Proportional Marginal Effects to Quantify Importance

Cooperative Game Theory and Global Sensitivity Analysis

[1]**EDF R&D - Lab Chatou** - PRISME Department
[2]**Institut de Mathématiques de Toulouse**
[3]**SINCLAIR AI Lab**

Forum MobiliT.AI
MEET, Toulouse, France
May 30-31, 2023

Marouane Il Idrissi[123], Margot Herin, Vincent Chabridon[13], Bertrand Iooss[123].

Context

Given **random inputs** $X_1, \ldots, X_d$ and a **random output** $G(X_1, \ldots, X_d)$, how much each **input contribute** to $\mathbb{V}(G(X_1, \ldots, X_d))$?

Context

Given **random inputs** $X_1, \dots, X_d$ and a **random output** $G(X_1, \dots, X_d)$, how much each **input contribute** to $\mathbb{V}\left(G(X_1, \dots, X_d)\right)$?

G is a **deterministic black-box model**:

- **Numerical model** (e.g., simulation codes)
- **Learned ML/DL models** (e.g., post-hoc interpretations).

The inputs $X_1, \dots, X_d$ are **not necessarily mutually independent.**

Context

Given **random inputs** $X_1, \dots, X_d$ and a **random output** $G(X_1, \dots, X_d)$, how much each **input contribute** to $\mathbb{V}\left(G(X_1, \dots, X_d)\right)$?

G is a **deterministic black-box model**:

- **Numerical model** (e.g., simulation codes)
- **Learned ML/DL models** (e.g., post-hoc interpretations).

The inputs $X_1, \dots, X_d$ are **not necessarily mutually independent.**

Global sensitivity analysis (GSA) provides an answer: the **Shapley effects** (Owen 2014). However, **exogenous inputs can be granted a non-zero contribution**.

Context

Given **random inputs** $X_1, \dots, X_d$ and a **random output** $G(X_1, \dots, X_d)$, how much each **input contribute** to $\mathbb{V}(G(X_1, \dots, X_d))$?

G is a **deterministic black-box model**:

- **Numerical model** (e.g., simulation codes)
- **Learned ML/DL models** (e.g., post-hoc interpretations).

The inputs $X_1, \dots, X_d$ are **not necessarily mutually independent.**

Global sensitivity analysis (GSA) provides an answer: the **Shapley effects** (Owen 2014). However, **exogenous inputs can be granted a non-zero contribution**.

Can we build **interpretable indices** that **circumvent this drawback**?

Context

Given **random inputs** $X_1, \dots, X_d$ and a **random output** $G(X_1, \dots, X_d)$, how much each **input contribute** to $\mathbb{V}(G(X_1, \dots, X_d))$?

G is a **deterministic black-box model**:

- **Numerical model** (e.g., simulation codes)
- **Learned ML/DL models** (e.g., post-hoc interpretations).

The inputs $X_1, \dots, X_d$ are **not necessarily mutually independent.**

Global sensitivity analysis (GSA) provides an answer: the **Shapley effects** (Owen 2014). However, **exogenous inputs can be granted a non-zero contribution**.

Can we build **interpretable indices** that **circumvent this drawback**?

Idea: Use a **different allocation** than the Shapley values.

Cooperative game theory

“Cooperative game theory = The art of cutting a cake”.

Cooperative game theory

"Cooperative game theory = The art of cutting a cake".

Formally, given:

- A **set of players** $D = \{1, \ldots, d\}$, and the subsequent **set of coalitions** $\mathcal{P}(D)$.
- A **value function** $v : \mathcal{P}(D) \to \mathbb{R}$ quantifying the **value produced** by each **coalition**.

(D, v) defines a **cooperative game**.

"Cooperative game theory = The art of cutting a cake".

Formally, given:

- A **set of players** $D = \{1, \ldots, d\}$, and the subsequent **set of coalitions** $\mathcal{P}(D)$.
- A **value function** $v : \mathcal{P}(D) \to \mathbb{R}$ quantifying the **value produced** by each **coalition**.

(D, v) defines a **cooperative game**.

Main question: How can we redistribute $v(D)$ among the players ?

Cooperative game theory

"Cooperative game theory = The art of cutting a cake".

Formally, given:

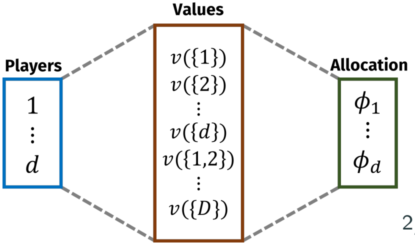
- A **set of players** $D = \{1, \ldots, d\}$, and the subsequent **set of coalitions** $\mathcal{P}(D)$.
- A **value function** $v : \mathcal{P}(D) \to \mathbb{R}$ quantifying the **value produced** by each **coalition**.

(D, v) defines a **cooperative game**.

Main question: How can we redistribute $v(D)$ among the players ?

Answer: By using allocations !

Allocation: description of the "cake-cutting" process.



Random order allocations

Allocating "the **whole cake** and **nothing but the cake**" is ensured by two criteria:

- **Efficiency**: $\sum_{i=1}^{d} \phi_i = v(D)$ (The whole cake).
- **Nonnegativity**: $\forall i \in D, \quad \phi_i \geq 0$ (Nothing but the cake).

Random order allocations

Allocating "the **whole cake** and **nothing but the cake**" is ensured by two criteria:

- **Efficiency**: $\sum_{i=1}^{d} \phi_i = v(D)$ (The whole cake).
- **Nonnegativity**: $\forall i \in D, \quad \phi_i \geq 0$ (Nothing but the cake).

Random order allocations (or the Weber (1988) set) are a **class of allocations** that are **always efficient**. They can be written $\forall i \in D$, as:

$$\phi_i = \sum_{\pi \in \mathcal{S}_D} p(\pi) \left[v\left(C_{\pi(i)}(\pi)\right) - v\left(C_{\pi(i)-1}(\pi)\right)\right].$$

where $\mathcal{S}_D$ is the set of permutations of D, and where

- $C_{\pi(i)-1}(\pi)$ is the set of players **before** i **in** π.
- $C_{\pi(i)}(\pi) = C_{\pi(i)-1}(\pi) \cup \{i\}$

$p(\pi)$ assigns a **probability to every permutation** $\pi \in \mathcal{S}_D$.

Random order allocations

Allocating “the **whole cake** and **nothing but the cake**” is ensured by two criteria:

- **Efficiency**: $\sum_{i=1}^{d} \phi_i = v(D)$ (The whole cake).
- **Nonnegativity**: $\forall i \in D, \quad \phi_i \geq 0$ (Nothing but the cake).

Random order allocations (or the Weber (1988) set) are a **class of allocations** that are **always efficient**. They can be written $\forall i \in D$, as:

$$\phi_i = \sum_{\pi \in \mathcal{S}_D} p(\pi) \left[v\left(C_{\pi(i)}(\pi)\right) - v\left(C_{\pi(i)-1}(\pi)\right) \right].$$

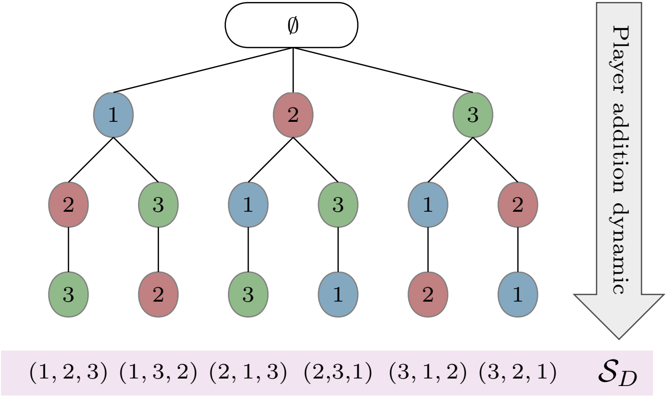
where $\mathcal{S}_D$ is the set of permutations of D, and where

- $C_{\pi(i)-1}(\pi)$ is the set of players **before** i **in** π.
- $C_{\pi(i)}(\pi) = C_{\pi(i)-1}(\pi) \cup \{i\}$

$p(\pi)$ assigns a **probability to every permutation** $\pi \in \mathcal{S}_D$.

A choice of $p \implies$ **An efficient allocation**

Random order allocations



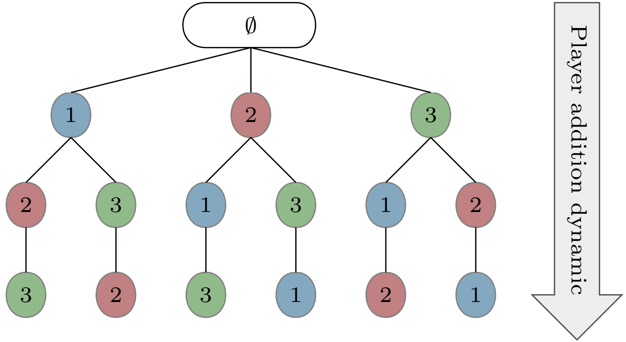
$$\Delta_\pi(i) := v\left(C_{\pi(i)}(\pi)\right) - v\left(C_{\pi(i)-1}(\pi)\right)$$

$$\phi_1 = \sum_{\pi \in \mathcal{S}_D} p(\pi)\Delta_\pi(1)$$

$$\phi_2 = \sum_{\pi \in \mathcal{S}_D} p(\pi)\Delta_\pi(2)$$

$$\phi_3 = \sum_{\pi \in \mathcal{S}_D} p(\pi)\Delta_\pi(2)$$

Random order allocations



$(1,2,3)$ $(1,3,2)$ $(2,1,3)$ $(2,3,1)$ $(3,1,2)$ $(3,2,1)$ $\mathcal{S}_D$

$$\Delta_\pi(i) := v\left(C_{\pi(i)}(\pi)\right) - v\left(C_{\pi(i)-1}(\pi)\right)$$

$$\phi_1 = \sum_{\pi \in \mathcal{S}_D} p(\pi)\Delta_\pi(1)$$

$$\phi_2 = \sum_{\pi \in \mathcal{S}_D} p(\pi)\Delta_\pi(2)$$

$$\phi_3 = \sum_{\pi \in \mathcal{S}_D} p(\pi)\Delta_\pi(2)$$

If v is **monotonic** (i.e., $\forall B \subseteq A \in \mathcal{P}(D)$, $v(B) \leq v(A)$), every random order allocation is also **nonnegative**.

Shapley values

The **Shapley values** is a **random order allocation** with the choice:

$$p(\pi) = \frac{1}{d!}, \quad \forall \pi \in \mathcal{S}_D,$$

and they can be interpreted as

"[...] *an a priori assessment of the situation, based on either ignorance or disregard of the social organization of the players."* - L. S. Shapley (2016)

They are a **uniform prior on the underlying redistribution process**, leading to an **egalitarian allocation principle.**

Shapley effects

By **analogy between players and inputs**, Owen (2014) proposed to study the game:

$$(D, S^T), \quad \text{where } \forall A \in \mathcal{P}(D), \quad S_A^T = \frac{\mathbb{E}\left[\mathbb{V}\left(G(X) \mid X_{\overline{A}}\right)\right]}{\mathbb{V}\left(G(X)\right)}.$$

The **Shapley effects** are the Shapley values of (D, S^T) (Song, Nelson, and Staum 2016).

Shapley effects

By **analogy between players and inputs**, Owen (2014) proposed to study the game:

$$(D, S^T), \quad \text{where } \forall A \in \mathcal{P}(D), \quad S^T_A = \frac{\mathbb{E}\left[\mathbb{V}\left(G(X) \mid X_{\overline{A}}\right)\right]}{\mathbb{V}\left(G(X)\right)}.$$

The **Shapley effects** are the Shapley values of (D, S^T) (Song, Nelson, and Staum 2016).

Since S^T is monotonic and $S^T_D = 1$, the **Shapley effects are efficient and nonnegative**: **They can be interpreted as shares of variance allocated to each input.** They are a **suitable solution** for **importance quantification with dependent inputs**.

Shapley effects

By **analogy between players and inputs**, Owen (2014) proposed to study the game:

$$(D, S^T), \quad \text{where } \forall A \in \mathcal{P}(D), \quad S^T_A = \frac{\mathbb{E}\left[\mathbb{V}\left(G(X) \mid X_{\overline{A}}\right)\right]}{\mathbb{V}\left(G(X)\right)}.$$

The **Shapley effects** are the Shapley values of (D, S^T) (Song, Nelson, and Staum 2016).

Since S^T is monotonic and $S^T_D = 1$, the **Shapley effects are efficient and nonnegative**: **They can be interpreted as shares of variance allocated to each input.** They are a **suitable solution** for **importance quantification with dependent inputs**.

However, **they do not detect exogenous inputs:**

$$G(X) = X_1 + X_2, \quad X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \rho \\ 0 & 1 & 0 \\ \rho & 0 & 1 \end{pmatrix} \right),$$

$$\mathrm{Sh}_1 = 0.5 - \rho^2/4, \quad \mathrm{Sh}_2 = 0.5, \quad \mathrm{Sh}_3 = \rho^2/4 > 0 \text{ if } \rho \neq 0.$$

Proportional values

Is it possible to find a suitable p in order to produce **interpretable indices** that **detect exogenous inputs** ?

Proportional values

Is it possible to find a suitable p in order to produce **interpretable indices** that **detect exogenous inputs** ?

The **proportional values** (Ortmann 2000) can be interpreted as a redistribution such that

> "[...] *each player gains in **equal proportion** to that which could be obtained by each alone."* - B. Feldman (1999)

They are based on a **proportional allocation principle** for **positive games**.

If $\forall A \in \mathcal{P}(D), v(A) > 0$, the choice of p is:

$$p(\pi) = \frac{L(\pi)}{\sum_{\sigma \in \mathcal{S}_D} L(\sigma)}, \quad L(\pi) = \exp\left(-\sum_{j \in D} \log\left(v\left(C_j(\pi)\right)\right)\right)$$

Proportional marginal effects

We extended the proportional values to **nonnegative value functions** (Herin et al. 2022).

The **proportional marginal effects (PME)** are the (extended) proportional values of the game (D, S^T).

Proposition *(Exogeneity detection (Herin et al. 2022)).*

Let $E \in \mathcal{P}(D)$. If X_E is the largest set of exogenous inputs, then:

$$\forall i \in E, \quad PME_i = 0, \quad \forall j \in \overline{E}, \quad PME_j > 0.$$

They are **efficient** and **nonnegative**: **interpretation as shares of the output variance.**

Estimation

Estimating the PME/Shapley effects $\iff$ Estimating S_A^T for every $A \in \mathcal{P}(D)$.

It can be achieved:

- Via **Monte Carlo sampling**: Requires a number proportional to $d!(d-1)$ model evaluations (Song, Nelson, and Staum 2016).
- **Given-data** (i.i.d. input-output sample) via a **nearest-neighbor procedure**: Requires 2^d estimates (Broto, Bachoc, and Depecker 2020).

These methods are time-consuming and do not scale with the number of inputs, but the **estimates can be recycled to compute both indices at once.**

Ishigami Model - Exogeneity detection

The (modified) Ishigami model is given by

$$G(X) = \sin(X_1) + 7\sin^2(X_2) + 0.1{X_3}^4\sin(X_1)$$

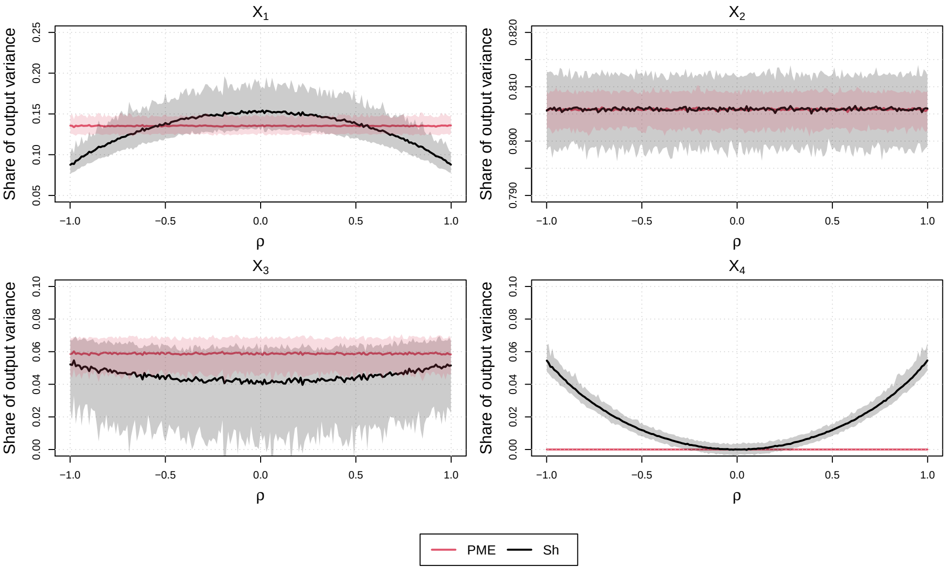
where

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} (\pi/3)^2 & 0 & 0 & \rho \\ 0 & (\pi/3)^2 & 0 & 0 \\ 0 & 0 & (\pi/3)^2 & 0 \\ \rho & 0 & 0 & (\pi/3)^2 \end{pmatrix}\right).$$

where X_4 is exogenous.

Estimation using Monte Carlo sampling, 200 repetitions.

Ishigami Model - Exogeneity detection



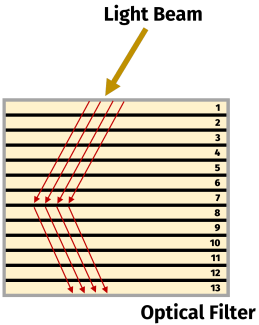
Optical filter transmittance - Feature selection

Quantification of the **transmittance performance** of an **optical filter** composed of 13 consecutive layers (Vasseur et al. 2010).

The inputs $I_1, \ldots, I_{13}$ represent the **refractive index error** of each filter ($\mathcal{U}([-0.05, 0.05])$)

These errors are (highly) correlated due to the manufacturing process (Gaussian copula, $\rho = 0.95$).

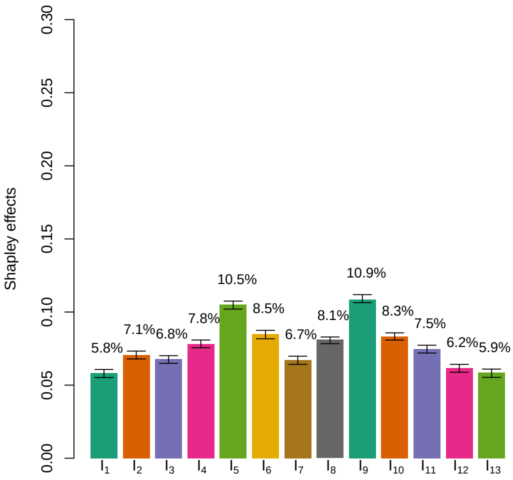
The black-box model computes the **transmittance error w.r.t. the "perfect filter"** over several wavelengths.

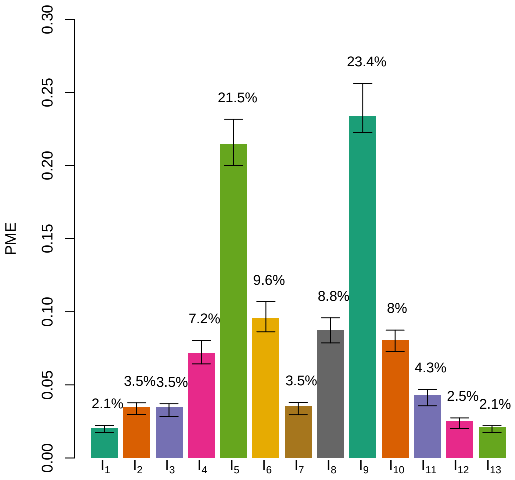


We only have access to an i.i.d. input-output sample ($n = 1000$).

The indices are computed using the nearest-neighbors approach (6 neighbors).

Optical filter transmittance - Feature selection





Optical filter transmittance - Feature selection

Scenario: We want to build a surrogate model (Gaussian process*) of this numerical model.

Using the whole dataset: $Q^2 = 99.48\%$.

Feature selection:

- First threshold: 2.5% importance.
  - **Shapley effects:** No features removed.
  - **PME:** l_1 and l_3 are removed, $Q^2 = 99.14\%$.
- Second threshold: 5% importance.
  - **Shapley effects:** No features removed.
  - **PME:** 7 inputs are removed, $Q^2 = 98.79\%$.

* $5/2$ Matérn covariance kernel, constant trend.

Conclusion

Conclusion:

- **Cooperative game theory:** resourceful for building interpretable importance indices.
- **Random order allocations:** reduce the allocation problem to a choice of p.
- **Shapley effects:** Subject to correlation distortion (equalize importance) (Verdinelli and Wasserman 2023), and “Shapley’s joke”.
- **PMEs:** Exogenous input detection and discriminative power.
- **Estimation:** Costly, but both indices can be estimated at once.

Software: Shapley effects and PMEs can be estimated using the R package `sensitivity`.

Conclusion

For a more in-depth discussion and additional analytical and empirical results, check out our pre-print (HAL/arXiv/ResearchGate):

Proportional marginal effects for global sensitivity analysis

Margot Herin[a], Marouane Il Idrissi[b,c,d], Vincent Chabridon[b,c], Bertrand Iooss[b,c,d,e]

[a] *Sorbonne Université, Laboratoire d'Informatique de Paris 6, 4 place Jussieu, 75005 Paris, France.*
[b] *EDF Lab Chatou, 6 Quai Watier, 78401 Chatou, France*
[c] *SINCLAIR AI Lab., Saclay, France*
[d] *Institut de Mathématiques de Toulouse, 31062 Toulouse, France*
[e] *Corresponding Author - Email: bertrand.iooss@edf.fr*

References i

Broto, B., F. Bachoc, and M. Depecker. 2020. "Variance reduction for estimation of Shapley effects and adaptation to unknown input distribution." *SIAM/ASA Journal on Uncertainty Quantification* 8:693–716.

Feldman, B. E. 1999. "The proportional value of a cooperative game." *Manuscript. Chicago: Scudder Kemper Investments.*

Herin, M., M. I., V. Chabridon, and B. Iooss. 2022. *Proportional marginal effects for global sensitivity analysis* [in en], October. https://hal.science/hal-03825935.

Ortmann, K. M. 2000. "The proportional value for positive cooperative games" [in English]. *Mathematical Methods of Operations Research (ZOR)* 51, no. 2 (April). ISSN: 1432-2994, 1432-5217, accessed December 4, 2020. https://doi.org/10.1007/s001860050086.

Owen, Art B. 2014. "Sobol' Indices and Shapley Value" [in en]. *SIAM/ASA Journal on Uncertainty Quantification* 2, no. 1 (January): 245–251. ISSN: 2166-2525. https://doi.org/10.1137/130936233. http://epubs.siam.org/doi/10.1137/130936233.

Shapley, L. S. 2016. "17. A Value for n-Person Games." In *Contributions to the Theory of Games (AM-28), Volume II,* edited by Harold William Kuhn and Albert William Tucker, 307–318. Princeton University Press. https://doi.org/doi:10.1515/9781400881970-018.

Song, E., B.L. Nelson, and J. Staum. 2016. "Shapley effects for global sensitivity analysis: Theory and computation." *SIAM/ASA Journal on Uncertainty Quantification* 4:1060–1083.

References ii

Vasseur, O., M. Claeys-Bruno, M. Cathelinaud, and M. Sergent. 2010. "High-dimensional sensitivity analysis of complex optronic systems by experimental design: applications to the case of the design and the robustness of optical coatings." *Chinese Optics Letters* 8(s1):21–24.

Verdinelli, I., and Larry Wasserman. 2023. *Feature Importance: A Closer Look at Shapley Values and LOCO.* ArXiv:2303.05981 [stat], March. https://doi.org/10.48550/arXiv.2303.05981. http://arxiv.org/abs/2303.05981.

Weber, R. J. 1988. "Probabilistic values for games." Chap. 7 in *The Shapley value: essays in honor of Lloyd S. Shapley,* edited by A. E. Roth, 101–120. New York, NY: Cambridge University Press.

Thank you for your attention!

Any questions?

Exogeneity

Definition *(L^2-exogeneity).* *Let $X = (X_1, \ldots, X_d)$ be random inputs of a model $G : \mathbb{R}^d \mapsto \mathbb{R}$ such that $Y = G(X)$, with Y the random output. Let $E \subset D$. The subset of random inputs X_E are said to be (L^2-)exogenous to G if, $\exists f \in L^2(P_{X_{\overline{E}}})$ such that:*

$$Y = f(X_{\overline{E}}) \quad a.s.$$

Proportional values extension

Theorem *(PV extension to monotonic nonnegative games).* *Let (D, v) be a nonnegative and monotonic game with value function $v : \mathcal{P}(D) \to \mathbb{R}^+$. Denote $\mathcal{K}$ the set of largest (w.r.t. their cardinality) zero coalitions, i.e., $\mathcal{K} = \underset{A \in \mathcal{P}(D)}{\operatorname{argmax}} \{|A| : v(A) = 0\}$. Additionally, the sets of largest zero coalitions that do not contain $i \in D$ is denoted by $\mathcal{K}_{-i}$, i.e., $\mathcal{K}_{-i} = \underset{A \in \mathcal{P}(D)}{\operatorname{argmax}} \{|A| : v(A) = 0, i \notin A\}$. Define, for any $A \in \mathcal{K}$, the positive set function:*

$$\begin{aligned} v_A : \mathcal{P}(D \setminus A) &\to \mathbb{R}_*^+ \\ B &\mapsto v(B \cup A). \end{aligned}$$

Let $\mathrm{PV}^0((D, v)) = \left(\mathrm{PV}_1^0, \dots, \mathrm{PV}_d^0\right)$ be the allocation defined as:

$$\mathrm{PV}_i^0 = \frac{\sum_{A \in \mathcal{K}_{-i}} R\left(D_{-i} \setminus A, v_A\right)^{-1}}{\sum_{A \in \mathcal{K}} R\left(D \setminus A, v_A\right)^{-1}} \quad \text{if } \mathcal{K}_{-i} \neq \emptyset \text{ and } \quad \mathrm{PV}_i^0 = 0 \text{ otherwise.} \tag{1}$$

Then, PV^0 is a continuous extension of PV to the set of nonnegative monotonic games, i.e., for a positive monotonic game (D, v),

$$\mathrm{PV}^0((D, v)) = \mathrm{PV}((D, v)).$$

Ratio potential computation

First, recall that for any value function v, $R(\emptyset, v) = 1$ and for any $i \in D$, $R(i, v) = v(\{i\})$. The computation of $R(A, v)$ can be broken down as follows:

1. Let $A \in \mathcal{P}(D)$, $A \neq \emptyset$, $|A| \geq 2$.
2. Compute $v(B)$, for every $B \in \mathcal{P}(A)$.
3. For $m = 1, \ldots, |A| - 1$:

   $\hookrightarrow$ For $B \subseteq A$ such that $|B| = m$:

   $\hookrightarrow$ Compute $R(B, v) = v(B) \left(\sum_{j \in B} R(B_{-j}, v)^{-1} \right)^{-1}$.

4. Compute $R(A, v) = v(A) \left(\sum_{j \in A} R(A_{-j}, v)^{-1} \right)^{-1}$.

Following this algorithm and given conditional element estimates, one can then compute $R\left(A, \widehat{S^T}\right)$ for any $A \in \mathcal{P}(D)$.

Computing the PME

Define the function, $\forall A \in \mathcal{P}(D)$:

$$\widehat{\zeta_A} : \mathcal{P}(D \setminus A) \to \mathbb{R}^+$$
$$B \mapsto \widehat{\zeta_A}(B) := \widehat{S^T_{A \cup B}}$$

The PME computation can then be broken down as follows:

1. Compute $\widehat{S^T_A}$, for every $A \in \mathcal{P}(D)$.
2. Compute $\mathcal{K} = \underset{A \in \mathcal{P}(D)}{\text{argmax}} \left\{ |A| : \widehat{S^T_A} = 0 \right\}$.
3. For every $A \in \mathcal{K}$, compute $R\left(D \setminus A, \widehat{\zeta_A}\right)$.
4. Let $R_{\mathcal{K}} = \sum_{A \in \mathcal{K}} R\left(D \setminus A, \widehat{\zeta_A}\right)^{-1}$.
5. For $i = 1, \dots, d$:
    5.1 Compute $\mathcal{K}_{-i} = \underset{A \in \mathcal{P}(D)}{\text{argmax}} \{|A| : v(A) = 0, i \notin A\}$.
    5.2 If $\mathcal{K}_{-i} = \emptyset$, set $\text{PME}_i = 0$.
    5.3 If $\mathcal{K}_{-i} \neq \emptyset$:
        5.3.1 For every $A \in \mathcal{K}_{-i}$, compute $R\left(D_{-i} \setminus A, \widehat{\zeta_A}\right)$.
        5.3.2 Let $\text{PME}_i = \sum_{A \in \mathcal{K}_{-i}} R\left(D_{-i} \setminus A, \widehat{\zeta_A}\right)^{-1} / R_{\mathcal{K}}$.