NEW ADVANCED IN SENSITIVITY ANALYSIS BASED ON HSIC DEPENDENCE MEASURES

**Hilbert Schmidt Independence Criterion*

FROM RESEARCH TO INDUSTRY

Amandine MARREL

CEA DES/IRESNE/DER, IMT Toulouse

In collaboration with M. Albert, V. Chabridon, M. De Lozzo, R. El Amri, B. Laurent-Bonneau, A. Meynaoui, G. Sarazin.

LIKE Workshop 2022

IRESNE / DER / SESI / LEMS

Research Institute on Nuclear Systems for Low Carbone Energy Production

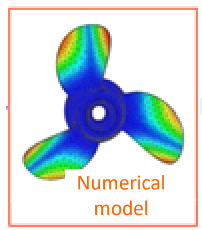
Sensitivity Analysis in Uncertainty treatment
⇒ *In numerical simulation*

- **Numerical simulators:** fundamental tools to model & predict physical phenomena.
- **Large number of input parameters**, characterizing the studied phenomenon or related to its physical and numerical modelling.
- **Uncertainty on some input parameters** → impacts the **uncertainty on the output**
- **Black-box and time-expensive simulators** → limited number of simulations

?

Uncertain input parameters

$\boldsymbol{X} = (X_1, \ldots, X_d)$



Numerical simulator

Uncertain output

$\boldsymbol{Y} = \mathcal{M}(X_1, \ldots, X_d)$

⇒ Quantify how the variability of the input parameters influences the output

→ Aim of **Sensitivity Analysis (SA)** *Saltelli et al. [2000]*

Sensitivity Analysis (SA) in Uncertainty treatment
⇒ *In numerical simulation*

- **Quantitative SA and Ranking purpose:**
  - Quantify the impact of each uncertain input and interaction → Ranking

  → Identify the variables to be fixed or further characterized in order to obtain the largest reduction of the output uncertainty

- **Screening purpose. Separate the inputs into two groups: influential and non-influential**
  - Non-influential variables fixed without consequences on the output uncertainty
  - In support of model reduction
  - To build a simplified model, a metamodel

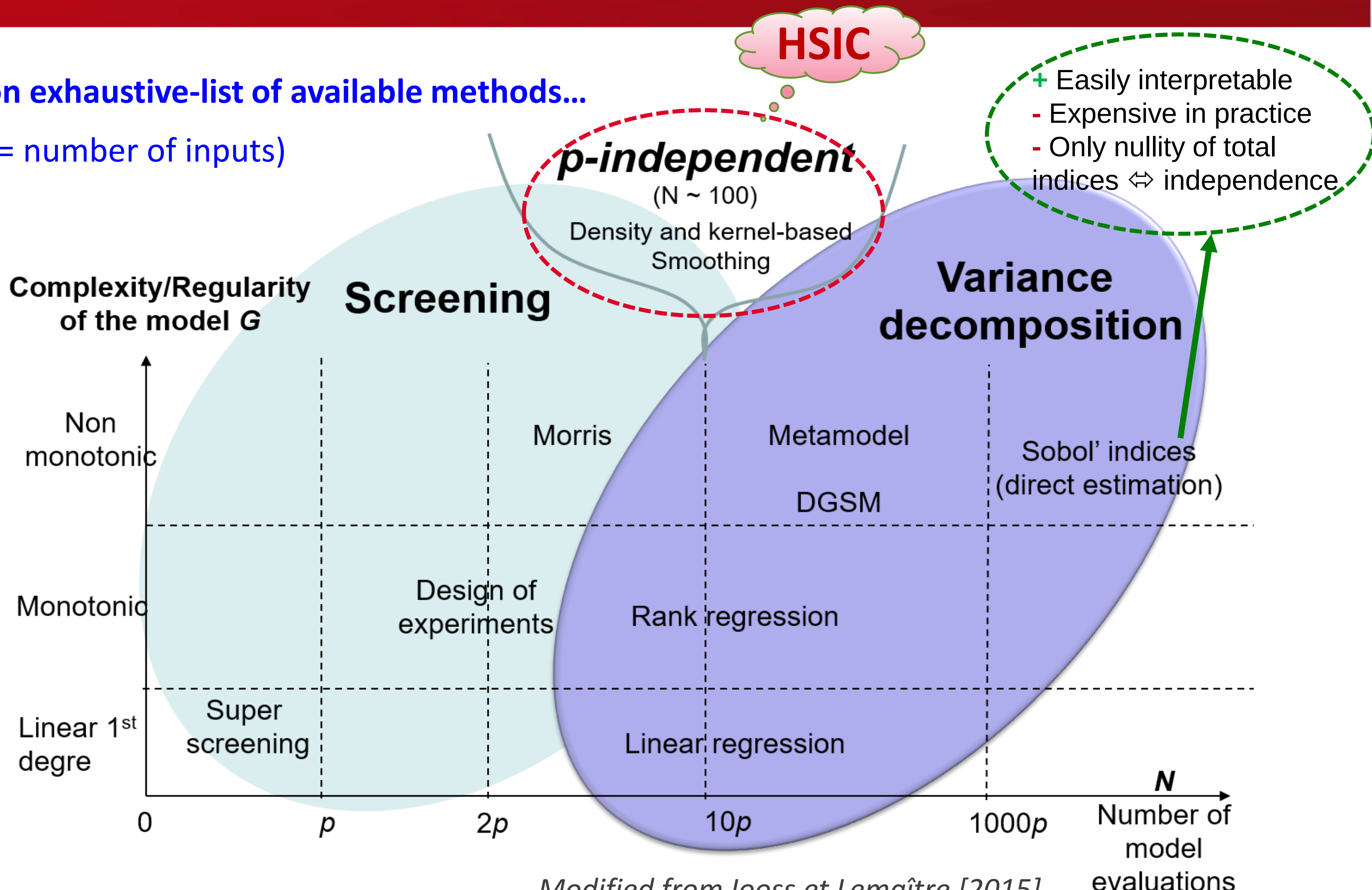
Global SA within a probabilistic framework

→ Valuable information to understand $\mathcal{M}$ and underlying phenomenon

Global Sensitivity Analysis of numerical simulators

Non exhaustive-list of available methods…

(p = number of inputs)



Modified from Iooss et Lemaître [2015]

HSIC* Review

**Hilbert Schmidt Independence Criterion*

IRESNE | DER | SESI | LEMS

Research Institute on Nuclear Systems for Low Carbone Energy Production

A few notations

Black-box model

$$\boldsymbol{Y} = \mathcal{M}(X_1, \dots, X_d)$$

- $X_1, \dots, X_d$ are *d* independent inputs, evolving in domain $\mathcal{X}_1, \dots, \mathcal{X}_d$
- $\boldsymbol{Y}$ evolves in domain $\mathcal{Y}$
- $\boldsymbol{P_X}$ denotes the probability distribution of X and $\boldsymbol{p_X}$ its density if is X continuous
- $\boldsymbol{P_{Y|X}}$: the conditional distribution of Y given X
- $\boldsymbol{P_{X,Y}}$: the joint probability measure and $\boldsymbol{P_Y} \otimes \boldsymbol{P_X}$ the product of marginal distributions

$\mathcal{M}$ unknown, only a ***n*-sample of simulations** $\left(\boldsymbol{X}^{(j)}, \boldsymbol{Y}^{(j)}\right)_{1 \le j \le n}$ where $Y^{(j)} = \mathcal{M}\left(X^{(j)}\right)$ for $j = 1, \dots, n$

GSA built upon RKHS framework

- **How to evaluate the sensitivity in a probabilistic way? ⇔ independence**

→ By comparing $P_{X,Y}$ with $P_X \otimes P_Y$

$$S_i = d\left(P_{X_i,Y}, P_{X_i} \otimes P_Y\right)$$

where *d* a dissimilarity measure between two probablity distributions

d can be based on **Maximum Mean Discrepancy**:

$$\mathrm{MMD}(\mathbb{P}, \mathbb{Q}) = \sup_{f \in \mathcal{F}} \left[\mathbb{E}_{\mathbb{P}} f(Y) - \mathbb{E}_{\mathbb{Q}} f(Y)\right]$$

with $\mathcal{F}$ = **unit ball in a (characteristic) Reproducing Kernel Hilbert Space (RKHS)**

(Sriperumbudur et al. [2008])

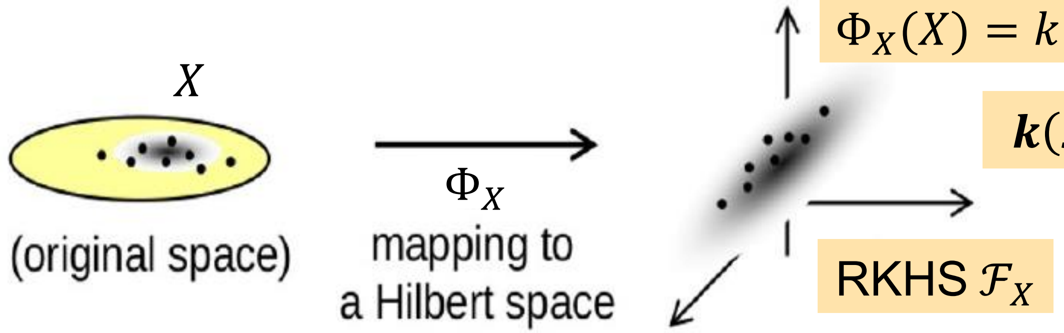
$$\Rightarrow \mathrm{S_i} = MMD^2\left(P_{X,Y}, P_Y \otimes P_X\right) = HSIC(X,Y)$$

Hilbert-Schmidt Independence Criterion

GSA built upon RKHS framework

▶ Link between embeddings in RKHS and MMD

Association of RKHS $\mathcal{F}_X$ to X: *with* Φ_X mapping function from $\mathcal{X}$ to $\mathcal{F}_X$ where the inner product is defined by kernel k (s.d.f. function).



$\Phi_X(X) = k(\bullet, X)$ *- canonical feature*

$$\boldsymbol{k(x, x') = \langle \Phi(x), \Phi(x') \rangle}$$

Picture modified from Arlot's slides [2014]

Define embedding of $\mathbb{P}$ in RKHS: $\boldsymbol{\mu_{\mathbb{P}}} \in \mathcal{F}_X$ such as $\boldsymbol{E_{\mathbb{P}}[f(X)] = \langle f, \mu_{\mathbb{P}} \rangle_{\mathcal{F}_X}}$ for all $f \in \mathcal{F}_X$

Reproducing property $\Rightarrow \boldsymbol{\mu_{\mathbb{P}} = E_{\mathbb{P}}[k(\bullet, X)] = E_{\mathbb{P}}[\Phi_X(X)]}$

$$\Rightarrow \mathrm{MMD}(\mathbb{P}, \mathbb{Q}) = \sup_{f \in \mathcal{F}_X, \|f\|_{\mathcal{F}_X} \leq 1} \left[\mathbb{E}_{\mathbb{P}} \boldsymbol{f(Y)} - \mathbb{E}_{\mathbb{Q}} \boldsymbol{f(Y)}\right] = \left\|\boldsymbol{\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}}\right\|_{\boldsymbol{\mathcal{F}}}$$

Kernel trick (reproducing prop.) $\Rightarrow \mathrm{MMD}^2(\mathbb{P}, \mathbb{Q}) = \mathbb{E}[k(X, X')] + \mathbb{E}[k(Y, Y')] - 2\mathbb{E}[k(X, Y)]$

where $X, X' \sim \mathbb{P}$ and $Y, Y' \sim \mathbb{Q}$

HSIC definition

▶ **MMD² applied between $P_{X,Y}$ and $P_Y \otimes P_X \Rightarrow HSIC(X,Y)_{\mathcal{F}_{X_i},\mathcal{F}_Y}$**

With $\mathcal{F}_{X_i}$ and $\mathcal{F}_Y$ **RKHS** associated to X_i and Y, resp.

k_X, k_Y, kernels on $\mathcal{X}$, $\mathcal{Y}$ associated with RKHS $\mathcal{F}_X$,$\mathcal{F}_Y$

$k_X k_Y$ kernel on $\mathcal{X} \times \mathcal{Y}$ associated with RKHS $\mathcal{F}_X \otimes \mathcal{F}_Y$

Steinwart and Christmann [2008]

$$\Rightarrow HSIC(X,Y)_{\mathcal{F}_{X_i},\mathcal{F}_Y} = MMD^2_{\mathcal{F}_{X_i},\mathcal{F}_Y}\left(P_{X,Y}, P_Y \otimes P_X\right) = \left\|\mu_{P_{X,Y}} - \mu_{P_Y \otimes P_X}\right\|^2_{\mathcal{F}_X,\mathcal{F}_Y}$$

Hilbert-Schmidt Independence Criterion

Gretton et al. [2005]

Kernel trick (reproducing property) again yields:

$$\mathrm{HSIC}(X,Y) = \mathbb{E}[k_X(X,X')k_Y(Y,Y')] + \mathbb{E}[k_X(X,X')]\mathbb{E}\left[k_y(Y,Y')\right] - 2\mathbb{E}\left[\mathbb{E}[k_X(X,X')|X]\mathbb{E}[k_Y(Y,Y')|Y]\right]$$

where (X',Y') is an independent and identically distributed copy of $(X,Y) \sim P_{X,Y}$.

HSIC definition

➢ **HSIC is in fact a *"super generalized"* covariance** *(Gretton et al. [2005])*

1. Cross-covariance operator $C_{X,Y}$ between the feature maps

$$C_{X,Y} = \mathbb{E}_{X,Y}[\Phi_X(X)\otimes\Phi_Y(Y)] - \mathbb{E}_X[\Phi_X(X)]\otimes\mathbb{E}_Y[\Phi_Y(Y)] = \mathbb{E}_{X,Y}[\Phi_X(X)\otimes\Phi_Y(Y)] - \mu_{P_X}\otimes\mu_{P_Y}$$

2. HSIC corresponds to the squared **Hilbert-Schmidt norm of** $C_{X,Y}$

$$\mathrm{HSIC}(X,Y)_{\mathcal{F}_X,\mathcal{F}_Y} = \left\|C_{X,Y}\right\|_{HS}^2 = \left\|\boldsymbol{\mu}_{\boldsymbol{P}_{X,Y}} - \boldsymbol{\mu}_{\boldsymbol{P}_Y\otimes\boldsymbol{P}_X}\right\|^2$$

⇨ **A larger panel of input-output dependency can be captured by this operator,**

HSIC somehow "summarizes" the cross-cov between feature maps applied to *X* and *Y*

► ***Characteristic kernels and RKHS*** ⇨ *Injective feature map*

⇨ ***Equivalence to independence:*** $\boldsymbol{HSIC(X,Y) = 0 \iff X \perp Y}$

Ex: Gaussian Kernel

$$k(x_i, x_i') = exp\left(-\frac{(x_i - x_i')^2}{2\lambda^2}\right)$$

HSIC estimation

➢ **HSIC** expression with only expectations of kernels

$$\mathrm{HSIC}(X,Y) = \mathbb{E}[k_X(X,X')k_Y(Y,Y')] + \mathbb{E}[k_X(X,X')]\mathbb{E}[k_y(Y,Y')] - 2\mathbb{E}\left[\mathbb{E}[k_X(X,X')|X]\mathbb{E}[k_Y(Y,Y')|Y]\right]$$

where (X',Y') is an independent and identically distributed copy of (X,Y).

- **Monte-Carlo estimator** from a ***n*-sample of simulations** $\left(\boldsymbol{X_i^{(j)}}, \boldsymbol{Y^{(j)}}\right)_{\boldsymbol{1\leq j\leq n}}$

$$\widehat{\textbf{HSIC}}(\boldsymbol{X},\boldsymbol{Y}) = \frac{\mathbf{1}}{\boldsymbol{n^2}}\boldsymbol{Tr}(\boldsymbol{K_X H L_Y H})$$

(Gretton et al. [2005])

where $H = I_n - \frac{1}{n}$, $K_X = \left(k_X\left(X^{(j)}, X^{(j')}\right)\right)_{1\leq j,j'\leq n}$ and $L_Y = \left(k\left(Y^{(j)}, Y^{(j')}\right)\right)_{1\leq j,j'\leq n}$

- **Statistical properties of $\widehat{\textbf{HSIC}}$:**
  - Asymptotically unbiased, variance of order *O*(1/n)
  - If $\boldsymbol{X} \perp \boldsymbol{Y}$, $n\widehat{\textbf{HSIC}}(\boldsymbol{X},\mathbf{Y})$ converges asymptotically to a Gamma distribution

HSIC for sensitivity analysis

IRESNE | DER | SESI | LEMS

Research Institute on Nuclear Systems for Low Carbone Energy Production

HSIC-based sentivity indices and independence test

- **Normalization for sensitivity analysis:**

(Da Veiga [2015])

$$\boxed{R^2_{HSIC,i} = \frac{HSIC\,(X_i,Y)}{\sqrt{HSIC(X_i,X_i)HSIC\,(Y,Y)}}}$$

⇨ $R^2_{HSIC} \in [0,1]$ for easier interpretation

$\Rightarrow$ **Use for ranking of inputs**

$$\textbf{Influence}(X_{[1]}) > \textbf{Influence}(X_{[2]}) > \cdots > \textbf{Influence}(X_{[d]})$$

Where order $[\cdot]$ is such that $\widehat{R^2_{H,X_{[1]}}} > \widehat{R^2_{H,X_{[2]}}} > \cdots > \widehat{R^2_{H,X_{[d]}}}$

- **Several illustrations on analytical examples and industrial applications**
  - HSIC detect non-influential factors easily and robustly, even with small sample size
  - HSIC indices can capture a large spectrum of dependence

    → Good ranking on usual GSA functions from sample size n~ 100

    → Efficiency for **screening** → Even better: **HSIC independence test**

HSIC-based Independence test

- **Use HSIC for screening → with Independence test**

$HSIC(X_i, Y) = 0 \Leftrightarrow X_i \perp Y$ *(with characteristic kernels!)*

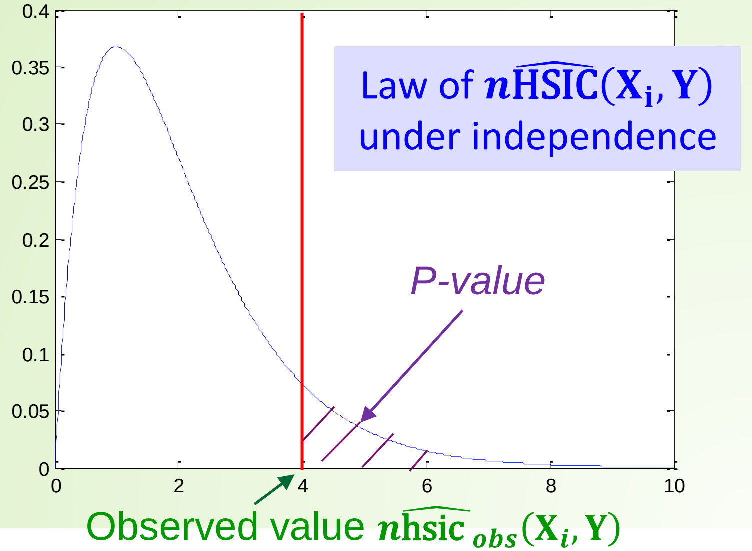
- Null hypothesis: $\mathcal{H}_0 : X_i \perp Y_i$ against $\mathcal{H}_1 : X_i \not\perp Y_i$
- Test statistics: $n\widehat{\mathrm{HSIC}}(X_i, Y)$
- Decision rule to obtain a test of level $\alpha = \mathbb{P}_{\mathcal{H}_0}[\text{reject } \mathcal{H}_0]$ (α fixed at 5% or 10%)

$\mathcal{H}_{0,i}$ rejected iff $n\widehat{\mathrm{HSIC}}(X_i, Y) > q_{1-\alpha}$ where $q_{1-\alpha}$ is the $(1-\alpha)$quantile under $\mathcal{H}_{0,i}$

→**Test function:** $\Delta_\alpha = \mathbf{1}_{n\widehat{\mathrm{HSIC}}(X_i,Y) > q_{1-\alpha}}$

- In practice, computation of p-value:

p-value $= \mathbb{P}[\widehat{\mathrm{HSIC}}(X_i, Y) > \widehat{\mathrm{hsic}}_{obs}(\mathrm{X_i}, \mathrm{Y})]$



⇒ **How to have the distribution of $n\widehat{\mathrm{HSIC}}(\mathrm{X_i}, \mathrm{Y})$ under $\mathcal{H}_0$?**

HSIC-based Independence test

⇒ How to have the distribution $n\widehat{\mathrm{HSIC}}(\mathrm{X_i}, \mathrm{Y})$ under $\mathcal{H}_0$ to compute *p-value*?

- **Asymptotic** computation with Gamma approximation for large *n* *(Gretton et al. (2008])*

- **Permutation-based** approximation for smaller sample size *n* *(De Lozzo & Marrel (2016a], Meynaoui et al. [2019])*

Algorithm 1 – Permutation-based independence test (for each X_i)

Require: The learning sample $(\mathbf{X}_i, \mathbf{Y})$ of n inputs/outputs $\{(X_i^{(1)}, Y^{(1)}), \ldots, (X_i^{(n)}, Y^{(n)})\}$, B and α

1: Compute $\widehat{\mathrm{HSIC}}_{obs}(X_i, Y)$ from Eq. (2)

2: Generate B permutation-based samples $(\mathbf{X}_i, \mathbf{Y}_{[b]})_{1 \leq b \leq B}$

3: Compute the B permutation-based estimators $\left(\widehat{\mathrm{HSIC}}_b(X_i, Y)\right)_{1 \leq b \leq B}$ by replacing $\mathbf{Y}$ by $\mathbf{Y}_{[b]}$ in Eq. (2)

4: Estimate the p-value by Monte-Carlo estimator $\hat{p}^B_{val,i} = \dfrac{1}{B}\sum_{b=1}^{B} \mathbb{1}_{\widehat{\mathrm{HSIC}}_b(X_i,Y) > \widehat{\mathrm{HSIC}}_{obs}(X_i,Y)}$

5: **if** $\hat{p}^B_{val,i} < \alpha$ **then**

6: **return** reject $(\mathscr{H}_0^i)$

7: **else**

8: **return** accept $(\mathscr{H}_0^i)$

9: **end if**

HSIC-based Independence test

⇒ How to have the distribution $n\widehat{\mathrm{HSIC}}(\mathbf{X_i}, \mathbf{Y})$ under $\mathcal{H}_0$ to compute *p-value*?

- **Asymptotic** computation with Gamma approximation for large *n* *(Gretton et al. (2008])*
- **Permutation-based** approximation for smaller sample size *n* *(De Lozzo & Marrel (2016a], Meynaoui et al. [2019])*

→ Theoretical demonstration: permuted-test is of level α *(Meynaoui et al. [2019])*

→ Empirically observed: power of asymptotic and permutation test equivalent for a sufficient number *B* of permutations

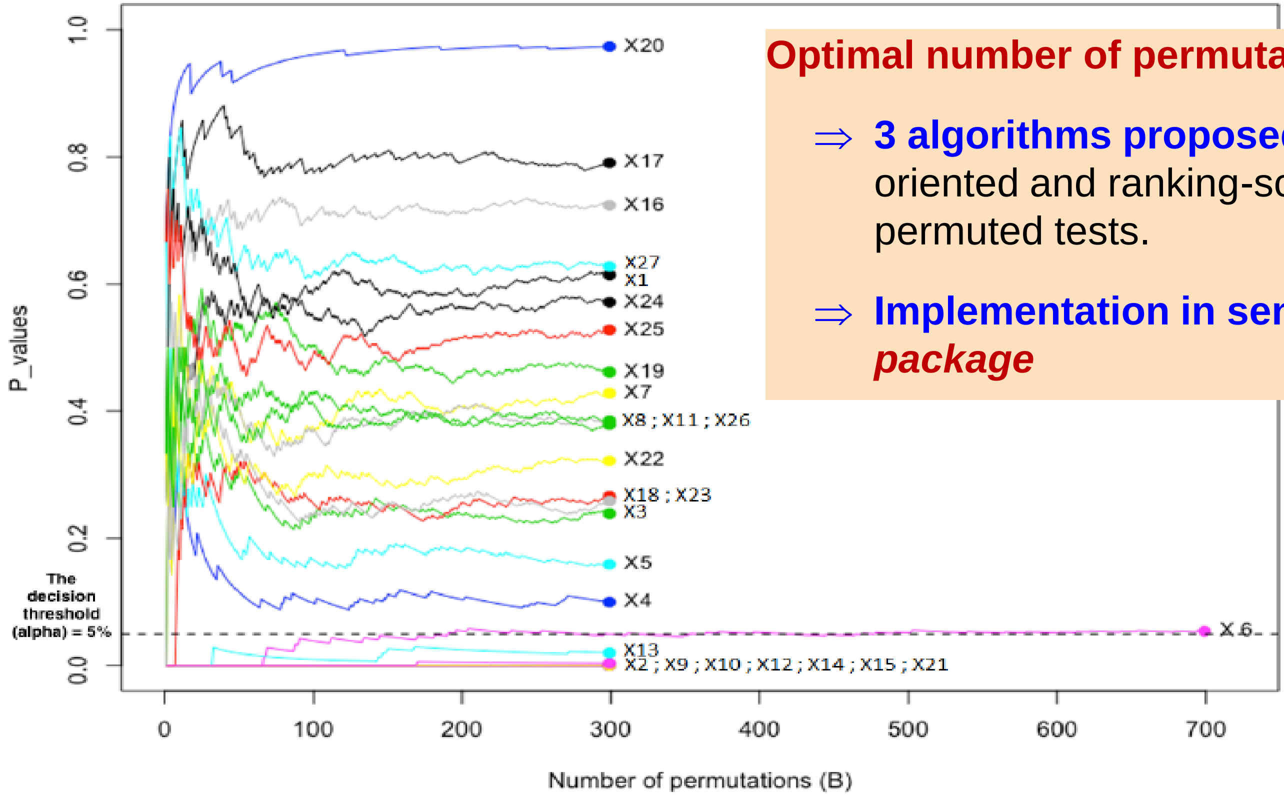
→ Extension for non independent identically distributed samples: *(El Amri & Marrel [2021b])*

- ✓ ok for Latin Hypercube sample (LHS)
- ✓ Correction with Conditional Randomization Test for scrambled low-discrepancy sequence *(El Amri & Marrel [2021b])*
- ☒ Not possible for space-filling LHS

HSIC-based Independence test

- **Permutation-based** approximation *(El Amri & Marrel [2021a])*

In practice, which number *B* of permutations to accurately estimate p-value?



Optimal number of permutations according to **the final goal**

⇒ **3 algorithms proposed**: screening-oriented, ranking-oriented and ranking-screening-oriented sequential permuted tests.

⇒ ***Implementation in sensiHSIC of R Sensitivity package***

In practice reduction of *B* to few hundreds.

⇒ Sensitivity studies and convergence studies more tractable

HSIC-based Independence test

- Estimation of p-value with **Pearson III**-approximation *(El Amri & Marrel [2021b])*

$$\widehat{\mathrm{HSIC}}(X,Y) = \frac{1}{n^2} Tr(K_X H L_Y H) = \frac{1}{n^2} Tr(\tilde{K}_X \tilde{L}_Y)$$

With $\tilde{K}_X = (HK_XH)$ and $\tilde{L}_Y$=(HL_YH) double centered matrices

$A \in \mathbb{R}^{n\times n}$ symmetric and with centered rows
$W = UU^T$ with $U \in \mathbb{R}^{n\times p}$ and centered columns
$S = Tr(AW)$

$\Rightarrow$ Distribution of S when rows of U are randomly permuted can be **approximated by a Pearson III** + expression of 3 first moments

(Kazi-Aoual et al. [1994])

- Application to permuted $\widehat{\mathrm{HSIC}}_b(X,Y)$ statistics with A=$\tilde{K}_X$ and W = $\tilde{L}_Y$
  - $\Rightarrow$ Direct approximation, no permutation required
  - $\Rightarrow$ Fast and very efficient
  - $\Rightarrow$ **Best method for non-asymptotic framework**

anr agence nationale de la recherche SAMOURAI

Solutions for more powerful HSIC tests

IRESNE | DER | SESI | LEMS

Research Institute on Nuclear Systems for Low Carbone Energy Production

Solutions for more powerful HSIC-based test

- **Limitation: as mentionned before, usual kernels often depend on parameter**

For 1-D Gaussian kernel: bandwidth parameter λ ($\lambda > 0$), estimated in practice either with empirical standard deviation, or empirical median

$$k_G(x_i, x_i') = exp\left(-\frac{(x_i - x_i')^2}{2\lambda^2}\right)$$

→ Heuristic choices without theoretical justifications

→ Impact on the **power** of the HSIC-based test (and quantitative interpretation of HSIC)

- **Solution 1: Aggregated HSIC-tests to take into account several kernels**

  → Improve the power of tests: theoretical demonstration on level and power

  → Improve the robustness of HSIC-based screening

Albert et al. [2021] and Meynaoui's PhD [2019]

Solution 1 for more powerful test: aggregation

- **Methodology of aggregated testing procedure of level α?**

1. Consider a countable **collection of positive bandwidths** $\Lambda = (\lambda_i)_i$ and $\mathcal{U} = (\mu_i)_i$ and a collection of positive weights $\{\omega_{\lambda,\mu} \;/\; (\lambda,\mu) \in \Lambda \times \mathcal{U}\}$ such that $\sum_{(\lambda,\mu)\in\Lambda\times\mathcal{U}} e^{-\omega_{\lambda,\mu}} \leq 1$

2. Define a level for each single test of the collection $(\lambda,\mu) \in \Lambda \times \mathcal{U}$: $\boldsymbol{u_\alpha e^{-\omega_{\lambda,\mu}}}$

   Single test rejects $\mathcal{H}_0 \iff \widehat{HSIC}_{\lambda,\mu} > q^{\lambda,\mu}_{1-u_\alpha e^{-\omega_{\lambda,\mu}}}$

   where $\boldsymbol{u_\alpha}$ is the less conservative value such that the aggregated test is of level $\boldsymbol{\alpha}$:

$$u_\alpha = \sup\left\{u > 0 \;;\; \mathbb{P}_{f_X \otimes f_Y}\left[\sup_{(\lambda,\mu)\in\Lambda\times\mathcal{U}}\left(\widehat{HSIC}_{\lambda,\mu} > q^{\lambda,\mu}_{1-u_\alpha e^{-\omega_{\lambda,\mu}}}\right) > 0\right] \leq \alpha\right\}$$

   $\Rightarrow$ *In practice $\boldsymbol{u_\alpha}$ estimated by permutation-based approach*

3. Reject independence if there is **at least one single test** that rejects $\mathcal{H}_0$

$$\Delta_\alpha = 1 \iff \sup_{(\lambda,\mu)\in\Lambda\times\mathcal{U}}\left(\widehat{HSIC}_{\lambda,\mu} > q^{\lambda,\mu}_{1-u_\alpha e^{-\omega_{\lambda,\mu}}}\right) > 0$$

Solution 2 for more powerful test: optimal bandwidths

- <u>Solution 2</u>: **HSIC-tests with optimal bandwidths** *(El Amri & Marrel [2021b])*

$$(\boldsymbol{\lambda}, \boldsymbol{\mu})_n^* = \underset{\boldsymbol{\lambda},\boldsymbol{\mu}}{\operatorname{argmax}} \ \widehat{\boldsymbol{HSIC}}_{\boldsymbol{\lambda},\boldsymbol{\mu}}$$

$$\widehat{\boldsymbol{HSIC}}^* = \widehat{\boldsymbol{HSIC}}_{(\boldsymbol{\lambda},\boldsymbol{\mu})_n^*}$$

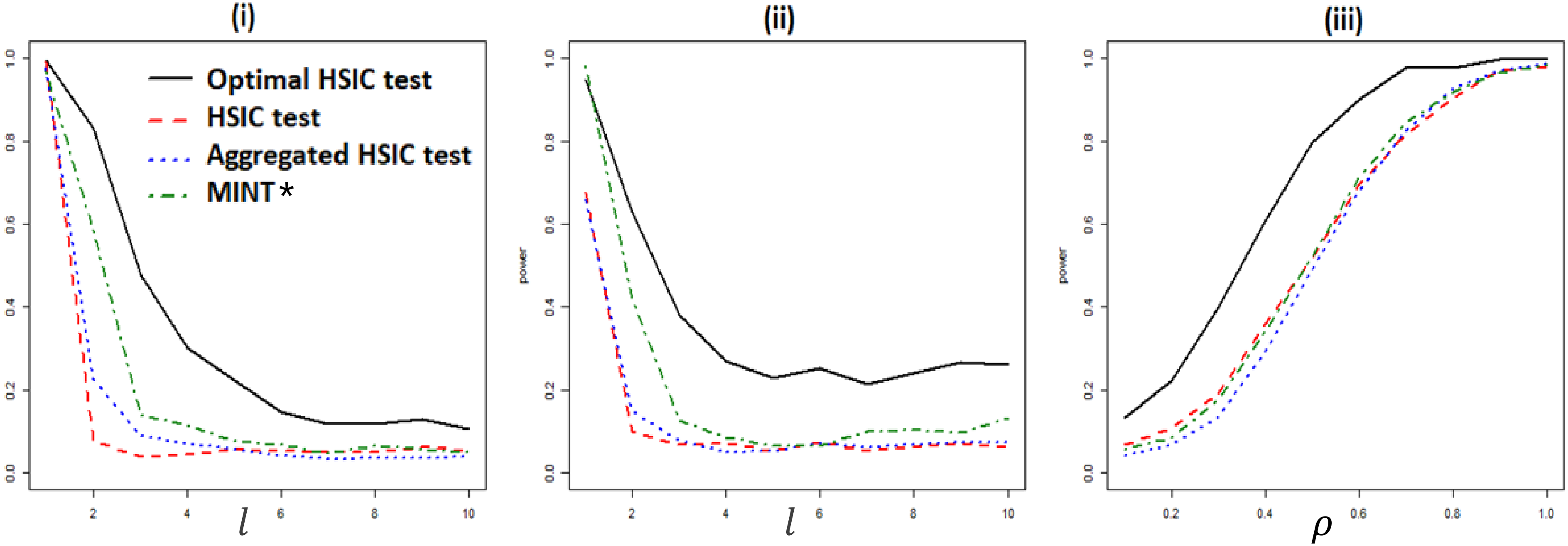
- **Methodology of optimal-bandwidth test:**
  - Optimization solved using HSIC gradient and kernel derivatives
  - Adaptation of permutation-based method by re-estimating the optimal bandwidths for each permuted sample
  - Use of sequential permutation to optimize the number of permutation

Numerical comparisons

▶ **Illustration on analytical examples** *(from Berreth & Samworth [2019], details in Appendix)*

Results obtained from 1000 i.i.d. Monte-Carlo samples of size *n=100*

Power curves of independence tests according to shape parameters l and ρ



**MINT : Mutual Information-based test*

Extension for Target and Functional Sensitivity Analysis

IRESNE I **DER** I **SESI** I **LEMS**

Research Institute on Nuclear Systems for Low Carbone Energy Production

Playing with kernels ⇨ large spectrum of applications

- **Goal-oriented SA for safety studies** *(Marrel & Chabridon [2021])*

  $\Rightarrow$ To measure the input influence in a **restricted output domain**: $Y \in \mathcal{C}$

  $\Rightarrow$ Numerous applications for safety and risk assessment ($\mathcal{C} = \{Y | Y > q_{0.9}\}$, e.g.)

- **Technical point: choose the characteristic kernel according to the type of data:**

  - Output = « *Is Y in domain $\mathcal{C}$?* »

    - Target SA: measures the influence of X over **the occurrence** of $Y \in \mathcal{C}$

    → Bernouilli output: $\mathbf{1}_{Y \in \mathcal{C}}(Y) \sim \mathcal{B}(p_{\mathcal{C}})$ with $p_{\mathcal{C}} = \mathbb{P}[Y \in \mathcal{C}]$ ⇨ Dirac Kernel

    - Conditional SA: performed within $\mathcal{C}$ only, ignoring what happens outside

    → Real output: $\boldsymbol{Y | Y \in \mathcal{C}}$ with $\mathbb{P}_{|Y \in \mathcal{C}}[\mathcal{A}] = \frac{\mathbb{P}[\mathcal{A} \cap Y \in \mathcal{C}]}{p_{\mathcal{C}}}$ ⇨ Gaussian kernel

Playing with kernels ⇨ large spectrum of applications

Goal-oriented HSIC for safety studies

(Marrel & Chabridon [2021])

⇒ **Brute versions:**

- Target SA: $\boldsymbol{HSIC(X, 1_{Y\in\mathcal{C}}(Y))}$ with Dirac Kernel
- Conditional SA: $\boldsymbol{HSIC(X, Y|Y \in \mathcal{C})}$ with Gaussian Kernel

⇒ **Smoother versions** to cope with the loss of information and take into account some information outside $\mathcal{C}$ → **Use of weight function $W_{\mathcal{C}}$ for relaxation**

$$W_{\mathcal{C}} : \mathcal{Y} \longrightarrow [0,1]$$

$$W_{\mathcal{C}}(y) = e^{-d_{\mathcal{C}}(y)/s} \text{ and } d_{\mathcal{C}}(y) = \inf_{y' \in \mathcal{C}} \|y - y'\|$$

→ $\boldsymbol{HSIC(X, W_{\mathcal{C}}(Y))}$ and $\boldsymbol{HSIC(X, W_{\mathcal{C}}(Y)Y|Y \in \mathcal{C})}$

Similar use for optimization purpose in Spagnol et al. [2019]

✓ **Implementation in sensiHSIC of *R Sensitivity package* :**

Estimators of $\boldsymbol{HSIC(X, W_{\mathcal{C}}(Y))}$ + asymptotic and permuted-based tests

Playing with kernels ⇨ large spectrum of applications

▶ SA for functional data

(El Amri & Marrel [2021b])

Output is a random function of time or space $\Rightarrow$ which kernel?

One solution: combine

1. Functional Principal Component Analysis (FPCA)
2. Truncation with the *q* first terms: $Y(t) - \mu(t) \approx \sum_{k=1}^{q} \boldsymbol{U}_k \, \varphi_k(t)$
3. Weigthed kernel based on the *q* FPCA (random) coefficients $(\boldsymbol{U}_k)_{\boldsymbol{k=1,\dots,q}}$

$$k(\mathbf{Y}^{(l)}, \mathbf{Y}^{(m)}) = \sum_{h=1}^{q} w_h k(||U_h^{(l)} - U_h^{(m)}||_2^2)$$

Where *k* is a usual kernel for real variables (Gaussian e.g.) and weights w_h correspond to the percentage of variance explained by each component U_h (cf. eigenvalues from FPCA)

Playing with kernels ⇨ large spectrum of applications

SA for functional data

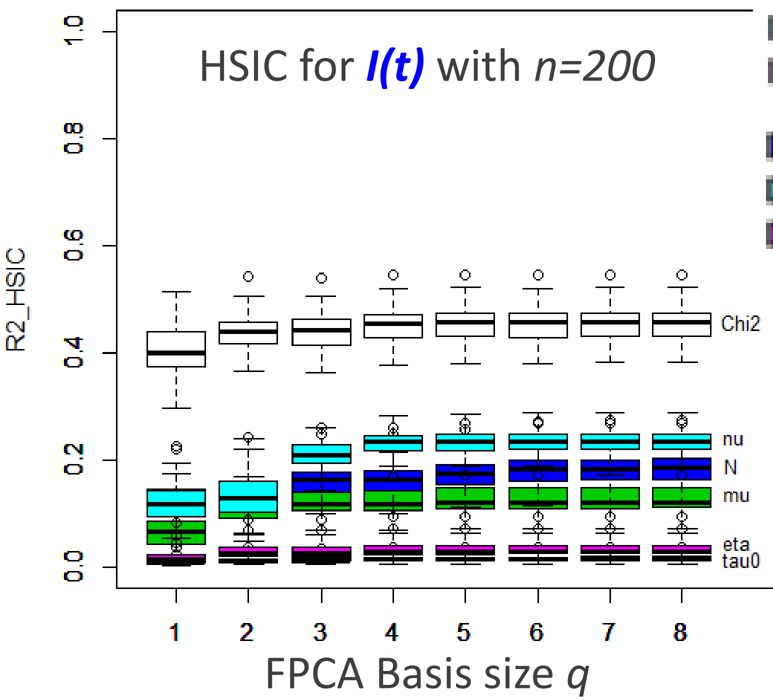
(El Amri & Marrel [2021b])

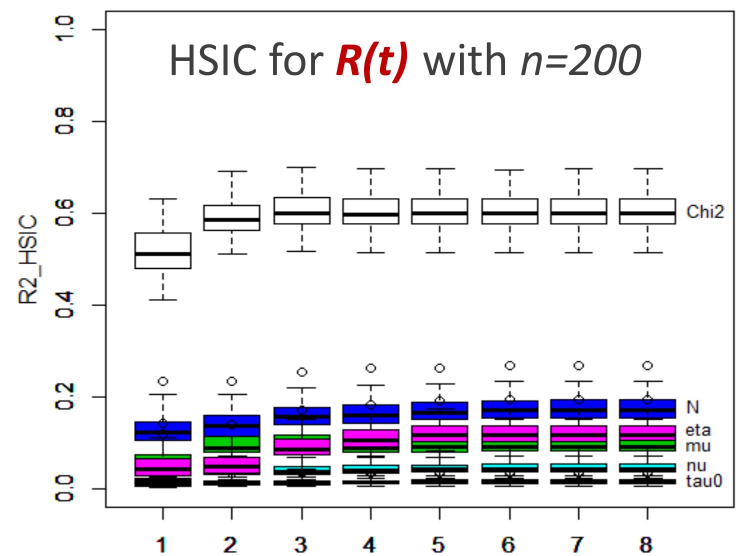
Illustration on compartmental epidemiologic model on COVID-19

Modified SIR model (Susceptible – Infected – Recovered) with 6 uncertain inputs

I(t) and *R(t)*: number of asymptomatic and reported symptomatic infectious individuals at time *t*

- Initial conditions, involved in $I(t_0)$
- Transmission rate before the epidemic outbreak
- Rate of the exponential decay of transmission as soon as social distancing and lockdown start
- Date of effect of government measures
- Average periods during which *I(t)* are infectious
- Average periods during which *R(t)* are infectious





Conclusion and prospects

IRESNE I DER I SESI I LEMS

Research Institute on Nuclear Systems for Low Carbone Energy Production

Conclusions

- **HSIC as indices of Sensitivity Analysis**
  - Focus the SA analysis on the difference between $P_{X,Y}$ with $P_X \otimes P_Y$
  - Power of RKHS → HSIC=one of the most successful non-parametric dependence measure
  - Capture a large spectrum of relationships
  - Able to deal with many factors and purposes (ranking, screening, goal-oriented SA)
  - **Characterize independence** → efficient for screening and building independence tests !

- **HSIC-tests of independence for screening**
  - Rigorous statistical framework, control of 1st and 2nd kind error
  - Asymptotic and several non-asymptotic versions
  - **P-value** of test → Really efficient for screening and for quantitative SA

→ **Efficiency demonstrated in numerous industrial applications, especially with small sample size and large dimension**

Prospects

Limitations and prospects remain in HSIC SA indices

- Decomposition into main effects & interactions must be investigated

  ⇒ Assess the use of **HSIC with ANOVA-like kernels** and **Shapley-HSIC for dependent inputs** (*Da Veiga [2021]*)

  ⇒ Build associated independence tests

  ⇒ Assess power for screening and relevancy for ranking

anr agence nationale de la recherche — SAMOURAI

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

- Invariance properties → Preliminary isoprobabilistic transformation? (*Poczos et al. (2018)*)
- Extension "functional" HSIC-tests with other reduction techniques like Dynamic Time Warping?

Some References 1/2

- Albert, M, Laurent, B., Marrel, A. and Meynaoui, A. Adaptive test of independence based on HSIC measures. To appear in Annals of Statistics, 2022.
- Baucells, M. and Borgonovo, E. Invariant probabilistic sensitivity analysis. Management Science, 59 (11), pp. 2536-2549, 2013.
- Berrett, T. B. and Samworth, R. J. Nonparametric independence testing via mutual information. Biometrika, 106(3):547{566, 2019.
- Borgonovo, E. A new uncertainty importance measure, Reliability Engineering & System Safety, 92(6), 771–784, 2007.
- Da Veiga S., Global sensitivity analysis with dependence measures, *Journal of Statistical Computation and Simulation*, 85:1283-1305, 2015.
- Da Veiga S., Kernel-based anova decomposition and shapley effects–application to global sensitivity analysis. arXiv preprint arXiv:2101.05487. 2021.
- De Lozzo M. and Marrel A., New improvements in the use of dependence measures for sensitivity analysis and screening, *Journal of Statistical Computation and Simulation*, 86:3038-3058, 2016
- El Amri R. and Marrel A. HSIC-based independence tests with optimal sequential permutations: application for sensitivity analysis of numerical simulators, *Quality and Reliability Engineering International*, 2021a.
- El Amri R. and Marrel A., More powerful HSIC-based independence tests, extension to space-filling designs and functional data, https://hal-cea.archives-ouvertes.fr/cea-03406956, 2021b.
- Gretton, Bousquet, Smola and Schölkopf. Measuring statistical dependence with Hilbert-Schmidt norms. *In: Proceedings Algorithmic Learning Theory*, 2015.
- Iooss B. and Lemaître P. A review on global sensitivity analysis methods. In Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications, Springer, 2015.
- Kraskov, A., Stogbauer, H. and Grassberger, P. Estimating mutual information', Physical Review E 69(6), 066138.1–066138.16, 2004.

Some References 2/2

- Marrel A. and Chabridon V. Statistical developments for target and conditional sensitivity analysis: application on safety studies for nuclear reactor, Reliability Engineering and System Safety 214, 2021.
- Meynaoui, A. New developments around dependence measures for sensitivity analysis: application to severe accident studies for generation IV reactors. PhD thesis, University of Toulouse, 2019.
- Müller, A. Integral probability metrics and their generating classes of functions, Advances in Applied Probability 29(2), 429–443, 1997.
- Spagnol A., Le Riche R. and Da Veiga S. Global Sensitivity Analysis for Optimization with Variable Selection, SIAM/ASA Journal on Uncertainty Quantification, Vol. 7, No. 2 : pp. 417-443, 2019.
- Poczos, Z. Ghahramani and J. Schneider. Copula-based Kernel Dependency Measures. https://arxiv.org/abs/1206.4682, 2018.
- Saltelli, A., Chan, K., and Scott, E., editors. Sensitivity analysis. Wiley Series in Probability and Statistics. Wiley., 2000.
- Sobol, I. Sensitivity estimates for non linear mathematical models. Mathematical Modelling and Computational Experiments, 1:407–414, 1993.
- Sriperumbudur B.K., Gretton A., and Fukumizu K. Hilbert space embeddings and metrics on probability measures. Journal of Machine Learning Research, (11):1517-1561, 2010.
- Zhang, Q., Filippi, S., Gretton, A. et al. Large-scale kernel methods for independence testing. Stat Comput 28, 113–130, 2018.

Thank you for your attention !

DE LA RECHERCHE À L’INDUSTRIE

APPENDIX

Numerical comparisons of standard and solutions 1 & 2

- **Illustration on analytical examples** *(usually used in indep. test comparison)*

3 different mechanisms of dependence, varying by a shape parameter (l or ρ)

Berreth & Samworth [2019]

(i). Defining the joint density f_l, $l = 1, \dots, 10$ of (X, Y) on $[-\pi, \pi]$ by

$$f_l(x, y) = \frac{1}{4\pi^2}\{1 + \sin(lx)\sin(ly)\}$$

(ii). Considering X and Y as $X = L\cos\Theta + \frac{\varepsilon_1}{4}, \quad Y = L\sin\Theta + \frac{\varepsilon_2}{4}$,

where L, Θ, ε_1 and ε_2 are independent, with $L \sim \mathcal{U}\{1, \dots, l\}$

for $l = 1, \dots, 10$, $\Theta \sim \mathcal{U}[0, 2\pi]$ and $\varepsilon_1, \varepsilon_2 \sim \mathcal{N}(0, 1)$.

(iii). Defining $X \sim \mathcal{U}[-1, 1]$. For a given $\rho = 0.1, 0.2, \dots, 1$, Y is defined as $Y = |X|^\rho \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, 1)$ independent with X.