Beyond Shapley Values

Cooperative Games for the Interpretation of Machine Learning Models

[1]**Département de mathématiques** - *Université du Québec à Montréal*
[2]**Institut Intelligence et Données** - *Université Laval*

IJCAI 2025 - Workshop on Explainable Artificial Intelligence
Palais des congrès - Montréal, QC, Canada.
August 17, 2025

Marouane Il Idrissi[12], Agathe Fernandes-Machado[1], Arthur Charpentier[1]

Introduction

The **Shapley values** are a staple in post-hoc model-agnostic feature attribution methods

☞ **Not for the right reasons**

Introduction

The **Shapley values** are a staple in post-hoc model-agnostic feature attribution methods
☞ **Not for the right reasons**

They have been **heavily criticized** in the recent literature
☞ **Not for the right reasons either**

Introduction

The **Shapley values** are a staple in post-hoc model-agnostic feature attribution methods
☞ **Not for the right reasons**

They have been **heavily criticized** in the recent literature
☞ **Not for the right reasons either**

This is because **the theory of cooperative games** is often under-utilized
☞ The **Shapley values** are **not <u>the</u> solution**, **they are a starting point**

Introduction

The **Shapley values** are a staple in post-hoc model-agnostic feature attribution methods
☞ **Not for the right reasons**

They have been **heavily criticized** in the recent literature
☞ **Not for the right reasons either**

This is because **the theory of cooperative games** is often under-utilized
☞ The **Shapley values** are **not <u>the</u> solution**, **they are a starting point**

<u>In this presentation:</u>

How can we leverage the theory of cooperative games for the interpretation of black-box machine learning models?

And (better) understand **what the Shapley values are**

Cooperative Game Theory

"Cooperative game theory = The art of sharing a cake"

Cooperative Game Theory

"Cooperative game theory = The art of sharing a cake"

<u>Two ingredients:</u>

- $D = \{1, \ldots, d\}$, a **set of players**, and $\mathcal{P}_D$ the **set of coalitions of players**
- $v : \mathcal{P}_D \to \mathbb{R}$, a **value function**, that **assigns a value to each coalition**

☞ (D, v) formally defines a **cooperative game**
☞ $v(D)$ is the value of the "grand coalition" (the cake)

Cooperative Game Theory

"Cooperative game theory = The art of sharing a cake"

Two ingredients:

- $D = \{1, \ldots, d\}$, a **set of players**, and $\mathcal{P}_D$ the **set of coalitions of players**
- $v : \mathcal{P}_D \to \mathbb{R}$, a **value function**, that **assigns a value to each coalition**

☞ (D, v) formally defines a **cooperative game**
☞ $v(D)$ is the value of the "grand coalition" (the cake)

Main question:

How can we redistribute $v(D)$ to each of the d individual players?

Example: Lindeman, Merenda, and Gold (1980) indices

Example: Lindeman, Merenda, and Gold (1980) indices

Data: features $X = (X_1, \ldots, X_d)$, and target Y
Players: for $A \in \mathcal{P}_D$, X_A only contains features with indices in A
Model: fitted linear regression model $\widehat{f}(X) = \widehat{\beta}_0 + X^\top \widehat{\beta}$

Cooperative Game Theory and "Machine Learning"

Example: Lindeman, Merenda, and Gold (1980) indices

Data: features $X = (X_1, \ldots, X_d)$, and target Y
Players: for $A \in \mathcal{P}_D$, X_A only contains features with indices in A
Model: fitted linear regression model $\widehat{f}(X) = \widehat{\beta}_0 + X^\top \widehat{\beta}$

Cooperative game:
Features indices D and value function

$$v(A) = R^2_Y(X_A)$$

i.e., the R^2 coefficient of the linear regression of X_A on Y
R^2 is the % of variance explained by the model

Cooperative Game Theory and "Machine Learning"

Example: Lindeman, Merenda, and Gold (1980) indices

Data: features $X = (X_1, \dots, X_d)$, and target Y
Players: for $A \in \mathcal{P}_D$, X_A only contains features with indices in A
Model: fitted linear regression model $\widehat{f}(X) = \widehat{\beta}_0 + X^\top \widehat{\beta}$

Cooperative game:
Features indices D and value function

$$v(A) = R^2_Y(X_A)$$

i.e., the R^2 coefficient of the linear regression of X_A on Y
R^2 is the % of variance explained by the model

☞ The cake $v(D)$ is the R^2 of the **full model (all the features)**
☞ We have 2^d model R^2 coefficients
For 20 features, this is more than a million model R^2s

Cooperative Game Theory and "Machine Learning"

Example: Lindeman, Merenda, and Gold (1980) indices

Data: features $X = (X_1, \dots, X_d)$, and target Y
Players: for $A \in \mathcal{P}_D$, X_A only contains features with indices in A
Model: fitted linear regression model $\widehat{f}(X) = \widehat{\beta}_0 + X^\top \widehat{\beta}$

Cooperative game:
Features indices D and value function

$$v(A) = R^2_Y(X_A)$$

i.e., the R^2 coefficient of the linear regression of X_A on Y
R^2 is the % of variance explained by the model

☞ The cake $v(D)$ is the R^2 of the **full model (all the features)**
☞ We have 2^d model R^2 coefficients
For 20 features, this is more than a million model R^2s

How can we aggregate the information of all these R^2s into something more manageable?

Allocations as an aggregation

This is **exactly what an allocation does**!

It summarizes the 2^d evaluations of v into **one quantity for each player**

Allocations as an aggregation

This is **exactly what an allocation does**!
It summarizes the 2^d evaluations of v into **one quantity for each player**

It is a mapping $\phi : D \to \mathbb{R}$, that must ideally respect <u>**one**</u> criteria:

Efficiency: $\sum_{i \in D} \phi(i) = v(D)$

This ensures that **the we actually redistribute the cake**

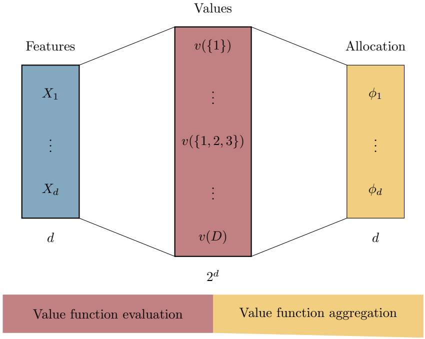
Allocations as an aggregation

This is **exactly what an allocation does**!
It summarizes the 2^d evaluations of v into **one quantity for each player**

It is a mapping $\phi : D \to \mathbb{R}$, that must ideally respect one criteria:

Efficiency: $\sum_{i \in D} \phi(i) = v(D)$

This ensures that **the we actually redistribute the cake**



In a nuthsell:

- Start with a learned model with d input features
- Chose a **value function** resulting in 2^d quantities
- Aggregate the 2^d quantities into d quantities using an **efficient allocation**

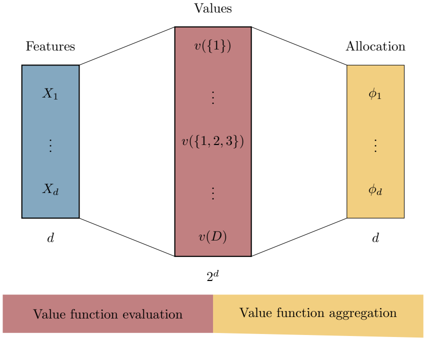
Allocations as an aggregation

This is **exactly what an allocation does**!
It summarizes the 2^d evaluations of v into **one quantity for each player**

It is a mapping $\phi : D \to \mathbb{R}$, that must ideally respect <u>one</u> criteria:

Efficiency: $\sum_{i \in D} \phi(i) = v(D)$

This ensures that **the we actually redistribute the cake**



In a nuthsell:

- Start with a learned model with d input features
- Chose a **value function** resulting in 2^d quantities
- Aggregate the 2^d quantities into d quantities using an **efficient allocation**

How can we define efficient allocations?

Allocations as a dividend sharing mechanism

The **Harsanyi (1963) dividends** of a cooperative game (D, v) are defined as:

$$\mathcal{D}_v(A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} v(B), \quad \text{or equivalently,} \quad \mathcal{D}_v(A) = v(A) - \sum_{B \in \mathcal{P}_A} \mathcal{D}_v(B)$$

Think of them as **the added-value of a coalition**: $\mathcal{D}_v(12) = v(12) - v(1) - v(2)$

Allocations as a dividend sharing mechanism

The **Harsanyi (1963) dividends** of a cooperative game (D, v) are defined as:

$$\mathcal{D}_v(A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} v(B), \quad \text{or equivalently,} \quad \mathcal{D}_v(A) = v(A) - \sum_{B \in \mathcal{P}_A} \mathcal{D}_v(B)$$

Think of them as **the added-value of a coalition**: $\mathcal{D}_v(12) = v(12) - v(1) - v(2)$

The **Harsanyi set** is a family of <u>**efficient**</u> **allocations** that **aggregate of the Harsanyi dividends**:

$$\phi(i) = \sum_{A \in \mathcal{P}_D \,:\, i \in A} \lambda_i(A) \mathcal{D}_v(A), \quad \text{where} \quad \begin{cases} \forall i \in D, \forall A \in \mathcal{P}_D, \ \lambda_i(A) \geq 0, \\ \forall A \in \mathcal{P}_D, \ \sum_{i \in D} \lambda_i(A) = 1 \end{cases}$$

parametrized by the **weight system** $\lambda : D \times \mathcal{P}_D \to \mathbb{R}$

Allocations as a dividend sharing mechanism

The **Harsanyi (1963) dividends** of a cooperative game (D, v) are defined as:

$$\mathcal{D}_v(A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} v(B), \quad \text{or equivalently,} \quad \mathcal{D}_v(A) = v(A) - \sum_{B \in \mathcal{P}_A} \mathcal{D}_v(B)$$

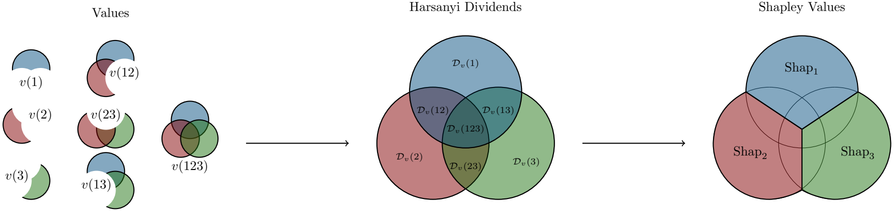
Think of them as **the added-value of a coalition**: $\mathcal{D}_v(12) = v(12) - v(1) - v(2)$

The **Harsanyi set** is a family of **efficient allocations** that **aggregate of the Harsanyi dividends**:

$$\phi(i) = \sum_{A \in \mathcal{P}_D \,:\, i \in A} \lambda_i(A) \mathcal{D}_v(A), \quad \text{where} \quad \begin{cases} \forall i \in D, \forall A \in \mathcal{P}_D, \ \lambda_i(A) \geq 0, \\ \forall A \in \mathcal{P}_D, \ \sum_{i \in D} \lambda_i(A) = 1 \end{cases}$$

parametrized by the **weight system** $\lambda : D \times \mathcal{P}_D \to \mathbb{R}$

In this setting, the **Shapley values are the egalitarian redistribution**, i.e., $\lambda_i(A) = 1/|A|$



Allocations using random orders

The **Weber (1988) set of allocations** relies on the notion of **random orders**

Allocations using random orders

The **Weber (1988) set of allocations** relies on the notion of **random orders**

Let $\mathcal{S}_D$ be the **set of permutations** $\pi = (\pi_1, \ldots, \pi_d)$ (i.e., orders) of players
For any $i \in D$, denote $\pi(i)$ the **position of player** i **in the permutation** π (i.e., $\pi_{\pi(i)} = i$)

Allocations using random orders

The **Weber (1988) set of allocations** relies on the notion of **random orders**

Let $\mathcal{S}_D$ be the **set of permutations** $\pi = (\pi_1, \ldots, \pi_d)$ (i.e., orders) of players
For any $i \in D$, denote $\pi(i)$ the **position of player** i **in the permutation** π (i.e., $\pi_{\pi(i)} = i$)

The **Weber (1988) set** is a family of **efficient allocations** as an average over the orderings

$$\begin{aligned}
\phi(i) &= \mathbb{E}_{\pi \sim p}\left[v\left(\left\{\pi_1, \ldots, \pi_{\pi(i)}\right\}\right) - v\left(\left\{\pi_1, \ldots, \pi_{\pi(i)-1}\right\}\right)\right] \\
&= \sum_{\pi \in \mathcal{S}_D} p(\pi)\left[v\left(\left\{\pi_1, \ldots, \pi_{\pi(i)}\right\}\right) - v\left(\left\{\pi_1, \ldots, \pi_{\pi(i)-1}\right\}\right)\right]
\end{aligned}$$

parametrized by a **probability mass function** p over the permutations $\mathcal{S}_D$.

Allocations using random orders

The **Weber (1988) set of allocations** relies on the notion of **random orders**

Let $\mathcal{S}_D$ be the **set of permutations** $\pi = (\pi_1, \dots, \pi_d)$ (i.e., orders) of players
For any $i \in D$, denote $\pi(i)$ the **position of player i in the permutation** π (i.e., $\pi_{\pi(i)} = i$)

The **Weber (1988) set** is a family of **efficient allocations** as an average over the orderings

$$\begin{aligned}
\phi(i) &= \mathbb{E}_{\pi \sim p}\left[v\left(\{\pi_1, \dots, \pi_{\pi(i)}\}\right) - v\left(\{\pi_1, \dots, \pi_{\pi(i)-1}\}\right)\right] \\
&= \sum_{\pi \in \mathcal{S}_D} p(\pi)\left[v\left(\{\pi_1, \dots, \pi_{\pi(i)}\}\right) - v\left(\{\pi_1, \dots, \pi_{\pi(i)-1}\}\right)\right]
\end{aligned}$$

parametrized by a **probability mass function** p over the permutations $\mathcal{S}_D$.

In this setting, the **Shapley values are <u>the uniform distribution over the permutations</u>**, i.e., $p(\pi) = 1/d!$

$$\mathrm{Shap}(i) = \frac{1}{d!} \sum_{\pi \in \mathcal{S}_D} \left[v\left(\{\pi_1, \dots, \pi_{\pi(i)}\}\right) - v\left(\{\pi_1, \dots, \pi_{\pi(i)-1}\}\right)\right]$$

The recipe

Overall blueprint for using cooperative games for XAI:

1. **Step 1: Identify a quantity of interest**
   Choose **a cake worth cutting**, e.g., point predictions $f(x)$, model variance $\mathbb{V}(f(X))$...
   ☞ Guides the **interpretation of the extracted insights**

The recipe

Overall blueprint for using cooperative games for XAI:

1. **Step 1: Identify a quantity of interest**
   Choose **a cake worth cutting**, e.g., point predictions $f(x)$, model variance $\mathbb{V}(f(X))$...
   ☞ Guides the **interpretation of the extracted insights**

2. **Step 2: Pick a value function v**
   And make sure that $v(D)$ **is equal to the quantity of interest**, e.g.,
   $\mathbb{E}[f(X) \mid X_A = x_A]$ for $f(x)$, $\mathbb{V}(\mathbb{E}[f(X) \mid X_A])$ for $\mathbb{V}(f(X))$...
   ☞ **This step is the most important** (garbage in - garbage out)

The recipe

Overall blueprint for using cooperative games for XAI:

1. **Step 1: Identify a quantity of interest**
   Choose **a cake worth cutting**, e.g., point predictions $f(x)$, model variance $\mathbb{V}(f(X))$...
   ☞ Guides the **interpretation of the extracted insights**

2. **Step 2: Pick a value function v**
   And make sure that $v(D)$ **is equal to the quantity of interest**, e.g.,
   $\mathbb{E}[f(X) \mid X_A = x_A]$ for $f(x)$, $\mathbb{V}(\mathbb{E}[f(X) \mid X_A])$ for $\mathbb{V}(f(X))$...
   ☞ **This step is the most important** (garbage in - garbage out)

3. **Step 3: Pick an efficient allocation**
   In order to summarize the information of the 2^d evaluations of v
   ☞ Less crucial and can **highlight some model behavior**

Picking a value function

Ok, but **how do I pick a value function?**

Picking a value function

Ok, but **how do I pick a value function?**

A **bad choice** can lead to **misleading insights**

e.g., **correlation/concurvity identifiability issues** (Zhang, Martinelli, and John 2024), **lack of purity** Köhler, Rügamer, and Schmid (2024)...

Picking a value function

Ok, but **how do I pick a value function?**

A **bad choice** can lead to **misleading insights**

e.g., **correlation/concurvity identifiability issues** (Zhang, Martinelli, and John 2024), **lack of purity** Köhler, Rügamer, and Schmid (2024)...

☞ **Recent theoretical developments offer an answer!...**
But several practical challenges still remain (estimation)

Picking a value function

Ok, but **how do I pick a value function?**

A **bad choice** can lead to **misleading insights**

e.g., **correlation/concurvity identifiability issues** (Zhang, Martinelli, and John 2024), **lack of purity** Köhler, Rügamer, and Schmid (2024)...

☞ **Recent theoretical developments offer an answer!...**
But several practical challenges still remain (estimation)

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

Picking a value function

Ok, but **how do I pick a value function?**

A **bad choice** can lead to **misleading insights**

e.g., **correlation/concurvity identifiability issues** (Zhang, Martinelli, and John 2024), **lack of purity** Köhler, Rügamer, and Schmid (2024)...

☞ **Recent theoretical developments offer an answer!...**
But several practical challenges still remain (estimation)

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

Conditional expectation

$$v(A) = \mathbb{E}\left[f(X) \mid X_A = x_A\right]$$

$$\mathcal{D}_v(1) = x_1 + \rho(x_1 + x_1^2 - 1) \quad \mathcal{D}_v(2) = x_2 + \rho(x_2 + x_2^2 - 1)$$

$$\mathcal{D}_v(12) = x_1 x_2 - \rho(x_1 + x_1^2 + x_2 + x_2^2 - 1)$$

$$\mathrm{Shap}_v(\{1\}) = x_1 + \frac{\rho}{2}(x_1 + x_1^2 - x_2 - x_2^2 - 1) + \frac{x_1 x_2}{2}$$

$$\mathrm{Shap}(\{2\}) = x_2 + \frac{\rho}{2}(x_2 + x_2^2 - x_1 - x_1^2 - 1) + \frac{x_1 x_2}{2}$$

Picking a value function

Ok, but **how do I pick a value function?**

A **bad choice** can lead to **misleading insights**

e.g., **correlation/concurvity identifiability issues** (Zhang, Martinelli, and John 2024), **lack of purity** Köhler, Rügamer, and Schmid (2024)...

☞ **Recent theoretical developments offer an answer!...**
But several practical challenges still remain (estimation)

$$G(X) = X_1 + X_2 + X_1X_2, \quad X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

Conditional expectation

$v(A) = \mathbb{E}\left[f(X) \mid X_A = x_A\right]$

$$\mathcal{D}_v(1) = x_1 + \rho(x_1 + x_1^2 - 1) \quad \mathcal{D}_v(2) = x_2 + \rho(x_2 + x_2^2 - 1)$$

$$\mathcal{D}_v(12) = x_1x_2 - \rho(x_1 + x_1^2 + x_2 + x_2^2 - 1)$$

$$\mathrm{Shap}_v(\{1\}) = x_1 + \frac{\rho}{2}(x_1 + x_1^2 - x_2 - x_2^2 - 1) + \frac{x_1x_2}{2}$$

$$\mathrm{Shap}(\{2\}) = x_2 + \frac{\rho}{2}(x_2 + x_2^2 - x_1 - x_1^2 - 1) + \frac{x_1x_2}{2}$$

Oblique projections

Il Idrissi et al. (2025)

$$\mathcal{D}_v(1) = x_1 \quad \mathcal{D}_v(2) = x_2$$

$$\mathcal{D}_v(12) = x_1x_2$$

$$\mathrm{Shap}_v(\{1\}) = x_1 + \frac{x_1x_2}{2}$$

$$\mathrm{Shap}(\{2\}) = x_2 + \frac{x_1x_2}{2}$$

Picking an allocation

Ok, but **how do I pick an allocation?**

Picking an allocation

Ok, but **how do I pick an allocation?**

No definite answer (yet)

☞ It is hard to compare because there is a lack of metrics between allocations

Picking an allocation

Ok, but **how do I pick an allocation?**

No definite answer (yet)
☞ It is hard to compare because there is a lack of metrics between allocations

But, sometimes they can have interesting properties

Picking an allocation

Ok, but **how do I pick an allocation?**

No definite answer (yet)
☞ It is hard to compare because there is a lack of metrics between allocations

But, sometimes they can have interesting properties

Example: Proportional Marginal Effects (Herin et al. 2024)

- **Quantity of interest:** $\mathbb{V}\left(f(X)\right)$
- **Value function:** $v(A) = \mathbb{E}\left[\mathbb{V}\left(f(X) \mid X_{D \setminus A}\right)\right]$
- **Allocation:** *Proportional values*

$$p(\pi) = \frac{L(\pi)}{\sum_{\sigma \in \mathcal{S}_D} L(\sigma)}, \quad L(\pi) = \exp\left(-\sum_{j \in D} \log\left(v\left(\{\pi_1, \dots \pi_{\pi(j)}\}\right)\right)\right)$$

Picking an allocation

Ok, but **how do I pick an allocation?**

No definite answer (yet)

☞ It is hard to compare because there is a lack of metrics between allocations

But, sometimes they can have interesting properties

Example: Proportional Marginal Effects (Herin et al. 2024)

- **Quantity of interest:** $\mathbb{V}\left(f(X)\right)$
- **Value function:** $v(A) = \mathbb{E}\left[\mathbb{V}\left(f(X) \mid X_{D\setminus A}\right)\right]$
- **Allocation:** *Proportional values*

$$p(\pi) = \frac{L(\pi)}{\sum_{\sigma \in \mathcal{S}_D} L(\sigma)}, \quad L(\pi) = \exp\left(-\sum_{j \in D} \log\left(v\left(\{\pi_1, \dots \pi_{\pi(j)}\}\right)\right)\right)$$

Proposition *(Exogeneity detection).*

$$PME_i = 0 \iff X_i \text{ is not in the model.}$$

Conclusion

Key take-aways:

- **Three ingredients** to using cooperative game for XAI:
  ☞ The **quantity of interest**, the **value function**, and the **allocation**

Conclusion

Key take-aways:

- **Three ingredients** to using cooperative game for XAI:
  ☞ The **quantity of interest**, the **value function**, and the **allocation**
- The **Shapley values** are **one example of allocation**
  ☞ **Egalitarian redistribution** (Harsanyi set), **uniformly likely orders** (Weber set)

Conclusion

Key take-aways:

- **Three ingredients** to using cooperative game for XAI:
  ☞ The **quantity of interest**, the **value function**, and the **allocation**
- The **Shapley values** are **one example of allocation**
  ☞ **Egalitarian redistribution** (Harsanyi set), **uniformly likely orders** (Weber set)
- The **choice of value function is crucial**
  ☞ Risk of misleading decompositions

Conclusion

Key take-aways:

- **Three ingredients** to using cooperative game for XAI:
  ☞ The **quantity of interest**, the **value function**, and the **allocation**
- The **Shapley values** are **one example of allocation**
  ☞ **Egalitarian redistribution** (Harsanyi set), **uniformly likely orders** (Weber set)
- The **choice of value function is crucial**
  ☞ Risk of misleading decompositions
- The **choice of allocation is secondary**
  ☞ It amounts to choosing an aggregation flavor

References i

Harsanyi, J. C. 1963. "A Simplified Bargaining Model for the n-Person Cooperative Game." Publisher: [Economics Department of the University of Pennsylvania, Wiley, Institute of Social and Economic Research, Osaka University], *International Economic Review* 4 (2): 194–220. ISSN: 0020-6598. https://doi.org/10.2307/2525487. https://www.jstor.org/stable/2525487.

Herin, M., M. Il Idrissi, V. Chabridon, and B. Iooss. 2024. "Proportional Marginal Effects for Global Sensitivity Analysis" [in en]. *SIAM/ASA Journal on Uncertainty Quantification* 12, no. 2 (June): 667–692. ISSN: 2166-2525. https://doi.org/10.1137/22M153032X. https://epubs.siam.org/doi/10.1137/22M153032X.

Il Idrissi, Marouane, Nicolas Bousquet, Fabrice Gamboa, Bertrand Iooss, and Jean-Michel Loubes. 2025. "Hoeffding decomposition of functions of random dependent variables." *Journal of Multivariate Analysis* 208 (July): 105444. ISSN: 0047-259X. https://doi.org/10.1016/j.jmva.2025.105444. https://www.sciencedirect.com/science/article/pii/S0047259X25000399.

Köhler, D., D. Rügamer, and M. Schmid. 2024. *Achieving interpretable machine learning by functional decomposition of black-box models into explainable predictor effects.* ArXiv:2407.18650 [cs, stat], July. Accessed August 26, 2024. http://arxiv.org/abs/2407.18650.

Lindeman, R. H., P. F. Merenda, and R. Z. Gold. 1980. *Introduction to Bivariate and Multivariate Analysis* [in English]. Scott, Foresman. ISBN: 978-0-673-15099-8. https://books.google.cz/books?id=-hfvAAAAMAAJ.

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.

References ii

Zhang, Xinyu, Julien Martinelli, and S. T. John. 2024. "Challenges in interpretability of additive models." In *Proceedings of the XAI Workshop @ IJCAI 2024.* ArXiv:2504.10169 [cs]. arXiv. https://doi.org/10.48550/arXiv.2504.10169. http://arxiv.org/abs/2504.10169.

Thank you for your attention!

Any questions?

MAROUANEILIDRISSI.COM

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC).
Nous remercions le Conseil de Recherches en Sciences Naturelles et en Génie du Canada (CRSNG) de son soutien.