SOBOL' INDICES, SHAPLEY EFFECTS

AND A (NEW) PATH TOWARDS HANDLING DEPENDENT INPUTS

[1]**Université du Québec à Montréal** (UQÀM)
[2]**Institut Intelligence et Données** (IID) - Universite Laval
[3]**EDF R&D** - PRISME Department
[4]**Institut de Mathématiques de Toulouse**
[5]**SINCLAIR AI Lab**

Sensitivity Analysis Discord Group
Online
September 18, 2024

Marouane IL IDRISSI[12(345)], Nicolas BOUSQUET[34], Fabrice GAMBOA[4], Bertrand IOOSS[345], Jean-Michel LOUBES[4]

Marouane IL IDRISSI

ilidrissi.m@gmail.com - marouaneilidrissi.com

☞ **Before:** Ph.D. Candidate (2021-2024)

EDF R&D - Institut de Mathématiques de Toulouse

Nicolas Bousquet, Fabrice Gamboa, Bertrand Iooss, Jean-Michel Loubes

Interpretability methods for certifying machine learning models applied to critical systems

☞ **Now:** Postdoctoral Researcher (2024-2026+)

UQÀM - IID (ULaval)

Arthur Charpentier (UQÀM), Marie-Pier Côté (ULaval)

Interpretability, fairness and causal inference of black-box models

(Some) Topics of interest:

XAI • *Uncertainty quantification* • *Sensitivity analysis* • *Industrial risks* • *Probabilistic modelling* • *Statistics* • *Statistical Learning* • *Applied cooperative game theory* • *Functional analysis*

Introduction

☞ Being able to **understand how black-box models behave under uncertainty** is crucial for **integrating AI models to critical systems**

Industrial processes, energy production, medicine, cars, planes, the stock market...

Introduction

☞ Being able to **understand how black-box models behave under uncertainty** is crucial for **integrating AI models to critical systems**

Industrial processes, energy production, medicine, cars, planes, the stock market...

☞ The decision-making process must be **justifiable** and **justified**

Empirical statements (e.g. SOTA) are notoriously **not enough** to convince **safety/regulation authorities**

Introduction

☞ Being able to **understand how black-box models behave under uncertainty** is crucial for **integrating AI models to critical systems**

Industrial processes, energy production, medicine, cars, planes, the stock market...

☞ The decision-making process must be **justifiable** and **justified**

Empirical statements (e.g. SOTA) are notoriously **not enough** to convince **safety/regulation authorities**

We must turn to theoretical arguments

☞ Being able to **understand how black-box models behave under uncertainty** is crucial for **integrating AI models to critical systems**
Industrial processes, energy production, medicine, cars, planes, the stock market...

☞ The decision-making process must be **justifiable** and **justified**
Empirical statements (e.g. SOTA) are notoriously **not enough** to convince **safety/regulation authorities**

We must turn to theoretical arguments

☞ **But**, some of the theoretical developments often rely on **unrealistic assumptions**
Usually, mutual independence of the inputs

Introduction

☞ Being able to **understand how black-box models behave under uncertainty** is crucial for **integrating AI models to critical systems**

Industrial processes, energy production, medicine, cars, planes, the stock market...

☞ The decision-making process must be **justifiable** and **justified**

Empirical statements (e.g. SOTA) are notoriously **not enough** to convince **safety/regulation authorities**

We must turn to theoretical arguments

☞ **But**, some of the theoretical developments often rely on **unrealistic assumptions**

Usually, mutual independence of the inputs

How to go beyond mutually independent inputs

Introduction

In this talk:

We are not going to talk about estimators

Introduction

In this talk:

We are not going to talk about estimators

We are going to **explore the behavior** of the **theoretical quantities** under **input dependence**

Introduction

In this talk:

We are not going to talk about estimators

We are going to **explore the behavior** of the **theoretical quantities** under **input dependence**

Quick outline:

- ☞ **From Hoeffding's decomposition to Sobol' indices**
- ☞ **Shapley effects and the "cooperative games" solution**
- ☞ **Generalized Hoeffding decomposition: disentangling interaction and dependence**

Hoeffding's decomposition

- ☞ Let $D = \{1, \ldots, d\}$ and let $\mathcal{P}_D$ denote the **power-set** of D
- ☞ Let $X = (X_1, \ldots, X_d)$ be a random vector of **mutually independent inputs**
- ☞ For every $A \in \mathcal{P}_D$, $A \neq \emptyset$, let X_A be a **subset of the inputs**
- ☞ Let G be a **given "black-box" model** such that $\mathbb{V}(G(X)) < \infty$

Hoeffding's decomposition

- ☞ Let $D = \{1, \ldots, d\}$ and let $\mathcal{P}_D$ denote the **power-set** of D
- ☞ Let $X = (X_1, \ldots, X_d)$ be a random vector of **mutually independent inputs**
- ☞ For every $A \in \mathcal{P}_D$, $A \neq \emptyset$, let X_A be a **subset of the inputs**
- ☞ Let G be a **given "black-box" model** such that $\mathbb{V}(G(X)) < \infty$

Hoeffding (1948): We can **uniquely** write

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A),$$

where $G_\emptyset$ is a constant, and the **representants** are all **pairwise orthogonal**, i.e.,

$$\forall A, B \in \mathcal{P}_D, A \neq B, \quad \mathbb{E}\left[G_A(X_A) G_B(X_B)\right] = 0$$

Hoeffding's decomposition

- Let $D = \{1, \ldots, d\}$ and let $\mathcal{P}_D$ denote the **power-set** of D
- Let $X = (X_1, \ldots, X_d)$ be a random vector of **mutually independent inputs**
- For every $A \in \mathcal{P}_D$, $A \neq \emptyset$, let X_A be a **subset of the inputs**
- Let G be a **given "black-box" model** such that $\mathbb{V}(G(X)) < \infty$

Hoeffding (1948): We can **uniquely** write

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A),$$

where $G_\emptyset$ is a constant, and the **representants** are all **pairwise orthogonal**, i.e.,

$$\forall A, B \in \mathcal{P}_D, A \neq B, \quad \mathbb{E}\left[G_A(X_A) G_B(X_B)\right] = 0$$

Moreover, we can characterize

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{E}\left[G(X) \mid X_B\right], \quad \forall A \in \mathcal{P}_D$$

Hoeffding's decomposition

☞ The terms $G_A(X_A)$ can be interpreted as **interactions induced by the model** G

This is great for **post-hoc interpretability purposes**!

Hoeffding's decomposition

☞ The terms $G_A(X_A)$ can be interpreted as **interactions induced by the model** G

This is great for **post-hoc interpretability purposes**!

For example:

$$G(X) = X_1 + X_2X_3, \quad X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right)$$

In this case, we have that

$$G_1(X_1) = X_1 \quad G_2(X_2) = 0, \quad G_3(X_3) = 0,$$
$$G_{12}(X_{12}) = 0, \quad G_{13}(X_{13}) = 0, \quad G_{23}(X_{23}) = X_2X_3,$$
$$G_{123}(X_{123}) = 0$$

Hoeffding's decomposition

☞ The terms $G_A(X_A)$ can be interpreted as **interactions induced by the model** G
This is great for **post-hoc interpretability purposes**!

For example:

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

In this case, we have that

$$G_1(X_1) = X_1 \quad G_2(X_2) = 0, \quad G_3(X_3) = 0,$$
$$G_{12}(X_{12}) = 0, \quad G_{13}(X_{13}) = 0, \quad G_{23}(X_{23}) = X_2 X_3,$$
$$G_{123}(X_{123}) = 0$$

We can retrieve the full model by only having access to the **representants**

Sobol' indices

☞ In **Global Sensitivity Analysis (GSA)** we want to **quantify the importance** of a set of inputs.

An input is important $\iff$ It **contributes** to the model's **uncertainty**

The **amount of uncertainty in the model**: $\mathbb{V}\left(G(X)\right)$

Sobol' indices

☞ In **Global Sensitivity Analysis (GSA)** we want to **quantify the importance** of a set of inputs.

An input is important $\iff$ It **contributes** to the model's **uncertainty**

The **amount of uncertainty in the model**: $\mathbb{V}(G(X))$

<u>Solution:</u> Decomposing $\mathbb{V}(G(X))$ w.r.t. the subsets of inputs!

Sobol' indices

☞ In **Global Sensitivity Analysis (GSA)** we want to **quantify the importance** of a set of inputs.

An input is important $\iff$ It **contributes** to the model's **uncertainty**

The **amount of uncertainty in the model**: $\mathbb{V}\left(G(X)\right)$

<u>Solution:</u> Decomposing $\mathbb{V}\left(G(X)\right)$ w.r.t. the subsets of inputs!

Sobol (2001) proposed to use

$$S_A = \frac{\mathbb{V}\left(G_A(X_A)\right)}{\mathbb{V}\left(G(X)\right)} = \frac{1}{\mathbb{V}\left(G(X)\right)} \times \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{V}\left(\mathbb{E}\left[G(X) \mid X_B\right]\right)$$

based on the rationale of **Hoeffding's decomposition**

$$\mathbb{V}\left(G(X)\right) = \mathbb{V}\left(\sum_{A \in \mathcal{P}_D} G_A(X_A)\right) \overset{\perp}{=} \sum_{A \in \mathcal{P}_D} \mathbb{V}\left(G_A(X_A)\right)$$

Sobol' indices

☞ In **Global Sensitivity Analysis (GSA)** we want to **quantify the importance** of a set of inputs.

An input is important $\iff$ It **contributes** to the model's **uncertainty**

The **amount of uncertainty in the model**: $\mathbb{V}\left(G(X)\right)$

Solution: Decomposing $\mathbb{V}\left(G(X)\right)$ w.r.t. the subsets of inputs!

Sobol (2001) proposed to use

$$S_A = \frac{\mathbb{V}\left(G_A(X_A)\right)}{\mathbb{V}\left(G(X)\right)} = \frac{1}{\mathbb{V}\left(G(X)\right)} \times \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{V}\left(\mathbb{E}\left[G(X) \mid X_B\right]\right)$$

based on the rationale of **Hoeffding's decomposition**

$$\mathbb{V}\left(G(X)\right) = \mathbb{V}\left(\sum_{A \in \mathcal{P}_D} G_A(X_A)\right) \overset{\perp}{=} \sum_{A \in \mathcal{P}_D} \mathbb{V}\left(G_A(X_A)\right)$$

The model's uncertainty is equal to the sum of its representant's uncertainty

Sobol' indices

The Sobol' indices are **amazing**:

☞ They are **always positive** ($S_A \geq 0$) and they **sum to 1**

$\Longrightarrow$ Interpretation **as a percentage of the model's overall uncertainty!**

Sobol' indices

The Sobol' indices are **amazing**:

☞ They are **always positive** ($S_A \geq 0$) and they **sum to 1**

$\Longrightarrow$ Interpretation **as a percentage of the model's overall uncertainty!**

But, they require the inputs to be **mutually independent**...

Sobol' indices

The Sobol' indices are **amazing**:

☞ They are **always positive** ($S_A \geq 0$) and they **sum to 1**

$\Longrightarrow$ Interpretation **as a percentage of the model's overall uncertainty!**

But, they require the inputs to be **mutually independent**...

If we **keep the same formula** : $S_A = \frac{1}{\mathbb{V}(G(X))} \times \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{V}\left(\mathbb{E}\left[G(X) \mid X_B\right]\right)$

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

Mutually independent case ($\rho = 0$)

$$S_1 = 0.5 \quad S_2 = 0, \quad S_3 = 0,$$
$$S_{12} = 0, \quad S_{13} = 0, \quad S_{23} = 0.5,$$
$$S_{123} = 0$$

Correlated case ($\rho \neq 0$)

$$S_1 = 0.5 \quad S_2 = 0, \quad S_3 = \rho^2/2,$$
$$S_{12} = \rho^2/2, \quad S_{13} = -\rho^2/2, \quad S_{23} = 0.5,$$
$$S_{123} = -\rho^2/2$$

☞ They still **sum to 1** but they can be **negative**

But their overall interpretation as (pure) interaction effects does not hold anymore...

Sobol' indices computed on dependent inputs

There are (mainly) 2 problems:

1. **Negative effects** (dependence)
2. **Unclear interpretation** (Hoeffding's decomposition)

Sobol' indices computed on dependent inputs

There are (mainly) 2 problems:

1. **Negative effects** (dependence)
2. **Unclear interpretation** (Hoeffding's decomposition)

But, there are solutions!

1. **Cooperative games' allocations** (e.g., Shapley values)
2. **Beyond Hoeffding's original decomposition**

Shapley effects

The Shapley effects (Owen 2014): **Egalitarian redistribution of the Sobol' indices.**

Shapley effects

The Shapley effects (Owen 2014): **Egalitarian redistribution of the Sobol' indices.**

Recipe:

1. Compute the Sobol' indices (*even if the inputs are dependent*)

Shapley effects

The Shapley effects (Owen 2014): **Egalitarian redistribution of the Sobol' indices.**

Recipe:

1. Compute the Sobol' indices (*even if the inputs are dependent*)
2. For every $A \in \mathcal{P}_D$ give $\frac{S_A}{|A|}$ to every input in X_A

Shapley effects

The Shapley effects (Owen 2014): **Egalitarian redistribution of the Sobol' indices.**

Recipe:

1. Compute the Sobol' indices (*even if the inputs are dependent*)
2. For every $A \in \mathcal{P}_D$ give $\frac{S_A}{|A|}$ to every input in X_A
   For example, X_1 is attributed S_1, $\frac{S_{12}}{2}$, $\frac{S_{123}}{3}$...

Shapley effects

The Shapley effects (Owen 2014): **Egalitarian redistribution of the Sobol' indices.**

Recipe:

1. Compute the Sobol' indices (*even if the inputs are dependent*)
2. For every $A \in \mathcal{P}_D$ give $\frac{S_A}{|A|}$ to every input in X_A
   For example, X_1 is attributed S_1, $\frac{S_{12}}{2}$, $\frac{S_{123}}{3}$ ...
3. And... that's it!

Shapley effects

The Shapley effects (Owen 2014): **Egalitarian redistribution of the Sobol' indices.**

Recipe:

1. Compute the Sobol' indices (*even if the inputs are dependent*)
2. For every $A \in \mathcal{P}_D$ give $\frac{S_A}{|A|}$ to every input in X_A
   For example, X_1 is attributed S_1, $\frac{S_{12}}{2}$, $\frac{S_{123}}{3}$...
3. And... that's it!

The resulting indices can be written as

$$\forall i \in D, \quad Sh_i = \sum_{A \in \mathcal{P}_D \,:\, i \in A} \frac{S_A}{|A|}$$

Shapley effects

The Shapley effects (Owen 2014): **Egalitarian redistribution of the Sobol' indices.**

Recipe:

1. Compute the Sobol' indices (*even if the inputs are dependent*)
2. For every $A \in \mathcal{P}_D$ give $\frac{S_A}{|A|}$ to every input in X_A
   For example, X_1 is attributed S_1, $\frac{S_{12}}{2}$, $\frac{S_{123}}{3}$...
3. And... that's it!

The resulting indices can be written as

$$\forall i \in D, \quad Sh_i = \sum_{A \in \mathcal{P}_D \,:\, i \in A} \frac{S_A}{|A|}$$

They are nonnegative and **sum to one**

First problem solved!

Shapley effects

The Shapley effects (Owen 2014): **Egalitarian redistribution of the Sobol' indices.**

Recipe:

1. Compute the Sobol' indices (*even if the inputs are dependent*)
2. For every $A \in \mathcal{P}_D$ give $\frac{S_A}{|A|}$ to every input in X_A
   For example, X_1 is attributed S_1, $\frac{S_{12}}{2}$, $\frac{S_{123}}{3}$ ...
3. And... that's it!

The resulting indices can be written as

$$\forall i \in D, \quad Sh_i = \sum_{A \in \mathcal{P}_D \,:\, i \in A} \frac{S_A}{|A|}$$

They are nonnegative and **sum to one**

First problem solved!

But let's take a step back...

Cooperative game theory

The **Shapley effects** are derived from the **Shapley (1951) values**

Stems from **cooperative game theory**

Cooperative game theory

The **Shapley effects** are derived from the **Shapley (1951) values**

Stems from **cooperative game theory**

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

Cooperative game theory

The **Shapley effects** are derived from the **Shapley (1951) values**

Stems from **cooperative game theory**

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

Let $D = \{1, \dots, d\}$ be a **set of players**, and $\mathcal{P}_D$ the **set of coalitions**

Let $v : \mathcal{P}_D \to \mathbb{R}$ be a **chosen value function**

☞ (D, v) formally defines a **cooperative game**

Cooperative game theory

The **Shapley effects** are derived from the **Shapley (1951) values**

Stems from **cooperative game theory**

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

Let $D = \{1, \dots, d\}$ be a **set of players**, and $\mathcal{P}_D$ the **set of coalitions**

Let $v : \mathcal{P}_D \to \mathbb{R}$ be a **chosen value function**

☞ (D, v) formally defines a **cooperative game**

Main question:

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

Cooperative game theory

The **Shapley effects** are derived from the **Shapley (1951) values**
Stems from **cooperative game theory**

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

Let $D = \{1, \ldots, d\}$ be a **set of players**, and $\mathcal{P}_D$ the **set of coalitions**
Let $v : \mathcal{P}_D \to \mathbb{R}$ be a **chosen value function**
☞ (D, v) formally defines a **cooperative game**

Main question:

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

Hint: by using an **allocation**

Cooperative Game Theory

An **allocation** associates a share of $v(D)$ to **inputs individually**

It is a mapping $\psi : D \to \mathbb{R}$

Cooperative Game Theory

An **allocation** associates a share of $v(D)$ to **inputs individually**

It is a mapping $\psi : D \to \mathbb{R}$

For **importance quantification** two criteria are required:

- **Efficiency:** $\sum_{i \in D} \psi(i) = v(D)$
- **Nonnegativity:** $\forall i \in D, \quad \psi(i) \geq 0$

We redistribute **the whole cake and nothing but the cake**

Cooperative Game Theory

An **allocation** associates a share of $v(D)$ to **inputs individually**

It is a mapping $\psi : D \to \mathbb{R}$

For **importance quantification** two criteria are required:

- **Efficiency:** $\sum_{i \in D} \psi(i) = v(D)$
- **Nonnegativity:** $\forall i \in D, \quad \psi(i) \geq 0$

We redistribute **the whole cake and nothing but the cake**

☞ In this case, ψ can be interpreted **as a percentage of variance allocated to an input**

Cooperative Game Theory

An **allocation** associates a share of $v(D)$ to **inputs individually**

It is a mapping $\psi : D \to \mathbb{R}$

For **importance quantification** two criteria are required:

- **Efficiency:** $\sum_{i \in D} \psi(i) = v(D)$
- **Nonnegativity:** $\forall i \in D, \quad \psi(i) \geq 0$

We redistribute **the whole cake and nothing but the cake**

☞ In this case, ψ can be interpreted **as a percentage of variance allocated to an input**

How can we define efficient and nonnegative allocations?

The Harsanyi set

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

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

It is a mapping $\mathcal{D}_v(A) : \mathcal{P}_D \to \mathbb{R}$

☞ They can be interpreted as the **added value** produced by each **coalition**

☞ They **always** sum-up to $v(D)$:

$$\sum_{A \in \mathcal{P}_D} \mathcal{D}_v(A) = v(D)$$

The Harsanyi set

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

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

It is a mapping $\mathcal{D}_v(A) : \mathcal{P}_D \to \mathbb{R}$

☞ They can be interpreted as the **added value** produced by each **coalition**

☞ They **always** sum-up to $v(D)$:

$$\sum_{A \in \mathcal{P}_D} \mathcal{D}_v(A) = v(D)$$

The **Harsanyi set of allocations** (Vasil'ev and Laan 2001) are **aggregations of the Harsanyi dividends**:

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

The Harsanyi set

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

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

It is a mapping $\mathcal{D}_v(A) : \mathcal{P}_D \to \mathbb{R}$

☞ They can be interpreted as the **added value** produced by each **coalition**

☞ They **always** sum-up to $v(D)$:

$$\sum_{A \in \mathcal{P}_D} \mathcal{D}_v(A) = v(D)$$

The **Harsanyi set of allocations** (Vasil'ev and Laan 2001) are **aggregations of the Harsanyi dividends**:

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

☞ **They are always efficient**

☞ **Nonnegative if v is monotonic**

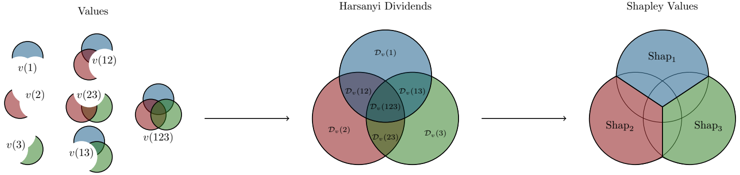
It is the case for $\mathbb{V}\left(\mathbb{E}\left[G(X) \mid X_A\right]\right)$

Egalitarian redistribution: the Shapley values

The **Shapley (1951) values** are the **egalitarian redistribution of the dividends**.

For a player $i \in D$,

$$\mathrm{Shap}_i = \sum_{A \in \mathcal{P}_D : i \in A} \frac{\mathcal{D}_v(A)}{|A|}.$$

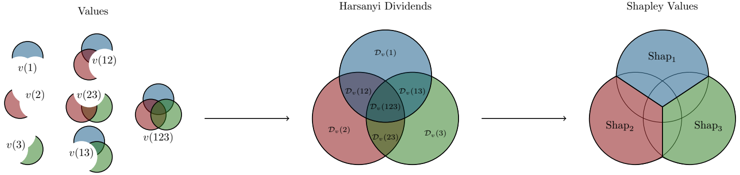


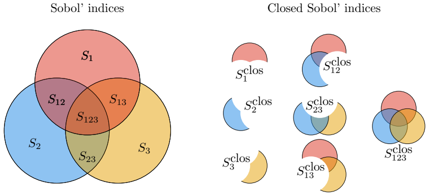
Egalitarian redistribution: the Shapley values

The **Shapley (1951) values** are the **egalitarian redistribution of the dividends**.

For a player $i \in D$,

$$\mathrm{Shap}_i = \sum_{A \in \mathcal{P}_D : i \in A} \frac{\mathcal{D}_v(A)}{|A|}.$$





Shapley effects (Owen 2014):

Shapley values with $v(A) = \mathbb{V}\left(\mathbb{E}\left[G(X) \mid X_A\right]\right) = S_A^{\mathrm{clos}}$

$$\mathrm{Sh}_i = \sum_{A \in \mathcal{P}_D : i \in A} \frac{S_A}{|A|}$$

The **Harsanyi dividends** become S_A: the **Sobol' indices**

Exogeneity detection and Shapley's joke

However, the **Shapley effects** have a **practical drawback**

Exogeneity detection and Shapley's joke

However, the **Shapley effects** have a **practical drawback**

An **exogenous input** can have a **non-zero share of importance**.

We call it **Shapley's joke** (Iooss and Prieur 2019)

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

$$\text{Sh}_1 = 0.5 - \rho^2/4, \quad \text{Sh}_2 = 0.5, \quad \underline{\text{Sh}_3 = \rho^2/4}$$

Exogeneity detection and Shapley's joke

However, the **Shapley effects** have a **practical drawback**

An **exogenous input** can have a **non-zero share of importance**.

We call it **Shapley's joke** (Iooss and Prieur 2019)

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

$$\text{Sh}_1 = 0.5 - \rho^2/4, \quad \text{Sh}_2 = 0.5, \quad \underline{\text{Sh}_3 = \rho^2/4}$$

To solve this issue, we proposed to use **a proportional redistribution of the dividends** (Ortmann 2000)

It led to the definition of the **proportional marginal effects (PME)** (Herin et al. 2023)

Exogeneity detection and Shapley's joke

However, the **Shapley effects** have a **practical drawback**

An **exogenous input** can have a **non-zero share of importance**.

We call it **Shapley's joke** (Iooss and Prieur 2019)

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

$$\mathrm{Sh}_1 = 0.5 - \rho^2/4, \quad \mathrm{Sh}_2 = 0.5, \quad \underline{\mathrm{Sh}_3 = \rho^2/4}$$

To solve this issue, we proposed to use **a proportional redistribution of the dividends** (Ortmann 2000)

It led to the definition of the **proportional marginal effects (PME)** (Herin et al. 2023)

> **Proposition** *(Exogeneity detection)*. Under mild assumptions on the probabilistic structure of X,
>
> $$PME_i = 0 \iff X_i \text{ is exogenous.}$$

In practice, they tend to **"discriminate more" than the Shapley values** when the inputs are **highly correlated**

Estimation

Estimating the PME/Shapley effects $\iff$ Estimating $v(A)$ for every $A \in \mathcal{P}_D$.

Two settings:

- **You can sample your model at will (Monte Carlo)**: Requires a number proportional to $d!(d-1)$ model evaluations (Song, Nelson, and Staum 2016).
  The estimation cost can be substantially lowered by giving-up precision.
- **You only have access to an i.i.d. sample (Given-data)**: **The nearest-neighbor procedure** requires 2^d estimates (Broto, Bachoc, and Depecker 2020a).

These methods are time-consuming and scale exponentially with the number of inputs, but the **estimates can be recycled to compute both indices at once.**
But there are many strategies to drive the computational burden down at the expense of precision

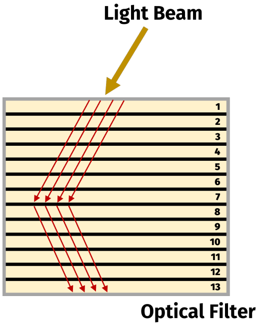
Optical filter transmittance - Feature selection

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 layer.

These errors are (highly) correlated due to the manufacturing process.

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



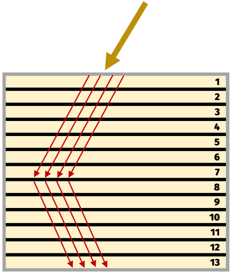
Optical filter transmittance - Feature selection

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 layer.

These errors are (highly) correlated due to the manufacturing process.

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



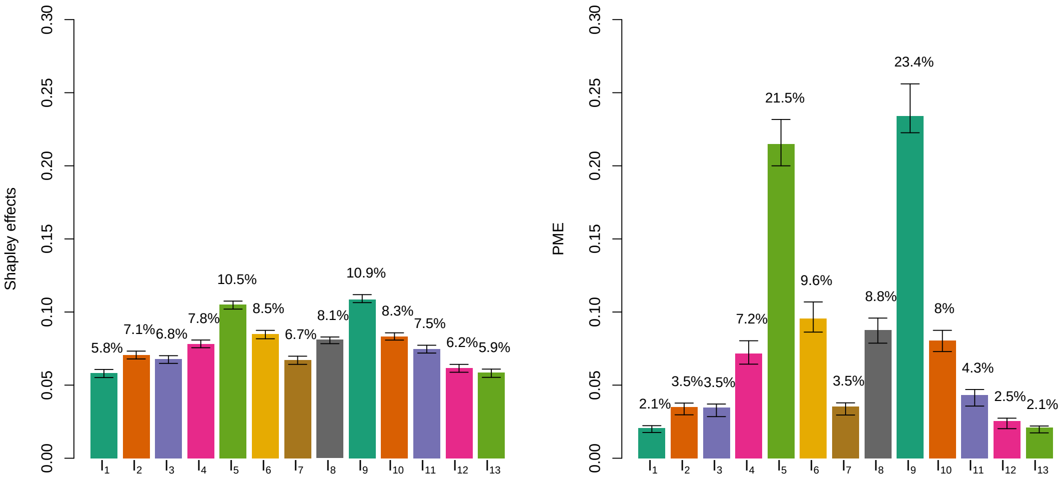
☞ **We only have access to an i.i.d. input-output sample** ($n = 1000$).

The indices are computed using a **nearest-neighbors approach** (Broto, Bachoc, and Depecker 2020b).

Parallelized implementation using the R package `sensitivity` ($\sim$ 4min runtine, 8 cores).

Arbitrarily chosen number of neighbors: 6.

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.

Partial conclusion

☞ **Sobol' indices** are an amazing tool to asses **the importance of the inputs of a black-box model**

Partial conclusion

☞ **Sobol' indices** are an amazing tool to asses **the importance of the inputs of a black-box model**

☞ Whenever the inputs are dependent, there are **two issues**:

1. The **Sobol' indices can be negative**
2. Their **intrinsic interpretation does not hold**

Partial conclusion

☞ **Sobol' indices** are an amazing tool to asses **the importance of the inputs of a black-box model**

☞ Whenever the inputs are dependent, there are **two issues**:

1. The **Sobol' indices can be negative**
2. Their **intrinsic interpretation does not hold**

☞ A **solution to the first issue** is to use **allocations** from cooperative game theory
e.g., the **Shapley effects** based on the egaliatarian aggregation a.k.a the **Shapley values**

Partial conclusion

☞ **Sobol' indices** are an amazing tool to asses **the importance of the inputs of a black-box model**

☞ Whenever the inputs are dependent, there are **two issues**:

1. The **Sobol' indices can be negative**
2. Their **intrinsic interpretation does not hold**

☞ A **solution to the first issue** is to use **allocations** from cooperative game theory
e.g., the **Shapley effects** based on the egaliatarian aggregation a.k.a the **Shapley values**

☞ A large variety of allocation can be understood as **aggregation of Harsanyi dividends**
For GSA, they are **aggregations of** the **Sobol' indices**

Partial conclusion

☞ **Sobol' indices** are an amazing tool to asses **the importance of the inputs of a black-box model**

☞ Whenever the inputs are dependent, there are **two issues**:

1. The **Sobol' indices can be negative**
2. Their **intrinsic interpretation does not hold**

☞ A **solution to the first issue** is to use **allocations** from cooperative game theory
e.g., the **Shapley effects** based on the egaliatarian aggregation a.k.a the **Shapley values**

☞ A large variety of allocation can be understood as **aggregation of Harsanyi dividends**
For GSA, they are **aggregations of** the **Sobol' indices**

☞ To detect **exogenous inputs** we can use the **proportional marginal effects**

Partial conclusion

☞ **Sobol' indices** are an amazing tool to asses **the importance of the inputs of a black-box model**

☞ Whenever the inputs are dependent, there are **two issues**:

1. The **Sobol' indices can be negative**
2. Their **intrinsic interpretation does not hold**

☞ A **solution to the first issue** is to use **allocations** from cooperative game theory
e.g., the **Shapley effects** based on the egaliatarian aggregation a.k.a the **Shapley values**

☞ A large variety of allocation can be understood as **aggregation of Harsanyi dividends**
For GSA, they are **aggregations of** the **Sobol' indices**

☞ To detect **exogenous inputs** we can use the **proportional marginal effects**

But what about the second issue?

Generalizing Hoeffding's decomposition to dependent inputs

Hoeffding's decomposition is the **key to the meaning** of the Sobol' indices...

Generalizing Hoeffding's decomposition to dependent inputs

Hoeffding's decomposition is the **key to the meaning** of the Sobol' indices...
...naturally, we would like to **generalize it to dependent inputs!**

Generalizing Hoeffding's decomposition to dependent inputs

Hoeffding's decomposition is the **key to the meaning** of the Sobol' indices...
...naturally, we would like to **generalize it to dependent inputs!**

We are the first to attempt this generalization:

e.g., Rabitz and Aliş (1999), Peccati (2004), Hooker (2007), Kuo et al. (2009), Hart and Gremaud (2018), and Chastaing, Gamboa, and Prieur (2012)

Generalizing Hoeffding's decomposition to dependent inputs

Hoeffding's decomposition is the **key to the meaning** of the Sobol' indices...
...naturally, we would like to **generalize it to dependent inputs!**

We are the first to attempt this generalization:

e.g., Rabitz and Aliş (1999), Peccati (2004), Hooker (2007), Kuo et al. (2009), Hart and Gremaud (2018), and Chastaing, Gamboa, and Prieur (2012)

But we believe we found an interesting approach to tackle it!

Generalizing Hoeffding's decomposition to dependent inputs

In general, **mutual independence** is actually **pretty restrictive**

It's much more restrictive than **pairwise independence**

Generalizing Hoeffding's decomposition to dependent inputs

In general, **mutual independence** is actually **pretty restrictive**

It's much more restrictive than **pairwise independence**

First challenge: define situations of **"non-mutual independence"**

In general, **mutual independence** is actually **pretty restrictive**

It's much more restrictive than **pairwise independence**

First challenge: define situations of **"non-mutual independence"**

We do this by imposing two assumptions:

- Non-perfect functional dependence between the inputs
- Non-degenerate stochastic dependence between the inputs

Generalizing Hoeffding's decomposition to dependent inputs

In general, **mutual independence** is actually **pretty restrictive**

It's much more restrictive than **pairwise independence**

First challenge: define situations of **"non-mutual independence"**

We do this by imposing two assumptions:

- Non-perfect functional dependence between the inputs
- Non-degenerate stochastic dependence between the inputs

Under these two assumptions **we can actually decompose the space of outputs**

Consequences

Under these two assumptions, any real-valued **random output** $G(X)$ with finite second moment can be **uniquely** decomposed:

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A),$$

where $G_A(X_A)$ are **hierarchically orthogonal**

As opposed to **pairwise orthogonal** with Hoeffding's classical result

Consequences

Under these two assumptions, any real-valued **random output** $G(X)$ with finite second moment can be **uniquely** decomposed:

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A),$$

where $G_A(X_A)$ are **hierarchically orthogonal**

As opposed to **pairwise orthogonal** with Hoeffding's classical result

They can be characterized using **oblique projections** $\mathbb{M}_A[G(X)]$ which **change w.r.t. the dependence structure**:

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{M}_B \left[G(X) \right], \quad \forall A \in \mathcal{P}_D$$

Consequences

Under these two assumptions, any real-valued **random output** $G(X)$ with finite second moment can be **uniquely** decomposed:

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A),$$

where $G_A(X_A)$ are **hierarchically orthogonal**

As opposed to **pairwise orthogonal** with Hoeffding's classical result

They can be characterized using **oblique projections** $\mathbb{M}_A[G(X)]$ which **change w.r.t. the dependence structure**:

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{M}_B \left[G(X)\right], \quad \forall A \in \mathcal{P}_D$$

And **when the inputs are mutually independent**, $\mathbb{M}_A\left[G(X)\right] = \mathbb{E}\left[G(X) \mid X_A\right]$, and

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{E}\left[G(X) \mid X_B\right], \quad \forall A \in \mathcal{P}_D$$

Consequences

Under these two assumptions, any real-valued **random output** $G(X)$ with finite second moment can be **uniquely** decomposed:

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A),$$

where $G_A(X_A)$ are **hierarchically orthogonal**

As opposed to **pairwise orthogonal** with Hoeffding's classical result

They can be characterized using **oblique projections** $\mathbb{M}_A[G(X)]$ which **change w.r.t. the dependence structure**:

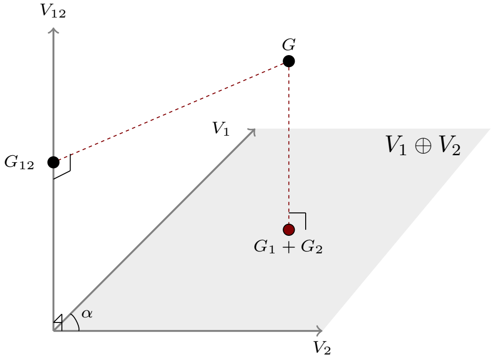
$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{M}_B\left[G(X)\right], \quad \forall A \in \mathcal{P}_D$$

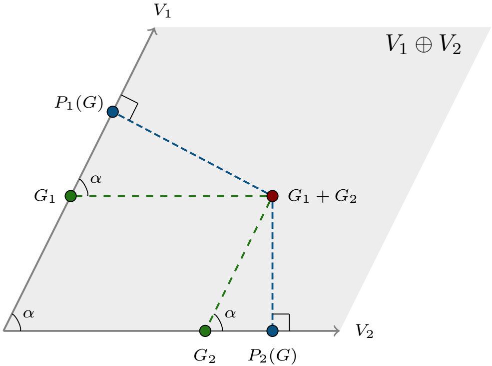
And **when the inputs are mutually independent**, $\mathbb{M}_A\left[G(X)\right] = \mathbb{E}\left[G(X) \mid X_A\right]$, and

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{E}\left[G(X) \mid X_B\right], \quad \forall A \in \mathcal{P}_D$$

☞ **We recover Hoeffding's classical decomposition**

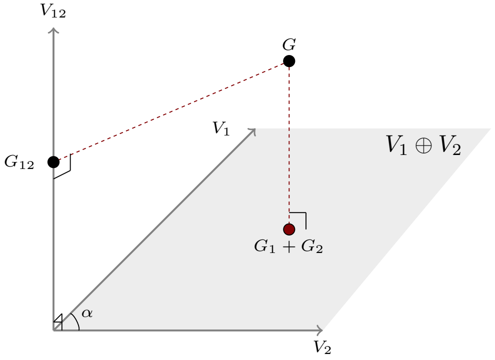
Illustration*: bivariate function

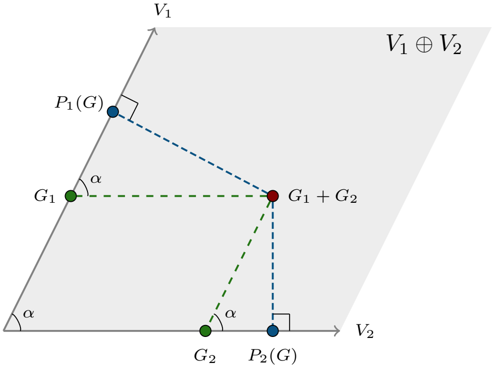




☞ The **oblique projections** usually differ from the **orthogonal projections** (i.e., conditional expectation)

Illustration*: bivariate function





☞ The **oblique projections** usually differ from the **orthogonal projections** (i.e., conditional expectation)

Variance decomposition

Let's talk about variance decomposition.

Variance decomposition

Let's talk about variance decomposition.

We propose **two complementary approaches** for decomposing $\mathbb{V}(G(X))$ based on this generalized decomposition, **inspired from the existing literature**

Variance decomposition

Let's talk about variance decomposition.

We propose **two complementary approaches** for decomposing $\mathbb{V}(G(X))$ based on this generalized decomposition, **inspired from the existing literature**

Organic variance decomposition: separate **pure interaction effects** to **dependence effects**. The dependence structure of X is **unwanted**, and one wishes to study its effects.

Let's talk about variance decomposition.

We propose **two complementary approaches** for decomposing $\mathbb{V}(G(X))$ based on this generalized decomposition, **inspired from the existing literature**

Organic variance decomposition: separate **pure interaction effects** to **dependence effects**. The dependence structure of X is **unwanted**, and one wishes to study its effects.

Orthocanonical variance decomposition: the dependence structure of X is **inherent in the uncertainty modeling** of the studied phenomenon. It amounts to quantify **structural** and **correlative** effects.

Organic variance decomposition: Pure interaction

The notion of pure interaction is intrinsically linked with the notion of mutual independence.

Let $\widetilde{X} = (\widetilde{X}_1, \ldots, \widetilde{X}_d)^\top$ be the random vector such that

$$\widetilde{X}_i \overset{d}{=} X_i, \quad \text{and } \widetilde{X} \text{ is mutually independent.}$$

Organic variance decomposition: Pure interaction

The notion of pure interaction is intrinsically linked with the notion of mutual independence.

Let $\widetilde{X} = (\widetilde{X}_1, \dots, \widetilde{X}_d)^\top$ be the random vector such that

$$\widetilde{X}_i \overset{d}{=} X_i, \quad \text{and } \widetilde{X} \text{ is mutually independent.}$$

Definition *(Pure interaction)*. For every $A \in \mathcal{P}_D$, define the **pure interaction of X_A on $G(X)$ as**

$$S_A = \frac{\mathbb{V}\left(P_A(G(\widetilde{X}))\right)}{\mathbb{V}\left(G(\widetilde{X})\right)} \times \mathbb{V}\left(G(X)\right).$$

These indices are the **Sobol' indices** computed on the mutually independent version of X.

Organic variance decomposition: Pure interaction

The notion of pure interaction is intrinsically linked with the notion of mutual independence.

Let $\widetilde{X} = (\widetilde{X}_1, \dots, \widetilde{X}_d)^\top$ be the random vector such that

$$\widetilde{X}_i \overset{d}{=} X_i, \qquad \text{and } \widetilde{X} \text{ is mutually independent.}$$

Definition *(Pure interaction)*. For every $A \in \mathcal{P}_D$, define the **pure interaction of X_A on $G(X)$ as**

$$S_A = \frac{\mathbb{V}\left(P_A(G(\widetilde{X}))\right)}{\mathbb{V}\left(G(\widetilde{X})\right)} \times \mathbb{V}\left(G(X)\right).$$

These indices are the **Sobol' indices** computed on the mutually independent version of X.

This approach **strongly resembles the "independent Sobol' indices"** proposed by Mara, Tarantola, and Annoni (2015).

(see, also, Lebrun and Dutfoy (2009a, 2009b))

Organic variance decomposition: Dependence effects

To define **dependence effects**, we measure **the distance between the orthogonal and oblique projections**

Organic variance decomposition: Dependence effects

To define **dependence effects**, we measure **the distance between the orthogonal and oblique projections**

Definition *(Dependence effects)*.

For every $A \in \mathcal{P}_D$, define the **dependence effects of X_A on $G(X)$ as**

$$S_A^D = \mathbb{E}\left[\left(Q_A(G(X)) - P_A(G(X))\right)^2\right].$$

Proposition . Under the two assumptions,

$$S_A^D = 0, \forall A \in \mathcal{P}_D, \quad \iff \quad X \text{ is mutually independent.}$$

Organic variance decomposition: Dependence effects

To define **dependence effects**, we measure **the distance between the orthogonal and oblique projections**

Definition *(Dependence effects)*.

For every $A \in \mathcal{P}_D$, define the **dependence effects of X_A on $G(X)$ as**

$$S_A^D = \mathbb{E}\left[\left(Q_A(G(X)) - P_A(G(X))\right)^2\right].$$

Proposition . Under the two assumptions,

$$S_A^D = 0, \forall A \in \mathcal{P}_D, \quad \iff \quad X \text{ is mutually independent.}$$

What do they sum up to ?...

Organic variance decomposition: Dependence effects

To define **dependence effects**, we measure **the distance between the orthogonal and oblique projections**

Definition *(Dependence effects)*.

For every $A \in \mathcal{P}_D$, define the **dependence effects of X_A on $G(X)$ as**

$$S_A^D = \mathbb{E}\left[\left(Q_A(G(X)) - P_A(G(X))\right)^2\right].$$

Proposition . Under the two assumptions,

$$S_A^D = 0, \forall A \in \mathcal{P}_D, \quad \iff \quad X \text{ is mutually independent.}$$

What do they sum up to ?...

Probably the *overall effect of the dependence on the model's uncertainty*!

Canonical variance decomposition

The structural effects represent the variance of each of the $G_A(X_A)$. It amounts to perform a **covariance decomposition** (Hart and Gremaud 2018; Da Veiga et al. 2021).

Definition *(Structural effects)*.

For every $A \in \mathcal{P}_D$, define the **structural effects of X_A on $G(X)$ as**

$$S_A^U = \mathbb{V}\left(G_A(X_A)\right).$$

The **correlative effects** represent the part of variance that is due to the correlation between the representants $G_A(X_A)$

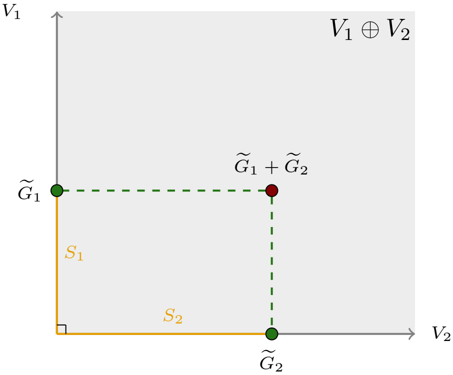
Definition *(Correlative effects)*.

For every $A \in \mathcal{P}_D$, define the **correlative effects of X_A on $G(X)$ as**

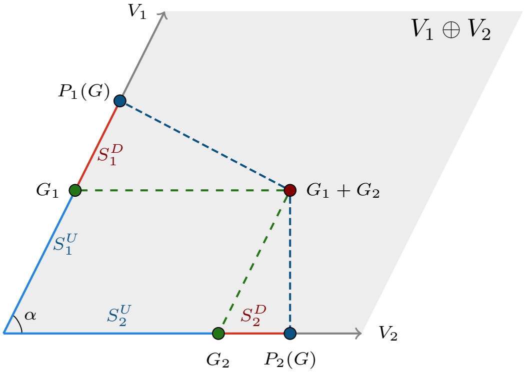
$$S_A^C = \mathrm{Cov}\left(G_A(X_A), \sum_{B \in \mathcal{P}_D : B \neq A} G_B(X_B)\right).$$

Variance decomposition: Intuition

Pure interaction effects

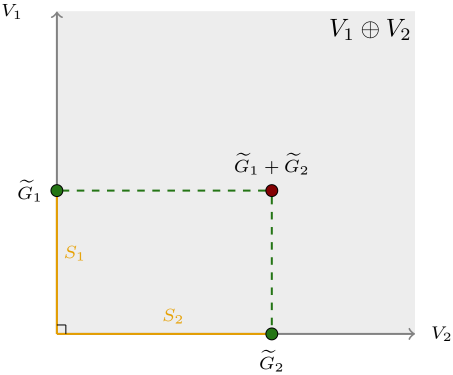


Structural and dependence effects

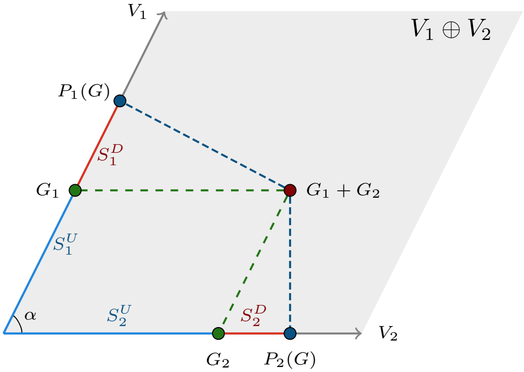


Variance decomposition: Intuition

Pure interaction effects



Structural and dependence effects



☞ **Maybe a good candidate to solve our second issue!**

Some perspectives

☞ **Main challenge: Estimating the oblique projections from observations**

They sure look a lot like conditional expectations...

Some perspectives

☞ **Main challenge: Estimating the oblique projections from observations**

They sure look a lot like conditional expectations...

☞ **Application to influence quantification**

Global and local sensitivity analysis, theoretically-grounded XAI, fairness assessment...

Some perspectives

☞ **Main challenge: Estimating the oblique projections from observations**

They sure look a lot like conditional expectations...

☞ **Application to influence quantification**

Global and local sensitivity analysis, theoretically-grounded XAI, fairness assessment...

☞ **Beyond hierarchical orthogonality**

e.g., Köhler, Rügamer, and Schmid (2024) with "stacked orthogonality" conditions

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

———. 2020b. "Variance Reduction for Estimation of Shapley Effects and Adaptation to Unknown Input Distribution." *SIAM/ASA Journal on Uncertainty Quantification* 8 (2): 693–716. ISSN: 2166-2525. https://doi.org/10.1137/18M1234631. https://epubs.siam.org/doi/10.1137/18M1234631.

Chastaing, G., F. Gamboa, and C. Prieur. 2012. "Generalized Hoeffding-Sobol decomposition for dependent variables - application to sensitivity analysis." Publisher: Institute of Mathematical Statistics and Bernoulli Society, *Electronic Journal of Statistics* 6, no. none (January): 2420–2448. ISSN: 1935-7524, 1935-7524. https://doi.org/10.1214/12-EJS749. https://projecteuclid.org/journals/electronic-journal-of-statistics/volume-6/issue-none/Generalized-Hoeffding-Sobol-decomposition-for-dependent-variables---application/10.1214/12-EJS749.full.

Da Veiga, S., F. Gamboa, B. Iooss, and C. Prieur. 2021. *Basics and Trends in Sensitivity Analysis. Theory and Practice in R.* SIAM. Computational Science / Engineering.

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.

Hart, J., and P. A. Gremaud. 2018. "An approximation theoretic perspective of Sobol' indices with dependent variables" [in English]. Publisher: Begel House Inc. *International Journal for Uncertainty Quantification* 8 (6). ISSN: 2152-5080, 2152-5099. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2018026498. https://www.dl.begellhouse.com/journals/52034eb04b657aea,23dc16a4645b89c9,61d464a51b6bf191.html.

Herin, M., M. I., V. Chabridon, and B. Iooss. 2023. "Proportional marginal effects for global sensitivity analysis." Preprint. https://hal.science/hal-03825935.

Hoeffding, W. 1948. "A Class of Statistics with Asymptotically Normal Distribution." *The Annals of Mathematical Statistics* 19 (3): 293–325. ISSN: 0003-4851, 2168-8990. https://doi.org/10.1214/aoms/1177730196. https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-19/issue-3/A-Class-of-Statistics-with-Asymptotically-Normal-Distribution/10.1214/aoms/1177730196.full.

Hooker, G. 2007. "Generalized Functional ANOVA Diagnostics for High-Dimensional Functions of Dependent Variables" [in en]. *Journal of Computational and Graphical Statistics* 16 (3): 709–732. http://www.jstor.org/stable/27594267.

Iooss, B., and C. Prieur. 2019. "Shapley effects for Sensitivity Analysis with correlated inputs : Comparisons with Sobol' Indices, Numerical Estimation and Applications." *International Journal for Uncertainty Quantification* 9 (5): 493–514.

Kuo, F. Y., I. H. Sloan, G. W. Wasilkowski, and H. Woźniakowski. 2009. "On decompositions of multivariate functions" [in en]. *Mathematics of Computation* 79, no. 270 (November): 953–966. ISSN: 0025-5718. https://doi.org/10.1090/S0025-5718-09-02319-9. http://www.ams.org/journal-getitem?pii=S0025-5718-09-02319-9.

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.

Lebrun, R., and A. Dutfoy. 2009a. "A generalization of the Nataf transformation to distributions with elliptical copula." *Probabilistic Engineering Mechanics* 24 (2): 172–178. ISSN: 0266-8920. https://doi.org/10.1016/j.probengmech.2008.05.001. https://www.sciencedirect.com/science/article/pii/S0266892008000507.

———. 2009b. "Do Rosenblatt and Nataf isoprobabilistic transformations really differ?" *Probabilistic Engineering Mechanics* 24 (4): 577–584. ISSN: 0266-8920. https://doi.org/10.1016/j.probengmech.2009.04.006. https://www.sciencedirect.com/science/article/pii/S0266892009000307.

Mara, Thierry A., Stefano Tarantola, and Paola Annoni. 2015. "Non-parametric methods for global sensitivity analysis of model output with dependent inputs" [in en]. *Environmental Modelling & Software* 72 (October): 173–183. ISSN: 13648152. https://doi.org/10.1016/j.envsoft.2015.07.010. https://linkinghub.elsevier.com/retrieve/pii/S1364815215300153.

Ortmann, K. M. 2000. "The proportional value for positive cooperative games." *Mathematical Methods of Operations Research (ZOR)* 51 (2): 235–248.

Owen, A. B. 2014. "Sobol' Indices and Shapley Value." *SIAM/ASA Journal on Uncertainty Quantification* 2 (1): 245–251. ISSN: 2166-2525. https://doi.org/10.1137/130936233. http://epubs.siam.org/doi/10.1137/130936233.

Peccati, Giovanni. 2004. "Hoeffding-ANOVA decompositions for symmetric statistics of exchangeable observations." Publisher: Institute of Mathematical Statistics, *The Annals of Probability* 32 (3): 1796–1829. ISSN: 0091-1798, 2168-894X. https://doi.org/10.1214/009117904000000405. https://projecteuclid.org/journals/annals-of-probability/volume-32/issue-3/Hoeffding-ANOVA-decompositions-for-symmetric-statistics-of-exchangeable-observations/10.1214/009117904000000405.full.

Rabitz, H., and O. Aliş. 1999. "General foundations of high-dimensional model representations" [in en]. *Journal of Mathematical Chemistry* 25 (2): 197–233. ISSN: 1572-8897. https://doi.org/10.1023/A:1019188517934. https://doi.org/10.1023/A:1019188517934.

Shapley, L. S. 1951. *Notes on the n-Person Game – II: The Value of an n-Person Game.* Research Memorandum ATI 210720. Santa Monica, California: RAND Corporation.

Sobol, I.M. 2001. "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates." *Mathematics and Computers in Simulation* 55 (1): 271–280. ISSN: 03784754.

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.

Vasil'ev, V., and G. van der Laan. 2001. *The Harsanyi Set for Cooperative TU-Games.* Working Paper 01-004/1. Tinbergen Institute Discussion Paper. https://www.econstor.eu/handle/10419/85790.

References v

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.

Thank you for your attention!

Any questions?