Hoeffding Decomposition, Revisited

And a generalization to dependent inputs

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

SIAM Conference on Uncertainty Quantification
Society for Industrial and Applied Mathematics - Trieste, Italy.
March 1st, 2024

Marouane Il Idrissi[1,2,3], Nicolas Bousquet[1,3], Fabrice Gamboa[2], Bertrand Iooss[1,2,3], Jean-Michel Loubes[2]

Some context

Does Hoeffding's functional decomposition hold when the inputs are not mutually independent?

Classical Hoeffding's decomposition: **Unique** decomposition $G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A)$ for any square-integrable $G(X)$, where the inputs X are **mutually independent**.

Does Hoeffding's functional decomposition hold when the inputs are not mutually independent?

Classical Hoeffding's decomposition: Unique decomposition $G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A)$ for any square-integrable $G(X)$, where the inputs X are **mutually independent**.

Yes, under **two reasonable assumptions on the inputs**:

- **Non-perfect functional dependence**.
- **Non-perfect stochastic dependence**.

Does Hoeffding's functional decomposition hold when the inputs are not mutually independent?

Classical Hoeffding's decomposition: **Unique** decomposition $G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A)$ for any square-integrable $G(X)$, where the inputs X are **mutually independent**.

Yes, under **two reasonable assumptions on the inputs**:

- **Non-perfect functional dependence**.
- **Non-perfect stochastic dependence**.

However... Achieving this result requires **an unusual methodological journey**.

Does Hoeffding's functional decomposition hold when the inputs are not mutually independent?

Classical Hoeffding's decomposition: **Unique** decomposition $G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A)$ for any square-integrable $G(X)$, where the inputs X are **mutually independent**.

Yes, under **two reasonable assumptions on the inputs**:

- **Non-perfect functional dependence**.
- **Non-perfect stochastic dependence**.

However... Achieving this result requires **an unusual methodological journey**.

In this talk: Mix the fields of **probability theory** and **functional analysis**, to **generalize Hoeffding's decomposition to dependent inputs**.

More context

We're not the first to have worked on this generalization.

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

More context

We're not the first to have worked on this generalization.

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

In **Chastaing, Gamboa, and Prieur (2012)**, the authors approached the problem by **considering subspaces of the Lebesgue space** $\mathbb{L}^2$.

They showed that the generalized decomposition hold, **but under fairly restrictive assumptions on the inputs**.

More context

We're not the first to have worked on this generalization.

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

In **Chastaing, Gamboa, and Prieur (2012)**, the authors approached the problem by **considering subspaces of the Lebesgue space** $\mathbb{L}^2$.

They showed that the generalized decomposition hold, **but under fairly restrictive assumptions on the inputs**.

Our approach: Understand the relationships between these subspaces of $\mathbb{L}^2$ when the inputs are **not mutually independent**.

Framework and notations

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

Framework and notations

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

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

> **Remark .** The **random inputs** are not necessarily $\mathbb{R}^d$-valued, and do not necessarily have a **density**.

Framework and notations

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

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

> **Remark** . The **random inputs** are not necessarily $\mathbb{R}^d$-valued, and do not necessarily have a **density**.

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

Framework and notations

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

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

> **Remark .** The **random inputs** are not necessarily $\mathbb{R}^d$-valued, and do not necessarily have a **density**.

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

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

Framework and notations

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

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

Remark . The **random inputs** are not necessarily $\mathbb{R}^d$-valued, and do not necessarily have a **density**.

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

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

☞ σ_X is the σ-algebra generated by X, and σ_A the one generated by X_A.

Lemma *(Doob-Dynkin Lemma).* If an $\mathbb{R}$-valued random variable Y is σ_X-measurable, then there exists some function $f : E \to \mathbb{R}$ such that $Y = G(X)$ a.s.

Framework and notations

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

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

Remark . The **random inputs** are not necessarily $\mathbb{R}^d$-valued, and do not necessarily have a **density**.

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

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

☞ σ_X is the σ-algebra generated by X, and σ_A the one generated by X_A.

Lemma *(Doob-Dynkin Lemma)*. If an $\mathbb{R}$-valued random variable Y is σ_X-measurable, then there exists some function $f : E \to \mathbb{R}$ such that $Y = G(X)$ a.s.

☞ $\sigma_\emptyset$ is the $\mathbb{P}$-trivial σ-algebra (it contains **every** null event of $\mathcal{F}$).

Proposition *(Resnick 2014)*. If an $\mathbb{R}$-valued random variable is $\sigma_\emptyset$-measurable, it is **constant a.e.**

Lebesgue spaces

Definition *(Lebesgue space)*. Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. Denote by $\mathbb{L}^2(\mathcal{G})$ the **Lebesgue space** containing every **real-valued random variables**, which are $\mathcal{G}$-measurable, and, if $Y \in \mathbb{L}^2(\sigma_{\mathcal{G}})$

$$\mathbb{E}\left[Y^2\right] = \int_\Omega Y(\omega)^2 d\mathbb{P}(\omega) < \infty.$$

☞ $\mathbb{L}^2(\sigma_X)$ is the space of **random outputs**: it contains random variables which are **functions of** X.

Lebesgue spaces

Definition *(Lebesgue space)*. Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. Denote by $\mathbb{L}^2(\mathcal{G})$ the **Lebesgue space** containing every **real-valued random variables**, which are $\mathcal{G}$-measurable, and, if $Y \in \mathbb{L}^2(\sigma_{\mathcal{G}})$

$$\mathbb{E}\left[Y^2\right] = \int_\Omega Y(\omega)^2 d\mathbb{P}(\omega) < \infty.$$

☞ $\mathbb{L}^2(\sigma_X)$ is the space of **random outputs**: it contains random variables which are **functions of** X.

☞ $\mathbb{L}^2(\sigma_X)$ is an (infinite-dimensional) Hilbert space, with inner product

$$\langle f(X), g(X)\rangle = \mathbb{E}\left[f(X)g(X)\right] = \int_E f(x)g(x)dP_X(x) = \int_\Omega f(X(\omega))g(X(\omega))d\mathbb{P}(\omega).$$

Lebesgue spaces

Definition *(Lebesgue space)*. Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. Denote by $\mathbb{L}^2(\mathcal{G})$ the **Lebesgue space** containing every **real-valued random variables**, which are $\mathcal{G}$-measurable, and, if $Y \in \mathbb{L}^2(\sigma_{\mathcal{G}})$

$$\mathbb{E}\left[Y^2\right] = \int_\Omega Y(\omega)^2 d\mathbb{P}(\omega) < \infty.$$

☞ $\mathbb{L}^2(\sigma_X)$ is the space of **random outputs**: it contains random variables which are **functions of** X.

☞ $\mathbb{L}^2(\sigma_X)$ is an (infinite-dimensional) Hilbert space, with inner product

$$\langle f(X), g(X)\rangle = \mathbb{E}[f(X)g(X)] = \int_E f(x)g(x)dP_X(x) = \int_\Omega f(X(\omega))g(X(\omega))d\mathbb{P}(\omega).$$

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

☞ $\left\{\mathbb{L}^2(\sigma_A)\right\}_{A\in\mathcal{P}_D}$ are **closed subspace of** $\mathbb{L}^2(\sigma_X)$.

Lebesgue spaces

Definition *(Lebesgue space)*. Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. Denote by $\mathbb{L}^2(\mathcal{G})$ the **Lebesgue space** containing every **real-valued random variables**, which are $\mathcal{G}$-measurable, and, if $Y \in \mathbb{L}^2(\sigma_{\mathcal{G}})$

$$\mathbb{E}\left[Y^2\right] = \int_\Omega Y(\omega)^2 d\mathbb{P}(\omega) < \infty.$$

☞ $\mathbb{L}^2(\sigma_X)$ is the space of **random outputs**: it contains random variables which are **functions of** X.

☞ $\mathbb{L}^2(\sigma_X)$ is an (infinite-dimensional) Hilbert space, with inner product

$$\langle f(X), g(X)\rangle = \mathbb{E}[f(X)g(X)] = \int_E f(x)g(x)dP_X(x) = \int_\Omega f(X(\omega))g(X(\omega))d\mathbb{P}(\omega).$$

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

☞ $\left\{\mathbb{L}^2(\sigma_A)\right\}_{A\in\mathcal{P}_D}$ are **closed subspace of** $\mathbb{L}^2(\sigma_X)$.

How are these subspaces related to each other?

Functional dependence

☞ First, we need to control **the functional dependence between the inputs**.
Otherwise, the subspaces $\mathbb{L}^2(\sigma_A)$ **cannot be distinct**.

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

Suppose that:

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

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

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**

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

In other words, **if Assumption 1 holds, then <u>the inputs cannot be functions of each other</u>**.

Angles between subspaces of Hilbert spaces

☞ **Conditions on the inner product** affect **the law of** X.

Angles between subspaces of Hilbert spaces

☞ **Conditions on the inner product** affect **the law of** X.

Theorem *(Malliavin 1995, Chapter 3)*. Let X and Y be two random elements. Then:

$$X \perp\!\!\!\perp Y \iff \forall f(X) \in \mathbb{L}^2_0(\sigma_X), \ \forall g(Y) \in \mathbb{L}^2_0(\sigma_Y), \quad \mathbb{E}[f(X), g(Y)] = 0,$$

$\mathbb{L}^2_0$ are the Lebesgue subspaces of **centered random variables**.

Angles between subspaces of Hilbert spaces

☞ **Conditions on the inner product** affect **the law of** X.

Theorem *(Malliavin 1995, Chapter 3)*. Let X and Y be two random elements. Then:

$$X \perp\!\!\!\perp Y \iff \forall f(X) \in \mathbb{L}^2_0(\sigma_X), \forall g(Y) \in \mathbb{L}^2_0(\sigma_Y), \quad \mathbb{E}[f(X), g(Y)] = 0,$$

$\mathbb{L}^2_0$ are the Lebesgue subspaces of **centered random variables**.

<u>**Intuition:**</u> Control the **dependence structure** by controlling the **magnitude of the inner product between the subspaces** $\{\mathbb{L}^2(\sigma_A)\}_{A \in \mathcal{P}_D}$?

Angles between subspaces of Hilbert spaces

☞ **Conditions on the inner product** affect **the law of** X.

Theorem *(Malliavin 1995, Chapter 3).* Let X and Y be two random elements. Then:

$$X \perp\!\!\!\perp Y \iff \forall f(X) \in \mathbb{L}_0^2(\sigma_X), \ \forall g(Y) \in \mathbb{L}_0^2(\sigma_Y), \quad \mathbb{E}[f(X), g(Y)] = 0,$$

$\mathbb{L}_0^2$ are the Lebesgue subspaces of **centered random variables**.

Intuition: Control the **dependence structure** by controlling the **magnitude of the inner product between the subspaces** $\{\mathbb{L}^2(\sigma_A)\}_{A \in \mathcal{P}_D}$?

To do that:

Definition *(Friedrichs' angle (Friedrichs 1937)).* Let M, N be **closed** subspaces of a Hilbert space H.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.

Assumptions

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

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

☞ Δ can be seen as a "generalization" of **precision matrices**.

☞ Its name comes from the (amazing) work of **Ivan Feshchenko (2020)**.

Why is this matrix interesting ?

Assumptions

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

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

☞ Δ can be seen as a "generalization" of **precision matrices**.

☞ Its name comes from the (amazing) work of **Ivan Feshchenko (2020)**.

Why is this matrix interesting ?

Proposition .

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

Assumptions

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

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

☞ Δ can be seen as a "generalization" of **precision matrices**.

☞ Its name comes from the (amazing) work of **Ivan Feshchenko (2020)**.

Why is this matrix interesting ?

Proposition .

$$\Delta = I_{2^d} \quad \Longleftrightarrow \quad X \text{ is mutually independent.}$$

Our second assumption

Assumption 2 *(Non-degenerate stochastic dependence)*. Δ is definite-positive.

Direct-sum decompositions

Definition *(Direct-sum decomposition)*. Let $W_1, \ldots, W_d$ be vector subspaces of a vector space W. W is said to admit a **direct-sum decomposition**, denoted:

$$W = \bigoplus_{i=1}^{d} W_i,$$

if any element $w \in W$ can be written **uniquely** as a sum of elements of the W_i.

Remark . Hilbert spaces are vector spaces.

Direct-sum decompositions

Definition *(Direct-sum decomposition)*. Let $W_1, \ldots, W_d$ be vector subspaces of a vector space W. W is said to admit a **direct-sum decomposition**, denoted:

$$W = \bigoplus_{i=1}^{d} W_i,$$

if any element $w \in W$ can be written **uniquely** as a sum of elements of the W_i.

Remark . Hilbert spaces are vector spaces.

Hence, a Hoeffding-like decomposition of a **black-box model** entails **finding a direct-sum decomposition for** $\mathbb{L}^2(\sigma_X)$, i.e., writing

$$\mathbb{L}^2(\sigma_X) = \bigoplus_{A \in \mathcal{P}_D} V_A,$$

where the V_A are subspaces that we need to characterize.

Main result

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

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

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

Corollary *(Orthocanonical decomposition)*. Under Assumptions 1 and 2, any $G(X) \in \mathbb{L}^2(\sigma_X)$ can be **uniquely decomposed** as

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

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

Main result

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

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

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

Corollary *(Orthocanonical decomposition)*. Under Assumptions 1 and 2, any $G(X) \in \mathbb{L}^2(\sigma_X)$ can be **uniquely decomposed** as

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

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

Intuition of the proof: **Inductive generalized centering**.

Intuition behind the result: 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** $\mathbb{L}^2(\sigma_i)$ (thus 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 $\mathbb{L}^2(\sigma_i) = V_\emptyset \oplus V_i$.
5. Since $V_\emptyset$ only contains constants, $V_i = \mathbb{L}_0^2(\sigma_i)$.

In other words, we just showed that any $f(X_i) \in \mathbb{L}^2(\sigma_i)$ can be written as

$$f(X_i) = \underbrace{\mathbb{E}[f(X_i)]}_{\in V_\emptyset} + \underbrace{\mathbb{E}[f(X_i) - \mathbb{E}[f(X_i)]]}_{\in V_i}.$$

Intuition behind the result: 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** $\mathbb{L}^2(\sigma_i)$ (thus 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 $\mathbb{L}^2(\sigma_i) = V_\emptyset \oplus V_i$.
5. Since $V_\emptyset$ only contains constants, $V_i = \mathbb{L}^2_0(\sigma_i)$.

In other words, we just showed that any $f(X_i) \in \mathbb{L}^2(\sigma_i)$ can be written as

$$f(X_i) = \underbrace{\mathbb{E}[f(X_i)]}_{\in V_\emptyset} + \underbrace{\mathbb{E}[f(X_i) - \mathbb{E}[f(X_i)]]}_{\in V_i}.$$

And note that we can do this for any $i \in D$.

Intuition behind the result: Two inputs

Two inputs:

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

Intuition behind the result: Two inputs

Two inputs:

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

And note that we can do this for any pair $i, j \in D$**.**

In essence, we **"centered" a bivariate function from its "univariate and constant parts".**

Intuition behind the result: Two inputs

Two inputs:

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

And note that we can do this for any pair $i, j \in D$.

In essence, we **"centered" a bivariate function from its "univariate and constant parts".**

And we can continue the same induction up to d inputs.

Projectors

Recall that for any $G(X) \in \mathbb{L}^2(\sigma_X)$, we have

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

Projectors

Recall that for any $G(X) \in \mathbb{L}^2(\sigma_X)$, we have

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

Oblique projections

Denote the operator

$$Q_A : \mathbb{L}^2(\sigma_X) \to \mathbb{L}^2(\sigma_X), \text{ such that } \quad Q_A(G(X)) = G_A(X_A).$$

Q_A is the (canonical) **oblique projection** onto V_A, parallel to $\bigoplus_{B \in \mathcal{P}_D : B \neq A} V_A$.

Projectors

Recall that for any $G(X) \in \mathbb{L}^2(\sigma_X)$, we have

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

Oblique projections

Denote the operator

$$Q_A : \mathbb{L}^2(\sigma_X) \to \mathbb{L}^2(\sigma_X), \text{ such that } \quad Q_A(G(X)) = G_A(X_A).$$

Q_A is the (canonical) **oblique projection** onto V_A, parallel to $\bigoplus_{B \in \mathcal{P}_D : B \neq A} V_A$.

Orthogonal projections

Denote the projector

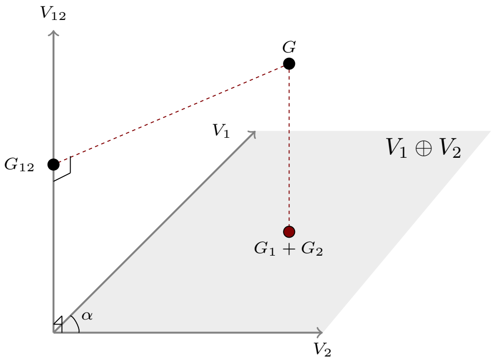
$$P_A : \mathbb{L}^2(\sigma_X) \to \mathbb{L}^2(\sigma_X), \text{ such that } \quad \mathrm{Ran}(P_A) = V_A, \mathrm{Ker}(P_A) = [V_A]^{\perp}.$$

the **orthogonal projection** onto V_A.

Illustration : $\mathbb{L}^2_0(\sigma_{12})$

Hence, for any $G(X) \in \mathbb{L}^2(\sigma_X)$, one has that, $\forall A \in \mathcal{P}_D$

$$G_A(X_A) = Q_A(G(X)).$$



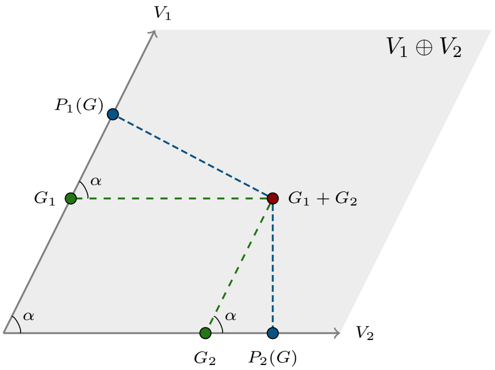
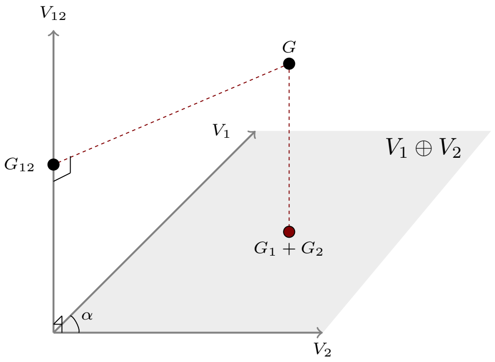
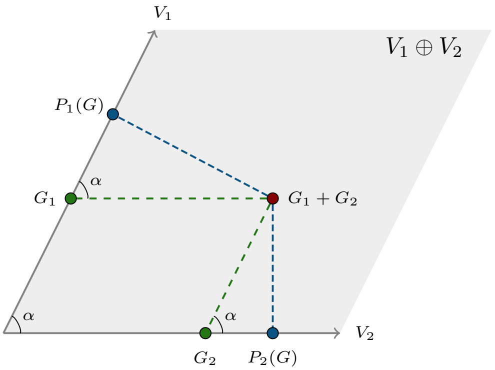


Illustration : $\mathbb{L}_0^2(\sigma_{12})$

Hence, for any $G(X) \in \mathbb{L}^2(\sigma_X)$, one has that, $\forall A \in \mathcal{P}_D$

$$G_A(X_A) = Q_A(G(X)).$$





The oblique projection Q_A usually differ from the oblique projections P_A

Variance decomposition

We propose two complementary approaches for decomposing $\mathbb{V}(G(X))$.

Variance decomposition

We propose two complementary approaches for decomposing $\mathbb{V}(G(X))$.

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

Variance decomposition

We propose two complementary approaches for decomposing $\mathbb{V}(G(X))$.

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

Canonical 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

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

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

Recall that **usually, $P_A(G(X))$ and $Q_A(G(X))$ differ**. In fact,

Proposition . Under Assumptions 1 and 2,

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

Which motivates the definition of dependence effects.

Organic variance decomposition: Dependence effects

Recall that **usually,** $P_A(G(X))$ **and** $Q_A(G(X))$ **differ**. In fact,

Proposition . Under Assumptions 1 and 2,

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

Which motivates the definition of dependence effects.

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 Assumptions 1 and 2,

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

Organic variance decomposition: Dependence effects

Recall that **usually,** $P_A(G(X))$ **and** $Q_A(G(X))$ **differ**. In fact,

Proposition . Under Assumptions 1 and 2,

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

Which motivates the definition of dependence effects.

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 Assumptions 1 and 2,

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

<u>Open question:</u> What do they sum up to ?...

Organic variance decomposition: Dependence effects

Recall that **usually,** $P_A(G(X))$ **and** $Q_A(G(X))$ **differ**. In fact,

Proposition . Under Assumptions 1 and 2,

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

Which motivates the definition of dependence effects.

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 Assumptions 1 and 2,

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

Open question: What do they sum up to ?...

Probably some interesting multivariate dependence measure!

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**

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

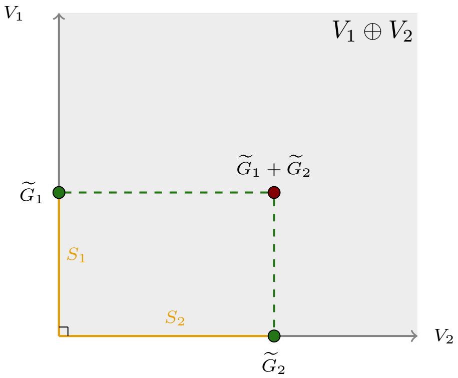
The **correlative effects** represent the part of variance that is due to the correlation between the $G_A(X_A)$.

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

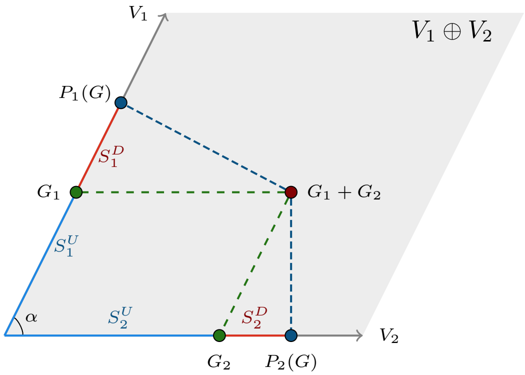
$$S_A^C = \mathrm{Cov}\left(G_A(X_A), \sum_{B \in \mathcal{P}_D : B \neq A} G_B(X_B)\right).$$

Variance decomposition: Intuition

Pure interaction effects



Structural and dependence effects



Conclusion

Main take-aways:

- Hoeffding-like decomposition of function with dependent inputs is **achievable under reasonable assumptions.**
- Mixing **probability, functional analysis (and combinatorics)** lead to an **interesting framework for studying multivariate stochastic problems**.
- We can define **meaningful (i.e., intuitive) decompositions of quantities of interest**, which **intrinsically encompasses the dependence between the inputs**.
- We proposed candidates to separate and quantify **pure interaction** from **dependence effects**.

Perspective

Main challenge: Estimation.

- We haven't found an **off-the-shelf method** to **estimate these oblique projections**...

Perspective

Main challenge: Estimation.

- We haven't found an **off-the-shelf method** to **estimate these oblique projections**...
- **But we have a lot of promising ideas :)**

Perspective

Main challenge: Estimation.

- We haven't found an **off-the-shelf method** to **estimate these oblique projections**...
- **But we have a lot of promising ideas :)**

A few perspectives:

- Links with already-established results (e.g., on copulas).
- Non $\mathbb{R}$-valued output.
- Many methodological questions that seemed unreachable so far, but appear approachable using this framework.

Checkout our pre-print!

To go further + illustrations (HAL/ResearchGate/arXiv)

Understanding black-box models with dependent inputs through a generalization of Hoeffding's decomposition

Marouane Il Idrissi[a,b,c,e], Nicolas Bousquet[a,b,d], Fabrice Gamboa[c], Bertrand Iooss[a,b,c], Jean-Michel Loubes[c]

References i

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* [in en]. Philadelphia, PA: Society for Industrial / Applied Mathematics, January. ISBN: 978-1-61197-668-7 978-1-61197-669-4, accessed November 22, 2022. https://doi.org/10.1137/1.9781611976694. https://epubs.siam.org/doi/book/10.1137/1.9781611976694.

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.

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.

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.

Joe, H. 1997. *Multivariate Models and Multivariate Dependence Concepts.* New York: Chapman / Hall/CRC. ISBN: 978-0-367-80389-6. https://doi.org/10.1201/9780367803896.

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.

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.

References iii

Malliavin, P. 1995. *Integration and Probability.* Vol. 157. Graduate Texts in Mathematics. New York, NY: Springer. ISBN: 978-1-4612-8694-3. https://doi.org/10.1007/978-1-4612-4202-4. http://link.springer.com/10.1007/978-1-4612-4202-4.

Mara, T. A., S. Tarantola, and P. Annoni. 2015. "Non-parametric methods for global sensitivity analysis of model output with dependent inputs." *Environmental Modelling & Software* 72:173–183. ISSN: 1364-8152. https://doi.org/10.1016/j.envsoft.2015.07.010. https://www.sciencedirect.com/science/article/pii/S1364815215300153.

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.

Resnick, S. I. 2014. *A Probability Path* [in en]. Boston, MA: Birkhäuser Boston. ISBN: 978-0-8176-8408-2 978-0-8176-8409-9. https://doi.org/10.1007/978-0-8176-8409-9. http://link.springer.com/10.1007/978-0-8176-8409-9.

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.

Thank you for your attention!

Any questions?

Example: Two Bernoulli inputs

Let $E = \{0,1\}^2$, and let $X = (X_1, X_2)$, where

$$X_1 \sim \mathcal{B}(q_1), \quad \text{and } X_2 \sim \mathcal{B}(q_2).$$

The joint law of X can be express using **three parameters**:

$$p_{00} = 1 - q_1 - q_2 + \rho, \quad p_{01} = q_2 - \rho, \quad p_{10} = q_1 - \rho, \quad p_{11} = \rho$$

where $p_{ij} = \mathbb{P}(\{X_1 = i\} \cap \{X_2 = j\})$.

Any function $G : \{0,1\}^2 \to \mathbb{R}$ can be expressed as the vector $G = (G_{00}, G_{01}, G_{10}, G_{11})^\top$.

Each value $G_{ij} = G(i,j)$, can be observed with probability p_{ij}.

Example: Two Bernoulli inputs

Let $E = \{0,1\}^2$, and let $X = (X_1, X_2)$, where

$$X_1 \sim \mathcal{B}(q_1), \quad \text{and } X_2 \sim \mathcal{B}(q_2).$$

The joint law of X can be express using **three parameters**:

$$p_{00} = 1 - q_1 - q_2 + \rho, \quad p_{01} = q_2 - \rho, \quad p_{10} = q_1 - \rho, \quad p_{11} = \rho$$

where $p_{ij} = \mathbb{P}(\{X_1 = i\} \cap \{X_2 = j\})$.

Any function $G : \{0,1\}^2 \to \mathbb{R}$ can be expressed as the vector $G = (G_{00}, G_{01}, G_{10}, G_{11})^\top$.

Each value $G_{ij} = G(i,j)$, can be observed with probability p_{ij}.

In this case, we can compute everything analytically.

It requires to solving 13 equations with 13 unknowns*.

*https://github.com/milidris/GeneralizedAnova

Feshchenko matrix and the Fréchet bounds

For the **Feshchenko matrix** Δ to be definite positive, one has that:

$$\max\left\{0, q_1 q_2 - \sqrt{q_1 q_2 (1-q_1)(1-q_2)}\right\} < \rho < \min\left\{1, q_1 q_2 - \sqrt{q_1 q_2 (1-q_1)(1-q_2)}\right\}.$$

However, the **classical Fréchet bounds for ρ for bivariate Bernoulli random variables** (Joe 1997, p.210) are equal to

$$\max\{0, q_1 + q_2 - 1\} \leq \rho \leq \min\{q1, q2\},$$

and are **more restrictive than the previous ones**.

ρ strictly contained in the Fréchet bounds $\implies$ Assumptions 1 and 2 hold.

Our decomposition hold for virtually any dependence structure between two Bernoullis.

More projectors

Recall that:

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

But what about projections onto the subspaces $\left\{\mathbb{L}^2(\sigma_A)\right\}_{A\in\mathcal{P}_D}$**?**

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

$$\begin{aligned}\mathbb{M}_A : \mathbb{L}^2(\sigma_X) &\to \mathbb{L}^2(\sigma_X)\\ G(X) &\mapsto \sum_{B\in\mathcal{P}_A} G_B(X_B)\end{aligned}$$

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

$$\mathbb{E}_A : \mathbb{L}^2(\sigma_X) \to \mathbb{L}^2(\sigma_X), \quad \text{such that } \mathrm{Ran}(\mathbb{E}_A) = \mathbb{L}^2(\sigma_A) \text{ and } \mathrm{Ker}(P_A) = \mathbb{L}^2(\sigma_A)^{\perp},$$

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

Is it possible to express the projections Q_A using $\mathbb{M}_A$?

Generalized Möbius inversion

Yes, because we're **working on the power-set** $\mathcal{P}_D$!

Corollary *(Möbius inversion on power-sets (Rota 1964)).* Let $D = \{1, \ldots, d\}$. For any two set functions:

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

where $\mathbb{A}$ is an **abelian group**, the following equivalence holds:

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

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

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

(This is what we call the **"model-centric"** approach)

Generalized Hoeffding decomposition

If the inputs are mutually independent, from Hoeffding (1948), we have that:

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

In our approach, 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,

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

Our approach actually generalizes Hoeffding's decomposition!