Towards more interpretable kernel-based sensitivity analysis

Gabriel Sarazin (DES/ISAS/DM2S/SGLS/LIAD)

- **Service de Génie Logiciel pour la Simulation**
- **Laboratoire d'Intelligence Artificielle et de sciences des Données**

Joint work with research partners from the SAMOURAI project:

- **CEA** : **Amandine Marrel** **(DES/IRESNE/DER/SESI/LEMS)**
- **CREST-ENSAI** : **Sébastien Da Veiga** **(Statistics Team)**
- **EDF R&D** : **Vincent Chabridon** **(PRISME Department)**

UQSay #68

January 25th, 2024 – Online seminar ☞ https://www.uqsay.org/2024/01/uqsay-68.html

About me…

- Hired as permanent **CEA research engineer** in **June 2023**.
  - ✓ Recruited to strengthen the **URANIE dev team** (currently upgrading my skills).
  - ✓ **5-year experience** in **uncertainty quantification (UQ)**:
    - **Areas of expertise:** sensitivity analysis & reliability assessment.
    - **Areas of interest:** kernel methods, stochastic modelling, copula theory…

Past positions

- **2012-2017** → Engineering student at **INSA Rennes** (Department of Applied Mathematics)

- **2017-2021** → PhD student at **ONERA Toulouse** (DTIS)
  - **Title:** *Reliability-oriented sensitivity analysis in presence of data-driven epistemic uncertainties.*
  - **Supervisors:** J. Morio (ONERA), A. Lagnoux (IMT), M. Balesdent (ONERA) & L. Brevault (ONERA).
  - **Keywords:** sensitivity analysis, rare-event probability estimation, extreme value theory, copula models…
  - **Applications:** buckling of a composite laminate plate + launch vehicle fallout in the atmosphere.

- **2021-2023** → Postdoctoral researcher at **CEA Cadarache** (DES/IRESNE/SESI/LEMS)
  - **Title:** *Surrogate modeling and optimization under uncertainty for high-dimensional problems.*
  - **Supervisors:** A. Marrel (CEA), S. Da Veiga (ENSAI) & V. Chabridon (EDF).
  - **Keywords:** sensitivity analysis, surrogate modelling, reproducing kernel theory, hypothesis testing.
  - **Application:** reliability assessment of nuclear power plants → study of accidental transients.

A few words on the SAMOURAI project…

- **4-year research project** launched in **March 2021** and funded by the **French National Research Agency**.

Simulation **A**nalytics and **M**etamodel-based solutions for **O**ptimization, **U**ncertainty and **R**eliability **A**nalys**I**s

☞ https://www.ifpenergiesnouvelles.fr/samourai

- ✓ **Industrial partners** → EDF R&D and Safran Tech
- ✓ **Public institution partners** → IFPEN and CEA
- ✓ **Academic partners** → Centrale Supélec, EMSE and Polytechnique Montréal

A few words on the SAMOURAI project…

- **4-year research project** launched in **March 2021** and funded by the **French National Research Agency**.

Simulation Analytics and Metamodel-based solutions for Optimization, Uncertainty and Reliability AnalysIs

☞ https://www.ifpenergiesnouvelles.fr/samourai

- The project is divided into **4 work packages**. Scientific coordination is ensured by Delphine Sinoquet (IFPEN).
  - ✓ **WP1: Metamodels for large-scale problems.**
    - **Investigators:** V. Chabridon (EDF), S. Da Veiga (ENSAI), A. Marrel (CEA) & B. Staber (Safran).
    - **Contributors:** R. Carpintero Perez (Safran), Y. Marnissi (Safran) & G. Sarazin (CEA).
  - ✓ **WP2: Enrichment strategies for RBI and RBDO.**
    - **Investigators:** J. Bect (Centrale Supélec) & E. Vasquez (Central Supélec).
    - **Contributors:** R. Abdelmalek-Lomenech (Centrale Supélec), V. Chabridon (EDF) & R. El Amri (IFPEN).
  - ✓ **WP3: Metamodels and optimization for mixed problems.**
    - **Investigators:** M. Keller (EDF) & R. Le Riche (EMSE).
    - **Contributors:** J. Pelamatti (EDF), B. Sow (EMSE) & S. Zannane (EDF).
  - ✓ **WP4: Dealing with hidden constraints.**
    - **Investigators:** S. Le Digabel (Polytechnique Montréal) & M. Munoz Zuniga (IFPEN).
    - **Contributors:** S. Jacquet (IFPEN) & M. Menz (IFPEN).

A few words on the SAMOURAI project…

- **4-year research project** launched in **March 2021** and funded by the **French National Research Agency**.

Simulation **A**nalytics and **M**etamodel-based solutions for **O**ptimization, **U**ncertainty and **R**eliability **A**nalys**I**s

☞ https://www.ifpenergiesnouvelles.fr/samourai

- As regards WP1, **4 research topics** were identified as priorities.
  - ✓ **Task 1.1:** Improve sparse Gaussian process (GP) regression and experiment modern kernel selection
  - ✓ **Task 1.2: Take advantage of state-of-the-art techniques in global sensitivity analysis (GSA)**
    - Better understand the mathematical foundings of the **HSIC-ANOVA decomposition**.
      - Investigate the existence of **explicit feature maps** for **Sobolev kernels**.
      - Establish connections between **feature functions** and **HSIC-ANOVA terms**.
    - Extend **HSIC-based independence test prodecures** to HSIC-ANOVA indices.
    - Compare numerically the information captured by **Sobol' indices** and **HSIC-ANOVA indices**.
    - Upgrade the **R package *sensitivity*** (especially the routines dedicated to **kernel-based GSA**).
  - ✓ **Task 1.3:** Make GP hyperparameter estimation more robust
  - ✓ **Task 1.4:** Extend and adapt all methodologies to (very) large databases

0. Introduction

GSA in support to metamodel construction

- In all four work packages, there is a need to construct **metamodels** for **high-dimensional design problems**.
  - Let $\boldsymbol{X} := [X_1, \dots, X_d]$ be a random vector with **independent** components ($d \approx 100$).
  - Let $Y := g(\boldsymbol{X})$ where $g: \mathcal{X}_1 \times \cdots \times \mathcal{X}_d \to \mathcal{Y}$ is a **computationally-expensive** simulation code.
  - $\boldsymbol{Z} = (\boldsymbol{X}, Y)$ is the **augmented vector** containing the input and output variables.

The design of experiments (DoE) consists of a number of input-output observations.

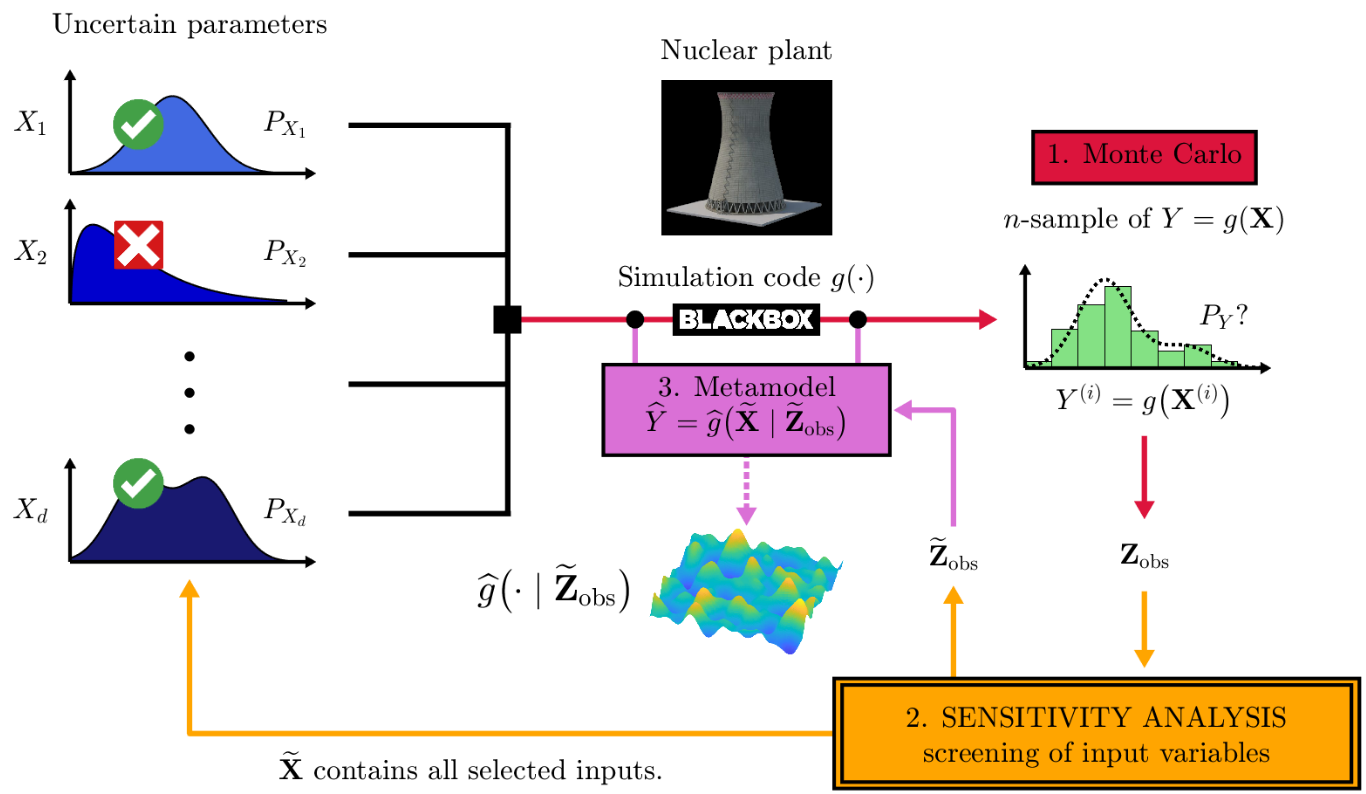
- The metamodel $\hat{g}$ must be constructed from $\boldsymbol{Z}_{\text{obs}} := \left\{\left(\boldsymbol{X}^{(i)}, Y^{(i)}\right)\right\}_{1 \le i \le n}$ with $n \le 10d$ → **SMALL DATA**.
- For a nice coverage of the input domain of variation, the DoE must be space-filling → **GIVEN DATA**.

Classical metamodeling techniques (such as **GP regression**) cannot be used directly.
Curse of dimensionality → too many **GP hyperparameters** have to be optimized!

Many existing strategies (screening, additive and ANOVA models, linear and nonlinear embeddings).
→ Binois & Wycoff (2022) for a comprehensive review.

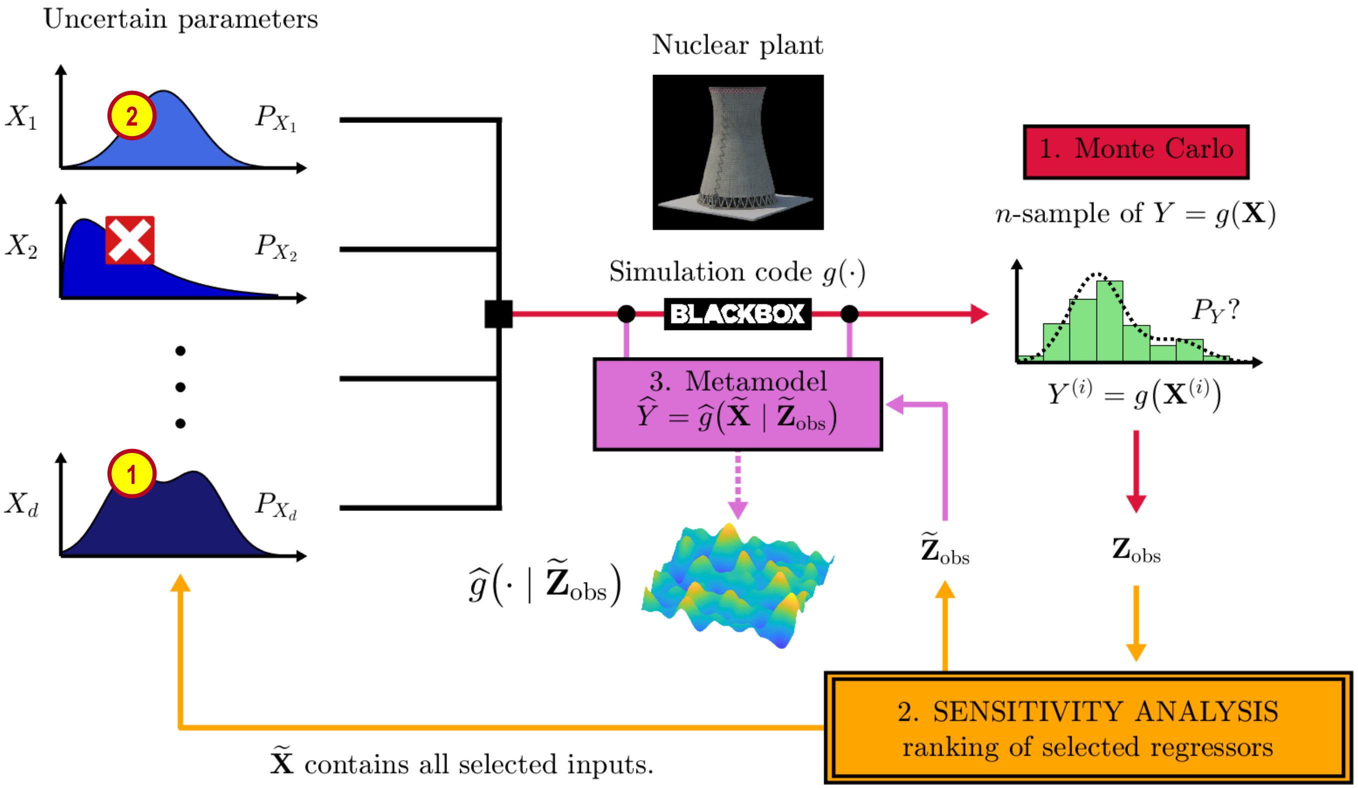
Focus on **SCREENING** → preliminary GSA for **variable selection** (and thus **dimension reduction**).

GSA in support to metamodel construction



- **Steps 2 and 3 of the ICSCREAM methodology** → Iooss & Marrel (2019) or Marrel *et al.* (2020)
  - ✓ **Identification of penalizing Configurations using SCREening And Metamodel**
- Performing a **preliminary GSA** has **two main advantages**.
  - **Screening-oriented GSA** → (crude) **dimension reduction** by discarding non-influential input variables.

GSA in support to metamodel construction



- **Steps 2 and 3 of the ICSCREAM methodology** → Iooss & Marrel (2019) or Marrel *et al.* (2020)
  - ✓ **Identification of penalizing Configurations using SCREening And Metamodel**
- Performing a **preliminary GSA** has **two main advantages**.
  - **Screening-oriented GSA** → (crude) **dimension reduction** by discarding non-influential input variables.
  - **Ranking-oriented GSA** → **sequential building process** of the GP metamodel.

Summary

1. Various concepts related to **kernels**
2. Sensitivity measures based on the **HSIC**
3. A bridge between two opposite worlds: **HSIC-ANOVA indices**
4. Is it relevant to talk about **interactions** for HSIC-ANOVA indices?
5. More about **Sobolev kernels** and their properties
6. Does all this benefit independence testing?

1. Various concepts related to kernels

Fundamentals of reproducing kernel theory

1 Reproducing kernel Hilbert space (RKHS) → Berlinet & Thomas-Agnan (2011)

→ Let $K: \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ be a function defined on $\mathcal{Z} \subseteq \mathbb{R}^p$ with $p \geq 1$.

K is said to be a **kernel** if it is **symmetric** and **positive definite**.

→ Let $(\mathcal{H}, \langle \cdot,\cdot \rangle_{\mathcal{H}})$ be a Hilbert space in $\mathbb{R}^{\mathcal{Z}}$ (the space of all functions defined from $\mathcal{Z}$ to $\mathbb{R}$).

A Hilbert space $(\mathcal{H}, \langle \cdot,\cdot \rangle_{\mathcal{H}})$ is said to be a **reproducing kernel Hilbert space (RKHS)** if:

$$\forall\, z \in \mathcal{Z}, \quad \exists\, C_z > 0, \quad \text{such that} \quad \forall\, h \in \mathcal{H}, \quad \left|h(z)\right| \leq C_z \left\|h\right\|_{\mathcal{H}} .$$

- Generally speaking, the smoother the functions, the smaller the function space.
  - ✓ **An RKHS is sufficiently big to remain complete.**
  - ✓ **An RKHS is sufficiently smooth to have interesting properties.**

Moore-Aronszajn theorem

There is a **one-to-one mapping** between reproducing kernels and RKHSs.

$K: \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$	↔	$(\mathcal{H}, \langle \cdot,\cdot \rangle_{\mathcal{H}})$ with $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{Z}}$
kernel	reproducing property	RKHS

$$\forall\, z \in \mathcal{Z}, \quad \forall\, h \in \mathcal{H}, \quad h(z) = \left\langle h, K(\cdot, z) \right\rangle_{\mathcal{H}}$$

Fundamentals of reproducing kernel theory

2 Kernel mean embeddings → Muandet *et al.* (2017)

→ Let $\mathcal{M}_1^+(\mathcal{Z})$ be the space of all probability measures defined on $\mathcal{Z} \subseteq \mathbb{R}^p$.
→ Let $K: \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ be a kernel and let $\mathcal{H}$ be the induced RKHS.

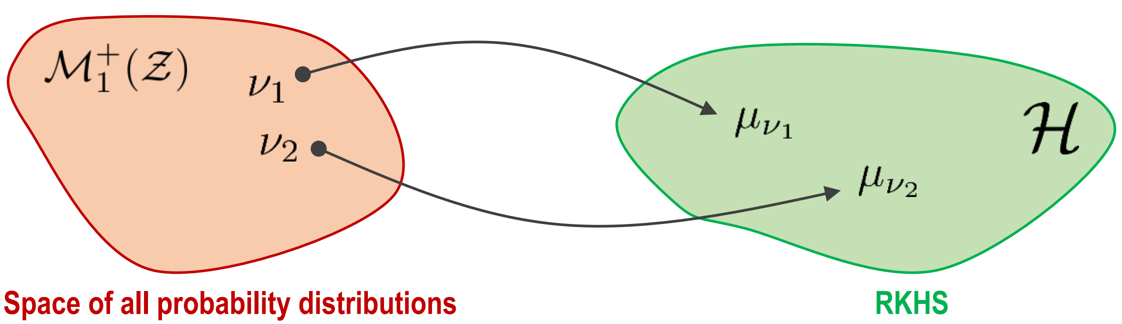
- Any probability measure $\nu \in \mathcal{M}_1^+(\mathcal{Z})$ can be represented by a (well-defined) function $\mu_\nu \in \mathcal{H}$.

$$\begin{aligned} \mu_\nu : \ \mathcal{Z} &\longrightarrow \mathbb{R} \\ z &\longmapsto \mu_\nu(z) = \mathbb{E}_\nu\left[K(z, Z)\right] = \int_{\mathcal{Z}} K(z,\zeta)\,\mathrm{d}\nu(\zeta) \end{aligned}$$

Assumptions

- K must be measurable
- $\mathbb{E}_\nu\left[\sqrt{K(Z,Z)}\right] < \infty$

- K is said to be a **characteristic kernel** if the map $\nu \mapsto \mu_\nu$ is **injective**.



- The **dissimilarity** between ν_1 and ν_2 can be measured through the **distance** in $\mathcal{H}$ between μ_{ν_1} and μ_{ν_2}.
  - ✓ Definition of a **kernel-based dissimilarity measure** on $\mathcal{M}_1^+(\mathcal{Z})$.

Fundamentals of reproducing kernel theory

2 Kernel mean embeddings → Muandet *et al.* (2017)

→ Let $\mathcal{M}_1^+(\mathcal{Z})$ be the space of all probability measures defined on $\mathcal{Z} \subseteq \mathbb{R}^p$.
→ Let $K: \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ be a kernel and let $\mathcal{H}$ be the induced RKHS.

➢ Any probability measure $\nu \in \mathcal{M}_1^+(\mathcal{Z})$ can be represented by a (well-defined) function $\mu_\nu \in \mathcal{H}$.

$$\begin{aligned} \mu_\nu : \ \mathcal{Z} &\longrightarrow \mathbb{R} \\ z &\longmapsto \mu_\nu(z) = \mathbb{E}_\nu\left[K(z, Z)\right] = \int_{\mathcal{Z}} K(z,\zeta)\,\mathrm{d}\nu(\zeta) \end{aligned}$$

Assumptions

- K must be measurable
- $\mathbb{E}_\nu\left[\sqrt{K(Z,Z)}\right] < \infty$

➢ K is said to be a **characteristic kernel** if the map $\nu \mapsto \mu_\nu$ is **injective**.

3 Maximum Mean Discrepancy (MMD) → Gretton *et al.* (2006)

$$\begin{aligned} \mathrm{MMD}^2(\nu_1,\nu_2) &= \boxed{\|\mu_{\nu_1} - \mu_{\nu_2}\|_{\mathcal{H}}^2} \\ &= \boxed{\mathbb{E}_{\nu_1\otimes\nu_1}\left[K(Z,Z')\right] + \mathbb{E}_{\nu_2\otimes\nu_2}\left[K(Z,Z')\right] - 2\,\mathbb{E}_{\nu_1\otimes\nu_2}\left[K(Z,Z')\right]} \end{aligned}$$

✓ **Definition resulting from the embedding mechanism**

✓ **Alternative formula paving the way to a simple estimation procedure**

Fundamentals of reproducing kernel theory

4 Feature maps → **Chapter 4** in Steinwart & Christmann (2008)

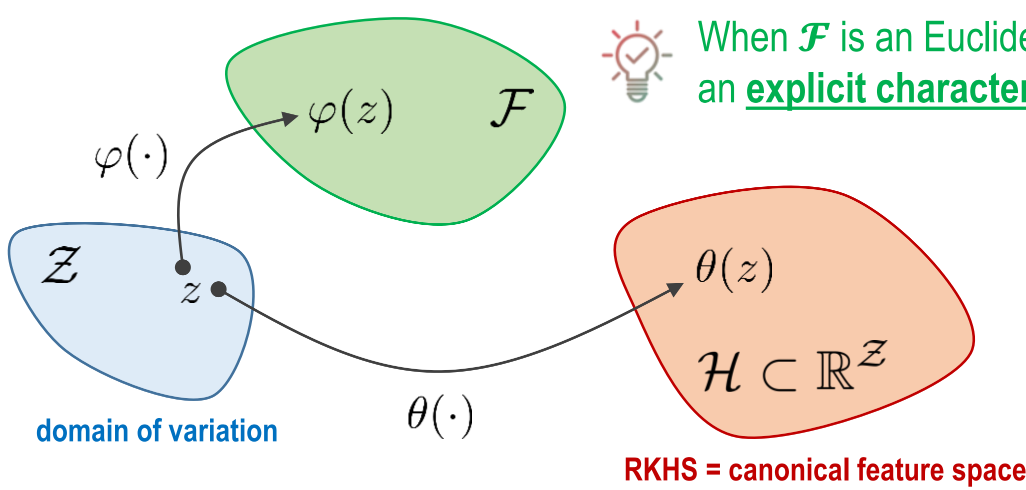
→ Let $K: \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ be a kernel and let $\mathcal{H}$ be the induced RKHS.

- Let us assume that there exist a **Hilbert space** $\mathcal{F}$ and a **map** $\varphi: \mathcal{Z} \to \mathcal{F}$ such that:

$$\boxed{\forall\, z, z' \in \mathcal{Z},\ K(z, z') = \langle \varphi(z), \varphi(z') \rangle_{\mathcal{F}}}$$

$\mathcal{F}$ is called a **feature space**. φ is called a **feature map**. Any object $\varphi(z)$ is called a **feature function**.

- <u>**Existence**</u> of at least one feature map.
  - ✓ The **canonical feature map** $\theta : \mathcal{Z} \to \mathcal{H}$ is thus defined by $\theta(z) := K(\cdot, z)$ for any $z \in \mathcal{Z}$.
- <u>**Non-unicity**</u> of the feature map.
  - ✓ There may exist a feature space where the kernel action is **much easier to understand**.



When $\mathcal{F}$ is an Euclidean space or a sequence space, an <u>**explicit characterization of the RKHS**</u> can be derived.

Most often, $\theta(\cdot)$ is **NOT informative!**

Fundamentals of reproducing kernel theory

5 **Feature-based characterization of the RKHS** → **Chapter 4** in Steinwart & Christmann (2008)

First, let us examine **two particular kernels!**

Example 1 ➢ The **polynomial kernel** with position parameter $c \geq 0$ and exponent $m \in \mathbb{N}^*$.

initial definition

$$\boxed{K_{\text{poly}}(x, x') := (xx' + c)^m} = \sum_{k=0}^{m} \binom{m}{k} x^k (x')^k c^{m-k}$$

$$= \langle \varphi_{\text{poly}}(x), \varphi_{\text{poly}}(x') \rangle_{\mathbb{R}^{m+1}} \quad \text{with} \quad \boxed{\varphi_{\text{poly}}(x) = \left[(\sqrt{c})^{m-k} \sqrt{\binom{m}{k}} x^k \right]_{0 \leq k \leq m}}$$

finite number of polynomial features

✓ The **binomial theorem** reveals a feature map φ_{poly} from $\mathbb{R}$ to the **Euclidean** space $\mathbb{R}^{m+1}$.

Fundamentals of reproducing kernel theory

5 **Feature-based characterization of the RKHS** → **Chapter 4** in Steinwart & Christmann (2008)

First, let us examine **two particular kernels!**

Example 2 ➢ The **Gaussian kernel** with scale parameter $\gamma > 0$.

initial definition

$$\boxed{K_\gamma(x,x') := e^{-\frac{1}{2}\left(\frac{x-x'}{\gamma}\right)^2}} = e^{-\frac{1}{2}\left(\frac{x}{\gamma}\right)^2} e^{-\frac{1}{2}\left(\frac{x'}{\gamma}\right)^2} \sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{x}{\gamma}\right)^k \left(\frac{x'}{\gamma}\right)^k$$

$$= \langle \varphi_\gamma(x'), \varphi_\gamma(x) \rangle_{\ell^2} \quad \text{with} \quad \boxed{\varphi_\gamma(x) := e^{-\frac{1}{2}\left(\frac{x}{\gamma}\right)^2} \left[\frac{1}{\sqrt{k}}\left(\frac{x}{\gamma}\right)^k\right]_{k\geq 0}}$$

infinite number of damped polynomial features

✓ The **Taylor series expansion** reveals a feature map φ_γ from $\mathbb{R}$ into the **Hilbert** space $\ell^2(\mathbb{N})$.

Fundamentals of reproducing kernel theory

5 **Feature-based characterization of the RKHS** → **Chapter 4** in Steinwart & Christmann (2008)

- As shown in these two examples, a **kernel expansion** allows to identify a **feature map.**
  - ✓ More importantly, it provides **all-in-one characterization** of the RKHS.

→ Let $K: \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ be a kernel and let $\mathcal{H}$ be the induced RKHS.

- It is assumed that it can be expanded as a sum (or series) of **symmetric** and **separable** functions.

$$\boxed{\forall\, z, z' \in \mathcal{Z}, \quad K(z,z') = \sum_{i\in I} g_i(z)\, g_i(z')}$$

Polynomial kernel → $I = \{0, \dots, m\}$
Gaussian kernel → $I = \mathbb{N}$

- ✓ The functions $(g_i)_{i\in I}$ are the **features**. They must be **linearly independent** (in the ℓ^2-sense).

❶ $$\mathcal{H} = \left\{ h \in \mathbb{R}^{\mathcal{Z}} \;\middle|\; h(\cdot) = \sum_{i\in I} a_i\, g_i(\cdot) \;\text{ with }\; (a_i)_{i\in I} \in \ell^2(I, \mathbb{R}) \right\}$$

❷ $$\begin{array}{rcccl} \langle\cdot,\cdot\rangle_{\mathcal{H}} : & \mathcal{H} & \times & \mathcal{H} & \longrightarrow \mathbb{R} \\ & \Big(h_1(\cdot) = \sum_{i\in I} a_i\, g_i(\cdot) & , & h_2(\cdot) = \sum_{i\in I} b_i\, g_i(\cdot) \Big) & \longmapsto \sum_{i\in I} a_i\, b_i \end{array}$$

❸ The functions $(g_i)_{i\in I}$ form an **orthonormal basis (ONB)** of $\mathcal{H}$.

2. Sensitivity measures based on the HSIC

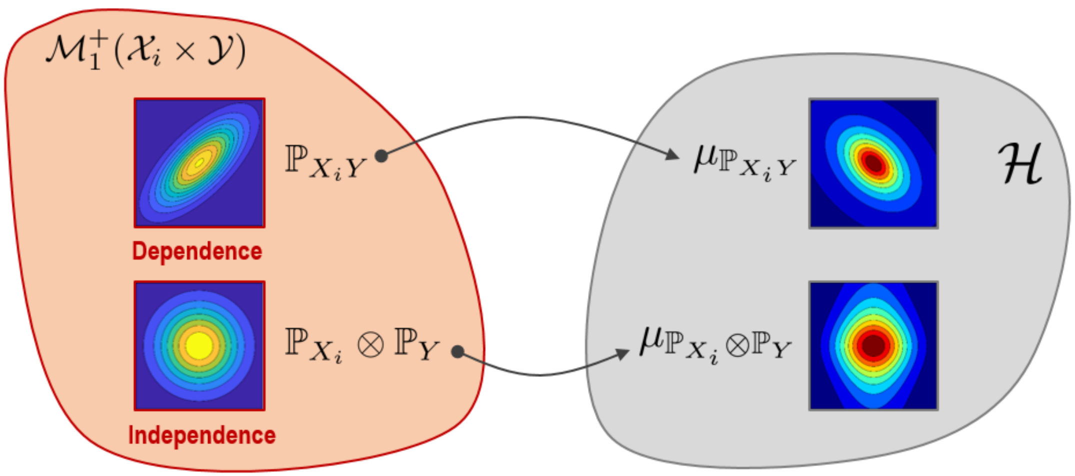
Several views on HSIC indices

1 Kernel-based dependences measures → Da Veiga (2015)

- $\mathbb{P}_{X_iY}$ → Joint distribution of (X_i, Y) → True influence of X_i on Y
- $\mathbb{P}_{X_i} \otimes \mathbb{P}_Y$ → Independence within (X_i, Y) → Hypothetical lack of influence

$$\boxed{S_i^{\Delta} := \Delta\left(\mathbb{P}_{X_iY}, \mathbb{P}_{X_i} \otimes \mathbb{P}_Y\right)}$$

How to measure the discrepancy?
What about using the MMD?



space of all probability distributions for the input-output pair (X_i, Y)

RKHS

- $K_i : \mathcal{X}_i \times \mathcal{X}_i \to \mathbb{R}$ assigned to X_i
- $K_Y : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ assigned to Y
- $K_i \otimes K_Y$ used to handle (X_i, Y)
- $\mathcal{H} := \mathcal{H}_i \otimes \mathcal{H}_Y$ induced by $K_i \otimes K_Y$

$$\boxed{\mathrm{HSIC}(X_i, Y) := \mathrm{MMD}^2\left(\mathbb{P}_{X_iY}, \mathbb{P}_{X_i} \otimes \mathbb{P}_Y\right) = \left\| \mu_{\mathbb{P}_{X_iY}} - \mu_{\mathbb{P}_{X_i} \otimes \mathbb{P}_Y} \right\|_{\mathcal{H}}^2}$$

How to estimate HSIC indices from a given dataset?

Several views on HSIC indices

2 Efficient estimation → Gretton *et al.* (2005, 2007) and Serfling (2009)

- The **alternative formula of the MMD** allows to rewrite the HSIC only in terms of **kernel-based moments**.

$$\mathrm{HSIC}(X_i, Y) = \mathbb{E}\Big[K_i(X_i, X_i')\ K_Y(Y, Y')\Big] + \mathbb{E}\Big[K_i(X_i, X_i')\ K_Y(Y'', Y''')\Big] - 2\,\mathbb{E}\Big[K_i(X_i, X_i')\ K_Y(Y, Y'')\Big]$$

$(X_i, Y) \perp (X_i', Y') \perp (X_i'', Y'') \perp (X_i''', Y''')$ follow the **joint input-output distribution** $\mathbb{P}_{X_i Y}$.

- **U-statistics** and **V-statistics** are well-adapted to estimate HSIC indices from a given DoE.

$N_{sim} = n$

$$\widehat{H}_i^U = \frac{1}{(n)_2} \sum_{1 \leq p \neq q \leq n} K_i\left(X_i^{(p)}, X_i^{(q)}\right) K_Y\left(Y^{(p)}, Y^{(q)}\right) + \frac{1}{(n)_4} \sum_{1 \leq p \neq q \neq r \neq s \leq n} K_i\left(X_i^{(p)}, X_i^{(q)}\right) K_Y\left(Y^{(r)}, Y^{(s)}\right) - \frac{2}{(n)_3} \sum_{1 \leq p \neq q \neq r \leq n} K_i\left(X_i^{(p)}, X_i^{(q)}\right) K_Y\left(Y^{(p)}, Y^{(r)}\right) \quad \text{with} \quad (n)_p = p! \binom{n}{p}$$

- $\widehat{H}_i^U$ denotes the **U-statistic** estimator of $\mathrm{HSIC}(X_i, Y)$ → **no bias** BUT **no guarantee of positivity**.
- $\widehat{H}_i^V$ denotes the **V-statistic** estimator of $\mathrm{HSIC}(X_i, Y)$ → **positivity** BUT **bias**.
- **Consistency** and **existence of a CLT** → convergence at rate $1/\sqrt{n}$.
- **Low computational complexity** → only $\mathcal{O}(n^2)$ operations are required to compute estimates.

Several views on HSIC indices

3 Cross-covariance operators → Gretton *et al.* (2005)

- Let $K_i : \mathcal{X}_i \times \mathcal{X}_i \to \mathbb{R}$ be the i-th input kernel (with RKHS denoted by $\mathcal{H}_i$).
- Let $K_Y : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ be the output kernel (with RKHS denoted by $\mathcal{H}_Y$).

➢ The knowledge of $\mathcal{H}_i$ and $\mathcal{H}_Y$ allows to rewrite $\mathrm{HSIC}(X_i, Y)$ as a kind of **generalized covariance**.

$$\boxed{\mathrm{HSIC}(X_i, Y) = \sum_k \sum_l \left|\mathrm{Cov}\left(v_{ik}(X_i), w_l(Y)\right)\right|^2} \text{ with } \boxed{\begin{cases} (v_{ik})_k & \text{an ONB of } \mathcal{H}_i \\ (w_l)_l & \text{an ONB of } \mathcal{H}_Y \end{cases}}$$

sum of covariances for different patterns **catalogues of transformations**

✓ **Aggregation of covariance terms** obtained after applying sequences of **preliminary basis transformations**.

✓ Each **pair of non-linear functions** $\left(v_{ik}(\cdot), w_l(\cdot)\right)$ corresponds to a **non-linear dependence pattern**.

Example ➢ **HSIC indices** computed with **Gaussian kernels** → $K_i = K_Y = K_\gamma$

$$K_\gamma(z, z') = e^{-\frac{1}{2}\left(\frac{z-z'}{\gamma}\right)^2} = \sum_{k=0}^{\infty} g_k(z)\, g_k(z') \quad \text{with} \quad \boxed{g_k(z) \propto e^{-\frac{1}{2}\left(\frac{z}{\gamma}\right)^2} z^k}$$

damped polynomial feature

$$\mathrm{HSIC}(X_i, Y) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \left|\mathrm{Cov}\Big(g_k(X_i), g_l(Y)\Big)\right|^2$$

✅ **Infinitely many damped polynomial transformations are applied to both X_i and Y.**

Several views on HSIC indices

4 Independence testing → Gretton *et al.* (2007)

→ The input kernel $K_i : \mathcal{X}_i \times \mathcal{X}_i \to \mathbb{R}$ is assumed to be **characteristic** to $\mathcal{M}_1^+(\mathcal{X}_i)$.

→ The output kernel $K_Y : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ is assumed to be **characteristic** to $\mathcal{M}_1^+(\mathcal{Y})$.

$$\boxed{X_i \perp Y \iff \mathrm{HSIC}(X_i, Y) = 0}$$

- **Testing independence** between X_i and Y is equivalent to **testing the nullity** of the HSIC.

$$(H_0) : \mathrm{HSIC}(X_i, Y) = 0 \quad \text{vs.} \quad (H_1) : \mathrm{HSIC}(X_i, Y) > 0$$

- **Test statistic** → either $\widehat{H}_i^U$ or $\widehat{H}_i^V$
- **Test procedure** → selected according to the **sample size** and the chosen **test statistic**
  - ✓ **Asymptotic test procedure** → Zhang *et al.* (2018)
  - ✓ **Permutation-based test procedure** → De Lozzo & Marrel (2016)
  - ✓ **Sequential permutation-based test procedure** → El Amri & Marrel (2022)
  - ✓ **Non-asymptotic Gamma test procedure** → El Amri & Marrel (2023)

5 Comparison with Sobol' indices

- Much harder to interpret → **no uniform bound** + **sum $\neq$ 1** + **non-trivial mathematical foundations**.
- Not conceptually tailored to ranking-oriented GSA → **no link with the output variability**.

Sobol' indices vs. HSIC indices

- HSIC indices perfectly meet the needs of **screening-oriented** GSA.
  - ✓ **The use of characteristic kernels allows to detect any type of input-output dependence.**
  - ✓ **Inference is an easy task (no need for specific data, big data or density estimation).**

GSA requirements	S_i	T_i	$\mathrm{HSIC}(X_i, Y)$
ANOVA decomposition → RANKING	✓	✓	✗
Characterize independence → SCREENING	✗	✓	✓
Estimation from GIVEN DATA	✓	✗	✓
Estimation from SMALL DATA	✓	✗	✓
Compatibility with DEPENDENT inputs	✗	✗	✓
INVARIANCE through monotonic transformations	✓	✓	✗

Still room to improve HSIC indices?

- HSIC indices <u>**lack interpretability**</u> and they are not tailored to perform **ranking-oriented** GSA.

⚠ **Sum not equal to 1.** ⚠ **No universal bound.** ⚠ **Different MMD scales.**

GSA requirements	S_i	T_i	$\mathrm{HSIC}(X_i, Y)$
ANOVA decomposition → RANKING	✔	✔	✘
Characterize independence → SCREENING	✘	✔	✔
Estimation from GIVEN DATA	✔	✘	✔
Estimation from SMALL DATA	✔	✘	✔
Compatibility with DEPENDENT inputs	✘	✘	✔
INVARIANCE through monotonic transformations	✔	✔	✘

🤔 How to do better on that point?

3. A bridge between two opposite worlds: HSIC-ANOVA indices

Taking inspiration from standard ANOVA…

- **ANOVA decomposition for Sobol' indices** → Sobol' (1993)
  - ✓ The output variance $\mathbb{V}(Y)$ is apportioned between **all subsets of inputs**.

$$\boxed{\mathbb{V}(Y) = \sum_{\boldsymbol{u} \subseteq \{1,\ldots,d\}} V_{\boldsymbol{u}} = \sum_{\boldsymbol{u} \subseteq \{1,\ldots,d\}} \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} \mathbb{V}\Big(\mathbb{E}[Y \mid X_{\boldsymbol{u}}]\Big)}$$

⚠ $X_1 \perp \cdots \perp X_d$

- **First-order and total-order Sobol' indices**
  - ✓ **First-order** Sobol' indices $(S_i)_{1 \le i \le d}$ → **main effects** only!
  - ✓ **Total-order** Sobol' indices $(T_i)_{1 \le i \le d}$ → **main effects** + **interactions**.

$$\boxed{\forall\, 1 \le i \le d, \quad S_i = \frac{\mathbb{V}(\mathbb{E}[Y \mid X_i])}{\mathbb{V}(Y)} \quad \text{and} \quad T_i = 1 - \frac{\mathbb{V}(\mathbb{E}[Y \mid \boldsymbol{X}_{-i}])}{\mathbb{V}(Y)}}$$

- **Constraints imposed on the sub-functions of the Sobol'-Hoeffding decomposition**

$$g(\boldsymbol{x}) = \sum_{\boldsymbol{u} \subseteq \{1,\ldots,d\}} \eta_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}}) \quad \text{such that} \quad \forall\, i \in \boldsymbol{u}, \ \boxed{\int_{\mathcal{X}_i} \eta_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}})\, \mathrm{d}\mathbb{P}_{X_i}(x_i) = 0}$$

… and bringing ANOVA into the HSIC paradigm

- **HISC-ANOVA decomposition** → Da Veiga (2021)
  - ✓ The quantity $\mathrm{HSIC}(\boldsymbol{X}, Y)$ is apportioned between **all subsets of inputs**.

$$\boxed{\mathrm{HSIC}(\boldsymbol{X}, Y) = \sum_{\boldsymbol{u}\subseteq\{1,\dots,d\}} H_{\boldsymbol{u}} = \sum_{\boldsymbol{u}\subseteq\{1,\dots,d\}} \sum_{\boldsymbol{v}\subseteq\boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|}\mathrm{HSIC}(\boldsymbol{X}_{\boldsymbol{v}}, Y)}$$

⚠ $X_1 \perp \cdots \perp X_d$

- **First-order and total-order HSIC-ANOVA indices**
  - ✓ **First-order** HSIC-ANOVA indices $\left(S_i^{\mathrm{HSIC}}\right)_{1\leq i\leq d}$ → **main effects** only!
  - ✓ **Total-order** HSIC-ANOVA indices $\left(T_i^{\mathrm{HSIC}}\right)_{1\leq i\leq d}$ → **main effects + interactions**.

$$\boxed{\forall\, 1\leq i\leq d, \quad S_i^{\mathrm{HSIC}} := \frac{\mathrm{HSIC}(X_i, Y)}{\mathrm{HSIC}(\mathbf{X}, Y)} \quad \text{and} \quad T_i^{\mathrm{HSIC}} := 1 - \frac{\mathrm{HSIC}(\mathbf{X}_{-i}, Y)}{\mathrm{HSIC}(\mathbf{X}, Y)}}$$

- **Constraints imposed on the input kernels**
  - ✓ Each input kernel K_i must be an **ANOVA kernel** (≈ a **constant** kernel + an **orthogonal** kernel).

$$K_i(x_i, x_i') = 1 + k_i(x_i, x_i') \quad \text{with} \quad \forall\, x_i \in \mathcal{X}_i, \quad \boxed{\int_{\mathcal{X}_i} k_i(x_i, x_i')\,\mathrm{d}\mathbb{P}_{X_i}(x_i') = 0}$$

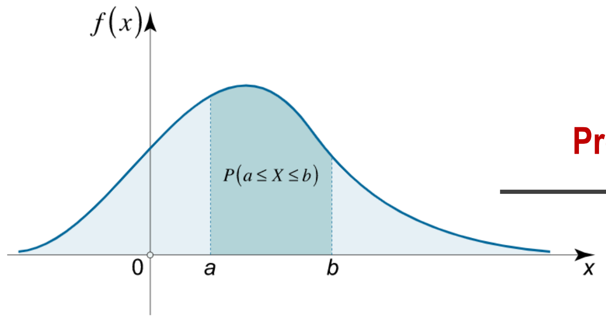
  - ✓ $\boldsymbol{\mathcal{H}}_i = \mathbb{R} \oplus \boldsymbol{\mathcal{G}}_i$ where $\boldsymbol{\mathcal{G}}_i$ is only composed of **zero-mean functions** (with respect to $\mathbb{P}_{X_i}$).

How to find ANOVA kernels?

For most **parametric families of distributions**, there is no well-known **characteristic ANOVA** kernel.

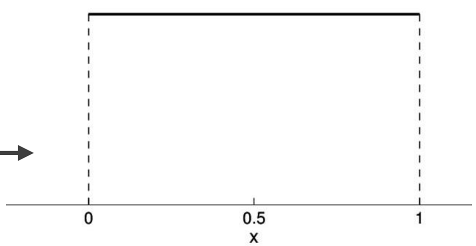
How to implement the HSIC-ANOVA decomposition in practice?

1. Transform each input distribution $\mathbb{P}_{X_i}$ into a standard uniform distribution $\mathcal{U}([0,1])$.



$$U_i = F_{X_i}(X_i)$$

Probability Integral Transform (PIT)



Density of the i-th input variable

Density of the uniform distribution

2. Assign a **Sobolev kernel** K^r_{Sob} to each new input variable $U_i := F_{X_i}(X_i)$.

$$\boxed{\forall\, u, u' \in [0,1], \quad K^r_{\text{Sob}}(u,u') := 1 + \sum_{i=1}^{r} \frac{B_i(u)\, B_i(u')}{(i!)^2} + \frac{(-1)^{r+1}}{(2r)!} B_{2r}(|u-u'|)}$$

- ✓ $r \in \mathbb{N}^*$ is an **integer parameter** indicating the **degree of smoothness** of the RKHS.
- ✓ The functions $(B_i)_{i\geq 1}$ are the **Bernoulli polynomials** → $\int_0^1 B_i(u)\, du = 0$.

A grey area around HSIC-ANOVA indices?

1. How do they measure sensitivity? How to distinguish between main effects and interactions?
2. Are they able to characterize independence?

GSA requirements	T_i	$\mathrm{HSIC}(X_i, Y)$	S_i^{HSIC} ?	T_i^{HSIC}
ANOVA decomposition → RANKING	✔	✘	?	?
Characterize independence → SCREENING	✔	✔	? ?	? ?
Estimation from GIVEN DATA	✘	✔	✔	✔
Estimation from SMALL DATA	✘	✔	✔	✔
Compatibility with DEPENDENT inputs	✘	✔	✘	✘
INVARIANCE through monotonic transformations	✔	✘	✘	✘

4. Is it relevant to talk about interactions for HSIC-ANOVA indices?

Focus on HSIC-ANOVA interactions

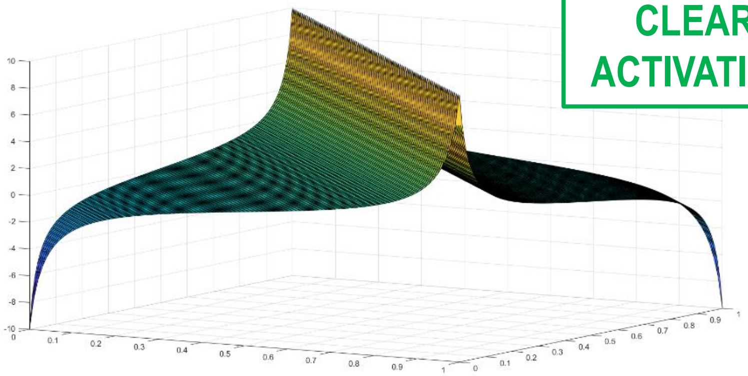
For most benchmark test cases, HSIC-ANOVA interactions are not significant.

Example → the **Ishigami** function

$$Y = g(X_1, X_2, X_3) = \sin(X_1) + \sin^2(X_2) + X_3^4 \sin(X_1) \quad \text{with} \quad X_i \sim \mathcal{U}([-\pi, \pi])$$

- Strong interaction between X_1 and X_3 in the **variance-based ANOVA** framework.
- No interaction between X_1 and X_3 in the **HSIC-ANOVA** framework.

Counterexample → Hand-made **pathological functions** (only for $d \approx 2$)



CLEAR ACTIVATION

Hull function

$$g(x_1, x_2) = -\tan\left[(2\sqrt{2})a\left|\frac{x_1 + x_2 - 1}{\sqrt{2}}\right| - a\right]$$

$$S_1^{\text{HSIC}} = S_2^{\text{HSIC}} = 17\%$$

$$T_1^{\text{HSIC}} = T_2^{\text{HSIC}} = 83\%$$

No clear explanation on why those functions lead to strong HSIC-ANOVA interactions.

The feature-based viewpoint on the HSIC allows to break the deadlock.

A detour through cross-covariance operators

1 HSIC indices

- Remember the **reformulation** of the HSIC as a sum of **covariance terms** (depending on the chosen kernels).

$$\boxed{\mathrm{HSIC}(X_1, Y) = \sum_k \sum_l \left|\mathrm{Cov}\big(v_{1k}(X_1), w_l(Y)\big)\right|^2} \quad \text{with} \quad \boxed{\begin{cases} (v_{1k})_k & \text{an ONB of } \mathcal{H}_1 \\ (w_l)_l & \text{an ONB of } \mathcal{H}_Y \end{cases}}$$

dependence patterns captured by K_1 and K_Y **catalogues of transformations**

A detour through cross-covariance operators

1 HSIC indices

- Remember the **reformulation** of the HSIC as a sum of **covariance terms** (depending on the chosen kernels).

$$\boxed{\mathrm{HSIC}(X_1,Y) = \sum_k \sum_l \left|\mathrm{Cov}\left(v_{1k}(X_1), w_l(Y)\right)\right|^2} \quad \text{with} \quad \boxed{\begin{cases} (v_{1k})_k & \text{an ONB of } \mathcal{H}_1 \\ (w_l)_l & \text{an ONB of } \mathcal{H}_Y \end{cases}}$$

dependence patterns captured by K_1 and K_Y — **catalogues of transformations**

2 First-order HSIC-ANOVA indices

- Application of the above formula in the case where K_1 is an <u>**ANOVA**</u> kernel.
  - ✓ The RKHS induced by $K_1 = 1 + k_1$ may be decomposed as $\boldsymbol{\mathcal{H}}_1 = \mathbb{R} \oplus \boldsymbol{\mathcal{G}}_1$.
  - ✓ All the functions in $\boldsymbol{\mathcal{G}}_1$ have **zero mean** (with respect to $\mathbb{P}_{X_1}$).
  - ✓ An ONB $(v_{1k})_k$ of $\boldsymbol{\mathcal{H}}_1$ can be obtained by taking $\{\mathbb{1}\,;(u_{1k})_k\}$ where $(u_{1k})_k$ is an ONB of $\boldsymbol{\mathcal{G}}_1$.

$$\boxed{S_1^{\mathrm{HSIC}} \propto \mathrm{HSIC}(X_1,Y) = \sum_k \sum_l \left|\mathrm{Cov}\left(u_{1k}(X_1), w_l(Y)\right)\right|^2} \quad \text{with} \quad \boxed{\begin{cases} (u_{1k})_k & \text{an ONB of } \mathcal{G}_1 \\ (w_l)_l & \text{an ONB of } \mathcal{H}_Y \end{cases}}$$

dependence patterns captured by k_1 and K_Y

🤔 **How to extend this reasoning to higher-order HSIC-ANOVA indices?**

A detour through cross-covariance operators

3 **HSIC-ANOVA decomposition**

- For the sake of clarity, it is assumed that $d = 2$.
  - ✓ **No loss of generality. Everything remains true in higher dimension!**

$$\mathrm{HSIC}(\boldsymbol{X}, Y) = \sum_{\boldsymbol{u} \subseteq \{1,\ldots,d\}} \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} \mathrm{HSIC}(\boldsymbol{X}_{\boldsymbol{v}}, Y)$$

$$= \mathrm{HSIC}(X_1, Y) + \mathrm{HSIC}(X_2, Y) + \ldots$$

$$\boxed{\mathrm{HSIC}(\boldsymbol{X}, Y) - \mathrm{HSIC}(X_1, Y) - \mathrm{HSIC}(X_2, Y)}$$

HSIC-ANOVA interaction term

Now, let us rewrite the left-hand term in the HSIC-ANOVA decomposition.

A detour through cross-covariance operators

3 HSIC-ANOVA decomposition

- For the sake of clarity, it is assumed that $d = 2$.
  - ✓ **No loss of generality. Everything remains true in higher dimension!**

$$\mathrm{HSIC}(\boldsymbol{X}, Y) = \sum_{\boldsymbol{u} \subseteq \{1,\dots,d\}} \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} \mathrm{HSIC}(\boldsymbol{X}_{\boldsymbol{v}}, Y)$$

$$= \mathrm{HSIC}(X_1, Y) + \mathrm{HSIC}(X_2, Y) + \dots$$

$$\mathrm{HSIC}(\boldsymbol{X}, Y) - \mathrm{HSIC}(X_1, Y) - \mathrm{HSIC}(X_2, Y)$$

- **Step A** → Identify the input and output kernels
  - ✓ For the random **INPUT vector** $\boldsymbol{X} = [X_1, X_2]$ → $K_1 \otimes K_2$ with RKHS $\mathcal{H}_1 \otimes \mathcal{H}_2$
  - ✓ For the random **OUTPUT variable** Y → K_Y with RKHS $\mathcal{H}_Y$

A detour through cross-covariance operators

3 HSIC-ANOVA decomposition

- For the sake of clarity, it is assumed that $d = 2$.
  - ✓ **No loss of generality. Everything remains true in higher dimension!**

$$\mathrm{HSIC}(\boldsymbol{X}, Y) = \sum_{\boldsymbol{u} \subseteq \{1,\dots,d\}} \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} \mathrm{HSIC}(\boldsymbol{X}_{\boldsymbol{v}}, Y)$$

$$= \mathrm{HSIC}(X_1, Y) + \mathrm{HSIC}(X_2, Y) + \dots$$

$$\mathrm{HSIC}(\boldsymbol{X}, Y) - \mathrm{HSIC}(X_1, Y) - \mathrm{HSIC}(X_2, Y)$$

- **Step A** → Identify the input and output kernels.
- **Step B** → Find an ONB for each input RKHS.

$$\boxed{\left(v_{1k}\right)_k = \left\{ \mathbb{1} \; ; \; \left(u_{1k}\right)_k \right\} \quad \text{and} \quad \left(v_{2k}\right)_k = \left\{ \mathbb{1} \; ; \; \left(u_{2k}\right)_k \right\}}$$

A detour through cross-covariance operators

3 HSIC-ANOVA decomposition

- For the sake of clarity, it is assumed that $d = 2$.
  - ✓ **No loss of generality. Everything remains true in higher dimension!**

$$\mathrm{HSIC}(\boldsymbol{X}, Y) = \sum_{\boldsymbol{u} \subseteq \{1,\ldots,d\}} \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} \mathrm{HSIC}(\boldsymbol{X}_{\boldsymbol{v}}, Y)$$

$$= \mathrm{HSIC}(X_1, Y) + \mathrm{HSIC}(X_2, Y) + \ldots$$

$$\mathrm{HSIC}(\boldsymbol{X}, Y) - \mathrm{HSIC}(X_1, Y) - \mathrm{HSIC}(X_2, Y)$$

- **Step A** → Identify the input and output kernels.
- **Step B** → Find an ONB for each input RKHS.
- **Step C** → Build an ONB of the product RKHS.

$$\begin{aligned} \left(v_{1i} \otimes v_{2j}\right)_{i,j\geq 0} &= \left\{ \mathbb{1} \otimes \mathbb{1} \ ; \ \left(u_{1i} \otimes \mathbb{1}\right)_{i\geq 1} \ ; \ \left(\mathbb{1} \otimes u_{2j}\right)_{j\geq 1} \ ; \ \left(u_{1i} \otimes u_{2j}\right)_{i,j\geq 1} \right\} \\ &= \left\{ \boldsymbol{x} \mapsto 1 \ ; \ \Big(\boldsymbol{x} \mapsto u_{1i}(x_1)\Big)_{i\geq 1} \ ; \ \Big(\boldsymbol{x} \mapsto u_{2j}(x_2)\Big)_{j\geq 1} \ ; \ \Big(\boldsymbol{x} \mapsto u_{1i}(x_1)\, u_{2j}(x_2)\Big)_{i,j\geq 1} \right\} \end{aligned}$$

A detour through cross-covariance operators

3 HSIC-ANOVA decomposition

- For the sake of clarity, it is assumed that $d = 2$.
  - ✓ **No loss of generality. Everything remains true in higher dimension!**

$$\mathrm{HSIC}(\boldsymbol{X}, Y) = \sum_{\boldsymbol{u} \subseteq \{1,\dots,d\}} \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} \mathrm{HSIC}(\boldsymbol{X}_{\boldsymbol{v}}, Y)$$

$$= \mathrm{HSIC}(X_1, Y) + \mathrm{HSIC}(X_2, Y) + \dots$$

$$\mathrm{HSIC}(\boldsymbol{X}, Y) - \mathrm{HSIC}(X_1, Y) - \mathrm{HSIC}(X_2, Y)$$

$$= \sum_i \sum_j \sum_k \left|\mathrm{Cov}\left(v_{1i}(X_1)\, v_{2j}(X_2), w_k(Y)\right)\right|^2$$

$$= \underbrace{\sum_i \sum_k \left|\mathrm{Cov}\left(u_{1i}(X_1), w_l(Y)\right)\right|^2}_{\mathrm{HSIC}(X_1,Y)} + \underbrace{\sum_j \sum_k \left|\mathrm{Cov}\left(u_{2j}(X_2), w_k(Y)\right)\right|^2}_{\mathrm{HSIC}(X_1,Y)} + \dots$$

$$\sum_i \sum_j \sum_l \left|\mathrm{Cov}\left(u_{1i}(X_1)\, u_{2j}(X_2), w_k(Y)\right)\right|^2$$

A detour through cross-covariance operators

3 HSIC-ANOVA decomposition

- For the sake of clarity, it is assumed that $d = 2$.
  - ✓ No loss of generality. Everything remains true in higher dimension!

$$\mathrm{HSIC}(\boldsymbol{X}, Y) = \sum_{\boldsymbol{u} \subseteq \{1,\dots,d\}} \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}|-|\boldsymbol{v}|} \mathrm{HSIC}(\boldsymbol{X}_{\boldsymbol{v}}, Y)$$

$$= \mathrm{HSIC}(X_1, Y) + \mathrm{HSIC}(X_2, Y) + \dots$$

$$\boxed{\mathrm{HSIC}(\boldsymbol{X}, Y) - \mathrm{HSIC}(X_1, Y) - \mathrm{HSIC}(X_2, Y)}$$

ANOVA viewpoint

$$= \sum_i \sum_j \sum_k \left| \mathrm{Cov}\left(v_{1i}(X_1)\, v_{2j}(X_2), w_k(Y) \right) \right|^2$$

$$= \underbrace{\sum_i \sum_k \left| \mathrm{Cov}\left(u_{1i}(X_1), w_l(Y) \right) \right|^2}_{\mathrm{HSIC}(X_1,Y)} + \underbrace{\sum_j \sum_k \left| \mathrm{Cov}\left(u_{2j}(X_2), w_k(Y) \right) \right|^2}_{\mathrm{HSIC}(X_1,Y)} + \dots$$

$$\boxed{\sum_i \sum_j \sum_l \left| \mathrm{Cov}\left(u_{1i}(X_1)\, u_{2j}(X_2), w_k(Y) \right) \right|^2}$$

feature-based viewpoint

A detour through cross-covariance operators

4 HSIC-ANOVA indices

$$S_1^{\mathrm{HSIC}} + S_2^{\mathrm{HSIC}} + \Delta_{12}^{\mathrm{HSIC}} = 1$$

$$S_1^{\mathrm{HSIC}} \propto \sum_i \sum_k \left|\mathrm{Cov}\left(u_{1i}(X_1), w_k(Y)\right)\right|^2$$

with

$$\begin{cases} (u_{1i})_i & \text{an ONB of } \mathcal{G}_1 \\ (w_k)_k & \text{an ONB of } \mathcal{H}_Y \end{cases}$$

dependence patterns captured by k_1 and K_Y

$$\Delta_{12}^{\mathrm{HSIC}} \propto \sum_i \sum_j \sum_k \left|\mathrm{Cov}\left(u_{1i}(X_1)\, u_{2j}(X_2), w_k(Y)\right)\right|^2$$

with

$$\begin{cases} (u_{1i})_i & \text{an ONB of } \mathcal{G}_1 \\ (u_{2j})_j & \text{an ONB of } \mathcal{G}_2 \\ (w_k)_k & \text{an ONB of } \mathcal{H}_Y \end{cases}$$

dependence patterns captured by $k_1 \otimes k_2$ and K_Y

A detour through cross-covariance operators

4 HSIC-ANOVA indices

$$S_1^{\text{HSIC}} + S_2^{\text{HSIC}} + \Delta_{12}^{\text{HSIC}} = 1$$

$$S_1^{\text{HSIC}} \propto \sum_i \sum_k \left|\text{Cov}\left(u_{1i}(X_1), w_k(Y)\right)\right|^2$$

with

$$\begin{cases} (u_{1i})_i & \text{an ONB of } \mathcal{G}_1 \\ (w_k)_k & \text{an ONB of } \mathcal{H}_Y \end{cases}$$

?

dependence patterns captured by k_1 and K_Y

$$\Delta_{12}^{\text{HSIC}} \propto \sum_i \sum_j \sum_k \left|\text{Cov}\left(u_{1i}(X_1)\, u_{2j}(X_2), w_k(Y)\right)\right|^2$$

with

$$\begin{cases} (u_{1i})_i & \text{an ONB of } \mathcal{G}_1 \\ (u_{2j})_j & \text{an ONB of } \mathcal{G}_2 \\ (w_k)_k & \text{an ONB of } \mathcal{H}_Y \end{cases}$$

?

dependence patterns captured by $k_1 \otimes k_2$ and K_Y

- Remember the **simplest solution** to compute HSIC-ANOVA indices.
  - ✓ **Uniform inputs** → $U_1 \perp U_2 \sim \mathcal{U}([0,1])$
  - ✓ **Sobolev kernels for the inputs** → $K_1 = K_2 = K_{\text{Sob}}^r$
  - ✓ **Gaussian kernel for the output** → $K_Y = K_\gamma$

$$\forall\, u, u' \in [0,1], \quad K_{\text{Sob}}^r(u, u') := 1 + \sum_{k=1}^{r} \frac{B_k(u)\, B_k(u')}{(k!)^2} + \frac{(-1)^{r+1}}{(2r)!} B_{2r}\left(|u - u'|\right)$$

🤔 **What are the basis functions in the case of Sobolev kernels?**

5. More about Sobolev kernels and their properties

Sobolev kernels and their feature maps

Many questions at the beginning of this work…

1. What is the RKHS $\boldsymbol{\mathcal{H}}_{\mathrm{Sob}}^{r}$ induced by K_{Sob}^{r}?
2. Is K_{Sob}^{r} a characteristic kernel?
3. Is there an explicit and easily interpretable feature map $\varphi_{\mathrm{Sob}}^{r} : [0,1] \to \boldsymbol{\mathcal{F}}_{\mathrm{Sob}}^{r}$?
4. How to identify an ONB of $\boldsymbol{\mathcal{H}}_{\mathrm{Sob}}^{r}$? Is there a link with feature maps?
5. How to choose r in practice?

Sobolev kernels and their feature maps

Many questions at the beginning of this work…

1. What is the RKHS $\mathcal{H}^r_{\mathrm{Sob}}$ induced by K^r_{Sob}? → see Gu (2013) or Kuo *et al.* (2010)

- **A standard function space:** the Sobolev space of order r defined on $[0,1]$ for the L^2-norm.

$$\boxed{H^r([0,1]) := \left\{ h \in \mathbb{R}^{[0,1]} \;\middle|\; \forall 0 \leq k \leq r, \quad D^k h \in L^2([0,1]) \right\}}$$

- **A specific inner product:**

$$\boxed{\langle f, g \rangle_{\mathcal{H}^r_{\mathrm{Sob}}} := \sum_{k=0}^{r-1} \left(\int_0^1 D^k f(x)\,\mathrm{d}x \right) \left(\int_0^1 D^k g(x)\,\mathrm{d}x \right) + \int_0^1 D^r f(x)\, D^r g(x)\,\mathrm{d}x}$$

Sobolev kernels and their feature maps

Many questions at the beginning of this work…

1. What is the RKHS $\mathcal{H}^r_{\text{Sob}}$ induced by K^r_{Sob}?

2. Is K^r_{Sob} a characteristic kernel?

- **YES!** Simply because $H^r([0,1])$ is **uniformly dense** in $C([0,1])$.

- **Major consequence**
  - ✓ The **HSIC-ANOVA indices** based on **Sobolev kernels** are able to characterize independence.

$$\boxed{X_i \perp Y \iff S_i^{\text{HSIC}} = 0 \iff T_i^{\text{HSIC}} = 0}$$

⚠ This is different from what happens for **Sobol' indices**.

$$\boxed{S_i = 0 \not\Longrightarrow X_i \perp Y} \quad \text{while} \quad \boxed{X_i \perp Y \iff T_i = 0}$$

Sobolev kernels and their feature maps

Many questions at the beginning of this work…

1. What is the RKHS $\mathcal{H}^r_{\text{Sob}}$ induced by K^r_{Sob}?
2. Is K^r_{Sob} a characteristic kernel?
3. Is there an explicit and easily interpretable feature map $\varphi^r_{\text{Sob}} : [0,1] \to \mathcal{F}^r_{\text{Sob}}$?

$$\boxed{K^r_{\text{Sob}}(x, x') = \left\langle \varphi^r_{\text{Sob}}(x), \varphi^r_{\text{Sob}}(x') \right\rangle_{\mathcal{F}^r_{\text{Sob}}}}$$

- For $r = 1$, the **Mercer expansion** of K^1_{Sob} is actually known. → Dick *et al.* (2014, 2015)

$$\boxed{K^1_{\text{Sob}}(x, x') := 1 + \sum_{k=1}^{\infty} \frac{1}{(k\pi)^2} c_k(x)\, c_k(x') \quad \text{with} \quad c_k(x) := \sqrt{2}\cos(k\pi x)}$$

- For $r \geq 2$, a **series expansion** of K^2_{Sob} is also mentioned in the literature. → Baldeaux *et al.* (2009)

$$\boxed{K^r_{\text{Sob}}(x, x') := 1 + \sum_{k=1}^{r} \frac{B_k(x)\, B_k(x')}{(k!)^2} + \sum_{k=1}^{\infty} \frac{1}{(2k\pi)^{2r}} \left[c_{2k}(x)\, c_{2k}(x') + s_{2k}(x)\, s_{2k}(x') \right] \quad \text{with} \quad \begin{cases} c_{2k}(x) := \sqrt{2}\cos(2k\pi x) \\ s_{2k}(x) := \sqrt{2}\sin(2k\pi x) \end{cases}}$$

Sobolev kernels and their feature maps

Many questions at the beginning of this work…

1. What is the RKHS $\mathcal{H}^r_{\mathrm{Sob}}$ induced by K^r_{Sob}?
2. Is K^r_{Sob} a characteristic kernel?
3. Is there an explicit and easily interpretable feature map $\varphi^r_{\mathrm{Sob}} : [0,1] \to \mathcal{F}^r_{\mathrm{Sob}}$?
4. How to identify an ONB of $\mathcal{H}^r_{\mathrm{Sob}}$? Is there a link with feature maps?

- **Mercer expansion** of K^1_{Sob} →

$$\boxed{K^1_{\mathrm{Sob}}(x,x') := 1 + \sum_{k=1}^{\infty} \frac{1}{(k\pi)^2}\, c_k(x)\, c_k(x')}$$

- **ONB** of the **RKHS** $\mathcal{H}^1_{\mathrm{Sob}}$ →

$$\boxed{\left\{ \mathbf{1};\ \left(\frac{c_k(\cdot)}{k\pi}\right)_{k\geq 1} \right\}}$$

- **Series expansion** of K^r_{Sob} →

$$\boxed{K^r_{\mathrm{Sob}}(x,x') := 1 + \sum_{k=1}^{r} \frac{B_k(x)\, B_k(x')}{(k!)^2} + \sum_{k=1}^{\infty} \frac{1}{(2k\pi)^{2r}} \left[c_{2k}(x)\, c_{2k}(x') + s_{2k}(x)\, s_{2k}(x') \right]}$$

- **ONB** of the **RKHS** $\mathcal{H}^r_{\mathrm{Sob}}$ →

$$\boxed{\left\{ \mathbf{1}\,;\ \left(\frac{B_k(\cdot)}{k!}\right)_{1\leq k\leq r};\ \left(\frac{c_{2k}(\cdot)}{(2k\pi)^r}\right)_{k\geq 1};\ \left(\frac{s_{2k}(\cdot)}{(2k\pi)^r}\right)_{k\geq 1} \right\}}$$

Sobolev kernels and their feature maps

Many questions at the beginning of this work…

1. What is the RKHS $\mathcal{H}^r_{\mathrm{Sob}}$ induced by K^r_{Sob}?
2. Is K^r_{Sob} a characteristic kernel?
3. Is there an explicit and easily interpretable feature map $\varphi^r_{\mathrm{Sob}} : [0,1] \to \mathcal{F}^r_{\mathrm{Sob}}$?
4. How to identify an ONB of $\mathcal{H}^r_{\mathrm{Sob}}$? Is there a link with feature maps?
5. How to choose r in practice?

- Taking $r = 1$ is recommended!
- For $r \geq 2$, $K^r_{\mathrm{Sob}}(x, x') \approx 1 + k_{\mathrm{lin}}(x, x')$ → **poor numerical performance** for screening!

Sobolev kernels and their feature maps

Many questions at the beginning of this work…

1. What is the RKHS $\mathcal{H}^r_{\mathrm{Sob}}$ induced by K^r_{Sob}?
2. Is K^r_{Sob} a characteristic kernel?
3. Is there an explicit and easily interpretable feature map $\varphi^r_{\mathrm{Sob}} : [0,1] \to \mathcal{F}^r_{\mathrm{Sob}}$?
4. How to identify an ONB of $\mathcal{H}^r_{\mathrm{Sob}}$? Is there a link with feature maps?
5. How to choose r in practice?

What is the point of these theoretical results?

- Remember the pure interaction term $\Delta^{\mathrm{HSIC}}_{12}$.
- Apply with $K_1 = K_2 = K^1_{\mathrm{Sob}}$ now that an ONB of $\mathcal{H}^1_{\mathrm{Sob}}$ is explicitly known.

$$\Delta^{\mathrm{HSIC}}_{12} \propto \sum_i \sum_j \sum_k \left|\mathrm{Cov}\left(u_{1i}(X_1)\, u_{2j}(X_2), w_k(Y)\right)\right|^2 = \boxed{\sum_i^{\infty} \sum_j^{\infty} \sum_k \frac{1}{ij\,\pi^2} \left|\mathrm{Cov}\left(c_i(X_1)\, c_j(X_2), w_k(Y)\right)\right|^2}$$

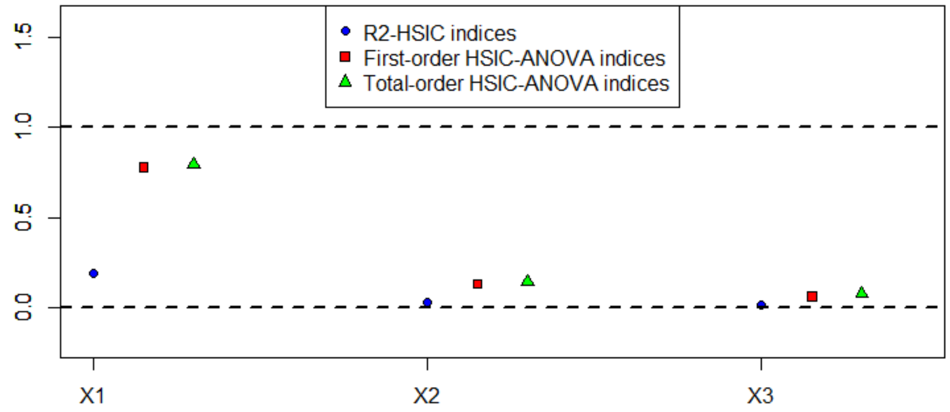
This provides the hint to design a toy case.

How to exacerbate HSIC-ANOVA interactions?

- **Back to the Ishigami function**
  - ✓ **Additional term** chosen to boost HSIC-ANOVA interactions.

$$Y = g(U_1, U_2, U_3) = \text{ishigami}(X_1, X_2, X_3) + \boldsymbol{\gamma} \cos(\pi U_1) \cos(\pi U_2) \quad \text{with} \quad \begin{array}{c} U_i \sim \mathcal{U}([0,1]) \\ X_i = \pi(2U_i - 1) \end{array}$$

- **Design parameter**
  - ✓ $\boldsymbol{\gamma = 0}$
- **Estimation of sensitivity measures**
  - ✓ Sample size $n = 500$
  - ✓ R^2-HSIC indices + HSIC-ANOVA indices



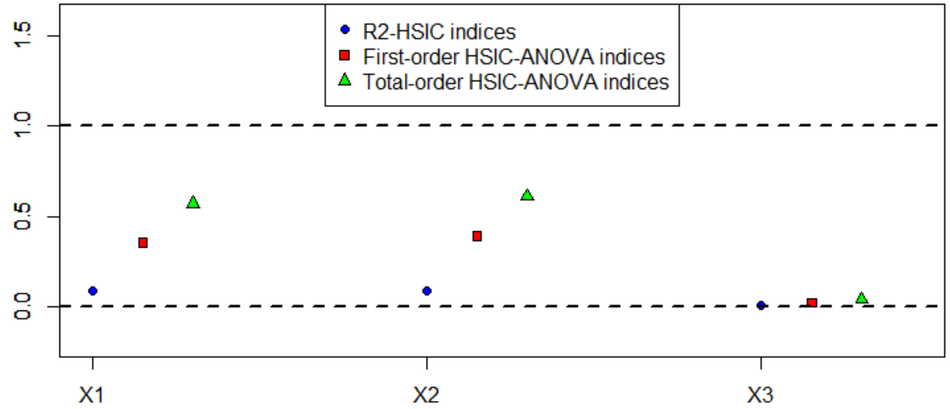
	U_1	U_2	U_3
R^2-HSIC	0.19	0.03	0.01
First-order	0.77	0.13	0.07
Total-order	0.79	0.14	0.08

How to exacerbate HSIC-ANOVA interactions?

- Back to the Ishigami function
  - ✓ Additional term chosen to boost HSIC-ANOVA interactions.

$$Y = g(U_1, U_2, U_3) = \text{ishigami}(X_1, X_2, X_3) + \boldsymbol{\gamma} \cos(\pi U_1) \cos(\pi U_2) \quad \text{with} \quad \begin{array}{c} U_i \sim \mathcal{U}([0,1]) \\ X_i = \pi(2U_i - 1) \end{array}$$

- Design parameter
  - ✓ $\boldsymbol{\gamma = 10}$
- Estimation of sensitivity measures
  - ✓ Sample size $n = 500$
  - ✓ R^2-HSIC indices + HSIC-ANOVA indices



	U_1	U_2	U_3
R^2-HSIC	0.05	0.08	0.01
First-order	0.25	0.40	0.02
Total-order	0.56	0.71	0.04

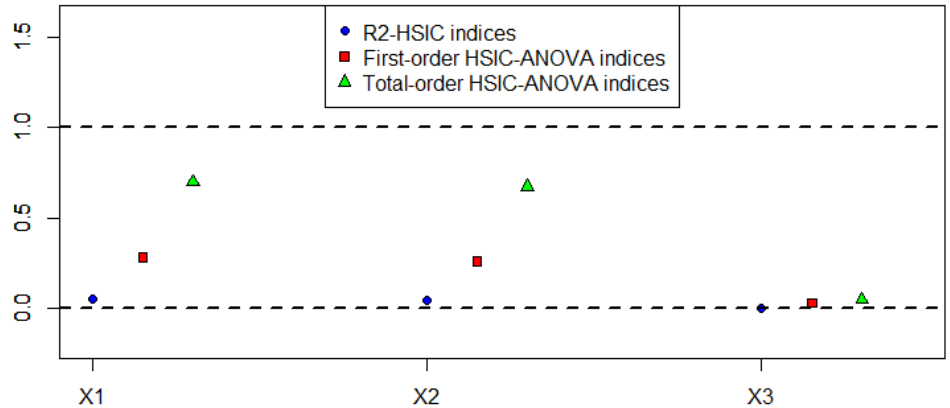
How to exacerbate HSIC-ANOVA interactions?

- Back to the Ishigami function
  - ✓ **Additional term** chosen to boost HSIC-ANOVA interactions.

$$Y = g(U_1, U_2, U_3) = \text{ishigami}(X_1, X_2, X_3) + \boldsymbol{\gamma} \cos(\pi U_1) \cos(\pi U_2) \quad \text{with} \quad \begin{array}{c} U_i \sim \mathcal{U}([0,1]) \\ X_i = \pi(2U_i - 1) \end{array}$$

- Design parameter
  - ✓ $\boldsymbol{\gamma = 100}$

- Estimation of sensitivity measures
  - ✓ Sample size $n = 500$
  - ✓ R^2-HSIC indices + HSIC-ANOVA indices



	U_1	U_2	U_3
R^2-HSIC	0.05	0.05	0.01
First-order	0.28	0.23	0.04
Total-order	0.72	0.66	0.05

How to use HSIC-ANOVA in practice?

1. How to build a test of independence? How to extend to the existing test procedures?
2. Is there any advantage to using the total-order HSIC-ANOVA index?

GSA requirements	T_i	$\mathrm{HSIC}(X_i, Y)$	S_i^{HSIC}	T_i^{HSIC}
ANOVA decomposition → RANKING	✓	✗	✓	✓
Characterize independence → SCREENING	✓	✓	✓ ?	✓ ?
Estimation from GIVEN DATA	✗	✓	✓	✓
Estimation from SMALL DATA	✗	✓	✓	✓
Compatibility with DEPENDENT inputs	✗	✓	✗	✗
INVARIANCE through monotonic transformations	✓	✗	✗	✗

6. Does all this benefit independence testing?

Testing independence with HSIC-ANOVA indices

- A **test of independence** consists in testing the **null hypothesis** $(H_0^i) : X_i \perp Y$.

$$X_i \perp Y \iff S_i^{\text{HSIC}} = 0 \iff T_i^{\text{HSC}} = 0$$

$$\iff \boxed{\begin{array}{c}\text{HSIC}(X_i, Y) = 0 \\ \text{with } K_{\text{Sob}}^1 \otimes K_Y\end{array}} \iff \boxed{\begin{array}{c}\text{HSIC}(\boldsymbol{X}, Y) - \text{HSIC}(\boldsymbol{X}_{-i}, Y) = 0 \\ \text{with } K_{\text{Sob}}^1 \otimes \ldots \otimes K_{\text{Sob}}^1 \otimes K_Y\end{array}}$$

Numerator of the first-order index

Numerator of the total-order index

Testing independence with HSIC-ANOVA indices

- A **test of independence** consists in testing the **null hypothesis** $(H_0^i) : X_i \perp Y$.

$$X_i \perp Y \iff S_i^{\text{HSIC}} = 0 \iff T_i^{\text{HSC}} = 0$$

$$\iff \boxed{\begin{array}{c}\text{HSIC}(X_i, Y) = 0 \\ \text{with } K_{\text{Sob}}^1 \otimes K_Y\end{array}} \iff \boxed{\begin{array}{c}\text{HSIC}(\boldsymbol{X}, Y) - \text{HSIC}(\boldsymbol{X}_{-i}, Y) = 0 \\ \text{with } K_{\text{Sob}}^1 \otimes \ldots \otimes K_{\text{Sob}}^1 \otimes K_Y\end{array}}$$

Numerator of the first-order index

Numerator of the total-order index

V-statistic estimator

$$\boxed{\widehat{S}_i(\boldsymbol{Z}_{\text{obs}}) = \widehat{\text{HSIC}}_v(X_i, Y)}$$

Apply existing test procedures with $K_i = K_{\text{Sob}}^1$

🤔 **Is there a reason to hope for higher statistical power?**

Actually, NO!

Testing independence with HSIC-ANOVA indices

- A **test of independence** consists in testing the **null hypothesis** $(H_0^i) : X_i \perp Y$.

$$X_i \perp Y \iff S_i^{\text{HSIC}} = 0 \iff T_i^{\text{HSC}} = 0$$

$$\iff \boxed{\begin{array}{c}\text{HSIC}(X_i, Y) = 0 \\ \text{with } K_{\text{Sob}}^1 \otimes K_Y\end{array}} \iff \boxed{\begin{array}{c}\text{HSIC}(\boldsymbol{X}, Y) - \text{HSIC}(\boldsymbol{X}_{-i}, Y) = 0 \\ \text{with } K_{\text{Sob}}^1 \otimes \ldots \otimes K_{\text{Sob}}^1 \otimes K_Y\end{array}}$$

Numerator of the first-order index

Numerator of the total-order index

V-statistic estimator

$$\widehat{S}_i(\boldsymbol{Z}_{\text{obs}}) = \widehat{\text{HSIC}}_v(X_i, Y)$$

Much more appealing!

Apply existing test procedures with $K_i = K_{\text{Sob}}^1$

Is there a reason to hope for higher statistical power?

Actually, NO!

V-statistic estimator

$$\widehat{T}_i = \widehat{\text{HSIC}}_v(\mathbf{X}, Y) - \widehat{\text{HSIC}}_v(\boldsymbol{X}_{-i}, Y)$$

Computing this test statistic is slightly more expensive.

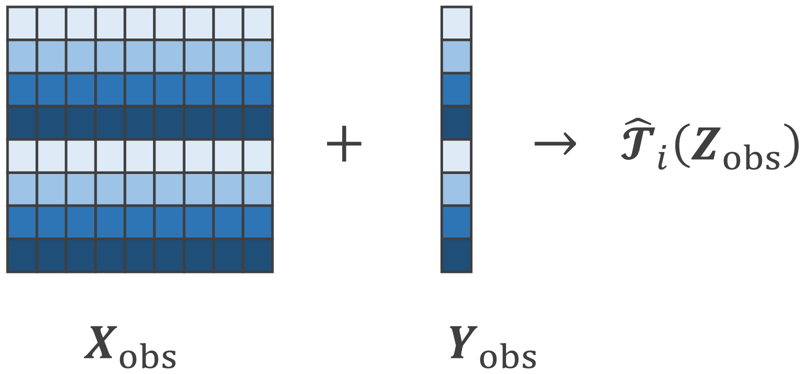
Is there a reason to hope for higher statistical power?

Let us see!

Testing independence with the total-order index

- The distribution of $\widehat{\mathcal{T}}_i(\boldsymbol{Z}_{\text{obs}})$ under (H_0^i) can be **simulated from the available data**.

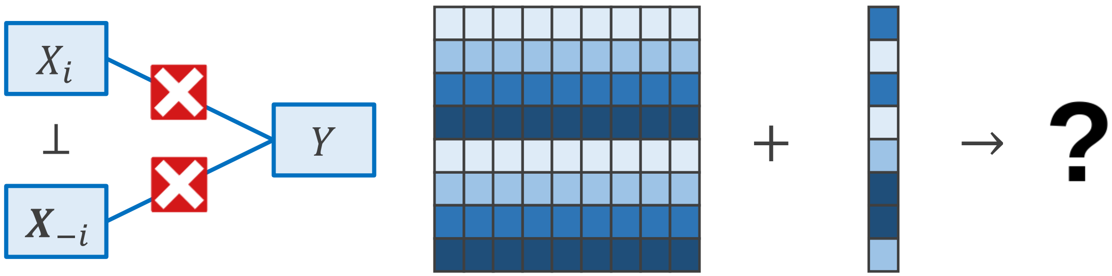
All the columns of the DoE are required to compute the test statistic.



Testing independence with the total-order index

- The distribution of $\widehat{\mathcal{T}}_i(\mathbf{Z}_{\text{obs}})$ under (H_0^i) can be **simulated from the available data**.

  All the columns of the DoE are required to compute the test statistic.



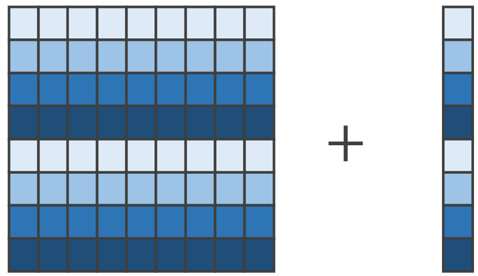
$\mathbf{X}_{\text{obs}}$ $\qquad$ $\mathbf{Y}_{\text{obs}}^{\sigma} := \left\{Y^{(\sigma(k))}\right\}_{1\leq k\leq n}$

- Permuting $\mathbf{Y}_{\text{obs}}$ leads to **eliminate dependence** between the joint observations $\left(\mathbf{X}^{(k)}, Y^{(k)}\right)$.
  - ✓ This boils down to testing $(H_0) : \mathbf{X} \perp Y$ and this is not what is desired!

Testing independence with the total-order index

- The distribution of $\widehat{\mathcal{T}}_i(\boldsymbol{Z}_{\mathrm{obs}})$ under (H_0^i) can be **simulated from the available data**.

Instead, the trick is to permute the observations of the input variable.

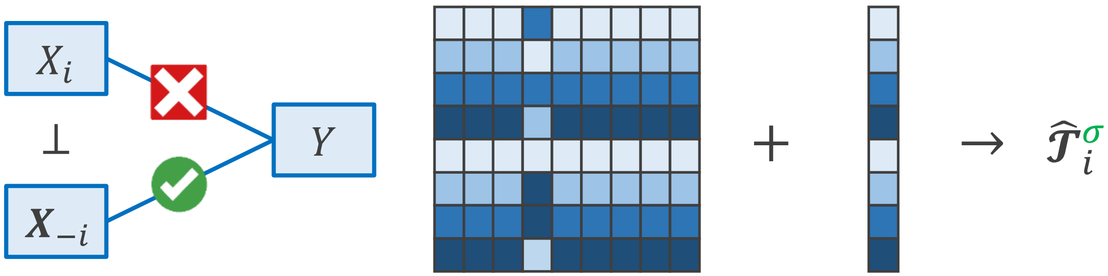


$$\boldsymbol{X}_{\mathrm{obs}} + \boldsymbol{Y}_{\mathrm{obs}} \rightarrow \widehat{\mathcal{T}}_i(\boldsymbol{Z}_{\mathrm{obs}})$$

Testing independence with the total-order index

- The distribution of $\widehat{\mathcal{T}}_i(\boldsymbol{Z}_{\text{obs}})$ under (H_0^i) can be **simulated from the available data**.

  Instead, the trick is to permute the observations of the input variable.



$$\boldsymbol{X}_{\text{obs}}^{\sigma} \coloneqq \left\{\left(X_i^{(\sigma(k))}, \boldsymbol{X}_{-i}^{(k)}\right)\right\}_k \qquad \boldsymbol{Y}_{\text{obs}}$$

Testing independence with the total-order index

- The distribution of $\widehat{\mathcal{T}}_i(\boldsymbol{Z}_{\mathrm{obs}})$ under (H_0^i) can be **simulated from the available data**.

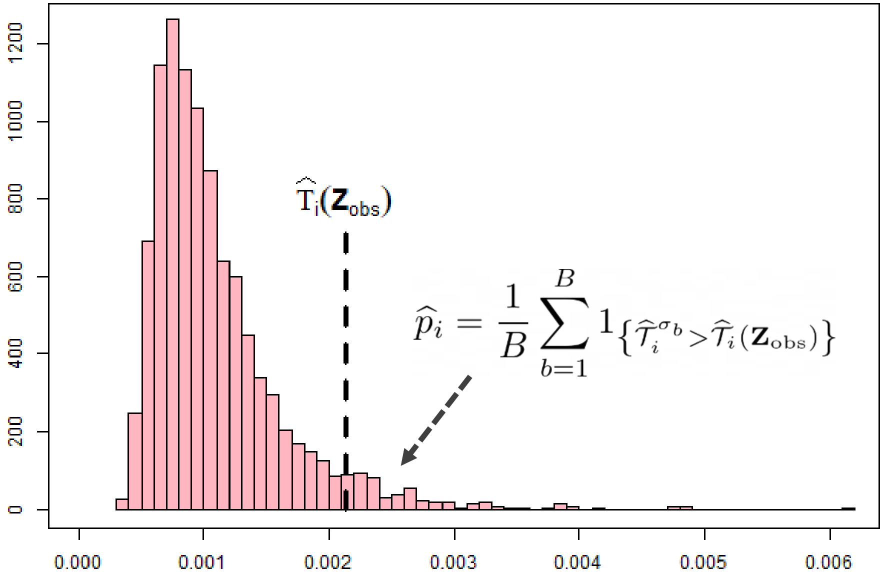
Instead, the trick is to permute the observations of the input variable.

Permutation-based test procedure

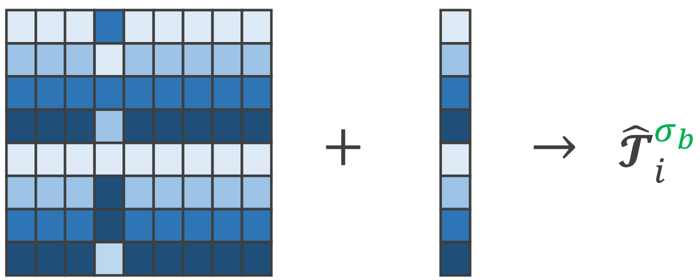
- <u>**Step A**</u> → Perform a sequence $\{\sigma_b\}_{1\leq b\leq B}$ of **random permutations** on the i-th column of $\boldsymbol{X}_{\mathrm{obs}}$.
- <u>**Step B**</u> → Compute the value $\widehat{\mathcal{T}}_i^{\sigma_b}$ of the test statistic for each permuted design.
- <u>**Step C**</u> → Derive a **non-parametric estimate** of the p-value $p_i := \mathbb{P}\left(\widehat{\mathcal{T}}_i > \widehat{\mathcal{T}}_i(\boldsymbol{Z}_{\mathrm{obs}})\right)$.

- **Default value:** $B \approx 10^3$
- **Complexity:** $(d^2 + 7Bd)\, n^2$

Simulation of the test statistic under the null hypothesis



Permutation scheme



$\boldsymbol{X}_{\mathrm{obs}}^{\sigma_b} := \left\{\left(X_i^{(\sigma_b(k))}, \boldsymbol{X}_{-i}^{(k)}\right)\right\}_k$ $\quad \boldsymbol{Y}_{\mathrm{obs}}$

Numerical study of the statistical power

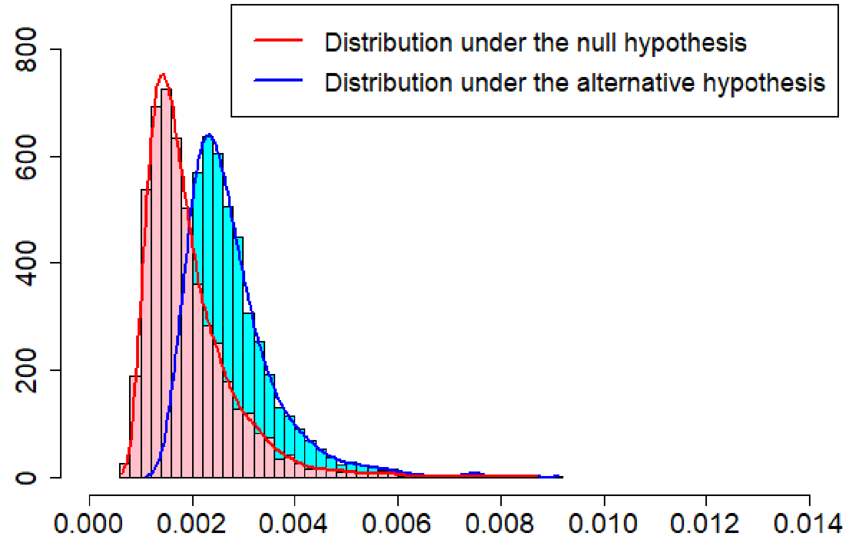
- Back to the Ishigami function
  - ✓ Additional term chosen to boost HSIC-ANOVA interactions.

$$Y = g(U_1, U_2, U_3) = \text{ishigami}(X_1, X_2, X_3) + \boldsymbol{\gamma} \cos(\pi U_1) \cos(\pi U_2) \quad \text{with} \quad \begin{matrix} U_i \sim \boldsymbol{\mathcal{U}}([0,1]) \\ X_i = \pi(2U_i - 1) \end{matrix}$$

- Design parameter
  - ✓ $\boldsymbol{\gamma = 0}$
- Study of the statistical power
  - ✓ Sample size $n = 50$
  - ✓ Number of replicates $M = 200$

	U_1	U_2	U_3
HSIC	0.88	0.07	0.22
Total-order	0.87	0.19	0.19

- Separation rate
  - ✓ Distributions of $\widehat{\boldsymbol{\mathcal{T}}}_i(\boldsymbol{Z}_{\text{obs}})$ under (H_0^i) et (H_1^i)



Numerical study of the statistical power

- **Back to the Ishigami function**
  - ✓ **Additional term** chosen to boost HSIC-ANOVA interactions.

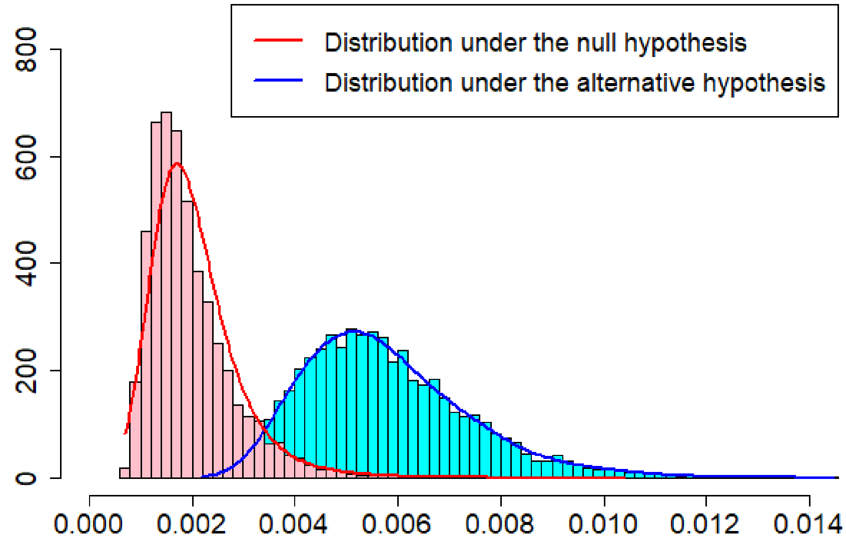
$$Y = g(U_1, U_2, U_3) = \text{ishigami}(X_1, X_2, X_3) + \boldsymbol{\gamma} \cos(\pi U_1)\cos(\pi U_2) \quad \text{with} \quad \begin{array}{c} U_i \sim \mathcal{U}([0,1]) \\ X_i = \pi(2U_i - 1) \end{array}$$

- **Design parameter**
  - ✓ $\boldsymbol{\gamma = 10}$

- **Study of the statistical power**
  - ✓ **Sample size** $n = 50$
  - ✓ **Number of replicates** $M = 200$

	U_1	U_2	U_3
HSIC	0.59	0.63	0.05
Total-order	0.92	0.94	0.07

- **Separation rate**
  - ✓ Distributions of $\widehat{\mathcal{T}}_i(\mathbf{Z}_{\text{obs}})$ under (H_0^i) et (H_1^i)



✅ **Increased power** when $S_i^{\text{HSIC}} \ll T_i^{\text{HSIC}}$

✅ **Same power** when $S_i^{\text{HSIC}} \approx T_i^{\text{HSIC}}$

Numerical study of the statistical power

- Back to the Ishigami function
  - ✓ <u>Additional term</u> chosen to boost HSIC-ANOVA interactions.

$$Y = g(U_1, U_2, U_3) = \text{ishigami}(X_1, X_2, X_3) + \boldsymbol{\gamma} \cos(\pi U_1)\cos(\pi U_2) \quad \text{with} \quad \begin{matrix} U_i \sim \mathcal{U}([0,1]) \\ X_i = \pi(2U_i - 1) \end{matrix}$$

- Design parameter
  - ✓ $\boldsymbol{\gamma = 100}$

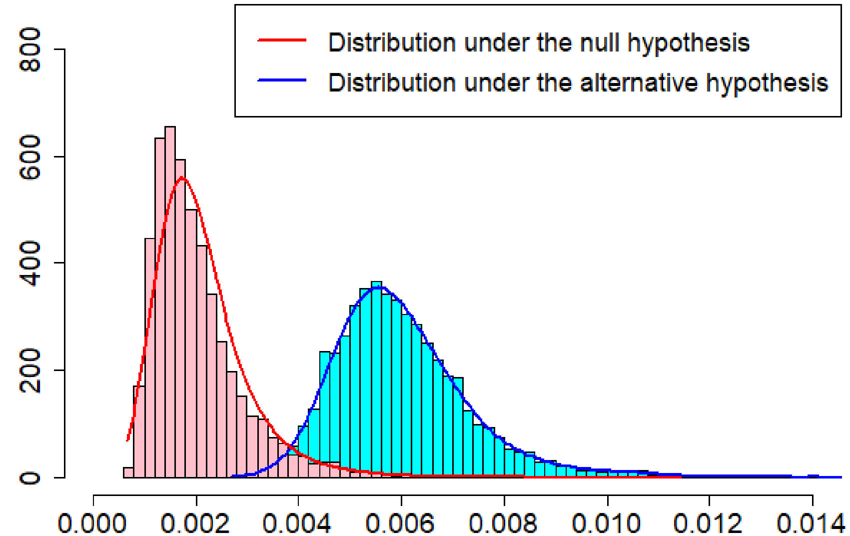
- Study of the statistical power
  - ✓ Sample size $n = 50$
  - ✓ Number of replicates $M = 200$

	U_1	U_2	U_3
HSIC	0.65	0.70	0.07
Total-order	1.00	1.00	0.06

✓ Increased power when $S_i^{\text{HSIC}} \ll T_i^{\text{HSIC}}$

✓ Same power when $S_i^{\text{HSIC}} \approx T_i^{\text{HSIC}}$

- Separation rate
  - ✓ Distributions of $\widehat{\mathcal{T}}_i(\mathbf{Z}_{\text{obs}})$ under (H_0^i) et (H_1^i)



Benefits brought by HSIC-ANOVA indices in GSA

- ✔ HSIC-ANOVA indices are <u>**fully transparent**</u> sensitivity measures **able to perform screening and ranking!**
- ✔ In many situations, the test of independence based on T_i^{HSIC} is **more powerful!**

GSA requirements	T_i	$\mathrm{HSIC}(X_i, Y)$	S_i^{HSIC}	T_i^{HSIC}
ANOVA decomposition → RANKING	✔	✘	✔	✔
Characterize independence → SCREENING	✔	✔	✔	✔ ✔
Estimation from GIVEN DATA	✘	✔	✔	✔
Estimation from SMALL DATA	✘	✔	✔	✔
Compatibility with DEPENDENT inputs	✘	✔	✘	✘
INVARIANCE through monotonic transformations	✔	✘	✘	✘

7. Conclusion

Conclusion

- The very recent **HSIC-ANOVA indices** have enabled **significant progress in GSA** since they combine the advantages of **Sobol' indices** (variance-based GSA) and those of **HSIC indices** (kernel-based GSA).

HSIC - ANOVA

HSIC-like properties

- ✓ **Low estimation cost.**
- ✓ **Characterization of independence.**
- ✓ **Unified framework to handle continuous and discrete variables.**

Sobol'-like properties

- ✓ **Interpretation as percentages of influence.**
- ✓ **Main effects vs. Interaction effects.**
- • **Need for mutually independent inputs**

Conclusion

- The very recent **HSIC-ANOVA indices** have enabled **significant progress in GSA** since they combine the advantages of **Sobol' indices** (variance-based GSA) and those of **HSIC indices** (kernel-based GSA).

- The **HSIC-ANOVA decomposition** requires the use of **characteristic ANOVA kernels** for the input variables.
  - ✓ For the standard uniform distribution, it is recommended to take the **Sobolev kernel** K^1_{Sob}.
  - ✓ For other distributions, **orthogonalization techniques** can be used to build suitable kernels.

Conclusion

- The very recent **HSIC-ANOVA indices** have enabled **significant progress in GSA** since they combine the advantages of **Sobol' indices** (variance-based GSA) and those of **HSIC indices** (kernel-based GSA).

- The **HSIC-ANOVA decomposition** requires the use of **characteristic ANOVA kernels** for the input variables.

- The way sensitivity is measured by HSIC-ANOVA indices is driven by the kernel feature maps.
  - ✓ The **first-order** index S_1^{HSIC} scans **all possible dependence patterns** between X_1 and Y.
  - ✓ The **second-order** index S_{12}^{HSIC} also scans **all possible dependence patterns** between (X_1, X_2) and Y.

Conclusion

- The very recent **HSIC-ANOVA indices** have enabled **significant progress in GSA** since they combine the advantages of **Sobol' indices** (variance-based GSA) and those of **HSIC indices** (kernel-based GSA).

- The **HSIC-ANOVA decomposition** requires the use of **characteristic ANOVA kernels** for the input variables.

- The way sensitivity is measured by HSIC-ANOVA indices is driven by the kernel feature maps.

- **Variable selection** can be performed with **test procedures** based on HSIC-ANOVA indices.
  - ✓ For the **first-order** index S_1^{HSIC} → **The existing test procedures can be applied directly.**
  - ✓ For the **total-order** index T_1^{HSIC} → **The existing test procedures need to be adapted!**

Conclusion

- The very recent **HSIC-ANOVA indices** have enabled **significant progress in GSA** since they combine the advantages of **Sobol' indices** (variance-based GSA) and those of **HSIC indices** (kernel-based GSA).
- The **HSIC-ANOVA decomposition** requires the use of **characteristic ANOVA kernels** for the input variables.
- The way sensitivity is measured by HSIC-ANOVA indices is driven by the kernel feature maps.
- **Variable selection** can be performed with **test procedures** based on HSIC-ANOVA indices.
- Using the total-order HSIC-ANOVA indices leads to **more powerful** test procedures.

Traditional benchmarks

- ✓ **Ishigami, Friedman, Morris…**

$$\boxed{S_i^{HSIC} \lesssim T_i^{HSIC}}$$

$$\boxed{\mathrm{Power}\big(\widehat{\mathcal{S}}_i\big) \approx \mathrm{Power}\big(\widehat{\mathrm{HSIC}}_{\mathcal{N}}\big) \approx \mathrm{Power}\big(\widehat{\mathcal{T}}_i\big)}$$

Specific benchmarks

- ✓ **Hand-made use cases.**
- ✓ **Test functions in optimization.**
- ✓ **Flexible metafunction framework.**

$$\boxed{S_i^{HSIC} \ll T_i^{HSIC}}$$

$$\boxed{\mathrm{Power}\big(\widehat{\mathcal{S}}_i\big) \ll \mathrm{Power}\big(\widehat{\mathrm{HSIC}}_{\mathcal{N}}\big) \ll \mathrm{Power}\big(\widehat{\mathcal{T}}_i\big)}$$

Conclusion

- The very recent **HSIC-ANOVA indices** have enabled **significant progress in GSA** since they combine the advantages of **Sobol' indices** (variance-based GSA) and those of **HSIC indices** (kernel-based GSA).
- The **HSIC-ANOVA decomposition** requires the use of **characteristic ANOVA kernels** for the input variables.
- The way sensitivity is measured by HSIC-ANOVA indices is driven by the kernel feature maps.
- **Variable selection** can be performed with **test procedures** based on HSIC-ANOVA indices.
- Using the total-order HSIC-ANOVA indices leads to **more powerful** test procedures.

Publications

- **Preprint** → https://cea.hal.science/cea-04320711/document
- **Conference paper** → https://cea.hal.science/cea-03701170v1/document

Codes

- **Two dedicated routines the R package sensitivity**
  - ✓ **sensiHSIC** → https://rdrr.io/cran/sensitivity/man/sensiHSIC.html
  - ✓ **testHSIC** → https://rdrr.io/cran/sensitivity/man/testHSIC.html

10. References

Key papers (1/4)

- **Baldeaux, J. & Dick, J. (2009).** QMC rules of arbitrary high order: reproducing kernel Hilbert space approach. *Constructive Approximation*, *30*, 495-527.
- **Becker W. (2020).** Metafunctions for benchmarking in sensitivity analysis, *Reliability Engineering & System Safety*, 204, 107189.
- **Berlinet, A. & Thomas-Agnan, C. (2011).** *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media.
- **Broto, B., Bachoc, F. & Depecker, M. (2020).** Variance reduction for estimation of Shapley effects and adaptation to unknown input distribution. *SIAM/ASA Journal on Uncertainty Quantification*, *8*(2), 693-716.
- **Chatterjee, S. (2021).** A new coefficient of correlation. *Journal of the American Statistical Association*, *116*(536), 2009-2022.
- **Da Veiga, S. (2015).** Global sensitivity analysis with dependence measures. *Journal of Statistical Computation and Simulation*, *85*(7), 1283-1305.
- **Da Veiga, S. (2021).** Kernel-based ANOVA decomposition and Shapley effects - Application to global sensitivity analysis. Preprint arXiv:2101.05487.
- **Da Veiga, S., Gamboa, F., Lagnoux, A., Klein, T. & Prieur, C. (2023).** New estimation of Sobol' indices using kernels. Preprint *arXiv:2303.17832*.
- **De Lozzo, M. & Marrel, A. (2016).** New improvements in the use of dependence measures for sensitivity analysis and screening. *Journal of Statistical Computation and Simulation*, *86*(15), 3038-3058.
- **Devroye, L., Györfi, L., Lugosi, G. & Walk, H. (2018).** A nearest neighbor estimate of the residual variance. *Electronic Journal of Statistics*, *12*, 1752-1778.
- **Dick, J., Nuyens, D. & Pillichshammer, F. (2014).** Lattice rules for nonperiodic smooth integrands. *Numerische Mathematik*, *126*, 259-291.
- **Dick, J., Hinrichs, A. & Pillichshammer, F. (2015).** Proof techniques in quasi-Monte Carlo theory. *Journal of Complexity*, *31*(3), 327-371.

Key papers (2/4)

- **El Amri, M. R. & Marrel, A. (2022).** Optimized HSIC-based tests for sensitivity analysis: Application to thermal-hydraulic simulation of accidental scenario on nuclear reactor. *Quality and Reliability Engineering International*, *38*(3), 1386-1403.
- **El Amri, M. R. & Marrel, A. (2023).** More powerful HSIC-based independence tests, extension to space-filling designs and functional data. Preprint available at: https://cea.hal.science/cea-03406956/
- **Gamboa, F., Janon, A., Klein, T., Lagnoux, A. & Prieur, C. (2016).** Statistical inference for Sobol' pick-freeze Monte Carlo method. *Statistics*, *50*(4), 881-902.
- **Gamboa, F., Gremaud, P., Klein, T. & Lagnoux, A. (2022).** Global sensitivity analysis: A novel generation of mighty estimators based on rank statistics. *Bernoulli*, *28*(4), 2345-2374.
- **Ginsbourger, D., Roustant, O., Schuhmacher, D., Durrande, N. & Lenz, N. (2016).** On ANOVA decompositions of kernels and Gaussian random field paths. In *Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Leuven, Belgium, April 2014* (pp. 315-330). Springer International Publishing.
- **Gretton, A., Bousquet, O., Smola, A., & Schölkopf, B. (2005).** Measuring statistical dependence with Hilbert-Schmidt norms. In *Algorithmic Learning Theory: 16th International Conference, ALT 2005, Singapore, October 8-11, 2005. Proceedings 16* (pp. 63-77). Springer Berlin Heidelberg.
- **Gretton, A., Borgwardt, K., Rasch, M., Schölkopf, B. & Smola, A. (2006).** A kernel method for the two-sample-problem. *Advances in Neural Information Processing Systems*, *19*.
- **Gretton, A., Fukumizu, K., Teo, C., Song, L., Schölkopf, B. & Smola, A. (2007).** A kernel statistical test of independence. *Advances in Neural Information Processing Systems*, *20*.
- **Gu, C. (2013).** *Smoothing Spline ANOVA Models* (Vol. 297). Springer Science & Business Media.
- **Hoeffding, W. (1992).** A class of statistics with asymptotically normal distribution. *Breakthroughs in Statistics: Foundations and Basic Theory*, 308-334.
- **Iooss, B. & Marrel, A. (2019).** Advanced methodology for uncertainty propagation in computer experiments with large number of inputs. *Nuclear Technology*, *205*(12), 1588-1606.

Key papers (3/4)

- **Janon, A., Klein, T., Lagnoux, A., Nodet, M. & Prieur, C. (2014).** Asymptotic normality and efficiency of two Sobol' index estimators. *ESAIM: Probability and Statistics*, *18*, 342-364.
- **Kuo, F., Sloan, I., Wasilkowski, G. & Woźniakowski, H. (2010).** On decompositions of multivariate functions. *Mathematics of computation*, *79*(270), 953-966.
- **Marrel, A., Iooss, B. & Chabridon, V. (2022).** The ICSCREAM methodology: Identification of penalizing configurations in computer experiments using screening and metamodel - Applications in thermal-hydraulics. *Nuclear Science and Engineering*, *196*(3), 301-321.
- **Mazo, G. & Tournier, L. (2023).** An inference method for global sensitivity analysis.
- **McKay, M. D. (1996).** Variance-based methods for assessing uncertainty importance in NUREG-1150 analyses. *Los Alamos National Laboratory LA-UR-96-2695*.
- **Muandet, K., Fukumizu, K., Sriperumbudur, B. & Schölkopf, B. (2017).** Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends in Machine Learning*, *10*(1-2), 1-141.
- **Saltelli, A. (2002).** Making best use of model evaluations to compute sensitivity indices. *Computer physics communications*, *145*(2), 280-297.
- **Sarazin, G., Marrel, A., Da Veiga, S. & Chabridon, V. (2022, June).** Test d'indépendance basé sur les indices HSIC-ANOVA d'ordre total. In *53èmes Journées de Statistique de la SFdS*.
- **Serfling, R. J. (2009).** *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons.
- **Sobol', I. M. (1993).** Sensitivity analysis for non-linear mathematical models. *Mathematical Modeling and Computational Experiment*, *1*, 407-414.
- **Sobol, I. M. (2001).** Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, *55*(1-3), 271-280.

Key papers (4/4)

- **Steinwart, I. & Christmann, A. (2008).** *Support Vector Machines*. Springer Science & Business Media.
- **Wahba, G. (1990).** *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics.
- **Zhang, Q., Filippi, S., Gretton, A. & Sejdinovic, D. (2018).** Large-scale kernel methods for independence testing. *Statistics and Computing*, *28*, 113-130.