Measuring the interactions among variables of functions over $[0,1]^n$

Jean-Luc Marichal and Pierre Mathonet

University of Luxembourg

Measure of influence of variables

Problem

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

We want to measure the *influence* (*importance*) of x_k over

$$f(x_1, \ldots, x_k, \ldots, x_n)$$

Example: affine function

$$f(x_1, \ldots, x_n) = c_0 + c_1\, x_1 + \cdots + c_k\, x_k + \cdots + c_n\, x_n$$

Measure of influence of x_k:

coefficient c_k

Measure of influence of variables

More complicated functions...

$$
\begin{aligned}
f(\mathbf{x}) &= \prod_{i=1}^{n} x_i^{c_i} \\
f(\mathbf{x}) &= \left(\sum_{i=1}^{n} c_i \, x_i^2 \right)^{1/2} \\
f(\mathbf{x}) &= \max_{i=1..n} \min(c_i, x_i)
\end{aligned}
$$

etc.

Measure of influence of variables

A reasonable answer:

$\longrightarrow$ Approximation of f by an affine function

$$f_1(x_1, \ldots, x_n) = a_0 + a_1\, x_1 + \cdots + a_k\, x_k + \cdots + a_n\, x_n$$

Measure of influence of x_k over f:

coefficient a_k

Measure of interaction among variables

What about interactions among variables?

Example: multilinear function

$$f(x_1, x_2) = c_0 + c_1\, x_1 + c_2\, x_2 + c_{12}\, x_1 x_2$$

Measure of *interaction* between x_1 and x_2 within f:

coefficient c_{12}

$$\begin{array}{lcl} c_{12} = 0 & : & \text{zero interaction} \\ c_{12} > 0 & : & \text{positive interaction} \\ c_{12} < 0 & : & \text{negative interaction} \end{array}$$

Measure of interaction among variables

Let $S \subseteq N = \{1, \ldots, n\}$
Measure of interaction among variables $\{x_k : k \in S\}$

$\longrightarrow$ Approx. of f by a multilinear polynomial of degree $\leqslant s = |S|$

$$f_s(\mathbf{x}) = \sum_{\substack{T \subseteq N \\ |T| \leqslant s}} a_s(T) \prod_{i \in T} x_i$$

Measure of interaction among $\{x_k : k \in S\}$ inside f:

coefficient $a_s(S)$

(leading coefficients)

Approximation problem

Multilinear approximation

Denote by M_s the set of all multilinear polynomials $g\colon [0,1]^n \to \mathbb{R}$ of degree $\leqslant s$

$$g(\mathbf{x}) = \sum_{\substack{T\subseteq N\\ |T|\leqslant s}} a_s(T) \prod_{i\in T} x_i$$

We define the *best s-th approximation* of a function $f \in L^2([0,1]^n)$ as the multilinear polynomial $f_s \in M_s$ that minimizes

$$d(f,g)^2 = \int_{[0,1]^n} \big(f(\mathbf{x}) - g(\mathbf{x})\big)^2 \, d\mathbf{x}$$

among all $g \in M_s$

Interaction index

From the best s-th approximation f_s of f, we define the following *interaction index* (*power index* if $s = 1$)

$$\mathcal{I}(f, S) = a_s(S) \qquad (s = |S|)$$

$$= \text{coefficient of } \prod_{i \in S} x_i \text{ in the best } s\text{-th approximation } f_s \text{ of } f$$

Interaction index

Explicit expression

$$\mathcal{I}(f,S) = 12^s \int_{[0,1]^n} f(\mathbf{x}) \prod_{i\in S}\Big(x_i - \frac{1}{2}\Big)\, d\mathbf{x}$$

Question: How can we see that this index actually measures an interaction?

Interaction index

Interpretation:

$$D_k f(\mathbf{x}) = \textit{rate of change} \text{ w.r.t. } x_k \text{ of } f \text{ at } \mathbf{x}$$
$$= \textit{local contribution} \text{ of } x_k \text{ over } f \text{ at } \mathbf{x}$$

$$D_j D_k f(\mathbf{x}) = \textit{local interaction} \text{ between } x_j \text{ and } x_k \text{ within } f \text{ at } \mathbf{x}$$

Theorem

If f is sufficiently differentiable, then

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

where $q_S(\mathbf{x})$ is a probability density function

Interaction index

What if f is not differentiable?

S-derivative $\rightarrow$ *discrete S-derivative* (S-difference quotient)

Theorem

We have

$$\mathcal{I}(f,S) = \int_{\mathbf{x}\in[0,1]^n} \int_{\mathbf{h}_S\in[\mathbf{0},\mathbf{1}-\mathbf{x}_S]} p_S(\mathbf{h}) \, \frac{\Delta_{\mathbf{h}}^S f(\mathbf{x})}{\prod_{i\in S} h_i} \, d\mathbf{h}_S \, d\mathbf{x}$$

where $p_S(\mathbf{h})$ is a p.d.f. over the domain of integration

Interaction index

Examples:

$$f(\mathbf{x}) = \sum_{i=1}^{n} w_i \, x_i$$

- $\mathcal{I}(f, \{k\}) = w_k$
- $\mathcal{I}(f, S) = 0 \qquad |S| \geqslant 2$

$$f(\mathbf{x}) = \Big(\prod_{i=1}^{n} x_i \Big)^{1/n}$$

- $\mathcal{I}(f, S) = \left(\frac{n}{n+1}\right)^{n} \left(\frac{6}{n+2}\right)^{|S|}$

Interaction index

Further examples:

$$\boxed{f(\mathbf{x}) = \min_{i \in T} x_i \qquad (T \subseteq N)}$$

$$\mathcal{I}(f,S) = \begin{cases} 6^{|S|}\,\frac{|S|!\,|T|!}{(|S|+|T|+1)!}, & \text{if } T \supseteq S, \\ 0, & \text{otherwise.} \end{cases}$$

$$\boxed{f(\mathbf{x}) = \max_{i \in T} x_i \qquad (T \subseteq N)}$$

$$\mathcal{I}(f,S) = \begin{cases} (-1)^{|S|+1}\,6^{|S|}\,\frac{|S|!\,|T|!}{(|S|+|T|+1)!}, & \text{if } T \supseteq S, \\ 0, & \text{otherwise.} \end{cases}$$

Interaction index

Further examples:

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

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

$$\boxed{f(\mathbf{x}) = \sum_{T\subseteq N} c(T) \min_{i\in T} x_i}$$

$$\mathcal{I}(f,S) = 6^{|S|} \sum_{T\supseteq S} c(T) \frac{|S|!\,|T|!}{(|S|+|T|+1)!}$$

Some properties

A variable x_k is *inefficient* for f if f does not depend on x_k

$$f(x_k; \mathbf{x}_{N \setminus k}) - f(0_k; \mathbf{x}_{N \setminus k}) = 0$$

Property

If x_k is *inefficient* for f then

$$\mathcal{I}(f, S) = 0 \qquad \forall S \ni k$$

Some properties

A variable x_k is *dummy* for f if

$$f(x_k;\mathbf{x}_{N\smallsetminus k}) - f(0_k;\mathbf{x}_{N\smallsetminus k}) = f(x_k;\mathbf{0}_{N\smallsetminus k}) - f(0_k;\mathbf{0}_{N\smallsetminus k})$$

(the marginal contribution of x_k is independent of $\mathbf{x}_{N\smallsetminus k}$)

Property

If x_k is *dummy* for f then

$$\mathcal{I}(f,S) = 0 \qquad \forall S \ni k,\ |S| \geqslant 2$$

Some properties

A combination of variables $\{x_k : k \in S\}$ is *dummy* for f if

$$f(\mathbf{x}_S; \mathbf{x}_{N \setminus S}) - f(\mathbf{0}_S; \mathbf{x}_{N \setminus S}) = f(\mathbf{x}_S; \mathbf{0}_{N \setminus S}) - f(\mathbf{0}_S; \mathbf{0}_{N \setminus S})$$

Property

If $\{x_k : k \in S\}$ is *dummy* for f, then

$$\mathcal{I}(f, T) = 0 \qquad \forall\, T \text{ such that } T \cap S \neq \varnothing \text{ and } T \setminus S \neq \varnothing$$

Conclusion

Given a function $f \in L^2([0,1]^n)$, we have defined an index to measure

- the influence of variables over f
- the interaction among variables within f

Interpretations

- leading coefficients in multilinear approximation (multilinear regression)
- mean S-derivative or S-difference quotient

Properties

- inefficient and dummy variables
- ...

Thank you for your attention!

arXiv : 0912.1547