a probability distribution over 2^N and we can interpret $w(S)$ as the probability that coalition S forms, that is, $w(S) = \Pr(C = S)$, where C denotes a random coalition.

In the special case of equiprobability, the approximation above reduces to standard least squares, and a closed form expression of the approximation f_k of f was given in [11, 9] and it was shown that, writing

$$f_k(\mathbf{x}) = \sum_{\substack{S\subseteq N\\ |S|\leqslant k}} a_k(S) \prod_{i\in S} x_i, \tag{3}$$

we have

$$I_{\mathrm{B}}(f,S) = a_{|S|}(S). \tag{4}$$

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 $|S|$.

Now, suppose that the players behave independently of each other to form coalitions, which means that the events $(C \ni i)$, for $i \in N$, are independent. Under this assumption, the weight function w is completely determined by the vector $\mathbf{p} = (p_1,\ldots,p_n)$, where $p_i = \Pr(C \ni i) = \sum_{S\ni i} w(S)$ (we assume $0 < p_i < 1$), by the formula

$$w(S) = \prod_{i\in S} p_i \prod_{i\in N\setminus S} (1-p_i).$$

In this particular setting, the weighted approximation problem was presented and solved in [3] and [4, Theorem 4] by noticing that the distance in (2) is the natural L^2-distance associated with the measure w, with respect to the inner product

$$\langle f,g\rangle = \sum_{\mathbf{x}\in\{0,1\}^n} w(\mathbf{x})f(\mathbf{x})g(\mathbf{x}),$$

and that the functions

$$v_S\colon \{0,1\}^n \to \mathbb{R}\colon \mathbf{x} \mapsto \prod_{i\in S} \frac{x_i - p_i}{\sqrt{p_i(1-p_i)}}$$

form an orthonormal basis of the vector space of pseudo-Boolean functions.

Using these functions, we immediately obtain that f_k is of the form (3) where

$$a_k(S) = \sum_{\substack{T\supseteq S\\ |T|\leqslant k}} \frac{\prod_{i\in T\setminus S}(-p_i)}{\prod_{i\in T}\sqrt{p_i(1-p_i)}}\,\langle f, v_T\rangle.$$

Using this solution, we define the index by analogy with (4).

Definition 1. *The weighted Banzhaf interaction index associated to w is*

$$I_{\mathrm{B},\mathbf{p}}\colon \mathcal{G}^N \times 2^N \to \mathbb{R} : (f,S) \mapsto I_{\mathrm{B},\mathbf{p}}(f,S) = a_{|S|}(S) = \frac{\langle f, v_S\rangle}{\prod_{i\in S}\sqrt{p_i(1-p_i)}}.$$

Then we show that most of the properties of the standard Banzhaf index can be generalized to the weighted index. For instance, Formula (1) is a particular case of

$$I_{\mathrm{B},\mathbf{p}}(f,S) = \sum_{T\supseteq S} a(T) \prod_{i\in T\setminus S} p_i = \sum_{T\subseteq N\setminus S} p_T^S\,(\Delta^S f)(T),$$

where $p_T^S = \Pr(T \subseteq C \subseteq S\cup T) = \prod_{i\in T} p_i \prod_{i\in (N\setminus S)\setminus T}(1-p_i)$.

This shows that the weighted Banzhaf interaction index belongs to the class of probabilistic interaction indexes introduced in [6], and we can moreover provide a nice interpretation of the probabilities p_T^S as conditional probabilities.

We then analyze the behaviour of the index with respect to null or dummy players or more generally to dummy coalitions, and we show how to compute the weighted Banzhaf index in terms of Owen's multilinear extension $\overline{f}$ of the game f. We also provide conversion formulas between the indexes corresponding to different weights, and show how to recover f from the weighted Banzhaf index.

Finally, we show that the standard Banzhaf index is the average of the weighted Banzhaf indexes over all the possible weights and that the Shapley index is the average of the weighted Banzhaf indexes over all possible symmetric weights.

References

1. J. Banzhaf. Weighted voting doesn't work : A mathematical analysis. *Rutgers Law Review*, 19:317–343, 1965.
2. A. Charnes, B. Golany, M. Keane, and J. Rousseau. Extremal principle solutions of games in characteristic function form: core, Chebychev and Shapley value generalizations. In *Econometrics of planning and efficiency*, volume 11 of *Adv. Stud. Theoret. Appl. Econometrics*, pages 123–133. Kluwer Acad. Publ., Dordrecht, 1988.
3. Guoli Ding, R. F. Lax, Jianhua Chen, and P. P. Chen. Formulas for approximating pseudo-Boolean random variables. *Discrete Appl. Math.*, 156(10):1581–1597, 2008.
4. Guoli Ding, R. F. Lax, Jianhua Chen, and P. P. Chen. Transforms of pseudo-Boolean random variables. *Discrete Appl. Math.*, 158(1):13–24, 2010.
5. P. Dubey and L. S. Shapley. Mathematical properties of the Banzhaf power index. *Math. Oper. Res.*, 4(2):99–131, 1979.
6. K. Fujimoto, I. Kojadinovic, and J.-L. Marichal. Axiomatic characterizations of probabilistic and cardinal-probabilistic interaction indices. *Games Econom. Behav.*, 55(1):72–99, 2006.
7. M. Grabisch. Alternative representations of discrete fuzzy measures for decision making. *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems*, 5(5):587–607, 1997.
8. M. Grabisch. *k*-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92(2):167–189, 1997.
9. M. Grabisch, J.-L. Marichal, and M. Roubens. Equivalent representations of set functions. *Math. Oper. Res.*, 25(2):157–178, 2000.
10. M. Grabisch and M. Roubens. An axiomatic approach to the concept of interaction among players in cooperative games. *Internat. J. Game Theory*, 28(4):547–565, 1999.
11. P. L. Hammer and R. Holzman. Approximations of pseudo-Boolean functions; applications to game theory. *Z. Oper. Res.*, 36(1):3–21, 1992.
12. T. Murofushi and S. Soneda. Techniques for reading fuzzy measures (iii): Interaction index (in Japanese). In *Proceedings of the 9th Fuzzy Systems Symposium, Sapporo, Japan*, pages 693–696, 1993.
13. G. Owen. Multilinear extensions of games. *Management Sci.*, 18:P64–P79, 1971/72.
14. M. Roubens. Interaction between criteria and definition of weights in MCDA problems. In *Proceedings of the 44th Meeting of the European Working Group "Multiple Criteria Decision Aiding"*, pages 693–696, October 1996.
15. 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.
16. L. S. Shapley and M. Shubik. A method for evaluating the distribution of power in a committee system. *American Political Science Review*, 48:787–792, 1954.