Influence and interaction indexes in cooperative games : a unified least squares approach

Jean-Luc Marichal*   Pierre Mathonet**

*University of Luxembourg
**University of Liège

Cooperative games

Set of players

$$N = \{1, \ldots, n\}$$

Game on N

$$f : 2^N \to \mathbb{R}$$

(usually $f(\varnothing) = 0$)

For every coalition $S \subseteq N$,

$$f(S) = \text{the } \textit{worth} \text{ of } S$$

Cooperative games

We can identify

$$S \subseteq N \qquad \text{with} \qquad \mathbf{1}_S \in \{0,1\}^n$$

Example: $N = \{1,2,3\}$

$$S = \{2,3\} \qquad \mathbf{1}_S = (0,1,1)$$

Consequence

A game $f\colon 2^N \to \mathbb{R}$ can also be regarded as a function

$$f\colon \{0,1\}^n \to \mathbb{R}$$

Cooperative games

A game on N can always be represented as a *multilinear polynomial* of degree $\leqslant n$

$$f(\mathbf{x}) = \sum_{T \subseteq N} a(T) \prod_{i \in T} x_i \qquad \mathbf{x} \in \{0,1\}^n$$

Möbius transform $a\colon 2^N \to \mathbb{R}$

$$f(S) = \sum_{T \subseteq S} a(T)$$

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

Cooperative games

Multilinear extension of a game (Owen 1972)

Given a game f on N

$$f(\mathbf{x}) = \sum_{T \subseteq N} a(T) \prod_{i \in T} x_i \qquad \mathbf{x} \in \{0,1\}^n$$

we can define its *multilinear extension*

$$\overline{f}(\mathbf{x}) = \sum_{T \subseteq N} a(T) \prod_{i \in T} x_i \qquad \mathbf{x} \in [0,1]^n$$

Banzhaf power index

Marginal contribution of player $i \in N$ when joining a coalition T

$$\Delta_i f(T) = f(T \cup \{i\}) - f(T) \qquad T \subseteq N \setminus \{i\}$$

Banzhaf power index for player i (Banzhaf 1965)

$$I_B(f, i) = \frac{1}{2^{n-1}} \sum_{T \subseteq N \setminus \{i\}} \Delta_i f(T)$$

Banzhaf power index

Discrete derivative of f (resp. $\overline{f}$) with respect to the ith variable

$$\Delta_i f(\mathbf{x}) = f(\mathbf{x} \mid x_i = 1) - f(\mathbf{x} \mid x_i = 0)$$

$$\Delta_i \overline{f}(\mathbf{x}) = \overline{f}(\mathbf{x} \mid x_i = 1) - \overline{f}(\mathbf{x} \mid x_i = 0)$$

Banzhaf power index for player i

$$I_B(f, i) = \frac{1}{2^n} \sum_{\mathbf{x} \in \{0,1\}^n} \Delta_i f(\mathbf{x})$$

Banzhaf power index

Alternative forms (Owen 1972, Grabisch et al. 2000)

$$I_B(f,i) = \sum_{T \ni i} \left(\frac{1}{2}\right)^{|T|-1} a(T)$$

$$I_B(f,i) = \big(\Delta_i \overline{f}\big)\big(\tfrac{1}{2},\ldots,\tfrac{1}{2}\big)$$

$$I_B(f,i) = \int_{[0,1]^n} \Delta_i \overline{f}(\mathbf{x})\, d\mathbf{x}$$

Banzhaf interaction index

Marginal interaction among players i and j conditioned to the presence of a coalition $T \subseteq N \setminus \{i,j\}$

$$\Delta_{ij}\, f(T) = \underbrace{\Big(f(T \cup \{i,j\}) - f(T \cup \{i\})\Big)}_{\substack{\text{marginal contr. of } j \\ \text{in the presence of } i}} - \underbrace{\Big(f(T \cup \{j\}) - f(T)\Big)}_{\substack{\text{marginal contr. of } j \\ \text{in the absence of } i}}$$

Banzhaf interaction index (Owen 1972)

$$I_B(f, \{i,j\}) = \frac{1}{2^{n-2}} \sum_{T \subseteq N \setminus \{i,j\}} \Delta_{ij}\, f(T)$$

Banzhaf interaction index

In terms of discrete derivatives :

$$\Delta_{ij}\, f(\mathbf{x}) = \Delta_j\, \Delta_i\, f(\mathbf{x})$$

$$\Delta_{ij}\, \overline{f}(\mathbf{x}) = \Delta_j\, \Delta_i\, \overline{f}(\mathbf{x})$$

Banzhaf interaction index

$$I_B(f, \{i, j\}) = \frac{1}{2^n} \sum_{\mathbf{x}\in\{0,1\}^n} \Delta_{ij}\, f(\mathbf{x})$$

Banzhaf interaction index

Alternative forms (Grabisch et al. 2000)

$$I_B(f,\{i,j\}) = \sum_{T\supseteq\{i,j\}} \left(\frac{1}{2}\right)^{|T|-2} a(T)$$

$$I_B(f,\{i,j\}) = \left(\Delta_{ij}\,\overline{f}\right)\left(\tfrac{1}{2},\ldots,\tfrac{1}{2}\right)$$

$$I_B(f,\{i,j\}) = \int_{[0,1]^n} \Delta_{ij}\,\overline{f}(\mathbf{x})\,d\mathbf{x}$$

Banzhaf interaction index

Measure of the average *interaction among players in coalition S* (Roubens 1996)

$$I_B(f,S) = \frac{1}{2^{n-|S|}} \sum_{T \subseteq N \setminus S} \Delta_S f(T)$$

$$I_B(f,S) = \frac{1}{2^n} \sum_{\mathbf{x} \in \{0,1\}^n} \Delta_S f(\mathbf{x})$$

Special case: when $S = \{i\} \longrightarrow$ power index

Banzhaf interaction index

Alternative forms (Grabisch et al. 2000)

$$I_B(f,S) = \sum_{T \supseteq S} \left(\frac{1}{2}\right)^{|T|-|S|} a(T)$$

$$I_B(f,S) = \left(\Delta_S \,\overline{f}\right)\left(\tfrac{1}{2},\ldots,\tfrac{1}{2}\right)$$

$$I_B(f,S) = \int_{[0,1]^n} \Delta_S \,\overline{f}(\mathbf{x})\, d\mathbf{x}$$

S-approximation of a game

Given a game f on N and a coalition $S \subseteq N$, the *best S-approximation of f* is the unique game on S

$$f_S(\mathbf{x}) = \sum_{T \subseteq S} c(T) \prod_{i \in T} x_i$$

that minimizes the square distance

$$\sum_{\mathbf{x} \in \{0,1\}^n} \Big(f(\mathbf{x}) - g(\mathbf{x}) \Big)^2$$

among all games g on S

Theorem (Hammer-Holzman 1992, Grabisch et al. 2000, M.-M. 2011)

$$c(S) \; = \; I_B(f,S)$$

Weighted S-approximation of a game

Toward a weighted interaction index ?

Idea: consider a weighted square distance

$$\sum_{\mathbf{x}\in\{0,1\}^n} w(\mathbf{x})\left(f(\mathbf{x})-g(\mathbf{x})\right)^2$$

where $w\colon \{0,1\}^n \to\,]0,+\infty[$ is a weight function.

Since w is defined up to a multiplicative constant $r>0$, we can assume that

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

$\longrightarrow$ probability distribution over $\{0,1\}^n$

Weighted S-approximation of a game

A possible interpretation of w

Let C denote a random coalition in N

Define

$$w(S) = \Pr(C = S)$$

i.e., the probability that coalition S forms

Independence: Suppose that the players behave independently of each other to form coalitions

This means that the events " $C \ni i$ ", $i \in N$, are <u>independent</u>

Weighted S-approximation of a game

Example: $N = \{1, 2, 3\}$

$$\begin{aligned} w(\{2,3\}) &= \Pr(C = \{2,3\}) \\ &= \Pr(C \not\ni 1)\Pr(C \ni 2)\Pr(C \ni 3) \\ &= (1 - p_1)\, p_2\, p_3 \end{aligned}$$

In general: Introducing $\mathbf{p} = (p_1, \ldots, p_n)$ with

$$p_i = \Pr(C \ni i),$$

we have

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

Weighted Banzhaf interaction index

The *best S-approximation of f* is the unique game on S

$$f_S(\mathbf{x}) = \sum_{T \subseteq S} c(T) \prod_{i \in T} x_i$$

that minimizes the weighted square distance

$$\sum_{\mathbf{x} \in \{0,1\}^n} w(\mathbf{x}) \left(f(\mathbf{x}) - g(\mathbf{x}) \right)^2$$

among all games g on S

Definition

$$I_{B,\mathbf{p}}(f, S) \;=\; c(S)$$

Weighted Banzhaf interaction index

Theorem

We have

$$I_{B,\mathbf{p}}(f,S) \;=\; \sum_{T\subseteq N\setminus S} p_T^S\,\Delta_S f(T)$$

with

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

$$I_{B,\mathbf{p}}(f,S) = \sum_{\mathbf{x}\in\{0,1\}^n} w(\mathbf{x})\,\Delta_S\, f(\mathbf{x})$$

Weighted Banzhaf interaction index

Non-weighted least squares (uniform probability)

$$w(S) = \frac{1}{2^n} \iff p_i = \Pr(C \ni i) = \frac{1}{2}$$

In this special case:

$$\boxed{I_{B,\mathbf{p}}(f,S) = I_B(f,S)}$$

Theorem

We have

$$I_B(f,S) = \int_{[0,1]^n} I_{B,\mathbf{p}}(f,S)\, d\mathbf{p}$$

Weighted Banzhaf interaction index

Alternative forms

$$I_{B,\mathbf{p}}(f,S) = \sum_{T \supseteq S} a(T) \prod_{i \in T \setminus S} p_i$$

$$I_{B,\mathbf{p}}(f,S) = \big(\Delta_S \overline{f}\big)(\mathbf{p})$$

$$I_{B,\mathbf{p}}(f,S) = \int_{[0,1]^n} \Delta_S \overline{f}(\mathbf{x})\; dF_1(x_1) \cdots dF_n(x_n)$$

$$\text{with } p_i = \int_0^1 x\, dF_i(x)$$

Weighted Banzhaf interaction index

Example (majority game): $N = \{1, 2, 3\}$

$$f(x_1, x_2, x_3) = x_1x_2 + x_2x_3 + x_3x_1 - 2x_1x_2x_3$$

We have

$$\begin{array}{ll} f(0,0,0) = 0 & f(1,0,0) = 0 \\ f(1,1,0) = 1 & f(1,1,1) = 1 \end{array}$$

$$\Delta_{12}\, f(x_1, x_2, x_3) = \Delta_2(x_2 + x_3 - 2x_2x_3) = 1 - 2x_3$$

$$I_{B,\mathbf{p}}(f, \{1, 2\}) = \big(\Delta_{12}\, \overline{f}\big)(p_1, p_2, p_3) = 1 - 2p_3$$

Weighted Banzhaf interaction index

Theorem

The map $f \mapsto \{I_{B,\mathbf{p}}(f,S) : S \subseteq N\}$ is a linear bijection

The inverse bijection is given by

$$f(\mathbf{x}) = \sum_{T \subseteq N} I_{B,\mathbf{p}}(f,T) \prod_{i \in T}(x_i - p_i)$$

We also have a conversion formula between $I_{B,\mathbf{p}}(f,\cdot)$ and $I_{B,\mathbf{p}'}(f,\cdot)$ for every $\mathbf{p}'$

$$I_{B,\mathbf{p}'}(f,S) = \sum_{T \supseteq S} I_{B,\mathbf{p}}(f,T) \prod_{i \in T \setminus S}(p'_i - p_i)$$

Weighted Banzhaf interaction index

Define

$$\begin{aligned} E(f) &= \sum_{\mathbf{x}\in\{0,1\}^n} w(\mathbf{x})\, f(\mathbf{x}) \\ \sigma^2(f) &= \sum_{\mathbf{x}\in\{0,1\}^n} w(\mathbf{x})\, \big(f(\mathbf{x}) - E(f)\big)^2 \end{aligned}$$

Theorem

We have

$$\big|I_{B,\mathbf{p}}(f,S)\big| \leqslant \frac{\sigma(f)}{\prod_{i\in S}\sqrt{p_i(1-p_i)}}$$

and the inequality is tight

Banzhaf influence index

Marginal contribution of $\{i,j\}$ when joining a coalition T

$$\sigma_{ij} f(T) = f(T \cup \{i,j\}) - f(T) \qquad T \subseteq N \setminus \{i,j\}$$

Banzhaf influence index for $\{i,j\}$

$$\mathcal{I}_B(f, \{i,j\}) = \frac{1}{2^{n-2}} \sum_{T \subseteq N \setminus \{i,j\}} \sigma_{ij} f(T)$$

Banzhaf influence index

Marginal contribution of S when joining a coalition T

$$\sigma_S f(T) = f(T \cup S) - f(T) \qquad T \subseteq N \setminus S$$

Banzhaf influence index for S

$$\mathcal{I}_B(f,S) = \frac{1}{2^{n-|S|}} \sum_{T \subseteq N \setminus S} \sigma_S f(T)$$

Banzhaf influence index

Define

$$\sigma_S f(\mathbf{x}) = f(\mathbf{x} \mid x_i = 1,\ i \in S) - f(\mathbf{x} \mid x_i = 0,\ i \in S)$$

Banzhaf influence index for S

$$\mathcal{I}_B(f, S) = \frac{1}{2^n} \sum_{\mathbf{x} \in \{0,1\}^n} \sigma_S f(\mathbf{x})$$

Banzhaf influence index

Alternative forms

$$\mathcal{I}_B(f,S) = \sum_{\substack{T\subseteq N \\ T\cap \overline{S}\neq\varnothing}} \Big(\frac{1}{2}\Big)^{|T\setminus S|} a(T)$$

$$\mathcal{I}_B(f,S) = \big(\sigma_S\,\overline{f}\big)\big(\tfrac{1}{2},\ldots,\tfrac{1}{2}\big)$$

$$\mathcal{I}_B(f,S) = \int_{[0,1]^n} \sigma_S\,\overline{f}(\mathbf{x})\,d\mathbf{x}$$

Least squares approach

Let f_S be the *S-approximation* of a game f on N (with a non-weighted distance)

$$f_S(\mathbf{x}) = \sum_{T \subseteq S} c(T) \prod_{i \in T} x_i$$

Theorem

$$\mathcal{I}_B(f, S) \; = \; f_S(N) - f_S(\varnothing)$$

Weighted version of $\mathcal{I}_B$? $\quad \longrightarrow$ use a weighted distance

Weighted least squares

Let f_S be the *S-approximation* of a game f on N
with a distance weighted by a weight function $w\colon 2^N \to]0,+\infty[$,
defined by

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

Definition

$$\mathcal{I}_{B,\mathbf{p}}(f,S) \;=\; f_S(N) - f_S(\varnothing)$$

Weighted Banzhaf influence index

Theorem

We have

$$\mathcal{I}_{B,\mathbf{p}}(f,S) \;=\; \sum_{T\subseteq N\setminus S} p_T^S\,\sigma_S f(T)$$

with

$$p_T^S \;=\; \cdots \qquad \text{(as before)}$$

$$\mathcal{I}_{B,\mathbf{p}}(f,S) = \sum_{\mathbf{x}\in\{0,1\}^n} w(\mathbf{x})\,\sigma_S\, f(\mathbf{x})$$

Weighted Banzhaf influence index

Non-weighted least squares (uniform probability)

$$w(S) \;=\; \frac{1}{2^n} \qquad \Longleftrightarrow \qquad p_i \;=\; \frac{1}{2}$$

In this special case:

$$\boxed{\mathcal{I}_{B,\mathbf{p}}(f,S) = \mathcal{I}_B(f,S)}$$

Theorem

We have

$$\mathcal{I}_B(f,S) \;=\; \int_{[0,1]^n} \mathcal{I}_{B,\mathbf{p}}(f,S)\, d\mathbf{p}$$

Weighted Banzhaf influence index

Alternative forms

$$\mathcal{I}_{B,\mathbf{p}}(f,S) = \sum_{\substack{T\subseteq N\\ T\cap S\neq\varnothing}} a(T) \prod_{i\in T\setminus S} p_i$$

$$\mathcal{I}_{B,\mathbf{p}}(f,S) = \big(\sigma_S\,\overline{f}\big)(\mathbf{p})$$

$$\mathcal{I}_{B,\mathbf{p}}(f,S) = \int_{[0,1]^n} \sigma_S\,\overline{f}(\mathbf{x})\;dF_1(x_1)\cdots dF_n(x_n)$$

$$\text{with } p_i = \int_0^1 x\,dF_i(x)$$

Weighted Banzhaf influence index

Can we reconstruct the game from the influence index ?

We have

$$\mathcal{I}_B(f,S) = \sum_{\substack{T\subseteq S\\ |T| \text{ odd}}} \left(\frac{1}{2}\right)^{|T|-1} I_B(f,T)$$

Weighted Banzhaf influence index

When $|S|$ is even, we have

$$\mathcal{I}_B(f,S) = -\sum_{T\subsetneq S} E_{|S|-|T|}(0)\, 2^{|S|-|T|}\, \mathcal{I}_B(f,T)$$

where E_n is the nth Euler polynomial
(with $E_n(0) = 0$ for even $n > 1$)

The "even" influences can be obtained from the "odd" influences

$\Longrightarrow$ The map $f \mapsto \{\mathcal{I}_B(f,S) : S \subseteq N\}$ is not a bijection
(half of the information is lost)

Weighted Banzhaf influence index

In the weighted case: We have $\sigma_{\varnothing}\overline{f} \equiv 0$ and hence

$$\boxed{\mathcal{I}_{B,\mathbf{p}}(f,\varnothing) = 0}$$

$\Longrightarrow$ The map $f \mapsto \{\mathcal{I}_{B,\mathbf{p}}(f,S) : S \subseteq N\}$ is <u>not</u> a bijection (still a piece of the information is lost)

Weighted Banzhaf influence index

However... we can reconstruct $I_{B,\mathbf{p}}(f,S)$ whenever

$$\prod_{i\in S}(1-p_i)-\prod_{i\in S}(-p_i)\neq 0$$

$$I_{B,\mathbf{p}}(f,S)=\frac{1}{\prod_{i\in S}(1-p_i)-\prod_{i\in S}(-p_i)}\sum_{T\subseteq S}(-1)^{|S|-|T|}\,\mathcal{I}_{B,\mathbf{p}}(f,T)$$

Thus, for almost every $\mathbf{p}$, the knowledge of $\mathcal{I}_{B,\mathbf{p}}(f,\cdot)$ enables us to reconstruct almost all $I_{B,\mathbf{p}}(f,\cdot)$ and hence f

Open problem

An interesting question:

Define weighted interaction and influence indexes in the nonindependent case

Thank you for your attention !

Best approximation theorem

Consider a finite-dimensional subspace W of an inner product space V

Theorem

If $\mathbf{u} \in V$, then $\mathrm{proj}_W \mathbf{u}$ is the *best approximation* to $\mathbf{u}$ from W in the sense that

$$\|\mathbf{u} - \mathrm{proj}_W \mathbf{u}\| < \|\mathbf{u} - \mathbf{w}\|$$

for every $\mathbf{w} \in W$ such that $\mathbf{w} \neq \mathrm{proj}_W \mathbf{u}$

Theorem

If $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is an orthonormal basis for W, then for every $\mathbf{u} \in V$, we have

$$\mathrm{proj}_W \mathbf{u} = \sum_{i=1}^{r} \langle \mathbf{u}, \mathbf{v}_i \rangle \, \mathbf{v}_i$$