Kernel-based sensitivity analysis for (excursion) sets

N. Fellmann[1,2], C. Blanchet-Scalliet[1], C. Helbert[1], A. Spagnol[2], and D. Sinoquet[2]

[1]Univ. Lyon, École Centrale de Lyon, CNRS UMR 5208, Institut Camille Jordan, 36 Avenue Guy de Collongue, 69134 Écully, France,
[2]IFP Energies Nouvelles
{noe.fellmann,christophette.blanchet,celine.helbert}@ec-lyon.fr
{noe.fellmann,delphine.sinoquet,adrien.spagnol}@ifp.fr

Abstract

This work is motivated by goal-oriented sensitivity analysis of inputs/outputs of complex simulators. We are interested in the ranking of the uncertain input variables that impact the most a feasible design domain. Most sensitivity analysis methods deal with scalar outputs. In this paper, we propose a way to perform sensitivity analysis when dealing with set-valued outputs. Our new methodology is driven by sensitivity analysis on excursion sets but can also be applied to set-valued simulators as in viability field, or when dealing with maps such as pollutant concentration maps or flooding zone maps.

We propose a method based on the Hilbert Schmidt Independence Criterion (HSIC) with a kernel tailored to sets as outputs. A first contribution is the proof that this kernel is *characteristic* (i.e injectivity of the embedding in the associated Reproducing Kernel Hilbert Space), a required property for the HSIC interpretation in a sensitivity analysis context. We propose then to compute the HSIC-ANOVA indices which give a decomposition of the input contributions. Using these indices, we identify which inputs should be neglected (*screening*) and we rank the others by influence (*ranking*). The proposed method is tested on two test cases of excursion sets.

Keywords: Sensitivity indices, random sets, HSIC, kernel methods

1 Introduction

In many industrial applications, complex physical systems are modeled by time consuming numerical models. Solving associated optimization problems is thus challenging: the dimension of the input space can be very high, the inputs can be deterministic but also uncertain and the optimization problems must often be solved under constraints and in presence of uncertainties. In this framework of high dimensional optimization under constraints with uncertainties (Beyer and Sendhoff [2007]), the optimization requires too many model evaluations to be performed directly. Simplification of the model is then necessary prior to the optimization. It often means reducing the input dimension which can be done by selecting the most impacting input variables through sensitivity analysis (SA).

Global SA techniques study the influence of inputs on scalar-valued output (see Da Veiga et al. [2021] for a general review on sensitivity analysis methods). Most known SA methods are the screening methods with the Morris method for instance (Morris [1991]) and the variance-based sensitivity measures. The latter, called Sobol' indices, have emerged with Sobol's work in 1993, who introduced them later in Sobol [2001]. These indices quantify the part of output variance which can be attributed to one or a group of inputs. This variance decomposition is called the Hoeffding ANOVA decomposition. However, even if these indices are widely used, a main drawback is that they quantify input contributions to the output variance but not to the whole output distribution. To deal with this issue, new sensitivity indices have been developed such as the Borgonovo indices (Borgonovo [2007]) or Cramér von Mises indices (Gamboa et al. [2017]). Instead of comparing the variance with and without conditioning, they compare the whole distributions. However their computation require a heavier computational cost. To circumvent this issue, kernel-based sensitivity indices have been more recently introduced. Their computational cost is lightened by the *kernel trick*. The idea is to use a kernel function which embeds distributions in Reproducing Kernel Hilbert Space (RKHS) on which distances are easier to compute. By embedding the output distribution in the RKHS, it is then possible to quantify the influence of an input by measuring its impact on the embedded distributions (Gretton et al. [2005b]). An ANOVA-like decomposition also exists for kernel-based indices as shown in Da Veiga [2021]. Kernel-based methods also enable to easily deal with vector-valued outputs. Such extension is also possible with variance-based indices but requires more work (Gamboa et al. [2013]).

In the context of optimization, the aforementioned SA methods have been goal oriented in order to identify which inputs are influential on reaching the optimal point or area (Marrel and Chabridon [2021], Spagnol [2020]). However they only consider the optimization without uncertainties and aim at reducing the dimension of the design space. In the presence of uncertain inputs, we also want to reduce the dimension of the uncertain space. To do so, we propose to study the impact of the uncertain inputs on the feasible sets. This, however, mean that we now deal with sets as output on which we aim at conducting a sensitivity analysis.

In literature, several works on SA adapted to complex outputs exist. For example, in El Amri and Marrel [2021], the case of functional output is studied. [Fort et al. 2021] also proposed indices that only require the output space to have a metric. In the previously introduced list of standard sensitivity indices, kernel-based approaches are the least dependent on the type of output. Indeed they only rely on the definition of an appropriate kernel. Thus the extension of kernel-based SA for more complex outputs seems easier. In this article we introduce a new kernel-based SA adapted to the particular case of set-valued output. The key ingredient to perform SA for set-valued output is to have a kernel on sets. In Section 2 we introduce a particular kernel k_{set} which is the adaptation of the Euclidean Gaussian kernel to a space of sets. We show that it is characteristic which is needed for screening purposes. Using k_{set}, we propose a kernel-based index, that we call H_{set}, which is the adaptation of HSIC-ANOVA indices to set-valued output. An efficient estimation of this index is given in Section 3. Numerical results obtained from two test cases related to excursion sets are then given in Section 4. Some proofs and numerical results are given in appendix.

2 Kernel-based sensitivity analysis for sets

Let $X \subset \mathbb{R}^d$ be a compact space. Without loss of generality, we suppose $X = [0,1]^d$. Let λ be the Lebesgue measure on $\mathbb{R}^d$ and $\mathcal{B}(X)$ the associated Lebesgue σ-algebra. Let $\eta : \mathbb{R}^p \to \mathcal{B}(X)$ be a set-valued model defined as a function of p inputs $\boldsymbol{u} = (u_1, ..., u_p) \in \mathbb{R}^p$, with the associated output $\gamma = \eta(\boldsymbol{u}) \in \mathcal{B}(X)$. As in many sensitivity analysis framework, the inputs are assumed to be independent random variables $\boldsymbol{U} = (U_1, ..., U_p) \in \mathbb{R}^p$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with known distributions $\mathbb{P}_{U_i}$. The associated output $\Gamma = \eta(\boldsymbol{U})$ is then a random variable which support is a subspace of $\mathcal{B}(X)$. We will call it a random set.

2.1 A kernel between sets

The first requirement for applying kernel methods in our context is to define a kernel on sets. A kernel on $\mathcal{B}(X)$ is by definition a positive definite function $k : \mathcal{B}(X) \times \mathcal{B}(X) \to \mathbb{R}$. Let Δ be the symmetric difference between two sets A and B defined by $A \Delta B = (A \cup B) \backslash (A \cap B)$. Let $\lambda(A)$ be the volume of a set A defined by $\lambda(A) = \int_X \mathbb{1}_A d\lambda$. We propose the function k_{set} defined by:

$$\forall \gamma_1, \gamma_2 \in \mathcal{B}(X), \quad k_{set}(\gamma_1, \gamma_2) := e^{-\frac{\lambda(\gamma_1 \Delta \gamma_2)}{2\sigma^2}}, \tag{1}$$

with σ a positive scalar. k_{set} is inspired by the classical Gaussian kernel as the volume of the symmetric difference is equal to the L^2 norm of the difference between two indicator functions.

Proposition 2.1. *The function* $k_{set} : \mathcal{B}(X) \times \mathcal{B}(X) \to \mathbb{R}$ *defined by*

$$\forall \gamma_1, \gamma_2 \in \mathcal{B}(X), \quad k_{set}(\gamma_1, \gamma_2) = e^{-\frac{\lambda(\gamma_1 \Delta \gamma_2)}{2\sigma^2}},$$

is positive definite for any positive scalars σ *which mean that* k_{set} *is a kernel.*

Proof. The proof is the same as in Balança and Herbin [2012], Lemma 2.1., replacing the indexing collection $\mathcal{A}$ by $\mathcal{B}(X)$. □

2.2 Kernel mean embedding of random sets

In (1), the volume of the symmetric difference $\delta : \gamma_1, \gamma_2 \mapsto \lambda(\gamma_1 \Delta \gamma_2)$ is a pseudo-metric in $\mathcal{B}(X)$. Indeed, $\delta(\gamma_1, \gamma_2) = 0$ only implies that $\gamma_1 = \gamma_2$ λ-almost everywhere. By considering the quotient space $\mathcal{B} = \mathcal{B}(X)/\sim$ for the equivalence relation $\gamma_1 \sim \gamma_2$ iif $\gamma_1 = \gamma_2$ λ-a.e., δ becomes a metric sometimes called Fréchet-Nikodym-Aronszajn metric (Marczewski and Steinhaus [1958]). On this space $\mathcal{B}$, k_{set} is well defined and remains a kernel.

Proposition 2.2. *The kernel* k_{set} *is bounded and measurable from* $\mathcal{B} \times \mathcal{B}$ *(with the sigma algebra generated by the open sets generated by the open balls for the metric* δ*) to* $\mathbb{R}$*.*

Proof. k_{set} is bounded by 1 and is measurable because it is continuous as *exp* and δ are continuous (δ is a metric). □

Then the Moore-Aronszajn theorem (Aronszajn [1950]) gives the existence of a unique RKHS $\mathcal{H} \subset \mathcal{B}^{\mathbb{R}}$ of reproducing kernel k_{set}. We can now embed borelian sets distributions into the RKHS $\mathcal{H}$.

Definition 2.1. *Let $\mathcal{M}^1_+(\mathcal{B})$ be the space of probability measures on $\mathcal{B}$. The mean embedding of $\mathcal{M}^1_+(\mathcal{B})$ in $\mathcal{H}$ is defined as:*

$$\begin{array}{rccl} & \mathcal{M}^1_+(\mathcal{B}) & \to & \mathcal{H} \\ \mu \ : & \mathbb{P} & \mapsto & \mu_{\mathbb{P}} = \int_{\mathcal{B}} k_{set}(\gamma, \cdot) d\mathbb{P}(\gamma). \end{array}$$

In the sequel, for a given random set Γ of probability distribution $\mathbb{P}_\Gamma$, we will denote μ_Γ its mean embedding instead of $\mu_{\mathbb{P}_\Gamma}$.

The existence of this embedding and that $\mu(\mathcal{M}^1_+(\mathcal{B})) \subset \mathcal{H}$ is ensured by $\int_{\mathcal{B}} k_{set}(\gamma, \gamma) d\mathbb{P}(\gamma) = 1 < +\infty$ for any $\mathbb{P} \in \mathcal{M}^1_+(\mathcal{B})$ (Smola et al. [2007]).

Proposition 2.3. *The kernel k_{set} is characteristic, i.e. the embedding defined in Definition 2.1 is injective.*

Proof. The proof is based on the Proposition 5.2. of Ziegel et al. [2022]. We apply the Proposition with $\mathcal{X} = \mathcal{B}$, $H = L^2(X)$, $\varphi = \exp(-\frac{\cdot}{2\sigma^2})$ and T defined by $T(\gamma) := x \mapsto \mathbb{1}_\gamma(x)$ for any $\gamma \in \mathcal{B}$. $L^2(X)$ is a separable Hilbert space (as X is compact so separable) and T is injective and measurable (as $\delta(\gamma_1, \gamma_2) = ||\mathbb{1}_{\gamma_1} - \mathbb{1}_{\gamma_2}||_2^2$). We just need to verify that $\mathcal{B}$ is a Polish space i.e. separable and complete as it is already a metric space. To do so, we will see $\mathcal{B}$ as a subspace $\tilde{\mathcal{B}}$ of $L_2(X)$ through the homeomorphism $T : \mathcal{B} \to \tilde{\mathcal{B}} = \{f \in L_2(X),\ \exists \gamma \in \mathcal{B}(X) \text{ s.t. } f = \mathbb{1}_\gamma\ \lambda\text{-a.e.}\}$. $\tilde{\mathcal{B}}$ is separable as it is a subspace of $L_2(X)$. As $L_2(X)$ is complete, $\tilde{\mathcal{B}}$ is also complete if it is closed. Let's then show that $\tilde{\mathcal{B}}$ is closed. Let $\mathbb{1}_{\gamma_n} \underset{n\to+\infty}{\overset{L_2}{\to}} f$ with $\gamma_n \in \mathcal{B}(X)\ \forall n$. The $L_2(X)$ convergence implies that there is a subsequence $\mathbb{1}_{\gamma_{\phi(n)}}$ that converges a.e. pointwise to f. It means that there exists a λ-null set $\mathcal{N}$ s.t. $\forall x \notin \mathcal{N}, \mathbb{1}_{\gamma_{\phi(n)}}(x) \underset{n\to+\infty}{\to} f(x) \in \{0, 1\}$ (as it is a limit of a sequence of 0 and 1). So,

$f = \mathbb{1}_{f^{-1}(\{1\})}$ λ-a.e. and $f^{-1}(\{1\}) \in \mathcal{B}(X)$ as f is measurable. Thus k_{set} is integrally strictly positive definite with respect to $\mathcal{M}(\mathcal{B})$, the set of signed measure on $\mathcal{B}$, which implies that it is characteristic. $\square$

2.3 Maximal Mean Discrepancy between sets

With this injective embedding, taking the distance in the RKHS induces a distance between two elements of $\mathcal{M}_+^1(\mathcal{B})$. As it coincides with the Maximum Mean Discrepancy (MMD) taken in the RKHS $\mathcal{H}$, we will call it MMD_{set} (Borgwardt et al. [2006]).

Definition 2.2. *Let Γ_1, Γ_2 be two random closed sets of distribution $\mathbb{P}_{\Gamma_1}, \mathbb{P}_{\Gamma_2}$ with mean embeddings μ_{Γ_1} and μ_{Γ_2}. The Maximum Mean Discrepancy (MMD) taken on the space $\mathcal{H}$ is defined as the distance between their mean embeddings:*

$$\text{MMD}_{set}(\mathbb{P}_{\Gamma_1}, \mathbb{P}_{\Gamma_2}) := ||\mu_{\Gamma_1} - \mu_{\Gamma_2}||_{\mathcal{H}}.$$

It can be written as a sum of expectations of the kernel k_{set}.

Proposition 2.4. *Let Γ_1, Γ_2 be two random sets and Γ_1', Γ_2' be two independent copies of Γ_1, Γ_2. The squared MMD can then be expressed as:*

$$\text{MMD}_{set}^2(\mathbb{P}_{\Gamma_1}, \mathbb{P}_{\Gamma_2}) = \mathbb{E}[k_{set}(\Gamma_1, \Gamma_1')] + \mathbb{E}[k_{set}(\Gamma_2, \Gamma_2')] - 2\mathbb{E}[k_{set}(\Gamma_1, \Gamma_2)].$$

Proof. It is the same as in Gretton et al. [2006] given for scalars outputs.

$$\begin{aligned}
\mathrm{MMD}^2_{set}(\mathbb{P}_{\Gamma_1}, \mathbb{P}_{\Gamma_2}) &= ||\mu_{\Gamma_1} - \mu_{\Gamma_2}||^2_{\mathcal{H}} \\
&= \langle \mu_{\Gamma_1}, \mu_{\Gamma_1} \rangle + \langle \mu_{\Gamma_2}, \mu_{\Gamma_2} \rangle - 2\langle \mu_{\Gamma_1}, \mu_{\Gamma_2} \rangle \\
&= \mathbb{E}\langle k_{set}(\Gamma_1, \cdot), k_{set}(\Gamma_1', \cdot)\rangle + \mathbb{E}\langle k_{set}(\Gamma_2, \cdot), k_{set}(\Gamma_2', \cdot)\rangle \\
&\quad - 2\mathbb{E}\langle k_{set}(\Gamma_1, \cdot), k_{set}(\Gamma_2, \cdot)\rangle \\
&= \mathbb{E}k_{set}(\Gamma_1, \Gamma_1') + \mathbb{E}k_{set}(\Gamma_2, \Gamma_2') - 2\mathbb{E}k_{set}(\Gamma_1, \Gamma_2).
\end{aligned}$$

□

In the case of scalar outputs, the MMD between two random variable distributions has been used to define kernel-based sensitivity indices such as the Hilbert-Schmidt Independence Criterion (HSIC). The latter is designed for screening through independence testing. But recent works (Da Veiga [2021]) have shown that under some hypotheses, HSIC has an ANOVA-like decomposition which makes it also suited for ranking. For these reasons, based on the MMD_{set}, we propose an extension of the HSIC definition to set-valued outputs.

2.4 HSIC for random sets

From Gretton et al. [2005b], Gretton et al. [2005a], HSIC is defined as the MMD-based distance between the distribution of the couple (input,output) and the product of their marginal. To define this distance, a kernel on the joint space must be defined. The natural candidate is the tensor product between the input and output kernels.

Definition 2.3. *Let $k = k_A \otimes k_{set}$ be a kernel inducing the RKHS $\mathcal{H}_k$, k_A being a kernel* *on the input space $\mathcal{U}_A$. Let U_A be a group of inputs and Γ a random set output. The* HSIC *between U_A and Γ, denoted H_{set} is defined as:*

$$\mathrm{H}_{set}(U_A, \Gamma) := \mathrm{MMD}^2_{set}(\mathbb{P}_{U_A,\Gamma}, \mathbb{P}_{U_A}\mathbb{P}_\Gamma) = ||\mu_{(U_A,\Gamma)} - \mu_{U_A} \otimes \mu_\Gamma||^2_{\mathcal{H}_k}.$$

The product kernel k is characteristic as soon as k_A and k_{set} are characteristic. k_{set} is characteristic (Proposition 2.3), so, assuming that k_A is a characteristic kernel, k inherits of the characteristic property. This implies that $\mathrm{H}_{set}(U_A, \Gamma) = 0$ iif U_A and Γ are independent. Screening by independence testing is then possible through the test $\mathcal{H}_0 :\ \mathrm{H}_{set}(U_A, \Gamma) = 0$ versus $\mathcal{H}_1 :\ \mathrm{H}_{set}(U_A, \Gamma) > 0$.

Using Proposition 2.4, H_{set} can be expressed as a sum of expectations of kernels.

Proposition 2.5 (Gretton et al. [2005a] for scalar outputs)**.** *With the same notations as Definition 2.3, and with $({U_A}', \Gamma'$ an independent copy of (U_A, Γ),*

$$\begin{aligned}\mathrm{H}_{set}(U_A, \Gamma) =& \mathbb{E}[k_A(U_A, {U_A}')k_{set}(\Gamma, \Gamma')] + \mathbb{E}[k_A(U_A, {U_A}')]\mathbb{E}[k_{set}(\Gamma, \Gamma')] \\ &- 2\mathbb{E}[\mathbb{E}[k_A(U_A, {U_A}')|U_A]\mathbb{E}[k_{set}(\Gamma, \Gamma')|\Gamma]].\end{aligned}$$

This property is one of the strengths of the HSIC because it makes its estimation easy as we will see in the next part. Another advantage of this index is that an ANOVA-like decomposition exists (see Da Veiga [2021]). However, this requires strong assumptions especially on the input kernel which must be an ANOVA kernel.

Definition 2.4 (Orthogonal and ANOVA kernel). *Let $\mathcal{U}$ be a measurable space. A kernel $k : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$ is said to be orthogonal with respect to a measure $\nu \in \mathcal{M}_1^+(\mathcal{U})$ if:*

$$\forall u \in \mathcal{U}, \int_{\mathcal{U}} k(u,z)d\nu(z) = 0.$$

A kernel $K : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$ is said to be ANOVA w.r.t. ν if it can be decomposed as $K = 1 + k$ with k being orthogonal w.r.t. ν.

Theorem 2.1 (ANOVA decomposition of H_{set}). *Assuming that:*

i. The inputs $U_1, ..., U_p$ are mutually independent.

ii. Each input has an ANOVA kernel K_i w.r.t. the input distribution $\mathbb{P}_{U_i}$. For any group of inputs $\boldsymbol{U}_A$ with $A \subset \{1, ..., p\}$, the associated kernel is defined by:

$$K_A = \bigotimes_{i \in A} K_i.$$

iii. For any $A \subset \{1, ..., p\}$, $\mathbb{E}[K_A(\boldsymbol{U}_A, \boldsymbol{U}_A)] < +\infty$ and $\mathbb{E}[k_{set}(\Gamma, \Gamma)] < +\infty$.

Then the ANOVA decomposition of the H_{set} is given by:

$$\mathrm{H}_{set}(\boldsymbol{U}, \Gamma) = \sum_{A \subseteq \{1,\ldots,p\}} \sum_{B \subseteq A} (-1)^{|A|-|B|} \, \mathrm{H}_{set}\left(\boldsymbol{U}_B, \Gamma\right).$$

This theorem is proven in Da Veiga [2021] for scalar outputs but is also valid for other outputs as long as the Mercer theorem holds.

An ANOVA input kernel also simplifies the HSIC expression of Proposition 2.5 which becomes:

$$\mathrm{H}_{set}(U_A, \Gamma) = \mathbb{E}\left[(K_A(U_A, U_A{}') - 1)k_{set}(\Gamma, \Gamma')\right]. \tag{2}$$

Using the decomposition of Theorem 2.1, associated SA indices can be defined.

Definition 2.5 (HSIC-ANOVA indices on sets)**.** *Similarly to the ANOVA decomposition of the Sobol indices, the HSIC-ANOVA first and total indices on sets, can be defined by:*

$$\forall i \in \{1, ..., d\}\ S_i^{\mathrm{H}_{set}} := \frac{\mathrm{H}_{set}(U_i, \Gamma)}{\mathrm{H}_{set}(\boldsymbol{U}, \Gamma)} \textit{ and } S_{T_i}^{\mathrm{H}_{set}} := 1 - \frac{\mathrm{H}_{set}(\boldsymbol{U}_{-i}, \Gamma)}{\mathrm{H}_{set}(\boldsymbol{U}, \Gamma)},$$

with $\boldsymbol{U}_{-i} = (U_1, ..., U_{i-1}, U_{i+1}, ..., U_d)$*. This indices can be generalized for groups of inputs.*

It is important to note that the three assumptions of Theorem 2.1, and especially ii., restrict the choice of kernels but only of the input kernel. Hopefully, on the input space $\mathbb{R}^d$, some kernels are known to be ANOVA. For instance, Sobolev kernels are ANOVA. With $r = 1$, it is defined by:

$$\begin{aligned} k_{sob}(x, y) &= 1 + B_1(x)B_1(y) + \frac{1}{2}B_2(|x - y|) \\ &= 1 + (x - \frac{1}{2})(y - \frac{1}{2}) + \frac{1}{2}[(x - y)^2 - |x - y| + \frac{1}{6}], \end{aligned} \tag{3}$$

with B_i the Bernoulli polynomial of degree i. This kernel is ANOVA w.r.t. the uniform law on $[0, 1]$. However, it is also possible to construct ANOVA kernels from classical kernels. One of the possible transformations is proposed in Ginsbourger et al. [2016]. For a given kernel k, an ANOVA kernel k_{ANOVA} w.r.t. ν is defined by:

$$\begin{aligned} k_{ANOVA}(x,y) = 1 + k(x,y) - \int k(x,z)d\nu(z) - \int k(z,y)d\nu(z) \\ + \int\int k(z,z')d\nu(z)d\nu(z'). \end{aligned} \quad (4)$$

In the case of inputs uniformly distributed on $[0,1]$, the previous transformation is analytically known for some classical kernels given in Appendix of Ginsbourger et al. [2016].

Given input kernels that verify the previous assumptions, we are now able to rank the input influence on the set-valued output Γ by ranking either their first order index $S_i^{\mathrm{H}_{set}}$ or their total index $S_{T_i}^{\mathrm{H}_{set}}$. Before ranking the inputs, it is also possible to use this indices for screening in order to reduce more roughly the number of inputs. Indeed, as shown in Sarazin et al. [2022],

$$S_i^{\mathrm{H}_{set}} = 0 \Longleftrightarrow S_{T_i}^{\mathrm{H}_{set}} = 0 \Longleftrightarrow U_i \perp \Gamma.$$

The authors also propose associated independence tests which are better in term of power than the usual tests performed with HSIC.

Thus, HSIC-ANOVA indices on sets are an answer to screen and rank the inputs of a set-valued model. The question of their estimation and the difficulties resulting from the presence of sets are raised in the next part.

3 Estimation

In this section we first recall classical estimation methods of the HSIC-ANOVA index. Then we explain how we manage our set-valued outputs and how it impacts the estimation of the indices.

3.1 HSIC-ANOVA index estimation

Given an iid sample $(U^{(i)}, \Gamma^{(i)}), i = 1, ..., n$, classical estimators of HSIC can be found in Gretton et al. [2007] and Song et al. [2007]. Adapted to expression (2), the unbiased estimator $\widehat{\mathrm{H}_{set}}\left(U_A, \Gamma\right)$ is defined by

$$\widehat{\mathrm{H}_{set}}\left(U_A, \Gamma\right) = \frac{2}{n(n-1)} \sum_{i<j}^{n} \left(K_A \left(U_A^{(i)}, U_A^{(j)} \right) - 1 \right) k_{set} \left(\Gamma^{(i)}, \Gamma^{(j)} \right).$$

The estimator of the associated normalized index is denoted $\widehat{S}_A^{\mathrm{H}_{set}}$.

3.2 Estimation of $k_{set}(\Gamma^{(i)}, \Gamma^{(j)})$

In most cases, the quantity $k_{set}(\Gamma^{(i)}, \Gamma^{(j)})$ also needs to be estimated. In practice, the quantity $\lambda(\Gamma^{(i)} \Delta \Gamma^{(j)})$ is estimated by discretizing the space X. Using that $\lambda(\Gamma^{(i)} \Delta \Gamma^{(j)}) = \lambda(X) \mathbb{E}_{X \sim \mathcal{U}(X)}[\mathbb{1}_{\Gamma^{(i)} \Delta \Gamma^{(j)}}(X) | (\Gamma^{(i)}, \Gamma^{(j)})]$, we write $k_{set}(\Gamma^{(i)}, \Gamma^{(j)})$ as:

$$k_{set}(\Gamma^{(i)}, \Gamma^{(j)}) = \exp\left(-\frac{\lambda(X)}{2\sigma^2} \mathbb{E}[\mathbb{1}_{\Gamma^{(i)} \Delta \Gamma^{(j)}}(X) | (\Gamma^{(i)}, \Gamma^{(j)})] \right),$$

with $X \sim \mathcal{U}(X)$. Using the previous expression and having an iid sample $(X_{i,j}^{(1)}, ..., X_{i,j}^{(m)})$ of X, we can estimate $k_{set}(\Gamma^{(i)}, \Gamma^{(j)})$ by:

$$\widehat{k_{set}}(\Gamma^{(i)}, \Gamma^{(j)}) = \exp\left(-\frac{\lambda(X)}{2\sigma^2}\frac{1}{m}\sum_{k=1}^{m}\mathbb{1}_{\Gamma^{(i)}\Delta\Gamma^{(j)}}(X_{i,j}^{(k)})\right).$$

We now inject this estimator in our previous estimators which gives a Nested Monte Carlo (NMC) estimator.

3.3 Nested estimation of the indices

Let us have an iid sample $(U^{(i)}, \Gamma^{(i)}), i = 1, ..., n$ of (U, Γ) and independent iid samples $(X_{i,j}^{(1)}, ..., X_{i,j}^{(m)})$ of $X \sim \mathcal{U}(X)$, for all $i < j \in \{1, ..., n\}$. Injecting the previous estimation of the kernel k_{set}, the NMC estimator of $\mathrm{H}_{set}(U_A, \Gamma)$ is given by:

$$\widehat{\mathrm{H}_{set}}^{nest}(U_A, \Gamma) = \frac{2}{n(n-1)}\sum_{i<j}^{n}\left(K_A\left(U_A^{(i)}, U_A^{(j)}\right) - 1\right)e^{-\frac{\lambda(X)}{2\sigma^2}\frac{1}{m}\sum_{k=1}^{m}\mathbb{1}_{\Gamma^{(i)}\Delta\Gamma^{(j)}}(X_{i,j}^{(k)})}. \quad (5)$$

Note that this estimator is biased as any NMC estimator (Rainforth et al. [2016]). For each couple $(\Gamma^{(i)}, \Gamma^{(j)})$, the previous estimation requires to test if $X_{i,j}^{(k)} \in \Gamma^{(i)}\Delta\Gamma^{(j)}$ for each $k \in \{1, ..., m\}$. Each of such tests requires to check if $X_{i,j}^{(k)} \in \Gamma^{(i)}$ and $X_{i,j}^{(k)} \in \Gamma^{(j)}$ which means $n(n-1)m$ tests which is not affordable. To circumvent this issue, we propose to reuse the same $X^{(k)}$ for each (i, j). By doing so, we only need to test if $X^{(k)} \in \Gamma^{(i)}$ for each

k and i which reduces the number of tests to nm. The estimator is then defined by

$$\widehat{\mathrm{H}_{set}}\left(U_A, \Gamma\right) = \frac{2}{n(n-1)} \sum_{i<j}^{n} \left(K_A\left(U_A^{(i)}, U_A^{(j)}\right) - 1 \right) e^{-\frac{\lambda(X)}{2\sigma^2} \frac{1}{m} \sum_{k=1}^{m} \mathbb{1}_{\Gamma^{(i)} \Delta \Gamma^{(j)}}(X^{(k)})}.$$

By taking only one sample of $X^{(k)}$, $\widehat{\mathrm{H}_{set}}\left(U_A, \Gamma\right)$ is no longer a classical Nested Monte Carlo Estimator (NMC). Still, we show that its quadratic risk goes to 0 and we detail an upper bound.

Proposition 3.1. *With the previous notations, we have*

$$\begin{aligned} &\mathbb{E}\left(\widehat{\mathrm{H}_{set}}\left(U_A, \Gamma\right) - \mathrm{H}_{set}(U_A, \Gamma)\right)^2 \\ &\quad \leq 2\left(\frac{2\sigma_1^2}{n(n-1)} + \frac{4(n-2)\sigma_2^2}{n(n-1)} + \frac{L^2\sigma_3^2}{m}\right), \end{aligned}$$

with

- $\sigma_1^2 = Var\left(\left(K_A\left(U_A, U_A{}'\right) - 1\right) k_{set}(\Gamma, \Gamma')\right)$,
- $\sigma_2^2 \ = Var\left(\mathbb{E}\left[\left(K_A\left(U_A, U_A{}'\right) - 1\right) k_{set}(\Gamma, \Gamma') | (\Gamma, U_A)\right]\right)$,
- $L = \frac{\lambda(X)}{2\sigma^2}$,
- $\sigma_3^2 = Var(\left(K_A\left(U_A, U_A{}'\right) - 1\right) \mathbb{E}\left[\mathbb{1}_{\Gamma\Delta\Gamma'}(X) | (U_A, U_A{}', \Gamma, \Gamma')\right])$,

where $(U_A{}', \Gamma')$ *is an independent copy of* (U_A, Γ).

Proof. The proof is given in appendix. □

The quadratic risk has a rate of $\mathcal{O}(\frac{1}{n} + \frac{1}{m})$ which tends to say that we should use $n = m$. In the case of classical NMC (without reusing the same sample of X), the rate is $\mathcal{O}(\frac{1}{n} + \frac{1}{m^2})$. As the previous bound is only an upper bound, we can expect to approach this classical convergence rate. The classical HSIC estimators are also known to work even with small n ($n = 100$ for instance). This comes from the fact that the constant σ_2 can be very small. This means that we can expect to have reached the convergence of our estimator without requiring to take high n and m. This will be highlighted in the numerical results presented in the next section. Using the estimator $\widehat{\widehat{\mathrm{H}}_{set}}$ we denote the $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ and $\hat{\hat{S}}_{T_i}^{\mathrm{H}_{set}}$ the estimators of the first and total order HSIC-ANOVA indices on sets.

4 Application on excursion sets

In industrial applications such as optimization or inversion, we seek for the set of feasible solutions. In presence of uncertainties, this set is random.

Definition 4.1. *Let $g : X \times \mathbb{R}^p \to \mathbb{R}$ be measurable on $X \times \mathbb{R}^p$. With $\boldsymbol{U}$ a random vector of $\mathbb{R}^p$,*

$$\Gamma_U = \{x \in X, \quad g(x, U) \leq 0\} \in \mathcal{B},$$

is called an excursion set.

With such sets, our goal is to quantify the influence of the different inputs U on the output Γ_U. To do so, we will consider two test cases. The first one is an analytically known function from El Amri et al. [2021] with two dimensions in both spaces X and $\mathcal{U}$ (section 4.1). On this first toy case, we will also study numerically the bound of the quadratic risk given in Proposition 3.1. The second one is associated to an optimization problem involving a stationary harmonic oscillator, from Cousin et al. [2022] on which we want to quantify the impact of some uncertain inputs on feasible sets (section 4.2). For each example, our sensitivity analysis is made in two steps:

— Screening: we compute the p-values associated to the test $H_{set}(U_i, \Gamma) = 0$ versus $H_{set}(U_i, \Gamma) > 0$. We use a permutation-based estimation. If the p-value is higher than 0.05, the input is negligible and if it is below 0.05, the input is influential.

— Ranking: we compute the first order indices $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ and total indices $\hat{\hat{S}}_{T_i}^{\mathrm{H}_{set}}$ of all inputs. The first order indices $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ are used to rank the inputs. With the total indices $\hat{\hat{S}}_{T_i}^{\mathrm{H}_{set}}$ we can quantify the HSIC interaction effects in our examples.

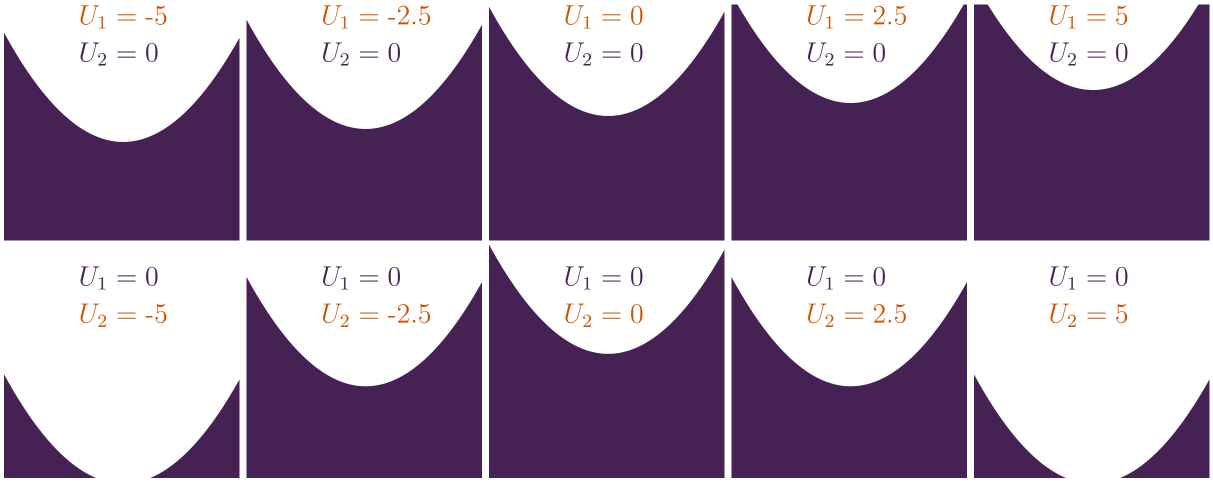

Figure 1 – Excursion set of the constraint $g \leq 0$ for $U_1 \in \{-5, -2.5, 0, 2.5, 5\}$ and $U_2 = 0$ (first row) and for $U_1 = 0$ and $U_2 \in \{-5, -2.5, 0, 2.5, 5\}$. (second row)

4.1 Toy function 1

In this part, we will estimate the previous HSIC-ANOVA indices on the excursion set Γ_U defined through the following function g from El Amri et al. [2021]

$$\forall x, u \in [-5, 5]^2 \times [-5, 5]^2 \quad g(x, u) = -x_1^2 + 5x_2 - u_1 + u_2^2 - 1.$$

Visual tests (Figure 1) highlight that u_2 seems to induce more changes in Γ_U than u_1. To verify this assumption, we compute the p-values and the indices $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$, $\hat{\hat{S}}_{T_i}^{\mathrm{H}_{set}}$ for each input $U_i \sim \mathcal{U}([-5, 5])$ for $i \in \{1, 2, 3\}$, U_3 being an additional dummy input which does not appear in function g. We use first $n = m = 100$ and then $n = m = 1000$. The hyperparameter σ^2 of the kernel k_{set} is taken equal to the empirical mean of $\lambda(\Gamma_i \Delta \Gamma_j)$ on $i > j$. This is a popular choice when using the Gaussian kernel for sensitivity analysis. We compute the indices for five characteristic ANOVA kernels. The first one is the Sobolev kernel with $r = 1$ presented in (3) and the four others are obtained using the transformation

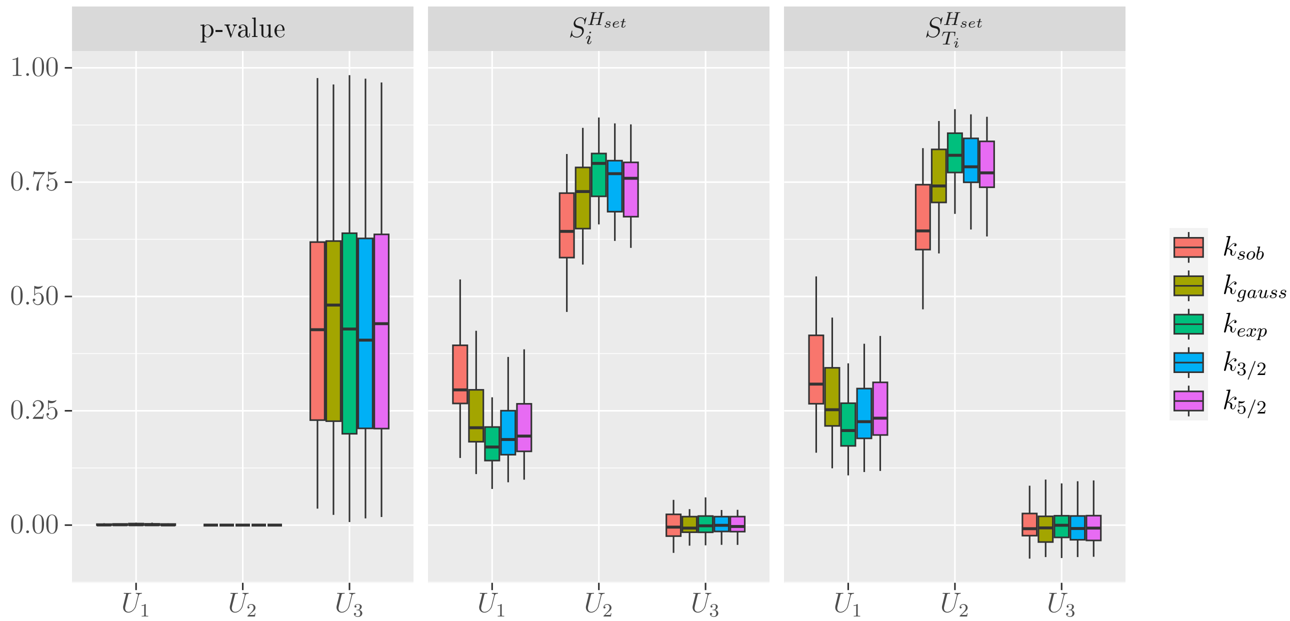

Figure 2 – Estimation of the p-values, $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ and $\hat{\hat{S}}_{T_i}^{\mathrm{H}_{set}}$ for the excursion set defined by the constraint $g \leq 0$ computed for 5 kernels with $n = 100$, $m = 100$ and repeated 20 times

given in (4) based on the following classical kernels:

— the Gaussian kernel, $k_{gauss}(x, y) = e^{-\frac{1}{2}\left(\frac{x-y}{\sigma}\right)^2}$ with $\sigma > 0$,

— the Laplace kernel, $k_{exp}(x, y) = e^{-\frac{|x-y|}{h}}$ with $h > 0$,

— the Matérn 3/2, $k_{3/2}(x, y) = \left(1 + \sqrt{3}\frac{|x-y|}{h}\right) e^{-\sqrt{3}\frac{|x-y|}{h}}$ with $h > 0$,

— the Matérn 5/2, $k_{5/2}(x, y) = \left(1 + \sqrt{5}\frac{|x-y|}{h} + \frac{5}{3}\frac{|x-y|}{h^2}\right) e^{-\sqrt{5}\frac{|x-y|}{h}}$ with $h > 0$.

As the inputs are not uniformly distributed on $[0, 1]$, we apply the inverse of the cumulative distribution function before the calculations. We compute the indices 20 times in order to obtain the boxplots given in Figure 2 for $n = m = 100$ and in Figure 3 for $n = m = 1000$.

With these results, we are able to class the inputs as influential or negligible (screening) and to rank them by influence (ranking). The p-values are below 0.05 for U_1 and U_2 which class them as influential. However, U_3 has always a p-value higher than 0.05 and is correctly

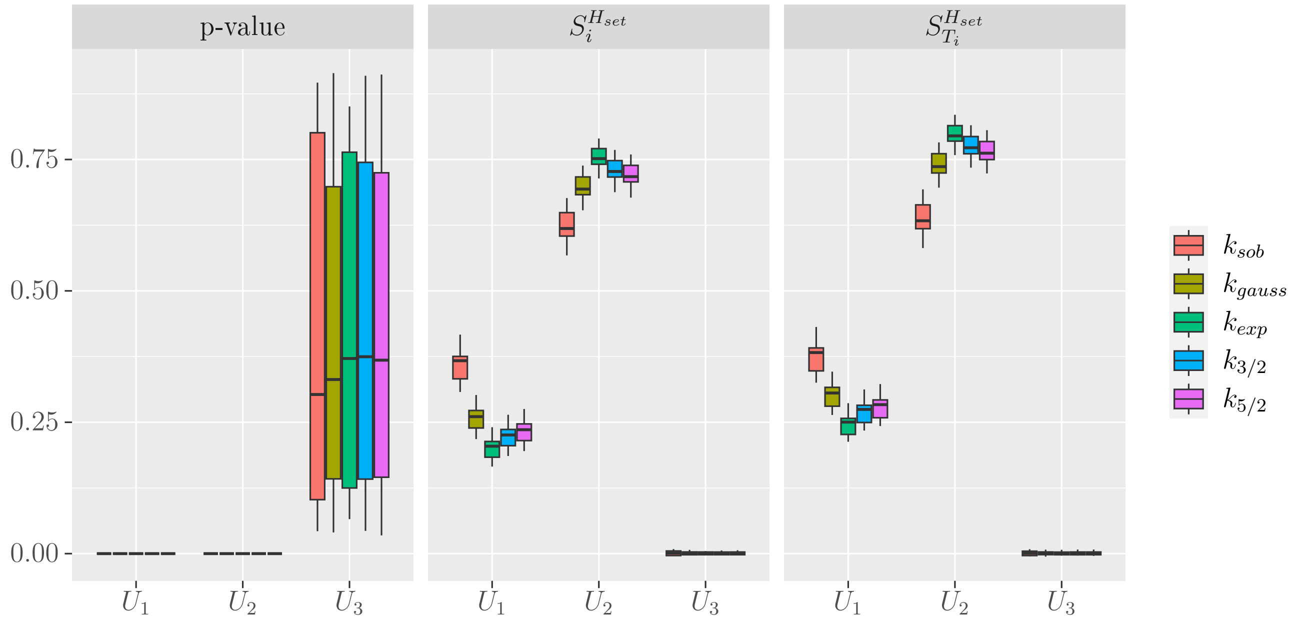

Figure 3 – Estimation of the p-values, $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$ and $\hat{\hat{S}}_{T_i}^{\mathrm{H}_{set}}$ for the excursion set defined by the constraint $g \leq 0$ computed for 5 kernels with $n = 1000$, $m = 1000$ and repeated 20 times

detected as negligible through independence testing. The first order and the total effect indices give the expected results that U_2 has a greater influence than U_1 which has a greater influence than U_3 on the excursion sets. More precisely, it shows that U_2 alone explains around 70% of $\mathrm{H}_{set}(U, \Gamma_U)$ (depending on the input kernel) when U_1 explains around 25%. The artificial input U_3 is responsible for 0% as expected. The rest correspond then to a kind of interaction between U_1 and U_2. This is confirmed when looking at the total indices which values increase a bit. However, this part of interaction must not be interpreted as in the case of Sobol indices. It is actually an open question to interpret what kind of interactions are detected by HSIC-ANOVA indices. The impact of the input kernel on the HSIC-ANOVA indices is also an open question. On these results, we can only say that the input kernel has a non-negligible impact on the values of the p-values and of the indices.

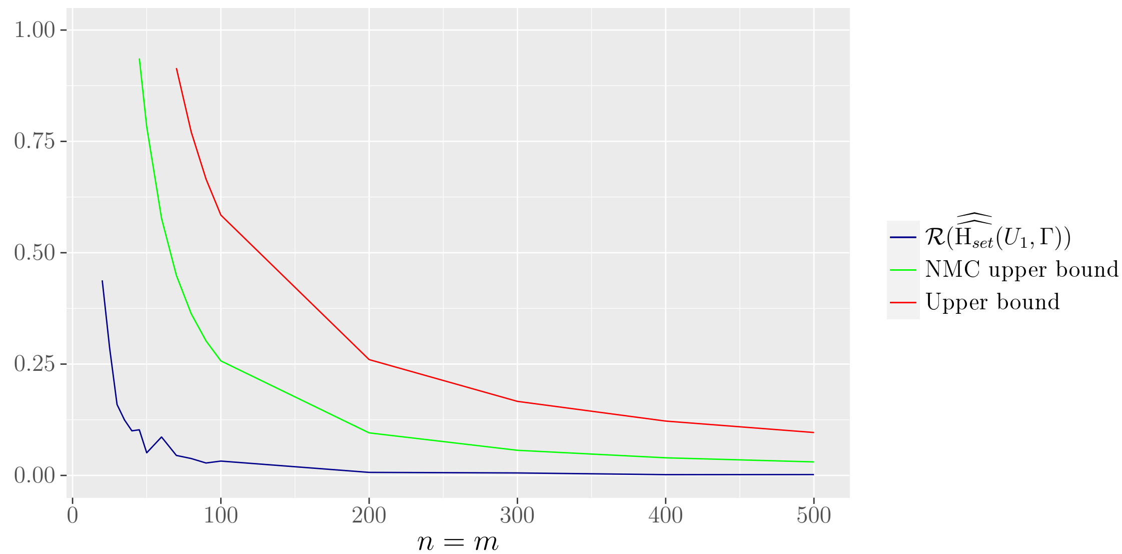

Figure 4 – Evolution of $\mathcal{R}(\widehat{\mathrm{H}_{set}}(U_1, \Gamma))$ and of the associated upper bounds for the excursion set of the constraint $g \leq 0$

However in this case, it does not change the results of the screening and of the ranking. It also seems that the Sobolev kernel behave differently than the other four kernels.

Taking $n = m = 1000$ in Figure 3 reduces the variance of the indices, but was not necessary to screen and rank the inputs. To support this remark, we compute the relative quadratic risk of the estimator $\widehat{\mathrm{H}_{set}}(U_1, \Gamma)$ defined by:

$$\mathcal{R}(\widehat{\mathrm{H}_{set}}(U_1, \Gamma)) = \mathbb{E}\left(\frac{\widehat{\mathrm{H}_{set}}(U_1, \Gamma) - \mathrm{H}_{set}(U_1, \Gamma)}{\mathrm{H}_{set}(U_1, \Gamma)}\right)^2.$$

We also compute the associated upper bound given in Proposition 3.1 and the upper bound for a classic NMC estimator (with independent $X_{i,j}^{(k)}$ for all $1 \leq i, j \leq n$, cf. (5)). The "true" value of H_{set}, used to compute the quadratic risk, and the constants σ_1, σ_2 and σ_3 are computed for $n = m = 3000$ and for the Sobolev kernel as input kernel. We plot the risk and the two bounds in Figure 4 from $n = m = 30$ to $n = m = 500$. We can observe that on this toy example, the quadratic risk of the estimator is as expected below the bound of Proposition 3.1 (in red), but also below the bound we would have by taking independent $X_{i,j}^{(k)}$ (in green). On this example, it confirms that in term of variance of the estimator, we did not loose too much by reusing the same $X^{(k)}$. It also seems that we do not necessary have to take high values of $n = m$. That is why in the next example, we will only compute the indices for $n = m = 100$.

4.2 Toy function 2

In Cousin et al. [2022], an optimization is carried out with three probabilistic constraints. We consider the two first constraints defined by the following functions g_1 and g_2:

$$g_1(x_1, x_2, u_1, u_2, u_p, u_{r_1}, u_{r_2}) = u_{r_1} - \max_{t \in [0,T]} \mathcal{Y}'(x_1 + u_1, x_2 + u_2, u_p; t),$$

$$g_2(x_1, x_2, u_1, u_2, u_p, u_{r_1}, u_{r_2}) = u_{r_2} - \max_{t \in [0,T]} \mathcal{Y}''(x_1 + u_1, x_2 + u_2, u_p; t),$$

with $\mathcal{Y}(x_1 + u_1, x_2 + u_2, u_p; t)$ the solution of the harmonic oscillator defined by:

$$(x_1 + u_1)\mathcal{Y}''(t) + u_p \mathcal{Y}'(t) + (x_2 + u_2)\mathcal{Y}(t) = \eta(t).$$

The deterministic input domain is $X = [1, 5] \times [20, 50]$. The uncertain input laws are given in Table 1. U_{r_3} is initially a random input associated to the third constraint. In our case, it will play the role of a dummy input to check if it is detected as negligible. We study the impact of the uncertain inputs on the excursion sets associated to the constraint $g_1 \leq 0$ and

Uncertainty	Distribution	Uncertainty	Distribution
U_1	$\mathcal{U}\,[-0.3, 0.3]$	U_{r_1}	$\mathcal{N}\,(1, 0.1^2)$
U_2	$\mathcal{U}\,[-1, 1]$	U_{r_2}	$\mathcal{N}\,(2.5, 0.25^2)$
U_p	$\mathcal{U}[0.5, 1.5]$	U_{r_3}	$\mathcal{N}\,(15, 3^2)$

Table 1 – Definition of the uncertain inputs

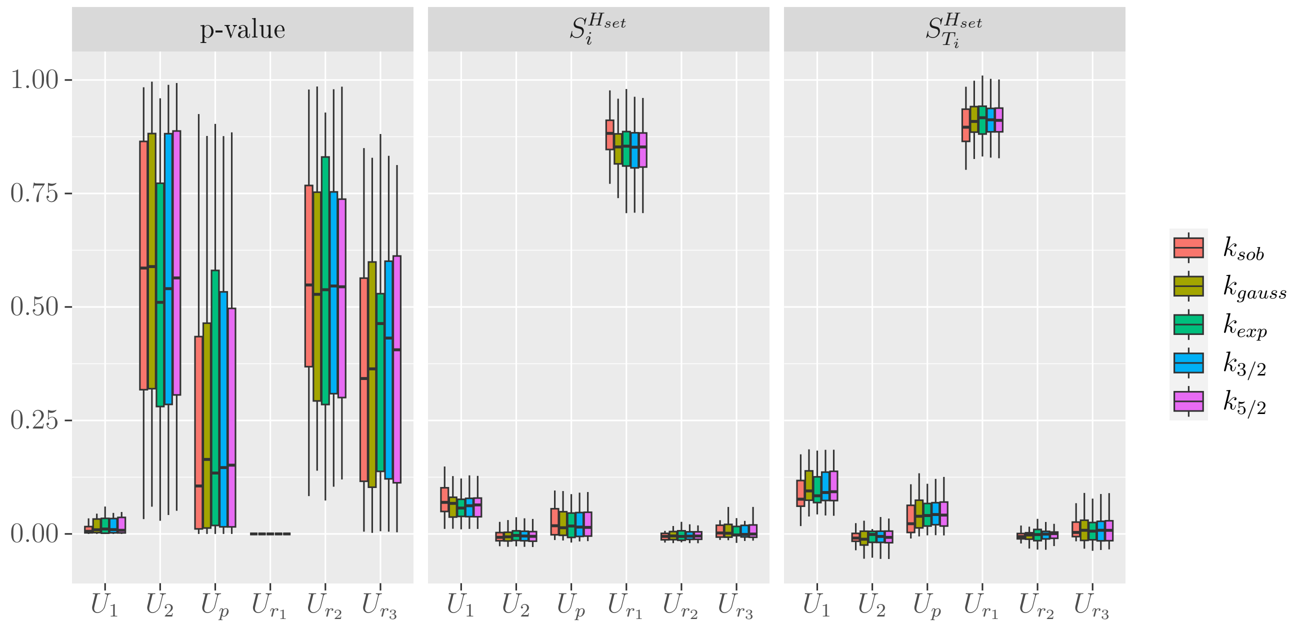

Figure 5 – First order HSIC-ANOVA indices for the first constraint $g_1 \leq 0$ computed for 5 kernels with $n = 100$, $m = 100$ and repeated 20 times

on the excursion sets associated to $g_2 \leq 0$. As kernel-based methods are suited for vectorial outputs, we also consider the case of an output being the couple of the two excursion sets, each one associated to one constraint. For each case, we compute $\hat{\hat{S}}_i^{\mathrm{H}_{set}}$, $\hat{\hat{S}}_{T_i}^{\mathrm{H}_{set}}$ and the associated p-values for each uncertain input with $n = 100$ and $m = 100$. We use again the output kernel k_{set} and $k_{set} \otimes k_{set}$ for the couple of excursion sets. The same 5 ANOVA kernels are used as in the previous example. We repeat the estimation 20 times to obtain the boxplots of Figures 5 to 7.

The analysis of the obtained boxplots makes it possible to screen the inputs and perform

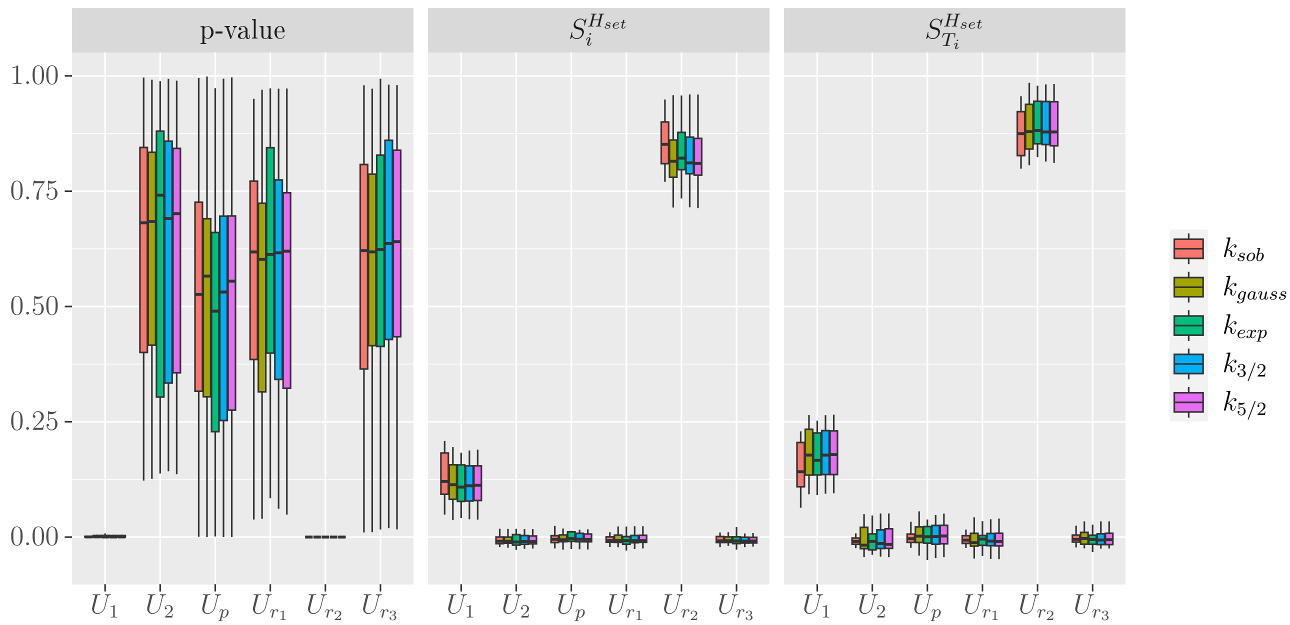

Figure 6 – First order HSIC-ANOVA indices for the constraint $g_2 \leq 0$ computed for 5 kernels with $n = 100$, $m = 100$ and repeated 20 times

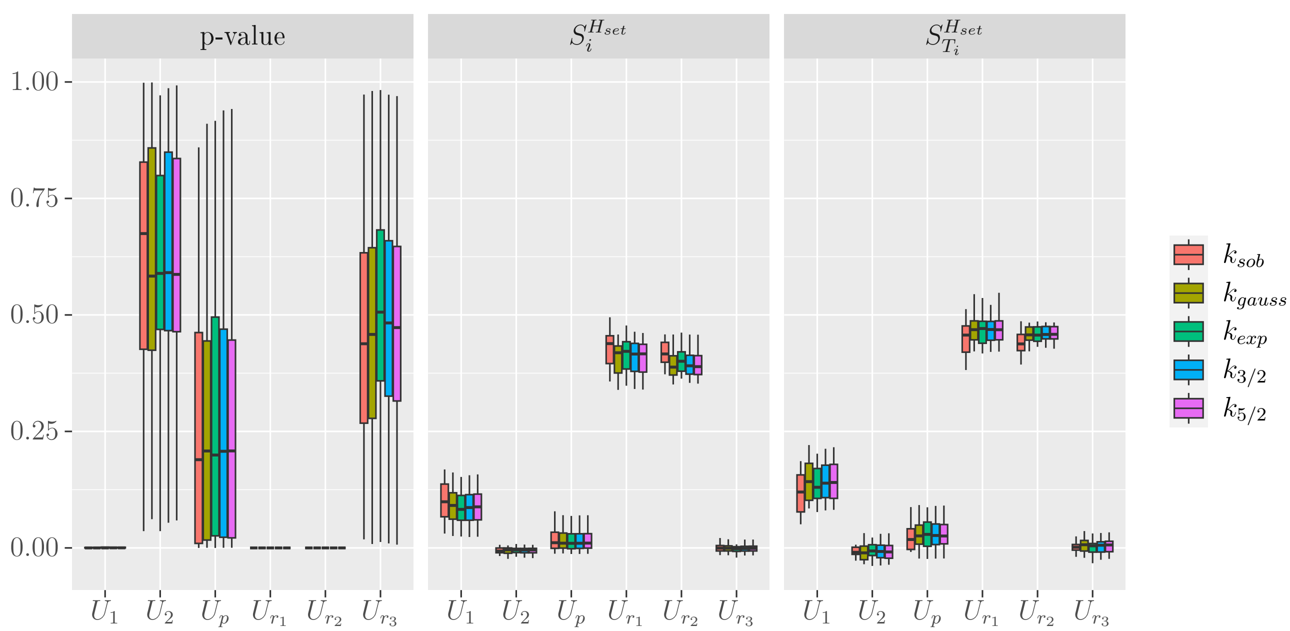

Figure 7 – 1st order HSIC-ANOVA indices for the couple of random set associated with the constraints ($g_1 \leq 0$ and $g_2 \leq 0$) computed for 5 kernels with $n = 100$, $m = 100$ and repeated 20 times

ranking. For the random set associated with the constraint $g_1 < 0$ (Figure 5), only U_1 and U_{r_1} are systematically detected as influential. U_p is also sometimes detected as influential. Between these three inputs, U_{r_1} is a lot more influential, as $\hat{\hat{S}}_{r_1}^{\mathrm{H}_{set}} \approx 80\%$. Then U_1 and U_p are responsible for approximately 10% and 3% of $H_{set}(U, \Gamma_U)$. For the second constraint $g_2 \leq 0$ (Figure 6), only U_1 and U_{r_2} are influential with their first order index respectively around 10% and 80%. When considering the couple of the two random sets (Figure 7), we get some kind of mixture between the two previous cases. U_1, U_{r_1} and U_{r_2} are detected as influential. In term of ranking, U_{r_1} and U_{r_2} have almost the same influence with they first order index around 40% and U_1 remains around 10%. It is important to note that considering the couple of random sets is totally different than considering the random set associated with the couple of constraints, i.e. the intersection of both sets. The results of the latter is given in appendix and are indeed very different from results in Figure 7.

On this test case, we observe that the choice of the input kernel has a limited impact on the indices. No matter what stage of screening or ranking is involved, the conclusions remain the same for the five kernels.

5 Conclusion

In this paper, we propose a method to conduct sensitivity analysis on set-valued outputs through kernel-based sensitivity analysis which relies on the choice of a kernel between sets. We introduce the kernel k_{set} which is based on the symmetric difference between two sets. We show that it is characteristic which is an essential property to perform screening. Then we adapt the recent HSIC-ANOVA index to set valued-outputs and we introduce an efficient estimator. Finally we compute the indices on two test cases. The proposed method allows to screen and rank the uncertain inputs according to their impact on excursion sets.

In the context of robust optimization, the presented method can be used to reduce dimension of the uncertain space by quantifying the impact of uncertain inputs on optimization constraints. Reducing the dimension of the uncertain space can then be useful to reduce the computation cost of a meta-model on the joint space.

More generally, sensitivity analysis on sets can also be used when dealing with numerical codes with set-valued outputs. This can appear in multiple fields : in viability field where the outputs are sets called viability kernels or, for instance, in flooding risk where the output is the map of flooded areas.

6 Aknowledgment

This research was conducted with the support of the consortium in Applied Mathematics CIROQUO, gathering partners in technological research and academia in the development of advanced methods for Computer Experiments.

References

Aronszajn, N. (1950). "Theory of reproducing kernels". In: *Transactions of the American Mathematical Society* 68.3, pp. 337–404. ISSN: 0002-9947, 1088-6850. DOI: 10.1090/S0002-9947-1950-0051437-7.

Balança, Paul and Erick Herbin (Jan. 2012). "A set-indexed Ornstein-Uhlenbeck process". In: *Electronic Communications in Probability* 17. DOI: 10.1214/ecp.v17-1903.

Beyer, Hans-Georg and Bernhard Sendhoff (2007). "Robust optimization – A comprehensive survey". In: *Computer Methods in Applied Mechanics and Engineering* 196.33, pp. 3190–3218. ISSN: 0045-7825. DOI: https://doi.org/10.1016/j.cma.2007.03.003.

Borgonovo, E. (2007). "A new uncertainty importance measure". In: *Reliability Engineering and System Safety* 92.6, pp. 771–784. ISSN: 0951-8320. DOI: https://doi.org/10.1016/j.ress.2006.04.015.

Borgwardt, Karsten M., Arthur Gretton, Malte J. Rasch, Hans-Peter Kriegel, Bernhard Schölkopf, and Alex Smola (2006). "Integrating structured biological data by Kernel Maximum Mean Discrepancy". In: *Bioinformatics* 22 14, e49–57.

Cousin, Alexis, Josselin Garnier, Martin Guiton, and Miguel Munoz Zuniga (Apr. 2022). "A Two-Step Procedure for Time-Dependent Reliability-Based Design Optimization Involving Piece-Wise Stationary Gaussian Processes". In: *Structural and Multidisciplinary Optimization* 65.4. Publisher: Springer Verlag (Germany), p. 120. DOI: 10.1007/s00158-022-03212-1.

Da Veiga, Sébastien (Jan. 2021). "Kernel-based ANOVA decomposition and Shapley effects - Application to global sensitivity analysis". working paper or preprint.

Da Veiga, Sébastien, Fabrice Gamboa, Bertrand Iooss, and Clémentine Prieur (2021). *Basics and Trends in Sensitivity Analysis*. Philadelphia, PA: Society for Industrial and Applied Mathematics. DOI: 10.1137/1.9781611976694. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611976694.

El Amri, Reda and Amandine Marrel (Oct. 2021). "More powerful HSIC-based independence tests, extension to space-filling designs and functional data". working paper or preprint.

El Amri, Reda, Rodolphe Le Riche, Céline Helbert, Christophette Blanchet-Scalliet, and Sébastien Da Veiga (2021). *A sampling criterion for constrained Bayesian optimization with uncertainties*. DOI: 10.48550/ARXIV.2103.05706.

Fort, Jean-Claude, Thierry Klein, and Agnès Lagnoux (2021). "Global Sensitivity Analysis and Wasserstein Spaces". In: *SIAM/ASA Journal on Uncertainty Quantification* 9.2, pp. 880–921. DOI: 10.1137/20M1354957. eprint: https://doi.org/10.1137/20M1354957.

Gamboa, Fabrice, Alexandre Janon, Thierry Klein, and Agnès Lagnoux (2013). "Sensitivity indices for multivariate outputs". In: *Comptes Rendus Mathematique* 351.7, pp. 307–310. ISSN: 1631-073X. DOI: https://doi.org/10.1016/j.crma.2013.04.016.

Gamboa, Fabrice, Thierry Klein, and Agnès Lagnoux (2017). *Sensitivity analysis based on Cramér von Mises distance*. arXiv: 1506.04133 [math.PR].

Ginsbourger, David, Olivier Roustant, Dominic Schuhmacher, Nicolas Durrande, and Nicolas Lenz (Jan. 2016). "On ANOVA Decompositions of Kernels and Gaussian Random Field Paths". In: pp. 315–330. ISBN: 978-3-319-33505-6. DOI: 10.1007/978-3-319-33507-0_15.

Gretton, Arthur, Karsten Borgwardt, Malte Rasch, Bernhard Schölkopf, and Alex Smola (2006). "A Kernel Method for the Two-Sample-Problem". In: *Advances in Neural Information Processing Systems*. Vol. 19. MIT Press.

Gretton, Arthur, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf (2005a). "Measuring Statistical Dependence with Hilbert-Schmidt Norms". In: *Algorithmic Learning Theory*. Ed. by Sanjay Jain, Hans Ulrich Simon, and Etsuji Tomita. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 63–77. ISBN: 978-3-540-31696-1.

Gretton, Arthur, Kenji Fukumizu, Choon Teo, Le Song, Bernhard Schölkopf, and Alex Smola (2007). "A Kernel Statistical Test of Independence". In: *Advances in Neural Information Processing Systems*. Vol. 20. Curran Associates, Inc.

Gretton, Arthur, Ralf Herbrich, Alexander Smola, Olivier Bousquet, and Bernhard Schölkopf (2005b). "Kernel Methods for Measuring Independence". In: *Journal of Machine Learning Research* 6.70, pp. 2075–2129.

Marczewski, Edward and Hugo Steinhaus (1958). "On a certain distance of sets and the corresponding distance of functions". In: *Colloquium Mathematicum* 6, pp. 319–327.

Marrel, Amandine and Vincent Chabridon (2021). "Statistical developments for target and conditional sensitivity analysis: application on safety studies for nuclear reactor". In: *Reliability Engineering and System Safety* 214, p. 107711.

Morris, Max D. (1991). "Factorial Sampling Plans for Preliminary Computational Experiments". In: *Technometrics* 33.2, pp. 161–174. ISSN: 00401706.

Rainforth, Tom, Robert Cornish, Hongseok Yang, Andrew Warrington, and Frank Wood (May 23, 2018). *On Nesting Monte Carlo Estimators*. arXiv: 1709.06181[stat].

Rainforth, Tom, Robert Cornish, Hongseok Yang, and Frank Wood (2016). *On the Pitfalls of Nested Monte Carlo*. arXiv: 1612.00951 [stat.CO].

Sarazin, Gabriel, Amandine Marrel, Sébastien da Veiga, and Vincent Chabridon (June 2022). "Test d'indépendance basé sur les indices HSIC-ANOVA d'ordre total". In: *53èmes Journées de Statistique de la SFdS*. Proceedings des 53èmes Journées de Statistique de la SFdS. Société Française de Statistique (SFdS) and Université Claude Bernard Lyon 1. Lyon, France.

Smola, Alex, Arthur Gretton, Le Song, and Bernhard Schölkopf (2007). "A Hilbert Space Embedding for Distributions". In: *Algorithmic Learning Theory*. Ed. by Marcus Hutter, Rocco A. Servedio, and Eiji Takimoto. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 13–31. ISBN: 978-3-540-75225-7.

Sobol, I. M (Feb. 15, 2001). "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates". In: *Mathematics and Computers in Simulation*. The Second IMACS Seminar on Monte Carlo Methods 55.1, pp. 271–280. ISSN: 0378-4754. DOI: 10.1016/S0378-4754(00)00270-6.

Song, Le, Alex Smola, Arthur Gretton, Karsten Borgwardt, and Justin Bedo (Apr. 20, 2007). *Supervised Feature Selection via Dependence Estimation.* DOI: 10.48550/arXiv.0704.2668. arXiv: 0704.2668[cs].

Spagnol, Adrien (July 2020). "Kernel-based sensitivity indices for high-dimensional optimization problems". Theses. Université de Lyon.

Ziegel, Johanna, David Ginsbourger, and Lutz Dümbgen (2022). *Characteristic kernels on Hilbert spaces, Banach spaces, and on sets of measures.* arXiv: 2206.07588 [stat.ML].

7 Annex

Proposition 3.1. *With the previous notations, we have*

$$\mathbb{E}\left(\widehat{\widehat{\mathrm{H}_{set}}}\,(U_A,\Gamma) - \mathrm{H}_{set}(U_A,\Gamma)\right)^2$$
$$\leq 2\left(\frac{2\sigma_1^2}{n(n-1)} + \frac{4(n-2)\sigma_2^2}{n(n-1)} + \frac{L^2\sigma_3^2}{m}\right),$$

with

— $\sigma_1^2 = Var\left(\left(K_A\left(U_A, U_A{}'\right) - 1\right) k_{set}(\Gamma,\Gamma')\right)$,

— $\sigma_2^2 \;= Var\left(\mathbb{E}\left[\left(K_A\left(U_A, U_A{}'\right) - 1\right) k_{set}(\Gamma,\Gamma')|(\Gamma, U_A)\right]\right)$,

— $L = \frac{\lambda(X)}{2\sigma^2}$,

— $\sigma_3^2 = Var(\left(K_A\left(U_A, U_A{}'\right) - 1\right)\mathbb{E}\left[\mathbb{1}_{\Gamma\Delta\Gamma'}(X)|(U_A, U_A{}', \Gamma, \Gamma')\right])$,

where $(U_A{}', \Gamma')$ *is an independent copy of* (U_A, Γ).

Proof. With $f(\boldsymbol{u}, z) = \left(K_A\left(u_1, u_2\right) - 1\right) e^{-\frac{\lambda(X)}{2\sigma^2}\boldsymbol{z}}$, and $\Phi(x, \gamma_1, \gamma_2) = \mathbb{1}_{\gamma_1\Delta\gamma_2}(x)$, we denote

$$\mathrm{H} = \mathrm{H}(U_A, \Gamma) = \mathbb{E}[f(U, U', \mathbb{E}[\Phi(X, \Gamma, \Gamma')|(\Gamma, \Gamma')]],$$

$$\mathrm{H}_{n,m} = \widehat{\mathrm{H}}_u\left(U_A, \Gamma\right) = \frac{2}{n(n-1)}\sum_{i<j}^{n} f(U_A^{(i)}, U_A^{(j)}, \frac{1}{m}\sum_{k=1}^{m}\Phi(X^{(k)}, \Gamma^{(i)}, \Gamma^{(j)}),$$

and

$$\mathrm{H}_n = \frac{2}{n(n-1)}\sum_{i<j}^{n} f(U_A^{(i)}, U_A^{(j)}, \mathbb{E}[\Phi(X, \Gamma^{(i)}, \Gamma^{(j)})|(\Gamma^{(i)}, \Gamma^{(j)})]).$$

First we split the risk into two terms:

$$\mathbb{E}|\,\mathrm{H}_{n,m} - \mathrm{H}\,|^2 \leq 2\mathbb{E}|\,\mathrm{H}_n - \mathrm{H}\,|^2 + 2\mathbb{E}|\,\mathrm{H}_n - \mathrm{H}_{n,m}\,|^2$$

The first term is the variance of a classic U-statistic of order 2:

$$\mathbb{E}|\,\mathrm{H}_n - \mathrm{H}\,|^2 = \frac{2\sigma_1^2}{n(n-1)} + \frac{4(n-2)\sigma_2^2}{n(n-1)}.$$

The second term can be developed:

$$\mathbb{E}|\,\mathrm{H}_n - \mathrm{H}_{nm}\,|^2 = \frac{4}{n^2(n-1)^2}\left(\sum_{i<j}^{n}\sum_{p<l}^{n}\mathbb{E}\left(E_{ij} - E_{ij,m}\right)\left(E_{pl} - E_{pl,m}\right)\right),$$

with

$$E_{ij} = f(U_A^{(i)}, U_A^{(j)}, \mathbb{E}[\Phi(X, \Gamma^{(i)}, \Gamma^{(j)})|(\Gamma^{(i)}, \Gamma^{(j)})])$$

and

$$E_{ij,m} = f(U_A^{(i)}, U_A^{(j)}, \frac{1}{m}\sum_{k=1}^{m}\Phi(X^{(k)}, \Gamma^{(i)}, \Gamma^{(j)}).$$

As $X^{(1)}, ..., X^{(m)}$ are common to each $E_{ij,m}$, the terms $E_{ij,m}$ and $E_{pl,m}$ are not independent even if i, j, p, l are pairwise distinct. We can still bound them but we will lose one order of convergence in m. We first have

$$
\begin{aligned}
\left|\mathbb{E}\left(E_{ij}-E_{ij,m}\right)\left(E_{pl}-E_{pl,m}\right)\right| &\leq \left(\mathbb{E}\left|E_{ij}-E_{ij,m}\right|^{2}\mathbb{E}\left|E_{pl}-E_{pl,m}\right|^{2}\right)^{\frac{1}{2}} \\
&= \left(\mathbb{E}\left|E_{12}-E_{12,m}\right|^{2}\right) \\
&\leq \frac{L^{2}}{m}\operatorname{Var}((K_{A}\left(U_{A},{U_{A}}'\right)-1)\,\mathbb{E}(\Phi(X,\Gamma,\Gamma')|(\Gamma,\Gamma'))),
\end{aligned}
$$

using that $z \rightarrow f(u,z)$ is L-lipschitz for all u. Then we have

$$
\mathbb{E}|\operatorname{H}_{n}-\operatorname{H}_{nm}|^{2} \leq \frac{L^{2}}{m}\operatorname{Var}((K_{A}\left(U_{A},{U_{A}}'\right)-1)\,\mathbb{E}(\Phi(X,\Gamma,\Gamma')|(\Gamma,\Gamma'))).
$$

Putting all results together, we get

$$
\mathbb{E}|\operatorname{H}_{n,m}-\operatorname{H}|^{2} \leq 2\left(\frac{2\sigma_{1}^{2}}{n(n-1)}+\frac{4(n-2)\sigma_{2}^{2}}{n(n-1)}+\frac{L^{2}\sigma_{3}^{2}}{m}\right).
$$

If the $X^{(k)}$ were drawn independently for each (i,j), we could obtain an asymptotic rate of $\mathcal{O}(\frac{1}{n}+\frac{1}{m^{2}})$ by using the independence in the upper-bound of $\left|\mathbb{E}\left(E_{ij}-E_{ij,m}\right)\left(E_{pl}-E_{pl,m}\right)\right|$ as done in Rainforth et al. [2018]. $\square$

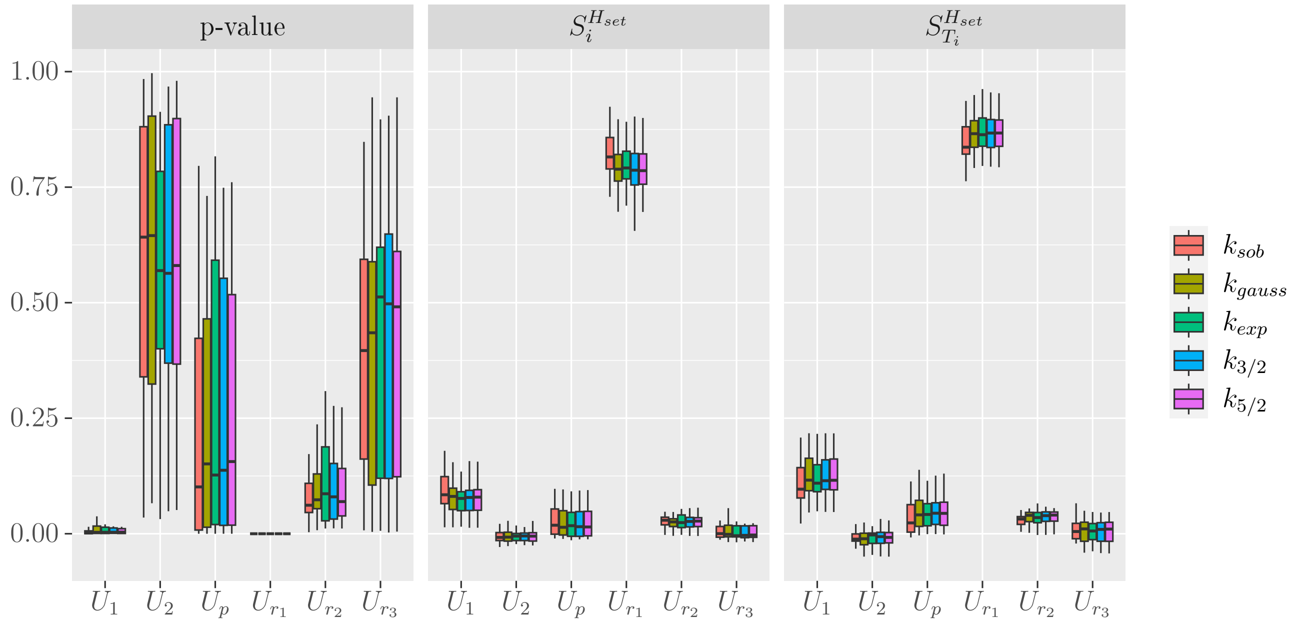

Figure 8 – First order HSIC-ANOVA indices for the constraint $g_1 \leq 0$ and $g_2 \leq 0$ computed for 5 kernels with $n = 100$, $m = 100$ and repeated 20 times