Weighted Banzhaf power and interaction indexes through weighted approximations of games

Pierre Mathonet and Jean-Luc Marichal

University of Luxembourg
Mathematics Research Unit, FSTC

Linz 2011

Cooperative games and pseudo-Boolean functions

Cooperative games

The set of players : $N = \{1, \ldots, n\}$. Game : $f : 2^N \to \mathbb{R}$. For a coalition S of players $f(S)$ is the *worth* of S (usually $f(\varnothing) = 0$, not additive).

Pseudo-Boolean functions

These are functions $f : \{0,1\}^n \to \mathbb{R}$.
We identify $S \subseteq N$ and $\mathbf{1}_S \in \{0,1\}^n$, for example

$$S = \{2,4\} \subset N = \{1,2,3,4\} \mapsto \mathbf{1}_S = (0,1,0,1).$$

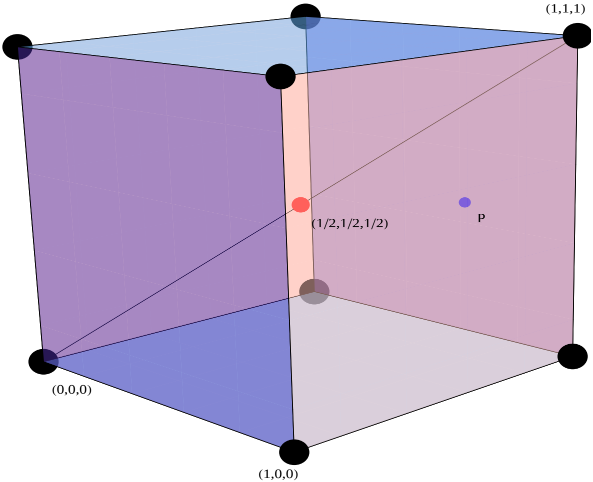
So, games are pseudo-Boolean functions. Such functions can be written

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

Multilinear extensions (Owen,1972)

$$\bar{f} : [0,1]^n \to \mathbb{R} : \mathbf{x} \mapsto \sum_{S \subseteq N} a(S) \prod_{i \in S} x_i.$$

The cube



Power indexes

Problem : Find the real influence/power of a player on the game (to share the benefits, or simply to analyze the game).

The Shapley power index (L.S. Shapley 1953):

$$\phi_{\mathrm{Sh}}(f,i) = \sum_{T \not\ni i} \frac{(n-t-1)!t!}{n!} \Delta^i f(T),$$

where

$$\Delta^i f(T) = f(T \cup i) - f(T \setminus i)$$

is the discrete derivative of f with respect to i at T.

The Banzhaf power index (J. Banzhaf 1965):

$$\phi_{\mathrm{B}}(f,i) = \frac{1}{2^{n-1}} \sum_{T \not\ni i} \Delta^i f(T).$$

There exist many axiomatic characterizations.

Interaction indexes I

Problem : The *influence* of a pair of players i, j is not the sum of their respective powers because of the *interactions*. Here we review concepts of interactions.

The Banzhaf interaction index (Owen (1972), Murofushi-Soneda (1993))

$$I_{\mathrm{B}}(f, \{i,j\}) = \frac{1}{2^{n-2}} \sum_{T \subset N \setminus \{i,j\}} (f(T \cup ij) - f(T \cup i) - f(T \cup j) + f(T)).$$

Note that, for $T \subseteq N \setminus \{i,j\}$,

$$\begin{aligned} \Delta^{ij} f(T) &= f(T \cup ij) - f(T \cup i) - f(T \cup j) + f(T) \\ &= (f(T \cup ij) - f(T)) - (f(T \cup i) - f(T)) - (f(T \cup j) - f(T)) \\ &= (f(T \cup ij) - f(T \cup j)) - (f(T \cup i) - f(T)). \end{aligned}$$

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

$$\Delta^S f(\mathbf{x}) = \sum_{T \supseteq S} a(T) \prod_{i \in T \setminus S} x_i$$

Interaction indexes II

To measure the interaction among players in coalition S:

The Banzhaf interaction index of S (Roubens (1996))

$$I_{\mathrm{B}}(f,S)=\frac{1}{2^{n-s}}\sum_{T\subset N\setminus S}\Delta^S f(T).$$

The Shapley interaction index of S (Grabisch (1997))

$$I_{\mathrm{Sh}}(f,S)=\sum_{T\subset N\setminus S}\frac{(n-t-s)!\,t!}{(n-s+1)!}\Delta^S f(T).$$

Probabilistic interaction index of S (Grabisch,Roubens, see also Fujimoto, Kojadinovic, Marichal (2006))

$$I(f,S)=\sum_{T\subset N\setminus S}p_T^S\Delta^S f(T),$$

with $p_T^S\geqslant 0$ and $\sum_T p_T^S=1$. Expected values of derivatives.

Alternative expressions of interactions

Expressions in terms of the Möbius transform

$$I_{\mathrm{B}}(f,S)=\sum_{T\supseteq S}\left(\frac{1}{2}\right)^{t-s}a(T),$$

$$I_{\mathrm{Sh}}(f,S)=\sum_{T\supseteq S}\frac{1}{t-s+1}a(T).$$

In terms of the derivatives of the Owen extension $\bar{f}$

$$I_{\mathrm{B}}(f,S)=(D^S\bar{f})(\frac{1}{2},\ldots,\frac{1}{2})=\int_{[0,1]^n}D^S\bar{f}(\mathbf{x})\,d\mathbf{x},$$

$$I_{\mathrm{Sh}}(f,S)=\int_{[0,1]}D^S\bar{f}(x,\ldots,x)\,dx.$$

We will interpret these integrals at the end of the talk.

Main properties

Alternative representations

The map $f \mapsto (I_B(f,S) : S \subseteq N)$ is a linear bijection :

$$\bar{f}(\mathbf{x}) = \sum_{S \subseteq N} I_B(f,S) \prod_{i \in S} (x_i - \frac{1}{2}).$$

Symmetry-anonymity

If $\pi \in S_n$ and $\pi(f)(x_1, \ldots, x_n) = f(x_{\pi(1)}, \ldots, x_{\pi(n)})$, then

$$I(\pi(f), \pi(S)) = I(f,S).$$

Dummy players

A player i is dummy in f if $f(T \cup i) = f(T) + f(i) - f(\varnothing)$ for $T \subseteq N \setminus i$.
If i is dummy, then

$$I(f,i) = f(i) - f(\varnothing) \quad \text{and} \quad I(f,S) = 0 \quad \forall S \ni i, S \neq \{i\}.$$

Some axiomatic characterizations use these properties.

Banzhaf power index and linear model

Rmk :If $\ell(\mathbf{x}) = \ell_\varnothing + \ell_1 x_1 + \cdots + \ell_n x_n$, we have $I_B(\ell, i) = \ell_i$.

Alternative definition of a power index

Given a pseudo-Boolean f, consider a linear model for f :

$$f_1(\mathbf{x}) = a_\varnothing + a_1 x_1 + \cdots + a_n x_n$$

Define the power of i in f by a_i.

Least squares method : find f_1 that minimizes

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

among all linear models g.
Note that this means that all the coalitions are on the same footing.

Theorem (Hammer-Holzman (1992))

In the solution of the least squares problem, $a_i = I_B(f, i)$.

Note that we could have used the model $f_{1,i}(\mathbf{x}) = a_\varnothing + a_i x_i$.

Banzhaf interaction index and multi-linear model

The setting

V_k : space of pseudo-Boolean functions of degree k at most

$$V_k = \{g : g(\mathbf{x}) = \sum_{S \subseteq N, s \leqslant k} c(S) \prod_{i \in S} x_i\}.$$

For each f, find $f_k \in V_k$ that minimizes $\sum_{\mathbf{x} \in \{0,1\}^n} (f(\mathbf{x}) - g(\mathbf{x}))^2$ among all $g \in V_k$.

Theorem (Grabisch-Marichal-Roubens (2000))

We have $f_k = \sum_{s \leqslant k} a_k(S) \prod_{i \in S} x_i$ *with*

$$a_k(S) = a(S) + (-1)^{k-s} \sum_{T \supseteq S, t > k} \binom{t-s-1}{k-s} \left(\frac{1}{2}\right)^{t-s} a(T).$$

$a_s(S) = I_B(f, S)$ *(!).*

The new recipe

To compute $I_B(f, S)$

- Look at the cardinality of S: $s = |S|$;
- Find the best approximation of f by a function of degree at most s;
- Collect the coefficient of this approximation along the monomial $\prod_{i \in S} x_i$ in this approximation.

Remarks :

1. The function $\mathbf{x} \mapsto \prod_{i \in S} x_i$ is the unanimity function w.r.t. S.
2. In this setting, all the coalitions are on the same footing, they are equally likely to form.

Weighted least squares

$w(S)$: the probability that coalition S forms : $w(S) = \Pr(C = S)$.

Under independence

$p_i = \Pr(C \ni i) = \sum_{S \ni i} w(S) \in (0,1)$ and

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

Associated weighted least squares problem

Find the unique $f_k \in V_k$ that minimizes the (squared) distance

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

among all functions $g \in V_k$.

Rmk : The distance is associated 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 w is defined by $\mathbf{p} = (p_1, \ldots, p_n)$.

First solution of the least squares problem

The use of independence Guoli Ding et al (2010)

$B_k = \{v_S : S \subseteq N,\ s \leqslant k\}$, where $v_S \colon \{0,1\}^n \to \mathbb{R}$ is given by

$$v_S(\mathbf{x}) = \prod_{i\in S} \frac{x_i - p_i}{\sqrt{p_i(1-p_i)}} = \sum_{T\subseteq S} \frac{\prod_{i\in S\setminus T}(-p_i)}{\prod_{i\in S}\sqrt{p_i(1-p_i)}} \prod_{i\in T} x_i$$

forms an orthonormal basis for V_k.

The projection

$$f_k = \sum_{\substack{T\subseteq N\\ t\leqslant k}} \langle f, v_T\rangle\, v_T .$$

The index

$$I_{\mathrm{B},\mathbf{p}}(f,S) = \frac{\langle f, v_S\rangle}{\prod_{i\in S}\sqrt{p_i(1-p_i)}}$$

First properties

The index characterizes the projection :

$f_k \in V_k$ is the best kth approximation of f iff

$$I_{\mathrm{B},\mathbf{p}}(f,S) = I_{\mathrm{B},\mathbf{p}}(f_k,S) \quad \forall\, S : s \leqslant k.$$

(Hint : this is equivalent to $\langle f, v_S\rangle = \langle f_k, v_S\rangle$.)

The map $f \mapsto I_{\mathrm{B},\mathbf{p}}(f,S)$ is linear.

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

$$f_k(\mathbf{x}) = \sum_{T\subseteq N, t\leqslant k} I_{\mathrm{B},\mathbf{p}}(f,T) \prod_{i\in T}(x_i - p_i), \quad k = n \quad (!)$$

The index and the multilinear extension of f :

$$I_{\mathrm{B},\mathbf{p}}(f,S) = (D^S\bar{f})(\mathbf{p})$$

Explicit formulas

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

Explicit expression of the index

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

Proof : Just compute the derivatives of $\bar{f}$.

Explicit expression of the approximation

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

$$a_k(S) = a(S) + (-1)^{k-s} \sum_{T\supseteq S, t>k} \binom{t-s-1}{k-t} \Big(\prod_{i\in T\setminus S} p_i\Big) a(T)$$

Proof : Use expression of f_k and $I_{\mathrm{B},\mathbf{p}}(f,S)$, expand and do some algebra.

The index as an expected value

An expected value of the discrete derivative

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

Proof : Use $\Delta^S f(\mathbf{x}) = \sum_{T\supseteq S} a(T)\prod_{i\in T\setminus S} x_i$, independence and explicit formula for $I_{\mathrm{B},\mathbf{p}}(f,S)$.

An average (As a probabilistic interaction index)

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

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

Interpretation : $p_T^S = \Pr(C = S\cup T \mid C\supseteq S) = \Pr(C = T \mid C\subseteq N\setminus S)$

Further properties

Null players

A player i is *null* for f if $f(T \cup i) = f(T)$ for all $T \subseteq N \setminus i$.
If S contains a null player then $I_{\mathrm{B},\mathbf{p}}(f, S) = 0$.

Dummy coalitions

$D \subseteq N$ is *dummy* for f if $f(T) = f(T \cap D) + f(T \cap (N \setminus D)) - f(\varnothing)$ for every $T \subseteq N$.
If D is dummy for f, if $K \cap D \neq \varnothing$ and $K \setminus D \neq \varnothing$: $I_{\mathrm{B},\mathbf{p}}(f, K) = 0$.

Symmetry of the index

An index I is symmetric if $I(\pi(f), \pi(S)) = I(f, S)$ for all permutations π.
$I_{\mathrm{B},\mathbf{p}}$ is symmetric if and only w or is symmetric i.e. $p_1 = \cdots = p_n$.

Back to Banzhaf and Shapley

A link with the Banzhaf index

$$I_{\mathrm{B},\mathbf{p}'}(f,S) = \sum_{T\supseteq S} I_{\mathrm{B},\mathbf{p}}(f,T) \prod_{i\in T\setminus S}(p'_i - p_i), \quad \text{set } p_i \text{ or } p'_i \text{ to } \tfrac{1}{2} \quad (!)$$

Another link with the Banzhaf index

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

Proof : Just use explicit expressions and integrate.
Interpretation : take an average over **p** if it is not known.

A link with the Shapley index

$$I_{\mathrm{Sh}}(f,S) = \int_0^1 D^S\overline{f}(p,\ldots,p)\,dp = \int_0^1 I_{\mathrm{B},(p,\ldots,p)}(f,S)\,dp.$$

Interpretation : average if the players behave in the same (unknown) way.

Thanks for your attention

arxiv:1001.3052