



**HAL**  
open science

# Test d'indépendance basé sur les indices HSIC-ANOVA d'ordre total

Gabriel Sarazin, Amandine Marrel, Sébastien da Veiga, Vincent Chabridon

► **To cite this version:**

Gabriel Sarazin, Amandine Marrel, Sébastien da Veiga, Vincent Chabridon. Test d'indépendance basé sur les indices HSIC-ANOVA d'ordre total. 53èmes Journées de Statistique de la SFdS, Société Française de Statistique (SFdS); Université Claude Bernard Lyon 1, Jun 2022, Lyon, France. cea-03701170

**HAL Id: cea-03701170**

**<https://hal-cea.archives-ouvertes.fr/cea-03701170>**

Submitted on 21 Jun 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# TEST D'INDÉPENDANCE BASÉ SUR LES INDICES HSIC-ANOVA D'ORDRE TOTAL

Gabriel Sarazin<sup>1</sup>, Amandine Marrel<sup>1-2</sup>, Sébastien Da Veiga<sup>3</sup> & Vincent Chabridon<sup>4</sup>

1. *CEA, DES, IRESNE, DER, Cadarache, 13108, Saint-Paul-Lez-Durance.*

2. *Institut de Mathématiques de Toulouse, 118 Route de Narbonne, 31062, Toulouse.*

3. *Safran Tech, Modeling & Simulation, Rue des Jeunes Bois, Châteaufort, 78114, Magny-Les-Hameaux.*

4. *EDF R&D, 6 Quai Watier, 78401, Chatou.*

**Résumé.** L'apprentissage statistique dans le cas de données simulées par un code de calcul industriel, aussi appelé "métamodélisation", est une tâche dont la difficulté de mise en œuvre croît avec la dimension du problème et le manque de données d'apprentissage. Une analyse de sensibilité préliminaire peut venir en soutien de la construction du métamodèle pour éliminer les variables les moins pertinentes et trier les variables restantes par ordre d'influence sur la sortie. Pour mener une analyse de sensibilité, l'approche historique de Sobol' offre un cadre conceptuel confortable qui est articulé autour de la décomposition de la variance de la sortie. Toutefois, l'estimation précise des indices associés n'est plus possible si l'échantillon d'apprentissage est de petite taille. Pour contourner cette difficulté, il est désormais fréquent d'utiliser une mesure de sensibilité basée sur le critère d'indépendance de Hilbert-Schmidt (notée HSIC). Elle est appliquée à chaque couple entrée-sortie, et permet ainsi de définir la collection des indices HSIC. Leur interprétation est généralement moins intuitive que celle des indices de Sobol car leur construction repose sur la théorie des espaces de Hilbert à noyaux reproduisants. Face à ce constat, les indices HSIC-ANOVA ont été récemment introduits et permettent une séparation des effets principaux et des interactions, à l'instar de la décomposition de Hoeffding dans le formalisme des indices de Sobol'. Cette avancée a été obtenue au prix d'une hypothèse d'indépendance mutuelle des entrées et sous réserve de l'utilisation de noyaux spécifiques, comme les noyaux de Sobolev. Dans ce travail, on commence par montrer que tout noyau de Sobolev est caractéristique, c'est-à-dire que la nullité d'un indice HSIC-ANOVA est équivalente à une situation d'indépendance au sein du couple formé par l'entrée considérée et la sortie. Dans un second temps, il est montré qu'un test d'indépendance peut être construit pour l'indice HSIC-ANOVA d'ordre total en s'inspirant de ce qui est fait pour l'indice HSIC traditionnel. Enfin, une étude numérique révèle empiriquement que le nouveau test d'indépendance est au moins aussi puissant que celui basé sur l'indice HSIC traditionnel, ce qui offre des perspectives intéressantes pour améliorer le processus de sélection des variables à impliquer dans la construction d'un métamodèle.

**Mots-clés.** Analyse de sensibilité, criblage, critère d'indépendance de Hilbert-Schmidt, test d'indépendance, analyse de la variance.

---

E-mails :

gabriel.sarazin@cea.fr  
sebastien.da-veiga@safrangroup.com

amandine.marrel@cea.fr  
vincent.chabridon@edf.fr

**Abstract.** Building a surrogate model for an industrial computationally-expensive simulation code is made difficult by the combined effect of the curse of dimensionality and the lack of input-output data. A preliminary sensitivity analysis may help discard non-influential inputs and rank the remaining inputs according to their impact on the output distribution. In order to perform sensitivity analysis, the historical approach proposed by Sobol' provides a convenient conceptual framework where the output variance is apportioned between input variables. However, the accurate estimation of the corresponding indices requires thousands of model evaluations, which is often unaffordable in an industrial context. To circumvent this pitfall, it has become quite common to resort to a sensitivity measure based on Hilbert-Schmidt independence criterion (denoted by HSIC). This measure is applied to all input-output pairs of variables and allows to define the so-called "HSIC indices". Their interpretation is much less intuitive than the one related to Sobol' indices since their foundations come from the theory of reproducing kernel Hilbert spaces. To ease interpretation, the HSIC-ANOVA indices have been recently introduced to allow for a strict separation of main effects and interactions, akin to what is proposed in Sobol' formalism with Hoeffding decomposition. This breakthrough was obtained after assuming mutual independence between inputs and provided that specific kernels, like Sobolev kernels, are used to compute HSIC-ANOVA indices. In this work, a first contribution consists in demonstrating that Sobolev kernels are characteristic. Because of this property, independence within input-output pairs of variables can be detected from the observed values of HSIC-ANOVA indices. Then, it is shown that a test of independence can be constructed for the total-order HSIC-ANOVA index after adapting existing methodologies in the HSIC-related literature. Finally, an extensive simulation study proves empirically that the newly-developed test of independence is, at least, as powerful as the older one based on the traditional HSIC index, which offers interesting prospects in order to improve the screening step performed before metamodeling.

**Keywords.** Sensitivity analysis, screening, Hilbert-Schmidt independence criterion, test of independence, analysis of variance.

# 1 Généralités sur l'analyse de sensibilité

## 1.1 Propagation d'incertitudes en milieu industriel

L'étude des performances d'un système complexe (comme une centrale nucléaire, un avion, une plate-forme pétrolière, un ouvrage d'art, etc.) requiert l'utilisation d'un code de simulation qui modélise les phénomènes physiques mis en jeu et reproduit les contraintes exercées sur le système. Ce code peut être vu comme une fonction déterministe  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  de type "boîte noire" dont les entrées  $\mathbf{x} := [x_1, \dots, x_d]$  sont des variables physiques incertaines et dont la sortie  $y$  est une variable d'intérêt qui résume l'état du système (valeur maximale d'une température, d'une pression, d'une déformation, etc.) et guide une éventuelle prise de décision. Dans un cadre probabiliste, les entrées du

code sont décrites par un vecteur aléatoire  $\mathbf{X}$  de loi conjointe  $\mathbb{P}_{\mathbf{X}}$  et de lois marginales  $\mathbb{P}_{X_1}, \dots, \mathbb{P}_{X_d}$ . La propagation des incertitudes à travers le code  $g$  conduit à l'obtention d'une loi de probabilité  $\mathbb{P}_Y$  en sortie. Comme chaque appel au code de calcul est très coûteux, la connaissance du fonctionnement interne de  $g$  se limite à un échantillon de simulations  $\mathbf{Z}_{\text{obs}} = \{(\mathbf{X}^{(k)}, Y^{(k)})\}_{k=1}^n$  appariant un petit nombre  $n$  d'entrées et de sorties de la boîte noire ( $n \approx 10d$ ). L'apprentissage statistique vise à construire un modèle de substitution (ou métamodèle)  $\hat{g}$  en s'appuyant sur la seule connaissance de l'échantillon  $\mathbf{Z}_{\text{obs}}$ . Lorsque la dimension augmente (plusieurs dizaines, voire une centaine d'entrées), les techniques usuelles de métamodélisation ne peuvent plus être appliquées car elles impliquent l'optimisation d'un trop grand nombre de paramètres et d'hyperparamètres du métamodèle  $\hat{g}$ . Face à ce problème, mener une analyse de sensibilité préliminaire permet d'écartier les entrées dont l'influence sur la sortie est négligeable (phase de **criblage**) et de disposer d'un classement des entrées restantes (phase de **hiérarchisation**) qui peut être exploité a posteriori si la construction du métamodèle obéit à une heuristique séquentielle. Cette approche est notamment mise en œuvre dans [Marrel et al. \(2021\)](#).

## 1.2 Émergence et utilisation grandissante des indices HSIC

Pour analyser la sensibilité de la sortie  $Y$  à des entrées  $X_1, \dots, X_d$  supposées indépendantes, l'approche fondatrice de Sobol s'appuie sur la décomposition de Hoeffding (ou "ANOVA fonctionnelle") de  $g$  afin de répartir la variance  $\mathbb{V}(Y)$  entre les différentes entrées contributrices. L'indice de Sobol' du premier ordre  $S_i$  quantifie la part de variance expliquée uniquement à partir de l'incertitude portée par  $X_i$ . L'indice de Sobol' d'ordre total  $S_i^T$  s'obtient en agglomérant l'effet principal et les interactions avec les autres entrées :

$$\forall 1 \leq i \leq d, S_i := \frac{\mathbb{V}(\mathbb{E}[Y | X_i])}{\mathbb{V}(Y)} \text{ et } S_i^T := 1 - \frac{\mathbb{V}(\mathbb{E}[Y | \mathbf{X}_{-i}])}{\mathbb{V}(Y)} \text{ avec } \mathbf{X}_{-i} := [X_k]_{\substack{1 \leq k \leq d \\ k \neq i}}.$$

La grande popularité de cette méthodologie ANOVA s'explique principalement par l'interprétation intuitive qui en découle, les indices de Sobol' s'apparentant à des pourcentages de  $\mathbb{V}(Y)$ . Toutefois, l'utilisation pratique de ces indices n'est pas toujours possible car leur définition mathématique à partir de variances conditionnelles implique des schémas d'estimation trop gourmands en termes de nombre d'appels à  $g$ . Pour remédier à cela, les indices HSIC (*Hilbert-Schmidt Independence Criterion*) de [Gretton et al. \(2005\)](#) ont été repris par [Da Veiga \(2015\)](#) pour l'analyse de sensibilité des simulateurs numériques. Ces indices sont d'une tout autre nature puisqu'ils puisent leurs fondements mathématiques dans la théorie des espaces de Hilbert à noyaux reproduisants. L'indice HSIC( $X_i, Y$ ) peut s'écrire comme une mesure de dissimilarité entre la loi conjointe  $\mathbb{P}_{X_i Y}$  décrivant la nature de la dépendance stochastique au sein de la paire  $(X_i, Y)$  et la loi  $\mathbb{P}_{X_i} \otimes \mathbb{P}_Y$  obtenue en cas d'indépendance parfaite. Pour définir cette mesure de dissimilarité, il faut commencer par affecter des noyaux  $K_1, \dots, K_d$  et  $K_Y$  aux entrées et à la sortie. On rappelle que pour tout vecteur aléatoire  $\mathbf{Z}$  à valeurs dans  $\mathcal{Z}$  et de loi  $\mathbb{P}_{\mathbf{Z}}$ , tout noyau  $K_{\mathbf{Z}} : \mathcal{Z} \times \mathcal{Z} \mapsto \mathbb{R}$

est en relation directe avec un espace fonctionnel  $\mathcal{H}_{\mathbf{Z}}$  appelé RKHS (*Reproducing Kernel Hilbert Space*) qui est susceptible d'offrir une certaine représentation de la loi  $\mathbb{P}_{\mathbf{Z}}$ . L'indice  $\text{HSIC}(X_i, Y)$  rend alors compte de la distance entre les représentants des lois  $\mathbb{P}_{X_i Y}$  et  $\mathbb{P}_{X_i} \otimes \mathbb{P}_Y$  dans le RKHS tensorisé  $\mathcal{H}_{X_i} \otimes \mathcal{H}_Y$ . Il est montré dans [Gretton et al. \(2007\)](#) que la quantité  $\text{HSIC}(X_i, Y)$  fait uniquement intervenir les espérances de variables aléatoires définies à partir de deux copies indépendantes  $(X_i, Y)$  et  $(X'_i, Y')$  de la loi  $\mathbb{P}_{X_i Y}$  :

$$\begin{aligned} \text{HSIC}(X_i, Y) &= \mathbb{E}_{\substack{X_i Y \\ X'_i Y'}} [K_i(X_i, X'_i) K_Y(Y, Y')] + \mathbb{E}_{X_i X'_i} [K_i(X_i, X'_i)] \mathbb{E}_{Y Y'} [K_Y(Y, Y')] \\ &\quad - 2 \mathbb{E}_{X_i Y} [\mathbb{E}_{X'_i} [K_i(X_i, X'_i)] \mathbb{E}_{Y'} [K_Y(Y, Y')]] . \end{aligned}$$

Un tel indice peut donc être estimé à l'aide de la V-statistique suivante :

$$\begin{aligned} \widehat{\text{HSIC}}_i(\mathbf{Z}_{\text{obs}}) &:= \frac{1}{n^2} \text{Tr}(\mathbf{L}_i \mathbf{H} \mathbf{L}_Y \mathbf{H}) \quad \text{avec} \quad \mathbf{L}_i := \left[ K_i(X_i^{(s)}, X_i^{(t)}) \right]_{1 \leq s, t \leq n} , \\ \mathbf{L}_Y &:= \left[ K_Y(Y^{(s)}, Y^{(t)}) \right]_{1 \leq s, t \leq n} \quad \text{et} \quad \mathbf{H} := [\delta_{st} - 1/n]_{1 \leq s, t \leq n} . \end{aligned}$$

En pratique, on obtient une précision suffisante dès lors que  $n \geq 100$ . L'indice  $\text{HSIC}(X_i, Y)$  correspond en fait à la norme de Hilbert-Schmidt d'un hyper-opérateur de covariance qui amalgame tout le catalogue des transformations constitutives (ou *feature functions*) des RKHS  $\mathcal{H}_{X_i}$  et  $\mathcal{H}_Y$ . Ainsi, lorsque les noyaux  $K_1, \dots, K_d$  et  $K_Y$  sont caractéristiques, le spectre des motifs de dépendance détectables devient suffisamment large pour que la nullité de  $\text{HSIC}(X_i, Y)$  traduise l'indépendance parfaite de  $X_i$  et  $Y$ .

Comme l'estimation de  $\text{HSIC}(X_i, Y)$  à partir de  $\mathbf{Z}_{\text{obs}}$  ne donne jamais zéro, et ce quelle que soit la nature de l'estimateur employé, on a besoin d'une procédure statistique pour tester l'hypothèse  $(H_0) : \text{HSIC}(X_i, Y) = 0$ . La loi asymptotique de l'estimateur  $\widehat{\text{HSIC}}_i(\mathbf{Z}_{\text{obs}})$  sous  $(H_0)$  étant connue, il est possible de s'appuyer sur cette statistique pour construire un test d'indépendance dont la conclusion doit aiguiller le choix de sélectionner ou d'évincer  $X_i$  lors du criblage des variables d'entrée. Toutefois, comme il est peu raisonnable en pratique de se considérer en régime asymptotique, un test alternatif décrit dans [De Lozzo et Marrel \(2016\)](#) permet de tester  $(H_0)$  quel que soit  $n$  en effectuant une séquence de permutations des observations  $Y^{(k)}$  de la sortie.

Bien que les indices HSIC apportent une solution à de nombreux écueils (entrées dépendantes, entrées fonctionnelles, faible budget d'observations), ils restent difficiles à appréhender pour les non-spécialistes car ils ne s'inscrivent pas dans un paradigme de type ANOVA. Pour améliorer leur interprétabilité, [Da Veiga \(2021\)](#) a démontré qu'une décomposition de la quantité  $\text{HSIC}(\mathbf{X}, Y)$  peut être mise en place, à condition que les entrées soient indépendantes et moyennant l'utilisation de noyaux spécifiques en entrée :

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

Lorsque les variables  $X_i$  sont toutes de loi uniforme sur  $[0, 1]$ , les noyaux de Sobolev vérifient toutes les conditions requises. Ces noyaux sont définis de la manière suivante :

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

avec  $(B_k)_{k \geq 1}$  les polynômes de Bernoulli et  $r$  un paramètre de lissage rendant compte de la régularité des fonctions dans le RKHS sous-jacent. Sur le modèle des indices de Sobol', on peut alors définir des indices HSIC-ANOVA du premier ordre et d'ordre total :

$$\forall 1 \leq i \leq d, S_i^{\text{HSIC}} := \frac{\text{HSIC}(X_i, Y)}{\text{HSIC}(\mathbf{X}, Y)} \text{ et } T_i^{\text{HSIC}} := 1 - \frac{\text{HSIC}(\mathbf{X}_{-i}, Y)}{\text{HSIC}(\mathbf{X}, Y)} = \frac{\tau_i^{\text{HSIC}}}{\text{HSIC}(\mathbf{X}, Y)} .$$

Dans ce contexte, l'objectif de ce travail est d'examiner le bienfondé théorique et la faisabilité algorithmique d'un test d'indépendance basé sur la nullité de l'indice  $\tau_i^{\text{HSIC}}$ .

## 2 Description de la contribution

On commence par démontrer que les noyaux de Sobolev sont caractéristiques. Il en découle un résultat intéressant qui marque la singularité de l'approche HSIC-ANOVA :

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

La première équivalence, bien que très naturelle du point de vue des indices HSIC, est fautive si on la transpose au cas des indices de Sobol'. Comme les indices HSIC-ANOVA basés sur les noyaux de Sobolev sont capables de caractériser l'indépendance au sein d'une paire de variables, on propose de les utiliser pour tester  $(H_0) : X_i \perp Y$ . Pour cela, on considère l'estimateur V-statistique du numérateur  $\tau_i^{\text{HSIC}}$  de l'indice  $T_i^{\text{HSIC}}$  :

$$\widehat{\tau}_i(\mathbf{Z}_{\text{obs}}) := \frac{1}{n^2} \text{Tr} \left[ \left( \left( \bigodot_{j \neq i} \mathbf{L}_j \right) \odot \mathbf{l}_i \right) \mathbf{H} \mathbf{L}_Y \mathbf{H} \right] \text{ avec } \begin{cases} \mathbf{L}_j := \left[ K_{\text{Sob}}^r(X_j^{(s)}, X_j^{(t)}) \right]_{1 \leq s, t \leq n} , \\ \mathbf{l}_i := \left[ k_{\text{Sob}}^r(X_i^{(s)}, X_i^{(t)}) \right]_{1 \leq s, t \leq n} , \end{cases}$$

où  $\odot$  est le produit de Hadamard. Pour simuler la loi de  $\widehat{\tau}_i(\mathbf{Z}_{\text{obs}})$  sous  $(H_0)$ , on construit une procédure de test articulée autour d'un schéma de permutation. Il s'agit plus précisément d'effectuer une séquence  $(\sigma_j)_{1 \leq j \leq B}$  de permutations aléatoires des observations de  $X_i$  au sein de  $\mathbf{Z}_{\text{obs}}$  puis de recalculer la statistique de test :

$$\widehat{\tau}_i^{[\sigma_j]} := \widehat{\tau}_i \left( \mathbf{Z}_{\text{obs}}^{[\sigma_j]} \right) \text{ avec } \mathbf{Z}_{\text{obs}}^{[\sigma_j]} := \left\{ (X_i^{(\sigma_j(k))}, \mathbf{X}_{-i}^{(k)}, Y^{(k)}) \right\}_{k=1}^n .$$

En reprenant les éléments de preuve de [Albert et al. \(2022\)](#), et en notant  $\widehat{q}_{1-\alpha}$  le quantile empirique d'ordre  $1 - \alpha$  de la séquence  $(\widehat{\tau}_i^{[\sigma_j]})_{1 \leq j \leq B+1}$  où  $\sigma_{B+1} = \text{Id}$  conduit à  $\widehat{\tau}_i(\mathbf{Z}_{\text{obs}})$ , il

est facile de montrer que  $\mathbb{P}_{H_0}(\hat{q}_{1-\alpha} \leq \alpha)$  et donc que le test est de bon niveau. En termes d'implémentation, la complexité de la procédure de test peut être limitée en adoptant la stratégie de **El Amri et Marrel (2021)** visant à minimiser le nombre  $B$  de permutations au regard de l'atteinte d'un objectif (stabilisation du criblage ou de la hiérarchisation). De surcroît, une astuce algébrique permet d'écrire que  $n^2 \hat{\tau}_i(\mathbf{Z}_{\text{obs}}) = \text{Tr}(\mathbf{A}\mathbf{W})$  avec  $\mathbf{A}$  et  $\mathbf{W}$  des matrices carrées symétriques. On montre qu'il est possible de calculer analytiquement les deux premiers moments de la variable  $\text{Tr}(\mathbf{A}\mathbf{W})$  lorsque  $\mathbf{W}$  subit des permutations aléatoires affectant simultanément ses lignes et ses colonnes et que la loi uniforme discrète modélise l'aléa sur l'ensemble  $\mathbb{S}_n$  des permutations. La p-valeur du test peut ensuite être estimée en supposant que la loi de  $n^2 \hat{\tau}_i(\mathbf{Z}_{\text{obs}})$ , sous l'effet des permutations, est une loi Gamma de paramètres calculés à l'aide de la méthode des moments exacts. On peut ainsi se départir de toute permutation au moment de tester ( $H_0$ ). Une étude par simulation des performances du test basé sur  $\hat{\tau}_i(\mathbf{Z}_{\text{obs}})$  est réalisée sur différents cas analytiques, dont certains ont été spécialement conçus et/ou sélectionnés pour leur propension à fortement activer les indices HSIC-ANOVA d'ordre total. Le test proposé s'avère au moins aussi puissant (quand il ne l'est pas beaucoup plus) que le test initial basé sur l'indice HSIC traditionnel avec utilisation de noyaux gaussiens.

## Bibliographie

- Albert, M., Laurent, B., Marrel, A. et Meynaoui, A. (2022)**, Adaptive test of independence based on HSIC measures, à paraître dans *The Annals of Statistics*.
- 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, arXiv preprint arXiv:2101.05487.
- De Lozzo, M. et 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.
- El Amri, M. R. et Marrel, A. (2021)**, Optimized HSIC-based tests for sensitivity analysis - Application to thermalhydraulic simulation of accidental scenario on nuclear reactor, *Quality and Reliability Engineering International 2021*, 1-18.
- Gretton, A., Bousquet, O., Smola, A. et Schölkopf, B. (2005)**, Measuring statistical dependence with Hilbert-Schmidt norms. *International conference on algorithmic learning theory*, 63-77, Springer, Berlin, Heidelberg.
- Gretton, A., Fukumizu, K., Teo, C., Song, L., Schölkopf, B. et Smola, A. (2007)**, A kernel statistical test of independence, *Advances in neural information processing systems*, 20.
- Marrel, A., Iooss, B. et Chabridon, V. (2021)**, The ICSCREAM Methodology: Identification of Penalizing Configurations in Computer Experiments Using Screening and Metamodel - Applications in Thermal-Hydraulics, *Nuclear Science and Engineering*, 1-21.

## Remerciements

Ce travail de recherche est mené dans le cadre du projet **SAMOURAI** (*Simulation Analytics and Meta-model-based solutions for Optimization, Uncertainty and Reliability Analysis*) financé par l'Agence Nationale de la Recherche (ANR-20-CE46-0013). Les auteurs sont très reconnaissants de ce soutien financier.



Tests of independence  
based on **HSIC-ANOVA** indices

DE LA RECHERCHE À L'INDUSTRIE

**53<sup>èmes</sup> Journées de Statistique de la SFdS, Lyon, France**

**G. Sarazin, A. Marrel, S. Da Veiga & V. Chabridon**

Thursday, March 16<sup>th</sup>, 2022

**IRESNE | DER | SESI | LEMS**

*Institut de recherche sur les systèmes nucléaires pour la production d'énergie bas carbone*

Commissariat à l'énergie atomique et aux énergies alternatives - [www.cea.fr](http://www.cea.fr)



## Summary

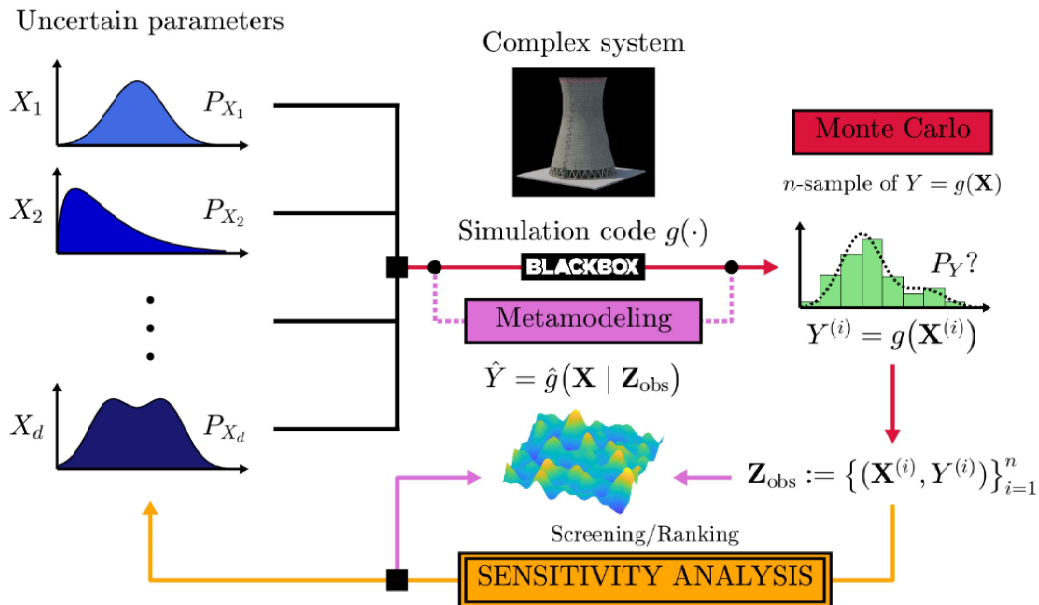
### Introduction

- 1) Useful insights about HSIC indices
- 2) Tests of independence based on HSIC indices
- 3) From HSIC indices to HSIC-ANOVA indices
- 4) Tests of independence based on HSIC-ANOVA indices

### Conclusion



- Let  $\mathbf{X} := [X_1, \dots, X_d]$  be a random vector with **independent** components ( $d \approx 100$ ).
- Let  $Y := g(\mathbf{X})$  where  $g: \mathbb{R}^d \rightarrow \mathbb{R}$  is a **computationally-expensive** simulation code.



- Total-order **Sobol' indices** cannot be estimated from a small amount of input-output samples.
- The **HSIC indices** were defined by [Gretton \*et al.\* \(2005\)](#) and promoted by [Da Veiga \(2015\)](#).

Useful insights about HSIC indices

## What are kernels used for in global sensitivity analysis?

- Let  $\mathbb{P}_X$  and  $\mathbb{P}_Y$  be the probability measures associated to  $X$  to  $Y$ .
- A kernel  $K_i : \mathcal{X}_i \times \mathcal{X}_i \rightarrow \mathbb{R}$  (with feature space  $\mathcal{F}_i$  and feature map  $\varphi_i$ ) is assigned to  $X_i$ .
- A kernel  $K_Y : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$  (with feature space  $\mathcal{G}$  and feature map  $\varphi_Y$ ) is assigned to  $Y$ .

$$\forall 1 \leq i \leq d, \quad K_i(x_i, x'_i) = \langle \varphi_i(x'_i), \varphi_i(x_i) \rangle_{\mathcal{F}_i} \quad \text{and} \quad K_Y(y, y') = \langle \varphi_Y(y'), \varphi_Y(y) \rangle_{\mathcal{G}}$$

### 0 Kernel embeddings

- Let  $\nu$  be a probability measure with support  $\mathcal{Z} \subseteq \mathbb{R}^p$ .
- Let  $K : \mathcal{Z} \times \mathcal{Z} \rightarrow \mathbb{R}$  be a kernel and let  $\mathcal{H}$  be its **reproducing kernel Hilbert space** (RKHS).
- $\nu$  can be represented in  $\mathcal{H}$  by the following function:

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

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

## The Hilbert-Schmidt independence criterion (HSIC)

### 1 Maximum Mean Discrepancy (MMD)

$$\text{HSIC}(X_i, Y) := \text{MMD}^2(\mathbb{P}_{X_i Y}, \mathbb{P}_{X_i} \otimes \mathbb{P}_Y) = \left\| \mu_{\mathbb{P}_{X_i Y}} - \mu_{\mathbb{P}_{X_i} \otimes \mathbb{P}_Y} \right\|_{\mathcal{F}_i \otimes \mathcal{G}}^2$$

- Kernel-based **dissimilarity measure** between  $\mathbb{P}_{X_i Y}$  and  $\mathbb{P}_{X_i} \otimes \mathbb{P}_Y$ .

### 2 Hilbert-Schmidt norm of a covariance operator

$$\text{HSIC}(X_i, Y) = \|C_{X_i Y}\|_{\text{HS}}^2 \quad \text{with} \quad \begin{aligned} C_{X_i Y} &: \mathcal{G} \longrightarrow \mathcal{F}_i \\ C_{X_i Y} &= \mathbb{E}[\varphi_i(X_i) \otimes \varphi_Y(Y)] - \mathbb{E}[\varphi_i(X_i)] \otimes \mathbb{E}[\varphi_Y(Y)] \end{aligned}$$

- Explicit knowledge about the **feature maps**  $\varphi_i$  and  $\varphi_Y$  helps to better understand HSIC indices.

### 3 Formula used in practice

$$\begin{aligned} \text{HSIC}(X_i, Y) &= \mathbb{E}_{\substack{X_i, Y \\ X'_i, Y'}} [K_i(X_i, X'_i) K_Y(Y, Y')] + \mathbb{E}_{X_i, X'_i} [K_i(X_i, X'_i)] \mathbb{E}_{Y, Y'} [K_Y(Y, Y')] \\ &\quad - 2 \mathbb{E}_{X_i, Y} [\mathbb{E}_{X'_i} [K_i(X_i, X'_i)] \mathbb{E}_{Y'} [K_Y(Y, Y')]] \quad \text{with} \quad (X_i, Y) \perp (X'_i, Y') \sim \mathbb{P}_{X_i Y} \end{aligned}$$

## 4 V-statistic estimator

computational complexity:  $\mathcal{O}(n^2)$ 

$$\widehat{\text{HSIC}}_v(X_i, Y) = \frac{1}{n^2} \sum_{1 \leq p, q \leq n} K_i(X_i^{(p)}, X_i^{(q)}) K_Y(Y^{(p)}, Y^{(q)})$$

$$- \frac{2}{n^3} \sum_{1 \leq p, q, r \leq n} K_i(X_i^{(p)}, X_i^{(q)}) K_Y(Y^{(p)}, Y^{(r)}) + \frac{1}{n^4} \sum_{1 \leq p, q, r, s \leq n} K_i(X_i^{(p)}, X_i^{(q)}) K_Y(Y^{(r)}, Y^{(s)})$$

✓ Algebraic formula  $\rightarrow \widehat{\text{HSIC}}_v(X_i, Y) = \frac{1}{n^2} \text{Tr}[\mathbf{L}_i \mathbf{H} \mathbf{L}_Y \mathbf{H}]$

Gram matrices  $\Rightarrow \mathbf{L}_i := [K_i(X_i^{(s)}, X_i^{(t)})]_{1 \leq s, t \leq n}$  and  $\mathbf{L}_Y := [K_Y(Y^{(s)}, Y^{(t)})]_{1 \leq s, t \leq n}$

Centering matrix  $\Rightarrow \mathbf{H} := \mathbf{I}_n - \frac{1}{n} \mathbf{1} \mathbf{1}^T$

## 5 Characterization of independence

➤ If  $K_1, \dots, K_d$  and  $K_Y$  are characteristic kernels, one has:  $X_i \perp Y \Leftrightarrow \text{HSIC}(X_i, Y) = 0$

✓ Testing  $(H_0^i) : X_i \perp Y$  is equivalent to testing  $(H_0^i) : \text{HSIC}(X_i, Y) = 0$ .

## Tests of independence based on HSIC indices

➤ To build a **test of independence**, two elements are required:

1. A **measure of association**  $\mathcal{M}(\cdot, \cdot)$  characterizing independence →
2. An estimator of  $\mathcal{M}(X_i, Y)$  with **known distribution** under  $(H_0^i)$ .

$$\begin{array}{c} \mathcal{M}(X_i, Y) = 0 \\ \Downarrow \\ X_i \perp Y \end{array}$$

□ What is the distribution of  $\widehat{\text{HSIC}}_v(X_i, Y)$  under  $(H_0^i)$  ?

**1** Asymptotic test → For LARGE sample sizes

Under  $(H_0^i)$ , the **asymptotic distribution** of  $\widehat{\text{HSIC}}_v(X_i, Y)$  is given by:

$$n\widehat{\text{HSIC}}_v(X_i, Y) \xrightarrow[\mathcal{L}]{n \rightarrow \infty} \Lambda := \sum_{k, l \geq 1} \lambda_k \mu_l N_{kl}^2$$

$(N_{kl})_{k, l \geq 1}$  is a collection of i.i.d. random variables with  $N_{kl} \sim \mathcal{N}(0, 1)$ .

$(\lambda_k)_{k \geq 1}$  and  $(\mu_l)_{l \geq 1}$  are the **eigenvalues** of the **integral operators** associated to  $\tilde{K}_i$  and  $\tilde{K}_Y$ .

- $\tilde{K}_i(x_i, x'_i) := \langle \tilde{\varphi}_i(x'_i), \tilde{\varphi}_i(x_i) \rangle_{\mathcal{F}_i}$  with  $\tilde{\varphi}_i(\cdot) := \varphi_i(\cdot) - \mu_{\mathbb{P}_{X_i}}$  ⇒  $\tilde{\mathbf{L}}_i := \mathbf{H}\mathbf{L}_i\mathbf{H}$
  - $\tilde{K}_Y(y, y') := \langle \tilde{\varphi}_Y(y'), \tilde{\varphi}_Y(y) \rangle_{\mathcal{G}}$  with  $\tilde{\varphi}_Y(\cdot) := \varphi_Y(\cdot) - \mu_{\mathbb{P}_Y}$  ⇒  $\tilde{\mathbf{L}}_Y := \mathbf{H}\mathbf{L}_Y\mathbf{H}$
- CENTERED**  
feature maps
**CENTERED**  
Gram matrices

- $\Lambda$  is a **positive linear combination** of independent **chi-squared** random variables.
- ✓ This **spectral distribution** can be approximated by a **Gamma distribution**.

## 1 Asymptotic test → computational complexity: $(8d + 5) n^2$

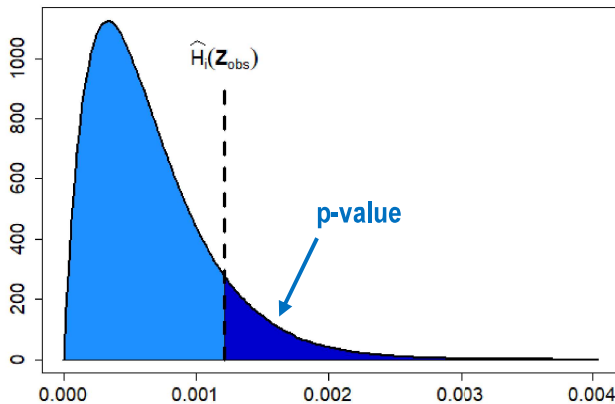
Under  $(H_0^i)$ , the asymptotic distribution of  $\widehat{\text{HSIC}}_v(X_i, Y)$  is given by:

$$n\widehat{\text{HSIC}}_v(X_i, Y) \xrightarrow[\mathcal{L}]{n \rightarrow \infty} \Lambda := \sum_{k,l \geq 1} \lambda_k \mu_l N_{kl}^2$$

$(N_{kl})_{k,l \geq 1}$  is a collection of i.i.d. random variables with  $N_{kl} \sim \mathcal{N}(0,1)$ .

$(\lambda_k)_{k \geq 1}$  and  $(\mu_l)_{l \geq 1}$  are the **eigenvalues** of the **integral operators** associated to  $\tilde{K}_X$  and  $\tilde{K}_Y$ .

### Gamma distribution



$$\bullet \mathbb{E}[\widehat{\text{HSIC}}_v(X_i, Y)] \approx \frac{1}{n^3} \text{Tr}(\tilde{\mathbf{L}}_X) \text{Tr}(\tilde{\mathbf{L}}_Y)$$

$$\bullet \mathbb{V}(\widehat{\text{HSIC}}_v(X_i, Y)) \approx \frac{2}{n^6} \text{Tr}(\tilde{\mathbf{L}}_X^2) \text{Tr}(\tilde{\mathbf{L}}_Y^2)$$

1. Fit based on the **method of moments**
2. **Parametric estimation** of the p-value

$$\hat{p}_i = 1 - \hat{F}_{\text{Ga}}(\hat{H}_i(\mathbf{Z}_{\text{obs}}))$$

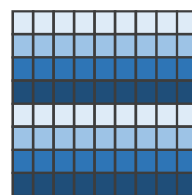
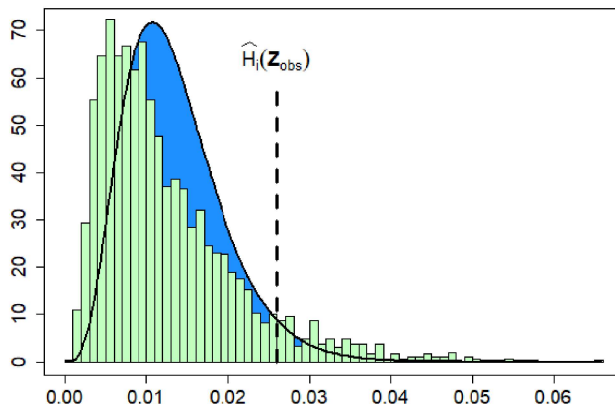
## 2 Permutation-based test → For **SMALL** sample sizes

➤ The distribution under  $(H_0^i)$  of  $\widehat{\text{HSIC}}_v(X_i, Y)$  can be **simulated** from the available data.

1. Perform a sequence  $\{\sigma_b\}_{1 \leq b \leq B}$  of **random permutations** on  $\mathbf{Y}_{\text{obs}} = \{Y^{(k)}\}_{1 \leq k \leq n}$ .
2. Recompute the test statistic  $\hat{H}_i^{\sigma_b}$  for each permuted output dataset  $\mathbf{Y}_{\text{obs}}^{\sigma_b}$ .

3. **Non-parametric estimation** of the p-value →

$$\hat{p}_i = \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\{\hat{H}_i^{\sigma_b} > \hat{H}_i(\mathbf{Z}_{\text{obs}})\}}$$



$\mathbf{X}_{\text{obs}}$



$\mathbf{Y}_{\text{obs}}$

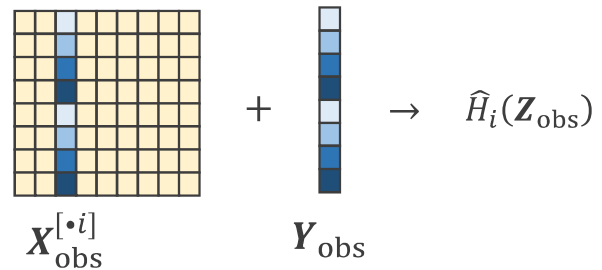
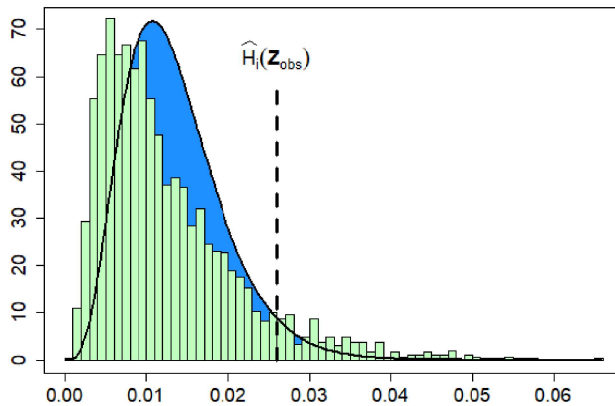
## 2 Permutation-based test → computational complexity: $(7Bd) n^2$

➤ The distribution under  $(H_0^i)$  of  $\widehat{\text{HSIC}}_v(X_i, Y)$  can be **simulated** from the available data.

1. Perform a sequence  $\{\sigma_b\}_{1 \leq b \leq B}$  of **random permutations** on  $\mathbf{Y}_{\text{obs}} = \{Y^{(k)}\}_{1 \leq k \leq n}$ .
2. Recompute the test statistic  $\widehat{H}_i^{\sigma_b}$  for each permuted output dataset  $\mathbf{Y}_{\text{obs}}^{\sigma_b}$ .

3. **Non-parametric estimation** of the **p-value** →

$$\widehat{p}_i = \frac{1}{B} \sum_{b=1}^B 1_{\{\widehat{H}_i^{\sigma_b} > \widehat{H}_i(\mathbf{Z}_{\text{obs}})\}}$$



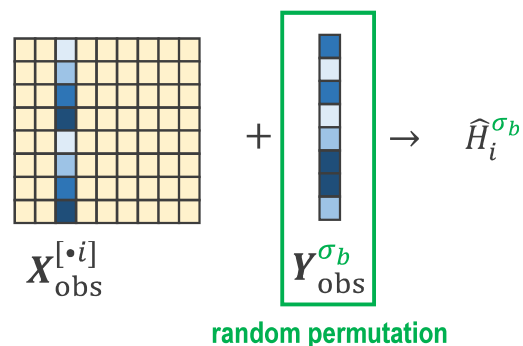
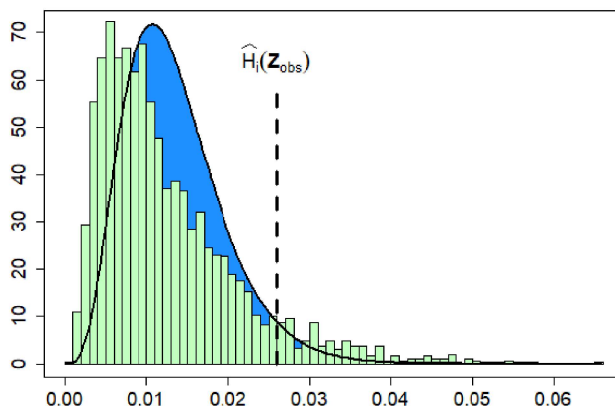
## 2 Permutation-based test → computational complexity: $(7Bd) n^2$

➤ The distribution under  $(H_0^i)$  of  $\widehat{\text{HSIC}}_v(X_i, Y)$  can be **simulated** from the available data.

1. Perform a sequence  $\{\sigma_b\}_{1 \leq b \leq B}$  of **random permutations** on  $\mathbf{Y}_{\text{obs}} = \{Y^{(k)}\}_{1 \leq k \leq n}$ .
2. Recompute the test statistic  $\widehat{H}_i^{\sigma_b}$  for each permuted output dataset  $\mathbf{Y}_{\text{obs}}^{\sigma_b}$ .

3. **Non-parametric estimation** of the **p-value** →

$$\widehat{p}_i = \frac{1}{B} \sum_{b=1}^B 1_{\{\widehat{H}_i^{\sigma_b} > \widehat{H}_i(\mathbf{Z}_{\text{obs}})\}}$$

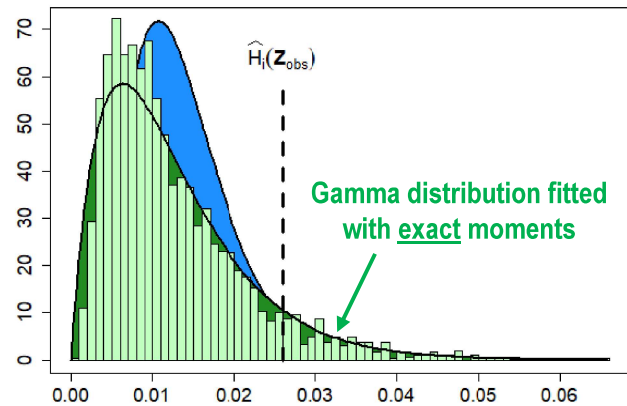
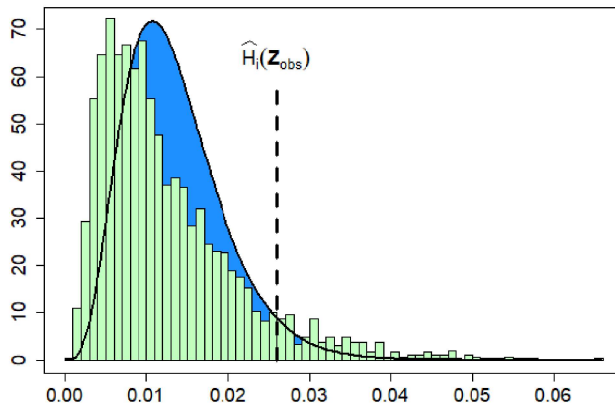


### 3 Non-asymptotic Gamma test → computational complexity: $6(d + 1) n^2$

- The histogram of the  $\{\widehat{H}_i^{\sigma_b}\}_{1 \leq b \leq B}$  is unimodal with support  $[0, +\infty[$  and positive skewness.
  - ✓ It looks like a **Gamma distribution** even if  $n$  is small.

$$n^2 \widehat{\text{HSIC}}_v(X_i, Y) = \text{Tr}(\widetilde{\mathbf{L}}_i \widetilde{\mathbf{L}}_Y) = \text{Tr}(\mathbf{AW})$$

- Applying  $\sigma(\cdot)$  on  $\mathbf{Y}_{\text{obs}}$  leads to apply simultaneously  $\sigma(\cdot)$  on the rows and columns of  $\mathbf{W}$ .
- The first two moments of  $\text{Tr}(\mathbf{AW})$  can be computed → see [Kazi-Aoual et al. \(1995\)](#)



## From HSIC indices to HSIC-ANOVA indices

## Towards an ANOVA decomposition in the HSIC paradigm

- **Sobol (1993)** → ANalysis Of VAriance :  $\mathbb{V}(Y)$  is apportioned between **independent** inputs.

$$\mathbb{V}(Y) = \sum_{\mathbf{u} \subseteq \{1, \dots, d\}} S_{\mathbf{u}} = \sum_{\mathbf{u} \subseteq \{1, \dots, d\}} \sum_{\mathbf{v} \subseteq \mathbf{u}} (-1)^{|\mathbf{u}| - |\mathbf{v}|} \mathbb{V}(\mathbb{E}[Y | \mathbf{X}_{\mathbf{v}}])$$

- The **first-order and total-order Sobol indices** are then defined by:

$$\forall 1 \leq i \leq d, \quad S_i := \frac{\mathbb{V}(\mathbb{E}[Y | X_i])}{\mathbb{V}(Y)} \quad \text{and} \quad S_i^T := 1 - \frac{\mathbb{V}(\mathbb{E}[Y | \mathbf{X}_{-i}])}{\mathbb{V}(Y)}$$

## Towards an ANOVA decomposition in the HSIC paradigm

- **Da Veiga (2021)** →  $\text{HSIC}(\mathbf{X}, Y)$  may be apportioned between **independent** inputs.

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

- The **first-order and total-order HSIC-ANOVA indices** are then defined by:

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



## Towards an ANOVA decomposition in the HSIC paradigm

- Da Veiga (2021) → HSIC( $\mathbf{X}, Y$ ) may be apportioned between independent inputs.

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

- The **first-order and total-order HSIC-ANOVA indices** are then defined by:

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

- Each input kernel  $K_i$  must be an **ANOVA kernel**.

✓  $K_i = 1 + k_i$  with  $k_i$  an **orthogonal kernel**.

$$\forall x_i \in \mathcal{X}_i, \quad \int_{\mathcal{X}_i} k_i(z, x_i) d\mathbb{P}_{\mathcal{X}_i}(z) = 0$$

- The kernel  $K_{\mathbf{u}}$  assigned to  $\mathbf{X}_{\mathbf{u}}$  (with  $\mathbf{u} := \{u_1, \dots, u_s\}$ ) is built as follows:

$$K_{\mathbf{u}} := K_{u_1} \otimes \dots \otimes K_{u_s} = (1 + k_{u_1}) \otimes \dots \otimes (1 + k_{u_s}) \quad \text{with} \quad \mathbf{u} := \{u_1, \dots, u_s\}$$

- The choice of  $K_Y$  remains free.

## Focus on Sobolev kernels

- They are very few **ANOVA kernels** among well-known kernels (e.g. Gaussian, Laplace, Matérn).
- One example is given by **Sobolev kernels** provided that  $\mathbb{P}_{\mathcal{X}_i} = \mathbf{u}([0,1])$ .

$$K_{\text{Sob}}^r(x, x') = 1 + k_{\text{Sob}}^r(x, x') = 1 + \sum_{j=1}^r \frac{B_j(x)B_j(x')}{(j!)^2} + \frac{(-1)^{r+1}}{(2r)!} B_{2r}(|x - x'|)$$

✓  $(B_j)_{1 \leq j \leq r}$  are the **Bernoulli polynomials**.

✓  $r \geq 1$  accounts for the **degree of smoothness** within the underlying RKHS.

For any  $r \geq 1$ , the **Sobolev kernel**  $K_{\text{Sob}}^r$  is **characteristic** on  $[0,1]$ .

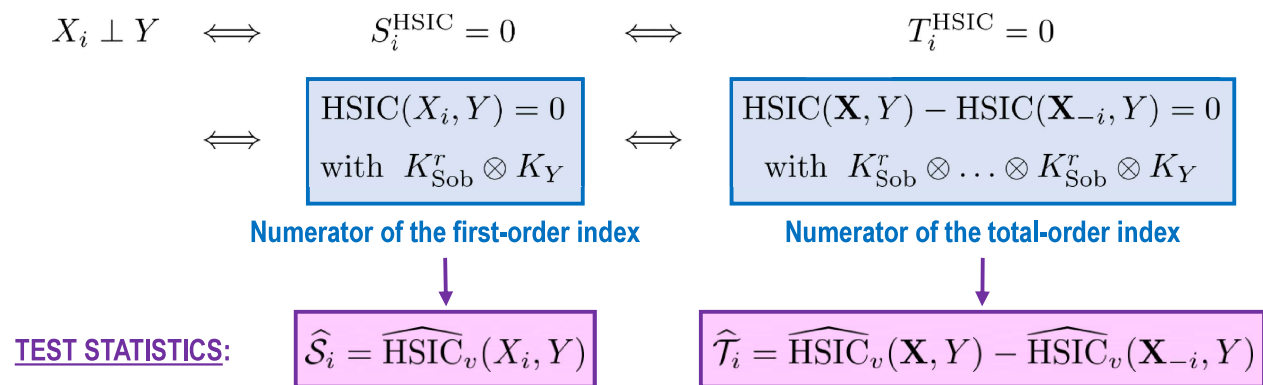
- Replacing  $B_{2r}$  by its **series expansion** leads to a feature map  $\varphi_{\text{Sob}}^r: [0,1] \rightarrow \ell^2(\mathbb{N}, \mathbb{R})$ .
- It can be shown that  $\mathcal{H}_{\text{Sob}}^r$  contains all **Bernoulli polynomials** (hence all polynomials).
- **Stone-Weierstrass theorem**  $\Rightarrow \mathcal{H}_{\text{Sob}}^r$  is **dense** in  $\mathcal{C}([0,1]) \Rightarrow K_{\text{Sob}}^r$  is **universal**.

## Focus on Sobolev kernels

- They are very few **ANOVA kernels** among well-known kernels (e.g. Gaussian, Laplace, Matérn).
- One example is given by **Sobolev kernels** provided that  $\mathbb{P}_{X_i} = \mathbf{u}([0,1])$ .

- They are very few **ANOVA kernels** among well-known kernels (e.g. Gaussian, Laplace, Matérn).
- One example is given by **Sobolev kernels** provided that  $\mathbb{P}_{X_i} = \mathbf{u}([0,1])$ .

For any  $r \geq 1$ , the Sobolev kernel  $K_{\text{Sob}}^r$  is characteristic on  $[0,1]$ .



## Tests of independence based on HSIC-ANOVA indices

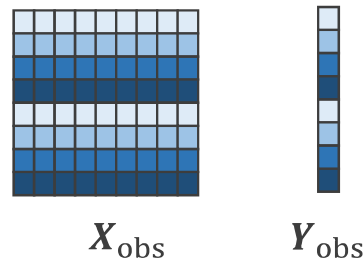
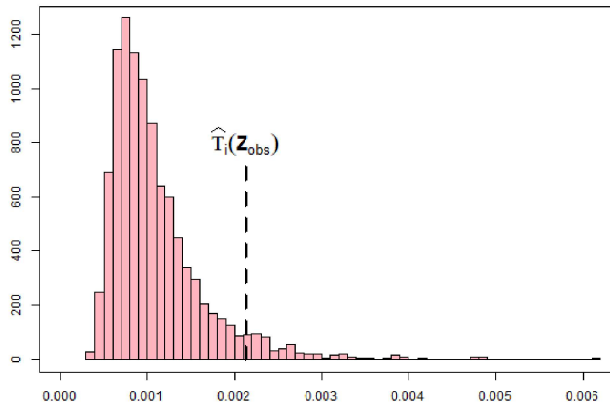
### 1 Permutation-based test → For **SMALL** sample sizes

➤ The distribution under  $(H_0^i)$  of  $\hat{\mathcal{J}}_i$  can be **simulated** from the available data.

1. Perform a sequence  $\{\sigma_b\}_{1 \leq b \leq B}$  of **random permutations** on  $\mathbf{X}_{\text{obs}}^{[i]} = \{X_i^{(k)}\}_{1 \leq k \leq n}$ .
2. Recompute the test statistic  $\hat{\mathcal{J}}_i^{\sigma_b}$  for each permuted input dataset  $\mathbf{X}_{\text{obs}}^{\sigma_b}$ .

3. **Non-parametric estimation** of the **p-value** →

$$\hat{p}_i = \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\{\hat{\mathcal{T}}_i^{\sigma_b} > \hat{\mathcal{T}}_i(\mathbf{z}_{\text{obs}})\}}$$



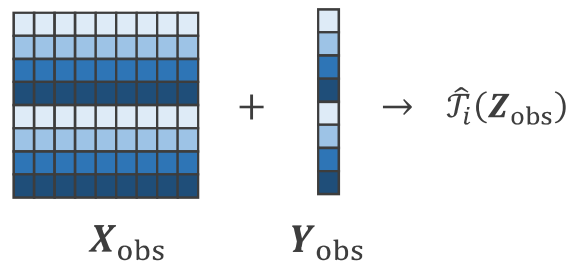
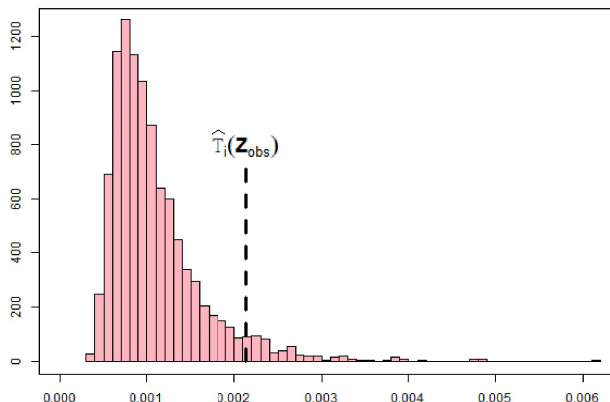
### 1 Permutation-based test → computational complexity: $(d^2 + 7Bd) n^2$

➤ The distribution under  $(H_0^i)$  of  $\hat{\mathcal{J}}_i$  can be **simulated** from the available data.

1. Perform a sequence  $\{\sigma_b\}_{1 \leq b \leq B}$  of **random permutations** on  $\mathbf{X}_{\text{obs}}^{[i]} = \{X_i^{(k)}\}_{1 \leq k \leq n}$ .
2. Recompute the test statistic  $\hat{\mathcal{J}}_i^{\sigma_b}$  for each permuted input dataset  $\mathbf{X}_{\text{obs}}^{\sigma_b}$ .

3. **Non-parametric estimation** of the **p-value** →

$$\hat{p}_i = \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\{\hat{\mathcal{T}}_i^{\sigma_b} > \hat{\mathcal{T}}_i(\mathbf{z}_{\text{obs}})\}}$$



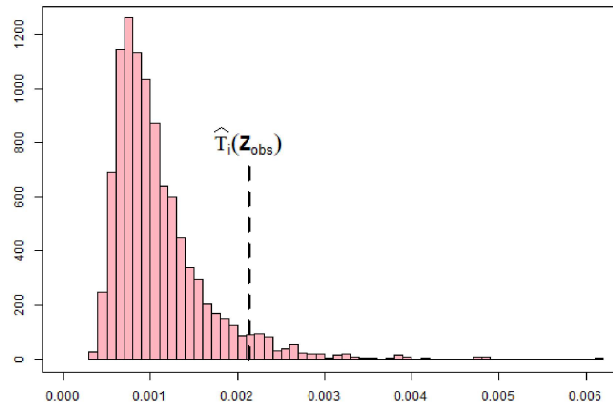
### 1 Permutation-based test → computational complexity: $(d^2 + 7Bd) n^2$

➤ The distribution under  $(H_0^i)$  of  $\hat{\mathcal{J}}_i$  can be simulated from the available data.

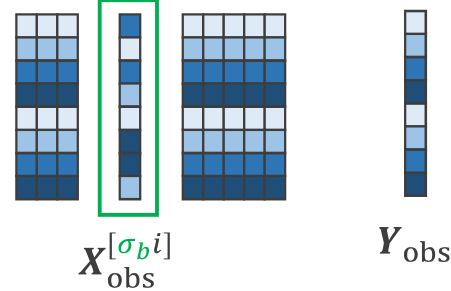
1. Perform a sequence  $\{\sigma_b\}_{1 \leq b \leq B}$  of random permutations on  $\mathbf{X}_{\text{obs}}^{[i]} = \{X_i^{(k)}\}_{1 \leq k \leq n}$ .
2. Recompute the test statistic  $\hat{\mathcal{J}}_i^{\sigma_b}$  for each permuted input dataset  $\mathbf{X}_{\text{obs}}^{\sigma_b}$ .

3. Non-parametric estimation of the p-value →

$$\hat{p}_i = \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\{\hat{\mathcal{J}}_i^{\sigma_b} > \hat{\mathcal{J}}_i(\mathbf{z}_{\text{obs}})\}}$$



random permutation



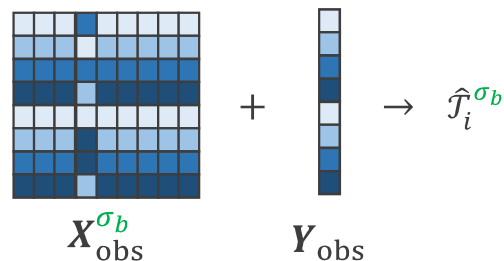
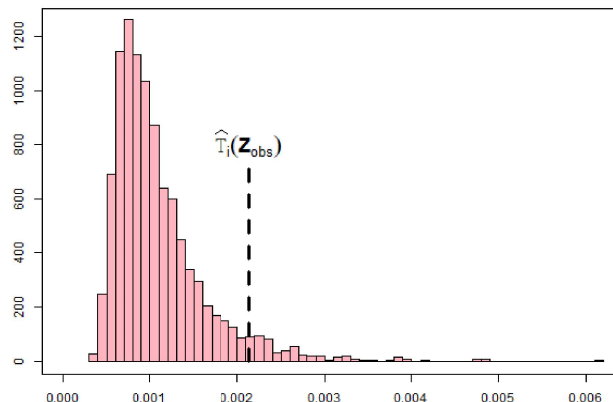
### 1 Permutation-based test → computational complexity: $(d^2 + 7Bd) n^2$

➤ The distribution under  $(H_0^i)$  of  $\hat{\mathcal{J}}_i$  can be simulated from the available data.

1. Perform a sequence  $\{\sigma_b\}_{1 \leq b \leq B}$  of random permutations on  $\mathbf{X}_{\text{obs}}^{[i]} = \{X_i^{(k)}\}_{1 \leq k \leq n}$ .
2. Recompute the test statistic  $\hat{\mathcal{J}}_i^{\sigma_b}$  for each permuted input dataset  $\mathbf{X}_{\text{obs}}^{\sigma_b}$ .

3. Non-parametric estimation of the p-value →

$$\hat{p}_i = \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\{\hat{\mathcal{J}}_i^{\sigma_b} > \hat{\mathcal{J}}_i(\mathbf{z}_{\text{obs}})\}}$$



## 2 Non-asymptotic Gamma test → For LARGE sample sizes

- The histogram of the  $\{\hat{\mathcal{T}}_i^{\sigma_b}\}_{1 \leq b \leq B}$  is **unimodal** with support  $[0, +\infty[$  and **positive skewness**.
  - ✓ It looks like a **Gamma distribution** even if  $n$  is small.

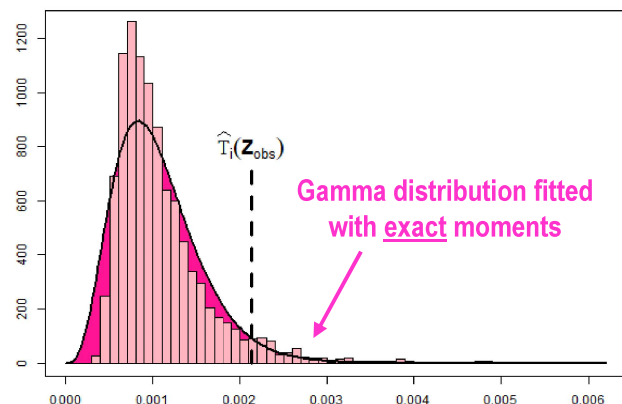
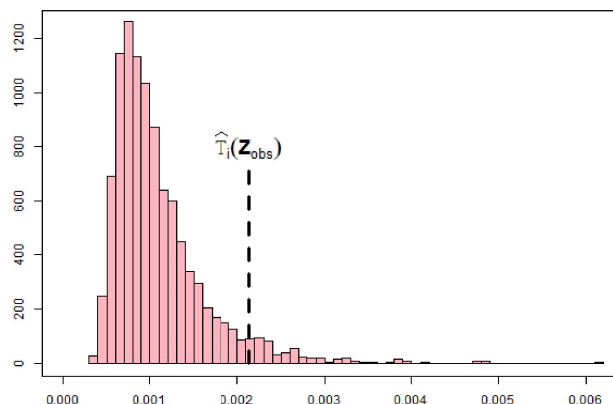
$$\begin{aligned} n^2 \hat{\mathcal{T}}_1 &= n^2 \left[ \widehat{\text{HSIC}}_v(\mathbf{X}, Y) - \widehat{\text{HSIC}}_v(\mathbf{X}_{-1}, Y) \right] = \text{Tr} \left( (\mathbf{L}_1 \odot \dots \odot \mathbf{L}_d) \tilde{\mathbf{L}}_Y \right) - \text{Tr} \left( (\mathbf{L}_2 \odot \dots \odot \mathbf{L}_d) \tilde{\mathbf{L}}_Y \right) \\ &= \text{Tr} \left( (\mathbf{I}_1 \odot \mathbf{L}_{2\dots d}) \tilde{\mathbf{L}}_Y \right) = \text{Tr} \left( (\tilde{\mathbf{L}}_Y \odot \mathbf{L}_{2\dots d}) \mathbf{I}_1 \right) = \text{Tr}(\mathbf{A}\mathbf{W}) \quad \text{with} \quad \begin{cases} \mathbf{L}_{2\dots d} & := \mathbf{L}_2 \odot \dots \odot \mathbf{L}_d \\ \mathbf{I}_1 & := \mathbf{L}_1 - 1 \end{cases} \end{aligned}$$

- Applying  $\sigma(\cdot)$  on  $\mathbf{X}_{obs}^1$  leads to apply simultaneously  $\sigma(\cdot)$  on the rows and columns of  $\mathbf{W}$ .
- The moments of  $\text{Tr}(\mathbf{A}\mathbf{W})$  are much harder to achieve.
  - **A** and **W** are **NOT double centered matrices!**
  - ✓ **Generalized formulas** to compute the first two moments.

## 2 Non-asymptotic Gamma test → computational complexity: $(d^2 + 14d + 4) n^2$

- The histogram of the  $\{\hat{\mathcal{T}}_i^{\sigma_b}\}_{1 \leq b \leq B}$  is **unimodal** with support  $[0, +\infty[$  and **positive skewness**.
  - ✓ It looks like a **Gamma distribution** even if  $n$  is small.

$$\begin{aligned} n^2 \hat{\mathcal{T}}_1 &= n^2 \left[ \widehat{\text{HSIC}}_v(\mathbf{X}, Y) - \widehat{\text{HSIC}}_v(\mathbf{X}_{-1}, Y) \right] = \text{Tr} \left( (\mathbf{L}_1 \odot \dots \odot \mathbf{L}_d) \tilde{\mathbf{L}}_Y \right) - \text{Tr} \left( (\mathbf{L}_2 \odot \dots \odot \mathbf{L}_d) \tilde{\mathbf{L}}_Y \right) \\ &= \text{Tr} \left( (\mathbf{I}_1 \odot \mathbf{L}_{2\dots d}) \tilde{\mathbf{L}}_Y \right) = \text{Tr} \left( (\tilde{\mathbf{L}}_Y \odot \mathbf{L}_{2\dots d}) \mathbf{I}_1 \right) = \text{Tr}(\mathbf{A}\mathbf{W}) \quad \text{with} \quad \begin{cases} \mathbf{L}_{2\dots d} & := \mathbf{L}_2 \odot \dots \odot \mathbf{L}_d \\ \mathbf{I}_1 & := \mathbf{L}_1 - 1 \end{cases} \end{aligned}$$



- Both **first-order** and **total-order** HSIC-ANOVA indices allow to **characterize independence**.

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



- ✓  $K_1, \dots, K_d$  must be **ANOVA kernels**  $\rightarrow K_i = K_{\text{Sob}}^r$  if  $\mathbb{P}_{X_i} = \mathcal{U}([0,1])$ .
- ✓  $K_1, \dots, K_d$  and  $K_Y$  must be **characteristic kernels**.

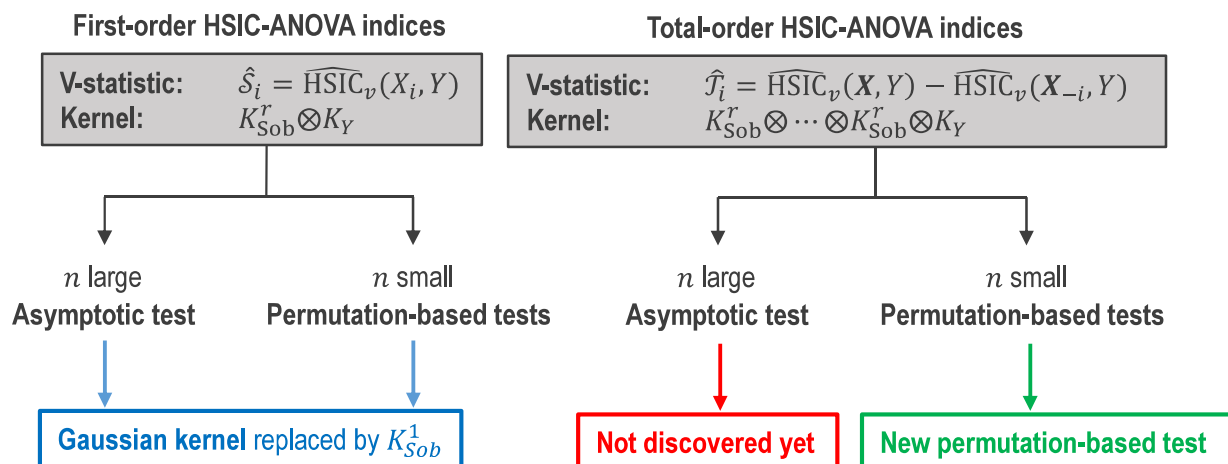
- Both **first-order** and **total-order** HSIC-ANOVA indices allow to **characterize independence**.

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



- ✓  $K_1, \dots, K_d$  must be **ANOVA kernels**  $\rightarrow K_i = K_{\text{Sob}}^r$  if  $\mathbb{P}_{X_i} = \mathcal{U}([0,1])$ .
- ✓  $K_1, \dots, K_d$  and  $K_Y$  must be **characteristic kernels**.

- Extension of HSIC-based **tests of independence** to HSIC-ANOVA indices.



- Both **first-order** and **total-order** HSIC-ANOVA indices allow to **characterize independence**.

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



- ✓  $K_1, \dots, K_d$  must be **ANOVA kernels**  $\rightarrow K_i = K_{\text{Sob}}^r$  if  $\mathbb{P}_{X_i} = \mathbf{u}([0,1])$ .
- ✓  $K_1, \dots, K_d$  and  $K_Y$  must be **characteristic kernels**.

- Extension of HSIC-based **tests of independence** to **HSIC-ANOVA indices**.

#### First-order HSIC-ANOVA indices

**V-statistic:**  $\hat{S}_i = \widehat{\text{HSIC}}_v(X_i, Y)$   
**Kernel:**  $K_{\text{Sob}}^r \otimes K_Y$

#### Total-order HSIC-ANOVA indices

**V-statistic:**  $\hat{T}_i = \widehat{\text{HSIC}}_v(\mathbf{X}, Y) - \widehat{\text{HSIC}}_v(\mathbf{X}_{-i}, Y)$   
**Kernel:**  $K_{\text{Sob}}^r \otimes \dots \otimes K_{\text{Sob}}^r \otimes K_Y$

- The **statistical power** of both tests is estimated through a **simulation study**:

✓ **Traditional benchmarks:**  $\text{Power}(\hat{S}_i) \approx \text{Power}(\widehat{\text{HSIC}}_{\mathcal{N}}) \approx \text{Power}(\hat{T}_i)$

✓ **Specific benchmarks:**  $\text{Power}(\hat{S}_i) < \text{Power}(\widehat{\text{HSIC}}_{\mathcal{N}}) < \text{Power}(\hat{T}_i)$

## References

- DA VEIGA, S. (2015), "Global sensitivity analysis with dependence measures", *Journal of Statistical Computation and Simulation*, vol. 85, n°7, pp. 1283-1305.
- DA VEIGA, S. (2016), "New perspectives for sensitivity analysis", *Proceedings of Mascot-Num 2016 Conference*, Toulouse, France.
- DA VEIGA, S. (2021), "Kernel-based ANOVA decomposition and Shapley effects - Application to global sensitivity analysis", [preprint](#).
- DE LOZZO, M. and MARREL, A. (2016), "New improvements in the use of dependence measures for sensitivity analysis and screening", *Journal of Statistical Computation and Simulation*, vol. 86, n°15, pp. 3038-3058.
- EL AMRI, M. R. and MARREL, A. (2021), "Optimized HSIC-based tests for sensitivity analysis - Application to thermalhydraulic simulation of accidental scenario on nuclear reactor", *Quality and Reliability Engineering International*.
- EL AMRI, M. R. and MARREL, A. (2021), "More powerful HSIC-based independence tests, extension to space-filling designs and functional data", [preprint](#).
- GRETTON, A., BOUSQUET, O., SMOLA, A. and SCHÖLKOPF, B. (2005), "Measuring statistical dependence with Hilbert-Schmidt norms", *International Conference on Algorithmic Learning Theory*, pp. 63-77, Springer.
- GRETTON, A., FUKUMIZU, K., TEO, C., SONG, L., SCHÖLKOPF, B. and SMOLA, A. (2007), "A kernel statistical test of independence", *Advances in Neural Information Processing Systems*, 20.
- KAZI-AOUAL, F., HITIER, S., SABATIER, R. and LEBRETON, J. D. (1995), "Refined approximations to permutation tests for multivariate inference", *Computational Statistics & Data Analysis*, vol. 20, n°6, pp. 643-656.

## References

- MARREL, A. and CHABRIDON, V. (2021), "Statistical developments for target and conditional sensitivity analysis - Application on safety studies for nuclear reactors", *Reliability Engineering & System Safety*, vol. 214, p. 107711.
- MARREL, A., IOOSS, B. and CHABRIDON, V. (2020), "Statistical identification of penalizing configurations in high-dimensional thermalhydraulic numerical experiments - The ICSCREAM methodology", [preprint](#).
- MEYNAOUI, A., MARREL, A. and LAURENT, B. (2019), "New statistical methodology for second level global sensitivity analysis", [preprint](#).
- SERFLING, R. J. (1980), "Approximation theorems of mathematical statistics" (vol. 162), John Wiley & Sons.
- SPAGNOL, A., LE RICHE, R. and DA VEIGA, S. (2019), "Bayesian optimization in effective dimensions via kernel-based sensitivity indices", *30th European Conference on Operational Research*, EURO2019.
- SOBOL, I. M., 1993, "Sensitivity analysis for non-linear mathematical models", *Mathematical Modelling and Computational Experiment*, vol. 1, pp. 407-414.
- ZHANG, K., PETERS, J., JANZING, D. and SCHÖLKOPF, B. (2011), "Kernel-based conditional independence test and application in causal discovery". *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, pp. 804-813.
- ZHANG, Q., FILIPPI, S., GRETTON, A. and SEJDINOVIC, D. (2018), "Large-scale kernel methods for independence testing", *Statistics and Computing*, vol. 28, n°1, pp. 113-130.



# THANKS FOR YOUR ATTENTION

Work funded and carried out within the  ANR research project.

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

