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*

Dalhousie Statistics Seminar Series
Dalhousie University - Halifax NS, Canada (Online)
November 7, 2025

Marouane Il Idrissi[1,2]



Marouane IL IDRISSI

il_idrissi.marouane@uqam.ca - marouaneilidrissi.com

CIFRE Ph.D. - Université de Toulouse and EDF R&D (2021-2024)
Postdoctoral Researcher (since Aug. 2024)
CANSSI Distinguished Postdoctoral Fellow (since Sept. 2025)

Département de Mathématiques, Université du Québec à Montréal
Institut Intelligence et Données, Université Laval

Project: *Interpretability of black-box machine learning models*
With Arthur Charpentier (UQÀM), Marie-Pier Côté (ULaval)

Research interests:

Statistical Learning • XAI • Uncertainty Quantification • Sensitivity Analysis • Probability Theory • Cooperative game theory • Functional analysis

Context

Why are ML models not widely integrated into critical systems yet?

Context

Why are ML models not widely integrated into critical systems yet?

Critical systems: *nuclear power plants, hydroelectric dams, the stock market, planes, the human body...*

Context

Why are ML models not widely integrated into critical systems yet?

Critical systems: *nuclear power plants, hydroelectric dams, the stock market, planes, the human body...*

Main reason: The decision-making process must be **justifiable** and **justified**.
e.g., statistical arguments

Context

Why are ML models not widely integrated into critical systems yet?

Critical systems: *nuclear power plants, hydroelectric dams, the stock market, planes, the human body...*

Main reason: The decision-making process must be **justifiable** and **justified**.
e.g., statistical arguments

However, **eXplainable AI (XAI) methods** are often backed by **empirical arguments**
"SOTA" methods, testing on limited benchmarks...

Context

Why are ML models not widely integrated into critical systems yet?

Critical systems: *nuclear power plants, hydroelectric dams, the stock market, planes, the human body...*

Main reason: The decision-making process must be **justifiable** and **justified**.
e.g., statistical arguments

However, **eXplainable AI (XAI) methods** are often backed by **empirical arguments**
"SOTA" methods, testing on limited benchmarks...

This is **not enough** to convince safety/regulatory authorities...

Our position:
Before choosing an XAI method, we need to understand it theoretically
☞ Understand the method before "explaining" the phenomenon

Context

In this talk:

Revisit **post-hoc model-agnostic** XAI methods based on **cooperative game theory**

Context

In this talk:
Revisit **post-hoc model-agnostic** XAI methods based on **cooperative game theory**

They promise a variable **influence quantification** of **non-linear** ML models
☞ **Not always ”well-defined”, especially with dependent covariates**

Context

In this talk:
Revisit **post-hoc model-agnostic** XAI methods based on **cooperative game theory**

They promise a variable **influence quantification** of **non-linear** ML models
☞ **Not always ”well-defined”, especially with dependent covariates**

They heavily rely on the notion of **the Shapley values**
☞ **Often misunderstood and misused in XAI**

Context

In this talk:
Revisit **post-hoc model-agnostic** XAI methods based on **cooperative game theory**

They promise a variable **influence quantification** of **non-linear** ML models
☞ **Not always "well-defined", especially with dependent covariates**

They heavily rely on the notion of **the Shapley values**
☞ **Often misunderstood and misused in XAI**

How can we leverage cooperative game theory to extract tractable insights on the behavior of black-box models?

Context

In this talk:
Revisit **post-hoc model-agnostic** XAI methods based on **cooperative game theory**

They promise a variable **influence quantification** of **non-linear** ML models
☞ **Not always ”well-defined”, especially with dependent covariates**

They heavily rely on the notion of **the Shapley values**
☞ **Often misunderstood and misused in XAI**

How can we leverage cooperative game theory to extract tractable insights on the behavior of black-box models?

Side quests:

- Understand **what the Shapley values are** and how to go **beyond them**
- Introduce some of **the open questions and challenges** that I will take on during the next two years

Framework and notations

☞ Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the sample (probability) space.

Framework and notations

☞ Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the sample (probability) space.

☞ Let $X = (X_1, \dots, X_d)$ be the **random inputs of a black-box model**.
A measurable mapping from Ω to a cartesian product of Polish spaces $E = \bigtimes_{i \in D} E_i$.

Framework and notations

☞ Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the sample (probability) space.

☞ Let $X = (X_1, \dots, X_d)$ be the **random inputs of a black-box model**.
A measurable mapping from Ω to a cartesian product of Polish spaces $E = \bigtimes_{i \in D} E_i$.

☞ Let $D = \{1, \dots, d\}$ and $\mathcal{P}_D$ be the **set of subsets (power-set)** of D.

Framework and notations

☞ Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the sample (probability) space.

☞ Let $X = (X_1, \dots, X_d)$ be the **random inputs of a black-box model**.
A measurable mapping from Ω to a cartesian product of Polish spaces $E = \bigtimes_{i \in D} E_i$.

☞ Let $D = \{1, \dots, d\}$ and $\mathcal{P}_D$ be the **set of subsets (power-set)** of D.

☞ For $A \in \mathcal{P}_D$, let X_A be a **subset of the inputs**.
A mapping from Ω to $E_A = \bigtimes_{i \in A} E_i$.

Framework and notations

☞ Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the sample (probability) space.

☞ Let $X = (X_1, \dots, X_d)$ be the **random inputs of a black-box model**.
A measurable mapping from Ω to a cartesian product of Polish spaces $E = \bigtimes_{i \in D} E_i$.

☞ Let $D = \{1, \dots, d\}$ and $\mathcal{P}_D$ be the **set of subsets (power-set)** of D.

☞ For $A \in \mathcal{P}_D$, let X_A be a **subset of the inputs**.
A mapping from Ω to $E_A = \bigtimes_{i \in A} E_i$.

☞ Let $\widehat{f}$ denote the **black-box ML model**, and $\widehat{f}(X)$ be the **random output**.
A measurable mapping from E to $\mathbb{R}$. $\widehat{f}(X)$ is a **random variable**.

Remark . We take a **post-hoc**, **model-agnostic** approach.

Cooperative Game Theory

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

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

Two ingredients:

- $D = \{1, \dots, d\}$, a **set of players**
  The power-set $\mathcal{P}_D$ is the **set of coalitions of players**
- $v : \mathcal{P}_D \to \mathbb{R}$, a **value function**
  It **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 = The art of sharing a cake"

Two ingredients:

- $D = \{1, \ldots, d\}$, a **set of players**
  The power-set $\mathcal{P}_D$ is the **set of coalitions of players**
- $v : \mathcal{P}_D \to \mathbb{R}$, a **value function**
  It **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 concern of the theory of cooperative games:

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

A famous example - LMG indices

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

A famous example - LMG indices

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

Data: covariates $X = (X_1, \ldots, X_d)$, and target Y
Model: estimated linear regression model $\widehat{f}(X) = \widehat{\beta}_0 + X^\top \widehat{\beta}$

A famous example - LMG indices

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

Data: covariates $X = (X_1, \ldots, X_d)$, and target Y
Model: estimated linear regression model $\widehat{f}(X) = \widehat{\beta}_0 + X^\top \widehat{\beta}$

Cooperative game:
Players: the covariates $(X_1, \ldots, X_d)$ (rather, their indices D)
Coalitions: for $A \in \mathcal{P}_D$, the subset of covariates X_A
Value function: $v(A) = R^2_Y(X_A)$ (the R^2 coefficient of the linear model only using X_A)

A famous example - LMG indices

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

Data: covariates $X = (X_1, \ldots, X_d)$, and target Y
Model: estimated linear regression model $\widehat{f}(X) = \widehat{\beta}_0 + X^\top \widehat{\beta}$

Cooperative game:
Players: the covariates $(X_1, \ldots, X_d)$ (rather, their indices D)
Coalitions: for $A \in \mathcal{P}_D$, the subset of covariates X_A
Value function: $v(A) = R^2_Y(X_A)$ (the R^2 coefficient of the linear model only using X_A)

☞ **The cake:** $v(D)$, the R^2 of the **full model**

A famous example - LMG indices

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

Data: covariates $X = (X_1, \dots, X_d)$, and target Y
Model: estimated linear regression model $\widehat{f}(X) = \widehat{\beta}_0 + X^\top \widehat{\beta}$

Cooperative game:
Players: the covariates $(X_1, \dots, X_d)$ (rather, their indices D)
Coalitions: for $A \in \mathcal{P}_D$, the subset of covariates X_A
Value function: $v(A) = R^2_Y(X_A)$ (the R^2 coefficient of the linear model only using X_A)

☞ **The cake:** $v(D)$, the R^2 of the **full model**

Evaluate the **value function** $\forall A \in \mathcal{P}_D \iff$ The R^2 of all the 2^d nested linear models
For 20 covariate: more than a million R^2 coefficients

A famous example - LMG indices

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

Data: covariates $X = (X_1, \dots, X_d)$, and target Y
Model: estimated linear regression model $\widehat{f}(X) = \widehat{\beta}_0 + X^\top \widehat{\beta}$

Cooperative game:
Players: the covariates $(X_1, \dots, X_d)$ (rather, their indices D)
Coalitions: for $A \in \mathcal{P}_D$, the subset of covariates X_A
Value function: $v(A) = R^2_Y(X_A)$ (the R^2 coefficient of the linear model only using X_A)

☞ **The cake:** $v(D)$, the R^2 of the **full model**

Evaluate the **value function** $\forall A \in \mathcal{P}_D \iff$ The R^2 of all the 2^d nested linear models
For 20 covariate: more than a million R^2 coefficients

☞ **This is way too much information to process...**

A famous example - LMG indices

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

Data: covariates $X = (X_1, \dots, X_d)$, and target Y
Model: estimated linear regression model $\widehat{f}(X) = \widehat{\beta}_0 + X^\top \widehat{\beta}$

Cooperative game:
Players: the covariates $(X_1, \dots, X_d)$ (rather, their indices D)
Coalitions: for $A \in \mathcal{P}_D$, the subset of covariates X_A
Value function: $v(A) = R^2_Y(X_A)$ (the R^2 coefficient of the linear model only using X_A)

☞ **The cake:** $v(D)$, the R^2 of the **full model**

Evaluate the **value function** $\forall A \in \mathcal{P}_D \iff$ The R^2 of all the 2^d nested linear models
For 20 covariate: more than a million R^2 coefficients

☞ **This is way too much information to process...**

Is it possible to aggregate this information (2^d coefficients) into something more manageable?

Allocations as an aggregation

This is **exactly the role of an allocation**!

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

Allocations as an aggregation

This is **exactly the role of an allocation**!
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:

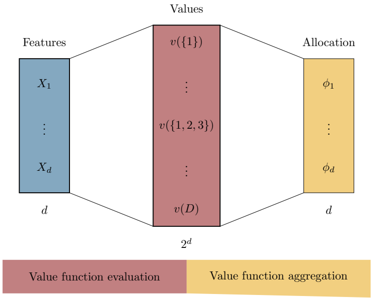
- **Efficiency:** $\sum_{i \in D} \phi(i) = v(D)$
  ☞ Ensures that **the we actually redistribute the cake**

Allocations as an aggregation

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

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

- **Efficiency:** $\sum_{i \in D} \phi(i) = v(D)$
  ☞ Ensures that **the we actually redistribute the cake**



In a nuthsell:

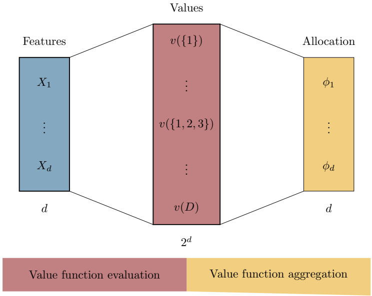
- Start with a learned model with d input features
- Chose a **value function** resulting in 2^d quantities
- Aggregate the 2^d quantities into d quantities using an **efficient allocation**

Allocations as an aggregation

This is **exactly the role of an allocation**!
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:** $\sum_{i \in D} \phi(i) = v(D)$
  ☞ Ensures that **the we actually redistribute the cake**



In a nuthsell:

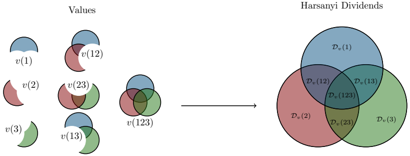
- Start with a learned model with d input features
- Chose a **value function** resulting in 2^d quantities
- Aggregate the 2^d quantities into d quantities using an **efficient allocation**

Is there a way to define efficient allocations?

Allocations as a dividend sharing mechanism

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

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

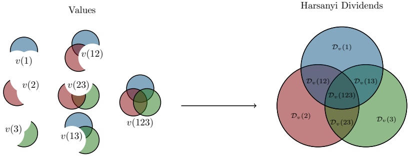


They quantify **the added-value of a coalition**:
$\mathcal{D}_v(12) = v(12) - v(1) - v(2)$

Allocations as a dividend sharing mechanism

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

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



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

They are the **Möbius inverse** of the **value function**

Proposition *(Möbius inversion on power-sets (Rota 1964; Kung, Rota, and Hung Yan 2012)).*

For any two set functions $v : \mathcal{P}_D \to \mathbb{R}$, $\mathcal{D} : \mathcal{P}_D \to \mathbb{R}$, the following equivalence holds:

$$\forall A \in \mathcal{P}_D, \quad v(A) = \sum_{B \in \mathcal{P}_A} \mathcal{D}(B), \quad \iff \quad \forall A \in \mathcal{P}_D, \quad \mathcal{D}(A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} v(B).$$

(i.e., generalized inclusion-exclusion principle)

Shapley values as the egalitarian dividend sharing mechanism

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** $\lambda : D \times \mathcal{P}_D \to \mathbb{R}$

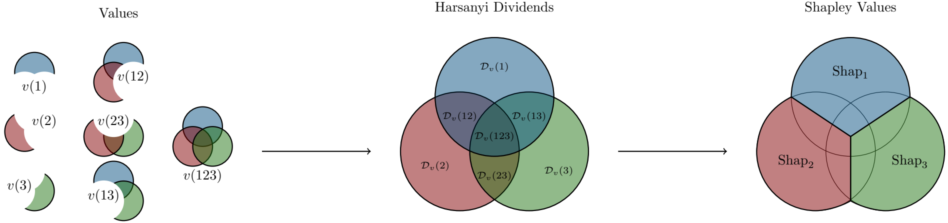
Shapley values as the egalitarian dividend sharing mechanism

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** 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** over the permutations $\mathcal{S}_D$.

In this setting, the **Shapley values are the uniform distribution over the permutations**, 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:
(I., Charpentier, and Fernandes Machado 2025)

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:
(I., Charpentier, and Fernandes Machado 2025)

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:
(I., Charpentier, and Fernandes Machado 2025)

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**

Challenge 1:
Choosing a value function

Picking a value function

How do we pick relevant value functions?

Picking a value function

How do we pick relevant value functions?

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

How do we pick relevant value functions?

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

How do we pick relevant value functions?

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)

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

Picking a value function

How do we pick relevant value functions?

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)

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

Conditional expectation

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

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

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

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

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

Picking a value function

How do we pick relevant value functions?

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)

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

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

I. et al. (2025a)

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

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

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

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

Oblique projections

These **oblique projections** generalize **conditional expectations** thanks to a **canonical direct-sum decomposition of** $\mathbb{L}^2$.

Oblique projections

These **oblique projections** generalize **conditional expectations** thanks to a **canonical direct-sum decomposition of** $\mathbb{L}^2$.

- ☞ Let σ_X be the σ-algebra generated by X
- ☞ Let σ_A be the σ-algebra generated by X_A, $A \in \mathcal{P}_D$
- ☞ Let $\sigma_\emptyset$ be the $\mathbb{P}$-trivial σ-algebra
- ☞ For a sub-σ-algebra $\mathcal{B} \subseteq \mathcal{F}$, denote $\mathbb{L}^2(\mathcal{B}) := \mathbb{L}^2\left((\Omega, \mathcal{B}, \mathbb{P})\right)$

Oblique projections

These **oblique projections** generalize **conditional expectations** thanks to a **canonical direct-sum decomposition of** $\mathbb{L}^2$.

- ☞ Let σ_X be the σ-algebra generated by X
- ☞ Let σ_A be the σ-algebra generated by X_A, $A \in \mathcal{P}_D$
- ☞ Let $\sigma_\emptyset$ be the $\mathbb{P}$-trivial σ-algebra
- ☞ For a sub-σ-algebra $\mathcal{B} \subseteq \mathcal{F}$, denote $\mathbb{L}^2(\mathcal{B}) := \mathbb{L}^2\left((\Omega, \mathcal{B}, \mathbb{P})\right)$

Theorem *(I. et al. 2025a)*. Under very mild assumptions on X, for every $A \in \mathcal{P}_D$,

$$\mathbb{L}^2\left(\sigma_A\right) = \bigoplus_{B \in \mathcal{P}_A} V_B.$$

where $V_\emptyset = \mathbb{L}^2\left(\sigma_\emptyset\right)$, and $V_B = \left[+_{C \in \mathcal{P}_B, C \neq B} V_C\right]^{\perp_B}$, with $\perp_B$ denoting the orthogonal complement in $\mathbb{L}^2\left(\sigma_B\right)$.

Oblique projections

These **oblique projections** generalize **conditional expectations** thanks to a **canonical direct-sum decomposition of** $\mathbb{L}^2$.

- Let σ_X be the σ-algebra generated by X
- Let σ_A be the σ-algebra generated by X_A, $A \in \mathcal{P}_D$
- Let $\sigma_\emptyset$ be the $\mathbb{P}$-trivial σ-algebra
- For a sub-σ-algebra $\mathcal{B} \subseteq \mathcal{F}$, denote $\mathbb{L}^2(\mathcal{B}) := \mathbb{L}^2\left((\Omega, \mathcal{B}, \mathbb{P})\right)$

Theorem *(I. et al. 2025a)*. Under very mild assumptions on X, for every $A \in \mathcal{P}_D$,

$$\mathbb{L}^2(\sigma_A) = \bigoplus_{B \in \mathcal{P}_A} V_B.$$

where $V_\emptyset = \mathbb{L}^2(\sigma_\emptyset)$, and $V_B = \left[+_{C \in \mathcal{P}_B, C \neq B} V_C\right]^{\perp_B}$, with $\perp_B$ denoting the orthogonal complement in $\mathbb{L}^2(\sigma_B)$.

- $\mathbb{E}[\cdot \mid X_A]$: Projection onto $\mathbb{L}^2(\sigma_A)$ parallel to $\mathbb{L}^2(\sigma_A)^\perp$ in $\mathbb{L}^2(\sigma_\mathcal{F})$
- **Oblique projection**: Projection onto $\mathbb{L}^2(\sigma_A)$ parallel to $\bigoplus_{B \in \mathcal{P}_D \setminus \mathcal{P}_A} V_A$

Oblique projections

These **oblique projections** generalize **conditional expectations** thanks to a **canonical direct-sum decomposition of** $\mathbb{L}^2$.

- Let σ_X be the σ-algebra generated by X
- Let σ_A be the σ-algebra generated by X_A, $A \in \mathcal{P}_D$
- Let $\sigma_\emptyset$ be the $\mathbb{P}$-trivial σ-algebra
- For a sub-σ-algebra $\mathcal{B} \subseteq \mathcal{F}$, denote $\mathbb{L}^2(\mathcal{B}) := \mathbb{L}^2\left((\Omega, \mathcal{B}, \mathbb{P})\right)$

Theorem *(I. et al. 2025a)*. Under very mild assumptions on X, for every $A \in \mathcal{P}_D$,

$$\mathbb{L}^2\left(\sigma_A\right) = \bigoplus_{B \in \mathcal{P}_A} V_B.$$

where $V_\emptyset = \mathbb{L}^2\left(\sigma_\emptyset\right)$, and $V_B = \left[+_{C \in \mathcal{P}_B, C \neq B} V_C\right]^{\perp_B}$, with $\perp_B$ denoting the orthogonal complement in $\mathbb{L}^2\left(\sigma_B\right)$.

- $\mathbb{E}\left[\cdot \mid X_A\right]$: Projection onto $\mathbb{L}^2\left(\sigma_A\right)$ parallel to $\mathbb{L}^2\left(\sigma_A\right)^\perp$ in $\mathbb{L}^2\left(\sigma_\mathcal{F}\right)$
- **Oblique projection**: Projection onto $\mathbb{L}^2\left(\sigma_A\right)$ parallel to $\bigoplus_{B \in \mathcal{P}_D \setminus \mathcal{P}_A} V_A$

Open question: Great contenders, but **we don't know how to estimate them**

Challenge 2:
Picking an allocation

Picking an allocation

How do we pick a ”good” allocation?

Picking an allocation

How do we pick a "good" allocation?

It is still an **open question**:

- Theory of cooperative games is a great source of inspiration for **allocations**
- There's limited research on a "goal-oriented" framework to fit different scenarios

Picking an allocation

How do we pick a "good" allocation?

It is still an **open question**:

- Theory of cooperative games is a great source of inspiration for **allocations**
- There's limited research on a "goal-oriented" framework to fit different scenarios

But, some choices can have interesting properties

Picking an allocation

How do we pick a "good" allocation?

It is still an **open question**:

- Theory of cooperative games is a great source of inspiration for **allocations**
- There's limited research on a "goal-oriented" framework to fit different scenarios

But, some choices 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

How do we pick a "good" allocation?

It is still an **open question**:

- Theory of cooperative games is a great source of inspiration for **allocations**
- There's limited research on a "goal-oriented" framework to fit different scenarios

But, some choices 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*

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

Proposition *(Exogeneity detection)*.

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

Challenge 3:
Computational aspects

Computational aspects - Monte-Carlo type estimates

We need to **evaluate the value function** for the 2^d coalitions

Computational aspects - Monte-Carlo type estimates

We need to **evaluate the value function** for the 2^d coalitions

For certain **value functions**, this implies **fitting** 2^d **regression models**
e.g., $v(A) = \mathbb{E}\left[\widehat{f}(X) \mid X_A\right]$ or $v(A) = \mathbb{V}\left(\mathbb{E}\left[\widehat{f}(X) \mid X_A\right]\right)$

Computational aspects - Monte-Carlo type estimates

We need to **evaluate the value function** for the 2^d coalitions

For certain **value functions**, this implies **fitting** 2^d **regression models**
e.g., $v(A) = \mathbb{E}\left[\widehat{f}(X) \mid X_A\right]$ or $v(A) = \mathbb{V}\left(\mathbb{E}\left[\widehat{f}(X) \mid X_A\right]\right)$

Leverage the Weber set and **sample the ordering of players** according to pmf p

Computational aspects - Monte-Carlo type estimates

We need to **evaluate the value function** for the 2^d coalitions

For certain **value functions**, this implies **fitting** 2^d **regression models**
e.g., $v(A) = \mathbb{E}\left[\widehat{f}(X) \mid X_A\right]$ or $v(A) = \mathbb{V}\left(\mathbb{E}\left[\widehat{f}(X) \mid X_A\right]\right)$

Leverage the Weber set and **sample the ordering of players** according to pmf p

Proposition *(I. et al. 2025b)*. Let p be a pmf over $\mathcal{S}_D$. Let $m > 0$ and $\pi_1, \ldots, \pi_m$ be an i.i.d. sample drawn from p. Assume that $0 < \mathbb{E}_p\left[\left(v(\pi^j) - v(\pi^j \setminus \{j\})\right)^2\right] < \infty$. Then, for every $j \in D$,

$$\widehat{\phi_v}(j) = \frac{1}{m} \sum_{i=1}^{m} \left[v(\pi^j) - v(\pi^j \setminus \{j\})\right],$$

is an unbiased, strongly consistent, and asymptotically normal estimators of $\phi_v(j) := \mathbb{E}_p[h(\pi)]$.

Generalize the estimator of the **Shapley values** by uniform sampling (Štrumbelj and Kononenko 2014)

Computational aspects - Monte-Carlo type estimates

We need to **evaluate the value function** for the 2^d coalitions

For certain **value functions**, this implies **fitting** 2^d **regression models**
e.g., $v(A) = \mathbb{E}\left[\widehat{f}(X) \mid X_A\right]$ or $v(A) = \mathbb{V}\left(\mathbb{E}\left[\widehat{f}(X) \mid X_A\right]\right)$

Leverage the Weber set and **sample the ordering of players** according to pmf p

> **Proposition** *(I. et al. 2025b)*. Let p be a pmf over $\mathcal{S}_D$. Let $m > 0$ and $\pi_1, \dots, \pi_m$ be an i.i.d. sample drawn from p. Assume that $0 < \mathbb{E}_p\left[\left(v(\pi^j) - v(\pi^j \setminus \{j\})\right)^2\right] < \infty$. Then, for every $j \in D$,
>
> $$\widehat{\phi_v}(j) = \frac{1}{m}\sum_{i=1}^{m}\left[v(\pi^j) - v(\pi^j \setminus \{j\})\right],$$
>
> is an unbiased, strongly consistent, and asymptotically normal estimators of $\phi_v(j) := \mathbb{E}_p[h(\pi)]$.

Generalize the estimator of the **Shapley values** by uniform sampling (Štrumbelj and Kononenko 2014)

☞ **Worst-case scenario:** $m \times d$ models to train instead of $d!$.

Computational aspects - Importance Sampling

We can **recycle the sampled value function evaluations** for several allocation estimates

Computational aspects - Importance Sampling

We can **recycle the sampled value function evaluations** for several allocation estimates

Proposition *(I. et al. 2025b)*. Let p and p' be pmfs over $\mathcal{S}_D$. Let $m > 0$ and $\pi_1, \ldots, \pi_m$ be an i.i.d. sample drawn from p. Assume that $0 < \mathbb{E}_p \left[\frac{p'(\pi)}{p(\pi)} \left(v(\pi^j) - v(\pi^j \setminus \{j\}) \right)^2 \right] < \infty$. Then, for every $j \in D$,

$$\widehat{\phi'_v}^{\mathrm{IS}}(j) = \frac{1}{m} \sum_{i=1}^{m} \frac{p'(\pi_i)}{p(\pi_i)} \left[v(\pi^j) - v(\pi^j \setminus \{j\}) \right]$$

are unbiased, strongly consistent, and asymptotically normal estimators of $\phi'_v(j) := \mathbb{E}_{p'}[h(\pi)]$.

Computational aspects - Importance Sampling

We can **recycle the sampled value function evaluations** for several **allocation** estimates

Proposition *(I. et al. 2025b)*. Let p and p' be pmfs over $\mathcal{S}_D$. Let $m > 0$ and $\pi_1, \dots, \pi_m$ be an i.i.d. sample drawn from p. Assume that $0 < \mathbb{E}_p\left[\frac{p'(\pi)}{p(\pi)}\left(v(\pi^j) - v(\pi^j \setminus \{j\})\right)^2\right] < \infty$. Then, for every $j \in D$,

$$\widehat{\phi'_v}^{\text{IS}}(j) = \frac{1}{m}\sum_{i=1}^{m}\frac{p'(\pi_i)}{p(\pi_i)}\left[v(\pi^j) - v(\pi^j \setminus \{j\})\right]$$

are unbiased, strongly consistent, and asymptotically normal estimators of $\phi'_v(j) := \mathbb{E}_{p'}[h(\pi)]$.

Open question: Any refinement possible?

Quasi-MC, control variates, sampling something else than permutations...

Computational aspects - Importance Sampling

We can **recycle the sampled value function evaluations** for several allocation estimates

Proposition *(I. et al. 2025b)*. Let p and p' be pmfs over $\mathcal{S}_D$. Let $m > 0$ and $\pi_1, \dots, \pi_m$ be an i.i.d. sample drawn from p. Assume that $0 < \mathbb{E}_p\left[\frac{p'(\pi)}{p(\pi)}\left(v(\pi^j) - v(\pi^j \setminus \{j\})\right)^2\right] < \infty$. Then, for every $j \in D$,

$$\widehat{\phi'_v}^{\text{IS}}(j) = \frac{1}{m}\sum_{i=1}^{m}\frac{p'(\pi_i)}{p(\pi_i)}\left[v(\pi^j) - v(\pi^j \setminus \{j\})\right]$$

are unbiased, strongly consistent, and asymptotically normal estimators of $\phi'_v(j) := \mathbb{E}_{p'}[h(\pi)]$.

Open question: Any refinement possible?

Quasi-MC, control variates, sampling something else than permutations...

Open question: Other types of computational improvement?

Maximal coalition cardinality, parallel and efficient computing, time and memorytrade-off...

Quick illustration - Conformal prediction decomposition

Uncertainty attribution based on conformal prediction (CP) intervals:

1. **Quantity of interest:** Width of the CP interval $\widehat{C}(x)$ at point x
2. **Value function:** Width of the CP interval $\widehat{C}_A(x)$ only using covariates X_A
3. **Efficient allocation:** Shapley values

Quick illustration - Conformal prediction decomposition

Uncertainty attribution based on conformal prediction (CP) intervals:

1. **Quantity of interest:** Width of the CP interval $\widehat{C}(x)$ at point x
2. **Value function:** Width of the CP interval $\widehat{C}_A(x)$ only using covariates X_A
3. **Efficient allocation:** Shapley values

Synthetic model: The Sobol-Levitan function (Sobol′ and Levitan 1999)

$$X = (X_1, \dots, X_{16})^\top \sim \mathcal{U}(0,1)^{\times 16}, \quad \beta \in \mathbb{R}^{16}, \quad \epsilon_Y \sim \mathcal{N}(0,1), \quad Y = \exp\left[\beta^\top X\right] + \prod_{i=1}^{16} \frac{\exp[\beta_i] - 1}{\beta_i} + \epsilon_Y$$

Here, $d = 16$, $d! \approx 2 \times 10^{13}$ and $2^{16} = 65\,536$

Quick illustration - Conformal prediction decomposition

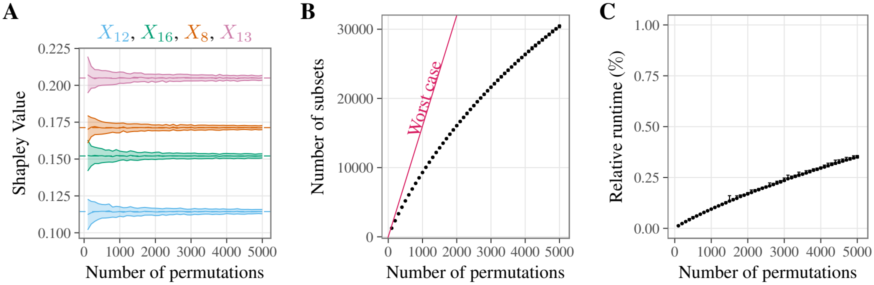
Uncertainty attribution based on conformal prediction (CP) intervals:

1. **Quantity of interest:** Width of the CP interval $\widehat{C}(x)$ at point x
2. **Value function:** Width of the CP interval $\widehat{C}_A(x)$ only using covariates X_A
3. **Efficient allocation:** Shapley values

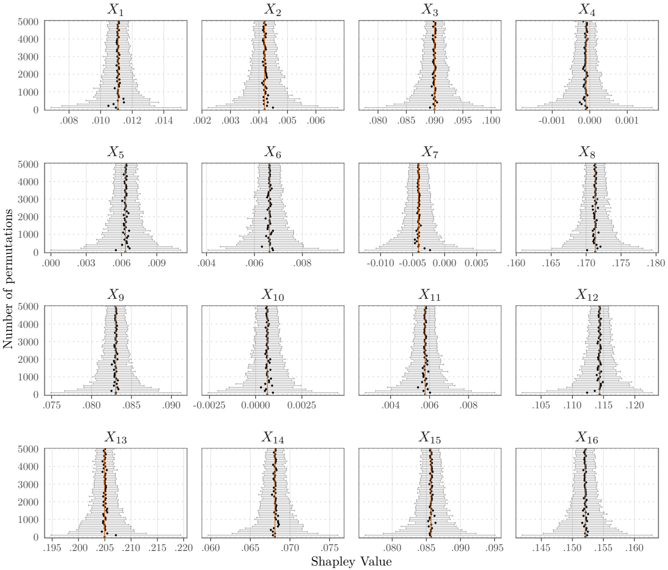
Synthetic model: The Sobol-Levitan function (Sobol′ and Levitan 1999)

$$X = (X_1, \ldots, X_{16})^\top \sim \mathcal{U}(0,1)^{\times 16}, \quad \beta \in \mathbb{R}^{16}, \quad \epsilon_Y \sim \mathcal{N}(0,1), \quad Y = \exp\left[\beta^\top X\right] + \prod_{i=1}^{16} \frac{\exp[\beta_i] - 1}{\beta_i} + \epsilon_Y$$

Here, $d = 16$, $d! \approx 2 \times 10^{13}$ and $2^{16} = 65\,536$



Quick illustration - Conformal prediction decomposition



Quick illustration - Conformal prediction decomposition

This **sampling strategy** allowed studying a **wide range of datasets** using allocations:

- The **Shapley values**: $\lambda_i(A) = \frac{1}{|A|}$
- The **proportional Shapley values**: $\lambda_i(A) = \frac{v(i)}{\sum_{j \in A} v(j)}$

Dataset	**Description**	n	d	Estimation
`bike`	Bike rental data	17,379	12	Exact
`blog`	Number of comments per blog posts	52,397	238	$m = 50$
`casp`	Physicochemical properties of proteins	45,730	9	Exact
`concrete`	Concrete compressive strength	1,030	8	Exact
`facebook`	Engagement of facebook posts	79,788	37	$m = 200$
`UScrime`	Crime data in the US	1,993	101	$m = 200$
`star`	Effect of reducing class size on test scores	2,161	38	$m = 200$

Quick illustration - Conformal prediction decomposition

This **sampling strategy** allowed studying a **wide range of datasets** using allocations:

- The **Shapley values**: $\lambda_i(A) = \frac{1}{|A|}$
- The **proportional Shapley values**: $\lambda_i(A) = \frac{v(i)}{\sum_{j \in A} v(j)}$

Dataset	**Description**	n	d	Estimation
`bike`	Bike rental data	17,379	12	Exact
`blog`	Number of comments per blog posts	52,397	238	$m = 50$
`casp`	Physicochemical properties of proteins	45,730	9	Exact
`concrete`	Concrete compressive strength	1,030	8	Exact
`facebook`	Engagement of facebook posts	79,788	37	$m = 200$
`UScrime`	Crime data in the US	1,993	101	$m = 200$
`star`	Effect of reducing class size on test scores	2,161	38	$m = 200$

More details in our most recent preprint:

Unveil Sources of Uncertainty: Feature Contribution to Conformal Prediction Intervals

Marouane Il Idrissi[a,b,e], Agathe Fernandes Machado[a], Ewen Gallic[c,d], Arthur Charpentier[a]

Conclusion

Key take-aways:

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

Conclusion

Key take-aways:

- **Three ingredients** to using cooperative game for ML interpretability: The **quantity of interest**, the **value function**, and the **allocation**
- The **Shapley values** are **one example of allocation**, and there are **many more**

Conclusion

Key take-aways:

- **Three ingredients** to using cooperative game for ML interpretability: The **quantity of interest**, the **value function**, and the **allocation**
- The **Shapley values** are **one example of allocation**, and there are **many more**
- The **choice of value function is crucial** for understanding the end product

Conclusion

Key take-aways:

- **Three ingredients** to using cooperative game for ML interpretability: The **quantity of interest**, the **value function**, and the **allocation**
- The **Shapley values** are **one example of allocation**, and there are **many more**
- The **choice of value function is crucial** for understanding the end product
- **An allocation** is (only) **an aggregation of information**

Conclusion

Key take-aways:

- **Three ingredients** to using cooperative game for ML interpretability: The **quantity of interest**, the **value function**, and the **allocation**
- The **Shapley values** are **one example of allocation**, and there are **many more**
- The **choice of value function is crucial** for understanding the end product
- **An allocation** is (only) **an aggregation of information**

Future work:

- **Oblique projection estimation:** Good contenders for **value functions**

Conclusion

Key take-aways:

- **Three ingredients** to using cooperative game for ML interpretability: The **quantity of interest**, the **value function**, and the allocation
- The **Shapley values** are **one example of allocation**, and there are **many more**
- The **choice of value function is crucial** for understanding the end product
- **An** allocation is (only) **an aggregation of information**

Future work:

- **Oblique projection estimation:** Good contenders for **value functions**
- **Refining existing sampling strategies**, and **finding new ones**

Conclusion

Key take-aways:

- **Three ingredients** to using cooperative game for ML interpretability: The **quantity of interest**, the **value function**, and the **allocation**
- The **Shapley values** are **one example of allocation**, and there are **many more**
- The **choice of value function is crucial** for understanding the end product
- **An allocation** is (only) **an aggregation of information**

Future work:

- **Oblique projection estimation:** Good contenders for **value functions**
- **Refining existing sampling strategies**, and **finding new ones**
- **Efficient implementations and open-source code** for a broader adoption

Conclusion

Key take-aways:

- **Three ingredients** to using cooperative game for ML interpretability: The **quantity of interest**, the **value function**, and the **allocation**
- The **Shapley values** are **one example of allocation**, and there are **many more**
- The **choice of value function is crucial** for understanding the end product
- **An allocation** is (only) **an aggregation of information**

Future work:

- **Oblique projection estimation:** Good contenders for **value functions**
- **Refining existing sampling strategies**, and **finding new ones**
- **Efficient implementations and open-source code** for a broader adoption
- **Decomposition of other quantity of interests** (e.g., fairness and calibration metrics)

Conclusion

Key take-aways:

- **Three ingredients** to using cooperative game for ML interpretability: The **quantity of interest**, the **value function**, and the **allocation**
- The **Shapley values** are **one example of allocation**, and there are **many more**
- The **choice of value function is crucial** for understanding the end product
- **An allocation** is (only) **an aggregation of information**

Future work:

- **Oblique projection estimation:** Good contenders for **value functions**
- **Refining existing sampling strategies**, and **finding new ones**
- **Efficient implementations and open-source code** for a broader adoption
- **Decomposition of other quantity of interests** (e.g., fairness and calibration metrics)
- **Inject causal interpretation in the construction of allocations**

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. I., 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.

I., M., Nicolas Bousquet, Fabrice Gamboa, Bertrand Iooss, and Jean-Michel Loubes. 2025a. "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.

I., M., Arthur Charpentier, and Agathe Fernandes Machado. 2025. "Beyond Shapley Values: Cooperative Games for the Interpretation of Machine Learning Models." In *International Joint Conference on Artificial Intelligence (IJCAI) - Workshop on Explainable Artificial Intelligence (XAI).* Montréal, Québec, Canada: Hendrik Baier and Tobias Huber and Mor Vered and Sarath Sreedharan and Katharina Weitz and Stylianos Loukas Vasileiou, August. https://hal.science/hal-05106257.

I., M., A. Fernandes Machado, E. Gallic, and A. Charpentier. 2025b. "Unveil Sources of Uncertainty: Feature Contribution to Conformal Prediction Intervals," arXiv: 2505.13118 [cs.AI]. https://arxiv.org/abs/2505.13118.

Kung, J. P. S., G. C. Rota, and C. Hung Yan. 2012. *Combinatorics: the Rota way.* OCLC: 1226672593. New York: Cambridge University Press. ISBN: 978-0-511-80389-5.

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.

Rota, G. C. 1964. "On the foundations of combinatorial theory I. Theory of Möbius Functions." *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 2 (4): 340–368. ISSN: 1432-2064. https://doi.org/10.1007/BF00531932.

Sobol', I.M., and Yu.L. Levitan. 1999. "On the use of variance reducing multipliers in Monte Carlo computations of a global sensitivity index." *Computer Physics Communications* 117 (1): 52–61. ISSN: 0010-4655. https://doi.org/https://doi.org/10.1016/S0010-4655(98)00156-8. https://www.sciencedirect.com/science/article/pii/S0010465598001568.

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.

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.

Štrumbelj, Erik, and Igor Kononenko. 2014. "Explaining prediction models and individual predictions with feature contributions." *Knowledge and Information Systems* 41, no. 3 (December): 647–665. ISSN: 0219-3116. https://doi.org/10.1007/s10115-013-0679-x. https://link.springer.com/article/10.1007/s10115-013-0679-x.

Thank you for your attention!

Any questions?

MAROUANEILIDRISSI.COM

We acknowledge the support of the Canadian Statistical Sciences Institute (CANSSI) and the Natural Sciences and Engineering Research Council of Canada (NSERC)

Nous reconnaissons le soutien de l'Institut Canadien des Sciences Statistiques (INCASS) et du Conseil de Recherches en Sciences Naturelles et en Génie du Canada (CRSNG)