

INTERPRÉTABILITÉ DES MODÈLES BOÎTES-NOIRES

AVEC VARIABLES DÉPENDANTES

¹Université du Québec à Montréal (UQÀM)

²Institut Intelligence et Données (IID) - Université Laval

³EDF R&D - PRISME Department

⁴Institut de Mathématiques de Toulouse

⁵SINCLAIR AI Lab

Séminaire STATQAM

UQÀM - Montréal, QC, Canada

3 Octobre 2024



Marouane IL IDRISSI

ilidrissi.m@gmail.com - marouaneilidrissi.com

☞ **Avant : Doctorat (2021-2024)**

EDF R&D - Institut de Mathématiques de Toulouse

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

Développement de méthodes d'interprétabilité pour la certification des IA

☞ **Maintenant : Chercheur Postdoctoral (2024-202X)**

UQÀM - IID (ULaval)

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

Interprétabilité, équité et causalité pour l'apprentissage automatique en actuariat

Thématiques de recherche :

IA Explicable • Quantification des incertitudes • Analyse de sensibilité • Modélisation statistique
• Probabilités • Apprentissage automatique • Théorie des jeux appliqués • Analyse fonctionnelle

➡ Être en mesure de **comprendre comment les boîtes-noires se comportent en situation d'incertitudes** est primordial pour **l'intégration des modèles d'IA dans des systèmes critiques**

Processus industriels, production d'électricité, médecine, transports, économie...

➡ Être en mesure de **comprendre comment les boîtes-noires se comportent en situation d'incertitudes** est primordial pour **l'intégration des modèles d'IA dans des systèmes critiques**
Processus industriels, production d'électricité, médecine, transports, économie...

➡ La prise de décision doit être **justifiable** et **justifiée**

Arguments empiriques (e.g. SOTA) ont **peu de poids** pour convaincre les **autorités de sûreté/régulation**

Contexte

➡ Être en mesure de **comprendre comment les boîtes-noires se comportent en situation d'incertitudes** est primordial pour **l'intégration des modèles d'IA dans des systèmes critiques**
Processus industriels, production d'électricité, médecine, transports, économie...

➡ La prise de décision doit être **justifiable** et **justifiée**

Arguments empiriques (e.g. SOTA) ont **peu de poids** pour convaincre les **autorités de sûreté/régulation**

➡ Malgré la **plus-value importante** que l'**apprentissage automatique** peut offrir :

Modélisation des **risques de crédits**, changement climatique et **primes d'assurance**, repérer les **fraudes**

➡ Être en mesure de **comprendre comment les boîtes-noires se comportent en situation d'incertitudes** est primordial pour **l'intégration des modèles d'IA dans des systèmes critiques**
Processus industriels, production d'électricité, médecine, transports, économie...

➡ La prise de décision doit être **justifiable** et **justifiée**

Arguments empiriques (e.g. SOTA) ont **peu de poids** pour convaincre les **autorités de sûreté/régulation**

➡ Malgré la **plus-value importante** que l'**apprentissage automatique** peut offrir :
Modélisation des **risques de crédits**, changement climatique et **primes d'assurance**, repérer les **fraudes**

➡ Plusieurs aspects doivent être pris en compte dans la décision :

Équité (variables protégées), **transparence de l'information**, **stabilité des marchés**

➡ Être en mesure de **comprendre comment les boîtes-noires se comportent en situation d'incertitudes** est primordial pour **l'intégration des modèles d'IA dans des systèmes critiques**
Processus industriels, production d'électricité, médecine, transports, économie...

➡ La prise de décision doit être **justifiable et justifiée**

Arguments empiriques (e.g. SOTA) ont **peu de poids** pour convaincre les **autorités de sûreté/régulation**

➡ Malgré la **plus-value importante** que l'**apprentissage automatique** peut offrir :
Modélisation des **risques de crédits**, changement climatique et **primes d'assurance**, repérer les **fraudes**

➡ Plusieurs aspects doivent être pris en compte dans la décision :

Équité (variables protégées), **transparence de l'information**, **stabilité des marchés**

Comment justifier d'une décision soutenue par un modèle boîte-noire ?

Par exemple, un modèle issu de l'apprentissage automatique

Deux domaines de recherche

☞ L'intelligence artificielle explicable (XAI) pour comprendre l'IA

Issue de la littérature ML/DL, avec une forte vision appliquée

Beaucoup de méthodes (LIME, SHAP...), mais justifiées empiriquement

Deux domaines de recherche

☞ L'**intelligence artificielle explicable (XAI)** pour comprendre l'IA

Issue de la **littérature ML/DL**, avec une forte **vision appliquée**

Beaucoup de méthodes (LIME, SHAP...), mais **justifiées empiriquement**

☞ L'**analyse de sensibilité (SA)** pour interpréter les modèles de simulation numériques

Issue de la **quantification des incertitudes (UQ)**, avec un fort **formalisme mathématique**

Compréhension théorique des méthodes, mais **sous des hypothèses peu réalistes**

Deux domaines de recherche

☞ L'**intelligence artificielle explicable (XAI)** pour comprendre l'IA

Issue de la **littérature ML/DL**, avec une forte **vision appliquée**

Beaucoup de méthodes (LIME, SHAP...), mais **justifiées empiriquement**

☞ L'**analyse de sensibilité (SA)** pour interpréter les modèles de simulation numériques

Issue de la **quantification des incertitudes (UQ)**, avec un fort **formalisme mathématique**

Compréhension théorique des méthodes, mais **sous des hypothèses peu réalistes**

Un objectif commun :

Comprendre les modèles de prévision pour valider leur utilisation

Notre ambition :

Étendre le formalisme de la SA, pour répondre aux enjeux de l'apprentissage automatique

Notre ambition :

Étendre le formalisme de la SA, pour **répondre aux enjeux de l'apprentissage automatique**

Concrètement, ça veut dire quoi ?

☞ Aller **au-delà des hypothèses** classiques en SA
e.g., Indépendance mutuelle des variables explicatives...

☞ **Compréhension théorique** des méthodes en XAI
Définir les quantités théoriques d'intérêt avant de proposer des estimateurs

☞ Proposer une **méthodologie générale** unificatrice
Tirer profit du meilleur des deux mondes

Position et velléité

Notre ambition :

Étendre le formalisme de la SA, pour **répondre aux enjeux de l'apprentissage automatique**

Concrètement, ça veut dire quoi ?

☞ Aller **au-delà des hypothèses** classiques en SA
e.g., Indépendance mutuelle des variables explicatives...

☞ **Compréhension théorique** des méthodes en XAI
Définir les quantités théoriques d'intérêt avant de proposer des estimateurs

☞ Proposer une **méthodologie générale** unificatrice
Tirer profit du meilleur des deux mondes

Pourquoi ?

S'inspirer de l'**avènement des modèles numériques** pour **simuler les phénomènes physiques**
Modèles **boîtes-noires** qui **servent déjà pour les domaines critiques** (e.g., *dossiers de sûreté*)

Quelques clarification

Dans la suite :

👉 **Boîte-noire** $\neq$ **modèle complexe**

Prise en compte d'une **large gamme de modèles** (approches agnostique au modèle)

Quelques clarifications

Dans la suite :

☞ **Boîte-noire** $\neq$ **modèle complexe**

Prise en compte d'une **large gamme de modèles** (approches agnostique au modèle)

☞ **Position *post-hoc***

Le modèle est **déjà entraîné/établi**, et on peut l'évaluer

Quelques clarifications

Dans la suite :

☞ **Boîte-noire** $\neq$ **modèle complexe**

Prise en compte d'une **large gamme de modèles** (approches agnostique au modèle)

☞ **Position *post-hoc***

Le modèle est **déjà entraîné/établi**, et on peut l'évaluer

☞ **Interprétation vs. explication**

Les méthodes servent **à interpréter**, les **explications peuvent en découler**

Quelques clarification

Dans la suite :

☞ **Boîte-noire $\neq$ modèle complexe**

Prise en compte d'une **large gamme de modèles** (approches agnostique au modèle)

☞ **Position *post-hoc***

Le modèle est **déjà entraîné/établi**, et on peut l'évaluer

☞ **Interprétation vs. explication**

Les méthodes servent **à interpréter**, les **explications peuvent en découler**

☞ **Entrées du modèle = variables explicatives**

Jargon SA (mais je vais essayer de faire attention)

Comment **mesurer l'influence** des variables sur une **quantité d'intérêt** d'un modèle boîte-noire ?

Comment **mesurer l'influence** des variables sur une **quantité d'intérêt** d'un modèle boîte-noire ?

☞ **Quantité d'intérêt (Qoi) :**

Indicateur de performance, variance, prévision ponctuelle...

Comment **mesurer l'influence** des variables sur une **quantité d'intérêt** d'un modèle boîte-noire ?

👉 **Quantité d'intérêt (Qoi) :**

Indicateur de performance, variance, prévision ponctuelle...

Pourquoi ?

- 👉 Vérifier et valider la **cohérence du modèle avec l'attendu pratique**
- 👉 Sélectionner les variables **avec des arguments tangibles**
- 👉 S'assurer de la **non-utilisation de variables protégées/sensibles**
- 👉 Mieux comprendre les **rouages internes de la boîte-noire**
- 👉 Prioriser l'acquisition de **certaines données pertinentes**
- 👉 Aide au **design des IAs** (bugs, dégénérescence...)

Quantification de l'influence

Comment **mesurer l'influence** des variables sur une **quantité d'intérêt** d'un modèle boîte-noire ?

🔊 **Quantité d'intérêt (Qoi) :**

Indicateur de performance, variance, prévision ponctuelle...

Pourquoi ?

- 🔊 Vérifier et valider la **cohérence du modèle avec l'attendu pratique**
- 🔊 Sélectionner les variables **avec des arguments tangibles**
- 🔊 S'assurer de la **non-utilisation de variables protégées/sensibles**
- 🔊 Mieux comprendre les **rouages internes de la boîte-noire**
- 🔊 Prioriser l'acquisition de **certaines données pertinentes**
- 🔊 Aide au **design des IAs** (bugs, dégénérescence...)

Comment ?

Décomposer la Qoi de sorte à en attribuer une part à chaque ensemble de variables

Notations

☞ $X = (X_1, \dots, X_d)$ dénotent **les variables (aléatoires) d'un modèle boîte-noire**

Notations

☞ $X = (X_1, \dots, X_d)$ dénotent **les variables (aléatoires) d'un modèle boîte-noire**

☞ On pose $D = \{1, \dots, d\}$ et $\mathcal{P}_D$ est **l'ensemble des sous-ensembles (power-set)** de D

Pour $d = 2$, on a $D = \{1, 2\}$, et $\mathcal{P}_D = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$

Notations

- ☞ $X = (X_1, \dots, X_d)$ dénotent **les variables (aléatoires) d'un modèle boîte-noire**
- ☞ On pose $D = \{1, \dots, d\}$ et $\mathcal{P}_D$ est **l'ensemble des sous-ensembles (power-set)** de D
Pour $d = 2$, on a $D = \{1, 2\}$, et $\mathcal{P}_D = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$
- ☞ Pour tout $A \in \mathcal{P}_D \setminus \{\emptyset, D\}$, X_A est un **sous-ensemble des variables**

Notations

- ☞ $X = (X_1, \dots, X_d)$ dénotent **les variables (aléatoires) d'un modèle boîte-noire**
- ☞ On pose $D = \{1, \dots, d\}$ et $\mathcal{P}_D$ est **l'ensemble des sous-ensembles (power-set)** de D
Pour $d = 2$, on a $D = \{1, 2\}$, et $\mathcal{P}_D = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$
- ☞ Pour tout $A \in \mathcal{P}_D \setminus \{\emptyset, D\}$, X_A est un **sous-ensemble des variables**
- ☞ G est un **modèle boîte-noire**, et $G(X)$ est la **sortie aléatoire**
On suppose que $G(X)$ est une **variable aléatoire** (à valeur dans $\mathbb{R}$)

Notations

- ☞ $X = (X_1, \dots, X_d)$ dénotent **les variables (aléatoires) d'un modèle boîte-noire**
 - ☞ On pose $D = \{1, \dots, d\}$ et $\mathcal{P}_D$ est **l'ensemble des sous-ensembles (power-set)** de D
Pour $d = 2$, on a $D = \{1, 2\}$, et $\mathcal{P}_D = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$
 - ☞ Pour tout $A \in \mathcal{P}_D \setminus \{\emptyset, D\}$, X_A est un **sous-ensemble des variables**
 - ☞ G est un **modèle boîte-noire**, et $G(X)$ est la **sortie aléatoire**
On suppose que $G(X)$ est une **variable aléatoire** (à valeur dans $\mathbb{R}$)
- Analyse de sensibilité : $\mathbb{V}(G(X))$ est **la quantité d'incertitude** du modèle $G(X)$
- ☞ Décomposer $\mathbb{V}(G(X)) \iff$ Quantifier l'importance des variables

Notations

- ☞ $X = (X_1, \dots, X_d)$ dénotent **les variables (aléatoires) d'un modèle boîte-noire**
- ☞ On pose $D = \{1, \dots, d\}$ et $\mathcal{P}_D$ est **l'ensemble des sous-ensembles (power-set)** de D
Pour $d = 2$, on a $D = \{1, 2\}$, et $\mathcal{P}_D = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$
- ☞ Pour tout $A \in \mathcal{P}_D \setminus \{\emptyset, D\}$, X_A est un **sous-ensemble des variables**
- ☞ G est un **modèle boîte-noire**, et $G(X)$ est la **sortie aléatoire**
On suppose que $G(X)$ est une **variable aléatoire** (à valeur dans $\mathbb{R}$)

Analyse de sensibilité : $\mathbb{V}(G(X))$ est **la quantité d'incertitude** du modèle $G(X)$

- ☞ Décomposer $\mathbb{V}(G(X)) \iff$ Quantifier l'importance des variables

XAI : Pour une réalisation x de X , $G(x)$ est une prévision ponctuelle

- ☞ Décomposer $G(x) \iff$ "Expliquer" une prévision

Notations

- ☞ $X = (X_1, \dots, X_d)$ dénotent **les variables (aléatoires) d'un modèle boîte-noire**
- ☞ On pose $D = \{1, \dots, d\}$ et $\mathcal{P}_D$ est **l'ensemble des sous-ensembles (power-set)** de D
Pour $d = 2$, on a $D = \{1, 2\}$, et $\mathcal{P}_D = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$
- ☞ Pour tout $A \in \mathcal{P}_D \setminus \{\emptyset, D\}$, X_A est un **sous-ensemble des variables**
- ☞ G est un **modèle boîte-noire**, et $G(X)$ est la **sortie aléatoire**
On suppose que $G(X)$ est une **variable aléatoire** (à valeur dans $\mathbb{R}$)

Analyse de sensibilité : $\mathbb{V}(G(X))$ est **la quantité d'incertitude** du modèle $G(X)$

- ☞ Décomposer $\mathbb{V}(G(X)) \iff$ Quantifier l'importance des variables

XAI : Pour une réalisation x de X , $G(x)$ est une prévision ponctuelle

- ☞ Décomposer $G(x) \iff$ "Expliquer" une prévision

Ces deux décompositions prennent racine dans un résultat commun :

La décomposition fonctionnelle d'Hoeffding (1948)

Décomposition d'Hoeffding

Si l'on suppose que :

☞ Les variables X sont **mutuellement indépendantes**

☞ $\mathbb{V}(G(X)) < \infty$

Décomposition d'Hoeffding

Si l'on suppose que :

☞ Les variables X sont **mutuellement indépendantes**

☞ $\mathbb{V}(G(X)) < \infty$

Alors, il existe une **unique** décomposition

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

où $G_\emptyset$ est constant, et où les **représentants** sont tous **deux-à-deux orthogonaux**, i.e.,

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

Décomposition d'Hoeffding

Si l'on suppose que :

☞ Les variables X sont **mutuellement indépendantes**

☞ $\mathbb{V}(G(X)) < \infty$

Alors, il existe une **unique** décomposition

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

où $G_\emptyset$ est constant, et où les **représentants** sont tous **deux-à-deux orthogonaux**, i.e.,

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

De plus, ils sont caractérisés par

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

Hoeffding's decomposition

☞ Les **représentants** $G_A(X_A)$ sont les **interactions pures** induites par le modèle
Ce qui répond parfaitement à notre question !

Hoeffding's decomposition

Les **représentants** $G_A(X_A)$ sont les **interactions pures** induites par le modèle
Ce qui répond parfaitement à notre question !

Par exemple :

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

Dans ce cas, nous avons que :

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

Hoeffding's decomposition

Les **représentants** $G_A(X_A)$ sont les **interactions pures** induites par le modèle
Ce qui répond parfaitement à notre question !

Par exemple :

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

Dans ce cas, nous avons que :

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

On peut **recomposer le modèle** en ayant accès qu'à ses **représentants**

Quantifier l'importance

Pour quantifier l'importance, on veut décomposer $\mathbb{V}(G(X))$

Quantifier l'importance

Pour quantifier l'importance, on veut décomposer $\mathbb{V}(G(X))$

Les **indices de Sobol (2001)** (FANOVA) le permettent :

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

en se basant sur la **décomposition d'Hoeffding**

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

L'incertitude du modèle est égale à l'incertitude de ses **représentants**

Quantifier l'importance

Pour quantifier l'importance, on veut décomposer $\mathbb{V}(G(X))$

Les **indices de Sobol (2001)** (FANOVA) le permettent :

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

en se basant sur la **décomposition d'Hoeffding**

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

L'incertitude du modèle est égale à l'incertitude de ses **représentants**

C'est un **outil puissant** pour interpréter les modèles boîtes-noires :

☞ Ils sont **toujours positifs** ($S_A \geq 0$) et **somment à 1**

Quantifier l'importance

Pour quantifier l'importance, on veut décomposer $\mathbb{V}(G(X))$

Les **indices de Sobol (2001)** (FANOVA) le permettent :

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

en se basant sur la **décomposition d'Hoeffding**

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

L'incertitude du modèle est égale à l'incertitude de ses **représentants**

C'est un **outil puissant** pour interpréter les modèles boîtes-noires :

☞ Ils sont **toujours positifs** ($S_A \geq 0$) et **somment à 1**

⇒ Interprétation **comme un pourcentage de l'incertitude du modèle**

“Expliquer” une prévision

Pour “expliquer” une prévision, on veut décomposer $G(x)$ pour une réalisation x de X

“Expliquer” une prévision

Pour “expliquer” une prévision, on veut décomposer $G(x)$ pour une réalisation x de X

Grâce à la décomposition d’Hoeffding, on a que :

$$E_A = G_A(x_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{E}[G(X) \mid X_B = x_B]$$

La prévision du modèle est égale l’évaluation de ses **représentants**

“Expliquer” une prévision

Pour “expliquer” une prévision, on veut décomposer $G(x)$ pour une réalisation x de X

Grâce à la décomposition d’Hoeffding, on a que :

$$E_A = G_A(x_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{E}[G(X) \mid X_B = x_B]$$

La prévision du modèle est égale l’évaluation de ses **représentants**

C’est (aussi) un **outil puissant** pour interpréter les modèles boîtes-noires :

👉 Ils **somment à la prévision** $G(x)$!

“Expliquer” une prévision

Pour “expliquer” une prévision, on veut décomposer $G(x)$ pour une réalisation x de X

Grâce à la décomposition d’Hoeffding, on a que :

$$E_A = G_A(x_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{E}[G(X) \mid X_B = x_B]$$

La prévision du modèle est égale l’évaluation de ses **représentants**

C’est (aussi) un **outil puissant** pour interpréter les modèles boîtes-noires :

👉 Ils **somment à la prévision** $G(x)$!

⇒ Interprétation **comme une contribution à la prévision**

Dépendance

Mais, ces méthodes requièrent l'indépendance mutuelle des variables...

Mais, ces méthodes requièrent l'indépendance mutuelle des variables...

Si on garde **les mêmes formules** :

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

$$E_A = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{E}[G(X) \mid X_B = x_B]$$

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

Indices de Sobol'

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

Décomposition de $G(x_1, x_2, x_3)$

$$\begin{aligned} E_1 &= x_1 & E_2 &= 0, & E_3 &= \rho x_3, \\ E_{12} &= \rho x_1 x_2, & E_{13} &= -\rho x_3, & E_{23} &= x_2 x_3, \\ E_{123} &= -\rho x_1 x_2 \end{aligned}$$

☞ **Somment toujours aux bonnes choses mais la décomposition n'est pas claire**

+ Indices de Sobol' négatifs difficilement interprétables

Décomposition d'Hoeffding et dépendance

Il y a deux grand problèmes :

1. **Indices de Sobol' négatifs**
2. **Interprétation des décompositions peu claire**

Décomposition d'Hoeffding et dépendance

Il y a deux grand problèmes :

1. **Indices de Sobol' négatifs**
2. **Interprétation des décompositions peu claire**

Mais, il y a des solutions !

1. **Allocations de jeux coopératifs** (e.g., valeurs de Shapley)
2. **Généralisation de la décomposition d'Hoeffding**

Valeurs de Shapley

Effets de Shapley (Owen 2014) : **Redistribution égalitaire des indices de Sobol'** S_A

$$Sh_1 = \sum_{A \in \mathcal{P}_D : \{1\} \subset A} \frac{S_A}{|A|} = S_1 + \frac{S_{12}}{2} + \frac{S_{13}}{2} + \dots + \frac{S_{123}}{3} + \dots + \frac{S_{1234}}{4} + \dots$$

Ils sont positifs et somment à 1

Premier problème résolu !

Valeurs de Shapley

Effets de Shapley (Owen 2014) : **Redistribution égalitaire des indices de Sobol'** S_A

$$Sh_1 = \sum_{A \in \mathcal{P}_D : \{1\} \subset A} \frac{S_A}{|A|} = S_1 + \frac{S_{12}}{2} + \frac{S_{13}}{2} + \dots + \frac{S_{123}}{3} + \dots + \frac{S_{1234}}{4} + \dots$$

Ils sont positifs et somment à 1

Premier problème résolu !

SHAP (Lundberg et Lee 2017) : **Redistribution égalitaire des** E_A

$$SHAP_1 = \sum_{A \in \mathcal{P}_D : \{1\} \subset A} \frac{E_A}{|A|} = E_1 + \frac{E_{12}}{2} + \frac{E_{13}}{2} + \dots + \frac{E_{123}}{3} + \dots + \frac{E_{1234}}{4} + \dots$$

Ils somment toujours à $G(x)$!

Valeurs de Shapley

Effets de Shapley (Owen 2014) : **Redistribution égalitaire des indices de Sobol'** S_A

$$Sh_1 = \sum_{A \in \mathcal{P}_D : \{1\} \subset A} \frac{S_A}{|A|} = S_1 + \frac{S_{12}}{2} + \frac{S_{13}}{2} + \dots + \frac{S_{123}}{3} + \dots + \frac{S_{1234}}{4} + \dots$$

Ils sont positifs et somment à 1

Premier problème résolu !

SHAP (Lundberg et Lee 2017) : **Redistribution égalitaire des** E_A

$$SHAP_1 = \sum_{A \in \mathcal{P}_D : \{1\} \subset A} \frac{E_A}{|A|} = E_1 + \frac{E_{12}}{2} + \frac{E_{13}}{2} + \dots + \frac{E_{123}}{3} + \dots + \frac{E_{1234}}{4} + \dots$$

Ils somment toujours à $G(x)$!

Revenons un peu en arrière...

Théorie des jeux coopératifs

SHAP et les **effets de Shapley** sont tirés des **valeurs de Shapley (1951)**

Qui font partie de **la théorie des jeux coopératifs**

Théorie des jeux coopératifs

SHAP et les **effets de Shapley** sont tirés des **valeurs de Shapley (1951)**

Qui font partie de **la théorie des jeux coopératifs**

**Théorie des jeux coopératifs =
Art de partager un gâteau**



Théorie des jeux coopératifs

SHAP et les **effets de Shapley** sont tirés des **valeurs de Shapley (1951)**

Qui font partie de **la théorie des jeux coopératifs**

**Théorie des jeux coopératifs =
Art de partager un gâteau**



Soit $D = \{1, \dots, d\}$ un **ensemble de joueurs**, et $\mathcal{P}_D$ l'**ensemble des coalitions**

On **choisit** $v : \mathcal{P}_D \rightarrow \mathbb{R}$ comme **fonction de valeur**

Assigne une valeur à une coalition de joueurs

👉 (D, v) définit formellement un **jeu coopératif**

Théorie des jeux coopératifs

SHAP et les **effets de Shapley** sont tirés des **valeurs de Shapley (1951)**

Qui font partie de **la théorie des jeux coopératifs**

**Théorie des jeux coopératifs =
Art de partager un gâteau**



Soit $D = \{1, \dots, d\}$ un **ensemble de joueurs**, et $\mathcal{P}_D$ l'**ensemble des coalitions**

On **choisit** $v : \mathcal{P}_D \rightarrow \mathbb{R}$ comme **fonction de valeur**

Assigne une valeur à une coalition de joueurs

👉 (D, v) définit formellement un **jeu coopératif**

Question centrale :

Comment redistribuer $v(D)$ à chacun des d joueurs?

$v(D)$ = gâteau = la valeur de tous les joueurs

Allocations

Une **allocation** associe une part de $v(D)$ à **chaque** joueurs

C'est une fonction $\psi : D \rightarrow \mathbb{R}$

Allocations

Une **allocation** associe une part de $v(D)$ à **chaque** joueurs

C'est une fonction $\psi : D \rightarrow \mathbb{R}$

For **quantifier l'importance** (décomposer $\mathbb{V}(G(X))$), deux critères sont primordiaux :

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

On redistribue tout le gâteau et rien que le gâteau

Allocations

Une **allocation** associe une part de $v(D)$ à **chaque** joueurs

C'est une fonction $\psi : D \rightarrow \mathbb{R}$

For **quantifier l'importance** (décomposer $\mathbb{V}(G(X))$), deux critères sont primordiaux :

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

On redistribue tout le gâteau et rien que le gâteau

☞ Dans ce cas, ψ s'interprète comme **un pourcentage d'incertitude due à chaque variable**

Et non plus à chaque sous-ensemble de variables

Allocations

Une **allocation** associe une part de $v(D)$ à **chaque** joueurs

C'est une fonction $\psi : D \rightarrow \mathbb{R}$

For **quantifier l'importance** (décomposer $\mathbb{V}(G(X))$), deux critères sont primordiaux :

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

On redistribue tout le gâteau et rien que le gâteau

☞ Dans ce cas, ψ s'interprète comme **un pourcentage d'incertitude due à chaque variable**

Et non plus à chaque sous-ensemble de variables

Comment trouver des allocations efficaces et positives ?

L'ensemble d'Harsanyi

Les **dividendes d'Harsanyi (1963)** d'un jeu coopératif (D, v) sont définis par :

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

C'est une fonction de $\mathcal{D}_v(A) : \mathcal{P}_D \rightarrow \mathbb{R}$

☞ Interprétation comme la **plus-value** produite par chaque **coalition**

☞ Ils somment **toujours** à $v(D)$:

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

L'ensemble d'Harsanyi

Les **dividendes d'Harsanyi (1963)** d'un jeu coopératif (D, v) sont définis par :

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

C'est une fonction de $\mathcal{D}_v(A) : \mathcal{P}_D \rightarrow \mathbb{R}$

☞ Interprétation comme la **plus-value** produite par chaque **coalition**

☞ Ils somment **toujours** à $v(D)$:

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

L'**ensemble de Harsanyi** (Vasil'ev et Laan 2001) est un ensemble **d'allocations** qui agrègent les dividendes d'Harsanyi :

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

L'ensemble d'Harsanyi

Les **dividendes d'Harsanyi (1963)** d'un jeu coopératif (D, v) sont définis par :

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

C'est une fonction de $\mathcal{D}_v(A) : \mathcal{P}_D \rightarrow \mathbb{R}$

☞ Interprétation comme la **plus-value** produite par chaque **coalition**

☞ Ils somment **toujours** à $v(D)$:

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

L'**ensemble de Harsanyi** (Vasil'ev et Laan 2001) est un ensemble **d'allocations** qui agrègent les dividendes d'Harsanyi :

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

☞ **Allocations toujours efficaces**

☞ **Positives si v est monotone**

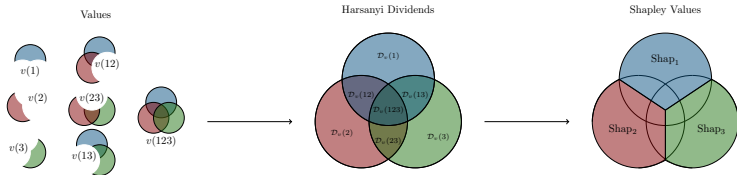
C'est le cas pour $\mathbb{V}(\mathbb{E}[G(X) \mid X_A])$

Agrégation égalitaire : les valeurs de Shapley

Les **valeurs de Shapley** (1951) sont l'agrégation égalitaire des dividendes.

Pour chaque joueur $i \in D$,

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

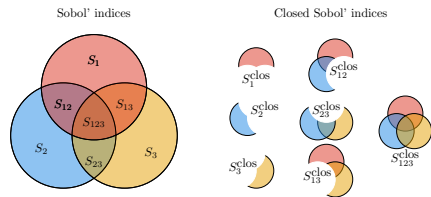
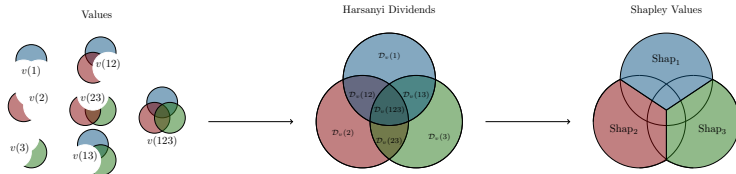


Agrégation égalitaire : les valeurs de Shapley

Les **valeurs de Shapley (1951)** sont l'agrégation égalitaire des dividendes.

Pour chaque joueur $i \in D$,

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



Effets de Shapley : choisir $v(A) = \mathbb{V}(\mathbb{E}[G(X) \mid X_A])$

Le gâteau devient $\mathbb{V}(G(X))$

Les dividendes d'Harsanyi deviennent S_A

SHAP : choisir $v(A) = \mathbb{E}[G(X) \mid X_A = x_A]$

Le gâteau devient $G(x)$

Les dividendes d'Harsanyi deviennent E_A

Détection de variables exogènes

Mais, les **effets de Shapley** ont un **comportement problématique**
(pour la décomposition de la variance)

Détection de variables exogènes

Mais, les **effets de Shapley** ont un **comportement problématique**

(pour la décomposition de la variance)

Une **variable exogène** peut **recevoir une part d'importance**

C'est la **blague de Shapley**

(looss et Prieur 2019)

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

$$Sh_1 = 0.5 - \rho^2/4, \quad Sh_2 = 0.5, \quad \underline{Sh_3 = \rho^2/4}$$

Détection de variables exogènes

Mais, les **effets de Shapley** ont un **comportement problématique**

(pour la décomposition de la variance)

Une **variable exogène** peut **recevoir une part d'importance**

C'est la **blague de Shapley**

(looss et Prieur 2019)

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

$$Sh_1 = 0.5 - \rho^2/4, \quad Sh_2 = 0.5, \quad \underline{Sh_3 = \rho^2/4}$$

Pour résoudre ce problème, on peut se tourner vers **un agrégation proportionnelle des dividendes** (Ortmann 2000)

Ainsi, on a pu définir les **effets marginaux proportionnels (PME)** (Herin et al. 2023)

Détection de variables exogènes

Mais, les **effets de Shapley** ont un **comportement problématique**

(pour la décomposition de la variance)

Une **variable exogène** peut **recevoir une part d'importance**

C'est la **blague de Shapley**

(looss et Prieur 2019)

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

$$Sh_1 = 0.5 - \rho^2/4, \quad Sh_2 = 0.5, \quad \underline{Sh_3 = \rho^2/4}$$

Pour résoudre ce problème, on peut se tourner vers **un agrégation proportionnelle des dividendes** (Ortmann 2000)

Ainsi, on a pu définir les **effets marginaux proportionnels (PME)** (Herin et al. 2023)

Proposition (*Detection de l'exogénéité*). Sous de légères hypothèses sur la structure probabiliste de X ,

$$PME_i = 0 \iff X_i \text{ est exogène.}$$

En pratique, indices **plus discriminants que les effets de Shapley** en situation de forte corrélation

Estimer les PME/Effets de Shapley $\iff$ Estimer $v(A)$ pour tout $A \in \mathcal{P}_D$

Deux cas pratiques :

- **On peut échantillonner les variables (Monte Carlo)** : Nécessite un nombre proportionnel à $d!(d-1)$ d'appel au modèle boîte-noire (Song, Nelson et Staum 2016)
Le coût computationnel peut être réduit au prix de la précision
- **On a seulement accès à un échantillon i.i.d** : Approche “plus-proches voisins” nécessite 2^d estimations (Broto, Bachoc et Depecker 2020).

Estimation TRÈS chronophage et augmente de manière exponentielle avec d , mais les **estimations peuvent être recyclées pour calculer les deux allocations d'un coup**

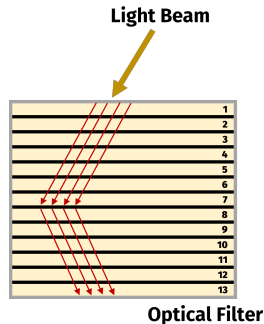
Transmission d'un filtre optique

Performance en transmission d'un **filtre optique** composé de 13 couches consécutives (Vasseur et al. 2010)

Les variables $I_1, \dots, I_{13}$ représentent l'**erreur de réfraction** des couches

Elles sont (très) corrélées due au processus de fabrication séquentiel

Le modèle numérique calcule l'**erreur de transmission par rapport à un filtre parfait** sur plusieurs longueurs d'onde



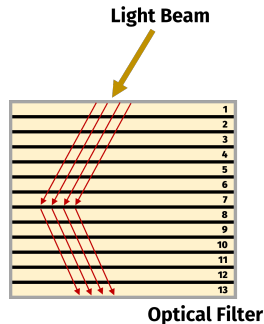
Transmission d'un filtre optique

Performance en transmission d'un **filtre optique** composé de 13 couches consécutives (Vasseur et al. 2010)

Les variables $I_1, \dots, I_{13}$ représentent l'**erreur de réfraction** des couches

Elles sont (très) corrélées due au processus de fabrication séquentiel

Le modèle numérique calcule l'**erreur de transmission par rapport à un filtre parfait** sur plusieurs longueurs d'onde



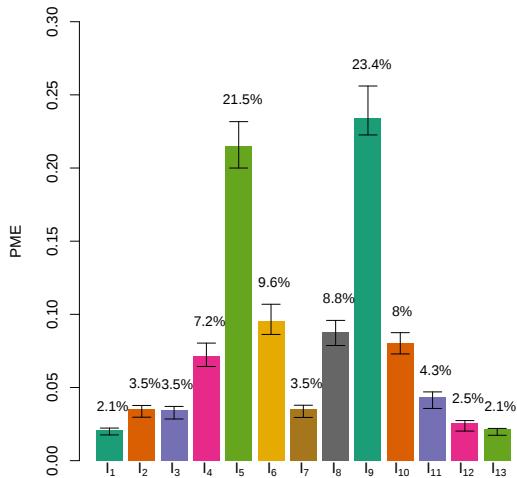
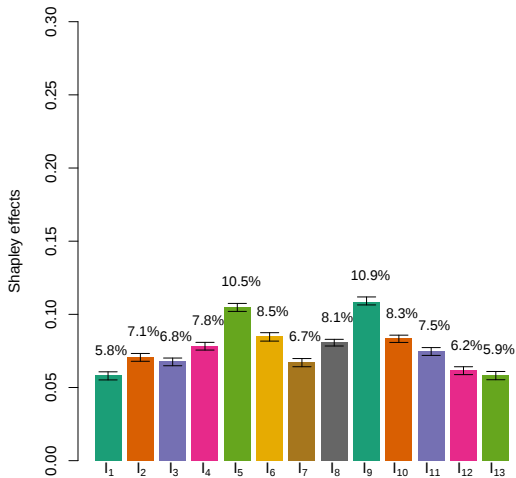
👉 **On a seulement accès à un jeu de données i.i.d.** ($n = 1000$).

Allocations calculées via approche **plus-proches voisins**

Implémentation parallélisée (package R *sensitivity*) (~ 4 min d'exécution, 8 coeurs)

Nombre de voisins choisis arbitrairement : 6

Transmission d'un filtre optique



Transmission d'un filtre optique - Sélection de variables

Scénario : On veut remplacer le modèle numérique par un méta-modèle (processus Gaussien*)

Sur toutes les variables : $Q^2 = 99.48\%$.

Sélection de variables :

- Premier seuil : 2.5% d'importance
 - **Effets de Shapley :** Aucune variable retirée
 - **PME :** I_1 et I_3 sont retirées, $Q^2 = 99.14\%$
- Deuxième seuil : 5% importance
 - **Effets de Shapley :** Aucune variable retirée
 - **PME :** 7 variables sont retirées, $Q^2 = 98.79\%$

Conclusion partielle

☞ Les **indices de Sobol'** sont un outil remarquable pour mesurer **l'importance des variables** d'un modèle boîte-noire

Conclusion partielle

☞ Les **indices de Sobol'** sont un outil remarquable pour mesurer **l'importance des variables d'un modèle boîte-noire**

☞ Lorsque les variables sont dépendantes, il y a **deux problèmes** :

1. Les **indices de Sobol'** peuvent être négatifs
2. Leur **interprétation en tant qu'interaction est moins claire**

Conclusion partielle

☞ Les **indices de Sobol'** sont un outil remarquable pour mesurer **l'importance des variables d'un modèle boîte-noire**

☞ Lorsque les variables sont dépendantes, il y a **deux problèmes** :

1. Les **indices de Sobol'** peuvent être négatifs
2. Leur **interprétation en tant qu'interaction est moins claire**

☞ **Solution au premier problème** : **allocations** issues de la théorie des jeux coopératifs
e.g., les **effets Shapley** inspirés de l'agrégation égalitaire, i.e., les **valeurs de Shapley**

Conclusion partielle

☞ Les **indices de Sobol'** sont un outil remarquable pour mesurer **l'importance des variables d'un modèle boîte-noire**

☞ Lorsque les variables sont dépendantes, il y a **deux problèmes** :

1. Les **indices de Sobol'** peuvent être négatifs
2. Leur **interprétation en tant qu'interaction est moins claire**

☞ **Solution au premier problème** : **allocations** issues de la théorie des jeux coopératifs
e.g., les **effets Shapley** inspirés de l'agrégation égalitaire, i.e., les **valeurs de Shapley**

☞ La plupart des allocation intéressantes sont **une agrégation des dividendes de Harsanyi**
Ce sont **des agrégations des indices de Sobol'** (importance), ou des E_A (prévision)

Conclusion partielle

☞ Les **indices de Sobol'** sont un outil remarquable pour mesurer **l'importance des variables d'un modèle boîte-noire**

☞ Lorsque les variables sont dépendantes, il y a **deux problèmes** :

1. Les **indices de Sobol'** peuvent être négatifs
2. Leur **interprétation en tant qu'interaction est moins claire**

☞ **Solution au premier problème** : **allocations** issues de la théorie des jeux coopératifs
e.g., les **effets Shapley** inspirés de l'agrégation égalitaire, i.e., les **valeurs de Shapley**

☞ La plupart des allocation intéressantes sont **une agrégation des dividendes de Harsanyi**
Ce sont **des agrégations des indices de Sobol'** (importance), ou des E_A (prévision)

☞ Pour identifier les **variables exogènes**, il vaut mieux utiliser les **effets marginaux proportionnels**

Conclusion partielle

☞ Les **indices de Sobol'** sont un outil remarquable pour mesurer l'importance des variables d'un modèle boîte-noire

☞ Lorsque les variables sont dépendantes, il y a **deux problèmes** :

1. Les **indices de Sobol'** peuvent être négatifs
2. Leur **interprétation en tant qu'interaction est moins claire**

☞ **Solution au premier problème** : **allocations** issues de la théorie des jeux coopératifs e.g., les **effets Shapley** inspirés de l'agrégation égalitaire, i.e., les **valeurs de Shapley**

☞ La plupart des allocation intéressantes sont **une agrégation des dividendes de Harsanyi**
Ce sont **des agrégations des indices de Sobol'** (importance), ou des E_A (prévision)

☞ Pour identifier les **variables exogènes**, il vaut mieux utiliser les **effets marginaux proportionnels**

Mais qu'en est-il du deuxième problème ?

Généraliser la décomposition d'Hoeffding pour des variables dépendantes

La **décomposition d'Hoeffding** garantit le **sens** des indices de Sobol'...

Généraliser la décomposition d'Hoeffding pour des variables dépendantes

La **décomposition d'Hoeffding** garantit le **sens** des indices de Sobol'...
... naturellement, on aimerait la **generaliser à des variables dépendantes!**

Généraliser la décomposition d'Hoeffding pour des variables dépendantes

La **décomposition d'Hoeffding** garantit le sens des indices de Sobol'...
... naturellement, on aimerait la **generaliser à des variables dépendantes!**

On est pas les premiers à s'être rendus compte de ça :

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

Généraliser la décomposition d'Hoeffding pour des variables dépendantes

La **décomposition d'Hoeffding** garantit le sens des indices de Sobol'...
... naturellement, on aimerait la **generaliser à des variables dépendantes!**

On est pas les premiers à s'être rendus compte de ça :

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

Mais on a découvert une approche intéressante :)

Généraliser la décomposition d'Hoeffding pour des variables dépendantes

Premier défi : définir des situations de “**non-indépendance mutuelle**”

Généraliser la décomposition d'Hoeffding pour des variables dépendantes

Premier défi : définir des situations de “**non-indépendance mutuelle**”

Or, l'**indépendance mutuelle** est plutôt **restrictive**

Bien plus que l'**indépendance deux-à-deux**

Généraliser la décomposition d'Hoeffding pour des variables dépendantes

Premier défi : définir des situations de “**non-indépendance mutuelle**”

Or, l'**indépendance mutuelle** est plutôt **restrictive**

Bien plus que l'**indépendance deux-à-deux**

On impose deux hypothèses “bien moins restrictives ” sur X :

- Dépendance fonctionnelle non-parfaite entre les variables
- Dépendance stochastique non-dégénérée entre les variables

Interprétation : On peut *distinguer* les variables (chacune apporte un signal différent)

Conséquences

Sous ces deux hypothèses, n'importe quelle **sortie aléatoire** $G(X)$ à second moment fini peut être décomposé de manière **unique** :

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

où les $G_A(X_A)$ sont **hiérarchiquement orthogonaux**

Et non plus **orthogonaux deux-à-deux** avec le résultat classique d'Hoeffding

Conséquences

Sous ces deux hypothèses, n'importe quelle **sortie aléatoire** $G(X)$ à second moment fini peut être décomposé de manière **unique** :

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

où les $G_A(X_A)$ sont **hiérarchiquement orthogonaux**

Et non plus **orthogonaux deux-à-deux** avec le résultat classique d'Hoeffding

Les **représentants** peuvent être caractérisés par des **projections obliques** $\mathbb{M}_A[G(X)]$ qui **changent en fonction de la structure de dépendance** :

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

Conséquences

Sous ces deux hypothèses, n'importe quelle **sortie aléatoire** $G(X)$ à second moment fini peut être décomposé de manière **unique** :

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

où les $G_A(X_A)$ sont **hiérarchiquement orthogonaux**

Et non plus **orthogonaux deux-à-deux** avec le résultat classique d'Hoeffding

Les **représentants** peuvent être caractérisés par des **projections obliques** $\mathbb{M}_A[G(X)]$ qui **changent en fonction de la structure de dépendance** :

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

Et quand les variables sont **mutuellement indépendantes**, $\mathbb{M}_A[G(X)] = \mathbb{E}[G(X) \mid X_A]$, et

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

Conséquences

Sous ces deux hypothèses, n'importe quelle **sortie aléatoire** $G(X)$ à second moment fini peut être décomposé de manière **unique** :

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

où les $G_A(X_A)$ sont **hiérarchiquement orthogonaux**

Et non plus **orthogonaux deux-à-deux** avec le résultat classique d'Hoeffding

Les **représentants** peuvent être caractérisés par des **projections obliques** $\mathbb{M}_A[G(X)]$ qui **changent en fonction de la structure de dépendance** :

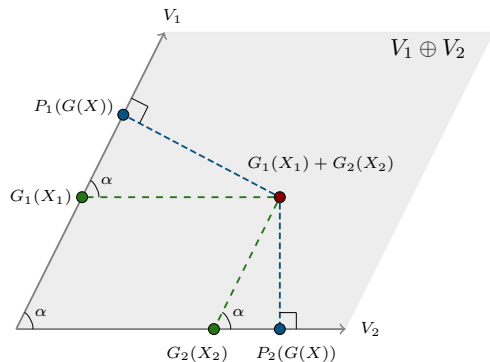
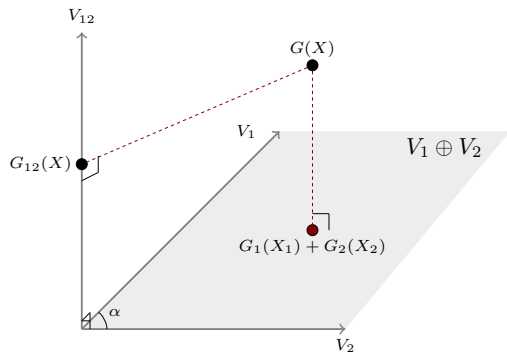
$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{M}_B[G(X)], \quad \forall A \in \mathcal{P}_D$$

Et quand les variables sont **mutuellement indépendantes**, $\mathbb{M}_A[G(X)] = \mathbb{E}[G(X) \mid X_A]$, et

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

👉 On retrouve le résultat d'Hoeffding !

Illustration* : modèle bivarié



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

Qu'est-ce qu'on a appris en développant cette généralisation :

Qu'est-ce qu'on a appris en développant cette généralisation :

☞ La dépendance introduit des **angles entre les fonctions des sous-ensembles de variables**

Indépendance mutuelle $\iff$ Tout le monde est orthogonal

Qu'est-ce qu'on a appris en développant cette généralisation :

- ☞ La dépendance introduit des **angles entre les fonctions des sous-ensembles de variables**
Indépendance mutuelle $\iff$ Tout le monde est orthogonal
- ☞ Pour une **décomposition unique**, il faut **corriger les projections en fonction de ces angles**
Espérance conditionnelle **en situation de dépendance** $\implies$ Interprétation floue du début

Qu'est-ce qu'on a appris en développant cette généralisation :

☞ La dépendance introduit des **angles entre les fonctions des sous-ensembles de variables**

Indépendance mutuelle $\iff$ Tout le monde est orthogonal

☞ Pour une **décomposition unique**, il faut **corriger les projections en fonction de ces angles**

Espérance conditionnelle **en situation de dépendance** $\implies$ Interprétation floue du début

☞ L'interprétation d'un modèle **change en fonction des données d'entrées**

Par ex. un modèle peut respecter une "équité" pour un jeu de données, mais pas pour un autre

Qu'est-ce qu'on a appris en développant cette généralisation :

☞ La dépendance introduit des **angles entre les fonctions des sous-ensembles de variables**

Indépendance mutuelle $\iff$ Tout le monde est orthogonal

☞ Pour une **décomposition unique**, il faut **corriger les projections en fonction de ces angles**

Espérance conditionnelle **en situation de dépendance** $\implies$ Interprétation floue du début

☞ L'interprétation d'un modèle **change en fonction des données d'entrées**

Par ex. un modèle peut respecter une "équité" pour un jeu de données, mais pas pour un autre

Mais il reste du travail avant sa mise en oeuvre pratique...

Quelques perspectives

👉 Défi principal : Estimer ces projections obliques à partir de d'observations

Mais on a plein de pistes prometteuses :)

👉 Défi principal : Estimer ces projections obliques à partir de d'observations

Mais on a plein de pistes prometteuses :)

👉 Applications pour la quantification d'influence

Evaluation de l'équité, Analyse de sensibilité locale et globale, XAI maîtrisée théoriquement,

👉 Défi principal : Estimer ces projections obliques à partir de d'observations

Mais on a plein de pistes prometteuses :)

👉 Applications pour la quantification d'influence

Evaluation de l'équité, Analyse de sensibilité locale et globale, XAI maîtrisée théoriquement,

👉 Implémentation informatique efficace

Ouvrir l'usage au plus grand nombre

References i

- Broto, B., F. Bachoc et M. Depecker. 2020. "Variance reduction for estimation of Shapley effects and adaptation to unknown input distribution". *SIAM/ASA Journal on Uncertainty Quantification* 8 : 693-716.
- Bryc, W. 1984. "Conditional expectation with respect to dependent sigma-fields". In *Proceedings of VII conference on Probability Theory*, 409-411. <https://homepages.uc.edu/~brycwz/preprint/Brasov-1982.pdf>.
- . 1996. "Conditional Moment Representations for Dependent Random Variables". Publisher : Institute of Mathematical Statistics and Bernoulli Society, *Electronic Journal of Probability* 1 (none) : 1-14. issn : 1083-6489, 1083-6489. <https://doi.org/10.1214/EJP.v1-7>. <https://projecteuclid.org/journals/electronic-journal-of-probability/volume-1/issue-none/Conditional-Moment-Representations-for-Dependent-Random-Variables/10.1214/EJP.v1-7.full>.
- Chastaing, G., F. Gamboa et 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, n° none (janvier) : 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 et C. Prieur. 2021. *Basics and Trends in Sensitivity Analysis. Theory and Practice in R*. SIAM. Computational Science / Engineering.
- Dauxois, J, G. M Nkiet et Y Romain. 2004. "Canonical analysis relative to a closed subspace". *Linear Algebra and its Applications*, Tenth Special Issue (Part 1) on Linear Algebra and Statistics, 388 : 119-145. issn : 0024-3795. <https://doi.org/10.1016/j.laa.2004.02.036>. <https://www.sciencedirect.com/science/article/pii/S0024379504001107>.

References ii

- Dixmier, J. 1949. "Étude sur les variétés et les opérateurs de Julia, avec quelques applications" [en fre]. *Bulletin de la Société Mathématique de France* 77 : 11-101. <http://eudml.org/doc/86830>.
- Feshchenko, I. 2020. *When is the sum of closed subspaces of a Hilbert space closed ?* <https://doi.org/10.48550/arXiv.2012.08688>. arXiv : 2012.08688 [math.FA].
- Friedrichs, K. 1937. "On Certain Inequalities and Characteristic Value Problems for Analytic Functions and For Functions of Two Variables". Publisher : American Mathematical Society, *Transactions of the American Mathematical Society* 41 (3) : 321-364. issn : 0002-9947. <https://doi.org/10.2307/1989786>. <https://www.jstor.org/stable/1989786>.
- Galántai, A. 2004. *Projectors and Projection Methods*. Boston, MA : Springer US. isbn : 978-1-4613-4825-2. <https://doi.org/10.1007/978-1-4419-9180-5>. <http://link.springer.com/10.1007/978-1-4419-9180-5>.
- Gebelein, H. 1941. "Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung" [en de]. *ZAMM - Zeitschrift für Angewandte Mathematik und Mechanik* 21 (6) : 364-379. issn : 00442267, 15214001. <https://doi.org/10.1002/zamm.19410210604>. <https://onlinelibrary.wiley.com/doi/10.1002/zamm.19410210604>.
- 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., et P. A. Gremaud. 2018. "An approximation theoretic perspective of Sobol' indices with dependent variables" [en 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 et 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" [en en]. *Journal of Computational and Graphical Statistics* 16 (3) : 709-732. <http://www.jstor.org/stable/27594267>.
- Iooss, B., et 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.
- Koyak, R. A. 1987. "On Measuring Internal Dependence in a Set of Random Variables". Publisher : Institute of Mathematical Statistics, *The Annals of Statistics* 15 (3) : 1215-1228. ISSN : 0090-5364, 2168-8966. <https://doi.org/10.1214/aos/1176350501>.
<https://projecteuclid.org/journals/annals-of-statistics/volume-15/issue-3/On-Measuring-Internal-Dependence-in-a-Set-of-Random-Variables/10.1214/aos/1176350501.full>.

- Kuo, F. Y., I. H. Sloan, G. W. Wasilkowski et H. Woźniakowski. 2009. "On decompositions of multivariate functions" [en en]. *Mathematics of Computation* 79, n° 270 (novembre) : 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>.
- Lebrun, R., et 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.pro bengmech.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.pro bengmech.2009.04.006>.
<https://www.sciencedirect.com/science/article/pii/S0266892009000307>.
- Lundberg, S., et S. Lee. 2017. "A Unified Approach to Interpreting Model Predictions". Décembre.
https://www.researchgate.net/publication/317062430_A_Unified_Approach_to_Interpreting_Model_Predictions.
- Mara, Thierry A., Stefano Tarantola et Paola Annoni. 2015. "Non-parametric methods for global sensitivity analysis of model output with dependent inputs" [en en]. *Environmental Modelling & Software* 72 (octobre) : 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.

References v

- 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., et O. Aliş. 1999. "General foundations of high-dimensional model representations" [en 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>.
- 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>.
- 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.
- Sidák, Z. 1957. "On Relations Between Strict-Sense and Wide-Sense Conditional Expectations". *Theory of Probability & Its Applications* 2 (2) : 267-272. issn : 0040-585X. <https://doi.org/10.1137/1102020>. <https://epubs.siam.org/doi/abs/10.1137/1102020>.
- 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 et J. Staum. 2016. "Shapley effects for global sensitivity analysis : Theory and computation". *SIAM/ASA Journal on Uncertainty Quantification* 4 : 1060-1083.
- Vasil'ev, V., et 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>.
- Vasseur, O., M. Claeys-Bruno, M. Cathelinaud et 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.

MERCI DE VOTRE ATTENTION !

DES QUESTIONS ?

MAROUANEILIDRISSI.COM

Non-perfect functional dependence

Assumption 1 (*Non-perfect functional dependence*).

- $\sigma_\emptyset \subset \sigma_i, i = 1, \dots, d$ (inputs are not constant).
- For $B \subset A, \sigma_B \subset \sigma_A$ (inputs add information).
- For every $A, B \in \mathcal{P}_D, A \neq B, \sigma_A \cap \sigma_B = \sigma_{A \cap B}$.

⇒ $\mathbb{L}^2(\sigma_\emptyset) \subset \mathbb{L}^2(\sigma_A)$, for every $A \in \mathcal{P}_D$

There are non-constant random variables in the Lebesgue spaces

⇒ For $B \subset A, \mathbb{L}^2(\sigma_B) \subset \mathbb{L}^2(\sigma_A)$

There are functions of X_A that are not functions of X_B

⇒ For any $A, B \in \mathcal{P}_D, \mathbb{L}^2(\sigma_A) \cap \mathbb{L}^2(\sigma_B) = \mathbb{L}^2(\sigma_{A \cap B})$

The functions of X_A and X_B are functions of $X_{A \cap B}$

Proposition . Under Assumption 1, for any $A, B \in \mathcal{P}_D$ such that $A \cap B \notin \{A, B\}$, **there is no mapping T such that $X_B = T(X_A)$ a.e.**

Non-perfect stochastic dependence

Definition (*Friedrichs (1937) angle*). The cosine of Friedrichs' angle is defined as

$$c(M, N) := \sup \left\{ |\langle x, y \rangle| : \begin{cases} x \in M \cap (M \cap N)^\perp, \|x\| \leq 1 \\ y \in N \cap (M \cap N)^\perp, \|y\| \leq 1 \end{cases} \right\},$$

where the orthogonal complement is taken w.r.t. to H .

☞ Analogous to the **maximal partial dependence** between random elements (Bryc 1984, 1996; Dauxois, Nkiet et Romain 2004)

Definition (*Feshchenko matrix*). Let Δ be the $(2^d \times 2^d)$, symmetric **set-indexed** matrix, defined element-wise, $\forall A, B \in \mathcal{P}_D$ as

$$\Delta_{AB} = \begin{cases} 1 & \text{if } A = B; \\ -c(\mathbb{L}^2(\sigma_A), \mathbb{L}^2(\sigma_B)) & \text{otherwise.} \end{cases}$$

Assumption 2 (*Non-degenerate stochastic dependence*). The Feshchenko matrix Δ of X is definite-positive.

Non-perfect functional dependence

- ☞ $\mathbb{L}^2(\sigma_X)$ contains random variables that are functions of X .
- ☞ For every $A \subset D$, $\mathbb{L}^2(\sigma_A)$ contains random variables that are functions of X_A .
- ☞ $\mathbb{L}^2(\sigma_\emptyset)$ contains constants.

Theorem Sidák (1957, Theorem 2). Let $\mathcal{G}_1, \mathcal{G}_2 \subseteq \mathcal{F}$, then

- If $\mathcal{G}_1 \subset \mathcal{G}_2$, then $\mathbb{L}^2(\mathcal{G}_1) \subset \mathbb{L}^2(\mathcal{G}_2) \subseteq \mathbb{L}^2(\mathcal{F})$;
- $\mathbb{L}^2(\mathcal{G}_1) \cap \mathbb{L}^2(\mathcal{G}_2) = \mathbb{L}^2(\mathcal{G}_1 \cap \mathcal{G}_2)$.

Assumption 1 (Non-perfect functional dependence).

- $\sigma_\emptyset \subset \sigma_i, i = 1, \dots, d$ (inputs are not constant).
- For $B \subset A$, $\sigma_B \subset \sigma_A$ (inputs add information).
- For every $A, B \in \mathcal{P}_D, A \neq B$,

$$\sigma_A \cap \sigma_B = \sigma_{A \cap B}.$$

Output space

Consequences of Assumption 1 :

☞ $\mathbb{L}^2(\sigma_\emptyset) \subset \mathbb{L}^2(\sigma_A)$, for every $A \in \mathcal{P}_D$

There are non-constant random variables in the Lebesgue spaces

☞ For $B \subset A$, $\mathbb{L}^2(\sigma_B) \subset \mathbb{L}^2(\sigma_A)$

There are functions of X_A that are not functions of X_B

☞ For any $A, B \in \mathcal{P}_D$,

$$\mathbb{L}^2(\sigma_A) \cap \mathbb{L}^2(\sigma_B) = \mathbb{L}^2(\sigma_{A \cap B})$$

The functions of X_A and X_B are in fact functions of $X_{A \cap B}$

Proposition . Suppose that Assumption 1 hold. Then, for any $A, B \in \mathcal{P}_D$ such that $A \cap B \notin \{A, B\}$, **there is no mapping T such that**

$$X_B = T(X_A) \text{ a.e.}$$

☞ Hence the name “**non-perfect functional dependence**”

Minimal angle, maximal correlation

Dixmier's angle : the **maximal value** the inner product can take **between the elements of two closed subspaces** of a Hilbert space

Definition (*Dixmier's angle* (Dixmier 1949)). Let M, N be **closed** subspaces of a Hilbert space H . The cosine of Dixmier's angle between M and N is defined as

$$c_0(M, N) := \sup \{ |\langle x, y \rangle| : x \in M, \|x\| \leq 1, \quad y \in N, \|y\| \leq 1 \}.$$

☞ Analogous to the **maximal correlation** in probability theory (Koyak 1987), as a dependence measure between **random elements**.

Definition (*Maximal correlation* (Gebelein 1941)). Let X, Y be two **random elements**. The **maximal correlation** between X and Y is

$$\rho_0(X, Y) := c_0(\mathbb{L}_0^2(\sigma_X), \mathbb{L}_0^2(\sigma_Y))$$

Remark . The independence relation from the previous slide can be written as :

$$X \perp\!\!\!\perp Y \iff c_0(\mathbb{L}_0^2(\sigma_X), \mathbb{L}_0^2(\sigma_Y)) = 0.$$

Maximal partial correlation

Friedrichs' angle : Restriction to the **elements orthogonal to the intersection of the subspaces**

Definition (*Friedrich's angle* (Friedrichs 1937)). The cosine of Friedrichs' angle is defined as

$$c(M, N) := \sup \left\{ |\langle x, y \rangle| : \begin{cases} x \in M \cap (M \cap N)^\perp, \|x\| \leq 1 \\ y \in N \cap (M \cap N)^\perp, \|y\| \leq 1 \end{cases} \right\},$$

where the orthogonal complement is taken w.r.t. to H .

👉 Analogous to the **maximal partial dependence** between random elements (Bryc 1984, 1996; Dauxois, Nkiet et Romain 2004).

Definition (*Maximal partial correlation*). The **maximal partial correlation** between X and Y is

$$\rho^*(X, Y) := c(\mathbb{L}^2(\sigma_X), \mathbb{L}^2(\sigma_Y))$$

Remark . It is closely related to the **commutativity of conditional expectations**.

$$c(\mathbb{L}^2(\sigma_X), \mathbb{L}^2(\sigma_Y)) = 0 \iff \mathbb{E}[\mathbb{E}[\cdot | X] | Y] = \mathbb{E}[\mathbb{E}[\cdot | Y] | X]$$

Feshchenko matrix

☞ For every $A \in \mathcal{P}_D$, $\mathbb{L}^2(\sigma_A)$ contains **the functions of** X_A

☞ Dixmier's and Friedrichs' angles to **pairwise** control the inner products in these spaces

Intuition : A **generalized precision matrix** to control the *global magnitude of all the angles*

Definition (*Maximal coalitional precision matrix*).

Let Δ be the $(2^d \times 2^d)$, symmetric **set-indexed** matrix, defined element-wise, $\forall A, B \in \mathcal{P}_D$ as

$$\Delta_{AB} = \begin{cases} 1 & \text{if } A = B; \\ -c(\mathbb{L}^2(\sigma_A), \mathbb{L}^2(\sigma_B)) & \text{otherwise.} \end{cases}$$

We use **Friedrichs' angles** (partial correlation), hence the **precision** part

☞ These matrices closely resemble the ones used by **Feshchenko (2020)** to study the **closedness of an arbitrary sum of closed subspaces** of a Hilbert space

We're calling them “Feshchenko matrices”.

Non-degenerate stochastic dependence

But why are Feshchenko matrices interesting ?

Proposition . Suppose that Assumption 1 hold. Then,

$$\Delta = I_{2d} \iff X \text{ is mutually independent.}$$

☞ X is valued in a product of Polish spaces, with **an arbitrary law**

Assumption 2 (*Non-degenerate stochastic dependence*).

The Feshchenko matrix Δ of the inputs X is definite-positive.

It's a **restriction on the inner product of $\mathbb{L}^2(\sigma_X)$** $\implies$ **A restriction on the law of X**

☞ Hence the *stochastic dependence* (in opposition to *functional dependence*).

Generalized Hoeffding decomposition

Theorem .

Under Assumptions 1 and 2, for every $A \in \mathcal{P}_D$, one has that

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

where $V_\emptyset = \mathbb{L}^2(\sigma_\emptyset)$, and

$$V_B = \left[\bigoplus_{C \in \mathcal{P}_B, C \neq B} V_C \right]^{\perp_B},$$

where $\perp_B$ denotes the orthogonal complement in $\mathbb{L}^2(\sigma_B)$.

Intuition of the proof :

Inductive functional centering

Intuition of the proof : One input

One input :

1. Let $i \in D$, and **fix** $\mathbb{L}^2(\sigma_i)$ **as the ambient space**
2. We have that $V_\emptyset := \mathbb{L}^2(\sigma_\emptyset)$ **is a closed subspace of** $\mathbb{L}^2(\sigma_i)$
(it is **complemented**)
3. Denote $V_i = [V_\emptyset]^{\perp_i}$, **the orthogonal complement of $V_\emptyset$ in $\mathbb{L}^2(\sigma_i)$**
4. One has that $\mathbb{L}^2(\sigma_i) = V_\emptyset \oplus V_i$

We just showed that any $f(X_i) \in \mathbb{L}^2(\sigma_i)$ can be written as

$$f(X_i) = \underbrace{\mathbb{E}[f(X_i)]}_{\in V_\emptyset} + \underbrace{\mathbb{E}[f(X_i) - \mathbb{E}[f(X_i)]]}_{\in V_i = \mathbb{L}_0^2(\sigma_i)}$$

And note that $\mathbb{L}^2(\sigma_i) = V_\emptyset \oplus V_i$ hold for any $i \in D$ (induction)

Intuition of the proof : Two inputs

Two inputs :

1. Let $i, j \in D$, and **fix** $\mathbb{L}^2(\sigma_{ij})$ **as the ambient space**
2. **Assumptions 1 and 2 imply that $\mathbb{L}^2(\sigma_i) + \mathbb{L}^2(\sigma_j)$ is closed in $\mathbb{L}^2(\sigma_{ij})$**
(it is **complemented**)
3. Notice **(previous step)** that $\mathbb{L}^2(\sigma_i) + \mathbb{L}^2(\sigma_j) = V_\emptyset + V_i + V_j$
4. Denote $V_{ij} = [V_\emptyset + V_i + V_j]^\perp$, **the orthogonal complement in $\mathbb{L}^2(\sigma_{ij})$**
5. We thus have that $\mathbb{L}^2(\sigma_{ij}) = V_\emptyset + V_i + V_j + V_{ij}$

And note that the decomposition hold for any pair $i, j \in D$

We “centered” a bivariate function from its “univariate and constant parts”

We continue the induction up to d inputs.

Orthocanonical decomposition

Corollary (*Orthocanonical decomposition*).

Suppose that Assumptions 1 and 2 hold.

Then, any random variable $G(X) \in \mathbb{L}^2(\sigma_X)$ can be **uniquely decomposed** as

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

where each $G_A(X_A) \in V_A$.

The term “**orthocanonical**” comes from **the choice** of the orthogonal complement in the “centering process”

The subspaces V_A contain **random variables that are functions of exactly X_A**

If one of their element can be expressed with fewer inputs, it is necessarily equal to 0

Is it possible to characterize the *representants* $G_A(X_A)$?

Projectors

☞ Let H be a Hilbert space, and let $P : H \rightarrow H$ be an operator

☞ $\text{Ran}(P)$ is the range of P , and $\text{Ker}(P)$ is its nullspace

If P is an idempotent ($P \circ P = P$), linear, and bounded operator, it is then called the **oblique projector onto $\text{Ran}(P)$ parallel to $\text{Ker}(P)$** and

$$H = \text{Ran}(P) \oplus \text{Ker}(P)$$

Conversely, suppose that

$$H = M \oplus N$$

then there exists a projector P such that $\text{Ran}(P) = M$ and $\text{Ker}(P) = N$ (Galántai 2004)

☞ P is called the **canonical oblique projector** w.r.t. to this direct-sum decomposition of H

If in addition $N = M^\perp$ (P is self-adjoint), then P is called the **orthogonal projector** onto M

☞ The orthogonal projector onto $\mathbb{L}^2(\sigma_A)$ is the conditional expectation $\mathbb{E}[\cdot \mid \sigma_A]$

Orthocanonical projectors

From the direct-sum decomposition of $\mathbb{L}^2(\sigma_X)$:

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

Oblique projection onto V_A

The operator

$$Q_A : \mathbb{L}^2(\sigma_X) \rightarrow \mathbb{L}^2(\sigma_X), \quad G(X) \mapsto G_A(X_A).$$

Q_A is the (canonical) **oblique projection** with

$$\text{Ran}(Q_A) = V_A, \text{ and } \text{Ker}(Q_A) = \bigoplus_{B \in \mathcal{P}_D : B \neq A} V_B$$

Orthogonal projections onto V_A

The projector

$$P_A : \mathbb{L}^2(\sigma_X) \rightarrow \mathbb{L}^2(\sigma_X), \text{ with } \text{Ran}(P_A) = V_A \text{ and } \text{Ker}(P_A) = [V_A]^\perp$$

is the **orthogonal projection** onto V_A .

Oblique and orthogonal projections

In fact,

Proposition .

Under Assumptions 1 and 2,

$$P_A(G(X)) = Q_A(G(X)) \text{ a.s. , } \forall A \in \mathcal{P}_D \iff X \text{ is mutually independent.}$$

This comes from the fact that **the subspaces V_A are all pairwise orthogonal if and only if the inputs are mutually independent**

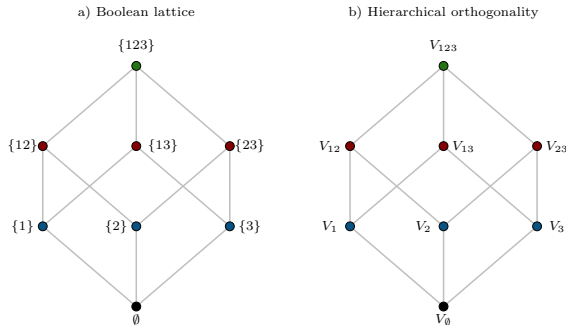
But, under Assumptions 1 and 2, they may not be all orthogonal

There is a more visual way to illustrate that

Boolean lattice and hierarchical orthogonality

Our decomposition is **over the power-set $\mathcal{P}_D$** , **and this is not trivial**

☞ Endowed with the **binary relation $\subseteq$** , $(\mathcal{P}_D, \subseteq)$ forms **a Boolean lattice**



The subspaces $\{V_A\}_{A \in \mathcal{P}_D}$ are **hierarchically orthogonal by design**

☞ They form the same algebraic structure **w.r.t. to $\perp$**

More projectors

Recall that :

- Q_A is the **canonical oblique projection** onto V_A
- P_A is the **orthogonal projection** onto V_A

But we're more familiar with projections onto $\mathbb{L}^2(\sigma_A)$...

☞ Conditional expectation operators, for example

(Canonical) oblique projection onto $\mathbb{L}^2(\sigma_A)$:

$$\mathbb{M}_A : \mathbb{L}^2(\sigma_X) \rightarrow \mathbb{L}^2(\sigma_X), \quad G(X) \mapsto \sum_{B \in \mathcal{P}_A} G_B(X_B)$$

Orthogonal projection onto $\mathbb{L}^2(\sigma_A)$:

$$\mathbb{E}_A : \mathbb{L}^2(\sigma_X) \rightarrow \mathbb{L}^2(\sigma_X), \quad \text{with } \text{Ran}(\mathbb{E}_A) = \mathbb{L}^2(\sigma_A) \text{ and } \text{Ker}(\mathbb{E}_A) = \mathbb{L}^2(\sigma_A)^\perp,$$

a.k.a **the conditional expectation w.r.t. to X_A** (i.e., $\mathbb{E}[\cdot | X_A]$).

Can we characterize Q_A w.r.t. $\mathbb{M}_A$?

Generalized Möbius inversion

Because $(\mathcal{P}_D, \subseteq)$ forms a **Boolean lattice**, yes!

Corollary (Möbius inversion on power-sets (Rota 1964)).

For any two set functions :

$$f : \mathcal{P}_D \rightarrow \mathbb{A}, \quad g : \mathcal{P}_D \rightarrow \mathbb{A},$$

valued in an abelian group $\mathbb{A}$, the following equivalence holds :

$$f(A) = \sum_{B \in \mathcal{P}_A} g(B), \quad \forall A \in \mathcal{P}_D \quad \Longleftrightarrow \quad g(A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} f(B), \quad \forall A \in \mathcal{P}_D.$$

☞ Analogous to the *inclusion-exclusion principle*

In our case, we have, **by definition of the oblique projection onto $\mathbb{L}^2(\sigma_A)$** , that

$$\mathbb{M}_A(G(X)) = \sum_{B \in \mathcal{P}_A} G_B(X_B), \quad \forall A \in \mathcal{P}_D,$$

which is equivalent to

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

Generalized Hoeffding decomposition

Hoeffding (1948) found that **for mutually independent inputs** :

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

Under Assumptions 1 and 2, we have that :

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

In addition :

Proposition .

Under Assumptions 1 and 2,

$$\mathbb{M}_A [G(X)] = \mathbb{E}_A [G(X)] \text{ a.s. , } \forall A \in \mathcal{P}_D \iff X \text{ is mutually independent.}$$

Our approach generalizes Hoeffding's original 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**

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

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

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

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

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 et Annoni (2015).

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

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

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

Proposition . Under the two assumptions,

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

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 et 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

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

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

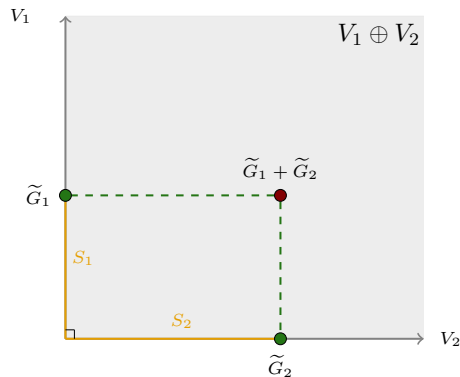
Definition (*Correlative effects*).

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

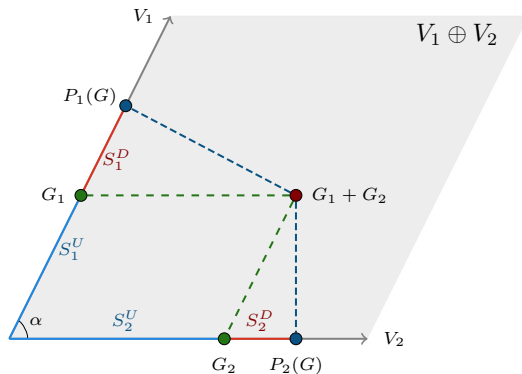
$$S_A^C = \text{Cov} \left(G_A(X_A), \sum_{B \in \mathcal{P}_D: B \neq A} G_B(X_B) \right).$$

Variance decomposition : Intuition

Pure interaction effects



Structural and dependence effects



👉 Maybe a good candidate to solve our second issue!