New developments around dependence measures for sensitivity analysis : application to severe accident studies for generation IV reactors

Anouar Meynaoui

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ FÉDÉRALE TOULOUSE MIDI-PYRÉNÉES

Délivré par:
Institut National des Sciences Appliquées de Toulouse

Discipline ou spécialité:
Domaine mathématiques – Mathématiques appliquées

Présentée et soutenue par
ANOUAR MEYNAOUI
le: 22 novembre 2019

Titre:
Nouveaux développements autour des mesures de dépendance pour l'analyse de sensibilité : application aux études d'accidents graves pour les réacteurs de génération IV

École doctorale:
Mathématiques Informatique Télécommunications de Toulouse (MITT)

Unité de recherche:
UMR 5219

Directrice de thèse:
Mme BEATRICE LAURENT-BONNEAU (Professeur, INSA Toulouse)

Encadrante CEA:
Mme AMANDINE MARREL (Ingénieur de recherche, CEA)

Rapporteurs:
Mme CRISTINA BUTUCEA (Professeur, ENSAE)
M. ARTHUR GRETTON (Professeur, University College London)

Président du jury:
M. FABRICE GAMBOA (Professeur, Université Paul Sabatier)

Examinateurs:
M. SEBASTIEN DA VEIGA (Ingénieur de recherche, Safran Tech)
M. GUILLAUME PERRIN (Ingénieur de recherche, CEA)

"No man ever steps in the same river twice, it's not the same river and he's not the same man"

$\sim$ Heraclitus of Ephesus

Remerciements

J'adresse mes plus profonds remerciements tout d'abord à deux personnes que je ne peux pas dissocier et à qui je dois énormément, ma directrice de thèse Béatrice et mon encadrante Amandine. Depuis que ma thèse a commencé, vous n'avez jamais hésité à me prêter main forte dans les moments où j'en avais besoin (et il y en avait tellement ...). Merci pour votre soutien, solidarité et grande disponibilité dans les moments difficiles. Je suis reconnaissant pour tous les efforts et le temps conséquents que vous m'avez accordés. J'ai tellement appris à vos côtés tant sur le plan scientifique qu'humain. Merci d'avoir rendu cette expérience si enrichissante et de m'avoir permis d'acquérir l'autonomie nécessaire pour avancer plus sereinement dans l'univers de la recherche. Enfin, c'était un véritable plaisir de travailler avec vous !

Je tiens également à remercier mes deux rapporteurs Cristina Butucea et Arthur Gretton pour avoir accepté d'évaluer mon travail. Merci pour vos rapports détaillés et bienveillants et pour vos remarques et suggestions pertinentes. C'est pour moi un grand honneur que vous soyez les rapporteurs de ma thèse. J'adresse également ma profonde gratitude à l'ensemble des membres du jury : Fabrice Gamboa, Sébastien Da Veiga et Guillaume Perrin, d'avoir accepté d'examiner mon travail.

Toujours sur le plan scientifique, un grand merci à toi Mélisande d'avoir accepté de travailler avec nous. J'ai beaucoup appris de ta rigueur scientifique et rédactionnelle. De plus, ta sympathie et ton sens de l'humour ont toujours rendu le travail avec toi très agréable. J'espère que nous aurons l'occasion de collaborer à nouveau dans l'avenir ! Un grand merci à toi aussi Jean-Baptiste pour ton aide conséquente sur le code MACARENa qui constitue un des fils conducteurs de ce manuscrit. Mes remerciements s'adressent aussi à Hugo Raguet que j'ai pu côtoyer pendant un an au SESI. Merci pour ta grande aide durant la première phase de ma thèse. J'ai beaucoup appris de ta façon d'aborder les problèmes scientifiques.

Mes remerciements s'adressent ensuite à tous les gens du SESI que j'ai pu rencontrer durant mon séjour à Cadarache. Merci tout spécialement à toi Michel d'avoir partagé le bureau avec moi pendant ces trois années. Nos discussions quotidiennes étaient pour moi une vraie drogue tout comme les tasses de café que j'ingurgitais à longueur de journée. Merci à toi Manuel, pour ton soutien, tes encouragements et ta disponibilité tout au long de ma thèse. Merci à toute l'équipe du "midi": Faouzi, Avent, Océane, Florence, Loïc Gautier et Loïc Augier. J'ai toujours apprécié nos grands débats philosophiques sur l'éthique, la conscience, l'épistémologie, la métaphysique et j'en passe. Nos discussions, nos balades rallongées ou encore nos petits pique-niques estivaux étaient un vrai répit pour moi dans la dernière phase de la thèse. Bref ! Merci à toutes les personnes du SESI pour votre accueil et bienveillance durant ces années.

Et comment oublier mes copains de l'IMT que je rencontrais lors de mes venues à Toulouse ou lors des conférences auxquelles j'ai pu participer. C'était toujours un régal de vous retrouver. Je pense en particulier à mes amis Albigeois et de longue date David et Florian. Merci pour toutes

ces expériences, ces moments agréables, ces discussions (et aussi ces parties d'échecs intenses ...). Merci aussi à tous les doctorants que je croisais en conférences, séminaires ou formations et avec qui on passait de très bons moments. Merci à toi Baptiste, Camille, Thrang et Eva. Merci à tous les autres doctorants que je n'ai pas pu citer et qui ont fait partie de cette belle expérience.

Pour finir avec les amis, je remercie mon ami d'enfance Jalil, qui m'a toujours encouragé et soutenu. Nos discussions à propos de tous ces sujets divers et variés sont un vrai plaisir, j'ai toujours apprécié ta vision du monde et ton optimisme. Je tiens aussi à remercier mon ami Taoufik d'avoir toujours su garder le contact et prendre de mes nouvelles malgré les distances.

Pour conclure ces remerciements, je m'adresse à présent à mes proches. À mes chers parents. Merci pour votre soutien indéfectible depuis ma plus tendre enfance. Je n'aurais jamais réussi sans vous. Je ne vous remercierai jamais assez pour ce que vous avez été pour moi et continuez à être. Aussi, c'est vous qui m'avez initié aux mathématiques, et vous m'y avez donné goût. Enfin, à ma petite sœur bien-aimée Nisrine. Merci à toi de m'avoir soutenu et encouragé depuis longtemps et d'avoir toujours été là dans les grands moments de ma vie. C'est un privilège d'avoir une sœur comme toi. Merci infiniment !

Contents

Chapter 1

Introduction (français)

1.1 Contexte

Dans le cadre des études de sûreté pour les réacteurs nucléaires, les codes de calcul (ou simulateurs numériques) sont essentiels pour comprendre, modéliser et prévoir des phénomènes physiques. Ces outils prennent en entrée un grand nombre de paramètres caractéristiques du phénomène étudié ou liés à sa modélisation physique et numérique. Les informations relatives à certains de ces paramètres sont souvent limitées ou incertaines, cela peut être dû à une absence ou manque de données, des erreurs de mesure ou de modélisation ou encore à une variabilité naturelle des paramètres. **Ces paramètres d'entrée, et par conséquent la sortie du simulateur sont donc entachés d'une certaine incertitude.** On parle alors de propagation des incertitudes. Il est donc important de considérer non seulement les valeurs nominales des entrées, mais aussi l'ensemble des valeurs possibles dans leurs domaines de variation. Une prise en compte des incertitudes des entrées ainsi que leurs effets sur l'incertitude de la sortie est donc une étape importante pour les études de sûreté.

La démarche générale de traitement des incertitudes dans les codes de calcul a fait l'objet de nombreux travaux dans les dernières décennies. Dans la littérature générale consacrée au sujet (De Rocquigny et al., 2008; Ghanem et al., 2017), la démarche méthodologique usuelle se décompose en quatre étapes clés. Cette démarche générique est illustrée par la Figure 2.1. **La première étape, étape A, est la spécification du problème,** qui consiste à définir le système à étudier (modèle, simulateur ou encore processus de mesure), identifier les variables d'entrée incertaines ou fixées, ainsi que les quantités d'intérêt à étudier (qui dérivent des variables de sortie du modèle). **L'étape B consiste ensuite à quantifier les incertitudes des variables d'entrée.** Dans le cadre probabiliste, les incertitudes des variables d'entrée aléatoires sont modélisées par des distributions de probabilités totalement ou partiellement connues (Helton, 1997; Oberkampf et al., 2001). Le choix de ces modèles probabilistes dépend des éventuelles données disponibles, des avis des experts ou encore de données bibliographiques. Récemment Bae et al. (2004) et Swiler et al. (2009) ont proposé d'autres méthodes de quantification pour les incertitudes épistémiques, *i.e.* liées au manque de connaissance plus qu'au caractère aléatoire du phénomène. Ces méthodes reposent entre autres sur la théorie de *l'évidence* aussi appelée théorie de *Dempster-Shafer* (Dempster, 1967; Shafer, 1976). **Dans l'étape C, les incertitudes sont propagées :** l'objectif est de quantifier comment les incertitudes en entrée se répercutent sur la ou les sorties prédites par le modèle, et plus précisément sur la quantité d'intérêt. Cette quantité d'intérêt qui dérive des sorties du modèle est étroitement liée aux objectifs de l'étude. Il peut s'agir de la moyenne ou de la dispersion de la sortie, d'une probabilité de dépassement

d'une valeur critique ou encore d'un quantile. Différentes approches spécifiques, déterministes ou reposant sur la simulation Monte-Carlo, ont été développées en fonction de la quantité d'intérêt considérée (Cannamela, 2007). En parallèle de la propagation des incertitudes, **une analyse de sensibilité, étape C' de la démarche,** peut aussi être réalisée. L'**analyse de sensibilité vise à déterminer comment la variabilité des paramètres en entrée influe sur la valeur de la sortie ou de la quantité d'intérêt** (Saltelli et al., 2004; Iooss, 2011)**.** Elle permet ainsi d'identifier et éventuellement quantifier, pour chaque paramètre d'entrée ou groupe de paramètres, sa contribution à la variabilité de la sortie. L'analyse de sensibilité peut avoir différents objectifs : hiérarchisation des paramètres en entrées par ordre d'influence sur la variabilité de la sortie, ou encore séparation des entrées en deux groupes, celles jugées significativement influentes sur l'incertitude de la sortie et celles dont l'influence peut être négligée. Cette séparation des variables d'entrée en deux groupes est appelée criblage (ou *screening*). Les résultats de l'analyse de sensibilité apportent des informations précieuses sur l'influence des entrées incertaines, la compréhension du modèle et du phénomène physique sous-jacent. Ils peuvent aussi être utilisés pour diverses fins : réduction des incertitudes en ciblant les efforts de caractérisation sur les entrées les plus influentes, simplification du modèle en fixant les entrées non-influentes à des valeurs de référence ou encore validation du modèle vis-à-vis du phénomène modélisé. Ces enjeux expliquent les nombreux travaux récents autour d'outils et méthodes statistiques pour l'analyse de sensibilité. L'une des méthodes les plus classiquement utilisées dans les applications industrielles repose sur une décomposition de la variance de la sortie (Hoeffding, 1992; Sobol, 1993), où chaque terme de la décomposition représente la part de la contribution d'une entrée ou d'un groupe d'entrées à la variance de la sortie. Cette approche conduit à l'obtention des indices de Sobol'. Ces indices facilement interprétables présentent en pratique plusieurs inconvénients (estimation coûteuse en nombre de simulations, information partielle apportée par la variance). Pour pallier ces limitations, d'autres approches basées sur des **mesures de dépendance** ont récemment été proposées (Da Veiga, 2015). Ces mesures présentent de nombreux avantages, qui seront exposés dans ce qui suit, et ont donné des résultats prometteurs sur plusieurs applications industrielles (De Lozzo and Marrel, 2016b).

Dans le cadre de l'analyse de sensibilité des simulateurs numériques, les travaux réalisés dans cette thèse ont pour objectif de proposer des nouvelles méthodes statistiques innovantes basées sur les mesures de dépendance, permettant de répondre efficacement aux problématiques posées par leur mise en œuvre sur des applications industrielles.

1.2 Analyse de sensibilité globale basée sur les mesures de dépendance

Comme mentionné précédemment, les méthodes de l'Analyse de Sensibilité (AS) visent à déterminer la façon dont la variabilité des entrées d'un modèle influe sur la variabilité de sa sortie. On distingue deux grands domaines : l'**Analyse de Sensibilité Locale** (ASL) et l'**Analyse de Sensibilité Globale** (ASG).

L'analyse de sensibilité locale étudie la variation de la sortie pour des petites variations des entrées autour de leurs valeurs de référence (aussi appelées valeurs nominales). Parmi les méthodes pour l'ASL, les principales sont celles basées sur les **dérivées partielles** (Alam et al., 2004; Pujol, 2009) et celles basées sur la **modélisation adjointe** (Hall et al., 1982; Cacuci, 1981, 2003). La première consiste à estimer les dérivées partielles du modèle numérique par rapport à chaque entrée, au point nominal. Ces dérivées partielles représentent l'effet de la perturbation

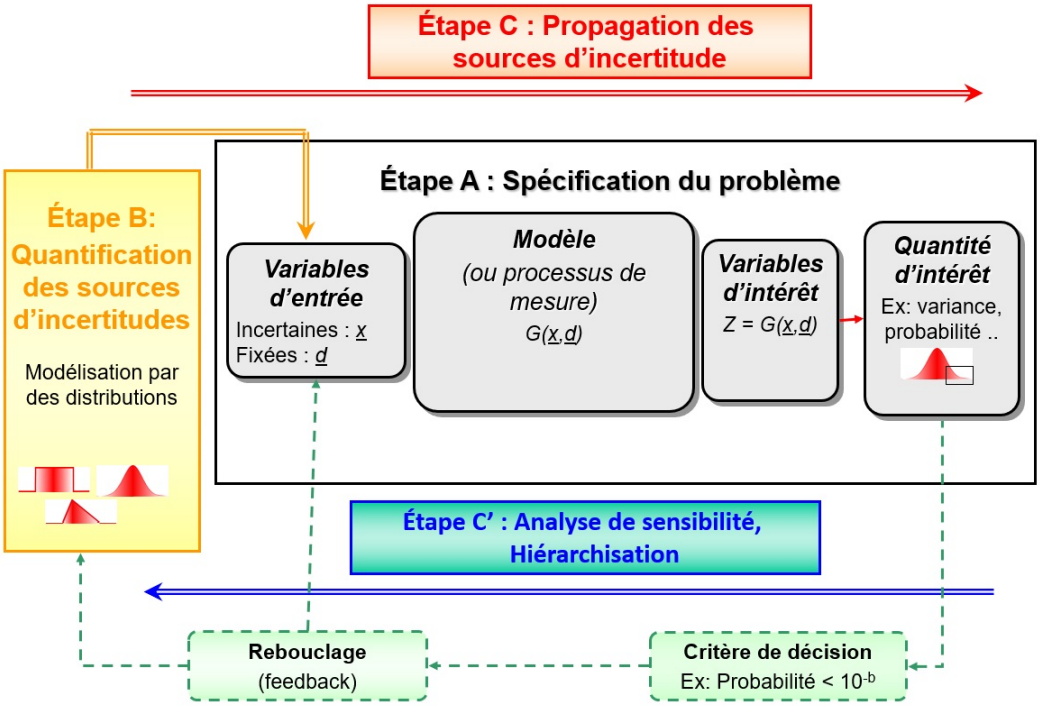

Figure 1.1 – Schéma général de la méthodologie de traitement des incertitudes issu de De Rocquigny et al. (2008).

de chaque entrée sur la perturbation totale de la sortie et sont directement interprétées comme des indices de sensibilité locale relatifs à chaque entrée. L'estimation de ces indices peut être réalisée grâce à des techniques de planification d'expériences de type *One-At-a-Time* (OAT) qui consistent à ne perturber qu'une entrée à la fois en fixant les autres entrées à leurs valeurs nominales (Morris, 1991). L'approche basée sur la modélisation adjointe est quant à elle une méthode purement analytique qui peut être utilisée lorsqu'une formule analytique du modèle est explicitement connue. La modélisation adjointe est intrusive d'un point de vue numérique, ce qui signifie que son application nécessite le développement d'un modèle de calcul des dérivées partielles suivant chaque direction. Cette méthode n'est donc pas utilisable dans le cas de simulateurs de type "boîte noire" où seules les entrées et sorties du modèle sont accessibles.

Ces méthodes d'ASL ne prennent donc pas en compte les incertitudes des variables d'entrée dans l'ensemble de leur domaine de variation. Pour évaluer et quantifier l'impact de l'incertitude globale de chaque entrée sur la sortie, des méthodes statistiques d'**Analyse de Sensibilité Globale (ASG)** ont été développées. Contrairement à l'ASL, l'approche globale nécessite de caractériser l'incertitude des entrées sur leur domaine de variation (étape B, Figure 2.1), via par exemple l'attribution d'une loi de probabilité au vecteur des entrées. Les méthodes statistiques pour l'ASG reposent ensuite le plus souvent sur des simulations de type Monte-Carlo du modèle, *i.e.* sur un échantillonnage aléatoire des variables d'entrées selon leurs distributions de probabilités. Parmi les méthodes usuelles pour l'ASG, figurent les **mesures d'ASG basées sur les dérivées** qui conduisent à l'obtention des indices DGSM, pour *Derivative-based Global Sensitivity Measures* (Kucherenko et al., 2009; Kucherenko and Iooss, 2017; Sobol and Kucherenko, 2010).

La construction de ces indices est basée sur une généralisation des mesures de sensibilité locale en moyennant les dérivées partielles par rapport à chaque entrée sur son domaine de variation Cependant, l'estimation de ces indices nécessite un grand nombre d'appels au code, ce qui limite considérablement son utilisation dans le cas de modèle coûteux[1]. Pour pallier cet inconvénient, des stratégies d'estimation basées sur l'utilisation de métamodèles pour approximer la sortie du modèle ont été proposées. On peut citer les travaux de Sudret and Mai (2015) basés sur des polynômes du chaos ou encore ceux de De Lozzo and Marrel (2016a) utilisant des métamodèles processus Gaussiens. Une autre approche classiquement utilisée pour l'ASG repose sur la décomposition de la variance de la sortie, où chaque terme de la décomposition représente la part de la contribution d'une entrée ou d'un groupe d'entrées à la variance de la sortie. Initialement introduite dans Hoeffding (1948a), cette décomposition est communément appelée : décomposition ANOVA (pour *ANalysis Of VAriance*). Des indices de sensibilités sont directement issus de cette décomposition : il s'agit des indices de Sobol' (Sobol, 1993), mentionnés précédemment. Les **indices de Sobol'** ont l'avantage d'être facilement interprétables mais leurs expressions font intervenir des intégrales multidimensionnelles dont l'estimation par des méthodes de type Monte-Carlo nécessitent en pratique un très grand nombre de simulations du modèle (plusieurs dizaines de milliers). Leur estimation directe est donc souvent impossible dans le cas de simulateurs coûteux en temps de calcul. Plusieurs travaux ont été développés pour réduire les efforts d'estimation de ces indices. D'autres approches supposant des régularités supplémentaires du modèle et basées sur des méthodes de décomposition spectrale ont aussi été proposées. On peut citer par exemple la méthode FAST (FAST pour *Fourier Amplitude Sensitivity Testing*) introduite dans Cukier et al. (1973) puis étudiée dans Lemaître (2014) et Iooss and Lemaître (2015). Des méthodes comme E-FAST (*Extended Fourier Amplitude Sensitivity Testing*) et RBD-FAST (*Random Balance Design Fourier Amplitude Sensitivity Testing*) respectivement introduites dans Saltelli et al. (1999) et Tarantola et al. (2006) proposent des améliorations de la méthode FAST classique. Le nombre d'appels au modèle avec ces méthodes demeure néanmoins important. Là aussi, une alternative possible consiste à estimer ces indices via des métamodèles : l'estimation des indices de Sobol' par des polynômes de chaos, des polynômes locaux ou encore des processus Gaussiens ont été respectivement proposés dans Sudret (2008), Da Veiga et al. (2009) et Marrel et al. (2009). Ces approches nécessitent cependant d'arriver à construire un métamodèle suffisamment prédictif, ce qui peut s'avérer compliqué dans le cas de simulateurs fortement non-linéaires et/ou dans le cas d'un grand nombre de variables d'entrées. Par ailleurs, indépendamment des difficultés liées à leur estimation, les indices de Sobol' ne considèrent que la variance de la sortie et n'évaluent donc pas l'influence de chaque entrée sur l'ensemble de la loi de probabilité de la sortie. Ils ne sont donc pas équivalents à l'indépendance entre la sortie et chacune des entrées (exception faite des indices de Sobol' totaux).

Les **mesures de dépendance,** récemment introduites pour l'ASG par Da Veiga (2015), permettent de pallier plusieurs des limitations listées précédemment. Tout d'abord, ces mesures quantifient d'un point de vue probabiliste la dépendance entre chaque entrée et la sortie. Ainsi la nullité d'une mesure de dépendance entre une entrée et la sortie est équivalente à l'indépendance de ces deux variables aléatoires. Ces mesures peuvent être utilisées d'un point de vue quantitatif pour hiérarchiser les entrées par ordre d'influence sur la sortie, aussi bien que d'un point vue qualitatif, pour effectuer un criblage des entrées, *via* des tests statistiques par exemple (De Lozzo and Marrel, 2016b). L'utilisation de tests statistiques pour identifier les variables non influentes offre un cadre statistique et mathématique plus rigoureux et précis qu'une simple appréciation et comparaison des mesures de sensibilité. Cela évite en particulier le choix arbitraire d'une

[1] Coûteux fait ici référence à la durée nécessaire pour chaque simulation du modèle ou code de calcul, qui limite le nombre total de simulations possibles.

valeur seuil pour les mesures de sensibilité, au-delà de laquelle une variable d'entrée est considérée comme influente. Parmi les mesures de dépendance existantes dans la littérature, on peut citer tout d'abord les **mesures de dissimilarité** introduites par Baucells and Borgonovo (2013). L'idée de construction de ces mesures est basée sur la comparaison de la distribution de probabilité de la sortie avec sa distribution lorsqu'une entrée donnée est fixée. Ces mesures appartiennent en fait à une plus large classe basée sur la f-divergence de Csiszàr (Csiszár, 1972). Cette dernière regroupe plusieurs notions de dépendance plus anciennes comme la distance d'Hellinger (Hellinger, 1909), la divergence de Kullback-Leibler (Kullback and Leibler, 1951) ou encore la distance de variation totale (Rudin et al., 1992). D'ailleurs, Da Veiga (2015) souligne aussi les liens entre la f-divergence de Csiszàr et l'information mutuelle introduite par Shannon (1948) ainsi qu'avec l'information mutuelle du carré de perte (Suzuki et al., 2009), ces mesures peuvant être interprétées comme des mesures de dissimilarité. À noter que les indices de Sobol' peuvent aussi être définis comme des mesures de dissimilarité (Chabridon, 2018). En dépit de leurs propriétés théoriques intéressantes, l'estimation des mesures basées sur la f-divergence de Csiszàr s'avère en pratique coûteuse en nombre de simulations, en particulier en grande dimension[2].

D'autres mesures de dépendance dont l'estimation souffre moins du "fléau de la dimension" ont aussi été proposées par Da Veiga (2015). Parmi elles, figure la **covariance de distance** basée sur l'utilisation des fonctions caractéristiques (Székely et al., 2007). Il a été démontré que cette mesure de dépendance possède des bonnes propriétés pour tester l'indépendance en grande dimension entre deux variables aléatoires (Székely and Rizzo, 2013; Yao et al., 2018). Il a aussi été mis en lumière que la covariance de distance fait partie d'une classe plus large de mesures de dépendance (Székely and Rizzo, 2013), basées sur des objets mathématiques appelés *noyaux caractéristiques* (Sriperumbudur et al., 2010). Ces mesures de dépendance s'avèrent en pratiques très efficaces pour tester l'indépendance entre des variables aléatoires de différente nature : variables scalaires, vectorielles, catégorielles, etc. Parmi elles, le critère d'indépendance de Hilbert–Schmidt noté **HSIC pour *Hilbert Schmidt Independence Criterion*** (Gretton et al., 2005a), généralise la notion de covariance entre deux variables aléatoires et permet ainsi de capturer un très large spectre de formes de dépendance entre les variables. Pour cette raison, Da Veiga (2015), puis De Lozzo and Marrel (2016b) se sont intéressés à l'utilisation des mesures HSIC pour l'ASG et les ont comparées aux indices de Sobol'. À noter que la mesure HSIC coïncide avec la covariance de distance pour un choix particulier de noyaux (Székely and Rizzo, 2013). Comme illustré par De Lozzo and Marrel (2016b), les indices HSIC présentent aussi l'avantage d'avoir un faible coût d'estimation (en pratique quelques centaines de simulations contre plusieurs dizaines de milliers pour les indices de Sobol') et leur estimation pour l'ensemble des entrées ne dépend pas du nombre d'entrées. De plus, des tests statistiques d'indépendance basés sur les mesures HSIC ont aussi été développés par Gretton et al. (2008), dans un cadre asymptotique. Plus récemment, une première extension à un cadre non-asymptotique a été proposée par De Lozzo and Marrel (2016b), qui ont aussi montré l'efficacité et le grand intérêt des tests statistiques basés sur les HSIC pour réaliser un criblage des variables d'entrée.

Pour toutes ces raisons, on s'intéresse dans le cadre de cette thèse aux mesures de dépendance de type HSIC pour l'ASG des simulateurs numériques. Plus précisément, l'objectif est de proposer des nouveaux développements théoriques, méthodologiques et applicatifs autour sur ces mesures.

[2]La grande dimension désigne ici un grand nombre de variables d'entrée.

1.3 Description du cas test applicatif

Cette thèse s'inscrit dans le cadre de la démonstration de sûreté et de maîtrise des risques des Réacteurs à Neutrons Rapides refroidis au sodium (RNR-Na, Figure 1.2) de Génération IV, menée par le CEA et ses partenaires. Comme leurs noms l'indiquent, les RNR utilisent l'énergie cinétique élevée des neutrons pour fusionner les noyaux d'uranium, par opposition aux réacteurs à neutrons thermiques (Réacteurs à Eau Pressurisée par exemple) où les neutrons sont ralentis pour augmenter la probabilité d'interagir avec les atomes d'uranium. Dans le cadre des études de sûreté, plusieurs scénarios d'accidents graves du réacteur sont étudiés à travers des essais expérimentaux et des simulations numériques. Les accidents graves sont ceux qui conduisent à une fusion partielle ou totale du cœur du réacteur. L'évolution dans le temps de diverses grandeurs physiques liées à l'accident (aussi appelés transitoires accidentels) permettent aux physiciens de mieux comprendre les phénomènes physiques mis en jeu et d'évaluer le comportement du cœur.

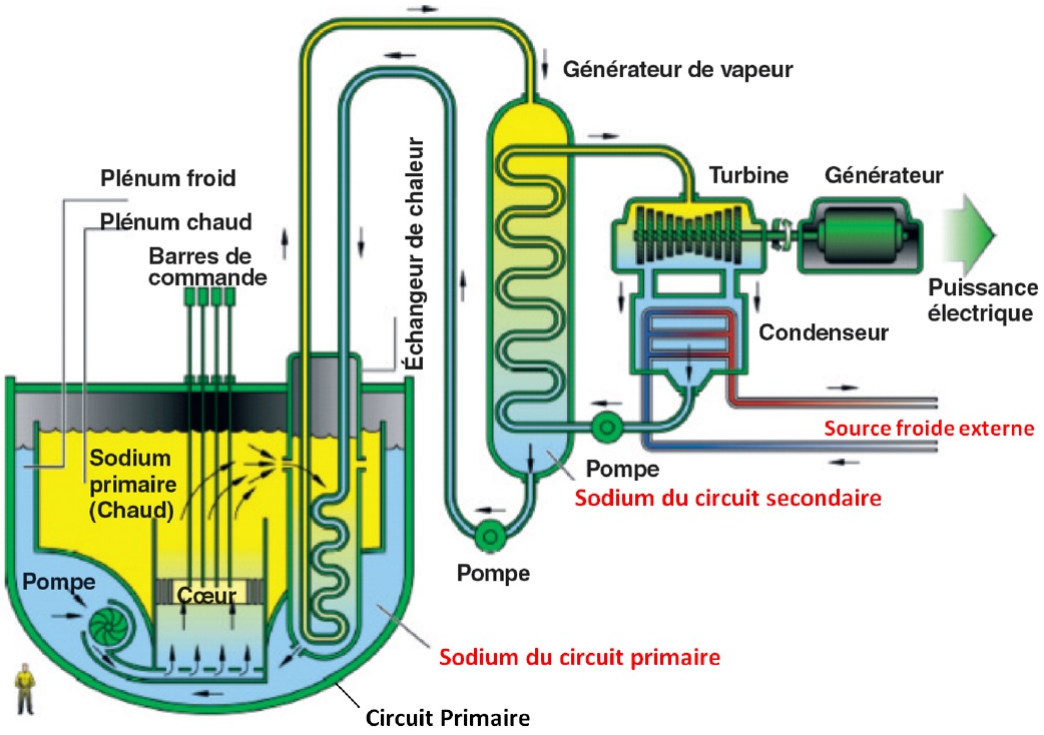

Figure 1.2 – Schéma général de fonctionnement d'un réacteur RNR-Na, extrait de Droin (2016).

1.3.1 Présentation du réacteur RNR-Na et de l'accident ULOF

Comme le montre la Figure 1.2, le fonctionnement général d'un réacteur nucléaire RNR-Na est basé sur des échanges thermiques produisant de l'énergie électrique. La chaleur produite par fission de l'uranium au cœur du réacteur est transmise composante par composante jusqu'à la turbine qui entraîne le générateur et permet la production d'énergie électrique. Trois circuits principaux assurent ces échanges thermiques :

- Le **circuit primaire (sodium).** La grande quantité de chaleur produite au cœur du réacteur fait augmenter la température du sodium qui circule à l'intérieur du cœur. Pour évacuer la puissance thermique, les pompes primaires envoient continuellement du sodium froid dans le cœur. La chaleur du circuit primaire est transférée au circuit secondaire par l'intermédiaire de l'échangeur de chaleur.

- Le **circuit secondaire (sodium).** La chaleur du circuit primaire transférée au secondaire est ensuite transmise au générateur de vapeur.

- Le **circuit vapeur (eau liquide – vapeur).** La détente de la vapeur générée entraîne la turbine.

- Le **circuit de refroidissement (eau).** La vapeur en sortie de turbine est condensée par le circuit de refroidissement (condenseur contenant de l'eau froide en provenance d'une source froide).

Dans le cadre des études d'accidents graves, on considère ici le scénario accidentel de type ULOF (*Unprotected Loss Of Flow*), qui correspond au transitoire de perte de débit primaire non protégé. Cette perte de débit, résulte du dysfonctionnement des pompes primaires sans reprise de secours ni chute des barres de contrôle. La perte du débit entraîne un échauffement progressif du cœur. Cette hausse de la température peut ensuite entraîner une ébullition du sodium accélérant la hausse de température, et pouvant conduire, *in fine*, à la fusion partielle ou totale du cœur.

1.3.2 Présentation de l'outil physique orienté conception MACARENa

En support à l'étude des scénarios accidentels tels que l'ULOF, le CEA a entrepris le développement d'outils de calculs analytiques simulant les différents phénomènes physiques régissant ces transitoires. Ces outils sont beaucoup plus rapides que des codes mécanistes : une ou deux heures pour une simulation avec les premiers, contre plusieurs jours ou semaines avec les seconds. Ainsi, ces codes rapides permettent d'envisager une prise en compte des incertitudes sur les paramètres d'entrée (variables physiques, variables de modèle ...), via des approches statistiques basées sur des simulations de type Monte-Carlo.

On considère ici l'outil orienté conception MACARENa (Modélisation de l'ACcident d'Arrêt des pompes d'un Réacteur refroidi au sodium) qui modélise la phase d'initiation et la phase primaire de l'accident ULOF. Cet outil, développé dans le cadre d'une précédente thèse au CEA, a été partiellement validé sur la base de données expérimentales et de résultats de simulation des codes mécanistes (Droin, 2016). Des études réalisées dans cette même thèse, ont montré que la séquence accidentelle prédite par le simulateur varie considérablement en fonction des paramètres d'entrée : paramètres liées à la conception ou la configuration du cœur avant l'accident, paramètres caractéristiques du déroulement du transitoire, paramètres des modèles physiques tels que les contre-réactions neutroniques, etc. **Il est donc essentiel de prendre en compte l'incertitude de ces paramètres et d'évaluer précisément, au travers d'une analyse de sensibilité, leur impact sur les résultats de l'outil.** Il s'agit entre autres **d'identifier les paramètres significativement influents** en vue par exemple d'une réduction des incertitudes dans de futures études. Ainsi, des premières études d'analyse de sensibilité ont été réalisées dans Droin (2016) en distinguant deux type d'incertitudes en entrée : les *incertitudes irréductibles* (ou aléatoires) inhérentes à la variabilité naturelle des phénomènes et les

incertitudes réductibles (ou *épistémiques*) liées au manque de connaissance[3] (Hora, 1996; Dantan et al., 2013). Dans le premier cas, ces incertitudes sont modélisées par une distribution de probabilité, estimées sur les données expérimentales, des données issues de simulations ou des données de conception du cœur. Dans le second cas, la modélisation des incertitudes ne s'appuie que sur des avis d'experts : il n'y a souvent pas de distribution de probabilité clairement identifiée, seulement un intervalle de variation. L'hypothèse d'une distribution uniforme sur cet intervalle est alors souvent réalisée dans Droin (2016). **Il est alors important d'évaluer l'impact de la méconnaissance de la loi de probabilité de ces variables ou du choix arbitraire d'une loi sur les résultats de l'analyse de sensibilité.**

Le scénario ULOF modélisé avec le simulateur MACARENa constitue ainsi le cas test fil rouge (désigné ULOF-MACARENa) sur lequel seront appliqués les méthodes et outils développés dans cette thèse.

1.4 Problématiques et objectifs

Comme expliqué précédemment, les mesures HSIC sont des outils efficaces dans le cadre de l'ASG. Suivant le cas d'étude, ces mesures peuvent être utilisées pour cribler ou hiérarchiser les entrées par ordre d'influence sur la sortie. Pour hiérarchiser les entrées par ordre d'influence, des indices de sensibilité normalisés ont été proposés par Da Veiga (2015). Pour réaliser un criblage des entrées, des tests d'indépendance basés sur la statistique HSIC sont individuellement effectués entre chaque entrée et la sortie (De Lozzo and Marrel, 2016b). À l'issue de ces tests, l'hypothèse de l'indépendance est retenue ou rejetée. Les entrées dont l'hypothèse d'indépendance avec la sortie est rejetée sont considérées comme significativement influentes sur la sortie. À la lumière de ces travaux récents sur les mesures HSIC pour l'ASG, nous proposons dans cette thèse des extensions et améliorations pour répondre aux deux objectifs suivants.

Analyse de sensibilité globale en présence d'incertitudes de second niveau. Les mesures HSIC sont particulièrement efficaces pour l'ASG lorsque les distributions de probabilité de toutes les entrées sont parfaitement connues. Cependant, dans certains cas, comme celui du cas test ULOF-MACARENa, des incertitudes sur le modèle probabiliste des entrées peuvent exister. Ces incertitudes proviennent généralement d'une divergence d'avis d'experts, d'un manque total ou partiel de données pour caractériser suffisamment les distributions ou encore d'un manque de confiance sur la qualité des données existantes. Ces incertitudes sur les distributions de probabilité seront qualifiées dans ce manuscrit d'incertitudes de second niveau, pour les dissocier des incertitudes sur les variables elle-mêmes (incertitudes de premier niveau). En présence d'incertitudes de second niveau, on désignera par ASG1 l'analyse de sensibilité de la sortie du simulateur en fonction des entrées incertaines lorsque le modèle probabiliste des entrées est connu et fixé. **On appellera alors ASG2, l'analyse de sensibilité visant à quantifier l'impact des incertitudes des lois des entrées sur les résultats d'ASG1.**

Dans ce contexte, un premier objectif de cette thèse est de proposer une méthodologie efficace pour l'ASG2 nécessitant un nombre raisonnable d'appels au code. Cette étude fera l'objet du chapitre 4 de ce manuscrit.

Amélioration de la qualité du criblage basé sur les mesures HSIC. Comme mentionné auparavant, un des objectifs de l'ASG peut être de réaliser un criblage des entrées, en utilisant des tests statistiques d'indépendance entre chaque entrée et la sortie. Un test statistique d'indépendance est une procédure de décision entre deux hypothèses : l'hypothèse nulle suivant

[3] Ce point sera discuté plus en détails dans les chapitres suivants.

laquelle une entrée donnée et la sortie sont indépendantes et son opposée, l'hypothèse alternative. Dans cette prise de décision et suivant la taille de l'échantillon disponible, cette décision statistique a une probabilité non nulle d'être fausse. La probabilité d'avoir tort sous l'hypothèse nulle est généralement appelée *erreur de première espèce* ou *niveau de test*. La probabilité que le test se trompe sous l'hypothèse alternative est quant à elle appelée *erreur de seconde espèce*. Le contrôle théorique et pratique du niveau des tests d'indépendance est possible et fixé généralement à un seuil de 5% ou 10%. En revanche, à l'heure actuelle il n'y a aucun contrôle théorique ni pratique de l'erreur de seconde espèce.

Pour les tests basés sur les mesures HSIC, on soulève ainsi deux points importants afin d'améliorer la robustesse des tests et de mieux contrôler l'erreur de seconde espèce. Le premier point est de s'affranchir du choix non justifié théoriquement des noyaux associés aux HSIC. En effet, des choix heuristiques sont généralement adoptés pour la définition de ces noyaux et peuvent impacter les résultats des tests. Le deuxième point d'amélioration consiste à contrôler et idéalement diminuer l'erreur de seconde espèce des tests afin d'augmenter la probabilité d'atteindre un *criblage parfait*.

Ainsi, le second objectif de cette thèse est de proposer une procédure de test agrégeant plusieurs tests unitaires basés sur des mesures HSIC avec des noyaux différents. Les résultats théoriques et numériques de cette méthodologie seront présentés au chapitre 5.

1.5 Organisation du document

Dans l'objectif de répondre aux deux problématiques introduites dans la précédente section, ce document sera organisé comme suit. Après un chapitre reprenant l'introduction en version anglaise, le chapitre 3 présente une revue théorique et méthodologique des mesures HSIC. De nouveaux développements autour de leur estimation à partir d'un échantillon généré suivant une loi de probabilité différente de celle des entrées (loi alternative) sont ensuite proposés. Ensuite, l'accent sera mis sur les tests d'indépendance basés sur les mesures HSIC. Des généralités sur les tests statistiques d'indépendance et en particulier la *vitesse de séparation uniforme* qui permet de juger de la qualité d'un test donné sont présentées. Enfin, les tests d'indépendance construits à partir de la statistique HSIC sont introduits, d'abord en version asymptotique avant de proposer une version non-asymptotique de ces tests.

À la lumière des techniques d'estimation proposées dans le chapitre 3, une méthodologie pour l'ASG2 utilisant un seul échantillon (bien choisi) est proposée dans le chapitre 4. L'efficacité de la méthodologie est illustrée sur un exemple analytique et plusieurs choix méthodologiques possibles sont comparés. Une application sur le cas test du transitoire ULOF-MACARENa est réalisée afin de prendre en compte l'incertitude sur les lois de certains paramètres d'entrée et d'évaluer leur impact sur l'ASG1. Enfin, pour ouvrir de nouvelles perspectives applicatives, la méthodologie d'ASG2 est étendue au traitement des incertitudes épistémiques et comparée à l'approche Dempster-Shafer.

Dans le chapitre 5, une procédure innovante d'agrégation de plusieurs tests HSIC est développée. Il s'agit plus précisément d'agréger plusieurs paramétrisations des mesures HSIC. Cette proposition s'appuie sur une étude préalable de l'erreur de second ordre du test unitaire basé sur la mesure HSIC et plus particulièrement sur la vitesse de séparation du test. A partir de là, un test agrégé est proposé et l'on démontre que cette procédure peut être quasiment optimale pour un choix adéquat de la collection de paramètres à agréger. Des exemples numériques sont implémentés et permettent d'un côté, de comparer les différents choix méthodologiques, et d'un

autre côté, d'illustrer l'efficacité de la procédure en la comparant à d'autres tests de la littérature. Enfin, la méthodologie est appliquée au cas test du transitoire ULOF-MACARENa afin de réaliser un criblage des entrées incertaines.

En conclusion, les chapitres 6 et 7 présentent respectivement en versions anglaise et française, une synthèse des nouvelles méthodes développées dans ce document en support à l'analyse de sensibilité des simulateurs numériques. Les perspectives de ces travaux et les possibles améliorations sont aussi discutées.

Chapter 2

Introduction (english)

2.1 Context

As part of safety studies for nuclear reactors, computation codes (or numerical simulators) are fundamental for understanding, modelling and predicting physical phenomena. These tools take a large number of input parameters, characterizing the studied phenomenon or related to its physical and numerical modelling. The information related to some of these parameters is often limited or uncertain, this can be due to the lack or absence of data, measurement or modelling errors or even a natural variability of the parameters. **These input parameters, and consequently the simulator output, are thus uncertain.** This is referred to as uncertainty propagation. It is important to consider not only the nominal values of inputs, but also the set of all possible values in the variation range of each uncertain parameter. It is therefore important to take into account the input uncertainties and their effects on the output, which constitutes a major step for safety studies.

The generic approach to deal with uncertainties in computation codes has been extensively studied in the past few decades. In the general literature on the subject (De Rocquigny et al., 2008; Ghanem et al., 2017), the usual methodological approach is divided into four key steps. This generic approach is illustrated by Figure 2.1. **The first step, step A, is the specification of the problem,** which consists in defining the system to be studied (model, simulator or measurement process), identifying uncertain or fixed input variables, as well as the quantities of interest to be studied (derived from the model output variables). **Step B then aims to quantifying the input uncertainties.** In the probabilistic framework, these uncertainties are modelled by fully or partially known probability distributions (Helton, 1997; Oberkampf et al., 2001). The selection of such probabilistic models depends on eventual available data, expert opinions or bibliographic databases. Recently, Bae et al. (2004) and Swiler et al. (2009) propose alternative quantification methods for epistemic uncertainties, i.e. more related to the lack of knowledge than the randomness of the phenomenon. One of the main approaches used by these methods is the theory of *evidence*, also known as the *Dempster-Shafer* theory (Dempster, 1967; Shafer, 1976). **At step C, uncertainties are propagated:** the objective is to quantify how input uncertainties affect the output(s) predicted by the model, and more precisely the quantity of interest. This quantity of interest deriving from the model outputs is directly linked to the objectives of the study. This may include the output mean or dispersion, a probability of exceeding a critical value or a quantile. Various specific approaches, deterministic or based on Monte-Carlo simulations, have been developed according to the studied quantity of interest (Cannamela, 2007). Alongside uncertainty propagation, **a sensitivity analysis, step C' of**

the approach, can be conducted. The **sensitivity analysis aims to determine how the variability of the input parameters affects the value of the output or the quantity of interest** (Saltelli et al., 2004; Iooss, 2011)**.** It thus allows to identify and perhaps quantify, for each input parameter or group of parameters, its contribution to the variability of the output. The purpose of sensitivity analysis can be to prioritize input parameters by order of influence on the output variability, or to separate the inputs into two groups: those which mostly influence the output uncertainty and those whose influence can be neglected. This input splitting into two groups is known as "screening". The sensitivity analysis results provide valuable information for the impact of uncertain inputs, the comprehension of the model and the underlying physical phenomenon. It can also be used for various purposes: reducing uncertainties by targeting characterization efforts on most influential inputs, simplifying the model by setting non-influential inputs to reference values, or validating the model with respect to the modeled phenomenon. These issues explain the amount of recent studies on statistical tools and methods for sensitivity analysis. One of the most commonly used methods in industrial applications is based on a decomposition of the output variance (Hoeffding, 1992; Sobol, 1993), each term of the decomposition represents the contribution share of an input or a group of inputs to the output variance. As a result of this approach, Sobol's indices are obtained. These easy-to-interpret indices have several practical drawbacks (expensive estimation in terms of the number of the code simulations, partial information provided by the variance). To overcome these limitations, other approaches based on **dependence measures** have recently been proposed (Da Veiga, 2015). These measures have several advantages, which are described below, and have produced promising results in several industrial applications (De Lozzo and Marrel, 2016b).

In the scope of sensitivity analysis for numerical simulators, the work carried out in this thesis seeks to propose new innovative statistical methods based on dependence measures, to effectively address some issues raised by their implementation on industrial applications.

2.2 Global sensitivity analysis based on dependence measures

As previously stated, Sensitivity Analysis (SA) methods aim to determine how the variability of a model's inputs affects its output variability. Two main fields are distinguished: **Local Sensitivity Analysis** (LSA) and **Global Sensitivity Analysis** (GSA).

Local sensitivity analysis studies the output variation for small input shifts near their reference values (also called nominal values). Among LSA methods, the principal ones are those based on **partial derivatives** (Alam et al., 2004; Pujol, 2009) and those based on **adjoint modeling** (Hall et al., 1982; Cacuci, 1981, 2003). The first involves estimating the partial derivatives of the numerical model with respect to each input at its nominal point. These partial derivatives represent the effect of each input perturbation on the total output perturbation and are directly interpreted as local sensitivity indices. These indices can be estimated using *One-At-a-Time* (OAT) experimental design techniques, which consist of perturbating only one input at a time by fixing the other inputs to their nominal values (Morris, 1991). The adjoint modeling approach is a purely analytical method that can be used when an analytical formula of the model is explicitly known. The adjoint modeling is numerically intrusive, which means that its application requires the development of a model for computing partial derivatives in each direction. This method is therefore not applicable in the case of "black box" simulators where only the inputs and outputs of the model are accessible.

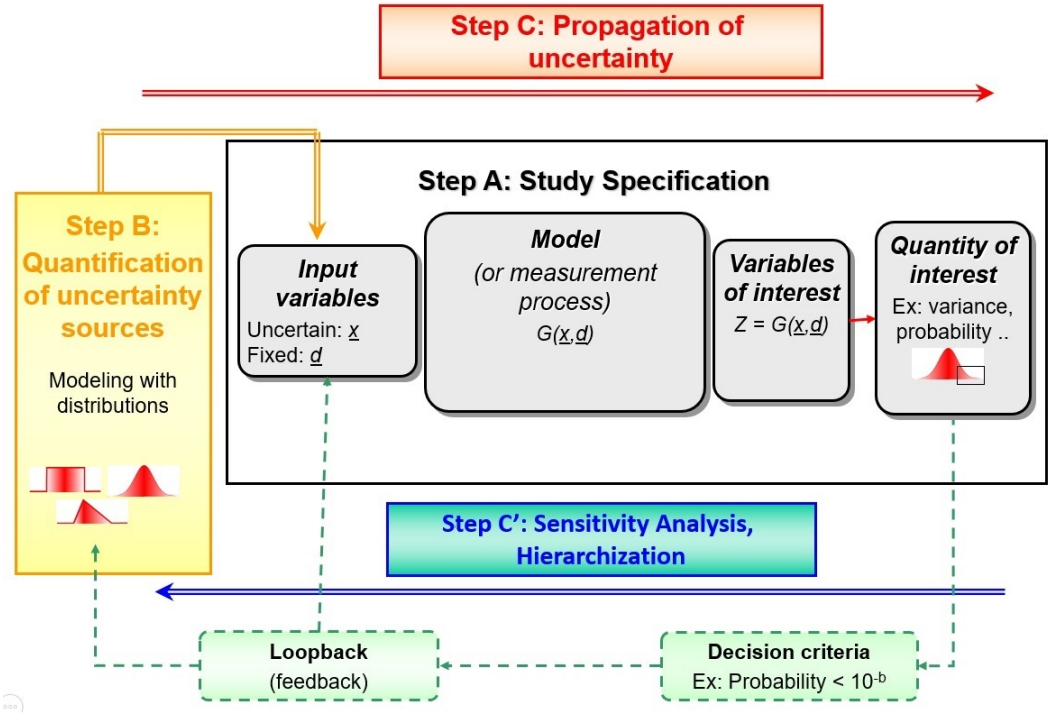

Figure 2.1 – General scheme for the methodology of uncertainty treatment from De Rocquigny et al. (2008).

All these LSA methods thus fail to consider the input uncertainties over their whole variation range. To assess and quantify the global impact of each input uncertainty on the output, statistical methods of **Global Sensitivity Analysis (GSA)** have been developed. In contrast to LSA, the global approach requires characterizing the input uncertainties over their variation range (step B, Figure 2.1), for example by assigning a probability distribution to the input vector. The statistical methods for GSA are mostly based on Monte Carlo simulations of the model, i.e. on a random sampling of inputs according to their probability distributions. Common GSA methods include the *Derivative-based Global Sensitivity Measures*, also called DGSM indices (Kucherenko et al., 2009; Kucherenko and Iooss, 2017; Sobol and Kucherenko, 2010). The construction of these indices is based on a generalization of local sensitivity measures by averaging partial derivatives with respect to each input over its range of variation. However, estimating these indices requires a large number of code calls, which considerably limits its use in the case of expensive models[1]. To overcome this disadvantage, estimation strategies based on the use of metamodels approximating the model output have been proposed. We can mention the works of Sudret and Mai (2015) based on chaos polynomials or those of De Lozzo and Marrel (2016a) using Gaussian process metamodels. Another approach conventionally used for the GSA is based on the decomposition of the output variance, where each term of the decomposition represents the part of the contribution of an input or a group of inputs to the output variance. Originally introduced in Hoeffding (1948a), this decomposition is commonly called : ANOVA decomposition (for *ANalysis Of VAriance*). Sensitivity indices are directly derived from this de-

[1]Expensive refers here to the time spent on each simulation of the model or computation code, which limits the total number of possible simulations.

composition: these are the Sobol' (Sobol, 1993) indices, mentioned above. **Sobol's indices** are easily interpretable, but their expressions involve multidimensional integrals whose estimation by Monte-Carlo methods requires in practice a very large number of model simulations (several tens of thousands). Their direct estimation is therefore very often impossible for time-consuming simulators. Several studies have been developed to reduce the estimation budget of these indices. Other approaches requiring additional model regularities and based on spectral decomposition methods were also considered. Examples include the FAST method (FAST for *Fourier Amplitude Sensitivity Testing*) introduced in Cukier et al. (1973) and then studied in Lemaître (2014) and Iooss and Lemaître (2015). Methods such as E-FAST (*Extended Fourier Amplitude Sensitivity Testing*) and RBD-FAST (*Random Balance Design Fourier Amplitude Sensitivity Testing*) introduced in Saltelli et al. (1999) and Tarantola et al. (2006) respectively suggest some improvements of the classical FAST method. Nevertheless, the number of model calls using these methods is still very high. Here again, a possible option is to estimate these indices using metamodels: the estimation of Sobol' indices by chaos polynomials, local polynomials or Gaussian processes have been respectively proposed in Sudret (2008), Da Veiga et al. (2009) and Marrel et al. (2009). Such approaches, however, require the ability to construct a sufficiently predictive metamodel, which can be complicated for highly non-linear simulators and/or for a large number of input variables. Moreover, regardless of the difficulties associated with their estimation, Sobol' indices only consider the variance of the output and do not evaluate the impact of each input on the whole probability distribution of the output. They are thus not equivalent to the independence between the output and each input (except for the total Sobol' indices).

The **dependence measures** recently introduced for the GSA by Da Veiga (2015), make it possible to overcome several of the limitations listed above. First, these measures quantify from a probabilistic point of view the dependence between each input and output. Thus, the nullity of a dependence measure between an input and the output is equivalent to the independence of these two random variables. These measures can be used quantitatively to prioritize the inputs in order of influence on the output, as well as qualitatively to perform the screening of inputs, for instance by using statistical tests like those in De Lozzo and Marrel (2016b). The use of statistical tests to identify non-influential variables provides a more rigorous and accurate statistical and mathematical framework than a simple assessment and comparison of sensitivity measures. In particular, this avoids the arbitrary choice of a threshold value for sensitivity measures, beyond which an input variable is considered influential. Among the existing dependence measures in the literature, we can first mention the **dissimilarity measures** introduced by Baucells and Borgonovo (2013). The idea of constructing these measures is based on comparing the probability distribution of the output with its distribution when a given input is fixed. These measures actually belong to a broader class based on Csiszàr's f-divergence (Csiszár, 1972). This latter includes several older notions of dependence such as Hellinger's distance (Hellinger, 1909), Kullback-Leibler's divergence (Kullback and Leibler, 1951) or the total variation distance (Rudin et al., 1992). Moreover, Da Veiga (2015) also highlights the links between Csiszàr's f-divergence and the mutual information introduced by Shannon (1948) as well as with the mutual information of the loss square (Suzuki et al., 2009), these measures can be interpreted as dissimilarity measures. Note that Sobol' indices can also be defined as dissimilarity measures (Chabridon, 2018). Despite their interesting theoretical properties, the estimation of measures based on Csiszàr's f-divergence is in practice costly in terms of the number of simulations, particularly in large dimension[2].

Other dependence measures whose estimation suffers less from the "Curse of dimensionality" have also been proposed by Da Veiga (2015). Among them is the **distance covariance** based

[2]The large dimension here refers to a large number of input variables.

on the characteristic functions (Székely et al., 2007). It has been shown that this dependence measure has good properties for testing the independence between two random variables in large dimensions (Székely and Rizzo, 2013; Yao et al., 2018). It has also been shown that the distance covariance is part of a larger class of dependence measures (Székely and Rizzo, 2013), based on mathematical objects called *characteristic kernels* (Sriperumbudur et al., 2010). These dependence measures are highly effective for testing the independence between random variables of various types: scalar, vector, categorical, etc. Among them, the Hilbert-Schmidt Independence Criterion denoted **HSIC** (Gretton et al., 2005a), generalizes the notion of covariance between two random variables and thus makes it possible to capture a very wide spectrum of forms of dependence between the variables. For this reason, Da Veiga (2015), then De Lozzo and Marrel (2016b) investigated the use of HSIC measures for GSA and compared them to Sobol' indices. Note that the HSIC measures is identical to the distance covariance for a particular choice of kernels (Székely and Rizzo, 2013). As illustrated by De Lozzo and Marrel (2016b), HSIC indices also have the advantage of having a low estimation cost (in practice a few hundred simulations compared to several tens of thousands for Sobol' indices) and their estimation for all inputs does not depend on the number of inputs. In addition, statistical independence tests based on HSIC measures have also been developed by Gretton et al. (2008), in an asymptotic framework. More recently, a first extension to a non-asymptotic framework has been proposed by De Lozzo and Marrel (2016b), which have also shown the effectiveness and great interest of HSIC-based statistical tests to screen input variables.

For all these reasons, this thesis focuses on HSIC-type dependence measures for the GSA of numerical simulators. More precisely, the objective is to propose new theoretical, methodological and applicative developments around these measures.

2.3 Description of test case application

This thesis is part of the demonstration of safety and risk control of the Generation IV sodium-cooled Fast Neutron Reactors (RNR-Na, Figure 2.2), conducted by the CEA and its partners. As their names imply, RNRs use the high kinetic energy of neutrons to fuse uranium nuclei, in contrast to thermal neutron reactors (Pressurized Water Reactors, for example) where neutrons are slowed down to increase the probability of interacting with uranium atoms. As part of the safety studies, several severe reactor accident scenarios are studied through experimental tests and numerical simulations. Serious accidents are defined as those that lead to partial or total fusion of the reactor core. The temporal evolution of various accident-related physical quantities (also known as accidental transients) allows physicists to better understand the physical phenomena involved and to evaluate the behaviour of the core.

2.3.1 Presentation of the RNR-Na reactor and the ULOF accident

As shown in Figure 2.2, the general operation of an RNR-Na nuclear reactor is based on heat exchanges producing electrical energy. The heat produced by fission of uranium in the reactor core is transmitted to the turbine component by component, which powers the generator and produces electrical energy. Three main circuits ensure these heat exchanges:

- The **primary circuit (sodium).** The large amount of heat produced in the reactor core increases the temperature of the sodium flowing inside the core. To evacuate the thermal power, the primary pumps continuously inject cold sodium into the core.

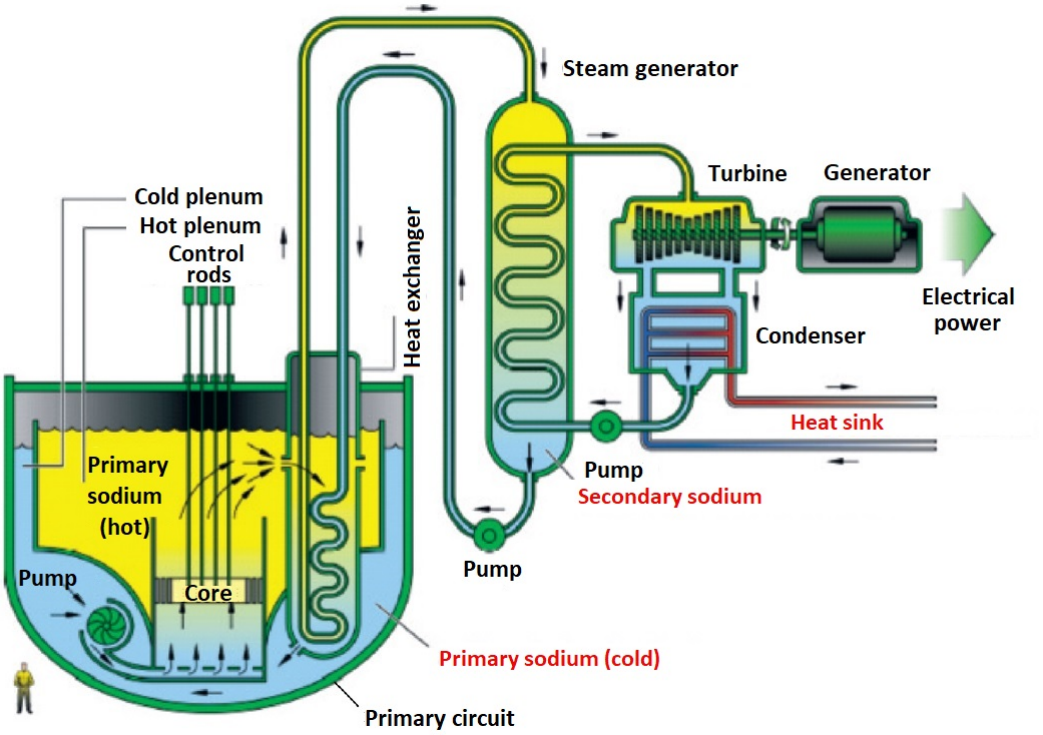

Figure 2.2 – General operating scheme of an RNR-Na reactor, from Droin (2016).

- The **secondary circuit (sodium).** The heat from the primary circuit transferred to the secondary circuit is then transmitted to the steam generator.
- The **steam circuit (liquid water - steam).** The expansion of the generated steam powers the turbine.
- The **cooling circuit (water).** The steam at the turbine outflow is condensed by the cooling circuit (condenser containing cold water from a cold source).

In severe accident studies, we consider here the ULOF accident scenario (*Unprotected Loss Of Flow*), which corresponds to the transient of unprotected primary flow loss. This loss of flow rate is due to the dysfunction of the primary pumps without emergency restart or fall of the control rods. The loss of flow leads to a gradual heating of the core. This temperature increase can then lead to sodium boiling, accelerating the temperature increase, and may lead, *in fine*, to partial or total fusion of the core.

2.3.2 Presentation of the MACARENa design-oriented physical tool

In support of the study of accident scenarios such as ULOF, the CEA has started the development of analytical computational tools simulating various physical phenomena ruling these transients. These tools are much faster than mechanistic codes: one or two hours for a simulation using the first tools, compared to several days or weeks with the second ones. These fast codes thus make

it possible to take into account input uncertainties (physical variables, model variables, etc.), via statistical approaches based on Monte Carlo simulations.

We consider here the design-oriented tool MACARENa (French: *Modélisation de l'ACcident d'Arrêt des pompes d'un Réacteur refroidi au sodium*) which models the initiation and the primary phases of the ULOF accident. This tool, previously developed as part of a PhD thesis at the CEA, has been partially validated using experimental data and simulation results from mechanistic codes (Droin, 2016). Studies carried out in this same thesis have shown that the accident sequence predicted by the simulator varies considerably according to the inputs: parameters related to the design or configuration of the core before the accident, parameters characteristic of the transient sequence, parameters of physical models such as neutronic back-reactions, etc. **It is consequently crucial to take into account the uncertainty of these parameters and to accurately assess, through a sensitivity analysis, their impact on the simulator results.** This includes **identifying significantly influential parameters**, in order for example to reduce uncertainties in upcoming studies. Thus, first sensitivity analysis studies were carried out in Droin (2016) by distinguishing two types of input uncertainties: the *irreducible* (or random) uncertainties inherent in the natural variability of phenomena and the *reducible* (or *epistemic*) uncertainties related to lack of knowledge [3] (Hora, 1996; Dantan et al., 2013). In the first case, the uncertainties are modelled by a probability distribution, estimated from experimental data, simulation data or core design data. In the second case, uncertainty modelling is based only on expert opinion: there is often no clearly identified probability distribution, only a range of variation. The hypothesis of a uniform distribution over this interval is then often assumed in Droin (2016). **It is therefore important to evaluate the impact of a lack of knowledge of these variables probability distribution or the arbitrary choice of a distribution on the sensitivity analysis results.**

The ULOF scenario modelled by the MACARENa simulator thus constitutes the main thread test case (called ULOF-MACARENa) on which the methods and tools developed in this thesis will be applied.

2.4 Issues and objectives

As explained above, HSIC measures are effective tools for GSA purpose. Depending on the study case, these measures can be used either to screen or prioritize inputs in order of influence on the output. To prioritize the inputs in order of influence, normalized sensitivity indices have been proposed by Da Veiga (2015). To perform input screening, independence tests based on HSIC statistics are performed individually between each input and the output (De Lozzo and Marrel, 2016b). At the end of these tests, the hypothesis of independence is either accepted or rejected. Inputs whose independence assumption with the output is rejected are considered to have a significant influence on the output. In the light of these recent works on HSIC measures for the GSA, we propose in this thesis some extensions and improvements to adress the following two objectives.

Global sensitivity analysis for second-level uncertainties. HSIC measures are effective for GSA when the probability distributions of all inputs are fully known. However, in some cases, such as the ULOF-MACARENa test case, uncertainties about the probabilistic input model may exist. These uncertainties generally stem from a divergence of expert opinions, a total or partial lack of data to sufficiently characterize the distributions or a lack of confidence in the quality of existing data. These uncertainties on probability distributions will be referred to in this

[3] This point will be discussed in more detail in the following chapters.

manuscript as second-level uncertainties, to dissociate them from uncertainties on the variables themselves (first-level uncertainties). In the presence of second-level uncertainties, the sensitivity analysis of the simulator output will be referred to as GSA1, when the probabilistic input model is known and fixed. **We will then call GSA2, the sensitivity analysis aiming to quantify the impact of uncertainties of input distributions on GSA1 results.**

In this context, a first objective of this thesis is to propose an efficient methodology for GSA2 requiring a reasonable number of code calls. This study will be the subject of chapter 4 of this manuscript.

Improvement of the quality of screening based on HSIC measures. As mentioned above, one of the objectives of GSA may be to perform input screening, using statistical tests of independence between each input and the output. A statistical independence test is a decision-making procedure between two hypotheses: the null hypothesis that a given input and the output are independent and its opposite, the alternative hypothesis. In this decision making and depending on the size of the available sample, this statistical decision has a non-zero probability of being false. The probability of being wrong under the null hypothesis is generally called *first-kind error* or *level of test*. The probability that the test is wrong under the alternative hypothesis is called *second-kind error*. Theoretical and practical control of the level of independence tests is possible and generally set at a threshold of 5% or 10%. By contrast, there is currently no theoretical or practical control of the second-kind error.

For tests based on HSIC measures, two important points are raised in order to improve the robustness of the tests and better control the second-kind error. The first point is to avoid the theoretically unjustified choice of the kernels associated to HSIC measures. Indeed, heuristic choices are generally adopted for the definition of these kernels and can impact the test results. The second point for improvement is to control and ideally reduce the second-kind error of the tests, in order to increase the probability of achieving a *perfect screening*.

Thus, the second objective of this thesis is to propose a test procedure that aggregates several unit tests based on HSIC measures with different kernels. The theoretical and numerical results of this methodology will be presented in chapter 5.

2.5 Organization of the document

In order to address the two issues introduced in the previous section, this document will be organized as follows. After a chapter with the introduction written in English, chapter 3 presents a theoretical and methodological review of HSIC measures. New developments around their estimation from a sample generated according to a probability distribution different from the prior one of the inputs (alternative distribution) are then proposed. Then, the focus will be on independence tests based on HSIC measures. General background on statistical independence tests and in particular the *uniform separation rates* over classes of regular alternatives, allowing to adjudge the quality of a given test is presented. Finally, statistical independence tests based on HSIC statistics are introduced, first in the asymptotic then non-asymptotic frameworks.

In light of the estimation techniques proposed in Chapter 3, a methodology for GSA2 using a well-chosen single sample is proposed in Chapter 4. The effectiveness of the methodology is illustrated with an analytical example and several possible methodological choices are compared. An application on the test case of the ULOF-MACARENa transient is performed, in order to take into account the distribution uncertainties of some inputs and to evaluate their impact on

GSA1. Finally, to open up new application perspectives, the GSA2 methodology is extended to the treatment of epistemic uncertainties and compared to the Dempster-Shafer approach.

In chapter 5, an innovative procedure for aggregating several HSIC tests is developed. More precisely, it involves aggregating several parameterizations of HSIC measures. This procedure is based on a preliminary study of the second-kind error of single tests based on HSIC measures and more particularly on the separation rate of these tests over classes of regular alternatives. From this point on, an aggregated test is proposed and it is shown that this procedure can be nearly optimal for an appropriate choice of the collection of parameters to be aggregated. Numerical examples are implemented and allow, on the one hand, to compare the different methodological choices and, on the other hand, to illustrate the effectiveness of the procedure by comparing it with other tests in the literature. Finally, the methodology is applied to the ULOF-MACARENa transient test case to perform a screening of uncertain inputs.

In conclusion, chapters 6 and 7 respectively present in English and French versions, a synthesis of the new methods developed in this document in support of the sensitivity analysis of numerical simulators. The prospect for this work and some possible improvements are also discussed.

Chapter 3

Review and theoretical developments around Hilbert-Schmidt dependence measures (HSIC)

3.1 Introduction and motivations

Since the earlier work of Sobol (1993), theoretical, methodological and applicative works in support of Global Sensitivity Analysis (GSA) for numerical simulators have grown increasingly. Several approaches and procedures have been proposed and further developed. Among them, *variance-based* methods follow the perspective of Sobol (1993), by computing the impact of each input on the variance of the output. Statistical estimation of Sobol' indices (Saltelli et al., 2010; Owen, 2013) as well as the properties of the associated estimators have been widely investigated (Janon et al., 2014; Da Veiga and Gamboa, 2013). However, despite their good theoretical and practical properties as well as the ease of their interpretation, Sobol' indices suffer from several limitations. First, due to the multidimensional integrations in their analytical formulas, the estimation of each Sobol' index requires in practice a large number of simulations (several thousands), which prevents their direct use for time-consuming codes. In such cases, alternative methods of estimation using surrogate models have been developed (Oakley and O'Hagan, 2004; Da Veiga et al., 2009; Marrel et al., 2009). But, as explained in the Introduction (Chapter 1 in French and Chapter 2 in English), the construction of a surrogate model can be complicated in many cases. Moreover, the computation cost of all Sobol' indices depends directly on the number of inputs, which makes them not convenient to perform a preliminary screening in high dimension. In addition to that, a noteworthy point raised by Da Veiga (2015) is that this approach only focuses on the variance of the output. Yet, the variance is partially informative on the output distribution. Other less used approaches such as *derivative measures* (Kucherenko et al., 2009; Sobol and Kucherenko, 2010) and *dissimilarity measures* (Baucells and Borgonovo, 2013; Csiszár, 1972) have been explored. All these measures are good indicators for the global impact of input uncertainties on the output. Moreover, from a theoretical point of view, these measures and their associated estimators have good properties. Nevertheless, practically speaking, a common drawback of these measures is the slowness of the convergence of estimators in high dimension,

also known as the "curse of dimensionality".

In the light of these elements, Da Veiga (2015) recently proposes a very interesting approach to deal with the limitations of Sobol' indices and other usual GSA methods. This new approach is based on mathematical tools called *dependence measures*. As its name implies, a dependence measure between two random variables is zero if and only if these random variables are independent. The definition of a dependence measure includes many well-known other notions related to the independence. Among them, we mention all the classical measures based on the f-divergence of Csiszàr (Csiszár, 1972), the mutual information based on the notion of entropy (Shannon, 1948) and the distance covariance based on characteristic functions (Székely et al., 2007). The most valuable family of dependence measures are those based on *Reproducing Kernel Hilbert Spaces* (RKHS, Aronszajn, 1950). Originally used in machine learning, these measures offer several advantages comparing to other dependence measures. Indeed, they are easy to adapt for multidimensional random variables and are cheap to estimate[1] comparing to other existing measures. In addition, they can be generalizable to other types of random variables (categorical variables, permutations, graphs, etc.). One of the earlier RKHS dependence measures is the Kernel Canonical Correlation (KCC), introduced in Bach and Jordan (2002). Unfortunately, the estimation of the KCC is not practical, as it requires an extra regularization, which has to be adjusted. Other dependence measures based on RKHS, easier to estimate have been proposed later. For instance, the Kernel Mutual Information (KMI, Gretton et al., 2003, 2005b) and the COnstrained COvariance (COCO, Gretton et al., 2005c,b), which are relatively easy to interpret and implement, have been widely used. Last but not least, one of the most interesting kernel dependence measure is the Hilbert-Schmidt Independence Criterion (HSIC, Gretton et al., 2005a). The HSIC has a very low computational cost and seems to numerically outperform all the previous RKHS measures (Gretton et al., 2005a). This is the reason why we focus our attention on this dependence measure for GSA.

3.2 Definition of HSIC and link with independence

Throughout the rest of this document, the numerical model is represented by the relation:

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

where $X_1, \ldots, X_d$ and Y are respectively the d uncertain inputs and the uncertain output, evolving in one-dimensional real areas respectively denoted $\mathcal{X}_1, \ldots, \mathcal{X}_d$ and $\mathcal{Y}$. $\mathcal{M}$ denotes the numerical simulator. We note $\mathbf{X} = (X_1, \ldots, X_d)$ the vector of uncertain inputs. As part of the probabilistic approach, the d inputs are assumed to be continuous and independent random variables with known densities. These densities are respectively denoted $f_1, \ldots, f_d$. Finally, $f(x_1, \ldots, x_d) = f_1(x_1) \times \ldots \times f_d(x_d)$ denotes the density of the random vector $\mathbf{X}$. As the model $\mathcal{M}$ is not known analytically, a direct computation of the output probability density as well as dependence measures between $\mathbf{X}$ and Y is impossible. Only observations (or realizations) of $\mathcal{M}$ are available. It is therefore assumed in the following that we have a n-sample of inputs and associated outputs $\left(\mathbf{X}^{(i)}, Y^{(i)}\right)_{1 \leq i \leq n}$, where $Y^{(i)} = \mathcal{M}(\mathbf{X}^{(i)})$ for $i = 1, \ldots, n$.

3.2.1 General principle and definition

The idea of constructing the HSIC measure (Gretton et al., 2005a) between an input X_k and the output Y, is based on a generalization of the "classical" notion of covariance between these

[1] Cheap here means that the required number of observations (here code simulations) is small.

random variables. Covariance only detects linear dependence and its nullity is not equivalent to independence. In contrast, **HSIC measures allow to simultaneously take into account many forms of dependence between X_k and Y, relying on particular Hilbert spaces called Reproducing Kernel Hilbert Spaces (RKHS)**. The reader can refer to Aronszajn (1950) for a complete bibliography on RKHS spaces.

Definition 3.1. *Let $\mathcal{S}$ be an arbitrary set and $\mathcal{H}$ be a Hilbert space of real-valued functions on $\mathcal{S}$ with a scalar product denoted $\langle\ ,\ \rangle_{\mathcal{H}}$. The Hilbert space $\mathcal{H}$ is said to be a RKHS if, for all s in $\mathcal{S}$: the application $h \in \mathcal{H} \mapsto h(s)$ is a continuous linear form.*

The particularity and interest of RKHS spaces is the Riesz representation theorem. This representation consists in associating to each value of the starting set, a function in the RKHS. Each element is then represented by a random functional variable belonging to a space having good properties.

Proposition 3.1 (Riesz representation theorem)**.** *Let $\mathcal{H}$ be a RKHS space associated to a set $\mathcal{S}$ and with a scalar product denoted $\langle\ ,\ \rangle_{\mathcal{H}}$. Then, for all s in $\mathcal{S}$ there is a unique φ_s in $\mathcal{H}$ such that: $h(s) = \langle h, \varphi_s \rangle_{\mathcal{H}}$, for all h in $\mathcal{H}$.*

According to Riesz theorem, we associate the variation domain $\mathcal{X}_k$ (resp. $\mathcal{Y}$) of X_k (resp. Y) with a RKHS space denoted $\mathcal{H}_k$ (resp. $\mathcal{G}$). We denote respectively by ϕ_k and ψ, the functional random variables representing X_k and Y in the RKHS spaces $\mathcal{H}_k$ and $\mathcal{G}$. It is then possible to define an operator between the random variables X_k and Y by defining an operator between ϕ_k and ψ. The idea of Gretton et al. (2005a) is to use the covariance operator in the RKHS spaces. Introduced in Baker (1973) and studied in Fukumizu et al. (2004), this operator is defined in a similar way of the usual covariance. The definition of such an operator requires first defining a product operator between two elements belonging to two different RKHS. This product, called the *tensor product*, is defined as follows:

Definition 3.2 (Tensor product)**.** *We define the tensor product between ϕ_k in $\mathcal{H}_k$ and ψ in $\mathcal{G}$ as being the operator:*

$$\begin{array}{rccc} \phi_k \otimes \psi & : & \mathcal{G} & \to & \mathcal{H}_k \\ & & h & \mapsto & \langle \psi, h \rangle_{\mathcal{G}}\ \phi_k. \end{array}$$

From the previous definition, the covariance operator between two elements ϕ_k and ψ is defined by analogy with the usual notion of covariance as

$$C_k := \mathbb{E}\left[\phi_k \otimes \psi\right] - \mathbb{E}\left[\phi_k\right] \otimes \mathbb{E}\left[\psi\right].$$

Note that, the last operator is well-defined as shown in Gretton (2019, Section 3). An interesting property of this operator is that it takes into account all the transformations ϕ_k of X_k and ψ of Y respectively belonging to the RKHS $\mathcal{H}_k$ and $\mathcal{G}$, through the following formula

$$\langle \phi_k, C_k \psi \rangle_{\mathcal{H}_k} = \mathrm{Cov}\left(\phi_k(X_k), \psi(Y)\right).$$

This means that, if the operator C_k is identically equal to zero and if the RKHS associated with X_k and Y are sufficiently rich[2], then C_k can be used to characterize the independence between X_k and Y. Assuming that C_k is *independence-characterizing*, it remains to statistically check its nullity. A direct verification from the operator expression being difficult, Gretton et al. (2005a) define an associated measure based on the Hilbert-Schmidt norm of the operator.

[2]This point will be detailed later.

Definition 3.3 (Norm of Hilbert-Schmidt)**.** *Let $\mathcal{G}$ and $\mathcal{H}$ be two Hilbert spaces and Λ an operator mapping from $\mathcal{G}$ to $\mathcal{H}$. The Hilbert-Schmidt norm of the operator Λ is defined as*

$$\|\Lambda\|_{\mathrm{HS}}^2 = \sum_{i,j} \langle u_i, \Lambda(v_j)\rangle_{\mathcal{H}}^2,$$

where $(u_i)_{i\geq 0}$ and $(v_j)_{j\geq 0}$ are orthonormal bases respectively of $\mathcal{H}$ and $\mathcal{G}$. In addition, if $\|\Lambda\|_{\mathrm{HS}}$ is finite, then the operator Λ is reffered to as a "Hilbert-Schmidt operator".

In particular, the covariance operator C_k is a Hilbert-Schmidt operator, as demonstrated for example in Gretton (2019, Section 3). Thus, the Hilbert-Schmidt Independence Criterion between X_k and Y is defined as the square of the Hilbert-Schmidt norm of C_k:

$$\mathrm{HSIC}(X_k, Y)_{\mathcal{H}_k,\mathcal{G}} = \|C_k\|_{\mathrm{HS}}^2.$$

Remark. *In the following, the notation* $\mathrm{HSIC}(X_k, Y)_{\mathcal{H}_k,\mathcal{G}}$ *is replaced by* $\mathrm{HSIC}(X_k, Y)$ *in order to lighten the expressions.*

Remark. *It is also possible and interesting to consider the greatest singular value of the operator C_k to study the dependence between X_k and Y. This notion is called the Constrained Covariance (Gretton et al., 2005b,c). The Constrained Covariance can be valuable to seek the transformations of X and Y maximizing the covariance, which correspond to the singular functions associated to the largest singular value of C_k. This can be particularly useful for detecting particular forms of dependence such as linear one.*

3.2.2 Kernel-based representation and characterization of independence

As explained in the previous section, the construction of HSIC measures first requires to respectively associate RKHS spaces $\mathcal{H}_k$ to the variation domains $\mathcal{X}_k, k = 1, \ldots, d$ and $\mathcal{G}$ to $\mathcal{Y}$. The HSIC characteristics depend entirely on the choice of these RKHS. This choice consists in associating a mapping function that assigns to each element of the domain a representative functional in the RKHS and a scalar product that defines the nature of the relationships between the representatives and so between the elements of the domain. The application that defines this scalar product is called *kernel* and is defined as follows:

Definition 3.4. *Let $(\mathcal{H}, \langle\ ,\ \rangle_{\mathcal{H}})$ be a RKHS space associated with a set $\mathcal{S}$. For all $s \in \mathcal{S}$, we denote by φ_s the representative functional of s in $\mathcal{H}$. The RKHS kernel associated with the couple $(\mathcal{S}, \mathcal{H})$ is the symmetrical application defined by:*

$$\begin{array}{rccc} l_{\mathcal{H}} & : & \mathcal{S}\times\mathcal{S} & \to & \mathbb{R} \\ & & (s, s') & \mapsto & \langle\varphi_s, \varphi_{s'}\rangle_{\mathcal{H}}. \end{array}$$

Unless otherwise stated, the kernels associated with the input $X_k, k = 1, \ldots, d$, will be denoted $l_k, k = 1, \ldots, d$, while the kernel associated with the output Y will be denoted l.

Reformulation of HSIC. The authors of Gretton et al. (2005a) show that the HSIC measure between an input X_k and the output Y can be expressed using the kernels l_k and l in a more convenient form:

$$\begin{aligned} \mathrm{HSIC}(X_k, Y) = {} & \mathbb{E}\left[l_k(X_k, X_k')l(Y, Y')\right] + \mathbb{E}\left[l_k\left(X_k, X_k'\right)\right]\mathbb{E}\left[l\left(Y, Y'\right)\right] \\ & - 2\mathbb{E}\left[\mathbb{E}\left[l_k\left(X_k, X_k'\right) \mid X_k\right]\mathbb{E}\left[l\left(Y, Y'\right) \mid Y\right]\right], \end{aligned} \tag{3.1}$$

where $(X'_1, \ldots, X'_d)$ is an independent and identically distributed copy of $(X_1, \ldots, X_d)$ and $Y' = \mathcal{M}(X'_1, \ldots, X'_d)$.

Among the most frequently used RKHS kernels in the literature, we can mention the linear, polynomial, Gaussian, Laplacian and Bergman kernels (Berlinet and Thomas-Agnan, 2011; Schölkopf et al., 2004).

Independence with universal kernels. The nullity of $\text{HSIC}(X_k, Y)$ is not always equivalent to the independence between X_k and Y: this characteristic depends on the RKHS associated to X_k and Y. In particular, if the kernels l_k and l belong to the specific class of *universal kernels* (Micchelli et al., 2006), the nullity of HSIC is equivalent to the independence. A kernel is said to be universal if the associated RKHS is dense in the space of continuous functions w.r.t. the infinity norm. However, the universality is a very strong assumption, especially on non-compact spaces. Let us mention as example the Gaussian kernel (the most commonly used for real variables) which is universal only on compact subsets $\mathcal{Z}$ of $\mathbb{R}^q$ (Steinwart, 2001). This kernel is defined for a pair of variables $(z, z') \in \mathbb{R}^q \times \mathbb{R}^q$ by:

$$k_\lambda(z, z') = exp\left(-\lambda \|z - z'\|_2^2\right), \tag{3.2}$$

where λ is a fixed positive real parameter, also called the *bandwidth parameter of the kernel* and $\|.\|_2$ is the Euclidean norm in $\mathbb{R}^q$.

First referred to as *probability-determining kernels* by Fukumizu et al. (2004), the notion of *characteristic kernels* (Fukumizu et al., 2008), which is a weaker assumption than universality, has been lately introduced. It has been proven that when the kernels l and l_k are characteristic then, $\text{HSIC}(X_k, Y) = 0$ if and only if (iff) X_k and Y are independent (Gretton, 2015; Szabó and Sriperumbudur, 2018). In particular, the Gaussian kernel defined in Formula (3.2) is characteristic on the entire $\mathbb{R}^q$ (Fukumizu et al., 2008).

Remark. ***Despite that theoretically*** $\text{HSIC}(X_k, Y) = 0$ ***is equivalent to the independence between*** X_k ***and*** Y***, a good choice of the kernel bandwidths is required in practice.*** *Indeed, as it will be further investigated in Chapter 5, a wise choice of these parameters guarantees a better behavior of* HSIC *estimators and better properties of the associated independence tests. Unfortunately, the best choice is unknown in practice, as it depends on the joint density of* (X_k, Y)*. For this, intrinsic characteristics of these random variables can be used. In particular, two main options are usually adopted in practice for the adjustment of* λ *(resp.* μ*) the bandwidth of the kernel associated to* X_k *(resp.* Y*) in Equation (3.2): whether the inverse of empirical variance of* X_k *(resp.* Y*), or the inverse of empirical median of* $\|X_k - X'_k\|_2^2$ *(resp.* $\|Y - Y'\|_2^2$*), where* X'_k *(resp.* Y'*) is an independent copy of* X_k *(resp.* Y*, cf. Gretton et al., 2008; Sugiyama and Suzuki, 2011; Zhang et al., 2012). To deal with this problem and avoid heuristic choices, some existing works such as Sugiyama and Yamada (2012) propose methods based on cross-validation to suitably select bandwidths. On our side, we chose to explore another solution based on an aggregated HSIC-based test. Thus, as it will be described in Chapter 5 and in Meynaoui et al. (2019), a well-chosen collection of single HSIC tests is aggregated through a single statistical test to improve the power.*

3.2.3 Use for first-level GSA

Several methods based on the use of HSIC measures have been developed for first-level GSA (GSA1)[3]. In this paragraph, we mention three possible approaches: sensitivity indices proposed by Da Veiga (2015), asymptotic tests of Gretton et al. (2008) and permutation (also referred to as bootstrap) tests initially introduced by De Lozzo and Marrel (2016b) and further investigated in Meynaoui et al. (2019).

HSIC-based sensitivity indices. These indices directly derived from HSIC measures, classify the input variables $X_1, .., X_d$ by order of influence on the output Y. They are defined for all $k \in \{1, \dots, d\}$ by:

$$\mathrm{R}^2_{\mathrm{HSIC},k} = \frac{\mathrm{HSIC}(X_k, Y)}{\sqrt{\mathrm{HSIC}(X_k, X_k)\,\mathrm{HSIC}(Y, Y)}}. \tag{3.3}$$

The normalization in (3.3) implies that $\mathrm{R}^2_{\mathrm{HSIC},k}$ is bounded and included in the range $[0, 1]$ which makes its interpretation easier. Other similar HSIC-based sensitivity indices are available in the literature. Examples include the distance correlation defined in Székely et al. (2009, Section 2), and based on the distance covariance. Note that, the distance correlation is also the $\mathrm{R}^2_{\mathrm{HSIC}}$ indice when the HSIC measure is the distance covariance[4]. We also mention the optimised criterion of (Blaschko and Gretton, 2009) used for taxonomy clustering and the kernel alignment defined in (Cortes et al., 2012).

In practice, $\mathrm{R}^2_{\mathrm{HSIC},k}$ can be estimated using a plug-in approach:

$$\widehat{\mathrm{R}}^2_{\mathrm{HSIC},k} = \frac{\widehat{\mathrm{HSIC}}(X_k, Y)}{\sqrt{\widehat{\mathrm{HSIC}}(X_k, X_k)\widehat{\mathrm{HSIC}}(Y, Y)}}. \tag{3.4}$$

These indices can be used to rank inputs by order of impact and perform GSA1.

Other approaches based on statistical HSIC-tests of independence are also possible to perform GSA1. According to the available number of simulations, these tests are mainly used under two versions: **asymptotic** and **non-asymptotic** tests. This point will be detailed later in a dedicated section of this chapter.

3.3 Statistical inference around HSIC measures

The first aim of this section is to present the usual estimators of HSIC measures along with their properties. Thereafter, we introduce a new method for estimating these measures using alternative samples, generated according to a law different from the prior law of inputs. The characteristics of the obtained estimators are demonstrated.

[3] It is recalled that GSA1 here refers to the classical sensitivity analysis of the simulator output as a function of the uncertain inputs when the probabilistic model of the inputs is known and fixed (cf. Sections 1.4 and 2.4).

[4] Indeed, there exists a particular choice of kernels for which the HSIC is the distance covariance, as shown in (Sejdinovic et al., 2013).

3.3.1 Statistical estimation under prior distributions

In this paragraph, we present HSIC estimators, as well as their characteristics. As a reminder, we assume that we have a n-sample of independent realizations $\left(\mathbf{X}^{(i)}, Y^{(i)}\right)_{1\leq i\leq n}$ of the inputs/output couple $(\mathbf{X}, Y)$ where $\mathbf{X} = (X_1, \ldots, X_d)$, according to the prior law of inputs $f(x_1, \ldots, x_d) = f_1(x_1) \times \ldots \times f_d(x_d)$.

Monte Carlo estimation. From Formula (3.1), authors of Gretton et al. (2005a) propose to estimate each $\mathrm{HSIC}(X_k, Y)$ by

$$\widehat{\mathrm{HSIC}}(X_k, Y) = \frac{1}{n^2}\sum_{1\leq i,j\leq n}(L_k)_{i,j}L_{i,j} + \frac{1}{n^4}\sum_{1\leq i,j,q,r\leq n}(L_k)_{i,j}L_{q,r} - \frac{2}{n^3}\sum_{1\leq i,j,r\leq n}(L_k)_{i,j}L_{j,r}, \quad (3.5)$$

where L_k and L are the matrices defined for all $i, j \in \{1, \ldots, n\}$ by $(L_k)_{i,j} = l_k(X_k^{(i)}, X_k^{(j)})$ and $(L)_{i,j} = l\left(Y^{(i)}, Y^{(j)}\right)$.

Remark. *The estimator in Equation (5.5) is part of a class of estimators called* V-statistics *(on behalf of Richard Von Mises), which are biased (but asymptotically unbiased) by contrast with unbiased estimators called* U-statistics *(U for unbiased) where diagonal terms are removed. Moreover, these two estimators as well as the bias term, all have the same computation cost (Song et al., 2012). Table 3.1 describes the characteristics of these two types of estimators.*

U-statistic estimators	V-statistic estimators
Without bias	Asymptotically unbiased
Variance of order $1/n$	Variance of order $1/n$
Approximation of the asymptotic law by a Gamma distribution under independence	Approximation of the asymptotic law by a Gamma distribution under independence
Practical to numerically implement but less practical than V-statistic estimators	Very practical to numerically implement
Computational complexity is n^2	Computational complexity is n^2

Table 3.1 – Comparison of the characteristics of U-statistical and V-statistical HSIC estimators.

These V-statistic estimators can also be written in the following more compact form (see Gretton et al., 2005a):

$$\widehat{\mathrm{HSIC}}(X_k, Y) = \frac{1}{n^2}Tr(L_kHLH), \quad (3.6)$$

where H is the matrix defined for all $i, j \in \{1, \ldots, n\}$ by $H_{i,j} = \delta_{i,j} - 1/n$, with $\delta_{i,j}$ the Kronecker symbol between i and j which is equal to 1 if $i = j$ and 0 otherwise.

Characteristics of HSIC estimators. Under the assumption of independence between X_k and Y and the assumption $l_k(x_k, x_k) = l(y, y) = 1$ (as in the case of Gaussian kernels), the estimator $\widehat{\mathrm{HSIC}}(X_k, Y)$ is asymptotically unbiased, its bias converges in $O(\frac{1}{n})$, while its variance converges to 0 in $O(\frac{1}{n^2})$. Moreover, the asymptotic distribution of $n \times \widehat{\mathrm{HSIC}}(X_k, Y)$ is an infinite sum of independent χ^2 random variables, which can be approximated by a Gamma law (Serfling, 2009) with shape and scale parameters, respectively denoted γ_k and β_k:

$$\gamma_k \simeq \frac{e_k^2}{v_k}, \quad \beta_k \simeq \frac{n.v_k}{e_k}$$

where e_k and v_k are respectively the expectation and the variance of $\widehat{\text{HSIC}}(X_k, Y)$, i.e.

$$e_k = \mathbb{E}\left[\widehat{\text{HSIC}}(X_k, Y)\right], \quad v_k = \text{Var}\left(\widehat{\text{HSIC}}(X_k, Y)\right).$$

The reader can refer to Gretton et al. (2008) and De Lozzo and Marrel (2016b) for more details on e_k and v_k and their estimation.

3.3.2 Statistical estimation under alternative distributions

In this part, we first demonstrate that HSIC measures presented in Section 3.2.1 can be expressed and then estimated using a sample generated from a probability distribution of inputs which is not their prior distribution. This sampling distribution will be called "alternative law" or "modified law". The characteristics of these new HSIC estimators (bias, variance, asymptotic law) will be demonstrated. **These estimators will then be used in the proposed methodology for 2^{nd}-level global sensitivity analysis in Section 4.2.**

3.3.2.1 Expression and estimation of HSIC from a sample drawn with alternative distributions

The purpose of this paragraph is to express HSIC measures between the inputs $X_1, ..., X_d$ and the output Y, using d random variables $\widetilde{X}_1$, ..., $\widetilde{X}_d$ whose laws are different from those of X_1, ..., X_d. We assume that their densities denoted $\widetilde{f}_1$, $\widetilde{f}_2$, ..., $\widetilde{f}_d$ have respectively the same supports as f_1, ..., f_d. We denote in the following by $\widetilde{\mathbf{X}}$ and $\widetilde{Y}$ respectively the random vector $\widetilde{\mathbf{X}} = (\widetilde{X}_1, ..., \widetilde{X}_d)$ and the associated output $\widetilde{Y} = \mathcal{M}(\widetilde{\mathbf{X}})$. Finally, we designate by $\widetilde{f}(x_1, .., x_d) = \widetilde{f}_1(x_1) \times \widetilde{f}_2(x_2) \times ... \times \widetilde{f}_d(x_d)$ the density of $\widetilde{\mathbf{X}}$.

Changing the probability laws in HSIC expression is based on a technique commonly used in the context of importance sampling (see e.g. Cannamela, 2007). This technique consists in expressing an expectation $\mathbb{E}[g(Z)]$, where Z is a random variable with density f_Z, by using a random variable $\widetilde{Z}$ with density $f_{\widetilde{Z}}$ whose support is the same as that of f_Z. This gives the following expression for $\mathbb{E}[g(Z)]$:

$$\mathbb{E}[g(Z)] = \int_{\text{Supp(Z)}} g(z)\, f_Z(z)\, dz = \int_{\text{Supp(Z)}} g(z)\, \frac{f_Z(z)}{f_{\widetilde{Z}}(z)}\, f_{\widetilde{Z}}(z)\, dz = \mathbb{E}\left[g(\widetilde{Z})\, \frac{f_Z(\widetilde{Z})}{f_{\widetilde{Z}}(\widetilde{Z})}\right], \quad (3.7)$$

where the notation Supp(Z) designates the support of Z.

The HSIC measures, formulated as a sum of expectations in Equation (3.1), can then be expressed under the density $f_{\widetilde{Z}}$ by adapting Equation (3.7) to more general forms of expectations. Hence, we obtain:

$$\text{HSIC}(X_k, Y) = H_k^1 + H_k^2 H_k^3 - 2H_k^4, \quad (3.8)$$

where $(H_k^l)_{1 \leq l \leq 4}$ are the real numbers defined by:

$$H_k^1 = \mathbb{E}\left[l_k(\widetilde{X}_k, \widetilde{X}_k') l(\widetilde{Y}, \widetilde{Y}') w(\widetilde{X}) w(\widetilde{X}')\right]; \; H_k^2 = \mathbb{E}\left[l_k(\widetilde{X}_k, \widetilde{X}_k') w(\widetilde{X}) w(\widetilde{X}')\right];$$

$$H_k^3 = \mathbb{E}\left[l(\widetilde{Y}, \widetilde{Y}') w(\widetilde{X}) w(\widetilde{X}')\right] \text{ and } H_k^4 = \mathbb{E}\left[\mathbb{E}\left[l_k(\widetilde{X}_k, \widetilde{X}_k') w(\widetilde{X}') \mid \widetilde{X}_k\right] \mathbb{E}\left[l(\widetilde{Y}, \widetilde{Y}') w(\widetilde{X}') \mid \widetilde{Y}\right] w(\widetilde{X})\right],$$

where $\widetilde{\mathbf{X}}'$ is an independent and identically distributed copy of $\widetilde{\mathbf{X}}$, $\widetilde{Y}' = \mathcal{M}(\widetilde{\mathbf{X}}')$ and $w = \dfrac{f}{\widetilde{f}}$.

Formula (3.8) shows that $\text{HSIC}(X_k, Y)$ can then be estimated using a sample generated from $\widetilde{f}$, provided that $\widetilde{f}$ has the same support than the original density f. Thus, if we consider a n-sample of independent realizations $\left(\widetilde{\mathbf{X}}^{(i)}, \widetilde{Y}^{(i)}\right)_{1\leq i\leq n}$, where $\widetilde{\mathbf{X}}$ is generated from $\widetilde{f}$ and $\widetilde{Y}^{(i)} = \mathcal{M}(\widetilde{X}^{(i)})$ for $i = 1, \ldots, n$, we propose the following V-statistic estimator of $\text{HSIC}(X_k, Y)$:

$$\widetilde{\text{HSIC}}(X_k, Y) = \widetilde{H}_k^1 + \widetilde{H}_k^2\widetilde{H}_k^3 - 2\widetilde{H}_k^4, \tag{3.9}$$

where $(\widetilde{H}_k^l)_{1\leq l\leq 4}$ are the V-statistics estimators of $(H_k^l)_{1\leq l\leq 4}$.

Proposition 3.2. *Similarly to Equation (3.6), this estimator can be rewritten as:*

$$\widetilde{\text{HSIC}}(X_k, Y) = \frac{1}{n^2} Tr\left(W\widetilde{L}_k W H_1 \widetilde{L} H_2\right), \tag{3.10}$$

where W, $\widetilde{L}_k$, $\widetilde{L}$, H_1 and H_2 are the matrices defined by:

$$\widetilde{L}_k = \left(l_k(\widetilde{X}_k^{(i)}; \widetilde{X}_k^{(j)})\right)_{1\leq i,j\leq n}; \quad \widetilde{L} = \left(l(\widetilde{Y}^{(i)}; \widetilde{Y}^{(j)})\right)_{1\leq i,j\leq n}; \quad W = Diag\left(w(\widetilde{X}^{(i)})\right)_{1\leq i\leq n};$$
$$H_1 = I_n - \frac{1}{n}UW; \quad H_2 = I_n - \frac{1}{n}WU;$$

with I_n is the identity matrix of size n and U the matrix filled with 1.

The proof of this proposition is detailed in Appendix 3.6.1.

Remark. *Similarly to Equation (3.4), the sensitivity index $\text{R}^2_{\text{HSIC},k}$ can also be estimated using the sample $\left(\widetilde{\mathbf{X}}^{(i)}, \widetilde{Y}^{(i)}\right)_{1\leq i\leq n}$ by:*

$$\widetilde{\text{R}}^2_{\text{HSIC},k} = \frac{\widetilde{\text{HSIC}}(X_k, Y)}{\sqrt{\widetilde{\text{HSIC}}(X_k, X_k)\widetilde{\text{HSIC}}(Y, Y)}}. \tag{3.11}$$

3.3.2.2 Statistical properties of HSIC alternative estimators

In this section, we show that the estimator $\widetilde{\text{HSIC}}(X_k, Y)$ has asymptotic properties similar to those of the estimator $\widehat{\text{HSIC}}(X_k, Y)$: same asymptotic behaviors of expectation and variance and same type of asymptotic distribution. The properties presented in the following are proved in Appendices 3.6.2, 3.6.3 and 3.6.4.

Proposition 3.3 (Bias)**.** *The estimator $\widetilde{\text{HSIC}}(X_k, Y)$ is asymptotically unbiased and its bias converges in $O(\frac{1}{n})$. Moreover, under the hypothesis of independence between X_k and Y and the assumption $l_k(x_k, x_k) = l(y, y) = 1$, its bias is:*

$$\mathbb{E}\left[\widetilde{\text{HSIC}}(X_k, Y)\right] - \text{HSIC}\left(X_k, Y\right) = \frac{2}{n}(E_\omega^k - E_{x_k,\omega})(E_\omega^{-k} - E_{y,\omega}) - \frac{1}{n}(E_\omega - E_{x_k})(E_\omega - E_y)$$
$$+ \frac{1}{n}E_\omega(E_\omega - 1) + O(\frac{1}{n^2}), \tag{3.12}$$

where

$$
\begin{aligned}
E_{\omega} &= \mathbb{E}\left[\omega^2(\widetilde{X})\right], & E_{x_k} &= \mathbb{E}\left[l_k\left(\widetilde{X}_k, \widetilde{X}'_k\right)\ \omega_k(\widetilde{X}_k)\,\omega_k(\widetilde{X}'_k)\right],\\
E_{y} &= \mathbb{E}\left[l\left(\widetilde{Y}, \widetilde{Y}'\right)\ \omega_{-k}(\widetilde{X}_{-k})\,\omega_{-k}(\widetilde{X}'_{-k})\right], & E_{x_k,\omega} &= \mathbb{E}\left[l_k\left(\widetilde{X}_k, \widetilde{X}'_k\right)\ \omega_k^2(\widetilde{X}_k)\,\omega_k(\widetilde{X}'_k)\right],\\
E_{y,\omega} &= \mathbb{E}\left[l\left(\widetilde{Y}, \widetilde{Y}'\right)\ \omega^2_{-k}(\widetilde{X}_{-k})\,\omega_{-k}(\widetilde{X}'_{-k})\right], & E_{\omega}^{k} &= \mathbb{E}\left[\omega_k^2(\widetilde{X}_k)\right],\\
E_{\omega}^{-k} &= \mathbb{E}\left[\omega^2_{-k}(\widetilde{X}_{-k})\right],
\end{aligned}
$$

ω and ω_k respectively denote the functions $\frac{f}{\widetilde{f}}$, $\frac{f_k}{\widetilde{f}_k}$, $\widetilde{X}_{-k}$ is the random vector extracted from $\widetilde{\mathbf{X}}$ by removing the k-th coordinate, $\widetilde{X}'_{-k}$ an independent and identically distributed copy of $\widetilde{X}_{-k}$ and $\omega_{-k}(x_{-k}) = \prod_{\substack{r=1\\r\neq k}}^{d} \omega_r(x_r)$ with x_{-k} the vector extracted from the vector $(x_1, .., x_d)$ by removing the k-th coordinate.

Under the independence assumption, an asymptotically unbiased estimator of the bias of $\widetilde{\mathrm{HSIC}}(X_k, Y)$ can be obtained by replacing each expectation in (3.12) by its empirical estimator.

Proposition 3.4 (Variance)**.** *Under the independence hypothesis between X_k and Y, the variance of $\widetilde{\mathrm{HSIC}}(X_k, Y)$ (denoted here ϑ_k) converges to 0 in $O(\frac{1}{n^2})$. More precisely, the variance ϑ_k can be expressed as:*

$$\vartheta_k = \frac{72(n-4)(n-5)}{n(n-1)(n-2)(n-3)}\ \mathbb{E}_{1,2}\left[\mathbb{E}_{3,4}[\widetilde{h}_{1,2,3,4}]^2\right] + O(\frac{1}{n^3}), \tag{3.13}$$

where $\widetilde{h}_{1,2,3,4} = \frac{1}{4!}\sum_{(t,u,v,s)}^{(1,2,3,4)}\left[(\widetilde{l}_k)_{t,u}\,\widetilde{l}_{t,u} + (\widetilde{l}_k)_{t,u}\,\widetilde{l}_{v,s} - 2\,(\widetilde{l}_k)_{t,u}\,\widetilde{l}_{t,v}\right]$, where the notation corresponds to the sum over all permutations (t,u,v,s) of $(1,2,3,4)$. The notations $(\widetilde{l}_k)_{p,q}$, $\widetilde{l}_{p,q}$ and w_p respectively denote $l_k\left(\widetilde{X}_k^{(p)}, \widetilde{X}_k^{(q)}\right)$, $l\left(\widetilde{Y}^{(p)}, \widetilde{Y}^{(q)}\right)$ and $\frac{f}{\widetilde{f}}(\widetilde{\mathbf{X}}^{(p)})$. Finally, the notation $\mathbb{E}_{p,q}$ means that the expectation is done by integrating only with respect to the variables indexed by p and q.

An estimator $\widetilde{\vartheta}_k$ of ϑ_k can be deduced from Equation (3.13):

$$\widetilde{\vartheta}_k = \frac{2(n-4)(n-5)}{n^3(n-1)(n-2)(n-3)}\mathbf{1}^T(\widetilde{B}\odot\widetilde{B})\mathbf{1}, \tag{3.14}$$

with $\odot$ is the Hadamard product and $\widetilde{B} = \sum_{r=1}^{4}\widetilde{B}_r$, where the matrices $\left(\widetilde{B}_r\right)_{1\leq r\leq 4}$ are defined for all $(i,j)\in\{1,\ldots,n\}^2$ by:

$$(\widetilde{B}_1)_{i,j} = (\widetilde{l}_k)_{i,j} w_i w_j \left(\widetilde{l}_{i,j} - \widetilde{l}_{i,.} - \widetilde{l}_{j,.} + \widetilde{l}\right), \qquad (\widetilde{B}_2)_{i,j} = -(\widetilde{l}_k)_{i,.} w_i \left(\widetilde{l}_{i,j} w_j - \widetilde{l}_{i,.} - \widetilde{l}_{j,.} w_j + \widetilde{l}\right),$$
$$(\widetilde{B}_3)_{i,j} = -(\widetilde{l}_k)_{.,j} w_j \left(\widetilde{l}_{i,j} w_i - \widetilde{l}_{i,.} w_i - \widetilde{l}_{j,.} + \widetilde{l}\right), \quad (\widetilde{B}_4)_{i,j} = (\widetilde{l}_k) \left(\widetilde{l}_{i,j} w_i w_j - \widetilde{l}_{i,.} w_i - \widetilde{l}_{j,.} w_j + \widetilde{l}\right),$$

where $w_i = W_{i,i}$ and the terms $(\widetilde{l}_k)_{i,.}$, $(\widetilde{l}_k)_{.,j}$, $(\widetilde{l}_k)$, $\widetilde{l}_{i,.}$, $\widetilde{l}_{.,j}$ and $\widetilde{l}$ are the empirical means (denoted with a bar above):

$$(\widetilde{l}_k)_{i,.} = \overline{L_k\left(\widetilde{X}_k^{(i)}, \widetilde{X}_k\right)\omega_k(\widetilde{X}_k)}, \qquad \widetilde{l}_{i,.} = \overline{L\left(\widetilde{Y}^{(i)}, \widetilde{Y}\right)\omega_{-k}(\widetilde{X}_{-k})},$$
$$(\widetilde{l}_k) = \overline{L_k\left(\widetilde{X}_k', \widetilde{X}_k\right)\omega_k(\widetilde{X}_k)}, \qquad \widetilde{l} = \overline{L\left(\widetilde{Y}', \widetilde{Y}\right)\omega_{-k}(\widetilde{X}_{-k})}.$$

Theorem 3.1 (Asymptotic law)**.** *In a similar way as* $n \times \widehat{\mathrm{HSIC}}(X_k, Y)$*, one can prove that the asymptotic distribution of* $n \times \widetilde{\mathrm{HSIC}}(X_k, Y)$ *can be approximated by a Gamma distribution, whose parameters* $\widetilde{\gamma}_k$ *and* $\widetilde{\beta}_k$ *are given by* $\widetilde{\gamma}_k = \dfrac{\varepsilon_k^2}{\vartheta_k}$ *and* $\widetilde{\beta}_k = \dfrac{n\vartheta_k}{\varepsilon_k}$*, where* ε_k *and* ϑ_k *are the expectation and variance of* $\widetilde{\mathrm{HSIC}}(X_k, Y)$*, i.e.* $\varepsilon_k = \mathbb{E}\left[\widetilde{\mathrm{HSIC}}(X_k, Y)\right]$ *and* $\vartheta_k = \mathrm{Var}\left(\widetilde{\mathrm{HSIC}}(X_k, Y)\right)$.

In practice, these parameters are respectively estimated by the empirical estimator for ε_k and the estimator given by Equation (3.13) for ϑ_k.

Remark. *From a practical point of view, the greater* $\mathrm{Var}\left(\omega_k(\widetilde{X}_k)\right)$*, the greater the number of simulations required to accurately estimate* $\mathrm{HSIC}(X_k, Y)$*. It is therefore highly recommended to check that* $\mathrm{Var}\left(\omega_k(\widetilde{X}_k)\right)$ *is finite. For instance, in the case of densities with compact supports, it is enough to check that* ω_k *is finite on its support.*

3.3.2.3 Illustration on an analytical example

In this paragraph, we illustrate via a numerical application the behavior and the convergence of the modified estimators $\widetilde{\mathrm{HSIC}}$, according to the size of the inputs/output sample. For that, we consider the analytic model $\mathcal{M}$ inspired from Ishigami's model (Ishigami and Homma, 1990) and defined by:

$$\mathcal{M}(X_1, X_2, X_3) = \sin(X_1) + 1.8\ \sin^2(X_2) + 0.5\ X_3^4\ \sin(X_1), \tag{3.15}$$

where the inputs X_1, X_2 et X_3 are assumed to be independent and follow a triangular distribution on $[0, 1]$ with a mode equal to $\dfrac{1}{2}$. We denote by Y the output variable $Y = \mathcal{M}(X)$.

We consider standardized Gaussian kernel HSIC measures (see remark 3.2.2) between each input X_k, $k = 1 \ldots 3$ and the output Y. The objective is to estimate these measures from samples where the inputs are independent and identically distributed but generated from a uniform distribution on $[0, 1]$ (here the modified law). For this, we consider Monte Carlo samples of size $n = 100$ to $n = 1500$ and for each sample size, the estimation process is repeated 200 times, with independent random samples.

Figure (3.1) presents as a boxplot the convergence graphs of the estimators $\widetilde{\text{HSIC}}(X_k, Y)$. Results for the estimator $\widehat{\text{HSIC}}(X_k, Y)$ computed with samples generated from the original law (namely triangular) are also given. Theoretical values of HSIC are represented in red dotted lines. We observe that for small sample sizes ($n < 500$), the modified estimators $\widetilde{\text{HSIC}}(X_k, Y)$ have more bias and variance than the estimators $\widehat{\text{HSIC}}(X_k, Y)$. But, from size $n = 700$, both estimators have similar behaviors.

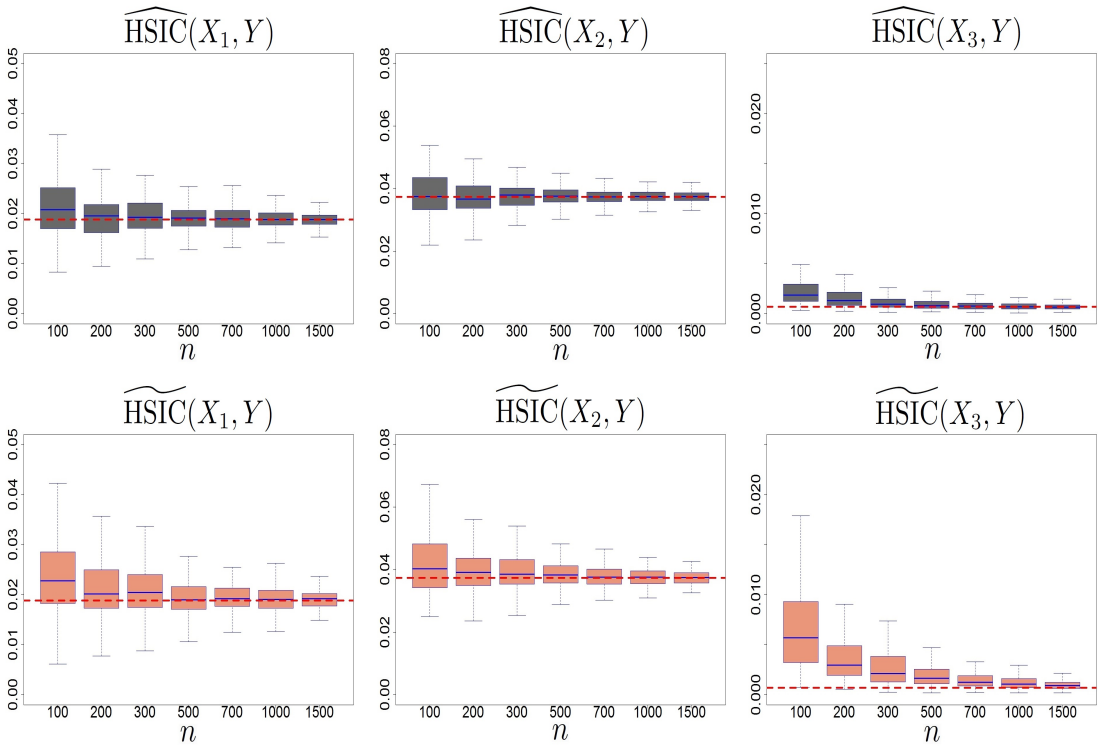

Figure 3.1 – Convergence plots of the estimators $\widehat{\text{HSIC}}(X_k, Y)$ and $\widetilde{\text{HSIC}}(X_k, Y)$ for Ishigami function, according to the sample size n. Theoretical values are represented in red dashed lines.

On this same example, we are now interested in the classification of the input variables by order of influence. For this, the sensitivity index $\text{R}^2_{\text{HSIC},k}$ $(k = 1, \ldots, d)$ is computed from both $\widetilde{\text{HSIC}}(X_k, Y)$ and $\widehat{\text{HSIC}}(X_k, Y)$ using Equations (3.11) and (3.4). The obtained sensitivity indices are respectively denoted $\widetilde{\text{R}}^2_{\text{HSIC}}$ and $\widehat{\text{R}}^2_{\text{HSIC}}$. In either case, the inputs are ordered by decreasing estimated values of R^2_{HSIC}. For each sample size, the rates of good ranking of inputs given by $\widetilde{\text{R}}^2_{\text{HSIC}}$ and $\widehat{\text{R}}^2_{\text{HSIC}}$ estimators are also computed. **Results are given in Table (3.2). As one might expect, prior estimators are more efficient for ranking the inputs. However, this example** illustrates that, even for small sample sizes (*e.g* $n = 200$), modified estimators $\widetilde{\text{R}}^2_{\text{HSIC}}$ have good ranking ability. Once again, for medium and high sample sizes ($n \geq 500$ here), these two types of estimators have the same ability to rank the inputs.

	$n = 100$	$n = 200$	$n = 300$	$n \geqslant 500$
$\widetilde{\mathrm{R}}^2_{\mathrm{HSIC}}$	88%	93.5%	97%	100%
$\widehat{\mathrm{R}}^2_{\mathrm{HSIC}}$	100%	100%	100%	100%

Table 3.2 – Good ranking rates of input variables using modified estimators $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC}}$ and prior estimators $\widehat{\mathrm{R}}^2_{\mathrm{HSIC}}$ for the Ishigami function, for different sample sizes n.

3.4 Statistical tests of independence based on HSIC

The purpose of this section is to present and discuss different options for independence tests using HSIC (and consequently, more carefully screen inputs). For this, we start by reminding the general principle of statistical tests and the characteristics to assess their quality. We then present a review of the most frequently known independence tests and the reasons for considering those based on HSIC. Thereafter, we present the two main approaches for HSIC tests, namely the *asymptotic* and the *non-asymptotic* tests. Finally, a new non-asymptotic HSIC-based test with more theoretical guarantees than those currently used, is proposed.

3.4.1 Review on non-parametric tests of independence

Let us first introduce some notations and assumptions. We consider here a numerical simulator of d scalar inputs which are not necessarily independent, and we assume that we have q scalar outputs which are viewed as a vector of q components denoted $Y = (Y_1, \ldots, Y_q)$. We also denote by $X = (X_1, \ldots, X_p)$, a group of p inputs viewed as a single vector. The couple (X, Y) is considered as continuous with a joint density $f_{X,Y}$. Moreover, the marginal densities of X and Y are respectively denoted f_X and f_Y. We also denote by $f_X \otimes f_Y$, the product of the marginal densities f_X and f_Y defined as follows:

$$f_X \otimes f_Y : (x, y) \in \mathbb{R}^p \times \mathbb{R}^q \mapsto f_X(x) f_Y(y).$$

The density $f_{X,Y}$ is assumed to be unknown as well as the marginals f_X and f_Y. We address here the question of testing the null hypothesis ($\mathcal{H}_0$): "X and Y are independent" against the alternative hypothesis ($\mathcal{H}_1$): "X and Y are dependent". Testing independence between X and Y is then equivalent to test ($\mathcal{H}_0$) : "$f_{X,Y} = f_X \otimes f_Y$" against ($\mathcal{H}_1$) : "$f_{X,Y} \neq f_X \otimes f_Y$". Moreover, we assume that we have an n-sample $(X^{(1)}, Y^{(1)}), \ldots, (X^{(n)}, Y^{(n)})$ of i.i.d. random variables with common density $f_{X,Y}$. The probability measure associated to this n-sample is denoted $P_{f_{X,Y}}$. By analogy, $P_{f_X \otimes f_Y}$ designates the probability measure associated to this n-sample with common density $f_X \otimes f_Y$. The densities $f_{X,Y}$, f_X and f_Y are assumed to be bounded and M_f denotes the maximum of their infinity norms[5]: $M_f = \max\{\|f_{X,Y}\|_\infty, \|f_X\|_\infty, \|f_Y\|_\infty\}$.

3.4.1.1 Generalities on statistical tests of independence

Basic definition. A test of independence Δ is a statistical procedure, testing two hypothesis. The first one, usually called the ***null hypothesis*** and denoted ($\mathcal{H}_0$), is defined as

($\mathcal{H}_0$): "X and Y are independent".

[5]The infinity norm of a bounded function g on a set S is defined as $\|g\|_\infty = \sup\{\, |g(t)| : t \in S \,\}$.

While the second one, called the ***alternative hypothesis*** and denoted $(\mathcal{H}_1)$, is defined as

$$(\mathcal{H}_1)\text{: “}X \text{ and } Y \text{ are dependent”.}$$

To achieve this, a ***test statistic*** $\hat{S}_\Delta$ is computed from the set of observations, which means that

$$\hat{S}_\Delta = \hat{S}_\Delta\Big((X^{(1)}, Y^{(1)}), \ldots, (X^{(n)}, Y^{(n)})\Big).$$

Depending on the value of the test statistic $\hat{S}_\Delta$, we reject or not the null hypothesis $(\mathcal{H}_0)$. For example, if $\hat{S}_\Delta$ is an estimator of a dependence measure between X and Y (whose nullity is equivalent to the independence between these two random variables), then the test Δ rejects $(\mathcal{H}_0)$ for large values of $\hat{S}_\Delta$. Therefore, the critical value from which the test rejects independence depends on the chosen test statistic. In addition, the test function Δ can be defined as

$$\Delta = \mathbb{1}_{\hat{S}_\Delta > C_{\hat{S}_\Delta}},$$

where $C_{\hat{S}_\Delta}$ is the critical value associated to test statistic $\hat{S}_\Delta$. The null hypothesis is then rejected if and only if $\Delta = 1$.

Errors of first and second-kind. The quality of a statistical test is naturally determined by the risks of misjudging dependence and independence between X and Y. The ***first-kind error*** of the test Δ is the probability of $(\mathcal{H}_0)$ being rejected, while X and Y are independent. The ***level*** of the test Δ is an upper bound of the first-kind error. Similarly, the ***second-kind error*** is the probability of $(\mathcal{H}_0)$ not being rejected, while X and Y are dependent. An independence test having a small second-kind error (lower than a chosen threshold β) is called $(1-\beta)$***-powerful*** (see Lehmann and Romano, 2006; Siegmund, 2013 for more details on theory of statistical tests). Thus, a “good” test of independence should ideally have small first and second-kind errors. Note that, 1st and 2nd kind errors can't be minimized at once, when either of them is reduced the other is raised. In practice, the tests of independence are built to have a small first-kind error(generally set at 5% or 10%). All the efforts are then concentrated to minimize the second-kind error.

Uniform separation rate. In a non-asymptotic framework, the uniform separation rate (Baraud, 2002) is a good indicator for the quality of a given statistical test of independence. As mentioned above, the first objective is to control the first kind error and this property is usually true by construction of the test. More precisely, if α in $(0,1)$ is the level of the test Δ, meaning that

$$P_{f_X \otimes f_Y}(\Delta = 1) \leq \alpha,$$

then the first-kind error is controlled by α. In other words, if X and Y are independent, the probability that the test Δ rejects the independence hypothesis is less than α (in practice the value of α is chosen at $\alpha = 5\%$ or $\alpha = 10\%$). The second objective is to control the 2nd-kind error. However, we already know that if X and Y are weakly dependent in the sense that $\|f_{X,Y} - f_X \otimes f_Y\|_2$ is very small, it is difficult to control the 2nd-kind error. This leads to the definition of the **uniform separation rate associated to the test Δ, which is the smallest distance $\|f_{X,Y} - f_X \otimes f_Y\|_2$ between $f_{X,Y}$ and $f_X \otimes f_Y$ from which dependence is detectable using the test Δ.** More precisely, the uniform separation rate $\rho(\Delta, \mathcal{C}_\delta, \beta)$ of the test Δ, over a class $\mathcal{C}_\delta$ of alternatives $f_{X,Y}$ such that $f_{X,Y} - f_X \otimes f_Y$ satisfies smoothness assumptions, w.r.t. the $\mathbb{L}_2$-norm, is defined for all β in $(0,1)$ by

$$\rho(\Delta, \mathcal{C}_\delta, \beta) = \inf\left\{\rho > 0, \sup_{f_{X,Y} \in \mathcal{F}_\rho(\mathcal{C}_\delta)} \mathbb{P}_{f_{X,Y}}(\Delta = 0) \leq \beta\right\}, \tag{3.16}$$

where $\mathcal{F}_\rho(\mathcal{C}_\delta) = \left\{f_{X,Y}, f_{X,Y} - f_X \otimes f_Y \in \mathcal{C}_\delta, \|f_{X,Y} - f_X \otimes f_Y\|_2 > \rho\right\}$. To avoid any misunderstanding, let us highlight that f_X and f_Y always denote the marginals of the function $f_{X,Y}$.

This definition extends to the non-asymptotic framework, the notion of *critical radius* introduced and studied for several examples in a series of Ingster papers (see e.g. Ingster, 1993a,b). A test of level α having the optimal performances, should then have the smallest possible uniform separation rate (up to a multiplicative constant) over $\mathcal{C}_\delta$. To quantify this, let us introduce, as in Baraud (2002), the *non-asymptotic minimax rate of testing*, defined by

$$\rho\left(\mathcal{C}_\delta, \alpha, \beta\right) = \inf_{\Delta_\alpha} \rho\left(\Delta_\alpha, \mathcal{C}_\delta, \beta\right), \tag{3.17}$$

where the infimum is taken over all α-level tests of $(\mathcal{H}_0)$ against $(\mathcal{H}_1)$. If the uniform separation rate of a test is upper-bounded up to a constant by the non-asymptotic minimax rate of testing, then this test is said to be *optimal in the minimax sense*. The problem of non-asymptotic minimax rate of testing based on HSIC measures will be more closely explored in Chapter 5.

3.4.1.2 Classical non-parametric tests of independence

To test independence between X and Y, many approaches have been explored in the last few decades. Among them, Hoeffding (1948b) proposes an independence test based on the difference between the distribution function of (X, Y) and the product of the marginal distribution functions. This test has good properties in the asymptotic framework, it is consistent and distribution-free under the null hypothesis. But, it is only designated to univariate continuous random variables. The authors of Bergsma and Dassios (2014) propose an improvement of Hoeffding's test, which is applicable to discrete, continuous or mixture of discrete and continuous distributions. Lately, Weihs et al. (2018) propose to extend Hoeffding's test to the case of multivariate random variables.

More efficient alternatives to test independence between X and Y are based on RKHS spaces. These tests can be implemented without any density estimation or high dimensional integration. As a result, they are more resistant to dimensionality. A first RKHS-based asymptotic independence test is performed using general large deviation inequalities (Gretton et al., 2005a). A more optimal asymptotic independence test based on a Gamma approximation of the distributions of HSIC estimators under $(\mathcal{H}_0)$ is proposed by Gretton et al. (2008). This last statistical test remains by far the most commonly used kernel-based test for independence. A generalization of this test for the joint and mutual independence of several random variables is presented in Pfister et al. (2018). We also mention the RKHS-based test of Póczos et al. (2012), based on a new dependence measure called Copula-based kernel dependency measure. Yet, this test is more conservative than the test of Gretton et al. (2008), since it is based on large deviation inequalities rather than the asymptotic distributions of the estimators under $(\mathcal{H}_0)$. Another worth-mentioning RKHS test is built from the distance covariance[6] (Székely and Rizzo, 2013; Yao et al., 2018). We also mention the statistical test of independence based on the kernel mutual information recently proposed by Berrett and Samworth (2017). This new statistical test seems to achieve comparable results with the classical tests based on HSIC.

One of the objectives of this thesis is to improve the performance of these HSIC statistical tests of independence for screening purposes.

[6] We recall that the distance covariance can be seen as an HSIC measure with specific kernel choices.

3.4.2 Existing HSIC-based statistical tests of independence

By now, we present the two manners to use HSIC measures to test the independence between X and Y: the asymptotic and the non-asymptotic approaches. The asymptotic approach is built upon the assumption that the law of HSIC estimator under the independence can be approached by the asymptotic Gamma distribution introduced in Section 3.3.1. Otherwise, if this assumption is not justified, it is necessary to use non-asymptotic tests, based on permutations.

A first theoretical test. Since the nullity of HSIC is equivalent to the independence between X and Y, testing independence is equivalent to test

$$(\mathcal{H}_0) : \mathrm{HSIC}(X,Y) \;=\; 0 \qquad \text{against} \qquad (\mathcal{H}_1) : \mathrm{HSIC}(X,Y) \;>\; 0.$$

The statistic estimator $\widehat{\mathrm{HSIC}}(X,Y)$ is then a natural choice to test independence between X and Y. The corresponding test rejects independence if $\widehat{\mathrm{HSIC}}(X,Y)$ is significantly large. Specifically, for $\alpha \in (0,1)$, we consider the statistical test which rejects $(\mathcal{H}_0)$ if

$$\widehat{\mathrm{HSIC}}(X,Y) > q_{1-\alpha},$$

where $q_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of $\widehat{\mathrm{HSIC}}(X,Y)$ under $(\mathcal{H}_0)$. The associated test function is defined by

$$\Delta_\alpha = \mathbb{1}_{\widehat{\mathrm{HSIC}}(X,Y) > q_{1-\alpha}}. \tag{3.18}$$

Then, the null hypothesis is rejected if and only if $\Delta_\alpha = 1$. By definition of the quantile, this theoretical test is of non-asymptotic level α, that is if $f_{X,Y} = f_X \otimes f_Y$,

$$P_{f_{X,Y}}\left(\Delta_\alpha = 1\right) \leq \alpha.$$

Note that analytical computation of the quantile $q_{1-\alpha}$ is not possible since its value depends on the unknown marginals f_X and f_Y of the couple (X,Y). The estimation of $q_{1-\alpha}$ therefore depends on the framework in which we are placed. Note also that the theoretical test can be defined in an equivalent way using the p-value[7]. To achieve this, the independence is rejected when the p-value is lower than α.

Asymptotic tests. In the asymptotic framework, the estimation of the quantile $q_{1-\alpha}$ is performed using the Gamma approximation of the distribution of $\widehat{\mathrm{HSIC}}(X,Y)$ under $(\mathcal{H}_0)$ as shown in Gretton et al. (2008). Similarly, the p-value (denoted P) of the test can be estimated in the asymptotic framework using the Gamma approximation (denoted G) as

$$\mathrm{P} \simeq 1 - F_G\left(n \times \widehat{\mathrm{HSIC}}(X,Y)_{obs}\right), \tag{3.19}$$

where F_G is the cumulative distribution function of G and $\widehat{\mathrm{HSIC}}(X,Y)_{obs}$ is the observed value of the random variable $\widehat{\mathrm{HSIC}}(X,Y)$.

Permutation tests. Outside the asymptotic framework, i.e. for small sample size n, the quantile $q_{1-\alpha}$ can be estimated using permutation-based technique as proposed by De Lozzo

[7]The p-value of the test is the probability that, under $(\mathcal{H}_0)$, the test statistic (in this case, $\widehat{\mathrm{HSIC}}(X,Y)$) is greater than or equal to the value observed on the data.

and Marrel (2016b). We denote $\mathbb{Z}_n = (X^{(i)}, Y^{(i)})_{1\leq i\leq n}$ the original sample and $\widehat{\mathrm{HSIC}}(\mathbb{Z}_n)$ the associated estimator. We consider B independent and uniformly distributed random permutations of $\{1, ..., n\}$, denoted $\tau_1, ..., \tau_B$, independent of $\mathbb{Z}_n$. Then, we define for each permutation τ_b the corresponding permuted sample $\mathbb{Z}_n^{\tau_b} = (X^{(i)}, Y^{(\tau_b(i))})_{1\leq i\leq n}$ and we estimate the following permuted test statistics

$$\widehat{H}^{\star b} = \widehat{\mathrm{HSIC}}(\mathbb{Z}_n^{\tau_b}),\ b = 1, \ldots, B.$$

Under $(\mathcal{H}_0)$, that is if $f_{X,Y} = f_X \otimes f_Y$, each permuted sample $\mathbb{Z}_n^{\tau_b}$ has the same distribution than the original sample $\mathbb{Z}_n$. Hence, the random variables $\widehat{H}^{\star b}, b = 1, \ldots, B$, have the same distribution as $\widehat{\mathrm{HSIC}}(X, Y)$. In order to estimate the quantile $q_{1-\alpha}$, we denote

$$\widehat{H}^{\star(1)} \leq \widehat{H}^{\star(2)} \leq \ldots \leq \widehat{H}^{\star(B)}$$

the order statistics associated to $\widehat{H}^{\star 1}, \ldots, \widehat{H}^{\star B}$. Therefore, the quantile $q_{1-\alpha}$ can be estimated using the permutation approach by $\widehat{q}_{1-\alpha}$ defined as

$$\widehat{q}_{1-\alpha} = \widehat{H}^{\star(\lceil B(1-\alpha)\rceil)}, \tag{3.20}$$

where $\lceil . \rceil$ denotes the ceiling function. The p-value (denoted p) of the permutation test can also be computed by:

$$\mathrm{p} = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}_{\widehat{H}^{\star b} > \widehat{\mathrm{HSIC}}(X,Y)}. \tag{3.21}$$

Although it is heuristically known that the level of the majority of permutation tests can be controlled, there is no theoretical guarantee in the specific case of HSIC tests. In the next section, we propose a modification of the "classical" version, so that we can show that its level can be controlled by an arbitrarily small α.

Other HSIC-based tests. Further HSIC-based independence tests are worth mentioning. More particularly, the wild bootstrap test of Chwialkowski et al. (2014) based on an external randomization of X and Y samples. This is performed by using random processes associated to X and Y which have some theoretical properties. The resulting test has asymptotically the good level. We also mention the asymptotic approach of Zhang et al. (2018). This last approach is spectral i.e. the asymptotic distribution of the HSIC estimator under $\mathcal{H}_0$ is approximated by estimating the eigenvalues of an operator called the "integral operator".

3.4.3 New version of non-asymptotic HSIC-based tests of independence

In order to control the level of HSIC tests in the non-asymptotic framework, we propose here a modification of the previous permutation test. To do so, we keep the same notations. If we denote by F the cumulative distribution function of $\widehat{\mathrm{HSIC}}(X, Y)$ under $f_X \otimes f_Y$, then F can be estimated by $\widehat{F}$ defined for all $t \in \mathbb{R}$ by

$$\widehat{F}(t) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}_{\widehat{H}^{\star b} \leq t}. \tag{3.22}$$

Yet, instead of considering the generalized inverse of $\widehat{F}$ to approximate the quantile, we apply a trick, based on Romano and Wolf (2005, Lemma 1), **which consists in adding the original**

sample to the Monte Carlo sample in order to obtain a test of non-asymptotic level α. For this, we denote

$$\widehat{H}^{\star B+1} = \widehat{\mathrm{HSIC}}(X,Y), \quad \text{and} \quad \widehat{H}^{\star(1)} \leq \widehat{H}^{\star(2)} \leq \ldots \leq \widehat{H}^{\star(B+1)}$$

the order statistic. Then, the permuted quantile with Monte Carlo approximation $\widehat{q}_{1-\alpha}$ is thus defined as

$$\widehat{q}_{1-\alpha} = \widehat{H}^{\star(\lceil (B+1)(1-\alpha) \rceil)}, \tag{3.23}$$

where $\lceil . \rceil$ denotes the ceiling function.

The permuted test with Monte Carlo approximation $\widehat{\Delta}_\alpha$ is then defined as

$$\widehat{\Delta}_\alpha = \mathbb{1}_{\widehat{\mathrm{HSIC}}(X,Y) > \widehat{q}_{1-\alpha}}. \tag{3.24}$$

Proposition 3.5. *Let* α *be in* $]0,1[$ *and* $\widehat{\Delta}_\alpha$ *the test defined by Equation* (3.24)*. Then, under* $(\mathcal{H}_0)$*, that is if* $f_{X,Y} = f_X \otimes f_Y$*, we have*

$$P_{f_{X,Y}}\left(\widehat{\Delta}_\alpha = 1\right) \leq \alpha, \tag{3.25}$$

this permuted test with Monte Carlo approximation is thus of prescribed non-asymptotic level α*.*

3.5 Synthesis

In this chapter, we detailed the Global Sensitivity Analysis (GSA) based on HSIC (Hilbert-Schmidt Independence Criterion). This choice is motivated by the advantages of HSIC comparing to other dependence measures: broad spectrum of captured dependence, good proprieties of estimators, easy to implement, low cost of estimation, etc. We started by reviewing basic background information on HSIC measures: principle and definition, which extends the well-known notion of linear covariance by projecting the code inputs and output into some Hilbert spaces having better characteristics. We also recalled sufficient conditions so that HSIC be independence-characterizing. The main approaches proposed in the literature for using HSIC for GSA were then presented: *normalized sensitivity indices*, *asymptotic* and *non-asymptotic tests of independence*. As a follow-up, the "usual" HSIC statistical estimators and their proprieties were described. The link between these estimators and the HSIC-based statistical tests was also highlighted. Thereafter, we presented a new method to estimate these measures using an "alternative sample" (with a different distribution than the prior one). The characteristics of these new estimators and the associated statistical tests were demonstrated. In the last section, we focus our attention to the HSIC-based statistical tests of independence. Indicators to assess the quality of the tests were presented, in particular the notion of uniform separation rate. The formulation of asymptotic and non-asymptotic versions of HSIC tests were subsequently detailed. Finally, an improvement in the commonly used non-asymptotic tests was proposed, this new version having more theoretical guarantees for testing the independence.

These different elements will be used later in the next two methodological chapters. In particular, **the statistical estimation under an alternative law will be helpful in Chapter 4 to build the methodology for GSA with second level uncertainty. The developments around HSIC tests, on the other hand, will be useful for our methodology of aggregated tests proposed in Chapter 5, for screening purpose.**

3.6 Proofs

3.6.1 Proof of Proposition 3.2

In this annex, we prove that:

$$\widetilde{\mathrm{HSIC}}(X_k, Y) = \frac{1}{n^2} \operatorname{Tr}\left(W\widetilde{L}_k W H_1 \widetilde{L} H_2\right).$$

Firstly, we evaluate the matrix $W\widetilde{L}_k W H_1 \widetilde{L} H_2$ coefficients before computing its trace. The matrix W being diagonal, we write for $i, j \in \{1, \ldots, n\}$:

$$(W\widetilde{L}_k W)_{i,j} = (\widetilde{L}_k)_{i,j} W_{i,i} W_{j,j}.$$

The coefficient of the matrix $W\widetilde{L}_k W H_1$ indexed by i and j can therefore be computed:

$$\begin{aligned}
(W\widetilde{L}_k W H_1)_{i,j} &= \sum_{r=1}^{n} (\widetilde{L}_k)_{i,r} W_{i,i} W_{r,r} (H_1)_{r,j} \\
&= \sum_{r=1}^{n} (\widetilde{L}_k)_{i,r} W_{i,i} W_{r,r} (\delta_{r,j} - \frac{1}{n} W_{j,j}) \\
&= (\widetilde{L}_k)_{i,j} W_{i,i} W_{j,j} - \frac{1}{n} \sum_{r=1}^{n} (\widetilde{L}_k)_{i,r} W_{i,i} W_{r,r} W_{j,j}.
\end{aligned}$$

Subsequently, the matrix $W\widetilde{L}_k W H_1 \widetilde{L}$ coefficients are obtained:

$$\begin{aligned}
(W\widetilde{L}_k W H_1 \widetilde{L})_{i,j} &= \sum_{r=1}^{n} (W\widetilde{L}_k W H_1)_{i,r} \widetilde{L}_{r,j} \\
&= \sum_{r=1}^{n} \left((\widetilde{L}_k)_{i,r} W_{i,i} W_{r,r} - \frac{1}{n} \sum_{s=1}^{n} (\widetilde{L}_k)_{i,s} W_{i,i} W_{s,s} W_{r,r} \right) \widetilde{L}_{r,j} \\
&= \sum_{r=1}^{n} (\widetilde{L}_k)_{i,r} \widetilde{L}_{r,j} W_{i,i} W_{r,r} - \frac{1}{n} \sum_{s=1}^{n} (\widetilde{L}_k)_{i,s} W_{i,i} W_{s,s} \sum_{r=1}^{n} \widetilde{L}_{r,j} W_{r,r}.
\end{aligned}$$

Finally,

$$
\begin{aligned}
(W\widetilde{L}_k W H_1 \widetilde{L} H_2)_{i,j} &= \sum_{r=1}^{n} (W\widetilde{L}_k W H_1 \widetilde{L})_{i,r} (H_2)_{r,j} \\
&= \sum_{r=1}^{n} (W\widetilde{L}_k W H_1 \widetilde{L})_{i,r} (\delta_{r,j} - \frac{1}{n} W_{r,r}) \\
&= (W\widetilde{L}_k W H_1 \widetilde{L})_{i,j} - \frac{1}{n} \sum_{r=1}^{n} (W\widetilde{L}_k W H_1 \widetilde{L})_{i,r} W_{r,r} \\
&= \sum_{r=1}^{n} (\widetilde{L}_k)_{i,r} \widetilde{L}_{r,j} W_{i,i} W_{r,r} - \frac{1}{n} \sum_{1 \leq r,s \leq n} (\widetilde{L}_k)_{i,s} \widetilde{L}_{r,j} W_{i,i} W_{s,s} W_{r,r} \\
&- \frac{1}{n} \sum_{r=1}^{n} \left(\sum_{s=1}^{n} (\widetilde{L}_k)_{i,s} \widetilde{L}_{s,r} W_{i,i} W_{s,s} - \frac{1}{n} \sum_{1 \leq p,q \leq n} (\widetilde{L}_k)_{i,q} \widetilde{L}_{p,r} W_{i,i} W_{q,q} W_{p,p} \right) W_{r,r} \\
&= \sum_{r=1}^{n} (\widetilde{L}_k)_{i,r} \widetilde{L}_{r,j} W_{i,i} W_{r,r} - \frac{1}{n} \sum_{1 \leq r,s \leq n} (\widetilde{L}_k)_{i,s} \widetilde{L}_{r,j} W_{i,i} W_{s,s} W_{r,r} \\
&- \frac{1}{n} \sum_{1 \leq r,s \leq n} (\widetilde{L}_k)_{i,s} \widetilde{L}_{s,r} W_{i,i} W_{s,s} W_{r,r} + \frac{1}{n^2} \sum_{1 \leq r,p,q \leq n} (\widetilde{L}_k)_{i,q} \widetilde{L}_{p,r} W_{i,i} W_{q,q} W_{p,p} W_{r,r}.
\end{aligned}
$$

Summing up the matrix $W\widetilde{L}_k W H_1 \widetilde{L} H_2$ diagonal terms, then dividing by n^2 gives:

$$
\begin{aligned}
\frac{1}{n^2} \operatorname{Tr}\left(W\widetilde{L}_k W H_1 \widetilde{L} H_2\right) &= \frac{1}{n^2} \sum_{1 \leq i,r \leq n} (\widetilde{L}_k)_{i,r} \widetilde{L}_{i,r} W_{i,i} W_{r,r} + \frac{1}{n^4} \sum_{1 \leq i,q \leq n} (\widetilde{L}_k)_{i,q} W_{i,i} W_{q,q} \sum_{1 \leq p,r \leq n} \widetilde{L}_{p,r} W_{p,p} W_{r,r} \\
&- \frac{2}{n^3} \sum_{1 \leq i,r,s \leq n} (\widetilde{L}_k)_{i,s} \widetilde{L}_{i,r} W_{i,i} W_{s,s} W_{r,r}.
\end{aligned}
$$

By definition of $\widetilde{L}_k$, $\widetilde{L}$ and W, the three terms of the last equation are respectively the estimators defined in Formula (3.9).

3.6.2 Proof of Proposition 3.3

To lighten formulas, we denote $(l_k)_{i,j} = (\widetilde{L}_k)_{i,j}$, $l_{i,j} = \widetilde{L}_{i,j}$ and $w_i = W_{i,i}$. We also denote $\widetilde{\mathrm{HSIC}}_{\mathrm{U}}(X_k, Y)$ the U-statistic associated to the estimator $\widetilde{\mathrm{HSIC}}(X_k, Y)$ defined as follows

$$
\begin{aligned}
\widetilde{\mathrm{HSIC}}_{\mathrm{U}}(X_k, Y) &= \frac{1}{(n)_2} \sum_{(i,j) \in i_2^n} (l_k)_{i,j} l_{i,j} w_i w_j + \frac{1}{(n)_4} \sum_{(i,j,p,q) \in i_4^n} (l_k)_{i,j} l_{p,q} w_i w_j w_p w_q \\
&- \frac{2}{(n)_3} \sum_{(i,j,r) \in i_3^n} (l_k)_{i,j} l_{i,r} w_i w_j w_r \\
&:= \widetilde{\mathrm{H}}_{\mathrm{k}}^{1,\mathrm{U}} + \widetilde{\mathrm{H}}_{\mathrm{k}}^{2,\mathrm{U}} - 2\,\widetilde{\mathrm{H}}_{\mathrm{k}}^{3,\mathrm{U}},
\end{aligned}
$$

where $(n)_s = \frac{n!}{(n-s)!}$ and i_s^n is the set of all s-tuples drawn without replacement from the set $\{1,\dots,n\}$. We also recall that

$$\begin{aligned}\widetilde{\mathrm{HSIC}}(X_k,Y) =& \frac{1}{n^2}\sum_{1\le i,j\le n}(l_k)_{i,j}l_{i,j}w_iw_j + \frac{1}{n^4}\sum_{1\le i,j,p,q\le n}(l_k)_{i,j}l_{p,q}w_iw_jw_pw_q\\ &-\frac{2}{n^3}\sum_{1\le i,j,r\le n}(l_k)_{i,j}l_{i,r}w_iw_jw_r\\ :=& \widetilde{\mathrm{H}}_{\mathrm{k}}^1 + \widetilde{\mathrm{H}}_{\mathrm{k}}^2 - 2\,\widetilde{\mathrm{H}}_{\mathrm{k}}^3,\end{aligned}$$

Denote the null hypothesis ($\mathcal{H}_{0,k}$): "X_k and Y are independent". Then, under ($\mathcal{H}_{0,k}$) the estimator $\widetilde{\mathrm{HSIC}}_{\mathrm{U}}(X_k,Y)$ is centered. The bias of the estimator $\widetilde{\mathrm{HSIC}}(X_k,Y)$ is then equal to that of $\widetilde{\mathrm{HSIC}}(X_k,Y) - \widetilde{\mathrm{HSIC}}_{\mathrm{U}}(X_k,Y)$ under this same assumption. We first compute the expression of $\widetilde{\mathrm{HSIC}}(X_k,Y) - \widetilde{\mathrm{HSIC}}_{\mathrm{U}}(X_k,Y)$, before computing its expectation. Let us compute $\widetilde{\mathrm{HSIC}}(X_k,Y) - \widetilde{\mathrm{HSIC}}_{\mathrm{U}}(X_k,Y)$ term by term:

$$\begin{aligned}\widetilde{\mathrm{H}}_{\mathrm{k}}^1 - \widetilde{\mathrm{H}}_{\mathrm{k}}^{1,\mathrm{U}} =& \frac{1}{n^2}\sum_{i=1}^n (l_k)_{i,i}l_{i,i}w_i^2 - \frac{1}{n^2(n-1)}\sum_{1\le i\ne j\le n}(l_k)_{i,j}l_{i,j}w_iw_j,\\ \widetilde{\mathrm{H}}_{\mathrm{k}}^2 - \widetilde{\mathrm{H}}_{\mathrm{k}}^{2,\mathrm{U}} =& \frac{1}{n^4}\sum_{(i,j,q)\in i_3^n}\left[(l_k)_{i,i}l_{j,q}w_i^2w_jw_q + 4(l_k)_{i,j}l_{i,q}w_i^2w_jw_q + (l_k)_{i,j}l_{q,q}w_iw_jw_q^2\right]\\ &-\frac{6}{n(n)_4}\sum_{(i,j,p,q)\in i_4^n}(l_k)_{i,j}l_{p,q}w_iw_jw_pw_q + O(\frac{1}{n^2}),\\ \widetilde{\mathrm{H}}_{\mathrm{k}}^3 - \widetilde{\mathrm{H}}_{\mathrm{k}}^{3,\mathrm{U}} =& \frac{1}{n^3}\sum_{1\le i\ne j\le n}\left[(l_k)_{i,i}l_{i,j}w_i^2w_j + (l_k)_{i,j}l_{i,i}w_i^2w_j + (l_k)_{i,j}l_{i,j}w_iw_j^2\right]\\ &-\frac{3}{n(n)_3}\sum_{(i,j,r)\in i_3^n}(l_k)_{i,j}l_{i,r}w_iw_jw_r + O(\frac{1}{n^2}).\end{aligned}$$

These expressions can be simplified by replacing $(l_k)_{i,i} = l_{i,i} = 1$:

$$\begin{aligned}\widetilde{\mathrm{H}}_{\mathrm{k}}^1 - \widetilde{\mathrm{H}}_{\mathrm{k}}^{1,\mathrm{U}} =& \frac{1}{n^2}\sum_{i=1}^n w_i^2 - \frac{1}{n^2(n-1)}\sum_{1\le i\ne j\le n}(l_k)_{i,j}l_{i,j}w_iw_j,\\ \widetilde{\mathrm{H}}_{\mathrm{k}}^2 - \widetilde{\mathrm{H}}_{\mathrm{k}}^{2,\mathrm{U}} =& \frac{1}{n^4}\sum_{(i,j,q)\in i_3^n}\left[l_{j,q}w_i^2w_jw_q + 4(l_k)_{i,j}l_{i,q}w_i^2w_jw_q + (l_k)_{i,j}w_iw_jw_q^2\right]\\ &-\frac{6}{n(n)_4}\sum_{(i,j,p,q)\in i_4^n}(l_k)_{i,j}l_{p,q}w_iw_jw_pw_q + O(\frac{1}{n^2}),\\ \widetilde{\mathrm{H}}_{\mathrm{k}}^3 - \widetilde{\mathrm{H}}_{\mathrm{k}}^{3,\mathrm{U}} =& \frac{1}{n^3}\sum_{1\le i\ne j\le n}\left[l_{i,j}w_i^2w_j + (l_k)_{i,j}w_i^2w_j + (l_k)_{i,j}l_{i,j}w_iw_j^2\right]\\ &-\frac{3}{n(n)_3}\sum_{(i,j,r)\in i_3^n}(l_k)_{i,j}l_{i,r}w_iw_jw_r + O(\frac{1}{n^2}).\end{aligned}$$

By computing the expectation of these three estimators under $\mathcal{H}_{0,k}$, we have:

$$\begin{aligned}
\mathbb{E}\left[\widetilde{\mathrm{H}}_{\mathrm{k}}^{1}-\widetilde{\mathrm{H}}_{\mathrm{k}}^{1,\mathrm{U}}\right] &= \frac{1}{n}\left(\mathbb{E}_{\omega}-\mathbb{E}_{x_k}\,\mathbb{E}_{y}\right),\\
\mathbb{E}\left[\widetilde{\mathrm{H}}_{\mathrm{k}}^{2}-\widetilde{\mathrm{H}}_{\mathrm{k}}^{2,\mathrm{U}}\right] &= \frac{1}{n}\left(\mathbb{E}_{\omega}\,\mathbb{E}_{y}+4\,\mathbb{E}_{x_k,\omega}\,\mathbb{E}_{y,\omega}+\mathbb{E}_{\omega}\,\mathbb{E}_{x_k}\right)-\frac{6}{n}\;\mathbb{E}_{x_k}\,\mathbb{E}_{y}+O(\frac{1}{n^2}),\\
\mathbb{E}\left[\widetilde{\mathrm{H}}_{\mathrm{k}}^{3}-\widetilde{\mathrm{H}}_{\mathrm{k}}^{3,\mathrm{U}}\right] &= \frac{1}{n}\left(\mathbb{E}_{\omega}^{k}\,\mathbb{E}_{y,\omega}+\mathbb{E}_{\omega}^{-k}\,\mathbb{E}_{x_k,\omega}+\mathbb{E}_{x_k,\omega}\,\mathbb{E}_{y,\omega}\right)-\frac{3}{n}\,\mathbb{E}_{x_k}\,\mathbb{E}_{y}+O(\frac{1}{n^2}).
\end{aligned}$$

From these last equations, we obtain:

$$\begin{aligned}
\mathbb{E}\left[\widetilde{\mathrm{HSIC}}(X_k,Y)-\widetilde{\mathrm{HSIC}}_{\mathrm{U}}(X_k,Y)\right] &= \frac{2}{n}(\mathbb{E}_{\omega}^{k}-\mathbb{E}_{x_k,\omega})(\mathbb{E}_{\omega}^{-k}-\mathbb{E}_{y,\omega})-\frac{1}{n}(\mathbb{E}_{\omega}-\mathbb{E}_{x_k})(\mathbb{E}_{\omega}-\mathbb{E}_{y})\\
&\quad+\frac{1}{n}\,\mathbb{E}_{\omega}(\mathbb{E}_{\omega}-1)+O(\frac{1}{n^2}).
\end{aligned}$$

Finally, the bias of $\widetilde{\mathrm{HSIC}}(X_k,Y)$ under $(\mathcal{H}_{0,k})$ is written:

$$\begin{aligned}
\mathbb{E}[\widetilde{\mathrm{HSIC}}(X_k,Y)]-\mathrm{HSIC}(X_k,Y) &= \frac{2}{n}(\mathbb{E}_{\omega}^{k}-\mathbb{E}_{x_k,\omega})(\mathbb{E}_{\omega}^{-k}-\mathbb{E}_{y,\omega})-\frac{1}{n}(\mathbb{E}_{\omega}-\mathbb{E}_{x_k})(\mathbb{E}_{\omega}-\mathbb{E}_{y})\\
&\quad+\frac{1}{n}\,\mathbb{E}_{\omega}(\mathbb{E}_{\omega}-1)+O(\frac{1}{n^2}).
\end{aligned}$$

3.6.3 Proof of Proposition 3.4

In order to compute the variance of $\widetilde{\mathrm{HSIC}}(X_k,Y)$ and to determine its asymptotic distribution under $\mathcal{H}_{0,k}$, general theorems on V-statistics must be used. For this, we write this last estimator as a single V-statistic. By analogy with theorem 1 of Gretton et al. (2008), we have:

$$\widetilde{\mathrm{HSIC}}(X_k,Y)=\frac{1}{n^4}\sum_{1\leq i,j,q,r\leq n}\widetilde{h}_{ijqr},$$

where

$$\widetilde{h}_{i,j,q,r}=\frac{1}{4!}\sum_{(t,u,v,s)}^{(i,j,q,r)}(l_k)_{t,u}l_{t,u}w_tw_u+(l_k)_{t,u}l_{v,s}w_tw_uw_vw_s-2(l_k)_{t,u}l_{t,v}w_tw_uw_v,$$

the sum represents all ordered quadruples (t,u,v,s) drawn without replacement from (i,j,q,r).

This equality is easily obtained by decomposing the last sum into three sums, then by writing that:

$$\begin{aligned}
\frac{1}{n^4}\sum_{1\leq i,j,q,r\leq n}\frac{1}{4!}\sum_{(t,u,v,s)}^{(i,j,q,r)}(l_k)_{t,u}l_{t,u}w_tw_u &= \frac{1}{n^2}\sum_{1\leq i,j\leq n}(l_k)_{i,j}l_{i,j}w_iw_j,\\
\frac{1}{n^4}\sum_{1\leq i,j,q,r\leq n}\frac{1}{4!}\sum_{(t,u,v,s)}^{(i,j,q,r)}(l_k)_{t,u}l_{v,s}w_tw_uw_vw_s &= \frac{1}{n^4}\sum_{1\leq i,j,q,r\leq n}(l_k)_{i,j}l_{q,r}w_iw_jw_qw_r,\\
\frac{1}{n^4}\sum_{1\leq i,j,q,r\leq n}\frac{1}{4!}\sum_{(t,u,v,s)}^{(i,j,q,r)}(l_k)_{t,u}l_{t,v}w_tw_uw_v &= \frac{1}{n^3}\sum_{1\leq i,j,r\leq n}(l_k)_{i,j}l_{i,r}w_iw_jw_r.
\end{aligned}$$

The result is then obtained by combining the last three equalities.

Remark. *The U-statistic associated to the estimator* $\widetilde{\text{HSIC}}(X_k, Y)$ *is written:*

$$\widetilde{\text{HSIC}}_{\text{U}}(X_k, Y) = \frac{1}{(n)_4} \sum_{(i,j,q,r)\in i_4^n} \widetilde{h}_{ijqr}.$$

Under $\mathcal{H}_{0,k}$, the estimators $n \times \widetilde{\text{HSIC}}(X_k, Y)$ et $n \times \widetilde{\text{HSIC}}_{\text{U}}(X_k, Y)$ have the same asymptotic behavior (see e.g. Serfling, 2009). Moreover, Hoeffding variance decomposition of $\widetilde{\text{HSIC}}_{\text{U}}(X_k, Y)$ is written:

$$\text{Var}\left(\widetilde{\text{HSIC}}_{\text{U}}(X_k, Y)\right) = \binom{n}{4}^{-1} \sum_{r=1}^{4} \binom{r}{4}\binom{n-4}{4-r}\zeta_r,$$

where $\zeta_r = \text{Var}\left(\mathbb{E}[\widetilde{h}_{1,2,3,4} \mid X_1, \ldots, X_r]\right), r = 1, \ldots, 4.$

Moreover, under $\mathcal{H}_{0,k}$, the variance of $\widetilde{\text{HSIC}}_{\text{U}}(X_k, Y)$ converges to 0 in $O(\frac{1}{n^2})$:

$$\text{Var}(\widetilde{\text{HSIC}}_{\text{U}}(X_k, Y)) = \frac{72(n-4)(n-5)}{n(n-1)(n-2)(n-3)}\zeta_2 + O(\frac{1}{n^3}).$$

Under $\mathcal{H}_{0,k}$, $\zeta_2 = \mathbb{E}_{i,j}\left[\mathbb{E}_{q,r}[\widetilde{h}_{1,2,3,4}]^2\right]$, where the notation $\mathbb{E}_{r,s}$ designates the expectation by integrating only w.r.t variables X_r and X_s.

Moreover, by detailing the different terms of $\widetilde{h}_{i,j,p,q}$, we easily show that:

$$\begin{aligned}
6\widetilde{h}_{ijqr} &= (l_k)_{i,j} w_i w_j \left(l_{i,j} + l_{q,r} w_q w_r\right) + (l_k)_{i,q} w_i w_q \left(l_{i,q} + l_{j,r} w_j w_r\right) + (l_k)_{i,r} w_i w_r \left(l_{i,r} + l_{j,q} w_j w_q\right) \\
&+ (l_k)_{j,q} w_j w_q \left(l_{j,q} + l_{i,r} w_i w_r\right) + (l_k)_{j,r} w_j w_r \left(l_{j,r} + l_{q,i} w_q w_i\right) + (l_k)_{q,r} w_q w_r \left(l_{q,r} + l_{i,j} w_i w_j\right) \\
&- \frac{1}{2} \sum_{(t,u,v)}^{(i,j,q,r)} (l_k)_{t,u} w_t w_u \left(l_{t,v} w_v + l_{u,v} w_v\right)
\end{aligned}$$

where

$$\begin{aligned}
(l_k)_{i,.} &= \widetilde{\mathbb{E}}\left[\widetilde{L}_k\left(X_k^{(i)}, X_k\right)\omega_k(X_k)\right], & l_{i,.} &= \widetilde{\mathbb{E}}\left[\widetilde{L}\left(Y^{(i)}, Y\right)\omega_{-k}(X_{-k})\right], \\
(l_k)_{.,j} &= \widetilde{\mathbb{E}}\left[\widetilde{L}_k\left(X_k^{(j)}, X_k\right)\omega_k(X_k)\right], & l_{j,.} &= \widetilde{\mathbb{E}}\left[\widetilde{L}\left(Y^{(j)}, Y\right)\omega_{-k}(X_{-k})\right], \\
(l_k) &= \widetilde{\mathbb{E}}\left[\widetilde{L}_k\left(X_k', X_k\right)\omega_k(X_k)\right], & l &= \widetilde{\mathbb{E}}\left[\widetilde{L}\left(Y', Y\right)\omega_{-k}(X_{-k})\right].
\end{aligned}$$

We therefore write under $\mathcal{H}_{0,k}$:

$$\begin{aligned}
6E_{q,r}\left[\widetilde{h}_{ijqr}\right] &= (\widetilde{l}_k)_{i,j} w_i w_j \left(\widetilde{l}_{i,j} - \widetilde{l}_{i,.} - \widetilde{l}_{j,.} + \widetilde{l}\right) - (\widetilde{l}_k)_{i,.} w_i \left(l_{i,j} w_j - l_{i,.} - l_{j,.} w_j + l\right) \\
&- (l_k)_{.,j} w_j \left(l_{i,j} w_i - l_{i,.} w_i - l_{j,.} + l\right) - (l_k)\left(l_{i,j} w_i w_j - l_{i,.} w_i - l_{j,.} w_j + l\right).
\end{aligned}$$

$E_{q,r}\left[\widetilde{h}_{ijqr}\right]$ can be estimated empirically by $\frac{1}{6}(\widetilde{B})_{i,j}$, where $\widetilde{B}$ is the matrix defined in Formula (3.13). The variance ζ_2 can be estimated by $\widehat{\zeta_2} = \frac{1}{36n^2}\mathbf{1}^T(\widetilde{B}\odot\widetilde{B})\mathbf{1}$. Formula (3.13) is then obtained by replacing the expression of $\widehat{\zeta_2}$ in Hoeffding's decomposition.

3.6.4 Proof of Theorem 3.1

The asymptotic distribution of the V-statistic $n \times \widetilde{\text{HSIC}}(X_k, Y)$ (as well as the U-statistic $n \times \widetilde{\text{HSIC}}_{\text{U}}(X_k, Y)$) is given by Theorem 5.5.2, page 194 of Serfling (2009), which gives a formulation of the asymptotic distributions of degenerate V-statistics (and U-statistics). Indeed, we can easily show that under $\mathcal{H}_{0,k}$ the statistic $\widetilde{\text{HSIC}}(X_k, Y)$ is degenerate.

Theorem. *Under $\mathcal{H}_{0,k}$ we have the two following convergence theorems:*

$$n \times \widetilde{\text{HSIC}}^V(\widetilde{X}_k, \widetilde{Y}) \xrightarrow{\mathcal{L}} \sum_{l=1}^{+\infty} \lambda_l \widetilde{z}_l^2,$$

$$n \times \widetilde{\text{HSIC}}^U(\widetilde{X}_k, \widetilde{Y}) \xrightarrow{\mathcal{L}} \sum_{l=1}^{+\infty} \lambda_l \left(\widetilde{z}_l^2 - 1\right)$$

where $(\widetilde{z}_l)_{l\geq 1}$ are independent and identically distributed random variables of distribution $\mathcal{N}(0,1)$ and $(\lambda_l)_{l\geq 1}$ are the eigenvalues of the following operator:

$$A(g) : z \mapsto \int \widetilde{h}_{ijqr}(\widetilde{z}, \widetilde{z}_j, \widetilde{z}_q, \widetilde{z}_r)\, g(\widetilde{z}_j)\, dF_{jqr}$$

where the integral is w.r.t. the distributions of the random variables $\widetilde{z}_j$, $\widetilde{z}_q$ and $\widetilde{z}_r$.

To conclude, the distribution $\sum_{l=1}^{+\infty} \lambda_l \widetilde{z}_l^2$ can be approximated by a Gamma distribution according to Gretton et al. (2008). In fact, it is an infinite sum of independent random variables of distribution χ^2 (Chi two). The asymptotic distribution of the V-statistic $n \times \widetilde{\text{HSIC}}(X_k, Y)$ under $\mathcal{H}_{0,k}$ is a Gamma distribution, whose parameters can be estimated based on the empirical expectation and variance of $n \times \widetilde{\text{HSIC}}(X_k, Y)$ (see section 3.3.2.2).

3.6.5 Proof of Proposition 3.5

Let α be in $(0,1)$. In order to prove that the permuted test with Monte Carlo approximation $\widehat{\Delta}_\alpha$ is of prescribed level α, we use the following lemma of Romano and Wolf (2005).

Lemma 3.1 (Romano and Wolf (2005, Lemma 1))**.** *Let $R_1, \ldots, R_{B+1}$ be $(B+1)$ exchangeable random variables. Then, for all u in $(0,1)$*

$$\mathbb{P}\left(\frac{1}{B+1}\left(1 + \sum_{b=1}^{B} \mathbb{1}_{R_b \geq R_{B+1}}\right) \leq u\right) \leq u.$$

Recall that for all $1 \leq b \leq B$,

$$\widehat{H}^{\star b} = \widehat{\mathrm{HSIC}}\left(\mathbb{Z}_n^{\tau_b}\right) \quad \text{and} \quad \widehat{H}^{\star B+1} = \widehat{\mathrm{HSIC}}\left(\mathbb{Z}_n\right) = \widehat{\mathrm{HSIC}}\left(\mathbb{Z}_n^{\tau_{B+1}}\right),$$

where $\tau_{B+1} = \mathrm{id}$ is the identity permutation of $\{1, \ldots, B+1\}$ (deterministic).

Assume that $f_{X,Y} = f_X \otimes f_Y$. Then the random variables $\widehat{H}^{\star 1}, \ldots, \widehat{H}^{\star B}$ and $\widehat{H}^{\star B+1}$ are exchangeable. Indeed, let π be a (deterministic) permutation of $\{1, \ldots, B+1\}$ and let us prove that

$$\left(\widehat{H}^{\star 1}, \ldots, \widehat{H}^{\star B}, \widehat{H}^{\star B+1}\right) \quad \text{and} \quad \left(\widehat{H}^{\star \pi(1)}, \ldots, \widehat{H}^{\star \pi(B+1)}\right) \quad \text{have the same distribution.} \tag{3.26}$$

1st case: if $\pi(B+1) = B+1$. Then, since the permutations $(\tau_b)_{1 \leq b \leq B}$ are i.i.d., they are exchangeable. Hence, $(\tau_{\pi(1)}, \ldots, \tau_{\pi(B)})$ is an i.i.d. sample of uniform permutations of $\{1, \ldots, n\}$, independent of $\mathbb{Z}_n$ and (3.26) holds by construction.

2nd case: if $\pi(B+1) \neq B+1$. Then,

$$\widehat{H}^{\star \pi(B+1)} = \widehat{\mathrm{HSIC}}\left(\mathbb{Z}_n^{\tau_{\pi(B+1)}}\right) = \widehat{\mathrm{HSIC}}\left(\tilde{\mathbb{Z}}_n\right), \quad \text{where} \quad \tilde{\mathbb{Z}}_n = \mathbb{Z}_n^{\tau_{\pi(B+1)}}.$$

In particular, for all b in $\{1, \ldots, B\}$,

$$\begin{cases} \widehat{H}^{\star \pi(b)} = \widehat{\mathrm{HSIC}}\left(\mathbb{Z}_n^{\tau_{\pi(b)}}\right) = \widehat{\mathrm{HSIC}}\left(\tilde{\mathbb{Z}}_n^{\tau_{\pi(b)} \circ \tau_{\pi(B+1)}^{-1}}\right) & \text{if } \pi(b) \neq B+1, \\ \widehat{H}^{\star \pi(b)} = \widehat{\mathrm{HSIC}}\left(\mathbb{Z}_n\right) = \widehat{\mathrm{HSIC}}\left(\tilde{\mathbb{Z}}_n^{\mathrm{id} \circ \tau_{\pi(B+1)}^{-1}}\right) & \text{if } \pi(b) = B+1. \end{cases}$$

Therefore, in order to prove (3.26), it is sufficient to prove that $\{\tau_{\pi(1)} \circ \tau_{\pi(B+1)}^{-1}, \ldots, \tau_{\pi(B)} \circ \tau_{\pi(B+1)}^{-1}\}$ is an i.i.d. sample of uniform permutations of $\{1, \ldots, n\}$ independent of $\tilde{\mathbb{Z}}_n$. Let A be a mesurable set, and $\sigma_1, \ldots, \sigma_B$ be (fixed) permutations of $\{1, \ldots, n\}$. Then,

$$\begin{aligned} &\mathbb{P}\left(\tilde{\mathbb{Z}}_n \in A,\ \tau_{\pi(1)} \circ \tau_{\pi(B+1)}^{-1} = \sigma_1,\ \ldots,\ \tau_{\pi(B)} \circ \tau_{\pi(B+1)}^{-1} = \sigma_B\right) \\ &= \mathbb{P}\left(\mathbb{Z}_n^{\tau_{\pi(B+1)}} \in A, \tau_{\pi(1)} = \sigma_1 \circ \tau_{\pi(B+1)}, \ldots, \tau_{\pi(B)} = \sigma_B \circ \tau_{\pi(B+1)}\right) \\ &= \mathbb{E}\left[\mathbb{P}\left(\mathbb{Z}_n^{\tau_{\pi(B+1)}} \in A, \tau_{\pi(1)} = \sigma_1 \circ \tau_{\pi(B+1)}, \ldots, \tau_{\pi(B)} = \sigma_B \circ \tau_{\pi(B+1)} \middle| \tau_{\pi(B+1)}\right)\right]. \end{aligned}$$

This leads to

$$\begin{aligned} &\mathbb{P}\left(\tilde{\mathbb{Z}}_n \in A,\ \tau_{\pi(1)} \circ \tau_{\pi(B+1)}^{-1} = \sigma_1,\ \ldots,\ \tau_{\pi(B)} \circ \tau_{\pi(B+1)}^{-1} = \sigma_B\right) \\ &\quad = \mathbb{E}\left[\mathbb{P}(\mathbb{Z}_n \in A) \times \left(\prod_{\substack{b=1 \\ b \neq \pi^{-1}(B+1)}}^{B} \mathbb{P}\left(\tau_{\pi(b)} = \sigma_b \circ \tau_{\pi(B+1)} \middle| \tau_{\pi(B+1)}\right)\right) \times \right. \\ &\qquad \left. \mathbb{P}\left(\mathrm{id} = \sigma_{\pi^{-1}(B+1)} \circ \tau_{\pi(B+1)} \middle| \tau_{\pi(B+1)}\right)\right], \end{aligned} \tag{3.27}$$

where (3.27) holds by independence of all permutations τ_b and of $\mathbb{Z}_n$ and since, if $f_{X,Y} = f_X \otimes f_Y$, $\mathbb{Z}_n^{\tau_{\pi(B+1)}}$ and $\mathbb{Z}_n$ have the same distribution. Hence,

$$\begin{aligned}
&\mathbb{P}\big(\tilde{\mathbb{Z}}_n \in A,\ \tau_{\pi(1)} \circ \tau_{\pi(B+1)}^{-1} = \sigma_1,\ \ldots,\ \tau_{\pi(B)} \circ \tau_{\pi(B+1)}^{-1} = \sigma_B\big) \\
&= \mathbb{E}\left[\mathbb{P}(\mathbb{Z}_n \in A)\left(\frac{1}{n!}\right)^{B-1} \mathbb{P}\big(\mathrm{id} = \sigma_{\pi^{-1}(B+1)} \circ \tau_{\pi(B+1)} \big| \tau_{\pi(B+1)}\big)\right], \\
&= \mathbb{P}(\mathbb{Z}_n \in A)\left(\frac{1}{n!}\right)^{B-1} \mathbb{P}\Big(\tau_{\pi(B+1)} = \sigma_{\pi^{-1}(B+1)}^{-1}\Big), \\
&= \mathbb{P}(\mathbb{Z}_n \in A)\left(\frac{1}{n!}\right)^{B},
\end{aligned}$$

This ends the proof of the exchangeability of the $(\widehat{H}^{\star b}_{\lambda,\mu})_{1\le b\le B+1}$.

Then, by applying Lemma 3.1 to the $(\widehat{H}^{\star b})_{1\le b\le B+1}$, we obtain

$$\begin{aligned}
P_{f_X\otimes f_Y}\left(\widehat{\Delta}_\alpha = 1\right) &= P_{f_X\otimes f_Y}\left(\widehat{\mathrm{HSIC}} > \widehat{q}_{1-\alpha}\right) \\
&= P_{f_X\otimes f_Y}\left(\widehat{H}^{\star B+1} > \widehat{H}^{\star(\lceil (B+1)(1-\alpha)\rceil)}\right) \\
&= P_{f_X\otimes f_Y}\left(\sum_{b=1}^{B+1} \mathbb{1}_{\widehat{H}^{\star b} < \widehat{H}^{\star B+1}} \ge \lceil (B+1)(1-\alpha)\rceil\right) \\
&= P_{f_X\otimes f_Y}\left(\sum_{b=1}^{B+1} \mathbb{1}_{\widehat{H}^{\star b} \ge \widehat{H}^{\star B+1}} \le \lfloor \alpha(B+1)\rfloor\right) \qquad (3.28) \\
&= P_{f_X\otimes f_Y}\left(\sum_{b=1}^{B+1} \mathbb{1}_{\widehat{H}^{\star b} \ge \widehat{H}^{\star B+1}} \le \alpha(B+1)\right) \\
&= P_{f_X\otimes f_Y}\left(\frac{1}{B+1}\left(1 + \sum_{b=1}^{B} \mathbb{1}_{\widehat{H}^{\star b} \ge \widehat{H}^{\star B+1}}\right) \le \alpha\right) \\
&\le \alpha, \qquad (3.29)
\end{aligned}$$

where (3.28) comes from the fact that

$$B + 1 - \lceil (B+1)(1-\alpha)\rceil = \lfloor \alpha(B+1)\rfloor,$$

and (3.29) is obtained from Lemma 3.1.

Chapter 4

Global sensitivity analysis for second level uncertainties

4.1 Issues and objectives

As previously outlined in Chapter 1 (in French) and Chapter 2 (in English), numerical simulators are powerful tools for modeling, studying and better comprehending natural phenomena. For many reasons, the inputs of these simulators are often uncertain or poorly known. Studies are therefore carried out to take these uncertainties into account. In the probabilistic framework, the inputs and output are modeled by random variables. The probability distribution of each input may be ***perfectly known*** or ***itself uncertain***[1]. For the last case, we differentiate in this manuscript two "levels" of Global Sensitivity Analysis (GSA):

First-level uncertainties and GSA1. For given probability distributions of the inputs, we call *first-level* GSA or GSA1, the "usual" Global Sensitivity Analysis between the uncertain inputs of the model and its output. Using classical methods, GSA1 studies can only be performed when the input probability distributions are perfectly (or assumed to be) known. Figure 4.1 illustrates the general principle of GSA1.

Second-level input uncertainties and GSA2. In some cases, the probability distributions characterizing the uncertain inputs may themselves be uncertain. These uncertainties may be related to a divergence of expert opinion on the probability distribution assigned to each input or a lack of information to characterize this distribution. The modeling of this lack of knowledge on input laws can take many forms:

- the type of the input distribution is uncertain (uniform, triangular, normal law, ...);
- the distribution is known but its parameters are uncertain (e.g., known normal distribution with unknown mean and variance, eventually estimated on data).
- the lack of knowledge about the distribution is a mixture of uncertainties about its type and its parameters (e.g. discrete law on a set of distributions with uncertain parameters for each of them).

The resulting uncertainties on the input laws are referred to here as *second-level uncertainties*. To handle this type of uncertainties in GSA studies, two main approaches are possible: *Aggregating*

[1] Further clarifications below.

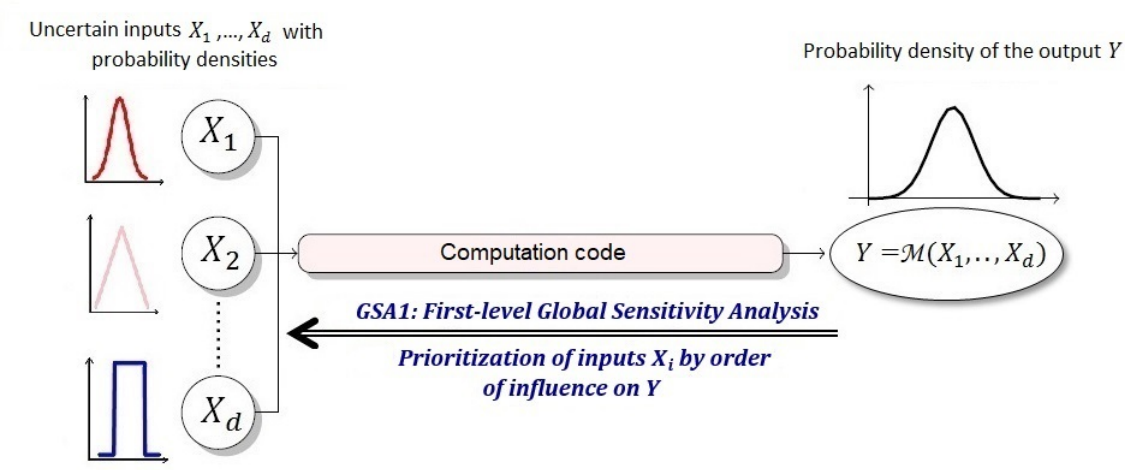

Figure 4.1 – General principle of GSA1.

uncertainties or *Separating them*. The first one consists in considering simultaneously both first and second level uncertainties, as a unique uncertainty. The lack of knowledge can then be summarized by a single-level uncertainty and represented by a well-defined probability distribution. In this case, the most intuitive choice for this distribution (and also the most commonly adopted) is the mixture distribution. Another solution for aggregating could be to integrate GSA1 results over the uncertainties on input distributions (Chabridon et al., 2017). By contrast, the principle of separating uncertainties is to dissociate the impacts of first and second uncertainties on the output. Certainly, the approach of separating uncertainties is more difficult to manipulate and requires more model simulations comparing to the approach of aggregating uncertainties. However, it provides a more complete picture and allows to avoid the loss of information due to the aggregation. For this reason, we choose to deal with second-level uncertainties by separating the input uncertainties from the probabilistic model uncertainties.

In this framework, 2^{nd}-level uncertainties can be modeled by a probability law on a set of possible probability laws of inputs or by a probability law on the parameters of a given input law (e.g. Gaussian distribution with probability law on mean and/or variance). In any case, these 2^{nd}-level uncertainties can significantly change the GSA1 results performed by HSIC or any other dependence measure. In Figure 4.2, we illustrate in a trivial way how the input distributions can change the GSA1 results. **The main purpose of *second-level GSA (GSA2)* is then to answer the following questions: «What impact do 2^{nd}-level uncertainties have on the GSA1 results ?» and «What are the most influential ones and those whose influence is negligible ?».** The GSA2 results and conclusion can then be used to prioritize the characterization efforts on the inputs whose uncertainties on probability laws have the greatest impact on GSA1 results. Figure 4.3 summarizes the general principle of GSA2.

Practical problems raised by GSA2. In practice, the realization of GSA2 raises several issues and technical obstacles. First, it is necessary to characterize GSA1 results, i.e. to define a quantity of interest which is representative of GSA1, in order to compare the results obtained for different laws of the inputs. Then, the impact of each uncertain input law on this quantity of interest has to be evaluated. For this, sensitivity indices measuring the dependence between GSA1 results and each input pdf have to be defined. We propose to call them *2^{nd}-level* indices

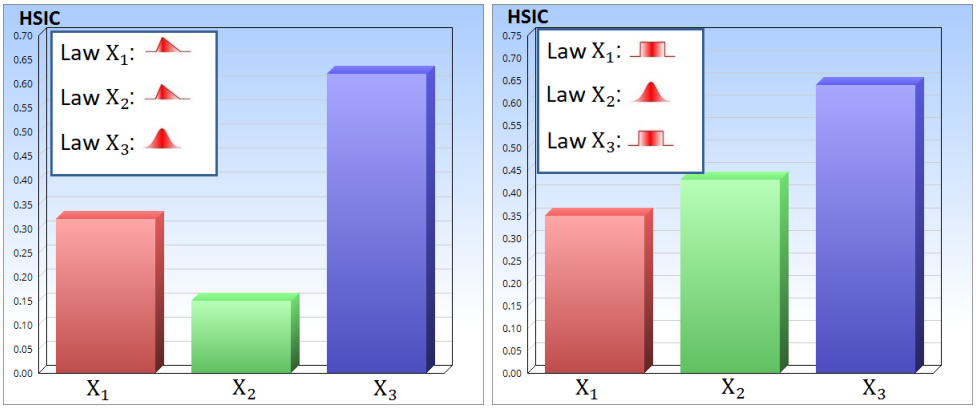

Figure 4.2 – Basic illustration of the impact of second-level uncertainties on GSA1 results.

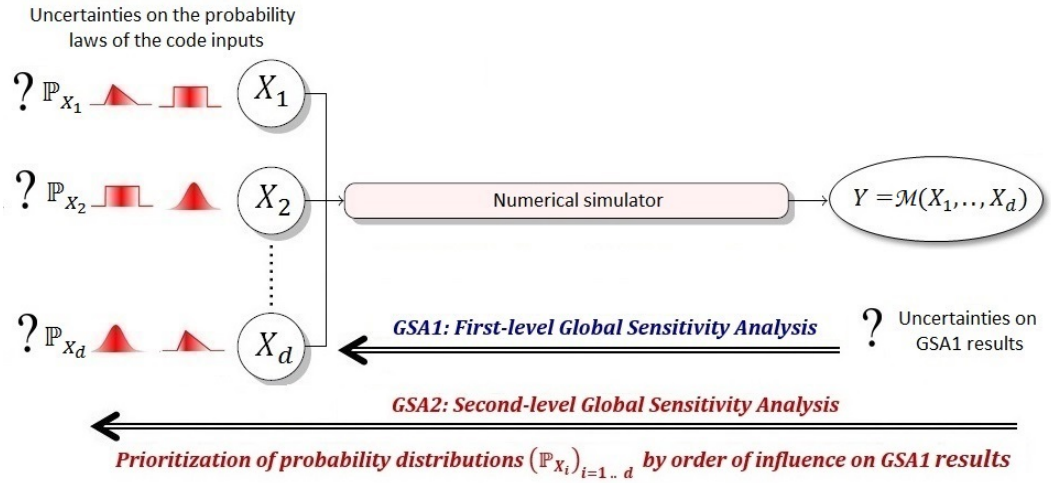

Figure 4.3 – General principle of GSA2.

(or GSA2 indices). In order to estimate these measures, an approach based on a "double Monte Carlo loop" could be considered. In the outer loop, a Monte Carlo sample of input laws is sorted, while the inner loop aims at evaluating the GSA1 results associated to each law. For each law selected in the outer loop, the inner loop consists in generating a Monte Carlo sample of code simulations (set of inputs/output) and to compute GSA1 results. The process is repeated for each input law. At the end of the outer loop, the impact of input pdf on the GSA1 results can be observed and quantified by computing GSA2 indices. Unfortunately, this type of double loop approach requires in practice a very large number of simulations which is intractable for time expensive computer codes. Therefore, other less expensive approaches must be developed.

To answer these different issues (choice of the quantity of interest, definition of GSA2 indices and reduction of the budget of simulations), we propose in this chapter a "single loop" Monte Carlo methodology for GSA2 based on both 1st-level and 2nd-level HSIC dependence measures.

This chapter is organized as follows. In Section 4.2, the full methodology for GSA2 is presented: a single inputs/output sample is used, taking advantage of the new HSIC estimators introduced in section 3.3.2. The GSA2 principle and the related practical issues are first detailed. The proposed general algorithm of GSA2 is then developed, followed by dedicated sections focusing on major technical elements. In Section 4.3, the methodology is illustrated on an analytical example, thereby comparing different options and technical choices of the methodology. The proposed methodology is applied to the accident scenario ULOF described in Chapters 1 (in French) and 2 (in English), simulated with MACARENa code. Finally, the methodology is compared with the extra-probabilistic approach of Dempster-Shafer which deals with epistemic uncertainties.

4.2 New methodology for second-level GSA

Let us first introduce some useful notations and assumptions. We assume here that the inputs are independent and continuous random variables with a probability density function, denoted pdf.

The inputs $X_1, \ldots, X_d$ vary according to unknown probability distributions respectively denoted $\mathbb{P}_{X_1}, \ldots, \mathbb{P}_{X_d}$ and we denote by $\mathbb{P}_X = \mathbb{P}_{X_1} \times \ldots \times \mathbb{P}_{X_d}$ the joint distribution of the vector $\mathbf{X}$ of inputs. Similarly, we denote by $\mathbb{P}_{\mathbb{P}_{X_1}}, \ldots, \mathbb{P}_{\mathbb{P}_{X_d}}$, the probability laws of the uncertain distributions $\mathbb{P}_{X_1}, \ldots, \mathbb{P}_{X_d}$.

4.2.1 Issues raised by GSA2

We present in what follows the different issues and technical locks raised by the realization of a GSA2.

4.2.1.1 Characterization of GSA1 results

The realization of GSA2 requires a prior characterization of GSA1 results. This characterization consists in associating to a given input distribution $\mathbb{P}_X = \mathbb{P}_{X_1} \times \ldots \times \mathbb{P}_{X_d}$, a measurable quantity $\mathcal{R}$ which represents GSA1 results. To choose this quantity of interest, we propose the following options introduced in §3.2.3 and all based on HSIC:

- **Vector $\mathrm{R}^{\mathbf{2}}_{\mathrm{HSIC}} = (\mathrm{R}^2_{\mathrm{HSIC},1}, \ldots, \mathrm{R}^2_{\mathrm{HSIC},d})$ of sensitivity indices.** The quantity of interest $\mathcal{R} = \mathrm{R}^{\mathbf{2}}_{\mathrm{HSIC}}$ is thereby a vector of d real components.
- **Ranking of inputs $X_1, \ldots, X_d$ using the indices $\mathrm{R}^2_{\mathrm{HSIC},1}, \ldots, \mathrm{R}^2_{\mathrm{HSIC},d}$.** In this case, the quantity of interest $\mathcal{R}$ is a permutation on the set $\{1, \ldots, d\}$, which verifies that $\mathcal{R}(k) = j$ if and only if the variable X_j is the k-th in the ranking.
- **Vector $\mathrm{P} = (\mathrm{P}_1, \ldots, \mathrm{P}_\mathrm{d})$ of p-values associated with asymptotic independence tests.** In this case, the quantity of interest $\mathcal{R} = \mathrm{P}$ is a vector of d components in $[0,1]^d$.
- **Vector $\mathrm{p} = (\mathrm{p}_1, \ldots, \mathrm{p}_\mathrm{d})$ of p-values associated with permutation independence tests.** The quantity of interest $\mathcal{R} = \mathrm{p}$ is a vector of d components in $[0,1]^d$.

4.2.1.2 Definition of GSA2 indices

By analogy with formulas (3.1), it is possible to build 2^{nd}-level HSIC measures between the probability distributions $\mathbb{P}_{X_1}, \ldots, \mathbb{P}_{X_d}$ and the quantity of interest $\mathcal{R}$. This involves to define

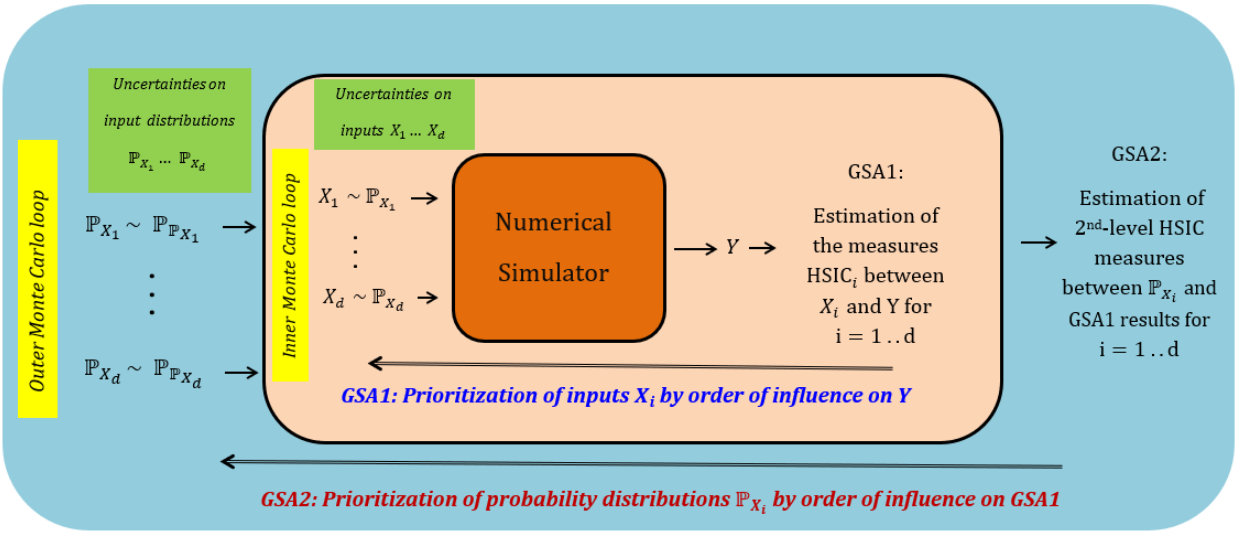

Figure 4.4 – General scheme for GSA2 using the "double loop"approach.

RKHS kernels $l_{\mathcal{D}_1}, \dots, l_{\mathcal{D}_\text{d}}$ on input distributions and a RKHS kernel $l_{\mathcal{R}}$ on the quantity of interest $\mathcal{R}$. This point will be further detailed in Section 4.2.3. Thus, assuming all the kernels are defined, we propose the 2nd-level HSIC measures defined for $k = 1..d$ by:

$$\begin{aligned}\text{HSIC}(\mathbb{P}_{X_k}, \mathcal{R}) = \mathbb{E}\left[l_{\mathcal{D}_\text{k}}(\mathbb{P}_{X_k}, \mathbb{P}'_{X_k}) l_{\mathcal{R}}(\mathcal{R}, \mathcal{R}')\right] + \mathbb{E}\left[l_{\mathcal{D}_\text{k}}\left(\mathbb{P}_{X_k}, \mathbb{P}'_{X_k}\right)\right] \mathbb{E}\left[l_{\mathcal{R}}\left(\mathcal{R}, \mathcal{R}'\right)\right] \\ - 2\mathbb{E}\left[\mathbb{E}\left[l_{\mathcal{D}_\text{k}}\left(\mathbb{P}_{X_k}, \mathbb{P}'_{X_k}\right) \mid \mathbb{P}_{X_k}\right] \mathbb{E}\left[l_{\mathcal{R}}\left(\mathcal{R}, \mathcal{R}'\right) \mid \mathcal{R}\right]\right], \end{aligned} \tag{4.1}$$

where $\mathbb{P}'_{X_k}$ is an independent and identically distributed copy of $\mathbb{P}_{X_k}$ and $\mathcal{R}'$ the GSA1 results associated to $\mathbb{P}'_{X_k}$.

From 2nd-level HSIC measures, we can define GSA2 indices by:

$$\text{R}^2_{\text{HSIC}}(\mathbb{P}_{X_k}, \mathcal{R}) = \frac{\text{HSIC}(\mathbb{P}_{X_k}, \mathcal{R})}{\sqrt{\text{HSIC}(\mathbb{P}_{X_k}, \mathbb{P}_{X_k})\,\text{HSIC}(\mathcal{R}, \mathcal{R})}}, \text{ for } k = 1..d. \tag{4.2}$$

4.2.1.3 Monte Carlo estimation

To estimate $\text{R}^2_{\text{HSIC}}(\mathbb{P}_{X_k}, \mathcal{R})$, for $k = 1..d$, one has to dispose of a n_1-sized sample $(\mathbb{P}^{(i)}_{\mathbf{X}}, \mathcal{R}^{(i)})_{1 \leq i \leq n_1}$ of $(\mathbb{P}_{\mathbf{X}}, \mathcal{R})$. For this, we could consider a double loop Monte Carlo approach as shown in Figure 4.4. In the outer loop, at each iteration i, a distribution $\mathbb{P}^{(i)}_{\mathbf{X}}$ is randomly generated from $\mathbb{P}_{\mathbb{P}_X} = \mathbb{P}_{\mathbb{P}_{X_1}} \times \dots \times \mathbb{P}_{\mathbb{P}_{X_d}}$. The quantity of interest $\mathcal{R}^{(i)}$ associated to this distribution $\mathbb{P}^{(i)}_{\mathbf{X}}$ is provided by a 2nd loop. This inner loop consists in generating a n_2-sized sample $(X_1^{(i,j)}, \dots, X_d^{(i,j)})_{1 \leq j \leq n_2}$ where $\mathbf{X}$ follows $\mathbb{P}^{(i)}_{\mathbf{X}}$. The n_2 corresponding outputs $(Y^{(i,j)})_{1 \leq j \leq n_2}$ are computed in this inner loop. Once this loop performed, the quantity of interest $\mathcal{R}^{(i)}$ is computed from $\mathcal{E}^{(i)} = (X_1^{(i,j)}, \dots, X_d^{(i,j)}, Y^{(i,j)})_{1 \leq j \leq n_2}$. This process is repeated for each $\mathbb{P}^{(i)}_{\mathbf{X}}$ of the outer loop. At the end, 2nd-level HSIC can be estimated from the sample $(\mathbb{P}^{(i)}_{\mathbf{X}}, \mathcal{R}^{(i)})_{1 \leq i \leq n_1}$ by:

$$\widehat{\text{HSIC}}(\mathbb{P}_{X_k}, \mathcal{R}) = \frac{1}{n_1^2} Tr(L_{\mathcal{D}_\text{k}} H L_{\mathcal{R}} H), \tag{4.3}$$

where $L_{\mathcal{D}_k}$ and $L_{\mathcal{R}}$ are the matrices defined for all $(i,j) \in \{1,\dots,n_1\}$ by: $(L_{\mathcal{D}_k})_{i,j} = l_{\mathcal{D}_k}(\mathbb{P}_{X_k}^{(i)}, \mathbb{P}_{X_k}^{(j)})$, $(L_{\mathcal{R}})_{i,j} = l_{\mathcal{R}}\left(\mathcal{R}^{(i)}, \mathcal{R}^{(j)}\right)$ and H the matrix defined in Formula (3.6).

From 2nd-level HSIC estimators, GSA2 indices can be estimated using plug-in and Formula (4.3) by:

$$\widehat{\mathrm{R}}^2_{\mathrm{HSIC}}(\mathbb{P}_{X_k}, \mathcal{R}) = \frac{\widehat{\mathrm{HSIC}}(\mathbb{P}_{X_k}, \mathcal{R})}{\sqrt{\widehat{\mathrm{HSIC}}(\mathbb{P}_{X_k}, \mathbb{P}_{X_k})\widehat{\mathrm{HSIC}}(\mathcal{R}, \mathcal{R})}}. \tag{4.4}$$

Consequently, this Monte Carlo double-loop approach requires a total of $n_1 n_2$ code simulations. For example, if $n_1 = 100$ and $n_2 = 1000$, the computation of GSA2 HSIC indices requires 10^5 code calls. This approach is therefore not tractable for CPU-time expensive simulators.

To overcome this problem and reduce the number of code-calls, we propose a single-loop Monte Carlo approach to obtain the sample $(\mathbb{P}_{\mathbf{X}}^{(i)}, \mathcal{R}^{(i)})_{1\leq i\leq n_1}$, which requires only n_2 simulations, and allows to consider a large sample $\mathcal{P}$ of distributions $\mathbb{P}_{\mathbf{X}}$. This new algorithm is detailed in the next section.

4.2.2 General algorithm for computing GSA2 indices with a single Monte Carlo loop

In this part, we detail our algorithm to estimate the GSA2 HSIC indices (and $\mathrm{R}^2_{\mathrm{HSIC}}$) from a unique inputs/output sample $\mathcal{E}$. We assume that inputs are generated from a unique and known probability distribution denoted $\overline{\mathbb{P}}_{\mathbf{X}} = \overline{\mathbb{P}}_{X_1} \times \ldots \times \overline{\mathbb{P}}_{X_d}$ with density denoted $\overline{f}(x_1,\dots,x_d) = \overline{f}_1(x_1)\times\ldots\times\overline{f}_d(x_d)$. The options for choosing $\overline{f}$ will be discussed in Section 4.2.4. The algorithm consists of 3 steps:

- **Step 1. Build a unique n_2-sized sample $\mathcal{E}$ from $\overline{f}$**

  In this step, we first draw a sample $\overline{\mathbf{X}} = \left(\mathbf{X}^{(i)}\right)_{1\leq i\leq n_2}$ according to $\overline{f}$, then we compute the associated outputs $\overline{Y} = \left(Y^{(i)}\right)_{1\leq i\leq n_2}$, to obtain a sample $\mathcal{E} = (\overline{\mathbf{X}}, \overline{Y})$ of inputs/output. Thus, in what follows, all the formulas for modified HSIC will be used with the alternative sample $\mathcal{E}$, $\overline{f}$ being the alternative distribution. Hence, in all the modified HSIC formulas, the alternative sample will be $(\widetilde{\mathbf{X}}, \widetilde{Y}) = (\overline{\mathbf{X}}, \overline{Y})$.

- **Step 2. Perform n_1 GSA1 from $\mathcal{E}$**

  First, we generate a n_1-sized sample of input distributions according to $\mathbb{P}_{\mathbb{P}_{\mathbf{X}}}$. This sample of distributions is denoted $\mathcal{P} = \left(\mathbb{P}_{\mathbf{X}}^{(i)}\right)_{1\leq i\leq n_1}$ and the density associated to each distribution $\mathbb{P}_{\mathbf{X}}^{(i)}$ is denoted $f^{(i)} = (f_1^{(i)},\dots,f_d^{(i)})$. The objective is then to compute the GSA1 results $\mathcal{R}^{(i)}$ associated to each distribution $\mathbb{P}_{\mathbf{X}}^{(i)}$, using only $\mathcal{E}$. The options proposed for $\mathcal{R}^{(i)}$ in Section 4.2.1.1 are distinguished:

  - **Vector $\mathcal{R}^{(i)} = (\mathrm{R}^{2,(i)}_{\mathrm{HSIC},1},\dots,\mathrm{R}^{2,(i)}_{\mathrm{HSIC},d})$ of sensitivity indices.** In this case, each $\mathrm{R}^{2,(i)}_{\mathrm{HSIC}}$ is estimated by $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},k}$ given by Equation (3.11) with $\mathcal{E} = (\overline{\mathbf{X}}, \overline{Y})$.
  - **Ranking of inputs $X_1,\dots,X_d$ using the indices $\mathrm{R}^2_{\mathrm{HSIC},1},\dots,\mathrm{R}^2_{\mathrm{HSIC},d}$.** These rankings are obtained by ordering the coordinates of $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC}}$ vectors; still estimated from $\mathcal{E}$ and Equation (3.11).

- **Vector $\mathcal{R}^{(i)} = (\mathrm{P}_1^{(i)}, \ldots, \mathrm{P}_d^{(i)})$ of p-values associated with asymptotic independence tests.** By analogy with Equation (3.19), each $\mathrm{P}_k^{(i)}$ is estimated thanks to the properties of the modified estimators:

$$\widetilde{\mathrm{P}}_k^{(i)} \simeq 1 - \widetilde{F}_{\mathrm{G_k}}\left(n_2 \times \widetilde{\mathrm{HSIC}}(X_k^{(i)}, Y)_{obs}\right), \ k = 1, \ldots, d \tag{4.5}$$

where $\tilde{F}_{\mathrm{G_k}}$ is the cumulative distribution function of Gamma law approximating the asymptotic law of $n_2 \times \widetilde{\mathrm{HSIC}}(X_k^{(i)}, Y)$.

- **Vector $\mathcal{R}^{(i)} = (\mathrm{p}_1^{(i)}, \ldots, \mathrm{p}_d^{(i)})$ of p-values associated with permutation independence tests.** Using the same notations as in Formula (3.21), each $\mathrm{p}_k^{(i)}$ is estimated by:

$$\widetilde{\mathrm{p}}_k^{(i)} = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}_{\widetilde{\mathrm{HSIC}}^{[b]}(X_k^{(i)}, Y) > \widetilde{\mathrm{HSIC}}(X_k^{(i)}, Y)}, \ k = 1, \ldots, d. \tag{4.6}$$

- **Step 3. Estimate GSA2 indices**

Finally, the GSA2 indices $\mathrm{R}^2_{\mathrm{HSIC}}(\mathbb{P}_{X_k}, \mathcal{R})$ are estimated with the sample $(\mathbb{P}_{\mathbf{X}}^{(i)}, \widetilde{\mathcal{R}}^{(i)})_{1 \leq i \leq n_1}$ using Formulas (4.3) and (4.4). The computation of matrices $L_{\mathcal{D}_{\mathrm{k}}}$, $k = 1, \ldots, d$ and $L_{\mathcal{R}}$ requires the definition of specific RKHS kernels $l_{\mathcal{D}_{\mathrm{k}}}$, $k = 1, \ldots, d$ and $l_{\mathcal{R}}$. This item is detailed in the next section.

4.2.3 Choice of characteristic kernels for probability distributions and for quantities of interest

In this part, we present examples of characteristic RKHS kernels for probability distributions and for the different quantities of interest $\mathcal{R}$, these kernels being involved in Formula (4.3) (and as a result in Equation (4.4)).

Characteristic RKHS kernel for probability distributions. Before defining a characteristic kernel for distributions, we first introduce the Maximum Mean Discrepancy (MMD) defined in Gretton et al. (2012). If we consider two distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ having the same support and if K denotes a RKHS kernel defined on the common support of $\mathbb{P}_1$ and $\mathbb{P}_2$, then the MMD between $\mathbb{P}_1$ and $\mathbb{P}_2$ induced by K is defined as:

$$\mathrm{MMD}_K(\mathbb{P}_1, \mathbb{P}_2) = \mathbb{E}[K(Z_1, Z_1')] - 2\mathbb{E}[K(Z_1, Z_2)] + \mathbb{E}[K(Z_2, Z_2')], \tag{4.7}$$

where Z_1, Z_2 are random variables respectively with laws $\mathbb{P}_1$, $\mathbb{P}_2$ and Z_1', Z_2' are independent and identically distributed copies respectively of Z_1, Z_2.

Authors of Gretton et al. (2012) establish that when K is characteristic, the MMD associated to K defines a distance. From MMD distance, Sriperumbudur et al. (2010) defines Gaussian RKHS kernels between probability distributions in a similar way to Formula (3.2):

$$l_{\mathcal{D}}(\mathbb{P}_1, \mathbb{P}_2) = \exp\left(-\lambda\, \mathrm{MMD}_K^2(\mathbb{P}_1, \mathbb{P}_2)\right), \tag{4.8}$$

where λ is a positive real parameter.

It has been shown in Christmann and Steinwart (2010) that when the common support of distributions is compact, the Gaussian MMD-based kernel is universal (and consequently characteristic). We can then define kernels $l_{\mathcal{D}_{\mathrm{k}}}$, $k = 1, \ldots, d$ introduced in Formula (4.1) by:

$$l_{\mathcal{D}_{\mathrm{k}}}(\mathbb{P}_{X_k}, \mathbb{P}_{X_k'}) = \exp\left(-\lambda_k\, \mathrm{MMD}_{l_k}^2(\mathbb{P}_{X_k}, \mathbb{P}_{X_k'})\right), \tag{4.9}$$

where $\lambda_k,\ k = 1, \ldots, d$ are positive real parameters.

From a practical point of view, one can choose λ_k as the inverse of s_k^2, the empirical variance w.r.t MMD distance (i.e. $\lambda_k = 1/s_k^2$):

$$s_k^2 = \frac{1}{n_1^2} \sum_{i=1}^{n_1} \mathrm{MMD}_{l_k}^2 \left(\mathbb{P}_{X_k}^{(i)}, \overline{\mathbb{P}}_{X_k} \right),$$

where the distribution $\overline{\mathbb{P}}_{X_k}$ is defined as, $\overline{\mathbb{P}}_{X_k} = \dfrac{1}{n_1} \sum_{i=1}^{n_1} \mathbb{P}_{X_k}^{(i)}$.

Characteristic RKHS kernel for permutations as quantity of interest. When GSA1 results $\mathcal{R}$ is a permutation (see Section 4.2.1.1), we propose to use Mallows kernel K_M (Jiao and Vert, 2018), the Mallows kernel is universal (and characteristic Mania et al., 2018). This kernel is given, for two permutations σ, σ' by:

$$K_M(\sigma, \sigma') = \exp\left(-\lambda n_d(\sigma, \sigma')\right), \tag{4.10}$$

where λ is a positive real parameter and n_d is the number of discordant pairs between σ and σ':

$$n_d(\sigma, \sigma') = \sum_{1 \leq r < s \leq d} \left[\mathbb{1}_{\{\sigma(r) < \sigma(s)\}} \mathbb{1}_{\{\sigma'(r) > \sigma'(s)\}} + \mathbb{1}_{\{\sigma(r) > \sigma(s)\}} \mathbb{1}_{\{\sigma'(r) < \sigma'(s)\}} \right]. \tag{4.11}$$

In practice, if a n_1-sample of σ is available, we propose to choose λ as the inverse of the empirical mean of $n_d(\sigma, \sigma')$ i.e.

$$\frac{1}{\lambda} = \frac{1}{n_1(n_1 - 1)} \sum_{i \neq j = 1}^{n_1} n_d(\sigma^{(i)}, \sigma^{(j)}),$$

where $\left(\sigma^{(i)}\right)_{1 \leq i \leq n_1}$ is the sample of permutations associated to the n_1-sample of distributions.

Characteristic RKHS kernel for real vectors as quantities of interest. In cases where $\mathcal{R}$ is a vector of $\mathbb{R}^d$, the usual Gaussian kernel defined in Formula (3.2) can be considered.

4.2.4 Possibilities for the unique sampling distribution

We propose three different possibilities for the single draw density $\overline{f}(x_1, \ldots, x_d) = \overline{f}_1(x_1) \times \ldots \times \overline{f}_d(x_d)$ which is used to generate the unique sample $\mathcal{E}$ in Step 1 of the algorithm (Section 4.2.2). Note here that the support of each $\overline{f}_k,\ k = 1, \ldots, d$ must be $\mathcal{X}_k$ (the variation domain of X_k, see Section 3.2). To have a density $\overline{f}_k$ close to the set of all possible densities and more particularly to the most likely ones, we propose to use either a mixture distribution or two barycenter distributions, namely the Wasserstein barycenter and the symmetric Kullback-Leibler barycenter. Figure 4.5 gives an example of these barycenter distributions.

Option 1: mixture distribution. The mixture density $\overline{f}_M$ (see e.g. Everitt and Hand, 1981; Titterington et al., 1985) of a random density probability f is defined by:

$$\overline{f}_M = \mathbb{E}_{\mathbb{F}}[f] = \int_{\mathcal{F}} f \, \mathrm{d}\mathbb{F}(f), \tag{4.12}$$

with f lying in $\mathcal{F}$ with probability distribution measure $\mathrm{d}\mathbb{F}$.

If $\mathbb{F}$ is discrete over a finite parametric set $\{f_{\theta_0}, \ldots, f_{\theta_m}\}$, the mixture density is written as

$$\overline{f}_M = \sum_{r=0}^{m} f_{\theta_r}\ \mathbb{F}(f_{\theta_r}). \tag{4.13}$$

If the density f depends on a parameter which is generated according to a continuous density π over Θ, the mixture density is defined by

$$\overline{f}_M = \int_{\Theta} f_\theta\ \pi(\theta)\ d\theta. \tag{4.14}$$

Option 2: symmetrized Kullback-Leibler barycenter. The symmetric Kullback-Leibler distance (Johnson and Sinanovic, 2001) is a distance obtained by symmetrizing the Kullback-Leibler divergence. It is defined for two real distributions μ and ν by:

$$D_K^S(\mu, \nu) = \frac{1}{2}\big(KL(\mu||\nu) + KL(\nu||\mu)\big), \tag{4.15}$$

where $KL(Q_1||Q_2) = \int \log(\frac{dQ_1}{dQ_2})dQ_1$ is the Kullback-Leibler divergence.

For a finite set of unidimensional and equiprobable densities $\{f_1, \ldots, f_m\}$, the symmetrized Kullback-Leibler barycenter $\overline{f}_K$ can not directly be expressed using densities. However, the distribution Q_K of density $\overline{f}_K$ defined by:

$$\overline{f}_K = \frac{1}{2m}\sum_{r=1}^{m} f_r + \frac{\left(\prod_{r=1}^{m} f_r\right)^{\frac{1}{m}}}{2\int \left(\prod_{r=1}^{m} f_r\right)^{\frac{1}{m}}}, \tag{4.16}$$

is a very good approximation of symmetrized Kullback-Leibler barycenter (see Veldhuis, 2002 for detailed proofs).

To generalize the Formula (4.16) to a probabilistic set of one-dimensional densities, we propose:

$$\overline{f}_K = \frac{1}{2}\mathbb{E}_{\mathbb{F}}\left[f\right] + \frac{e^{\mathbb{E}_{\mathbb{F}}[\ln f]}}{2\int e^{\mathbb{E}_{\mathbb{F}}[\ln f]}}. \tag{4.17}$$

where $\mathbb{E}_{\mathbb{F}}\left[f\right]$ and $\mathbb{E}_{\mathbb{F}}\left[\ln f\right]$ are mixture functions of random functions f and $\ln f$ (given by Equation 4.12).

Option 3: Wasserstein barycenter distribution. The Wasserstein distance (see e.g. Givens and Shortt, 1984; Villani, 2003) of order p between two real distributions μ and ν with the same support A is defined by:

$$W_p(\mu, \nu) = \left(\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{A\times A} |x - y|^p\, \mathrm{d}\gamma(x, y)\right)^{1/p}, \tag{4.18}$$

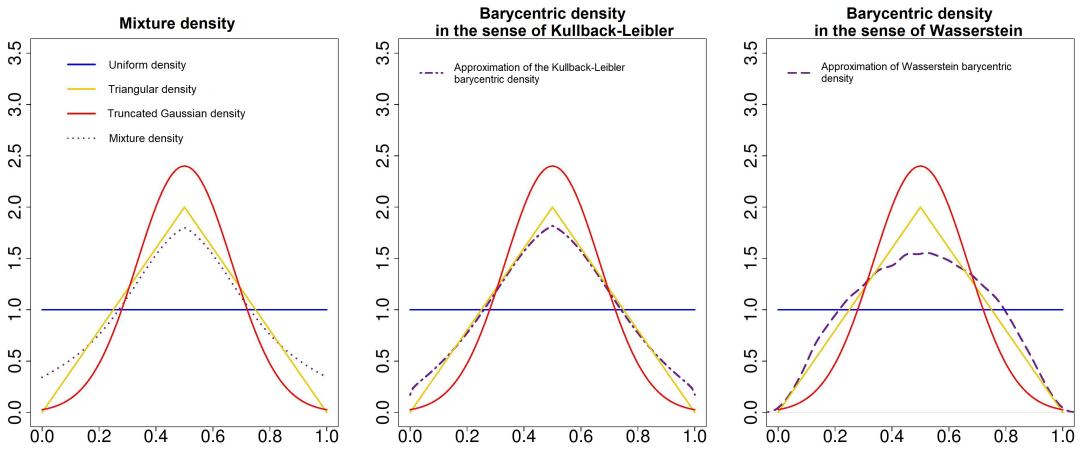

Figure 4.5 – Graphical representations of the mixture distribution, the barycentric distributions in the sense of Wasserstein and in the sense of Kullback-Leibler for three distributions: uniform, triangle with mode 0.5 and normal with mean 0.8 and standard deviation 0.1, which are equiprobable and defined on [0,1].

where $\Gamma(\mu, \nu)$ is the set of probability measures on $A \times A$ with marginals μ and ν.

Note that in the general case, when referring to the Wasserstein distance (without specifying the order) we refer to Wasserstein distance of order 2.

For a finite set $\mathcal{F}$ of unidimensional and equiprobable densities, the Wasserstein barycenter density (Agueh and Carlier, 2011) is the one whose quantile function is the mean of the quantile functions of the elements of the set $\mathcal{F}$:

$$\overline{q}_W = \frac{1}{|\mathcal{F}|} \sum_{f \in \mathcal{F}} q_f, \tag{4.19}$$

where $\overline{q}_W$ denotes the quantile function of Wasserstein barycenter, $|\mathcal{F}|$ the cardinal of the set $\mathcal{F}$ and q_f the quantile function associated to f.

To generalize Formula (4.19) to a probabilistic set of one-dimensional densities, we propose to use

$$\overline{q}_W = \mathbb{E}_{\mathbb{F}}[q_f], \tag{4.20}$$

where $\mathbb{E}_{\mathbb{F}}[q_f]$ is the mixture quantile function of the quantile functions $(q_f)_{f \in \mathcal{F}}$.

4.2.5 Discussion about the supports of the distributions

The reader might wonder if the methodology remains applicable when the potential distributions of inputs don't have a common support. Before answering this question, we mention that in real-applications, we generally deal with physical quantities having a bounded range of admissible values, linked to physical constraints (e.g. a chemical concentration lies between 0 and 1). For this reason, we use pdf with bounded support to represent the uncertainties of these quantities on their range of variations. Moreover, we focus here on equal supports but, it is also interesting to know if the methodology is still valid if we remove the hypothesis of common supports. We

already know that the only required condition to use the importance sampling technique is that the support of the prior distribution be contained in the support of the sampling one. Hence, in the case of different supports, choosing a sampling distribution with a support covering all those of the possible distributions resolves this limitation. This leads us to the second point: when supports are equal, all the points of the sample are used to estimate the HSIC for each assumed pdf. Conversely, a discordance of the supports leads to a loss of information, this loss being all the more important as the supports differ. To illustrate this, we consider the two extreme configurations given in Figure 4.6. On the left, the three densities have the same support, the budget of computation is used to compute the HSIC of each pdf. In contrast, on the right, the supports are totally disjoint (worst configuration) and the use of importance sampling technique is equivalent to dividing the simulation budget into three parts, only a third of simulations are used to estimate the HSIC for each case of pdf. Besides, in the case of disjoint supports, the "single loop" methodology turns into the usual "double loop" approach.

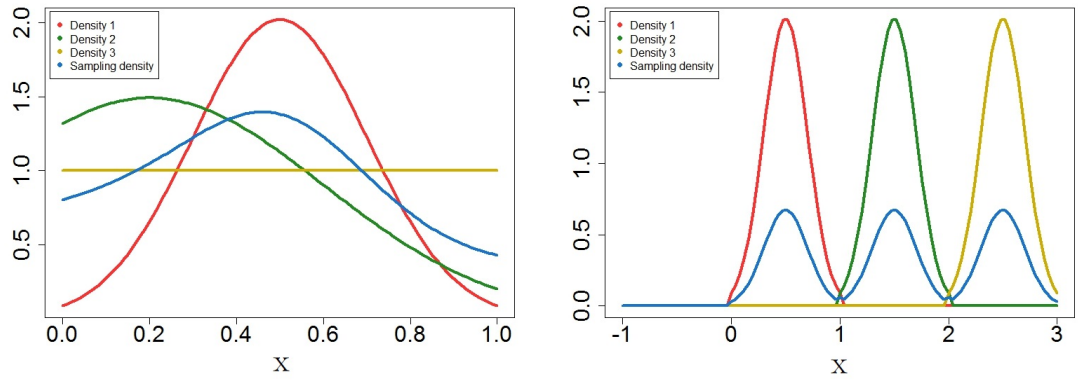

Figure 4.6 – Two examples of the sampling distribution for three possible densities. In the first case, the three densities have the same support, while in the second case, the three densities have totally disjoint supports.

4.3 Application of GSA2 methodology

In this part, our proposed methodology is first applied to an analytical model. The three different options proposed in Section 4.2.4 to generate the unique sample are studied and compared. Moreover, the benefit of this new methodology comparing to a "double loop" approach is highlighted. Thereafter, the whole methodology is applied to the nuclear ULOF-MACARENa test case.

4.3.1 Analytical example

Our proposed "single loop" methodology for GSA2 is first tested on the analytical model presented in Section 3.3.2.3. We recall that this model is defined on the set $[0,1]^3$ by $h(X_1, X_2, X_3) = \sin(X_1) + 1.5\ \sin^2(X_2) + 0.5\ X_3^4\ \sin(X_1)$. The inputs X_1, X_2 and X_3 are assumed here to be independent and their probability distributions $\mathbb{P}_{X_1}$, $\mathbb{P}_{X_2}$ et $\mathbb{P}_{X_3}$ can equiprobably be the laws $\mathbb{P}_U$, $\mathbb{P}_T$ et $\mathbb{P}_N$, where

- $\mathbb{P}_U$ is the uniform distribution on $[0,1]$,
- $\mathbb{P}_T$ is the triangular distribution on $[0,1]$ with mode 0.4,
- $\mathbb{P}_N$ is the truncated normal distribution on $[0,1]$ with mean 0.6 and standard deviation 0.2.

In practice, such second-level uncertainty as described above is encountered when there is a divergence of several expert opinions. More precisely, they all agree on the range of variation of an input but, their opinions differ on the type of the probability distribution. To illustrate this, consider here the three following expert opinions:

- The first expert claims that except the range of variation, no other information can be assumed on the uncertain variable;
- The second adds that the most likely value is 0.4;
- The third thinks that the mean and the standard deviation can respectively be assumed equal to 0.6 and 0.2.

Following the principle of maximum entropy for expert elicitation (Meyer and Booker, 2001; O'Hagan et al., 2006), the information provided by each expert can respectively be modeled by the distributions $\mathbb{P}_U$, $\mathbb{P}_T$ and $\mathbb{P}_N$. Thus, the uniform distribution on the set $\{\mathbb{P}_U, \mathbb{P}_T, \mathbb{P}_N\}$ seems to be a natural choice for the second-level uncertainty in this case.

The objective here is to estimate from a single inputs/output sample, the 2^{nd}-level GSA indices $\mathrm{R}^2_{\mathrm{HSIC}}(\mathbb{P}_{X_k}, \mathcal{R})_{k=1\ldots d}$ indices of the model $Y = h(\mathbf{X})$ for different sample sizes. For this, we use HSIC measures for GSA1 (and $\mathrm{R}^2_{\mathrm{HSIC}}(X_k, Y)_{k=1\ldots d}$ indices) with standardized Gaussian kernel. We characterize GSA1 results by the vector of 1^{st}-level $\mathrm{R}^2_{\mathrm{HSIC}}$ (option 1 in Section 4.2.2). To compute the 2^{nd}-level $\mathrm{R}^2_{\mathrm{HSIC}}$ indices, MMD-based kernels $l_{\mathcal{D}_k}$ (Equation 4.8) are used for input distributions and the standardized Gaussian kernel (Equation 3.2) is used for GSA1 results.

Remark. *The other quantities of interest characterizing GSA1 results could be studied in a similar way.*

4.3.1.1 Computation of theoretical values

In order to compute theoretical values of 2^{nd}-level HSIC and $\mathrm{R}^2_{\mathrm{HSIC}}$ indices, we consider the finite set of the $n_1 = 27$ possible triplet of input probability distributions. The 1^{st}-level $\mathrm{R}^2_{\mathrm{HSIC}}$ vector associated to each distribution is then computed with a sample of size $n_2 = 1000$ (which ensures the convergence of HSIC estimators). Theorical values of 2^{nd}-level HSIC are estimated with Formula (4.3):

- $\mathrm{HSIC}(\mathbb{P}_{X_1}, \mathcal{R}) = 0.0414$,
- $\mathrm{HSIC}(\mathbb{P}_{X_2}, \mathcal{R}) = 0.0261$,
- $\mathrm{HSIC}(\mathbb{P}_{X_3}, \mathcal{R}) = 0.0009$.

The theorical values of 2^{nd}-level $\mathrm{R}^2_{\mathrm{HSIC}}(\mathbb{P}_{X_k}, \mathcal{R})$ indices can also be computed:

- $\mathrm{R}^2_{\mathrm{HSIC}}(\mathbb{P}_{X_1}, \mathcal{R}) = 0.4152$,
- $\mathrm{R}^2_{\mathrm{HSIC}}(\mathbb{P}_{X_2}, \mathcal{R}) = 0.2516$,
- $\mathrm{R}^2_{\mathrm{HSIC}}(\mathbb{P}_{X_3}, \mathcal{R}) = 0.0086$.

We observe that $\mathrm{R}^2_{\mathrm{HSIC}}(\mathbb{P}_{X_1}, \mathcal{R})$ is considerably larger than $\mathrm{R}^2_{\mathrm{HSIC}}(\mathbb{P}_{X_2}, \mathcal{R})$, while $\mathrm{R}^2_{\mathrm{HSIC}}(\mathbb{P}_{X_3}, \mathcal{R})$ is negligible compared to the other two. In this example, the lack of knowledge on $\mathbb{P}_{X_3}$ has therefore no influence on 1^{st}-level $\mathrm{R}^2_{\mathrm{HSIC}}$. Furthermore, the uncertainty on $\mathbb{P}_{X_1}$ has a much higher impact than the one of $\mathbb{P}_{X_2}$, which remains non-negligible. Consequently, characterization efforts must be targeted in priority on $\mathbb{P}_{X_1}$, followed-up by $\mathbb{P}_{X_2}$.

4.3.1.2 GSA2 with our single loop approach

In the following, $\widetilde{\mathrm{HSIC}}_M(\mathbb{P}_{X_k}, \mathcal{R})$, $\widetilde{\mathrm{HSIC}}_W(\mathbb{P}_{X_k}, \mathcal{R})$ and $\widetilde{\mathrm{HSIC}}_K(\mathbb{P}_{X_k}, \mathcal{R})$, $k = 1, \ldots, 3$ denote the 2^{nd}-level HSIC measures respectively associated to mixture law, Wasserstein barycenter law and symmetrized Kullback-Leibler barycenter law. Similarly, $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},M}(\mathbb{P}_{X_k}, \mathcal{R})$, $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},W}(\mathbb{P}_{X_k}, \mathcal{R})$ and $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},K}(\mathbb{P}_{X_k}, \mathcal{R}), k = 1, \ldots, 3$ are the derived 2^{nd}-level $\mathrm{R}^2_{\mathrm{HSIC}}$ indices.

In this section, we apply the methodology proposed in Section 4.2.2 to estimate GSA2 HSIC-based indices. For this, we consider Monte Carlo samples of sizes $n_2 = 100$ to $n_2 = 1500$. The estimations are repeated independently 200 times from independent samples. The results obtained with the three modified laws are given by Figure 4.7. The theoretical values of $\mathrm{R}^2_{\mathrm{HSIC}}(\mathbb{P}_{X_k}, \mathcal{R})$ are represented in dotted lines. In this case, the estimators $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},M}(\mathbb{P}_{X_k}, \mathcal{R})$ and $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},K}(\mathbb{P}_{X_k}, \mathcal{R})$ seem to have similar behaviors for both small and higher sample sizes. The dispersion of these two estimators remains high for small sizes (especially for $n_2 \leq 200$) and becomes satisfying from $n_2 = 700$. The estimators $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},W}(\mathbb{P}_{X_k}, \mathcal{R})$ have a higher variance than the two previous estimators, particularly for small and medium sample sizes ($300 \leq n_2 \leq 700$).

In addition, we compare the ability of the three estimators to correctly order $\mathbb{P}_{X_k}, k = 1 \ldots 3$ by order of influence. For this, we compute for each sample size, the ratio of times when they give the good theoretical ranking. Table (4.1) gives the good ranking rates of 2^{nd}-level $\mathrm{R}^2_{\mathrm{HSIC}}$ estimators w.r.t the sample size. These results confirm that the estimators based on mixture and Kullback-Leibler barycenter laws outperform those based on Wasserstein barycenter law. Both $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},M}(\mathbb{P}_{X_k}, \mathcal{R})$ and $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},K}(\mathbb{P}_{X_k}, \mathcal{R})$ yield highly accurate ranking from $n_2 = 500$ against $n_2 = 700$ for $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},W}(\mathbb{P}_{X_k}, \mathcal{R})$. Furthermore, the Kullback-Leibler barycenter seems to give slightly better results for small samples $n_2 \leq 300$, this being reversed from $n_2 = 500$. The lower performance of Wasserstein barycenter law could be explained by the fact that the ratio $\frac{f}{f_W}$ becomes very high in the neighborhood of 0.

n_2	100	200	300	500	700	1000	1500
$\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},M}(\mathbb{P}_{X_k}, \mathcal{R})$	74%	79%	84%	94.5%	97%	100%	100%
$\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},K}(\mathbb{P}_{X_k}, \mathcal{R})$	75.5%	79%	87%	92%	97%	99.5%	99.5%
$\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},W}(\mathbb{P}_{X_k}, \mathcal{R})$	57.5%	71%	77%	82%	91%	93.5%	98%

Table 4.1 – Good ranking rates of $(\mathbb{P}_{X_i})_{i=1..3}$ given by the estimators $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},M}(\mathbb{P}_{X_k}, \mathcal{R})$, $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},K}(\mathbb{P}_{X_k}, \mathcal{R})$ and $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},W}(\mathbb{P}_{X_k}, \mathcal{R})$ for the model h, w.r.t the size n_2 of samples.

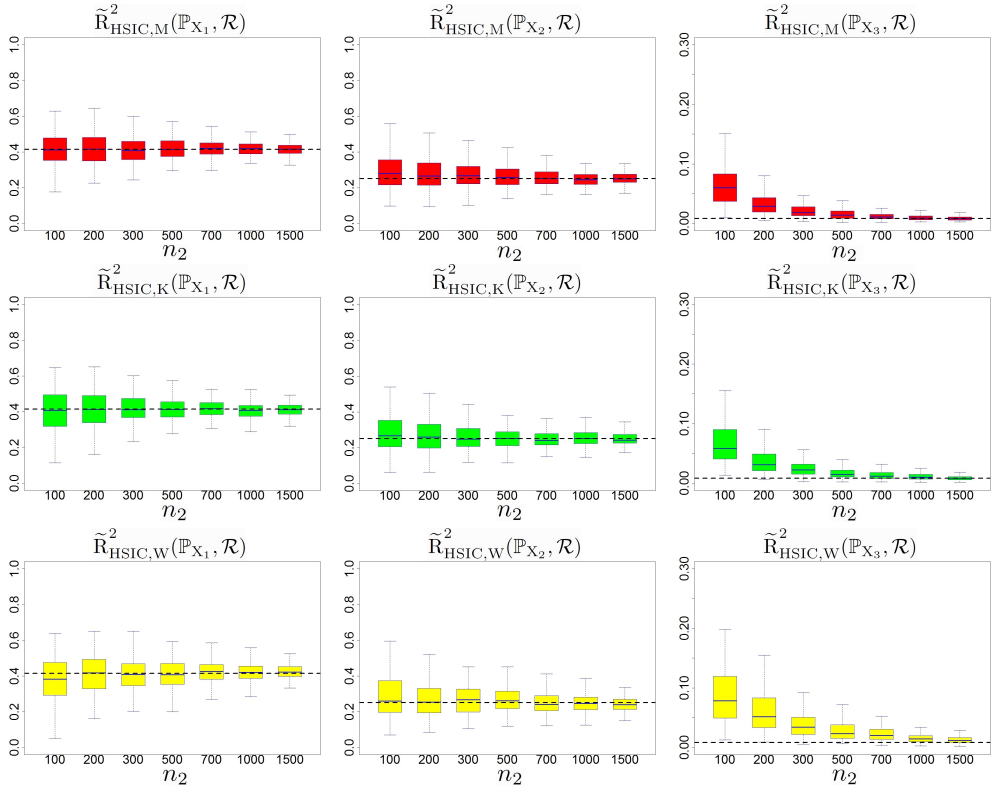

Figure 4.7 – Convergence plots of the estimators $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},M}(\mathbb{P}_{X_k},\mathcal{R})$, $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},W}(\mathbb{P}_{X_k},\mathcal{R})$ and $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},K}(\mathbb{P}_{X_k},\mathcal{R}), k = 1,\ldots,3$, for the model h and w.r.t the size n_2 of samples. Theoretical values are represented in dotted lines.

4.3.1.3 Comparison with Monte Carlo "double loop" approach

In this part, we compare the "single loop" estimation of 2^{nd}-level HSIC measures with the "double loop" estimation. For this, we consider a total budget $n = 1026$ simulations for both methods and propose the following test:

- **For the "double loop" approach,** a sample of size $n_2 = 38$ is generated for each triplet of input distributions ($n = n_1 \times n_2 = 1026$ simulations). The computed "double loop" estimators are denoted $\widehat{\mathrm{R}}^2_{\mathrm{HSIC}}(\mathbb{P}_{X_k},\mathcal{R}), k = 1\ldots 3$.
- **For the "single loop" approach,** we apply the proposed methodology with $n_2 = 1026$ to compute the "single loop" estimators $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},M}(\mathbb{P}_{X_k},\mathcal{R})$ and $\widetilde{\mathrm{R}}^2_{\mathrm{HSIC},K}(\mathbb{P}_{X_k},\mathcal{R})$, $k = 1\ldots 3$.

This numerical test is repeated 200 times with independent Monte Carlo samples. Figure 4.8 shows the dispersion of the obtained estimators. Theoretical values are shown in dotted lines. We observe that the "double loop" estimators have much more variability than "single loop" ones (especially for the distribution $\mathbb{P}_{X_3}$). We even observe a much larger bias (especially for $\mathbb{P}_{X_3}$) for the "double loop" approach. Good ranking rates are given by Table 4.2 and confirm that our proposed "single loop" approach significantly outperforms the "double loop" approach.

This example illustrates the interest of the "simple loop" approach which allows a much more accurate estimation of 2nd-level HSIC measures. Indeed, for a given total budget of n simulations, 1st-level HSIC are computed via modified HSIC from $n_2 = n$ simulations in our "single loop" approach against $n_2 = n/n_1$ in the "double loop" one. Even if classical estimators converge faster than modified ones, the number of simulations available for their estimation is drastically reduced with the double loop approach.

On this same analytical function, other numerical tests with different hypotheses on the input distributions (more different from each other) have been performed and yield similar results and conclusions.

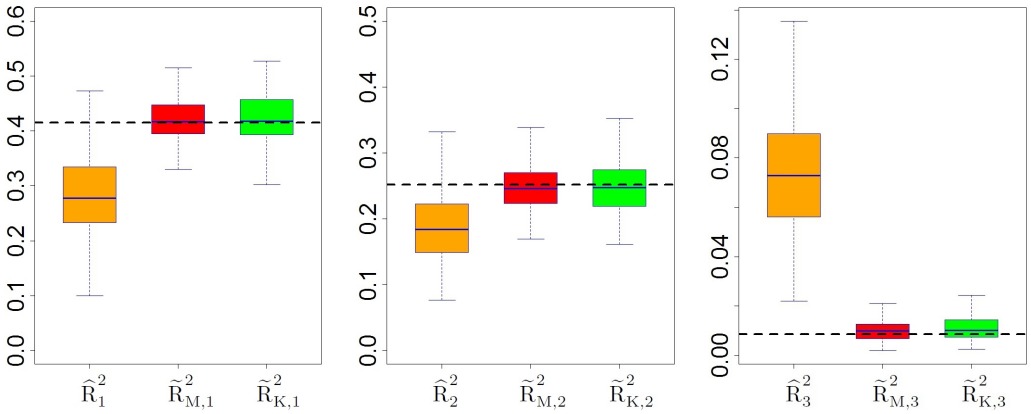

Figure 4.8 – Comparison of convergence plots of 2nd-level GSA indices by "double loop" approach ($\widehat{R}^2_k = \widehat{R}^2_{\text{HSIC}}(\mathbb{P}_{X_k}, \mathcal{R})$) and by "single loop" approach ($\widetilde{R}^2_{M,k} = \widetilde{R}^2_{\text{HSIC},M}(\mathbb{P}_{X_k}, \mathcal{R})$ and $\widetilde{R}^2_{K,k} = \widetilde{R}^2_{\text{HSIC},K}(\mathbb{P}_{X_k}, \mathcal{R})$) for the model h and $n = 1026$. Theorical values are represented in dotted lines.

Double loop	Single loop	
$\widehat{R}^2_{\text{HSIC}}(\mathbb{P}_{X_k}, \mathcal{R})$	$\widetilde{R}^2_{\text{HSIC},M}(\mathbb{P}_{X_k}, \mathcal{R})$	$\widetilde{R}^2_{\text{HSIC},K}(\mathbb{P}_{X_k}, \mathcal{R})$
67.5%	100%	99%

Table 4.2 – Comparison of good ranking rates of "double loop" and "single loop" estimators, for model h and $n = 1026$.

4.3.1.4 GSA2 using other quantities of interest

It is fair to wonder if GSA2 conclusions vary according to the chosen quantity of interest. For this, we focus here on the other quantities of interest previously defined. Namely, the ranking by R^2_{HSIC} indices and the P-values vector. In both cases, we consider the associated RKHS kernels as described in Section 4.2.3. Let us explore these two choices one-by-one.

• **Ranking by R^2_{HSIC}.** Since the convergence of 1st-level R^2_{HSIC} automatically implies the convergence of the ranking by these indices, the convergence of 2nd-level R^2_{HSIC} using the ranking

is at least as fast as the one using 1st-level R^2_{HSIC}. Based on the results of Section 4.3.1.2, we consider a unique sample of size $n_2 = 1000$ to accurately estimate the 2nd-level R^2_{HSIC} using the ranking. Denoting by $\mathcal{R}$ the input ranking, the obtained 2nd-level R^2_{HSIC} values are:

- $R^2_{HSIC}(\mathbb{P}_{X_1}, \mathcal{R}) = 0.3830$,
- $R^2_{HSIC}(\mathbb{P}_{X_2}, \mathcal{R}) = 0.0958$,
- $R^2_{HSIC}(\mathbb{P}_{X_3}, \mathcal{R}) \simeq 0$.

We notice that the gaps between these values are more significant than those using 1st-level R^2_{HSIC}. This is probably due to the stability of the ranking compared to the numerical values of 1st-level R^2_{HSIC}. Indeed, only significant shifts of 1st-level R^2_{HSIC} contribute to 2nd-level R^2_{HSIC} using the ranking. In this case, we can safely conclude that $\mathbb{P}_{X_1}$ is the major contributor of the ranking uncertainty. The robustess of GSA2 results can then considerably improved merely by reducing $\mathbb{P}_{X_1}$ uncertainties. Furthermore, as one can expect, less characterization efforts are required to increase the robustness of the ranking compared to 1st-level R^2_{HSIC}.

• **P-values vector.** By considering the p-values vector as the quantity of interest, two points are highlighted. Firstly, the 2nd-level R^2_{HSIC} estimators show a large variance, regardless of the estimation method of the p-values (i.e. by Gamma approximation or by permutations), even for very big sample sizes such as $n_2 = 5000$. In addition, the three estimated values are small (not exceeding 0.2). To better understand these results, we focus on the estimated p-values for each possible input distribution. For this, we use the permutation method with $B = 1000$ resamplings for each estimation. The obtained results show that the p-values associated to the inputs X_1 and X_2 are almost equal to zero (exactly zero numerically), regardless of the input distribution. Furthermore, the p-values associated to the input X_3 are low and in most cases below 10^{-5}. Thus, the high variance of 2nd-level R^2_{HSIC} is due to the difficulty of estimating the p-values for each given input distribution. Besides, GSA2 using the p-values is not relevant here. Indeed, the null hypothesis is not reliable in any case, GSA1 conclusions do not depend on the input distribution. The difference between these results and those based on 1nd-level R^2_{HSIC} can be explained by the fact that the p-values are mainly focused on assessing the confidence of the null hypothesis of each input considered individually, while 1nd-level R^2_{HSIC} are designed to compare the influence of inputs on the output.

4.3.2 Application on ULOF-MACARENa test case

Now, the objective is to apply the whole "single loop" methodology detailed in this chapter to the MACARENa simulator, presented in Chapters 1 and 2 (Sections 1.3.2 and 2.3.2 respectively), which simulates the ULOF (Unprotected Loss Of Flow) accident on sodium-cooled fast reactor. The simulator MACARENa has a total of 26 input parameters, which are either random or epistemic. In his work, Droin (2016) considered a total of 26 uncertain inputs of the simulator, which are either random or epistemic variables. They were assumed to be independent and their uncertainties were modeled with fixed probability distributions, listed in Table 4.3. Only first level of uncertainties were considered.

Among the outputs computed by the MACARENa simulator to describe the ULOF accident, the first instant of sodium boiling, denoted here Y, is of major interest for safety assessment. The preliminary GSA studies of Droin (2016) showed the predominant influence of 3 parameters on Y:

- X_1: external pressure drop discrepancy (Variable 4 in Table 4.3),

- X_2: primary mass flow rate (Variable 8 in Table 4.3),
- X_3: Lockart-Martinelli correction value (Variable 25 in Table 4.3).

However, due to lack of data and knowledge, uncertainty remains on the distributions $\mathbb{P}_{X_1}$, $\mathbb{P}_{X_2}$ and $\mathbb{P}_{X_3}$ respectively of X_1, X_2 and X_3. More precisely, in consultation with physical experts of the domain, we can reasonably assume (for these 3 inputs) the type of law as known but its parameters as uncertain, as described in Table 4.5. The notations $\mathcal{N}_t(a, b, m, \sigma)$, $\mathrm{T}(a, b, c)$ and $\mathcal{U}(a, b)$ are respectively, the truncated normal law of mean m and standard deviation σ on $[a, b]$, the triangular law on $[a, b]$ with mode c and the uniform law on $[a, b]$. The uncertainty on σ results from a prior knowledge (no available data), while the uncertainties on c and m are due to their estimation using few existing partial data.

In this framework, the objective is then to assess how each uncertainty on input pdf can impact the results of sensitivity analysis of Y.

Methodological choices. In order to perform GSA2, we apply our proposed algorithm with the following methodological choices (see Section 4.2.2):

- the unique sample for each input is generated according to the mixture law,
- the quantity of interest characterizing GSA1 results is the vector $\mathrm{R}^{\mathbf{2}}_{\mathrm{HSIC}}$,
- the RKHS kernel based on the MMD distance is used for input distributions and the standardized Gaussian kernel is used for GSA1 results.

Choices of sample sizes n_1 and n_2. We consider a Monte Carlo sample of size $n_2 = 1000$ for the unique sample. This choice is motivated by two main reasons, firstly the calculation time of one simulation of MACARENa (between 2 and 3 hours on average) which limits the total number of simulations and secondly the results obtained on the analytical three-dimensional example of Section 4.3.1. Furthermore, for the sample of distributions, we consider a Monte Carlo sample of $n_1 = 200$ triplets of pdf. These two choices for n_1 and n_2 will then be justified later in this section, by checking the convergence of estimators.

By applying our 2nd GSA methodology, with all these choices, we obtain the following 2nd-level sensitivity indices values:

- $\widetilde{\mathrm{R}}^{2}_{\mathrm{HSIC},M}(\mathbb{P}_{X_1}, \mathcal{R}) = 0.5341$,
- $\widetilde{\mathrm{R}}^{2}_{\mathrm{HSIC},M}(\mathbb{P}_{X_2}, \mathcal{R}) = 0.3317$,
- $\widetilde{\mathrm{R}}^{2}_{\mathrm{HSIC},M}(\mathbb{P}_{X_3}, \mathcal{R}) = 0.0753$.

Consequently, uncertainty on $\mathbb{P}_{X_1}$ mainly impacts GSA1 results, followed by $\mathbb{P}_{X_2}$, while $\mathbb{P}_{X_3}$ has a negligible impact. Therefore, the efforts of characterization must be targeted on $\mathbb{P}_{X_1}$ to improve the confidence in GSA1 results.

Remark. *A deeper analysis of the 200 GSA1 results shows that X_2 is almost all the time the predominant input (99% of cases). On the other hand, the rank of X_3 or X_1 varies: X_3 is the least influential input in 63% of cases, against 37% for X_1.*

Nominal core configuration					
Uncertain parameters	**Probability distribution**	**Distribution parameters**			
		P1	**P2**	**P3**	**P4**
1- Nominal core power (%P_{nom})	Truncated normal	1	3 σ	0.93	1.07
2- Irradiation cycle (-)	Discrete Uniform	1	2		
3- Cold sodium temperature in the secondary loops (K)	Truncated normal	618	3 σ	608	628
4- External pressure drop discrepancy (%)	Uniform	-0.1	+0.1		
5- Natural convection mass flow rate in the secondary loops (%)	Truncated normal	0.1	3 σ	0	0.2
6- Natural convection mass flow rate in the outer-subassembly flow (%)	Uniform	0.05	0.15		
7- Outer subassembly flow (%Q_{nom_ass})	Truncated normal	0.027	3 σ	0	0.054
Transient characteristics					
Uncertain parameters	**Probability distribution**	**Distribution parameters**			
		P1	**P2**	**P3**	**P4**
8- Primary mass flow rate halving time (s)	Triangular	5	10	15	
9- Secondary mass flow rate halving time (s)	Uniform	1	3		
Modelling uncertainties					
Uncertain parameters	**Probability distribution**	**Distribution parameters**			
		P1	**P2**	**P3**	**P4**
10- Steel penetration in the fertile inner fertile zone? (yes or no)	Discrete Uniform	0	1		
11- Upper steel plug thickness (m)	Uniform	0.01	0.07		
12- Discrepancy related to the Nusselt number calculation (%)	Truncated normal	0	3 σ	-0.2	+0.2
13- Discrepancy related to the calculation of axial fuel dilatation neutronic feedback coefficient (%)	Truncated normal	0	3 σ	-0.38	+0.38
14- Discrepancy related to the calculation of sodium dilatation/voiding neutronic feedback coefficient (%)	Truncated normal	0	3 σ	-0.30	+0.30
15- Discrepancy related to the calculation of the diagrid neutronic feedback coefficient (%)	Truncated normal	0	3 σ	-0.30	+0.30
16- Discrepancy related to the calculation of the Doppler neutronic constant (%)	Truncated normal	0	3 σ	-0.23	+0.23
17- Discrepancy related to the calculation of steel relocation neutronic feedback coefficient (%)	Truncated normal	0	3 σ	-0.30	+0.30
18- Discrepancy related to the calculation of fuel relocation neutronic feedback coefficient (%)	Truncated normal	0	3 σ	-0.30	+0.30
19- Discrepancy related to the calculation of thermal dilatation resulting in the control rods insertion in the core (%)	Truncated normal	0	3 σ	-0.30	+0.30
20- Discrepancy related to the residual power calculation (%)	Truncated normal	0	3 σ	-0.1	+0.1
21- Hot pool sodium volumic fraction affected by the ULOF transient (%)	Uniform	0.2	0.7		
22- Cold pool sodium volumic fraction affected by the ULOF transient (%)	Uniform	0.2	0.7		
23- Sodium vapour quality at dry-out (-)	Triangular	0.2	0.7	0.95	
24- Sodium overheating temperature (K)	Uniform	0	30		
25- Corrective factor applied on the Lockhart-Martinelli coefficient (-)	Uniform	0.5	3		
26- Discrepancy on the Wallis velocity calculation (%)	Uniform	-0.5	0.5		

Table 4.3 – List of the 26 uncertain inputs of MACARENa test case, as well as the associated probability distributions. The legend of the parameters of each distribution is presented in Table 4.4. The table is taken from Droin (2016).

Probability distribution	Distribution parameters			
	P1	P2	P3	P4
Uniform	Minimum	Maximum	-	-
Triangular	Minimum	Modal value	Maximum	-
Truncated Normal	Mean	Standard deviation	Minimum	Maximum

Table 4.4 – Parameters of uniform, triangular and truncated normal probability distributions of MACARENa test case. The table is taken from Droin (2016).

Law of input	Nature	Uncertain parameter
$\mathbb{P}_{X_1}$	$\mathcal{N}_t(-0.1, 0.1, 0, \sigma)$	$\sigma \sim \mathcal{U}(0.03, 0.05)$
$\mathbb{P}_{X_2}$	$\mathrm{T}(0, 20, c)$	$c \sim \mathcal{U}(8, 15)$
$\mathbb{P}_{X_3}$	$\mathrm{T}(0.8, 2, m)$	$m \sim \mathcal{U}(1, 1.5)$

Table 4.5 – Uncertainties on the laws $\mathbb{P}_{X_1}$,$\mathbb{P}_{X_2}$ and $\mathbb{P}_{X_3}$ for MACARENa test case.

In the light of GSA2 results, this alternation between the rank of X_3 and X_1 is therefore mainly driven by the uncertainty on $\mathbb{P}_{X_1}$, to a lesser extend by $\mathbb{P}_{X_2}$, while $\mathbb{P}_{X_3}$ has no impact. Moreover, X_2 whose distribution is not the most influential on GSA1 result, is surprisingly, the most influential input on Y. This example illustrates, if necessary, that GSA2 aims to capture an information that is different but complementary to that of GSA1.

In order to assess the accuracy of 2^{nd}-level $\mathrm{R}^2_{\mathrm{HSIC}}$ estimation, we use a non-asymptotic bootstrapping approach (see e.g. Efron and Tibshirani, 1994). For this, we first generate Monte Carlo subsamples with replacement from the initial sample (of 1000 simulations), then we re-estimate 2^{nd}-level $\mathrm{R}^2_{\mathrm{HSIC}}$ using these samples. We consider in particular subsamples of sizes $n_2 = 100$ to $n_2 = 800$. For each size, the estimation is repeated independently $B = 20$ times. Furthermore, to reduce computational efforts, we consider a sample of distributions of reduced size $n_1 = 30$ and generated with a *Space-Filling* approach. More precisely, the vector (σ, c, m) is sampled with a Maximum Projection Latin Hypercube Design (Joseph et al., 2015) of size $n_1 = 30$ and defined on the cubic domain $[0.03, 0.05] \times [8, 15] \times [1, 1.5]$.

Figure 4.9 presents as a boxplot the mismatch between the value estimated from the initial sample and the values estimated from subsamples. We first observe a robustness of estimation: the means of estimators seem to match the value given by the initial sample. We notice also high dispersions for small and medium sizes ($n_2 \leq 400$) and small dispersions for medium and big sizes ($n_2 \geq 500$). Therefore, it can be deduced that the estimations of GSA2 indices with the sample of $n_2 = 1000$ simulations have converged, the stabilization of the estimations being satisfactory from $n_2 = 700$.

We also test the robustness of the estimation in terms of ranking of input distributions. Table (4.6) gives for each subsample size, the rate of times that the ranking matches with the ranking obtained on the initial sample. The results given by Table (4.6) validate the conclusions drawn from convergence plots (4.9).

After considering the quantity of interest as the 1^{st}-level $\mathrm{R}^2_{\mathrm{HSIC}}$ vector, it would be interesting to know whether or not GSA2 conclusions change if we consider other quantities of interest. For

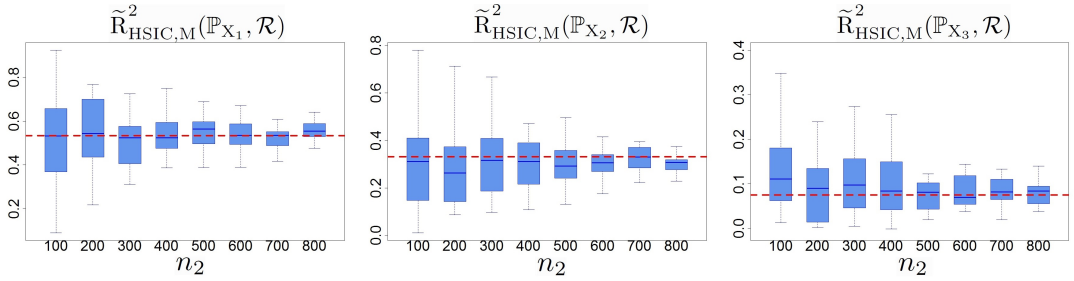

Figure 4.9 – Convergence plots of the estimators $\widetilde{R}^2_{HSIC,M}(\mathbb{P}_{X_k},\mathcal{R})$ for MACARENa test case, according to the sample size n_2. Theoretical values are represented in red dashed lines.

$n_2 = 100$	$n_2 = 200$	$n_2 = 300$	$n_2 = 400$	$n_2 = 500$	$n_2 = 600$	$n_2 \geq 700$
45%	55%	70%	75%	95%	95%	100%

Table 4.6 – Good ranking rates given by the estimators $\widetilde{R}^2_{HSIC,M}(\mathbb{P}_{X_k},\mathcal{R})$ for MACARENa test case w.r.t the size n_2 of the unique sample.

this, we present in Table 4.7 the 2nd-level R^2_{HSIC} indices using the three quantities of interest proposed in this chapter. We first remark that for the three chosen quantities of interest, the distribution $\mathbb{P}_{X_1}$ is always and significantly the predominant. In this application case, regardless of the chosen quantity of interest, the uncertainty on $\mathbb{P}_{X_1}$ has to be first and foremost reduced to stabilize GSA1 results.

$\mathcal{R}$	$\widetilde{R}^2_{HSIC,M}(\mathbb{P}_{X_1},\mathcal{R})$	$\widetilde{R}^2_{HSIC,M}(\mathbb{P}_{X_2},\mathcal{R})$	$\widetilde{R}^2_{HSIC,M}(\mathbb{P}_{X_3},\mathcal{R})$
Vector of R^2_{HSIC}	0.5341	0.3317	0.0753
Ranking by R^2_{HSIC}	0.2311	0.0848	0.0069
Vector of P-values	0.1821	0.0925	0.1064

Table 4.7 – The 2nd-level R^2_{HSIC} obtained according to each chosen quantity of interest, for MACARENa test case.

4.4 Conclusion and Prospect

In this chapter, we proposed a new methodology for second-level Global Sensitivity Analysis (GSA2) based on Hilbert-Schmidt Independence Criterion (HSIC). When input distributions are uncertain, GSA2 purpose is to assess the impact of these uncertainties on GSA results. In order to perform GSA2, we presented a new "single loop" Monte Carlo methodology to address problems raised by GSA2: characterization of GSA results, definition of 2nd-level HSIC measures and limitation of the calculation budget. This methodology calls for a single sample generated according to a "reference distribution" (related to the set of all possible distributions) and relies judiciously on the modified estimators introduced in Chapter 3 (Section 3.3.2). Three options have been proposed for the reference distribution: mixture law and barycentric laws w.r.t symmetrized Kullback-Leibler distance or Wasserstein distance. Analytical example show that

the estimation of 2^{nd}-level HSIC seems to be more accurate using the two first options rather than the Wasserstein barycenter. We also illustrated the great interest of the "single loop" approach compared to the "double loop" approach. Finally, the whole methodology has been applied to the MACARENa test case to take into account uncertain parameter of probability distributions of three inputs. This application illustrates how GSA2 can provide additional information to classical GSA.

Several points of the methodology could be more investigated in future research. First, we could focus on comparing *Space-Filling* Design (see e.g. Pronzato and Müller, 2012; Cioppa, 2002; Wang and Shan, 2007) techniques and Monte Carlo methods for the sampling of input distribution in the case of probabilistic densities (pdf) with uncertain parameters. Indeed, sampling the uncertain parameters of pdf following a *Space-Filling* design could improve the accuracy of the estimators of GSA2 indices. Another interesting perspective would be to build independence tests based on 2^{nd}-level HSIC estimators. This could be achieved by identifying the asymptotic distributions of these estimators under the assumption of independence between distributions and GSA1 results[2], or using the common permutation method.

Furthermore, this new approach for GSA2 could also be compared to the classical approach of epistemic GSA in the framework of Dempster-Shafer theory (see Smets, 1994; Alvarez, 2009). Indeed, Dempster-Shafer theory gives a description of random variables with epistemic uncertainty, which is to associate with an epistemic variable Z on a set A, a mass function representing a probability measure on the set $\mathcal{P}(A)$ of all A-subsets. This lack of knowledge is reflected in Dempster-Shafer theory by an upper and lower bound of the cumulative distribution function and can be viewed as 2^{nd}-level of uncertainty.

Another potential prospect could be to make the connection between our approach and Perturbed-Law based Indices (PLI, Lemaître et al., 2015). These indices are used to quantify the impact of a perturbation of an input density on the failure probability (probability that a model output exceeds a given threshold). To compare our GSA2 indices with PLI, the probability of failure could be considered as the quantity of interest characterizing GSA results in our methodology. Last but not least, GSA2 method can be compared to the approach proposed in Chabridon (2018) which models 2^{nd}-level uncertainties as a uni-level uncertainty on the vector (Θ, X), where Θ is the vector of uncertain parameters.

[2]Indeed, this is the null hypothesis for testing independence in the framework of GSA2.

Chapter 5

Aggregated tests of independence based on HSIC measures: theoretical properties and applications to Global Sensitivity Analysis

5.1 Issues and objectives

As previously explained in the Introduction 1.1 and 2.1, the screening can be the main objective of a GSA. To achieve this goal, statistical tests of independence based on HSIC measures are a efficient method. Depending on the available data size, HSIC tests can be performed either in an asymptotic or non-asymptotic framework (cf. Section 3.4). The asymptotic HSIC test proposed by Gretton et al. (2008) can be used when enough simulations of the code are available. This test is based on the asymptotic distribution of HSIC estimator under the assumption of independence. However, when the number of observations is too weak, the asymptotic distribution is no longer applicable. In this case, a non-asymptotic version of HSIC tests has to be developed, as introduced by De Lozzo and Marrel (2016b). Furthermore, the definition of HSIC measures as well as the associated tests requires a choice of RKHS kernels respectively associated to each input and to the output. One may then wonder about the choice of these kernels and their impact on the independence tests. Heuristic choices are adopted in practice but without any guarantee on the quality of the test.

In order to know what is a "relevant" choice of kernels (regarding the HSIC-based independence tests), one needs an objective evaluation criterion. As presented in Section 3.4.1.1, the criteria typically used to assess the quality of a statistical test are its *level* and *power*. The level is the probability of wrongly deciding the dependence (rejecting $\mathcal{H}_0$ while X and Y are independent) and the power is the probability of correctly deciding the dependence (rejecting $\mathcal{H}_0$ while X and Y are dependent). So, the power is directly linked to the second-kind error[1] (probability of complementary events). In practice, both HSIC-test versions, asymptotic and

[1] Probability of accepting the hypothesis $\mathcal{H}_0$ of independence while X and Y are dependent.

non-asymptotic, are built to control the level (frequently set at 5% or 10%). These tests have then a good ability to correctly detect the non-influential inputs. Therefore, a good remaining criterion for comparing tests of same levels but based on different kernel choices can be the power of the tests. This amounts to comparing the ability of tests to detect the influential inputs. Unfortunately, the optimal kernels (in terms of power) depend on the unknown joint distributions between each input and the output. **It will thus be valuable to further explore the link between these kernel choices and the power of associated tests, with the objective of providing a more robust methodology for input screening, and this, regardless of the number of available simulations.**

Theoretically speaking, since the power varies according to the unknown type of dependence between each input and the output, we propose to argue in terms of the non-asymptotic uniform separation rate introduced in Section 3.4.1.1. The first question to be addressed involves then the optimal HSIC test in the minimax sense (see Section 3.4.1.1 for the complete definition of minimax optimality). Theoretical advances to study the minimax testing in various contexts were proposed in many papers over the past years. Among them, we mention for example Ingster and Suslina (1998) and Laurent et al. (2012) for minimax signal detection testing. However, only few works exist for the problem of minimax independence testing. The notable works are those of Ingster (Ingster, 1989, 1993b) and those of Yodé (Yodé, 2004, 2011). Still, these works are provided in the asymptotic framework. As far as we know, no minimax rate of testing independence was yet proved in the non-asymptotic framework. **A primary goal of this chapter will be to provide a minimax theoretical test based on HSIC measure defined with Gaussian kernels.**

Beyond the problem of minimax rate for HSIC independence tests, the straightforward practical construction of a minimax test is impossible. Indeed, this construction depends on the unknown regularity of the joint density. The objective is then to construct a minimax test which does not need any smoothness property. These tests are generally called *minimax adaptive* (or assumption free). It has been shown that a standard logarithmic price is sometimes inevitable for adaptivity (Spokoiny, 1996). The problem of adaptivity has received a good attention in the literature. We mention for instance Baraud et al. (2003) for linear regression model testing with normal noise and Butucea and Tribouley (2006) for testing the equality of two samples densities. For the specific case of testing independence, the adaptive testing procedure introduced in Yodé (2011) seems to be the only currently existing. As mentioned above, this test is purely asymptotic, but we are interested here in the non-asymptotic framework. Recently, an interesting approach of testing developed in Fromont et al. (2013) consists in testing the equality of the intensities of two Poisson processes by aggregating several kernels in a unique testing procedure. It has been shown in Fromont et al. (2013) that this procedure is adaptive over several regularity spaces. Inspired by these works, and following the work of Gretton et al. (2008) and Gretton and Györfi (2010), **we consider in this chapter a procedure of testing independence which aggregates a collection of Gaussian-kernel HSIC tests.**

To study the adaptive properties of our procedure, we will consider the two popular spaces of regularity, namely Sobolev and Nikol'skii-Besov spaces. The proposed procedure is shown to be adaptive over the well-known *Sobolev spaces*. Moreover, the upper bound of its uniform separation rate over *Nikol'skii-Besov balls* seems optimal compared to "classical" testing rates in other frameworks. This suggests that this test may also be adaptive over Nikol'skii-Besov spaces.

The structure of this chapter is as follows. In Section 5.2, we first provide theoretical conditions to control the second-kind error of the non-asymptotic HSIC-test introduced in Section 3.4.3. This step leads us to sharply upper bound the uniform separation rate of any HSIC-test with a given Gaussian-kernel over *Sobolev* and *Nikol'skii-Besov* spaces. In Section 5.3, we

introduce our general HSIC-based aggregated procedure of testing, as well as its practical implementation. The aggregated procedure is shown to have the smallest possible upper bound of the uniform separation rate up to a small factor over Sobolev and Nikol'skii-Besov spaces. In addition, this upper bound is shown to be optimal over Sobolev spaces in Section 5.4, which means that the aggregated procedure is *adaptive* over these spaces. In Section 5.5, we first illustrate the procedure through some analytical examples. Different methodological choices are tested and compared. Subsequently, the performance of the procedure is compared to other "classical" independence tests. Finally, the methodology is applied to the test case ULOF-MACARENa introduced in Section 1.3 (in French) and 2.3 (in English). This real data case offers the opportunity to explore new methodological aspects of the procedure and to identify practical perspectives of improvement.

5.2 Performance of single HSIC-based tests of independence

All along this chapter, we will focus on HSIC measures based on Gaussian kernels. The objective here is to study the performance of single HSIC tests w.r.t. the bandwidth parameters of the chosen Gaussian kernels. This helps to better understand the behavior of HSIC tests w.r.t. the kernel parameters. For this, we will first provide theoretical guarantees for each single HSIC test to be at least of power β in $(0,1)$. Then, the final purpose of this first section will be to sharply upper bound the uniform separation rate w.r.t. the bandwidths. This allows to determine the optimal uniform separation rate to be achieved. As stated earlier, this optimal separation rate depends on the regularity of the joint density and, consequently, we will focus here on *Sobolev* and *Nikol'skii-Besov* regularity spaces.

5.2.1 Some notation and assumptions

For the sake of clarity and reduced complexity of demonstrations, we make some minor modifications to the assumptions and notations introduced so far. In particular, we consider a numerical simulator with d scalar inputs and q scalar outputs. We aim to study the independence of a subgroup of p inputs denoted $X = (X^{(1)}, \dots, X^{(p)})$ and the output denoted $Y = (Y^{(1)}, \dots, Y^{(q)})$. In addition, a sample $(X_i, Y_i)_{1 \leq i \leq n}$ of size n of (X, Y) is available. The random variables X and Y are assumed to be continuous with marginal densities respectively denoted f_1 and f_2, while the joint density of (X, Y) is denoted f with associated probability distribution P_f. We also denote by $f_1 \otimes f_2$, the product of the marginal densities f_1 and f_2 defined as

$$f_1 \otimes f_2 : (x, y) \in \mathbb{R}^p \times \mathbb{R}^q \mapsto f_1(x) f_2(y).$$

Let g_s be the density of the standard Gaussian distribution on $\mathbb{R}^s$ defined for all $x \in \mathbb{R}^s$ by

$$g_s(x) = \frac{1}{(2\pi)^{s/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{s} x_i^2\right). \tag{5.1}$$

For any bandwidths $\lambda = (\lambda_1, ..., \lambda_p) \in (0, +\infty)^p$ and $\mu = (\mu_1, ..., \mu_q) \in (0, +\infty)^q$, we define

for any $x \in \mathbb{R}^p$ and $y \in \mathbb{R}^q$,

$$\varphi_\lambda(x) = \frac{1}{\lambda_1 \dots \lambda_p} g_p\left(\frac{x_1}{\lambda_1}, \dots, \frac{x_p}{\lambda_p}\right), \tag{5.2}$$

$$\phi_\mu(y) = \frac{1}{\mu_1 \dots \mu_q} g_q\left(\frac{y_1}{\mu_1}, \dots, \frac{y_q}{\mu_q}\right). \tag{5.3}$$

We then define the Gaussian kernels, for $x, x' \in \mathbb{R}^p$ and $y, y' \in \mathbb{R}^q$,

$$k_\lambda(x, x') = \varphi_\lambda(x - x'), \quad l_\mu(y, y') = \phi_\mu(y - y'). \tag{5.4}$$

We denote by $\mathrm{HSIC}_{\lambda,\mu}(f)$ the HSIC measure defined in (3.1), where the kernels k and l are respectively the Gaussian kernels k_λ and l_μ. Furthermore, we consider here U-statistic estimators for $\mathrm{HSIC}_{\lambda,\mu}(f)$. For this, we introduce the following U-statistics, respectively of order 2, 3 and 4,

$$\widehat{\mathrm{HSIC}}^{(2)}_{\lambda,\mu} = \frac{1}{n(n-1)} \sum_{(i,j)\in \mathbf{i}_2^n} k_\lambda(X_i, X_j)\, l_\mu(Y_i, Y_j),$$

$$\widehat{\mathrm{HSIC}}^{(3)}_{\lambda,\mu} = \frac{1}{n(n-1)(n-2)} \sum_{(i,j,r)\in \mathbf{i}_3^n} k_\lambda(X_i, X_j)\, l_\mu(Y_j, Y_r),$$

and

$$\widehat{\mathrm{HSIC}}^{(4)}_{\lambda,\mu} = \frac{1}{n(n-1)(n-2)(n-3)} \sum_{(i,j,q,r)\in \mathbf{i}_4^n} k_\lambda(X_i, X_j)\, l_\mu(Y_q, Y_r),$$

where $\mathbf{i}_r^n$ is the set of all r-tuples drawn without replacement from the set $\{1, ..., n\}$. We then estimate $\mathrm{HSIC}_{\lambda,\mu}(f)$ by the U-statistic

$$\widehat{\mathrm{HSIC}}_{\lambda,\mu} = \widehat{\mathrm{HSIC}}^{(2)}_{\lambda,\mu} + \widehat{\mathrm{HSIC}}^{(4)}_{\lambda,\mu} - 2\widehat{\mathrm{HSIC}}^{(3)}_{\lambda,\mu}. \tag{5.5}$$

Finally, the theoretical independence test of level α introduced in Section 3.4.3 and associated to $\mathrm{HSIC}_{\lambda,\mu}(f)$ is denoted $\Delta_\alpha^{\lambda,\mu}$, while the permuted test is denoted $\widehat{\Delta}_\alpha^{\lambda,\mu}$.

5.2.2 Control of the second-kind error in terms of HSIC

For an arbitrarily small β given in $(0, 1)$, Lemma 5.1 provides a first non-asymptotic condition on the alternative f ensuring that the probability of second kind error of the theoretical test under P_f is at most equal to β. This condition is given for the value of $\mathrm{HSIC}_{\lambda,\mu}(f)$. It involves the variance of the estimator $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ which is finite since this estimator is a bounded random variable.

Lemma 5.1. *Let $(X_i, Y_i)_{1\le i\le n}$ be an i.i.d. sample with distribution P_f and consider the test statistic $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ defined by (5.5). Let α, β in $(0, 1)$, and $q_{1-\alpha}^{\lambda,\mu}$ be the $(1-\alpha)$-quantile of $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ under $P_{f_1\otimes f_2}$. Then $P_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu} \le q_{1-\alpha}^{\lambda,\mu}) \le \beta$ as soon as*

$$\mathrm{HSIC}_{\lambda,\mu}(f) \ge \sqrt{\frac{\mathrm{Var}_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu})}{\beta}} + q_{1-\alpha}^{\lambda,\mu}.$$

Lemma 5.1 gives a threshold for $\mathrm{HSIC}_{\lambda,\mu}(f)$ from which the dependence between X and Y is detectable with probability at least $1-\beta$ using Gaussian kernels k_λ and l_μ with given bandwidths λ and μ. Furthermore, it would be useful to give more explicit conditions w.r.t. the bandwidths λ and μ and the sample size n. The objective of this section is to provide a condition w.r.t. λ, μ and n on the theoretical value $\mathrm{HSIC}_{\lambda,\mu}$, so that the test $\Delta_\alpha^{\lambda,\mu}$ has a second kind error controlled by $\beta \in (0,1)$. For this, Lemma 5.1 already provides a condition involving $\mathrm{Var}_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu})$ and $q_{1-\alpha}^{\lambda,\mu}$. It is therefore necessary to establish sharp upper bounds for these two quantities w.r.t. λ, μ and n. Propositions 5.1 and 5.2 give these upper bounds.

Proposition 5.1. *Let $(X_i,Y_i)_{1\leq i\leq n}$ be an i.i.d. sample with distribution P_f and consider the test statistic $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ defined by (5.5). Assume that the densities f, f_1 and f_2 are bounded. Then,*

$$\mathrm{Var}_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu}) \leq C\left(M_f,p,q\right)\left\{\frac{1}{n}+\frac{1}{\lambda_1...\lambda_p\mu_1...\mu_q n^2}\right\},$$

where $M_f = \max\left(\|f\|_\infty,\|f_1\|_\infty,\|f_2\|_\infty\right)$.

Proposition 5.2. *Let $(X_i,Y_i)_{1\leq i\leq n}$ be an i.i.d. sample with distribution P_f and consider the test statistic $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ defined by (5.5). Let α in $(0,1)$ and $q_{1-\alpha}^{\lambda,\mu}$ be the $(1-\alpha)$-quantile of $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ under $P_{f_1\otimes f_2}$. Assuming that the densities f_1, f_2 are bounded,*

$$\max\left(\lambda_1...\lambda_p\ ,\ \mu_1...\mu_q\right)<1 \ \text{ and } \ n\sqrt{\lambda_1...\lambda_p\mu_1...\mu_q}\ >\log\left(\frac{1}{\alpha}\right)\ >1.$$

Then,

$$q_{1-\alpha}^{\lambda,\mu} \leq \frac{C\left(\|f_1\|_\infty,\|f_2\|_\infty,p,q\right)}{n\sqrt{\lambda_1...\lambda_p\mu_1...\mu_q}}\log\left(\frac{1}{\alpha}\right).$$

Combining Lemma 5.1, Propositions 5.1 and 5.2, we can then give a sufficient condition on $\mathrm{HSIC}_{\lambda,\mu}$ depending on the parameters λ, μ and the sample size n in order to control the second kind error by β. This result is presented in Corollary 5.1.

Corollary 5.1. *Let $(X_i,Y_i)_{1\leq i\leq n}$ be an i.i.d. sample with distribution P_f and consider the test statistic $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ defined by (5.5). Let α, β in $(0,1)$, and $q_{1-\alpha}^{\lambda,\mu}$ be the $(1-\alpha)$-quantile of $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ under $P_{f_1\otimes f_2}$. Assume that the densities f, f_1 and f_2 are bounded, and that*

$$\max\left(\lambda_1...\lambda_p\ ,\ \mu_1...\mu_q\right)<1 \ \text{ and } \ n\sqrt{\lambda_1...\lambda_p\mu_1...\mu_q}\ >\log\left(\frac{1}{\alpha}\right)\ >1.$$

Then, one has $P_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu} \leq q_{1-\alpha}^{\lambda,\mu}) \leq \beta$ as soon as

$$\mathrm{HSIC}_{\lambda,\mu}(f) > C\left(M_f,p,q,\beta\right)\left\{\frac{1}{\sqrt{n}}+\frac{1}{n\sqrt{\lambda_1...\lambda_p\mu_1...\mu_q}}\log\left(\frac{1}{\alpha}\right)\right\},$$

where $M_f = \max\left(\|f\|_\infty,\|f_1\|_\infty,\|f_2\|_\infty\right)$.

Note that the right hand term given in Corollary 5.1 is not computable in practice since it depends on the unknown density f. However, this dependence is weak since it only depends on the infinite norm of f and its marginals. For a given $\beta \in (0,1)$, Corollary 5.1 provides a condition on the value of $\mathrm{HSIC}_{\lambda,\mu}(f)$ ensuring that the probability of second kind error of the theoretical test under such f is at most equal to β. We now want to express such conditions in terms of the $\mathbb{L}_2$-norm of the function $f - f_1 \otimes f_2$, for the sake of interpretation, and in order to be able to determine separation rates with respect to this $\mathbb{L}_2$-norm for our test.

5.2.3 Control of the second-kind error in terms of $\mathbb{L}_2$-norm

In order to express a condition on the $\mathbb{L}_2$-norm of the difference $f - f_1 \otimes f_2$ ensuring a probability of second kind error controlled by β, we first give in Lemma 5.2 a link between $\mathrm{HSIC}_{\lambda,\mu}$ and $\|f - f_1 \otimes f_2\|_2^2$.

Lemma 5.2. *Let $\psi = f - f_1 \otimes f_2$. The* HSIC *measure $\mathrm{HSIC}_{\lambda,\mu}(f)$ associated to kernels k_λ and l_μ and defined in Equation (3.1) can be written as*

$$\mathrm{HSIC}_{\lambda,\mu}(f) = \langle \psi, \psi * (\varphi_\lambda \otimes \phi_\mu) \rangle_2,$$

where φ_λ and ϕ_μ are the functions respectively defined in Equations (5.2) and (5.3). Moreover, the notation $\langle ., . \rangle_2$ designates the usual scalar product in the space $\mathbb{L}_2$. One can easily deduce that

$$\mathrm{HSIC}_{\lambda,\mu}(f) = \frac{1}{2}\Big(\|\psi\|_2^2 + \|\psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 - \|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 \Big). \tag{5.6}$$

The Theorem 5.1 gives a sufficient condition on $\|f - f_1 \otimes f_2\|_2^2$, for the second kind error of the test $\Delta_\alpha^{\lambda,\mu}$ to be upper bounded by β.

Theorem 5.1. *Let $(X_i, Y_i)_{1 \le i \le n}$ be an i.i.d. sample with distribution P_f and consider the test statistic $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ defined by (5.5). Denote $\psi = f - f_1 \otimes f_2$. Let α, β in $(0,1)$, and $q_{1-\alpha}^{\lambda,\mu}$ be the $(1-\alpha)$-quantile of $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ under $P_{f_1 \otimes f_2}$. Assume that the densities f, f_1 and f_2 are bounded, and that*

$$\max(\lambda_1 ... \lambda_p \ , \ \mu_1 ... \mu_q) < 1 \ \text{ and } \ n\sqrt{\lambda_1 ... \lambda_p \mu_1 ... \mu_q} \ > \log\left(\frac{1}{\alpha}\right) \ > 1.$$

One has $P_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu} \le q_{1-\alpha}^{\lambda,\mu}) \le \beta$ as soon as

$$\|\psi\|_2^2 > \|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 + \frac{C(M_f, p, q, \beta)}{n\sqrt{\lambda_1 ... \lambda_p \mu_1 ... \mu_q}} \log\left(\frac{1}{\alpha}\right).$$

where $M_f = \max(\|f\|_\infty, \|f_1\|_\infty, \|f_2\|_\infty)$, and $C(M_f, p, q, \beta)$ denotes a positive constant depending only on its arguments.

In the condition given in Theorem 5.1, appears a compromise between a bias term $\|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2$ and a term induced by the square-root of the variance of the estimator $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$. Comparing the conditions on the HSIC given in Corollary 5.1 and on $\|f - f_1 \otimes f_2\|_2^2$ given in Theorem 5.1, the meticulous reader may notice that the term in $1/\sqrt{n}$ has been removed. This suppression seems to be necessary to obtain optimal separation rates according to the literature in other testing frameworks. This derives from quite tricky computations that we point out here and that directly prove Theorem 5.1. By combining Lemmas 5.1 and 5.2, direct computations lead to the condition

$$\|\psi\|_2^2 > \|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 - \|\psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 + 2\sqrt{\frac{\mathrm{Var}_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu})}{\beta}} + 2q_{1-\alpha}^{\lambda,\mu}. \tag{5.7}$$

If one directly considers the upper bound of the variance $\mathrm{Var}_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu})$ given in Proposition 5.1, one would get the unwanted $1/\sqrt{n}$ term. The idea is to take advantage of the negative term $-\|\psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2$ to compensate such term. To do so, we need a more refined control of the variance given in the technical Proposition 5.3.

Proposition 5.3. *Let $(X_i, Y_i)_{1\leq i\leq n}$ be an i.i.d. sample with distribution P_f and consider the test statistic $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ defined by (5.5). Assume that the densities f, f_1 and f_2 are bounded. Then,*

$$\mathrm{Var}_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu}) \leq C(M_f)\frac{\|\psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2}{n} + \frac{C\left(M_f, p, q\right)}{\lambda_1 ... \lambda_p \mu_1 ... \mu_q n^2},$$

where $M_f = \max\left(\|f\|_\infty, \|f_1\|_\infty, \|f_2\|_\infty\right)$.

Finally, using standard inequalities such as $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ and $2\sqrt{ab} \leq \delta a + b/\delta$ for all positive a, b and δ, one can prove

$$2\sqrt{\frac{\mathrm{Var}_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu})}{\beta}} \leq \|\psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 + \frac{C(M_f, \beta)}{n} + \frac{C\left(M_f, p, q, \beta\right)}{n\sqrt{\lambda_1 ... \lambda_p \mu_1 ... \mu_q}},$$

which leads to Theorem 5.1 when combined with Equation (5.7) and Proposition 5.2. Notice that such trick is already present in Fromont et al. (2013).

5.2.4 Uniform separation rate

The bias term in Theorem 5.1 comes from the fact that we do not estimate $\|f - f_1 \otimes f_2\|_2^2$ but $\mathrm{HSIC}_{\lambda,\mu}(f)$. In order to have a control of the bias term w.r.t λ and μ, we assume that $f - f_1 \otimes f_2$ belongs some class of regular functions. We introduce the two following classes: Sobolev balls (isotropic case) and Nikol'skii-Besov balls (anisotropic case).

5.2.4.1 Case Sobolev balls

For $d \in \mathbb{N}^*$, $\delta > 0$ and $R > 0$, the Sobolev ball $\mathcal{S}_d^\delta(R)$ is the set defined by

$$\mathcal{S}_d^\delta(R) = \left\{ s : \mathbb{R}^d \to \mathbb{R} \;\middle|\; s \in \mathbb{L}^1(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d), \int_{\mathbb{R}^d} \|u\|^{2\delta} |\hat{s}(u)|^2 du \leq (2\pi)^d R^2 \right\}, \tag{5.8}$$

where $\hat{s}$ denotes the Fourier transform of s defined by $\hat{s}(u) = \displaystyle\int_{\mathbb{R}^d} s(x) e^{i\langle x,u\rangle}\, \mathrm{d}x$, $\langle ., . \rangle$ denotes the usual scalar product in $\mathbb{R}^d$ and $\|.\|$ the Euclidean norm in $\mathbb{R}^d$.

Lemma 5.3 gives an upper bound for the bias term in the case when $f - f_1 \otimes f_2$ belongs to particular Sobolev balls.

Lemma 5.3. *Let $\psi = f - f_1 \otimes f_2$. We assume that $\psi \in \mathcal{S}_{p+q}^\delta(R)$, where $\delta \in (0, 2]$ and $\mathcal{S}_d^\delta(R)$ is defined by (5.8). Let φ_λ and ϕ_μ be the functions respectively defined in Equations (5.2) and (5.3). Then we have the following inequality,*

$$\|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 \leq C(p, q, \delta, R) \left[\sum_{i=1}^{p} \lambda_i^{2\delta} + \sum_{j=1}^{q} \mu_j^{2\delta} \right].$$

One can deduce from Theorem 5.1 upper bounds for the uniform separation rates (defined in (3.16)) of the test $\Delta_\alpha^{\lambda,\mu}$ over Sobolev balls.

Theorem 5.2. *Let $\alpha, \beta \in (0,1)$ and consider the same notation and assumptions as in Theorem 5.1. Let $\delta \in (0,2]$ and $R > 0$. Then, the uniform separation rate $\rho\left(\Delta_\alpha^{\lambda,\mu}, \mathcal{S}_{p+q}^{\delta}(R), \beta\right)$ defined in (3.16) over the Sobolev ball $\mathcal{S}_{p+q}^{\delta}(R)$ can be upper bounded as follows*

$$\left[\rho\left(\Delta_\alpha^{\lambda,\mu}, \mathcal{S}_{p+q}^{\delta}(R), \beta\right)\right]^2 \leq C(p,q,\delta,R)\left[\sum_{i=1}^{p}\lambda_i^{2\delta} + \sum_{j=1}^{q}\mu_j^{2\delta}\right] + \frac{C\left(M_f, p, q, \beta\right)}{n\sqrt{\lambda_1 \ldots \lambda_p \mu_1 \ldots \mu_q}}\log\left(\frac{1}{\alpha}\right), \tag{5.9}$$

where $M_f = \max\left(\|f\|_\infty, \|f_1\|_\infty, \|f_2\|_\infty\right)$, $C(M_f, p, q, \beta)$ and $C(p,q,\delta,R)$ are positive constants depending only on their arguments.

One can now determine optimal bandwidths (λ^*, μ^*) in order to minimize the right-hand side of Equation (5.9). To do so, the idea is to find for which (λ, μ) both terms in the right hand side of (5.9) are of the same order w.r.t. n. We also provide an upper bound for the uniform separation rate of the optimized test $\Delta_\alpha^{\lambda^*,\mu^*}$ on Sobolev balls.

Corollary 5.2. *Consider the assumptions of Theorem 5.2, and define for all i in $\{1,\ldots,p\}$ and for all j in $\{1,\ldots,q\}$,*

$$\lambda_i^* = \mu_j^* = n^{-2/(4\delta+p+q)}.$$

The uniform separation rate of the test $\Delta_\alpha^{\lambda^*,\mu^*}$ over the Sobolev ball $\mathcal{S}_{p+q}^{\delta}(R)$ is controlled as follows

$$\rho\left(\Delta_\alpha^{\lambda^*,\mu^*}, \mathcal{S}_{p+q}^{\delta}(R), \beta\right) \leq C\left(M_f, p, q, \alpha, \beta, \delta, R\right) n^{-2\delta/(4\delta+p+q)}. \tag{5.10}$$

Note that, in the definition of the Sobolev ball $\mathcal{S}_{p+q}^{\delta}(R)$, we have the same regularity parameter $\delta > 0$ for all the directions in $\mathbb{R}^{p+q}$. This corresponds to isotropic regularity conditions. We now introduce other classes of functions allowing to take into account possible anisotropic regularity properties.

5.2.4.2 Case of Nikol'skii-Besov balls

For $d \in \mathbb{N}^*$, $\delta = (\delta_1, ..., \delta_d) \in (0, +\infty)^d$ and $R > 0$, we consider the anisotropic Nikol'skii-Besov ball $\mathcal{N}_{2,d}^{\delta}(R)$ defined by

$$\mathcal{N}_{2,d}^{\delta}(R) = \Bigg\{ s : \mathbb{R}^d \to \mathbb{R} \;\Big|\; s \text{ has continuous partial derivatives } D_i^{\lfloor\delta_i\rfloor} \text{ of order } \lfloor\delta_i\rfloor \text{ w.r.t } u_i, \text{and}$$
$$\forall i = 1, ..., d,\ u_1, ..., u_d, v \in \mathbb{R}, \|D_i^{\lfloor\delta_i\rfloor} s(u_1, ..., u_i + v, ..., u_d) - D_i^{\lfloor\delta_i\rfloor} s(u_1, ..., u_d)\|_2 \leq R|v|^{\delta_i - \lfloor\delta_i\rfloor} \Bigg\}, \tag{5.11}$$

where $\lfloor\delta_i\rfloor$ denotes the floor function of δ_i if δ_i is not integer and $\lfloor\delta_i\rfloor = \delta_i - 1$ if δ_i is an integer. We give in the following proposition an upper bound of the bias term, similar to that of Lemma 5.3, in the case when $f - f_1 \otimes f_2$ belongs to particular Nikol'skii-Besov balls.

Lemma 5.4. *We assume that $\psi \in \mathcal{N}_{2,p+q}^{\delta}(R)$, where $\delta = (\nu_1, ..., \nu_p, \gamma_1, ..., \gamma_q) \in (0,2]^{p+q}$. Then, we have the following inequality,*

$$\|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 \leq C(R,\delta)\left[\sum_{i=1}^{p}\lambda_i^{2\nu_i} + \sum_{j=1}^{q}\mu_j^{2\gamma_j}\right].$$

As in Section 5.2.4.1, one can deduce from Theorem 5.1 upper bounds for the uniform separation rates of the test $\Delta_\alpha^{\lambda,\mu}$ over Nikol'skii-Besov balls.

Theorem 5.3. *Let $\alpha, \beta \in (0,1)$ and consider the same notation and assumptions as in Theorem 5.1. Let $\delta = (\nu_1, ..., \nu_p, \gamma_1, ..., \gamma_q) \in (0,2]^{p+q}$ and $R > 0$. Then, the uniform separation rate $\rho\left(\Delta_\alpha^{\lambda,\mu}, \mathcal{N}_{2,p+q}^{\delta}(R), \beta\right)$ defined in (3.16) over the Nikol'skii-Besov ball $\mathcal{N}_{2,p+q}^{\delta}(R)$ can be upper bounded as follows*

$$\left[\rho\left(\Delta_\alpha^{\lambda,\mu}, \mathcal{N}_{2,p+q}^{\delta}(R), \beta\right)\right]^2 \leq C(R,\delta)\left[\sum_{i=1}^{p}\lambda_i^{2\nu_i} + \sum_{j=1}^{q}\mu_j^{2\gamma_j}\right] + \frac{C\left(M_f, p, q, \beta\right)}{n\sqrt{\lambda_1 ... \lambda_p \mu_1 ... \mu_q}}\log\left(\frac{1}{\alpha}\right). \quad (5.12)$$

where $M_f = \max\left(\|f\|_\infty, \|f_1\|_\infty, \|f_2\|_\infty\right)$, $C\left(M_f, p, q, \beta\right)$ and $C(R,\delta)$ are positive constants depending only on their arguments.

As in Section 5.2.4.1, we now determine optimal bandwidths (λ^*, μ^*) which minimize the right-hand side of Equation (5.12) and compute an upper bound for the uniform separation rate of the optimized test $\Delta_\alpha^{\lambda^*,\mu^*}$ on Nikol'skii-Besov balls.

Corollary 5.3. *Consider the assumptions of Theorem 5.3, and define for all i in $\{1, \ldots, p\}$ and for all j in $\{1, \ldots, q\}$,*

$$\lambda_i^* = n^{-2\eta/[\nu_i(1+4\eta)]} \quad \text{and} \quad \mu_j^* = n^{-2\eta/[\gamma_j(1+4\eta)]},$$

where η is defined by $\dfrac{1}{\eta} = \displaystyle\sum_{i=1}^{p}\frac{1}{\nu_i} + \sum_{j=1}^{q}\frac{1}{\gamma_j}$,

The uniform separation rate of the test $\Delta_\alpha^{\lambda^*,\mu^*}$ over the Nikol'skii-Besov ball $\mathcal{N}_{2,p+q}^{\delta}(R)$ is controlled as follows

$$\rho\left(\Delta_\alpha^{\lambda^*,\mu^*}, \mathcal{N}_{2,p+q}^{\delta}(R), \beta\right) \leq C\left(M_f, p, q, \alpha, \beta, \delta\right) n^{-2\eta/(1+4\eta)}. \quad (5.13)$$

Notice that the upper bound obtained for Nikol'skii-Besov balls in Corollary 5.3 is analogue to that obtained for Sobolev balls in Corollary 5.2. Indeed, if we consider the same regularities in all directions in the case of Nikol'skii-Besov balls: $\nu_1 = \ldots = \nu_p = \gamma_1 = \ldots = \gamma_q$, we obtain a similar upper bound. These upper bounds obtained in Corollaries 5.2 and 5.3 remind the asymptotic minimax separation rate of testing independence w.r.t. the $\mathbb{L}_2$-norm over Hölder spaces Ingster (1989); Yodé (2004). However, the test having a rate with the smallest upper bound is not adaptive, it depends on the regularity parameter δ. In the next section, for the purpose of adaptivity, we build an aggregated testing procedure taking into account a collection of bandwidths $(\lambda, \mu) \in \Lambda \times U$. In particular, this avoids the delicate choice of arbitrary bandwidths. We then prove that the uniform separation rate of this aggregated procedure is of the same order as the smallest uniform separation rate of the chosen collection, up to a logarithmic term.

5.3 Aggregated non-asymptotic kernel-based test

In Section 5.2, we consider single tests based on Gaussian kernels associated to a particular choice of the bandwidths (λ, μ). However, applying such a procedure leads to the question of the choice

of these parameters. There is as yet no justified method to choose λ and μ. In many cases, authors choose these parameters w.r.t the available data $(X_i, Y_i)_{1\leq i\leq n}$ by taking for example λ (resp. μ) as the empirical median or standard deviation of the X_i's (resp. the Y_i's), which is not necessarily an optimal choice. To avoid this delicate choice, we consider in this section an aggregated testing procedure combining a collection of single tests based on different bandwithds.

5.3.1 The aggregated testing procedure

Consider now a finite or countable collection $\Lambda \times U$ of bandwidths in $(0,+\infty)^p \times (0,+\infty)^q$. Consider a collection of positive weights $\left\{\omega_{\lambda,\mu} \mid (\lambda,\mu) \in \Lambda \times U\right\}$ such that $\sum_{(\lambda,\mu)\in\Lambda\times U} e^{-\omega_{\lambda,\mu}} \leq 1$.

For a given $\alpha \in (0,1)$, we define the aggregated test which rejects $(\mathcal{H}_0)$ if there is at least one $(\lambda,\mu) \in \Lambda \times U$ such that

$$\widehat{\mathrm{HSIC}}_{\lambda,\mu} \;>\; q^{\lambda,\mu}_{1-u_\alpha e^{-\omega_{\lambda,\mu}}},$$

where u_α is the less conservative value such that the test is of level α, and is defined by

$$u_\alpha = \sup\left\{u>0 \;;\; P_{f_1\otimes f_2}\left(\sup_{(\lambda,\mu)\in\Lambda\times U}\left(\widehat{\mathrm{HSIC}}_{\lambda,\mu} - q^{\lambda,\mu}_{1-ue^{-\omega_{\lambda,\mu}}}\right)>0\right)\leq\alpha\right\}. \tag{5.14}$$

We should mention here that the supremum in Equation (5.14) exists. Indeed, one may notice that the function

$$u \mapsto P_{f_1\otimes f_2}\left(\sup_{(\lambda,\mu)\in\Lambda\times U}\left(\widehat{\mathrm{HSIC}}_{\lambda,\mu} - q^{\lambda,\mu}_{1-ue^{-\omega_{\lambda,\mu}}}\right)>0\right)$$

is well defined for u in the interval $(0, \inf\{\exp(\omega_{\lambda,\mu}); \lambda\in\Lambda, \mu\in U\})$, non-decreasing, and converges to 0 and 1 respectively at the boundaries of this interval.

The test function Δ_α associated to this aggregated test, takes values in $\{0,1\}$ and is defined by

$$\Delta_\alpha = 1 \quad\Longleftrightarrow\quad \sup_{(\lambda,\mu)\in\Lambda\times U}\left(\widehat{\mathrm{HSIC}}_{\lambda,\mu} - q^{\lambda,\mu}_{1-u_\alpha e^{-\omega_{\lambda,\mu}}}\right)>0. \tag{5.15}$$

It is easy to check that the test Δ_α is of level α, this is directly derived from the definition of u_α. Note that in order to guarantee the level of the aggregated procedure, the level α of each single test $\Delta^{\lambda,\mu}_\alpha$ is here replaced by the *corrected level* $u_\alpha \exp(-\omega_{\lambda,\mu})$.

For computational limitations, the collections Λ and U are finite in practice. Moreover, note that, as for the quantile, the correction u_α of the level is not analytically computable since it depends on the unknown marginals f_1 and f_2. In practice, it can also be approached by a permutation method with Monte Carlo approximation, as done in Albert (2015). More precisely, consider the notations of Section 3.4.2. First, generate B_1 independent and uniformly distributed random permutations of $\{1,...,n\}$, denoted $\tau_1,...,\tau_{B_1}$, independent of $\mathbb{Z}_n$ and compute for each $(\lambda,\mu)\in\Lambda\times U$ and each $u>0$ the permuted quantile with Monte Carlo approximation $\hat{q}^{\lambda,\mu}_{1-ue^{-\omega_{\lambda,\mu}}}$ as defined in (3.23).
Second, in order to estimate the probabilities under $P_{f_1\otimes f_2}$ in Equation (5.14), generate B_2 independent and uniformly distributed random permutations of $\{1,...,n\}$, denoted $\kappa_1,...,\kappa_{B_2}$, independent of $\mathbb{Z}_n$ and of $\tau_1,...,\tau_{B_1}$. Denote for all permutation κ_b, the corresponding permuted statistic

$$\widehat{H}^{\kappa_b}_{\lambda,\mu} = \widehat{\mathrm{HSIC}}_{\lambda,\mu}\left(\mathbb{Z}^{\kappa_b}_n\right)$$

Then, the correction u_α is approached by

$$\hat{u}_\alpha = \sup\left\{u>0\ ;\ \frac{1}{B_2}\sum_{b=1}^{B_2}\mathbb{1}_{\max_{(\lambda,\mu)\in\Lambda\times U}\left\{\widehat{H}^{\kappa_b}_{\lambda,\mu}-\hat{q}^{\lambda,\mu}_{1-ue^{-\omega_{\lambda,\mu}}}\right\}>0}\leq\alpha\right\}. \tag{5.16}$$

In the end, the aggregated testing procedure with permutation approach rejects the null hypothesis if

$$\max_{(\lambda,\mu)\in\Lambda\times U}\left(\widehat{\mathrm{HSIC}}_{\lambda,\mu}-\hat{q}^{\lambda,\mu}_{1-\hat{u}_\alpha e^{-\omega_{\lambda,\mu}}}\right)>0.$$

In the next section, we will provide a uniform separation rate similar to that of Corollaries 5.2 and 5.3 for the aggregated test Δ_α. This uniform separation rate will be given in the two cases mentioned earlier in Section 5.2.4 where $f-f_1\otimes f_2$ belongs to isotropic Sobolev balls or to anisotropic Nikol'skii-Besov balls.

5.3.2 Oracle type conditions for the second-kind error

As a reminder, our goal is to construct a testing procedure with a uniform separation rate as small as possible and whose implementation does not require any information about the regularity of the difference $f-f_1\otimes f_2$.

The main advantage of the aggregated procedure is that its second kind error is as small as the one of the single test corresponding to the best bandwidths in the collection $\Lambda\times U$ with a corrected level. The main argument is highlighted in Lemma 5.5.

Lemma 5.5. *Let α, β in $(0,1)$, and consider the aggregated test Δ_α defined in (5.15). Then, $u_\alpha\geq\alpha$ and*

$$P_f\left(\Delta_\alpha=0\right)\leq\inf_{(\lambda,\mu)\in\Lambda\times U}\left\{P_f\left(\Delta^{\lambda,\mu}_{\alpha e^{-\omega_{\lambda,\mu}}}=0\right)\right\}.$$

According to Lemma 5.5, if there exists at least one $(\lambda,\mu)\in\Lambda\times U$ such that the associated single test $\Delta^{\lambda,\mu}_{\alpha e^{-\omega_{\lambda,\mu}}}$ has a probability of second kind error at most equal to β, then the probability of the second kind error of the aggregated test Δ_α is at most equal to β.

We now give an oracle inequality for the uniform separation rate of the aggregation procedure Δ_α. This inequality given in the following theorem shows the interest of this testing procedure.

Theorem 5.4. *Let $\alpha,\beta\in(0,1)$, $\left\{(k_\lambda,l_\mu)\ \middle|\ (\lambda,\mu)\in\Lambda\times U\right\}$ a collection of Gaussian kernels and $\left\{\omega_{\lambda,\mu}\ \middle|\ (\lambda,\mu)\in\Lambda\times U\right\}$ a collection of positive weights, such that $\sum_{(\lambda,\mu)\in\Lambda\times U}e^{-\omega_{\lambda,\mu}}\leq 1$. We also assume that all bandwidths (λ,μ) in $\Lambda\times U$ verify the conditions given in Theorem 5.1, and that f, f_1 and f_2 are bounded. Then, the test Δ_α of level α defined in Equation (5.15) has a uniform separation rate $\rho\left(\Delta_\alpha,C_\delta,\beta\right)$ which can be upper bounded as follows*

- *If $C_\delta=\mathcal{S}^\delta_{p+q}(R)$, where $\delta\in(0,2]$ and $R>0$, then*

$$\left[\rho\left(\Delta_\alpha,\mathcal{S}^\delta_{p+q}(R),\beta\right)\right]^2\leq C\left(M_f,p,q,\beta,\delta\right)\inf_{(\lambda,\mu)\in\Lambda\times U}\left\{\frac{1}{n\sqrt{\lambda_1\ldots\lambda_p\mu_1\ldots\mu_q}}\left(\log\left(\frac{1}{\alpha}\right)+\omega_{\lambda,\mu}\right)+\left[\sum_{i=1}^p\lambda_i^{2\delta}+\sum_{j=1}^q\mu_j^{2\delta}\right]\right\},$$

where $M_f=\max\left(\|f\|_\infty,\|f_1\|_\infty,\|f_2\|_\infty\right)$ and $C\left(M_f,p,q,\beta,\delta\right)$ is a positive constant depending only on its arguments.

- *If* $C_\delta = \mathcal{N}^\delta_{2,p+q}(R)$, *where* $\delta = (\nu_1, ..., \nu_p, \gamma_1, ..., \gamma_q) \in (0,2]^{p+q}$ *and* $R > 0$, *then*

$$\left[\rho\left(\Delta_\alpha, \mathcal{N}^\delta_{2,p+q}(R), \beta\right)\right]^2 \leq C\left(M_f, p, q, \beta, \delta\right) \inf_{(\lambda,\mu)\in\Lambda\times U} \left\{ \frac{1}{n\sqrt{\lambda_1...\lambda_p\mu_1...\mu_q}} \left(\log\left(\frac{1}{\alpha}\right) + \omega_{\lambda,\mu}\right) + \left[\sum_{i=1}^{p} \lambda_i^{2\nu_i} + \sum_{j=1}^{q} \mu_j^{2\gamma_j}\right] \right\},$$

where $C\left(M_f, p, q, \beta, \delta\right)$ *is a positive constant depending only on its arguments.*

According to Theorem 5.4, the uniform separation rate of the aggregated testing procedure Δ_α is the infimum of all $(\lambda, \mu) \in \Lambda \times U$, up to the additional term $\omega_{\lambda,\mu}$. This theorem can also be interpreted as an oracle type condition for the second kind error of the test Δ_α. Indeed, without knowing the regularity of $f - f_1 \otimes f_2$, we prove that the uniform separation rate of Δ_α is of the same order as the smallest uniform separation rate over $(\lambda, \mu) \in \Lambda \times U$, up to $\omega_{\lambda,\mu}$.

5.3.3 Uniform separation rate over Sobolev balls and Nikol'skii-Besov balls

In this section, we provide an upper bound of the uniform separation rate $\rho\left(\Delta_\alpha, C_\delta, \beta\right)$ of the multiple testing procedure Δ_α over the classes of Sobolev balls and Nikol'skii-Besov balls. For this, we consider the collections Λ and U of parameters λ and μ respectively, defined by

$$\Lambda = \{(2^{-m_{1,1}}, \ldots, 2^{-m_{1,p}}) \ ; \ (m_{1,1}, \ldots, m_{1,p}) \in (\mathbb{N}^*)^p\}, \tag{5.17}$$

and

$$U = \{(2^{-m_{2,1}}, \ldots, 2^{-m_{2,q}}) \ ; \ (m_{2,1}, \ldots, m_{2,q}) \in (\mathbb{N}^*)^q\}. \tag{5.18}$$

In addition, we associate to every $\lambda = (2^{-m_{1,1}}, \ldots, 2^{-m_{1,p}})$ in Λ and $\mu = (2^{-m_{2,1}}, \ldots, 2^{-m_{2,q}})$ in U the positive weights

$$\omega_{\lambda,\mu} = 2\sum_{i=1}^{p} \log\left(m_{1,i} \times \frac{\pi}{\sqrt{6}}\right) + 2\sum_{j=1}^{q} \log\left(m_{2,j} \times \frac{\pi}{\sqrt{6}}\right), \tag{5.19}$$

so that $\sum_{(\lambda,\mu)\in\Lambda\times U} e^{-\omega_{\lambda,\mu}} = 1$. Corollary 5.4 provides these upper bounds.

Corollary 5.4. *Assuming that* $\log\log(n) > 1$, $\alpha, \beta \in (0,1)$ *and* Δ_α *the test defined in* (5.15), *with the particular choice of* Λ, U *and the weights* $(\omega_{\lambda,\mu})_{(\lambda,\mu)\in\Lambda\times U}$ *defined in* (5.17), (5.18) *and* (5.19). *Then, the uniform separation rate* $\rho\left(\Delta_\alpha, C_\delta, \beta\right)$ *of the aggregated test* Δ_α *can be upper bounded as follows.*

- *If* $C_\delta = \mathcal{S}^\delta_{p+q}(R)$, *where* $\delta \in (0,2]$ *and* $R > 0$, *then,*

$$\rho\left(\Delta_\alpha, \mathcal{S}^\delta_{p+q}(R), \beta\right) \leq C\left(M_f, p, q, \alpha, \beta, \delta\right) \left(\frac{\log\log(n)}{n}\right)^{2\delta/(4\delta+p+q)}, \tag{5.20}$$

where $M_f = \max\left(\|f\|_\infty, \|f_1\|_\infty, \|f_2\|_\infty\right)$.

- *If* $C_\delta = \mathcal{N}^\delta_{2,p+q}(R)$, *where* $\delta = (\nu_1, \dots, \nu_p, \gamma_1, \dots, \gamma_q) \in (0,2]^{p+q}$ *and* $R > 0$, *then,*

$$\rho\left(\Delta_\alpha, \mathcal{N}^\delta_{2,p+q}(R), \beta\right) \leq C\left(M_f, p, q, \alpha, \beta, \delta\right) \left(\frac{\log\log(n)}{n}\right)^{2\eta/(1+4\eta)}, \tag{5.21}$$

where $\dfrac{1}{\eta} = \displaystyle\sum_{i=1}^{p} \frac{1}{\nu_i} + \sum_{j=1}^{q} \frac{1}{\gamma_j}$ *and* $M_f = \max\left(\|f\|_\infty, \|f_1\|_\infty, \|f_2\|_\infty\right)$.

According to Corollary 5.4, the rate of the aggregation procedure over the classes of Sobolev balls and Nikol'skii-Besov balls is in the same order of the best rate of single tests (given in Theorem 5.1), up to a $\log\log(n)$ factor.

5.4 Lower bound for uniform separation rates over Sobolev balls.

In this section, we study the optimality of the single test introduced in Corollary 5.2 and of the aggregated testing procedure defined in Corollary 5.4 over Sobolev balls. For this, we first present a general method based on a Bayesian approach to lower bound the non-asymptotic minimax separation rate of testing as defined in (3.17). The general idea of this method is due to Ingster (1993a) and relies on Lemma 5.6.

Lemma 5.6. *Let* α, β *in* $(0,1)$ *such that* $\alpha + \beta < 1$. *Let* $\rho_* > 0$ *and* $\mathcal{C}_\delta$ *some regularity space. We recall that the set* $\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)$ *is the set defined as*

$$\mathcal{F}_{\rho_*}(\mathcal{C}_\delta) = \{f; f - f_1 \otimes f_2 \in \mathcal{C}_\delta, \|f - f_1 \otimes f_2\|_2 \geq \rho_*\}.$$

Let ν_{ρ_*} *be a probability measure on* $\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)$ *and* $P_{\nu_{\rho_*}}$ *the probability measure defined for all measurable set* A *in* $\mathbb{R}^{p+q}$ *by*

$$P_{\nu_{\rho_*}}(A) = \int P_f(A)\, \mathrm{d}\nu_{\rho_*}(f).$$

Denote for all $\rho > 0$,

$$\beta\big[\mathcal{F}_\rho(\mathcal{C}_\delta)\big] = \inf_{\Delta_\alpha} \sup_{f \in \mathcal{F}_\rho(\mathcal{C}_\delta)} P_f\left(\Delta_\alpha = 0\right),$$

where the infimum is taken over all α*-level tests of* $(\mathcal{H}_0)$ *against* $(\mathcal{H}_1)$. *Assume there exists a distribution* f_0 *that satisfies* $(\mathcal{H}_0)$ *such that the probability measure* $P_{\nu_{\rho_*}}$ *is absolutely continuous w.r.t.* P_{f_0} *and verifies*

$$\mathbb{E}_{P_{f_0}}\left[L^2_{\nu_{\rho_*}}(\mathbb{Z}_n)\right] < 1 + 4(1-\alpha-\beta)^2, \tag{5.22}$$

where the likelihood ratio $L_{\nu_{\rho_*}}$ *is defined by* $L_{\nu_{\rho_*}} = \dfrac{\mathrm{d}P_{\nu_{\rho_*}}}{\mathrm{d}P_{f_0}}$. *Then, for all* $\rho \leq \rho_*$ *we have that*

$$\beta\big[\mathcal{F}_\rho(\mathcal{C}_\delta)\big] > \beta.$$

It follows that

$$\rho\left(\mathcal{C}_\delta, \alpha, \beta\right) = \inf_{\Delta_\alpha} \rho\left(\Delta_\alpha, \mathcal{C}_\delta, \beta\right) \geq \rho_*.$$

We aim at proving that

$$\rho_n^* = Cn^{-2\delta/(4\delta+p+q)}$$

is a lower bound for the non-asymptotic minimax rate of testing over Sobolev balls $\mathcal{S}_{p+q}^{\delta}(R)$, for some positive constant C, that is, $\rho\left(\mathcal{S}_{p+q}^{\delta}(R), \alpha, \beta\right) \geq \rho_n^*$. According to Lemma 5.6, it is sufficient to find a probability distribution $\nu_{\rho_n^*}$ over $\mathcal{F}_{\rho_n^*}(\mathcal{S}_{p+q}^{\delta}(R))$ so that the condition (5.22) holds.

To do so, we generalize the construction of Butucea (2007) to our multidimensional framework. The idea is to construct a finite set of alternatives $(f_\theta)_\theta$ by perturbing the uniform density on $[0,1]^p \times [0,1]^q$, and define $\nu_{\rho_n^*}$ as a uniform mixture of these alternatives. For this, consider the function G defined for all t in $\mathbb{R}$ by

$$G(t) = \exp\left(-\frac{1}{1-(4t+3)^2}\right)\mathbb{1}_{(-1,-1/2)}(t) - \exp\left(-\frac{1}{1-(4t+1)^2}\right)\mathbb{1}_{(-1/2,0)}(t). \tag{5.23}$$

One may notice that G is continuous, with support in $[-1,0]$ and that $\int_{\mathbb{R}} G(t)\mathrm{d}t = 0$. The function G together with its Fourier transform has valuable properties for our study.

Let $(h_n)_n$ be a sequence of positive numbers to be specified later, and consider an integer M_n such the $M_n h_n = 1$ (possibly rounded to the nearest integer). Denote $I_{n,p,q} = \{1,\dots,M_n\}^p \times \{1,\dots,M_n\}^q$. For all $\theta = (\theta_{(j,l)})_{(j,l)\in I_{n,p,q}}$ in $\{-1,1\}^{M_n^{p+q}}$, define for all (x,y) in $\mathbb{R}^p \times \mathbb{R}^q$,

$$f_\theta(x,y) = \mathbb{1}_{[0,1]^{p+q}}(x,y) + h_n^{\delta+(p+q)} \sum_{(j,l)\in I_{n,p,q}} \theta_{(j,l)} \prod_{r=1}^{p} G_{h_n}(x_r - j_r h_n) \prod_{s=1}^{q} G_{h_n}(y_s - l_s h_n), \tag{5.24}$$

where for all $h>0$, $G_h(\cdot) = (1/h)\,G(\cdot/h)$. One may notice that for all θ, the alternative f_θ is supported in $[0,1]^{p+q}$. Moreover, since the integral of G over $\mathbb{R}$ equals 0, the marginals $f_{\theta,1}$ and $f_{\theta,2}$ of f_θ are respectively the uniform densities on $[0,1]^p$ and $[0,1]^q$. Proposition 5.4 justifies the choice of these alternatives.

Proposition 5.4. *Let $\delta > 0$ and $R > 0$. Fix a sequence $(h_n)_n$ of positive numbers and consider an integer M_n such the $M_n h_n = 1$. Then, for all $\theta = (\theta_{(j,l)})_{(j,l)\in I_{n,p,q}}$ in $\{-1,1\}^{M_n^{p+q}}$, the function f_θ defined in Equation (5.24) satisfies the following properties.*

1. *The function f_θ is a density function for h_n small enough.*
2. *The function $f_\theta - f_{\theta,1}\otimes f_{\theta,2}$ belongs to the Sobolev ball $\mathcal{S}_{p+q}^{\delta}(R)$ for n large enough.*
3. *The function f_θ is such that $\|f_\theta - f_{\theta,1}\otimes f_{\theta,2}\|_{\mathbb{L}_2} = C(p,q)h_n^{\delta}$.*

Let us now consider a uniform mixture $\nu_{\rho_n^*}$ of the alternatives (f_θ), for θ in $\{-1,1\}^{M_n^{p+q}}$. Note that this is equivalent to considering a random alternative f_Θ where $\Theta = (\Theta_{(j,l)})_{(j,l)\in I_{n,p,q}}$ with i.i.d. Rademacher components $\Theta_{(j,l)}$.
Following Lemma 5.6, let P_ν be the probability measure defined for all measurable set A in $\mathbb{R}^{p+q}$ by

$$P_\nu(A) = \int_{\{-1,1\}^{M_n^{p+q}}} P_{f_\theta}(A)\pi(\mathrm{d}\theta) = \frac{1}{2^{M_n^{p+q}}} \sum_{\theta\in\{-1,1\}^{M_n^{p+q}}} P_{f_\theta}(A), \tag{5.25}$$

where π is the distribution of a (M_n^{p+q})-sample of i.i.d. Rademacher random variables. Proposition 5.5 justifies the use of these alternatives and this probability measure to prove the lower bound.

Proposition 5.5. *Let α, β in $(0,1)$ such that $\alpha + \beta < 1$ and let $\delta > 0$. Denote f_0 the uniform density on $[0,1]^{p+q}$. There exists some positive constant $C(p,q,\alpha,\beta,\delta)$ such that, if we set*

$$h_n = C(p,q,\alpha,\beta,\delta) n^{-2/(4\delta+p+q)}, \tag{5.26}$$

and define P_ν by Equations (5.24) and (5.25), then we have

$$\mathbb{E}_{P_{f_0}}\left[\left(\frac{\mathrm{d}P_\nu}{\mathrm{d}P_{f_0}}(\mathbb{Z}_n)\right)^2\right] < 1 + 4(1-\alpha-\beta)^2,$$

for n large enough.

Finally, combining Lemma 5.6 with Propositions 5.4 and 5.5, we obtain a lower bound for the non-asymptotic minimax separation rate of testing in Theorem 5.5.

Theorem 5.5. *Consider α, β in $(0,1)$ such that $\alpha + \beta < 1$. Let $\delta > 0$ and $R > 0$. Then, there exists a positive constant $C(p,q,\alpha,\beta,\delta,R)$ such that*

$$\rho\left(\mathcal{S}^\delta_{p+q}(R), \alpha, \beta\right) \geq C(p,q,\alpha,\beta,\delta,R)\ n^{-2\delta/(4\delta+p+q)},$$

for n large enough.

Theorem 5.5 proves that each single test introduced in Corollary 5.2 is optimal in the minimax sense over Sobolev balls $\mathcal{S}^\delta_{p+q}(R)$ for δ in $(0,2]$ since the upper and lower bounds coincide up to constants. Moreover, the aggregated testing procedure defined in Corollary 5.4 is optimal up to a logarithmic term over Sobolev balls. Since it does not depend on the prior knowledge of the regularity parameter δ, it is adaptive.

5.5 Application of the HSIC-based testing procedure methodology

In this section, we illustrate the practical implementation of the proposed methodology. For this, we first rely on several analytical data generating mechanisms of dependence. Subsequently, an application of the methodology to the case of ULOF-MACARENa simulator is provided.

5.5.1 Numerical simulations

Now, several analytical tests based on intensive numerical simulations will be performed for two main goals. Firstly, to assess the effectiveness of this new testing approach and secondly to test and compare the possible methodological choices. To achieve this, we first numerically illustrate that the power of the approximated single-HSIC test $\hat{\Delta}^{\lambda,\mu}_\alpha$ is almost equal to the power of the theoretical test $\Delta^{\lambda,\mu}_\alpha$ defined in Equation (3.18). A similar verification is also carried out for the aggregated procedure. Secondly, we compare and discuss different strategies of aggregation. These strategies of aggregation are then compared to single-HSIC tests. Finally, a comparison with the non parametric independence procedure of Berrett and Samworth (2017) is proposed, based on the analytical test cases provided by the authors.

5.5.1.1 Assessment of the power of permuted HSIC-tests

In this section, we first numerically illustrate that permuted single HSIC tests approximate very well the power of the theoretical tests, as soon as enough permutations are used for the estimation of the threshold value (the quantile under the null hypothesis). The level of the permuted tests being guaranteed by Proposition 3.5, the permutation approach does not affect the quality of these tests. Thereafter, a similar study will be conducted for the aggregated testing procedure introduced in Section 5.3.1. All along this section, we rely on the following analytical model $\mathcal{M}$, inspired from the Ishigami function from Ishigami and Homma (1990). Let X_1, X_2 and X_3 be independent uniform random variables on $[0,1]$, we define the model $\mathcal{M}$ as follows:

$$\mathcal{M}(X_1, X_2, X_3) = \sin(X_1) + 4\sin^2(X_2) + 0.5X_3^4 \sin(X_1),$$

defining the random variable $Y = \mathcal{M}(X_1, X_2, X_3)$. The analytical illustrations of this section are focused on the random couple (X_1, Y).

• **For the single test.** In order to evaluate the approximation accuracy of permuted single HSIC tests, we consider here the heuristic kernel parameterization based on X_1 and Y intrinsic characteristics. More precisely, the kernel bandwidth associated to X_1 (resp. Y) is the empirical standard deviation of X_1 (resp. Y). In the following analytical illustrations, we draw results for three sample sizes: $n = 50, 100, 200$ and two test levels: $\alpha = 0.05$ and 0.001. For each sample size n and test level α, we first estimate the theoretical power of the test. To achieve this, a preliminary step is to estimate the theoretical $(1-\alpha)$-quantile (denoted $q_{1-\alpha}$) of the estimator $\widehat{\text{HSIC}}(X_1, Y)$ under the null hypothesis. For this, we re-estimate 500000 times $\widehat{\text{HSIC}}(X_1, Y)$, we use at each time a different n-sample $(X_1^{(i)}, Y^{(i)})_{1\leq i\leq n}$ of (X_1, Y) under the null hypothesis[2]. Once the quantile $q_{1-\alpha}$ is estimated, we generate 1000 different n-samples of (X_1, Y) according to its true distribution. From these 1000 samples, we estimate the "theoretical" power $\beta_{th}(n, \alpha)$ of this HSIC-test as being the ratio of times that the estimator $\widehat{\text{HSIC}}(X_1, Y)$ exceeds the quantile $q_{1-\alpha}$. The second step consists in estimating the power of permuted tests for different number of permutations B. The chosen values of B are: 10, 20, ..., 100, 200, ..., 2500. For each value of n, α and B, we generate 1000 different times a n-sample of (X_1, Y) according to its distribution. For each sample, we estimate $\widehat{\text{HSIC}}(X_1, Y)$, as well as the quantile $q_{1-\alpha}$ using B random permutations of the sample. From this, we deduce the power $\beta(B, n, \alpha)$ of the permuted test, as being the ratio of times the estimated value of $\widehat{\text{HSIC}}(X_1, Y)$ exceeds the estimated value of the quantile. To compare the powers of theoretical and permuted tests (resp. $\beta_{th}(n, \alpha)$ and $\beta(n, \alpha, B)$), we consider the relative absolute error $Err(n, \alpha, B)$ defined as

$$Err(n, \alpha, B) = \frac{|\beta(n, \alpha, B) - \beta_{th}(n, \alpha)|}{\beta_{th}(n, \alpha)}.$$

The results obtained for $\alpha = 0.05$ and different n values are given by Figure 5.1. We can see that the approximation accuracy of the permuted approach tends to increase as n increases. This can be explained by the fact that the power of the theoretical test increases as the sample size increases, the permuted test becomes then less and less sensitive to the error due to Monte Carlo estimation of the quantile $q_{1-\alpha}$. Moreover, the power of the theoretical test is more difficult to estimate for small sizes, which explains the fluctuations observed for $n = 50$. Generally,

[2] To generate an independent n-sample of (X_1, Y) under the null hypothesis, we first generate an independent $2n$-sample of X_1. Only the first n elements are used to compute the marginal sample of Y and the remaining n elements are considered to be the marginal sample of X_1. We thus obtain a n-sample of (X_1, Y) with realizations of X_1 independent of those of Y.

the permutation approach allows to obtain the theoretical test power with an acceptable level of precision, even for small values of B. In particular, we observe for $n = 50$ that aside from very small values of B and two outliers, the absolute relative error is always lower than 10%. Moreover, from $n = 100$ this error is mostly inferior than 10% and no observed error is over 5% for $n = 200$.

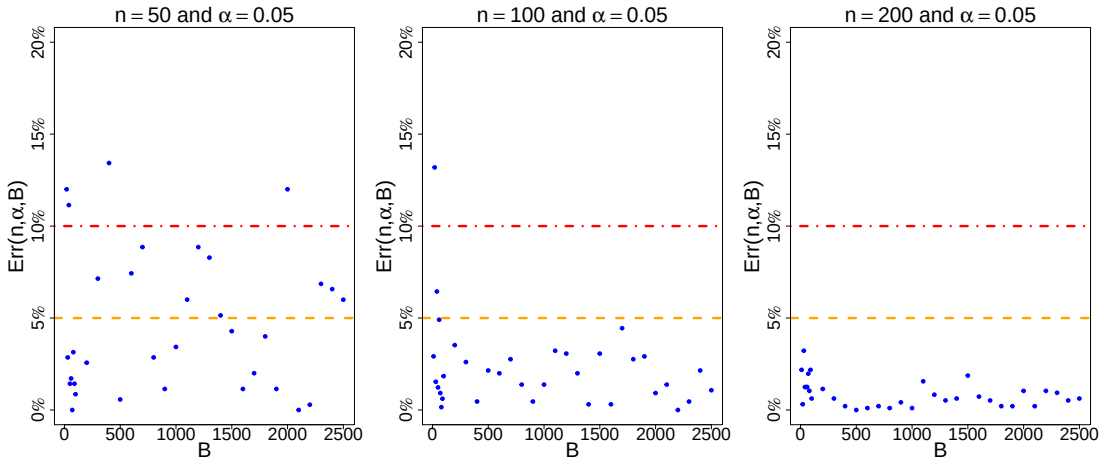

Figure 5.1 – Absolute relative error between the powers of the theoretical and permuted HSIC-tests, w.r.t the number B of permutations, for sample sizes $n = 50, 100$ and 200. The presumed level of tests is $\alpha = 0.05$. The red (resp. orange) dashed line represents the error threshold of 10% (resp. 5%).

In order to study the impact of changing the test level on the accuracy of the permutation approximation, we show in Figure 5.2 the relative absolute error of the power w.r.t. n and B for the extreme level value $\alpha = 0.001$. Contrary to the case $\alpha = 0.05$, we observe here much less precision of the power approximation. In particular, for $n = 50$, $B = 2000$ permutations are required to obtain satisfactory accuracy (against $B = 30$ for $\alpha = 0.05$). Similar observations are done for $n = 100$ and 200 with respectively $B = 1200$ and $B = 500$ permutations required (against $B = 30$ and $B = 10$ for $\alpha = 0.05$).

• **For the aggregated procedure.** The objective here is to check that the permutation approach does not impact the quality of the aggregated HSIC procedure. Before that, let us describe the practical implementation of the aggregated testing procedure, based on permutation approach. Considering the sample $(X_1^{(i)}, Y^{(i)})_{1\leq i\leq n}$ of (X_1, Y) and a collection of kernel bandwidths $(\lambda, \mu) \in \Lambda \times U$, this procedure relies on Algorithm 1. Note that, the computational complexity of Algorithm 1 is $O\left((B_1 + B_2) \times |\Lambda| \times |U| \times n^2\right)$, corresponding to the estimation of HSIC for different permutations and widths in Step 0. The two remaning steps do not add any additional cost w.r.t to the sample size n. Furthermore, the procedure requires the memory storage of an array of size $(B_1 + B_2 + 1) \times |\Lambda| \times |U| \times n^2$.

By now, let us go back to our initial goal: compare the power (of the aggregated test) using the permuted approach, with the theoretical power.

- Theoretical power. From a given sample of size n and for a given collection of bandwidths $\Lambda \times U$, we propose to compute the theoretical power of the aggregated test as

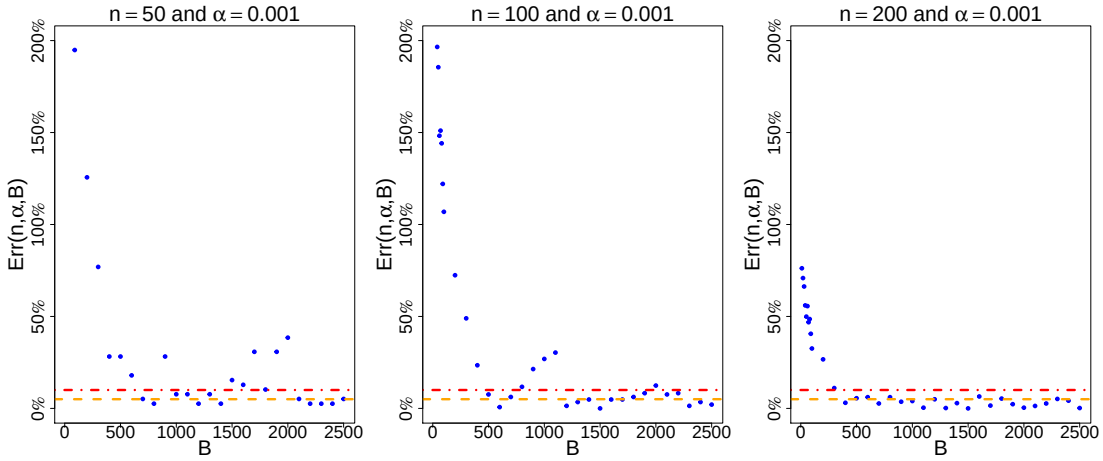

Figure 5.2 – Absolute relative error between the powers of the theoretical and the permuted HSIC-tests, w.r.t the number B of permutations, for sample sizes $n = 50, 100$ and 200. The presumed level of tests is $\alpha = 0.001$. The red (resp. orange) dashed line represents the error threshold of 10% (resp. 5%).

follows. The value of u_α is still computed by dichotomy but, using a "theoretical" value of P_u (for each possible u). For this, the quantile $q^{\lambda,\mu}_{1-ue^{-\omega_{\lambda,\mu}}}$ is computed for each (λ, μ), with a sample of 500000 values of $\widehat{\mathrm{HSIC}}_{\lambda,\mu}(X_1, Y)$ drawn under the null hypothesis. An estimation of P_u is then deduced by computing 1000 times $\widehat{\mathrm{HSIC}}_{\lambda,\mu}(X_1, Y)$ with different i.i.d samples generated under the null hypothesis. The value of P_u is the ratio of times that at least one $\widehat{\mathrm{HSIC}}_{\lambda,\mu}(X_1, Y)$ is greater than its associated quantile $q^{\lambda,\mu}_{1-ue^{-\omega_{\lambda,\mu}}}$. The stop criterion for u_α computation is the same as in Algorithm 1. Knowing the theoretical value of u_α, we estimate for each (λ, μ) the theoretical quantiles $q^{\lambda,\mu}_{1-u_\alpha e^{-\omega_{\lambda,\mu}}}$ by using the 500000 values of $\widehat{\mathrm{HSIC}}_{\lambda,\mu}(X_1, Y)$. We then generate 1000 i.i.d samples and for each one, we compute the $\widehat{\mathrm{HSIC}}_{\lambda,\mu}(X_1, Y)$ associated to all (λ, μ) in $\Lambda \times U$. The theoretical power is then considered to be the ratio of samples for which at least one $\widehat{\mathrm{HSIC}}_{\lambda,\mu}(X_1, Y)$ exceeds its associated quantile $q^{\lambda,\mu}_{1-u_\alpha e^{-\omega_{\lambda,\mu}}}$.

- Permuted power. For our numerical tests, we consider 6 different collections of bandwidths $(\Lambda_r \times U_r)_{2\leq r\leq 7}$, defined for all r by

$$\Lambda_r \times U_r = \left\{1, 1/2, \ldots, 1/2^{r-1}\right\}^2$$

and choose the uniform weights defined for all (λ, μ) as $\omega_{\lambda,\mu} = \log(r^2)$. Note that, the case $r = 1$ corresponds to the single test with $\lambda = \mu = 1$.

In all the following, the presumed level of the tests is set at $\alpha = 0.05$ and $B_2 = 500$ permutations are used to estimate the probabilities P_u. We consider different values of n and B_1: $n = 50, 100, 200$ and $B_1 = 100, 200, 500, 1000, ..., 5000$. For each triplet (r, n, B_1), the procedure of aggregated test with permutation is implemented for 1000 different samples according to Algorithm 1. The power of aggregated test with permutation is then the ratio of times that the null hypothesis is rejected. Results in terms of absolute relative error on power are given by Figure 5.3. Notice that, regardless of the n value, the required value of B_1 to well approximate

Algorithm 1 *Aggregated procedure using permutations*

Step 0. For each (λ, μ), we compute two sets of independent permutations of $(X_1^{(i)}, Y^{(i)})_{1\leq i\leq n}$ repectively denoted $\mathbb{Z}_n^{\tau_{1,b}}, b = 1 \dots B_1$ and $\mathbb{Z}_n^{\tau_{2,b}}, b = 1 \dots B_2$ (cf. notation in Section §5.3.1). Two samples of permuted $\widehat{\text{HSIC}}_{\lambda,\mu}(X_1, Y)$, denoted $\widehat{H}_{\lambda,\mu}^{\star b,1}, b = 1 \dots B_1$ and $\widehat{H}_{\lambda,\mu}^{\star b,2}, b = 1 \dots B_2$ are computed, respectively from $\mathbb{Z}_n^{\tau_{1,b}}, b = 1 \dots B_1$ and $\mathbb{Z}_n^{\tau_{2,b}}, b = 1 \dots B_2$. The role and use of these two sets of permuted HSIC is explained in the next step.

Step 1. This step aims at numerically estimating the u_α defined in Equation (5.14). A very straightforward approach to estimate u_α is to proceed by dichotomy on the interval $[\alpha, M]$, where $M = \inf_{(\lambda,\mu)\in\Lambda\times U} e^{\omega_{\lambda,\mu}}$ (u_α belonging to this interval as demonstrated in §5.3.1). For this, u_α is initialized at $u := \alpha(1 + M)/2$ and for each considered value of u, we estimate the following probability P_u:

$$P_u = P_{f_1\otimes f_2}\left(\sup_{(\lambda,\mu)\in\Lambda\times U}\left(\widehat{\text{HSIC}}_{\lambda,\mu} - q_{1-ue^{-\omega_{\lambda,\mu}}}^{\lambda,\mu}\right) > 0\right).$$

To achieve this, we first estimate for each (λ, μ) the quantile $q_{1-ue^{-\omega_{\lambda,\mu}}}^{\lambda,\mu}$ using the set $\widehat{H}_{\lambda,\mu}^{\star b,1}, b = 1 \dots B_1$ of permuted HSIC. Then, we estimate P_u according to Equation (5.16) and using the set $\widehat{H}_{\lambda,\mu}^{\star b,2}, b = 1 \dots B_2$. Thereafter, the dichotomy interval is updated by comparing P_u and α. Furthermore, we propose the following stop criterion: $u_{max} - u_{min} \leq 10^{-3}u_{min}$, where u_{max} and u_{min} are respectively the upper and the lower bounds of the current dichotomy interval. The u_α thus estimated is denoted $\hat{u}_\alpha$.

Step 2. For each couple (λ, μ), the estimated value of $\widehat{\text{HSIC}}_{\lambda,\mu}(X_1, Y)$ is compared with quantile $q_{1-\hat{u}_\alpha e^{-\omega_{\lambda,\mu}}}^{\lambda,\mu}$. The aggregated testing procedure rejects the independence hypothesis of X_1 and Y if there is at least one (λ, μ) such that

$$\widehat{\text{HSIC}}_{\lambda,\mu}(X_1, Y) > q_{1-\hat{u}_\alpha e^{-\omega_{\lambda,\mu}}}^{\lambda,\mu}.$$

the theoretical power increases with r. In fact, the supremum in Equation (5.16) becomes more difficult to estimate as the number r^2 of aggregated tests increases. Unsurprisingly, for a given B_1 value, the accuracy of the power estimation increases with n as in the case of single tests. In particular, we observe that for $n = 50$, the biggest error becomes less than 10% from $B_1 = 3500$ (case with $r = 7$), while this threshold seems to be achieved from $B_1 = 3000$ for $r = 4, 5, 6$ and from $B_1 = 500$ for $r = 2, 3$. For bigger sample sizes $n = 100$ and 200, a good approximation of the theoretical test is achieved from small values of B_1 ($B_1 < 1000$), even for a relatively large number of aggregated tests. In particular, for $n = 200$, an error smallest than 10% is reached for all tested B_1 values.

5.5.1.2 Performance of the aggregated procedure

The objective of this section is to compare the methodological choices of the aggregated procedure and to assess its performance. More precisely, we first investigate different choices of collections Λ and U, associated weights and sizes of each collection. A comparison with the single HSIC test is also carried out. Thereafter, the aggregated procedure is compared to the test of independence based on the mutual information, recently proposed by Berrett and Samworth (2017). Indeed,

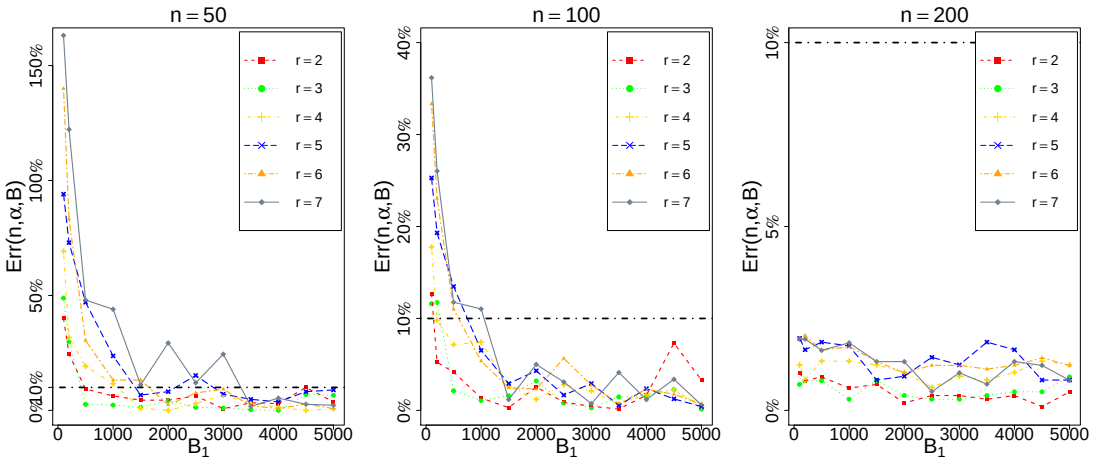

Figure 5.3 – Absolute relative error between the powers of the theoretical and permuted aggregated HSIC procedures w.r.t. the number B_1 of permutations, for the sample sizes $n = 50, 100$ and 200 and the collection Λ and U sizes r. The presumed level of tests is $\alpha = 0.05$. The black dashed line represents the error threshold of 10%.

the latter has been compared to some classical existing procedures (among which single HSIC-tests), through several analytical examples and seems to efficiently detect some forms of weak dependence. All along this section, we will consider the same analytical examples as Berrett and Samworth (2017). These examples consider the three following different mechanism of dependence (i), (ii) and (iii), each one with a varying parameter.

(i) Defining the joint density f_l of the couple (X, Y) for all (x, y) in $[-\pi, \pi]$ by

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

Densities f_l with $l = 1 \dots 10$ will be considered here.

(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 is uniform on $\{1, \dots, l\}$ for some l in $\mathbb{N}$, Θ is uniform on $[0, 2\pi]$ and ε_1, ε_2 are standard normal. In addition, the chosen values of l are $1, \dots, 10$.

(iii) Defining X as uniform on $[-1, 1]$. For a given $\rho \geq 0$, Y is defined as

$$Y = |X|^\rho \varepsilon,$$

where ε is standard normal independent with X. The considered values of ρ are $0.1, 0.2, \dots, 1$.

In addition, we will also consider the following bi-variate case: $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$ where (X_1, Y_1) is generated according to any described mechanisms above, while X_2, Y_2 are independent uniform distributions on $[0, 1]$ and independent from (X_1, Y_1).

Comparison of different strategies of aggregation. In our aggregated procedure, collections of bandwidths Λ and U have to be chosen. There is certainly no universal best collections that would ensure optimal test power; the optimal bandwidth depends on the intrinsic characteristics of X and Y and their dependency structure. Consequently, it seems relevant to consider the possible bandwidths relatively to the standard deviations of X and Y. Moreover, the standard deviation is the usual choice for bandwidth in the literature on single HSIC-test (Zhang et al., 2012). We assume here that the exact values of standard deviations of X and Y, respectively denoted s and s', are known[3]. Note that, in practice s and s' will be estimated most of the time, but experience feedback allows us to assume that the error of estimation of s and s' does not significantly impact the performance of the single test. Indeed, standard deviation estimators converge in most cases rapidly w.r.t. n. More particularly, this estimation error is small compared to the estimation error of the quantiles. Similar aggregated tests performance are then obtained using Monte Carlo estimated standard deviations.

First of all, we propose to illustrate the impact of bandwidth for single tests. For this, we consider the uni-variate mechanism of dependence (ii) with $l = 2$. Moreover, we consider, as possible bandwidths λ and μ respectively associated to X and Y, multiple or dyadic fractions of s and s' respectively. For each couple of (λ, μ), the power of single HSIC tests is computed. Figure 5.4 shows the obtained power maps w.r.t. (λ, μ), for different sample sizes. First, we can observe that bandwidths significantly impact the power: in this case, there is an optimal area around $(\lambda, \mu) = (s/4, s'/4)$ with a power close to one for $n = 200$. The power decreases progressively as we go away from this area, until being null for very high and very low values of bandwidths, this observation is also mentioned in Fukumizu et al. (2009). Moreover, we can see that the regularity of the maps increases with the sample-size (just like the power for each point, obviously). Similar conclusions were observed for the other analytical cases with one or several best-power areas. This illustrates that an arbitrary choice of bandwidths is not relevant and justifies the interest of considering several bandwidths through an aggregation strategy. From these results, we can consider aggregating procedures based on collections Λ and U of types

$$\Lambda = \left\{ s, s/2, \ldots, s/2^{r-1} \right\}, \quad U = \left\{ s', s'/2, \ldots, s'/2^{r-1} \right\}, \tag{5.27}$$

where r belongs to $\mathbb{N}^*$.

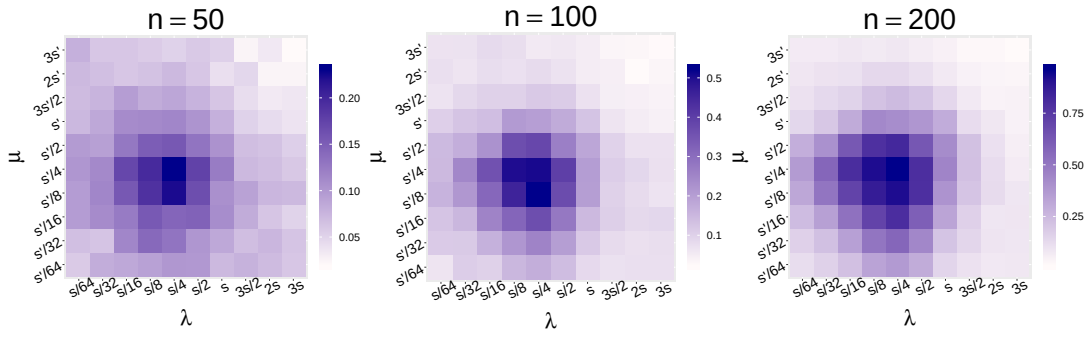

Figure 5.4 – Power map of single HSIC test w.r.t. to kernel widths λ and μ respectively associated to X and Y, for sample sizes $n = 50, 100$ and 200.

[3]In practice, when only a n-sample of (X, Y) is available, we estimate these standard deviations by usual Monte Carlo estimators.

Note that, generally speaking, small bandwidths intensify the local dissimilarities between the joint density and the product of the marginals. On the contrary, large bandwidth values are better to catch the global dependence. This explains the reason why small bandwidths values (w.r.t. standard deviations) give better results in Figure 5.4. Indeed, the dependence form (ii) with $l = 2$ is more local than global due to the presence of Cosine and Sine functions respectively in the equations defining X and Y. By now, let us compare the possible choice of weights associated to the collections Λ and U described above. We study in particular two types of weights: uniform and exponential weights. We recall that uniform weights depend only on the cardinals of Λ and U, they are defined for all (λ, μ) as

$$\omega_{\lambda,\mu} = \log(r^2).$$

Furthermore, the exponential weights are adapted to the collections defined in Equation (5.27). They are defined, by analogy with Equation (5.19), for all $(s/2^i, s'/2^j)$ as

$$\omega_{s/2^i,s'/2^j} = \log\left(\sum_{1\leq u,v\leq r} \frac{(i+1)^2(j+1)^2}{u^2v^2}\right). \tag{5.28}$$

The results obtained with the two types of weights are given by Figure 5.5, for different collection sizes r and sample sizes n. In this application, the uniform weights strategy seems to give a better power than the exponential weights one. But, we can observe a different behavior w.r.t. r. For the uniform weights, the power increases until a specific r ($r = 3$ or 4 w.r.t n), before decreasing with r, to being lower than power with exponential weights. On the contrary, the power with exponential weights has a more robust behavior, since it increases with r until it stabilizes. The uniform strategy is perhaps more convenient for small values of r (in practice $r \leq 4$), whereas the exponential strategy performs better for bigger values of r. We can also observe that the two aggregated strategies yield a greater power than the single test (which corresponds to the case $r = 1$), as soon as a sufficient sizes of collection are chosen. Indeed, in this case the dependence is more accurately captured with small bandwidth values (smaller than the standard deviations). Similar conclusions have been drawn from the other analytical examples, which are not presented here for the sake of brevity.

Comparison with Mutual Information Test. To complete these numerical tests, we propose to compare our aggregated procedure with some existing reference tests of independence. For this, we rely on the analytical examples of Berrett and Samworth (2017) presented above and consider a sample size $n = 200$. In their paper, Berrett and Samworth (2017) numerically compare the powers of several independence tests. More precisely, they compare their Mutual Information Test (called MINT) with that based on the copula defined in Kojadinovic and Holmes (2009), the distance covariance (Székely et al., 2007), the Kendall's tau introduced in Bergsma and Dassios (2014) and the single HSIC using the permutation method (Gretton et al., 2008) with $B = 1000$ permutations. For this last test, same kernel bandwidths defined for multivariate variables are chosen for X and for Y in Berrett and Samworth (2017):

$$\lambda^2 = \frac{1}{2} median\left\{\|X_i - X_j\|_2^2 : i < j\right\} \quad \text{and} \quad \mu^2 = \frac{1}{2} median\left\{\|Y_i - Y_j\|_2^2 : i < j\right\},$$

where p (resp. q) is the dimension of X (resp. Y) and $\|.\|_p$ (resp. $\|.\|_p$) is the Euclidean norm in dimension p (resp. q). For the sake of consistency, we make here slightly different choices for these bandwidths, by taking λ and μ such that

$$\widetilde{\lambda}^2 = \frac{1}{2} mean\left\{\|X_i - X_j\|_2^2 : i < j\right\} \quad \text{and} \quad \widetilde{\mu}^2 = \frac{1}{2} mean\left\{\|Y_i - Y_j\|_2^2 : i < j\right\}. \tag{5.29}$$

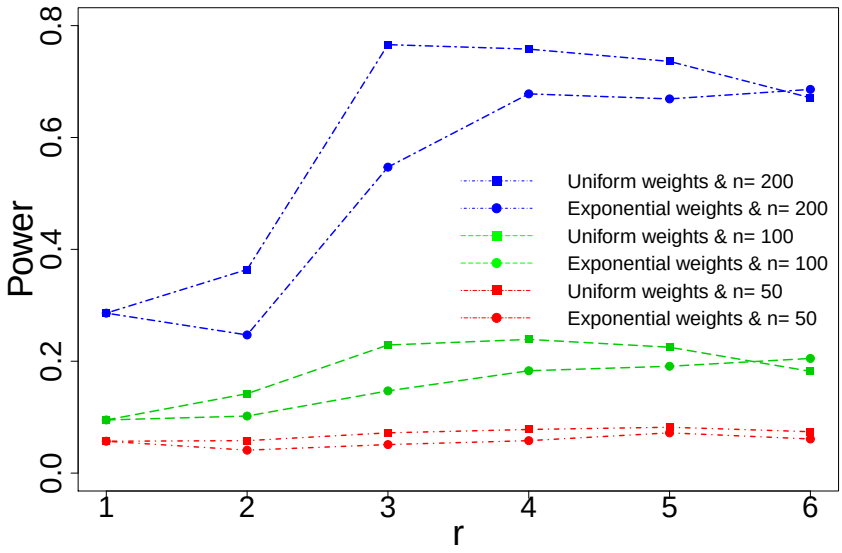

Figure 5.5 – Power of aggregated procedures with uniform and exponential weights, w.r.t. the number r of aggregated bandwidths in each direction, for sample sizes $n = 50, 100$ and 200.

Notice that, when X (resp. Y) is one-dimensional, then $\widetilde{\lambda}$ (resp. $\widetilde{\mu}$) is the empirical estimator of the standard deviation of X (resp. Y). In Berrett and Samworth (2017), the most powerful tests on the analytical uni-variate and bi-variate examples (i), (ii) and (iii) are the MINT and single HSIC test. Let us compare the performances of these two tests with our aggregated procedure with the following methodological choices:

- The bandwidths associated to X (resp. Y) are chosen to be the same for all directions and belong to the collection Λ (resp. U) defined as

$$\Lambda = (\widetilde{\lambda}, \widetilde{\lambda}/2, \ldots, \widetilde{\lambda}/2^6) \quad \text{and} \quad U = (\widetilde{\mu}, \widetilde{\mu}/2, \ldots, \widetilde{\mu}/2^6),$$

  where $\widetilde{\lambda}$ and $\widetilde{\mu}$ are the bandwidths introduced in Equation (5.29).

- Exponential weights of Equation (5.28) are chosen.

- Algorithm 1 is used with $B_1 = 3000$ and $B_2 = 500$.

For each example, the power of the different testing procedures is estimated using 1000 different samples of (X, Y). The obtained power curves are given by Figure 5.6, w.r.t parameters l and ρ. No procedure of testing constantly yields the best performances. For the case (i), the MINT and the HSIC aggregated procedure have competitive results, much better than single HSIC. For the mechanism (ii), the MINT is the most powerful method, then comes the aggregated procedure, single HSIC giving the worst results. For the last example (iii), results are opposite. Thus, the HSIC aggregated procedure seems to yield intermediate results between MINT and single HSIC: it provides better results on average, regardless of the mechanism of dependence between the variables. Moreover, in the majority of the presented examples, the HSIC aggregated procedure performs better than the single HISC test.

It may occur that the aggregation procedure with dyadic fractions of standard deviations as bandwidths yields weaker performance than single HSIC tests. This is the case when the

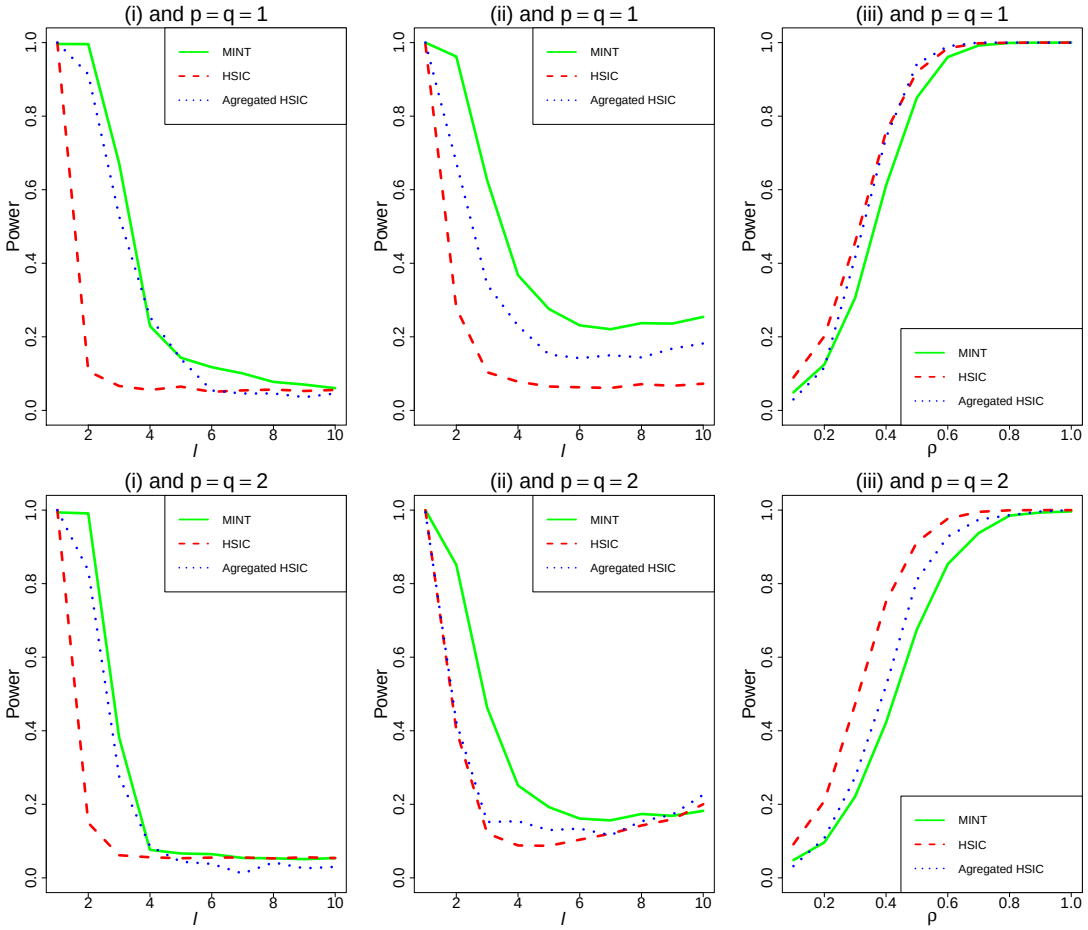

Figure 5.6 – Power curves of MINT, single HSIC test and aggregated procedure for the mechanisms of dependence (i), (ii) and (iii) in the uni-variate ($p = q = 1$) and the bi-variate ($p = q = 2$) cases.

dissimilarity between the joint density and the product of the marginals is very global. In such cases, it is more useful to take into account large bandwidth values (more precisely, higher than standard deviations). To illustrate this point, we consider a new modified version of Ishigami function, defined as

$$\mathcal{M}(X_1, X_2, X_3) = \sin(X_1) + 4\ \sin^2(X_2) + 0.2\ X_3^4\ \sin(X_1), \tag{5.30}$$

where the inputs X_1, X_2 et X_3 are independent and follow a uniform distribution on $[0, 1]$. Let Y be the output variable $Y = \mathcal{M}(X_1, X_2, X_3)$. Figure 5.7 shows the power map of the single HSIC test of the couple (X_1, Y) w.r.t. the bandwidth values λ and μ respectively associated to X_1 and Y. The sample size chosen for this analytical example is $n = 200$. Moreover, the power of the test associated to each couple of bandwidths (λ, μ) is estimated using 200 different samples. We observe here that the single tests associated to the bandwidths in the area beyond standard deviations have a good power. This area is then to be considered in this case. In order

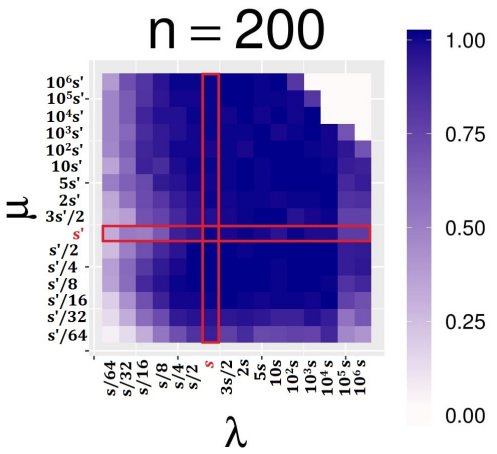

Figure 5.7 – Power map of single HSIC test w.r.t. the kernel bandwidth λ and μ respectively associated to X_1 and Y, for modified Ishigami function and a sample size $n = 200$. The standard deviations of X_1 and Y are respectively denoted s and s'.

to validate this intuition, we consider the collections Λ_r and U_r defined as

$$\Lambda_r = \{s, 2s, \ldots, 2^r s\} \qquad \text{and} \qquad U_r = \{s', 2s', \ldots, 2^r s'\},$$

where s and s' are respectively the standard deviations of X_1 and Y, while r in $\mathbb{N}^*$. Note that, as previously, the case $r = 1$ corresponds to the heuristic single HSIC test. For each r in $\{1, 2, 3, 4\}$ and n in $\{50, 100, 200\}$, we consider the aggregated procedure using the uniform weights and implemented with $B_1 = 500$ and $B_2 = 3000$. Table 5.1 shows the powers of the tests w.r.t. r and n, estimated each time using 200 different samples. For all the values of n, the aggregated procedure gives better results than single HSIC tests when $r = 3$ and has comparable power values for $r = 4$. This suggests that the aggregation in the increasing directions of bandwidths can be a good approach when the nature of the dependence is more global in nature.

	$r = 1$	$r = 2$	$r = 3$	$r = 4$
$n = 50$	0.42	0.44	0.47	0.42
$n = 100$	0.73	0.69	0.82	0.70
$n = 200$	0.96	0.98	1 .00	0.98

Table 5.1 – Power of the aggregated HSIC tests for modified Ishigami function w.r.t. r the number of aggregated bandwidths at each direction and n the sample size.

Depending on the nature of the dependence (local or global), it appears that it will be more relevant to aggregate either dyadic fractions or multiples of standard deviations. In practice, as the nature of the dependence (local or global) is often unknown, we can recommend two possible strategies of aggregation. The first approach could consist in considering both small and big bandwidth values around the standard deviation: for example, the collection will includes the standard deviation and its half and double in both directions.

Another suitable solution, if the sample size is large enough, could be to use a part of the initial sample to identify "relevant" bandwidth choices, while the rest of the sample is kept to perform the aggregated tests. On the first part of the sample, the relevant bandwidths could be the ones that lead to minimal p-values of single HSIC tests, among some possibly interesting values bandwidths (e.g. some multiples and dividers by 2 of the standard deviations). Note that the number of selected bandwidths could be also limited to 3 or 4, to avoid decreasing power of the aggregated procedure (especially if uniform weights are used). In this last approach, the choice of the proportion of simulations to be set aside to find the interesting bandwidths remains an interesting research prospect.

Even if the optimal choice of the elements of the collection depends on the test case and remains to be explored further, our aggregate procedure makes it possible to take into account several scales of dependence simultaneously, and thus makes it possible to detect a wider spectrum of dependency relationship.

5.5.2 Nuclear safety application

Our current objective is to apply and further explore the methodology and its various aspects by facing it with the real data case of MACARENa Simulator, presented in Chapters 1 and 2 (sections 1.3.2 and 2.3.2 respectively), which simulates the ULOF (Unprotected Loss Of Flow) accident on sodium-cooled fast reactor. For this, we consider here all the 26 uncertain and independent inputs identified by Droin (2016) and listed in Table 4.3. For the sake of clarity, the 26 inputs will be denoted $X_1, \ldots, X_{26}$ (numbering according to their order of appearance in Table 4.3). As in Section 4.3.2, the output of interest, denoted Y, is the first instant of sodium boiling. Note that, to focus on simulations with boiling, the support of variable 4, the external pressure drop discrepancy, is here reduced to $[-0.1, 0.04]$, still with a uniform distribution. For this study, a Monte Carlo sample of size $n = 697$ of inputs and associated outputs is available (from Droin, 2016). Based on this sample, we examine the performance of input screening based on HSIC-tests (single and aggregated), w.r.t. the number of simulations. For all the statistical independence HSIC-tests used in this section, the level is set at $\alpha = 5\%$.

Screening using the aggregation methodology. By now, we study the performance of the aggregated testing procedure proposed above. As we do not have any prior information on the nature (local or global) of the dependence between the inputs and the output, we propose here to consider the following bandwidth collections Λ and U respectively associated to X_k and Y defined as

$$\Lambda = \{s/2, s, 2s\} \quad \text{and} \quad U = \{s'/2, s', 2s'\},$$

where s and s' are respectively the standard deviations of X_k and Y. Moreover, the uniform strategy for weights is adopted. From the initial sample of size $n = 697$, we first apply the methodology using Algorithm 1 with $B_1 = 2000$ and $B_2 = 500$. The inputs selected by the aggregated tests of level $\alpha = 5\%$ are X_1, X_2, X_4, X_8 and X_{24}[4]. Let us now study the convergence and robustness of these screening results with respect to sample size n. For each given sample size $n = 50, 100, 150, 200, 300, 400, 500$ and each input identified as "influential", we estimate the selection rate[5] of the aggregated procedure using 200 subsamples of size n. Figure 5.8 represents the simulation results. As we can observe, the inputs X_4 and X_8 are quickly detected by the test even with few simulations (since $n = 50$). It seems reasonable to think that these inputs may also have the strongest and most global dependence with Y. In contrast, the dependence between

[4]The interpretation of these results in the physical sense can be further conducted.
[5]The selection rate indicates the percentage of times that the variable is selected for the screening.

the other selected inputs X_1, X_2, X_{24} and the output Y are more challenging to detect. We observe for instance that, to exceed the fifty-fifty chance to select X_1 as influential, the sample size must at least be equal to $n = 500$. Furthermore, medium sample sizes are sufficient to have a good chance to select the inputs X_2 and X_{24}. The inputs X_1, X_2, X_{24} have perhaps a lower or more local dependence with the output.

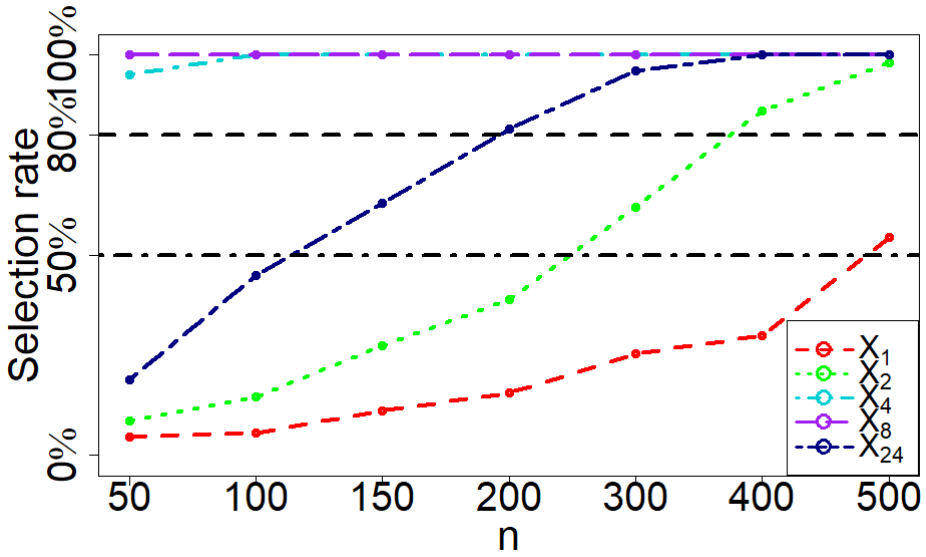

Figure 5.8 – Convergence of the selection rate for each "influential" input of MACARENa simulator based on aggregated HSIC tests for each input w.r.t. the sample size n.

Screening using heuristic single HSIC tests. Let us now compare the performance of the aggregated procedure with the single heuristic test. At the beginning, we use the total available sample to get an insight of the screening results using the single test. To do so, we rely on HSIC-based asymptotic tests of independence[6] presented in Section 3.4.2. The p-values associated to each of the 26 inputs are then estimated using the Gamma approximation of HSIC estimators under the null hypothesis. Table 5.2 indicates the inputs whose asymptotic p-values (estimated using the whole sample of size 697) are less than 5%. Thus, the inputs X_1, X_2, X_4, X_8 and X_{24} are selected by the single tests of level $\alpha = 5\%$. The screening results given with the single tests are then in line with those obtained using our aggregated procedure. By now, we compare the screening results with smaller sample sizes of the single test and the aggregated procedure. For this, we consider the same sizes as in the case of the aggregated procedure $n = 50, 100, 150, 200, 300, 400, 500$. In addition, we estimate for each size n the corresponding rate of selection using 200 different subsamples of size n. We show in Figure 5.9 the obtained results. We notice that generally, the results obtained by both tests are very similar. However, the single tests which focus only on the standard deviation as bandwidth seems on this example to detect influential variables a little faster than aggregated tests. In particular, the rate of

[6]Before applying HSIC-based independence tests with the Gamma approximation, we checked that the use of the asymptotic framework is justified.

selection exceeds 50% earlier for X_2 and X_{24} using the single test ($n = 200$ for X_2 and $n = 100$ for X_{24}).

Input	X_1	X_2	X_4	X_8	X_{24}
P-value	1.21×10^{-2}	2.84×10^{-5}	0	0	1.11×10^{-11}

Table 5.2 – Asymptotic p-values less than 10% among MACARENa inputs, from usual HSIC-test with Gaussian kernels and heuristic bandwidth parameters.

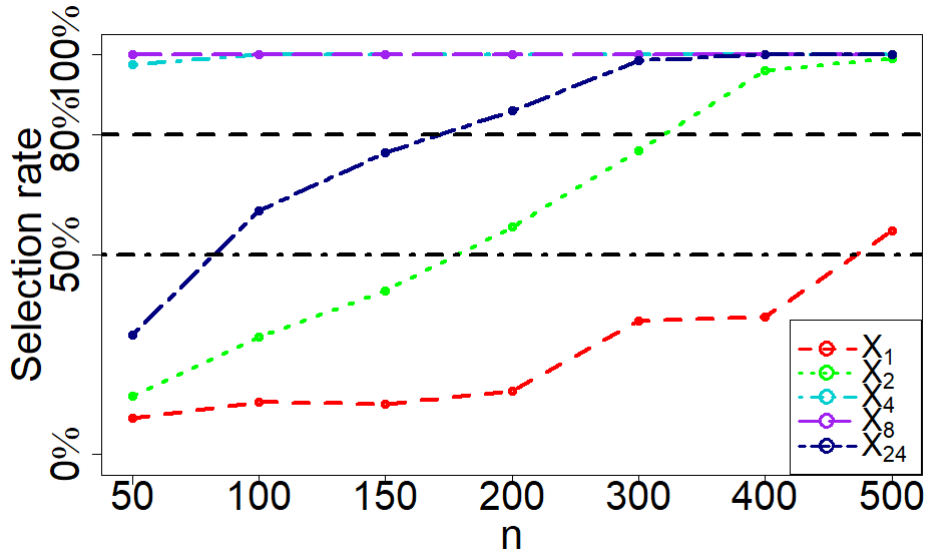

Figure 5.9 – Convergence of the selection rate for each "influential" input of MACARENa simulator based on single HSIC tests w.r.t. the sample size n.

Discussion. Assuming that the detected inputs are effectively influential (which is a quite fair assumption considering the convergence graphs), it would be instructive to understand why the heuristic test gives here better results than the aggregated procedure. For this, we represent in Figure 5.10 the maps of the probability that the HSIC estimators are lower than the observed value (one minus the p-value) for the inputs X_1, X_2 and X_{24} and using the initial sample of size $n = 697$. The graphics show that for these three inputs, the performance of single tests are maximized using bandwidths around the standard deviations. This means that in this case, the aggregation does not provide any new information, since the heuristic unit test is already well located. But once again, since the nature of the dependence (local/global) is unknown in the general case, the interest of the aggregated procedure is to take into account several dependence scales simultaneously. Even if we cannot totally avoid a heuristic choice since we consider collections of dyadic dividers and multiples of the standard deviations but, the impact of this choice is more relaxed and this is the whole interest of the aggregated procedure.

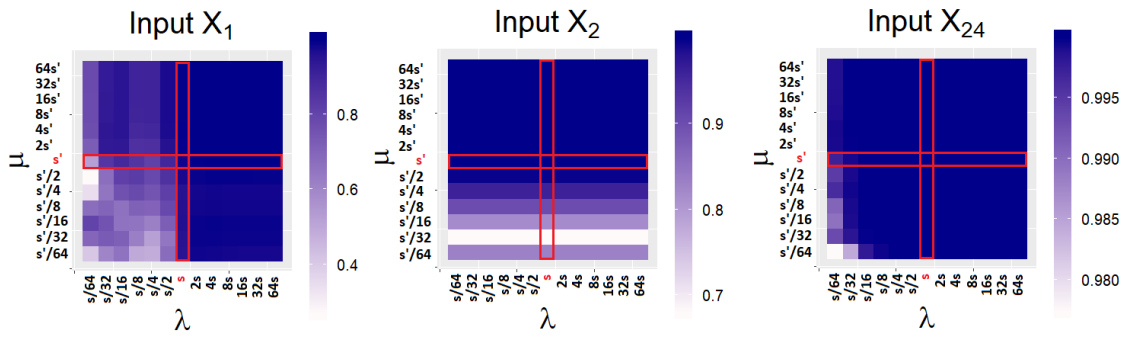

Figure 5.10 – Probability under the null hypothesis to be lower than the observed HSIC value (one minus the p-value) maps of the inputs X_1, X_2 and X_{24} using the total available sample of size $n = 697$.

5.6 Conclusion and Prospect

In this chapter, we present a new methodology to build statistical tests of independence based on HSIC measures. These tests are developped to screen the inputs of a numerical simulator. The principle of this methodology is based on the aggregation of several HSIC tests, in order to improve the ability of the test to detect the influential inputs. This methodology is motivated on one side by the efficiency of HSIC for screening, and on the other side by the unjustified heuristic choices for the kernels associated to these measures. To introduce this methodology, we consider the HSIC measures based on Gaussian kernels and focus on the bandwidth parameters involved in these kernels.

Theoretical characteristics of each independence test with a given Gaussian kernel (i.e. with given bandwidths) are first studied. In particular, sharp upper bounds of the uniform separation rate over classes of regular alternatives is demonstrated. We also establish lower bounds for these uniform separation rates, proving thus the optimality of our procedure. Subsequently, we propose the aggregated methodology based on a collection of single HSIC tests, each one of them is defined using a different bandwith in the Gaussian kernel. This procedure is shown to be nearly adaptive over Sobolev spaces, meaning that it achieves the optimal uniform separation rate over all these spaces simultaneously (up to a logarithmic term).

Numerical studies are performed to assess the performance of the methodology. In particular, different strategies of aggregation are studied and compared. In addition, the practical interest of this testing methodology is highlighted by comparison with single HSIC and other independence tests. Significant improvement of the power of single HSIC tests can be achieved. Moreover, the procedure performs well comparing to other existing independence tests.

Future works could attempt to improve some technical points of the methodology and to extend its scope. First, the choice of the bandwidth collections remains to explore. Indeed, it seems that small bandwidth values are more suitable to detect local dependence, while large ones are more adapted to capture more global dependence. An improvement of the methodology for capturing both types of dependence could be considered. In particular, we can study in more details the performance of collections with both small and large values or alternately set a proportion of the sample aside for bandwidth selection.

Another important improvement is to define and implement the computation of the p-value

associated to the aggregated procedure. In fact, the p-value can be seen as the minimal level value such that the null hypothesis is not rejected. The p-value of the aggregated procedure can then be estimated by dichotomy.

It would also be interesting to further study the link between the type of variables (continuous, discrete, etc.) and the bandwidth collections to be chosen. Furthermore, it would also be valuable to extend the aggregation to other types of characteristic kernels and not only Gaussian ones.

Another natural continuation of these works would be the asymptotic version of the aggregation procedure, where the permutation approach is replaced by the method based on asymptotic distributions of HSIC estimators under the hypothesis of independence.

Finally, in the framework of uncertainty treatment of expensive simulators with lot of uncertain inputs, *Quasi-Monte Carlo* (Niederreiter, 1992; Caflisch, 1998) and *Space-Filling* Design methods (Pronzato and Müller, 2012; Husslage et al., 2011) are often used to optimally cover the space of the inputs. These sampling techniques considerably improve the convergence rates of estimators based on expectations, such as the HSIC estimators. However, some statistical assumptions in the HSIC-based tests (independence of the observations) are not satisfied anymore. In particular, the estimation of quantiles should be corrected (like any estimators of order statistics). So, to increase the scope of application of our methodology, an attractive perspective would be to adapt the methodology to such kind of designs.

5.7 Proofs

All along the proofs, we set $Z = (X, Y)$ and $Z_i = (X_i, Y_i)$ for all i in $\{1, \ldots, n\}$. We also denote by A, B and C positive universal constants whose values may change from line to line.

5.7.1 Proof of Lemma 5.1

Let α and β be in $(0, 1)$. We aim here to give a condition on $\mathrm{HSIC}_{\lambda,\mu}(f)$ w.r.t. the variance $\mathrm{Var}_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu})$ and the quantile $q_{1-\alpha}^{\lambda,\mu}$, so that the statistical test $\Delta_\alpha^{\lambda,\mu}$ has a second kind error controlled by β. For this, we use Chebyshev's inequality. Since $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ is an unbiased estimator of $\mathrm{HSIC}_{\lambda,\mu}(f)$,

$$P_f\left(\left|\widehat{\mathrm{HSIC}}_{\lambda,\mu} - \mathrm{HSIC}_{\lambda,\mu}(f)\right| \geq \sqrt{\frac{\mathrm{Var}_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu})}{\beta}}\right) \leq \beta.$$

We then have the following inequality:

$$P_f\left(\widehat{\mathrm{HSIC}}_{\lambda,\mu} \leq \mathrm{HSIC}_{\lambda,\mu}(f) - \sqrt{\frac{\mathrm{Var}_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu})}{\beta}}\right) \leq \beta.$$

Consequently, one has $P_f\left(\widehat{\mathrm{HSIC}}_{\lambda,\mu} \leq q_{1-\alpha}^{\lambda,\mu}\right) \leq \beta$, as soon as

$$\mathrm{HSIC}_{\lambda,\mu}(f) \geq \sqrt{\frac{\mathrm{Var}_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu})}{\beta}} + q_{1-\alpha}^{\lambda,\mu}.$$

5.7.2 Proof of Proposition 5.1

In order to provide an upper bound of the variance $\mathrm{Var}_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu})$ w.r.t. the bandwidths λ, μ and the sample-size n, let us first give the following lemma for a general U-statistic of any order r in $\{1,\dots,n\}$.

Lemma 5.7. *Let h be a symmetric function with $r \leq n$ inputs, $V_1,\dots,V_n$ be independent and identically distributed random variables and U_n be the U-statistic defined by*

$$U_n = \frac{(n-r)!}{n!} \sum_{(i_1,\dots,i_r)\in \mathbf{i}_r^n} h(V_{i_1},\dots,V_{i_r}).$$

The following inequality gives an upper bound of the variance of U_n,

$$\mathrm{Var}(U_n) \leq C(r)\left(\frac{\sigma^2}{n} + \frac{s^2}{n^2}\right), \tag{5.31}$$

where $\sigma^2 = \mathrm{Var}\left(\mathbb{E}[h(V_1,\dots,V_r) \mid V_1]\right)$ and $s^2 = \mathrm{Var}\left(h(V_1,\dots,V_r)\right)$.

Proof of Lemma 5.7. First, using Hoeffding's decomposition (see e.g. (Serfling, 2009, Lemma A, p. 183)), the variance of U_n can be decomposed as

$$\mathrm{Var}(U_n) = \binom{n}{r}^{-1} \sum_{c=1}^{r} \binom{r}{c}\binom{n-r}{r-c}\zeta_c,$$

where $\zeta_c = \mathrm{Var}(\mathbb{E}[h(V_1,\dots,V_r) \mid V_1,\dots,V_c])$.

Let us now prove that, for all $n \in \mathbb{N}^*$, $r \in \{1,\dots,n\}$ and $c \in \{1,\dots,r\}$,

$$\binom{n}{r}^{-1}\binom{r}{c}\binom{n-r}{r-c} \leq \frac{C(r,c)}{n^c}. \tag{5.32}$$

We first write

$$\begin{aligned}\binom{n}{r}^{-1}\binom{r}{c}\binom{n-r}{r-c} &= \binom{r}{c} \times \frac{(n-r)!}{(r-c)!(n+c-2r)!} \times \frac{r!(n-r)!}{n!} \\ &= \binom{r}{c} \times \frac{r!}{(r-c)!} \times \frac{(n-r)!}{(n+c-2r)!} \times \frac{(n-r)!}{n!}.\end{aligned} \tag{5.33}$$

Moreover,

$$\begin{aligned} n! &= (n-r)! \times (n-r+1) \times \ldots \times (n-r+r) \\ &\geq (n-r)! \times (n-r+1)^r, \end{aligned}$$

and

$$\begin{aligned}(n-r)! &= (n-2r+c)! \times (n-2r+c+1) \times \ldots \times (n-2r+c+r-c) \\ &\leq (n-2r+c)! \times (n-r+1)^{r-c}.\end{aligned}$$

Then, we have

$$\frac{(n-r)!}{(n+c-2r)!} \times \frac{(n-r)!}{n!} \leq \frac{1}{(n-r+1)^c}.$$

Furthemore, using that $n \geq r$, one can write

$$\begin{aligned} \frac{n-r+1}{n} &= 1 - \frac{r-1}{n} \\ &\geq 1 - \frac{r-1}{r} \\ &= \frac{1}{r}. \end{aligned}$$

This leads to, $\frac{1}{n-r+1} \leq \frac{r}{n}$. Finally, Equation (5.33) leads to Equation (5.32).

By upper bounding each term in Hoeffding's decomposition of the variance of U_n according to Equation (5.32), we obtain the following inequality:

$$\mathrm{Var}(U_n) \leq C(r) \sum_{c=1}^{r} \frac{\zeta_c}{n^c}. \tag{5.34}$$

On the one hand, $\zeta_1 = \sigma^2$. On the other hand, using the law of total variance (see e.g. Weiss (2006)), for all c in $\{2, .., r\}$: $\zeta_c \leq s^2$. By injecting this last inequality in Equation (5.34), we obtain for all n in $\mathbb{N}^*$:

$$\mathrm{Var}(U_n) \leq C(r) \left(\frac{\sigma^2}{n} + \frac{s^2}{n^2} \right),$$

which achieves the proof of Lemma 5.7. □

Let us now apply Lemma 5.7 in order to control the variance of $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ w.r.t λ, μ and n. For this, we first recall that $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ can be written as a single U-statistic of order 4 Gretton et al. (2008) as

$$\widehat{\mathrm{HSIC}}_{\lambda,\mu} = \frac{1}{n(n-1)(n-2)(n-3)} \sum_{(i,j,q,r) \in \mathbf{i}_4^n} h_{i,j,q,r},$$

where the general term $h_{i,j,q,r}$ of $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ is defined as

$$h_{i,j,q,r} = \frac{1}{4!} \sum_{(t,u,v,w)}^{(i,j,q,r)} \left(k_{t,u} l_{t,u} + k_{t,u} l_{v,w} - 2 k_{t,u} l_{t,v} \right). \tag{5.35}$$

where $k_{t,u}$ (resp. $l_{t,u}$) is defined for all t, u in $\{1, \ldots, n\}$ as $k_{t,u} = k(X_t, X_u)$ (resp. $l_{t,u} = l(Y_t, Y_u)$) and the sum represents all ordered quadruples (t, u, v, w) drawn without replacement from (i, j, q, r).

Thus, using Lemma 5.7, the variance of $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ can be upper bounded as follows:

$$\mathrm{Var}_f \left(\widehat{\mathrm{HSIC}}_{\lambda,\mu} \right) \leq C \left(\frac{\sigma^2(\lambda,\mu)}{n} + \frac{s^2(\lambda,\mu)}{n^2} \right), \tag{5.36}$$

where recalling that $Z_i = (X_i, Y_i)$ for all i in $\{1, \ldots, n\}$, $\sigma^2(\lambda,\mu) = \mathrm{Var}_f \left(\mathbb{E}[h_{1,2,3,4} \mid Z_1] \right)$ and $s^2(\lambda,\mu) = \mathrm{Var}_f \left(h_{1,2,3,4} \right)$.

5.7.2.1 Upper bound of $\sigma^2(\lambda,\mu)$

By now, we upper bound $\sigma^2(\lambda,\mu)$ defined in Equation (5.36) w.r.t. λ and μ. For this, we first notice that in the cases when $k_\lambda(X_a,X_b)l_\mu(Y_c,Y_d)$ is independent from Z_1, the variance of its expectation conditionally on Z_1 equals 0. That are the cases when a,b,c and d are all different from 1. We then have the following inequality:

$$\sigma^2(\lambda,\mu) \leq C\sum_{i=1}^{6}\sigma_i^2(\lambda,\mu),$$

where

$$\begin{aligned}
&\sigma_1^2(\lambda,\mu) = \mathrm{Var}_f\left(\mathbb{E}[k_\lambda(X_1,X_2)l_\mu(Y_1,Y_2)\mid Z_1]\right), \quad \sigma_2^2(\lambda,\mu) = \mathrm{Var}_f\left(\mathbb{E}[k_\lambda(X_1,X_2)l_\mu(Y_3,Y_4)\mid X_1]\right),\\
&\sigma_3^2(\lambda,\mu) = \mathrm{Var}_f\left(\mathbb{E}[k_\lambda(X_3,X_4)l_\mu(Y_1,Y_2)\mid Y_1]\right), \quad \sigma_4^2(\lambda,\mu) = \mathrm{Var}_f\left(\mathbb{E}[k_\lambda(X_1,X_2)l_\mu(Y_1,Y_3)\mid Z_1]\right),\\
&\sigma_5^2(\lambda,\mu) = \mathrm{Var}_f\left(\mathbb{E}[k_\lambda(X_2,X_1)l_\mu(Y_2,Y_3)\mid X_1]\right), \quad \sigma_6^2(\lambda,\mu) =, \mathrm{Var}_f\left(\mathbb{E}[k_\lambda(X_2,X_3)l_\mu(Y_2,Y_1)\mid Y_1]\right).
\end{aligned}$$

Case 1. Upper bound of $\sigma_1^2(\lambda,\mu)$

$$\begin{aligned}
\sigma_1^2(\lambda,\mu) &\leq \mathbb{E}\left[\left(\mathbb{E}\left[k_\lambda(X_1,X_2)l_\mu(Y_1,Y_2)\mid Z_1\right]\right)^2\right]\\
&\leq \mathbb{E}\left[k_\lambda(X_1,X_2)l_\mu(Y_1,Y_2)k_\lambda(X_1,X_3)l_\mu(Y_1,Y_3)\right].
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
&\mathbb{E}\left[k_\lambda(X_1,X_2)k_\lambda(X_1,X_3)l_\mu(Y_1,Y_2)l_\mu(Y_1,Y_3)\right]\\
&= \int_{(\mathbb{R}^p\times\mathbb{R}^q)^3} k_\lambda(x_1,x_2)k_\lambda(x_1,x_3)l_\mu(y_1,y_2)l_\mu(y_1,y_3)\prod_{k=1}^{3} f(x_k,y_k)\mathrm{d}x_k\mathrm{d}y_k.
\end{aligned}$$

By upper bounding $f(x_2,y_2)$ and $f(x_3,y_3)$ by $\|f\|_\infty$, we have

$$\begin{aligned}
\sigma_1^2(\lambda,\mu) &\leq \|f\|_\infty^2 \int_{(\mathbb{R}^p\times\mathbb{R}^q)^3} k_\lambda(x_1,x_2)k_\lambda(x_1,x_3)l_\mu(y_1,y_2)l_\mu(y_1,y_3)\, f(x_1,y_1)\prod_{k=1}^{3}\mathrm{d}x_k\mathrm{d}y_k\\
&= \|f\|_\infty^2 \int_{\mathbb{R}^p\times\mathbb{R}^q}\left[\int_{\mathbb{R}^p} k_\lambda(x_1,x_2)\mathrm{d}x_2\right]\left[\int_{\mathbb{R}^p} k_\lambda(x_1,x_3)\mathrm{d}x_3\right]\left[\int_{\mathbb{R}^q} l_\mu(y_1,y_2)\mathrm{d}y_2\right]\\
&\qquad\left[\int_{\mathbb{R}^q} l_\mu(y_1,y_3)\mathrm{d}y_3\right] f(x_1,y_1)\mathrm{d}x_1\mathrm{d}y_1.
\end{aligned}$$

Finally, using that $\int_{\mathbb{R}^p} k_\lambda(.,x)\mathrm{d}x = \int_{\mathbb{R}^q} l_\mu(.,y)\mathrm{d}y = 1$, we write

$$\sigma_1^2(\lambda,\mu) \leq \|f\|_\infty^2. \tag{5.37}$$

Case 2. Upper bound of $\sigma_2^2(\lambda,\mu)$

$$\begin{aligned}
\sigma_2^2(\lambda,\mu) &\leq \mathbb{E}\left[\left(\mathbb{E}\left[k_\lambda(X_1,X_2)l_\mu(Y_3,Y_4)\mid X_1\right]\right)^2\right]\\
&\leq \mathbb{E}\left[\left(\mathbb{E}\left[k_\lambda(X_1,X_2)\mid X_1\right]\right)^2\right]\left(\mathbb{E}\left[l_\mu(Y_3,Y_4)\right]\right)^2\\
&\leq \mathbb{E}\left[k_\lambda(X_1,X_2)k_\lambda(X_1,X_3)\right]\left(\mathbb{E}\left[l_\mu(Y_3,Y_4)\right]\right)^2.
\end{aligned}$$

Moreover, it is easy to see that by upper bounding $f_1(x_2)$ and $f_1(x_3)$ by $\|f_1\|_\infty$, and recalling that $\int_{\mathbb{R}^p} k_\lambda(x_1, x)\mathrm{d}x = 1$, we have,

$$\begin{aligned}\mathbb{E}\left[k_\lambda(X_1, X_2)k_\lambda(X_1, X_3)\right] &= \int_{\mathbb{R}^p}\left[\int_{\mathbb{R}^p} k_\lambda(x_1, x_2)f_1(x_2)\mathrm{d}x_2\right]\left[\int_{\mathbb{R}^p} k_\lambda(x_1, x_3)f_1(x_3)\mathrm{d}x_3\right] f_1(x_1)\mathrm{d}x_1 \\ &\leq \|f_1\|_\infty^2.\end{aligned}$$

Besides, upper bounding $f_2(y_3)$ by $\|f_2\|_\infty$ in the integral form of $\mathbb{E}\left[l_\mu(Y_3, Y_4)\right]$ gives

$$\mathbb{E}\left[l_\mu(Y_3, Y_4)\right] \leq \|f_2\|_\infty.$$

By combining these inequalities, we obtain

$$\sigma_2^2(\lambda, \mu) \leq \|f_1\|_\infty^2\|f_2\|_\infty^2. \tag{5.38}$$

Case 3. Upper bound of $\sigma_3^2(\lambda, \mu)$
This case is similar to case 2 by exchanging X by Y and k_λ by l_μ. Thus, we have the inequality

$$\sigma_3^2(\lambda, \mu) \leq \|f_1\|_\infty^2\|f_2\|_\infty^2. \tag{5.39}$$

Case 4. Upper bound of $\sigma_4^2(\lambda, \mu)$

$$\begin{aligned}\sigma_4^2(\lambda, \mu) &\leq \mathbb{E}\left[\left(\mathbb{E}\left[k_\lambda(X_1, X_2)l_\mu(Y_1, Y_3) \mid Z_1\right]\right)^2\right] \\ &\leq \mathbb{E}\left[k_\lambda(X_1, X_2)k_\lambda(X_1, X_4)l_\mu(Y_1, Y_3)l_\mu(Y_1, Y_5)\right].\end{aligned}$$

By upper bounding $f_1(x_2)$, $f_1(x_4)$ by $\|f_1\|_\infty$ and $f_2(y_3)$, $f_2(y_5)$ by $\|f_2\|_\infty$ in the integral form of $\mathbb{E}\left[k_\lambda(X_1, X_2)k_\lambda(X_1, X_4)l_\mu(Y_1, Y_3)l_\mu(Y_1, Y_5)\right]$, we obtain

$$\sigma_4^2(\lambda, \mu) \leq \|f_1\|_\infty^2\|f_2\|_\infty^2. \tag{5.40}$$

Case 5. Upper bound of $\sigma_5^2(\lambda, \mu)$

$$\begin{aligned}\sigma_5^2(\lambda, \mu) &\leq \mathbb{E}\left[\left(\mathbb{E}\left[k_\lambda(X_2, X_1)l_\mu(Y_2, Y_3) \mid X_1\right]\right)^2\right] \\ &\leq \mathbb{E}\left[k_\lambda(X_2, X_1)k_\lambda(X_4, X_1)l_\mu(Y_2, Y_3)l_\mu(Y_4, Y_5)\right].\end{aligned}$$

By upper bounding $f(x_2, y_2)$ and $f(x_4, y_4)$ by $\|f\|_\infty$ in the integral form of the last expectation, we have

$$\sigma_5^2(\lambda, \mu) \leq \|f\|_\infty^2. \tag{5.41}$$

Case 6. Upper bound of $\sigma_6^2(\lambda, \mu)$
This case is similar to case 5 by exchanging X by Y and k_λ by l_μ. We have then the inequality

$$\sigma_6^2(\lambda, \mu) \leq \|f\|_\infty^2. \tag{5.42}$$

Finally, by combining inequalities (5.37), (5.38), (5.39), (5.40), (5.41) and (5.42), we have the following inequality

$$\sigma^2(\lambda, \mu) \leq C(M_f). \tag{5.43}$$

5.7.2.2 Upper bound of $s^2(\lambda, \mu)$

Let us first recall that the general term of the U-statistic $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ is written as

$$h_{1,2,3,4}(Z_1, Z_2, Z_3, Z_4) = \frac{1}{4!} \sum_{(u,v,w,t)}^{(1,2,3,4)} k_\lambda(X_u, X_v) \left[l_\mu(Y_u, Y_v) + l_\mu(Y_w, Y_t) - 2l_\mu(Y_u, Y_w) \right].$$

Moreover, all the terms of the last sum have the same distribution. We then have

$$\begin{aligned} s^2(\lambda, \mu) &= \mathrm{Var}_f \left(h_{1,2,3,4}(Z_1, Z_2, Z_3, Z_4) \right) \\ &\leq C \, \mathrm{Var}_f \left(k_\lambda(X_1, X_2) \left[l_\mu(Y_1, Y_2) + l_\mu(Y_3, Y_4) - 2l_\mu(Y_1, Y_3) \right] \right), \end{aligned}$$

It follows that,

$$\begin{aligned} \mathrm{Var}_f \left(h_{1,2,3,4}(Z_1, Z_2, Z_3, Z_4) \right) \leq C \big[& \mathrm{Var}_f \left(k_\lambda(X_1, X_2) l_\mu(Y_1, Y_2) \right) + \mathrm{Var}_f \left(k_\lambda(X_1, X_2) l_\mu(Y_3, Y_4) \right) \\ + \quad & \mathrm{Var}_f \left(k_\lambda(X_1, X_2) l_\mu(Y_1, Y_3) \right) \big] \\ \leq C \big(& \mathbb{E} \left[k_\lambda^2(X_1, X_2) l_\mu^2(Y_1, Y_2) \right] + \mathbb{E} \left[k_\lambda^2(X_1, X_2) l_\mu^2(Y_3, Y_4) \right] \\ + \quad & \mathbb{E} \left[k_\lambda^2(X_1, X_2) l_\mu^2(Y_1, Y_3) \right] \big), \end{aligned}$$

In order to bring back to multivariate normal densities, we express k_λ^2 and l_μ^2 as

$$k_\lambda^2 = \frac{k_{\lambda'}}{(4\pi)^{\frac{p}{2}} \lambda_1 ... \lambda_p} \text{ and } l_\mu^2 = \frac{l_{\mu'}}{(4\pi)^{\frac{q}{2}} \mu_1 ... \mu_q},$$

where $\lambda' = \dfrac{\lambda}{\sqrt{2}}$ and $\mu' = \dfrac{\mu}{\sqrt{2}}$.

Consequently, the expectation $\mathbb{E}\left[k_\lambda^2(X_1, X_2) l_\mu^2(Y_1, Y_2)\right]$ can be expressed as

$$\begin{aligned} \mathbb{E}\left[k_\lambda^2(X_1, X_2) l_\mu^2(Y_1, Y_2)\right] &= \frac{1}{(4\pi)^{\frac{p+q}{2}} \lambda_1 ... \lambda_p \mu_1 ... \mu_q} \mathbb{E}\left[k_{\lambda'}(X_1, X_2) l_{\mu'}(Y_1, Y_2)\right] \\ = \frac{1}{(4\pi)^{\frac{p+q}{2}} \lambda_1 ... \lambda_p \mu_1 ... \mu_q} & \int_{(\mathbb{R}^p \times \mathbb{R}^q)^2} k_{\lambda'}(x_1, x_2) l_{\mu'}(y_1, y_2) f(x_1, y_1) f(x_2, y_2) \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}y_1 \mathrm{d}y_2. \end{aligned}$$

By upper bounding $f(x_2, y_2)$ by $\|f\|_\infty$ in the last integral, we have

$$\begin{aligned} & \int_{(\mathbb{R}^p \times \mathbb{R}^q)^2} k_{\lambda'}(x_1, x_2) l_{\mu'}(y_1, y_2) f(x_1, y_1) f(x_2, y_2) \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}y_1 \mathrm{d}y_2 \\ \leq & \|f\|_\infty \int_{\mathbb{R}^p \times \mathbb{R}^q} \left[\int_{\mathbb{R}^p} k_{\lambda'}(x_1, x_2) \mathrm{d}x_2 \right] \left[\int_{\mathbb{R}^q} l_{\mu'}(y_1, y_2) \mathrm{d}y_2 \right] f(x_1, y_1) \mathrm{d}x_1 \mathrm{d}y_1 \\ = & \|f\|_\infty. \end{aligned}$$

This leads to,

$$\mathbb{E}\left[k_\lambda^2(X_1, X_2) l_\mu^2(Y_1, Y_2)\right] \leq \frac{\|f\|_\infty}{(4\pi)^{\frac{p+q}{2}} \lambda_1 ... \lambda_p \mu_1 ... \mu_q}. \tag{5.44}$$

We can easily show by similar argument that

$$\mathbb{E}\left[k_\lambda^2(X_1, X_2) l_\mu^2(Y_3, Y_4)\right] \leq \frac{\|f_1\|_\infty \|f_2\|_\infty}{(4\pi)^{\frac{p+q}{2}} \lambda_1 ... \lambda_p \mu_1 ... \mu_q}. \tag{5.45}$$

and

$$\mathbb{E}\left[k_\lambda^2(X_1,X_2)l_\mu^2(Y_1,Y_3)\right] \leq \frac{\|f\|_\infty}{(4\pi)^{\frac{p+q}{2}}\lambda_1...\lambda_p\mu_1...\mu_q}. \tag{5.46}$$

From Equations (5.44), (5.45) and (5.46), we have

$$s^2(\lambda,\mu) \leq \frac{C(M_f)}{(4\pi)^{\frac{p+q}{2}}\lambda_1...\lambda_p\mu_1...\mu_q}. \tag{5.47}$$

From Equations (5.43) and (5.47) we obtain the following inequality for $\mathrm{Var}_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu})$

$$\mathrm{Var}_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu}) \leq C\left(M_f,p,q\right)\left\{\frac{1}{n}+\frac{1}{\lambda_1...\lambda_p\mu_1...\mu_q n^2}\right\}.$$

5.7.3 Proof of Proposition 5.2

To give an upper bound for the quantile $q_{1-\alpha}^{\lambda,\mu}$ w.r.t λ and μ, we use concentration inequalities for general U-statistics. However, sharp upper bounds are obtained only for degenerate U-statistics (see e.g. Houdré and Reynaud-Bouret (2003)). We recall that, a U-statistic $U_n = U_n(V_1,...,V_r)$ is degenerate if $\mathbb{E}[U_n \mid V_1,...,V_i] = 0$ for all i in $\{1,...,r-1\}$. The first step to upper bound $q_{1-\alpha}^{\lambda,\mu}$ is then to write $\widehat{\mathrm{HSIC}}_{\lambda,\mu}$ as a sum of degenerate U-statistics. For this, we rely on the ANOVA-decomposition (ANOVA for ANalyse Of VAriance, see e.g. Sobol (2001)) of the symmetrical function $h_{i,j,q,r}$ introduced in Equation (5.35). We then write:

$$h_{i,j,q,r} = \frac{1}{2}\sum_{(t,u)}^{(i,j,q,r)} h_{t,u} + \frac{1}{6}\sum_{(t,u,v)}^{(i,j,q,r)} h_{t,u,v} + \widetilde{h}_{i,j,q,r}, \tag{5.48}$$

where the first (resp. the second) sum represents all ordered pairs (t,u) (resp. triplets (t,u,v)) drawn without replacement from (i,j,q,r) and the terms $h_{t,u}$, $h_{t,u,v}$ and $\widetilde{h}_{i,j,q,r}$ are defined as

$$\begin{aligned}
h_{t,u} &= \mathbb{E}\left[h_{i,j,q,r} \mid Z_t, Z_u\right], \\
h_{t,u,v} &= \mathbb{E}\left[h_{i,j,q,r} \mid Z_t, Z_u, Z_v\right] - \frac{1}{2}\sum_{(t',u')}^{(t,u,v)} h_{t',u'}, \\
\widetilde{h}_{i,j,q,r} &= h_{i,j,q,r} - \frac{1}{6}\sum_{(t,u,v)}^{(i,j,q,r)} h_{t,u,v} - \frac{1}{2}\sum_{(t,u)}^{(i,j,q,r)} h_{t,u}.
\end{aligned}$$

Hence, by summing all terms $h_{i,j,q,r}$ for (i,j,q,r) in $\mathbf{i}_4^n$ and then dividing by $n(n-1)(n-2)(n-3)$, we have

$$\widehat{\mathrm{HSIC}}_{\lambda,\mu} = 6\widehat{\mathrm{HSIC}}_{\lambda,\mu}^{(2,D)} + 4\widehat{\mathrm{HSIC}}_{\lambda,\mu}^{(3,D)} + \widehat{\mathrm{HSIC}}_{\lambda,\mu}^{(4,D)}, \tag{5.49}$$

where

$$\widehat{\mathrm{HSIC}}_{\lambda,\mu}^{(2,D)} = \frac{1}{n(n-1)}\sum_{(i,j)\in\mathbf{i}_2^n} h_{i,j} \quad , \quad \widehat{\mathrm{HSIC}}_{\lambda,\mu}^{(3,D)} = \frac{1}{n(n-1)(n-2)}\sum_{(i,j,q)\in\mathbf{i}_3^n} h_{i,j,q}$$

$$\widehat{\mathrm{HSIC}}_{\lambda,\mu}^{(4,D)} = \frac{1}{n(n-1)(n-2)(n-3)}\sum_{(i,j,q,r)\in\mathbf{i}_4^n} \widetilde{h}_{i,j,q,r}.$$

Lemma 5.8. *Let us assume that $f = f_1 \otimes f_2$. Then, the U-statistics $\widehat{\text{HSIC}}_{\lambda,\mu}^{(2,D)}$, $\widehat{\text{HSIC}}_{\lambda,\mu}^{(3,D)}$ and $\widehat{\text{HSIC}}_{\lambda,\mu}^{(4,D)}$ are degenerated.*

Proof. According to Theorem 2 of Gretton et al. (2008), if $f = f_1 \otimes f_2$, we have:

$$\mathbb{E}[h_{i,j,q,r} \mid Z_i] = 0.$$

We then easily show that $\widehat{\text{HSIC}}_{\lambda,\mu}^{(2,D)}$ is degenerated by writing

$$\mathbb{E}[h_{i,j} \mid Z_i] = \mathbb{E}[h_{i,j,q,r} \mid Z_i] = 0. \tag{5.50}$$

Moreover, to prove that $\widehat{\text{HSIC}}_{\lambda,\mu}^{(3,D)}$ is degenerated, we have

$$\begin{aligned}\mathbb{E}[h_{i,j,q} \mid Z_i, Z_j] =& \mathbb{E}[h_{i,j,q,r} \mid Z_i, Z_j] - \mathbb{E}[h_{i,j} \mid Z_i, Z_j] - \mathbb{E}[h_{i,q} \mid Z_i] - \mathbb{E}[h_{j,q} \mid Z_j] \\ =& h_{i,j} - h_{i,j} \quad \text{(by definition of } h_{i,j} \text{ and Equation (5.50))} \\ =& 0.\end{aligned} \tag{5.51}$$

Finally, to show that $\widehat{\text{HSIC}}_{\lambda,\mu}^{(4,D)}$ is degenerated, we write

$$\begin{aligned}\mathbb{E}[\widetilde{h}_{i,j,q,r} \mid Z_i, Z_j, Z_q] =& \mathbb{E}[h_{i,j,q,r} \mid Z_i, Z_j, Z_q] - h_{i,j,q} - h_{i,j} - h_{i,q} - h_{j,q} \\ =& 0.\end{aligned} \tag{5.52}$$

□

Once we have upper bounds of the $(1-\alpha)$-quantiles of $\widehat{\text{HSIC}}_{\lambda,\mu}^{(r,D)}$ with r in $\{2,3,4\}$ under the assumption $f = f_1 \otimes f_2$, an upper bound of the quantile $q_{1-\alpha}^{\lambda,\mu}$ is naturally obtained. In fact, we can easily show that,

$$q_{1-\alpha}^{\lambda,\mu} \leq 6q_{1-\alpha/3,2}^{\lambda,\mu} + 4q_{1-\alpha/3,3}^{\lambda,\mu} + q_{1-\alpha/3,4}^{\lambda,\mu}$$

where $q_{1-\alpha,r}^{\lambda,\mu}$ is the $(1-\alpha)$-quantiles of $\widehat{\text{HSIC}}_{\lambda,\mu}^{(r,D)}$ under the assumption $f = f_1 \otimes f_2$.

5.7.3.1 Upper bound of $q_{1-\alpha,2}^{\lambda,\mu}$

In this part, we give an upper bound of $q_{1-\alpha,2}^{\lambda,\mu}$. For this, we use the concentration inequality 3.5, page 15 of Giné et al. (2000), given for degenerated U-statistics of order 2. We write for all $t > 0$:

$$\mathbb{P}\left(\left|\sum_{(i,j)\in i_2^n} h_{i,j}\right| > t\right) \leq A \exp\left(-\frac{1}{A}\min\left[\frac{t}{M}, \left(\frac{t}{L}\right)^{2/3}, \left(\frac{t}{K}\right)^{1/2}\right]\right), \tag{5.53}$$

where A is an absolute positive constant,

$$K = \max_{(i,j)\in i_2^n} \|h_{i,j}\|_\infty,\ M^2 = \sum_{(i,j)\in i_2^n} \mathbb{E}[h_{i,j}^2]$$

$$L^2 = \max\left[\left\|\sum_{i=1}^n \mathbb{E}\left[h_{i,j}^2(Z_i, y)\right]\right\|_\infty, \left\|\sum_{j=1}^n \mathbb{E}\left[h_{i,j}^2(x, Z_j)\right]\right\|_\infty\right].$$

By setting $\varepsilon = \frac{t}{n^2}$, and using Equation (5.53), we obtain

$$\mathbb{P}\left(\frac{1}{n^2}|\sum_{(i,j)\in i_2^n} h_{i,j}| > \varepsilon\right) \leq A\exp\left(-\frac{1}{A}\min\left[\frac{n^2\varepsilon}{M}, \left(\frac{n^2\varepsilon}{L}\right)^{2/3}, \left(\frac{n^2\varepsilon}{K}\right)^{1/2}\right]\right).$$

Therefore, we have for all $\varepsilon > 0$,

$$\mathbb{P}\left(\frac{1}{n^2}|\sum_{(i,j)\in i_2^n} h_{i,j}| > \varepsilon\right) \leq A\max\left[\exp\left(-\frac{n^2\varepsilon}{AM}\right), \exp\left(-\frac{n^{4/3}\varepsilon^{2/3}}{AL^{2/3}}\right), \exp\left(-\frac{n\varepsilon^{1/2}}{AK^{1/2}}\right)\right].$$

By adjusting the constant A, we can replace in the last inequality $\frac{1}{n^2}\sum_{(i,j)\in i_2^n} h_{i,j}$ by $\widehat{\mathrm{HSIC}}_{\lambda,\mu}^{(2,D)}$,

$$\mathbb{P}\left(|\widehat{\mathrm{HSIC}}_{\lambda,\mu}^{(2,D)}| > \varepsilon\right) \leq A\max\left[\exp\left(-\frac{n^2\varepsilon}{AM}\right), \exp\left(-\frac{n^{4/3}\varepsilon^{2/3}}{AL^{2/3}}\right), \exp\left(-\frac{n\varepsilon^{1/2}}{AK^{1/2}}\right)\right].$$

Furthermore, if ε_α is a positive number verifying

$$\alpha = A\max\left[\exp\left(-\frac{n^2\varepsilon_\alpha}{AM}\right), \exp\left(-\frac{n^{4/3}\varepsilon_\alpha^{2/3}}{AL^{2/3}}\right), \exp\left(-\frac{n\varepsilon_\alpha^{1/2}}{AK^{1/2}}\right)\right].$$

Then, we can easily show the following inequality

$$q_{1-\alpha,2}^{\lambda,\mu} \leq \varepsilon_\alpha. \tag{5.54}$$

By now, we upper bound ε_α (and consequently $q_{1-\alpha,2}^{\lambda,\mu}$), in the 3 following cases.

Case 1. $\alpha = A\exp\left(-\frac{n^2\varepsilon_\alpha}{AM}\right)$

In this case, ε_α is expressed as

$$\varepsilon_\alpha = \frac{AM}{n^2}\left(\log\left(\frac{1}{\alpha}\right) + \log(A)\right).$$

We can then upper bound ε_α as

$$\varepsilon_\alpha \leq \frac{CM}{n^2}\left(\log\left(\frac{1}{\alpha}\right) + 1\right),$$

for some absolute positive constant C. Furthermore, considering the values of α such that $\log\left(\frac{1}{\alpha}\right) > 1$, we obtain

$$\varepsilon_\alpha \leq \frac{2CM}{n^2}\log\left(\frac{1}{\alpha}\right). \tag{5.55}$$

Let us upper bound M w.r.t λ, μ and n. For this, we first write

$$M^2 = \sum_{(i,j)\in i_2^n} \mathbb{E}[h_{i,j}^2] \leq n^2\,\mathbb{E}[h_{1,2}^2].$$

Moreover, using the law of total variance, we have under the hypothesis $f = f_1 \otimes f_2$,

$$\begin{aligned}\mathbb{E}[h_{1,2}^2] &= \mathrm{Var}\left(\mathbb{E}[h_{1,2,3,4} \mid Z_1, Z_2]\right) \\ &\leq \mathrm{Var}\left(h_{1,2,3,4}\right).\end{aligned}$$

Furthermore, we have shown in Annexe 5.7.2.2 that,

$$\mathrm{Var}\left(h_{1,2,3,4}\right) \leq \frac{C\left(M_f, p, q\right)}{\lambda_1 ... \lambda_p \mu_1 ... \mu_q}.$$

Hence, we can upper bound M as follows,

$$M \leq \frac{C\left(M_f, p, q\right) n}{\sqrt{\lambda_1 ... \lambda_p \mu_1 ... \mu_q}}. \tag{5.56}$$

Consequently, by combining Equations (5.55) and (5.56), we obtain

$$q_{1-\alpha,2}^{\lambda,\mu} \leq \frac{C\left(M_f, p, q\right)}{n\sqrt{\lambda_1 ... \lambda_p \mu_1 ... \mu_q}} \log\left(\frac{1}{\alpha}\right). \tag{5.57}$$

Case 2. $\alpha = A \exp\left(-\dfrac{n^{4/3}\varepsilon_\alpha^{2/3}}{AL^{2/3}}\right)$

In this case, ε_α verifies that,

$$\varepsilon_\alpha^{2/3} = \frac{AL^{2/3}}{n^{4/3}}\left(\log\left(\frac{1}{\alpha}\right) + \log\left(A\right)\right).$$

Thus, ε_α can be upper bounded as

$$\varepsilon_\alpha \leq \frac{CL}{n^2} \log\left(\frac{1}{\alpha}\right)^{3/2}, \tag{5.58}$$

Let us upper bound L w.r.t n, λ and μ. For this, knowing that $h_{i,j}$ is symmetrical we write

$$L^2 = \| \sum_{i=1}^{n} \mathbb{E}[h_{i,j}^2(Z_i, y)] \|_\infty.$$

Moreover, according to Gretton et al. (2008) page 10, we have under the hypothesis $f = f_1 \otimes f_2$,

$$h_{i,j}(Z_i, Z_j) = \frac{1}{6}\left[k_\lambda(X_i, X_j) + (k_\lambda)_{.,.} - (k_\lambda)_{i,.} - (k_\lambda)_{.,j}\right]\left[l_\mu(Y_i, Y_j) + (l_\mu)_{.,.} - (l_\mu)_{i,.} - (l_\mu)_{.,j}\right],$$

where $(k_\lambda)_{.,.} = \mathbb{E}[k_\lambda(X_i, X_j)]$, $(k_\lambda)_{i,.} = \mathbb{E}[k_\lambda(X_i, X_j) \mid X_i]$, $(k_\lambda)_{.,j} = \mathbb{E}[k_\lambda(X_i, X_j) \mid X_j]$ and $(l_\mu)_{.,.}$, $(l_\mu)_{i,.}$, $(l_\mu)_{.,j}$ are defined in a similar way.

Hence, we write for all $y = (y_1, y_2) \in \mathbb{R}^2$,

$$\begin{aligned}h_{i,j}^2(Z_i, y) = \frac{1}{36}&\left[k_\lambda(X_i, y_1) + (k_\lambda)_{.,.} - \mathbb{E}[k_\lambda(X_i, X_j) \mid X_i] - \mathbb{E}[k_\lambda(X_i, y_1)]\right]^2 \\ &\times \left[l_\mu(Y_i, y_2) + (l_\mu)_{.,.} - \mathbb{E}[l_\mu(Y_i, Y_j) \mid Y_i] - \mathbb{E}[l_\mu(Y_i, y_2)]\right]^2.\end{aligned}$$

Therefore, we have the following inequality for $h^2_{i,j}(Z_i, y)$,

$$h^2_{i,j}(Z_i,y) \leq C\left[k_\lambda(X_i,y_1)^2 + (k_\lambda)^2_{.,.} + (\mathbb{E}[k_\lambda(X_i,X_j) \mid X_i])^2 + (\mathbb{E}[k_\lambda(X_i,y_1])^2\right] \times\left[l_\mu(Y_i,y_2)^2 + (l_\mu)^2_{.,.} + (\mathbb{E}[l_\mu(Y_i,Y_j) \mid Y_i])^2 + (\mathbb{E}[l_\mu(Y_i,y_2])^2\right].$$

Using that $(X_1,\ldots,X_n)$ and $(Y_1,\ldots,Y_n)$ are independent, we write

$$L^2 \leq Cn \sup_{y_1,y_2\in\mathbb{R}} \mathbb{E}\left[k_\lambda(X_i,y_1)^2 + (k_\lambda)^2_{.,.} + (\mathbb{E}[k_\lambda(X_i,X_j) \mid X_i])^2 + (\mathbb{E}[k_\lambda(X_i,y_1])^2\right] \times\mathbb{E}\left[l_\mu(Y_i,y_2)^2 + (l_\mu)^2_{.,.} + (\mathbb{E}[l_\mu(Y_i,Y_j) \mid Y_i])^2 + (\mathbb{E}[l_\mu(Y_i,y_2])^2\right].$$

Each term can be upper bounded by similar arguments as 5.7.2.2, we then have

$$L^2 \leq C\left(M_f\right) n \left(1+\frac{1}{\lambda_1 ... \lambda_p}\right)\left(1+\frac{1}{\mu_1 ... \mu_q}\right).$$

Thus, using that $\lambda_1 ... \lambda_p < 1$ and $\mu_1 ... \mu_q < 1$, we obtain:

$$L \leq \frac{C\left(M_f\right)\sqrt{n}}{\sqrt{\lambda_1 ... \lambda_p \mu_1 ... \mu_q}}. \tag{5.59}$$

By combining Equations (5.58) and (5.59), we have

$$\varepsilon_\alpha \leq \frac{C\left(M_f\right)}{\sqrt{\lambda_1 ... \lambda_p \mu_1 ... \mu_q} n^{3/2}}\left[\log\left(\frac{1}{\alpha}\right)\right]^{3/2}.$$

Moreover, knowing that $\lambda_1 ... \lambda_p \mu_1 ... \mu_q < 1$, we obtain

$$\varepsilon_\alpha \leq \frac{C\left(M_f\right)}{(n\sqrt{\lambda_1 ... \lambda_p \mu_1 ... \mu_q})^{3/2}}\left[\log\left(\frac{1}{\alpha}\right)\right]^{3/2}. \tag{5.60}$$

Case 3. $\alpha = A\exp\left(-\frac{n\varepsilon_\alpha^{1/2}}{AK^{1/2}}\right)$

In this case, ε_α is expressed as

$$\varepsilon_\alpha^{1/2} = \frac{AK^{1/2}}{n}\left(\log\left(\frac{1}{\alpha}\right) + \log\left(A\right)\right).$$

Using that $\log\left(\frac{1}{\alpha}\right) > 1$ and by adjusting the value of A, we upper bound ε_α as

$$\varepsilon_\alpha \leq \frac{AK}{n^2}\left[\log\left(\frac{1}{\alpha}\right)\right]^2. \tag{5.61}$$

Morover, we can easily show that

$$K \leq \frac{4}{\lambda_1 ... \lambda_p \mu_1 ... \mu_q}. \tag{5.62}$$

By combining Equations (5.61) and (5.62), we obtain:

$$q^{\lambda,\mu}_{1-\alpha,2} \leq \frac{C}{\lambda_1 ... \lambda_p \mu_1 ... \mu_q n^2} \left[\log\left(\frac{1}{\alpha}\right)\right]^2. \tag{5.63}$$

using (5.57), (5.60) and (5.63) and the fact that $\frac{1}{\sqrt{\lambda_1 ... \lambda_p \mu_1 ... \mu_q} n} \log\left(\frac{1}{\alpha}\right) < 1$, we have the following inequality

$$q^{\lambda,\mu}_{1-\alpha,2} \leq \frac{C\left(\|f_1\|_\infty, \|f_2\|_\infty, p, q\right)}{n\sqrt{\lambda_1 ... \lambda_p \mu_1 ... \mu_q}} \log\left(\frac{1}{\alpha}\right). \tag{5.64}$$

5.7.3.2 Upper bound of $q^{\lambda,\mu}_{1-\alpha,3}$

In this part, we give an upper bound for the $(1-\alpha)$-quantile of $\widehat{\text{HSIC}}^{(3,D)}_{\lambda,\mu}$. For this, we use the concentration inequality (c), page 1501 of Arcones and Gine (1993). We write for all $t > 0$,

$$\mathbb{P}\left(n^{-3/2}|\sum_{(i,j,q)\in i^n_3} h_{i,j,q}| > t\right) \leq A \exp\left[-\frac{Bt^{2/3}}{M^{2/3} + K^{1/2} t^{1/6} n^{-1/4}}\right], \tag{5.65}$$

where $K = \|h_{i,j,q}\|_\infty$, $M^2 = \mathbb{E}[h^2_{1,2,3}]$ and B is an absolute positive constant.

By setting $\varepsilon = \frac{t}{n^{3/2}}$ and using Equation (5.65), we have

$$\mathbb{P}\left(\frac{1}{n^3}|\sum_{(i,j,q)\in i^n_3} h_{i,j,q}| > \varepsilon\right) \leq A \exp\left[-\frac{Bn\varepsilon^{2/3}}{M^{2/3} + K^{1/2}\varepsilon^{1/6}}\right].$$

Moreover, by adjusting the value of B, we can write

$$\mathbb{P}\left(|\widehat{\text{HSIC}}^{(3,D)}_{\lambda,\mu}| > \varepsilon\right) \leq A \exp\left[-\frac{Bn\varepsilon^{2/3}}{M^{2/3} + K^{1/2}\varepsilon^{1/6}}\right]. \tag{5.66}$$

Furthermore, if ε_α is a positive number verifying

$$A \exp\left[-\frac{Bn\varepsilon_\alpha^{2/3}}{M^{2/3} + K^{1/2}\varepsilon_\alpha^{1/6}}\right] = \alpha, \tag{5.67}$$

then, we have the following inequality

$$q^{\lambda,\mu}_{1-\alpha,3} \leq \varepsilon_\alpha.$$

In order to upper bound ε_α in (5.67), we set $\gamma_\alpha = \varepsilon_\alpha^{1/6}$ and we obtain

$$Bn\gamma_\alpha^4 = K^{1/2} \log\left(\frac{A}{\alpha}\right)\gamma_\alpha + M^{2/3} \log\left(\frac{A}{\alpha}\right). \tag{5.68}$$

The polynomial Equation (5.68) is not resolvable. However, it is possible to give an upper bound of its roots. Indeed,

$$Bn\gamma_\alpha^4 \leq 2\max\left[K^{1/2}\gamma_\alpha, M^{2/3}\right] \log\left(\frac{A}{\alpha}\right).$$

Case 1. $\max\left[K^{1/2}\gamma_\alpha, M^{2/3}\right] = K^{1/2}\gamma_\alpha$
In this case, γ_α verifies the following inequality,

$$\gamma_\alpha^3 \leq \frac{2K^{1/2}}{Bn}\log\left(\frac{A}{\alpha}\right).$$

Hence,

$$\varepsilon_\alpha \leq \frac{4K}{B^2n^2}\left(\log\left(\frac{A}{\alpha}\right)\right)^2.$$

Since $K \leq \frac{4}{\lambda_1\ldots\lambda_p\mu_1\ldots\mu_q}$, $\frac{1}{\sqrt{\lambda_1\ldots\lambda_p\mu_1\ldots\mu_q}n}\log\left(\frac{1}{\alpha}\right) < 1$ and $\log\left(\frac{1}{\alpha}\right) > 1$, we have

$$\varepsilon_\alpha \leq \frac{C}{n\sqrt{\lambda_1\ldots\lambda_p\mu_1\ldots\mu_q}}\log\left(\frac{1}{\alpha}\right).$$

Case 2. $\max\left[K^{1/2}\gamma_\alpha, M^{2/3}\right] = M^{2/3}$
In this case,

$$\gamma_\alpha^4 \leq \frac{M^{2/3}}{Bn}\left[\log\left(\frac{A}{\alpha}\right)\right].$$

Therefore, ε_α can be upper bounded as

$$\varepsilon_\alpha \leq \frac{M}{(Bn)^{3/2}}\left[\log\left(\frac{A}{\alpha}\right)\right]^{3/2}.$$

Moreover, using the law of total variance, it is easy to see that under the hypothesis $f = f_1 \otimes f_2$,

$$M^2 = \mathrm{Var}\left(h_{1,2,3}\right) \leq C\,\mathrm{Var}\left(h_{1,2,3,4}\right). \tag{5.69}$$

Then, according to Annexe 5.7.2.2, M can be upper bounded as

$$M \leq \frac{C(M_f, p, q)}{\sqrt{\lambda_1\ldots\lambda_p\mu_1\ldots\mu_q}}.$$

Hence,

$$\varepsilon_\alpha \leq \frac{C(M_f, p, q)}{n\sqrt{\lambda_1\ldots\lambda_p\mu_1\ldots\mu_q}}\log(\frac{1}{\alpha}),$$

since $\frac{1}{n}\log(\frac{1}{\alpha}) < 1$. To conclude, in all cases we have the following inequality for $q_{1-\alpha,3}^{\lambda,\mu}$

$$q_{1-\alpha,3}^{\lambda,\mu} \leq \frac{C(\|f_1\|_\infty, \|f_2\|_\infty, p, q)}{n\sqrt{\lambda_1\ldots\lambda_p\mu_1\ldots\mu_q}}\log\left(\frac{1}{\alpha}\right).$$

5.7.3.3 Upper bound of $q_{1-\alpha,4}^{\lambda,\mu}$

In this part, we give an upper bound for the $(1-\alpha)$-quantile of $\widehat{\mathrm{HSIC}}_{\lambda,\mu}^{(4,D)}$. For this, we use the concentration inequality (d), page 1501 of Arcones and Gine (1993). We have for all $t > 0$:

$$\mathbb{P}\left(\frac{1}{n^2}\Big|\sum_{(i,j,q,r)\in i_4^n}\widetilde{h}_{i,j,q,r}\Big| > t\right) \leq A\exp\left(-B\sqrt{\frac{t}{K}}\right),$$

where A and B are absolute positive constants and $K = \|\widetilde{h}_{1,2,3,4}\|_\infty$.

By setting $\varepsilon = \dfrac{t}{n^2}$, we have

$$\mathbb{P}\left(\frac{1}{n^4}|\sum_{(i,j,q,r)\in i_4^n}\widetilde{h}_{i,j,q,r}| > \varepsilon\right) \leq A\exp\left(-Bn\sqrt{\frac{\varepsilon}{K}}\right).$$

Furthermore, by adjusting the constant B, we replace $\dfrac{1}{n^4}\sum_{(i,j,q,r)\in i_4^n}\widetilde{h}_{i,j,q,r}$ by $\widehat{\mathrm{HSIC}}^{(4,D)}_{\lambda,\mu}$. We write

$$\mathbb{P}\left(|\widehat{\mathrm{HSIC}}^{(4,D)}_{\lambda,\mu}| > \varepsilon\right) \leq A\exp\left(-Bn\sqrt{\frac{\varepsilon}{K}}\right). \tag{5.70}$$

Moreover, if ε_α is a positive number verifying

$$A\exp\left(-Bn\sqrt{\frac{\varepsilon_\alpha}{K}}\right) = \alpha, \tag{5.71}$$

then,

$$q^{\lambda,\mu}_{1-\alpha,4} \leq \varepsilon_\alpha.$$

By resolving Equation (5.71), we obtain

$$\varepsilon_\alpha = \frac{BK}{n^2}\left[\log\left(\frac{A}{\alpha}\right)\right]^2.$$

Therefore, we can easily show that

$$\varepsilon_\alpha \leq \frac{CK}{n^2}\left[\log\left(\frac{1}{\alpha}\right)\right]^2.$$

Moreover, by using the inequality $K \leq \dfrac{4}{\lambda_1...\lambda_p\mu_1...\mu_q}$ we have

$$q^{\lambda,\mu}_{1-\alpha,4} \leq \frac{C}{\lambda_1...\lambda_p\mu_1...\mu_q n^2}\left(\log\left(\frac{1}{\alpha}\right)\right)^2.$$

Consequently,

$$q^{\lambda,\mu}_{1-\alpha,4} \leq \frac{C}{n\sqrt{\lambda_1...\lambda_p\mu_1...\mu_q}}\log\left(\frac{1}{\alpha}\right). \tag{5.72}$$

To conclude, under the hypothesis of Proposition 5.2, the quantile $q^{\lambda,\mu}_{1-\alpha}$ can be upper bounded under the hypothesis $f = f_1 \otimes f_2$ as follows,

$$q^{\lambda,\mu}_{1-\alpha} \leq \frac{C\left(\|f_1\|_\infty, \|f_2\|_\infty, p, q\right)}{n\sqrt{\lambda_1...\lambda_p\mu_1...\mu_q}}\log\left(\frac{1}{\alpha}\right). \tag{5.73}$$

5.7.4 Proof of Corollary 5.1

The proof of this corollary is immediately obtained from Lemma 5.1, Proposition 5.1 and Proposition 5.2.

5.7.5 Proof of Lemma 5.2

Recalling the formulation of $\mathrm{HSIC}_{\lambda,\mu}(f)$ given in Equation (3.1) with $k = k_\lambda$ and $l = l_\mu$, we obtain

$$\begin{aligned}\mathrm{HSIC}_{\lambda,\mu}(f) = &\int_{(\mathbb{R}^p\times\mathbb{R}^q)^2} k_\lambda(x,x')l_\mu(y,y')f(x,y)f(x',y')\mathrm{d}x\mathrm{d}y\mathrm{d}x'\mathrm{d}y' \\ &-2\int_{(\mathbb{R}^p\times\mathbb{R}^q)^2} k_\lambda(x,x')l_\mu(y,y')f(x,y)f_1(x')f_2(y')\mathrm{d}x\mathrm{d}y\mathrm{d}x'\mathrm{d}y' \\ &+\int_{(\mathbb{R}^p\times\mathbb{R}^q)^2} k_\lambda(x,x')l_\mu(y,y')f_1(x)f_2(y)f_1(x')f_2(y')\mathrm{d}x\mathrm{d}y\mathrm{d}x'\mathrm{d}y'.\end{aligned}$$

This expression can be factorized using the symmetry of the kernels k_λ and l_μ:

$$\begin{aligned}\mathrm{HSIC}_{\lambda,\mu}(f) &= \int_{(\mathbb{R}^p\times\mathbb{R}^q)^2} k_\lambda(x,x')l_\mu(y,y')\Big[f(x,y)-f_1(x)f_2(y)\Big]\Big[f(x',y')-f_1(x')f_2(y')\Big]\mathrm{d}x\mathrm{d}y\mathrm{d}x'\mathrm{d}y' \\ &= \int_{(\mathbb{R}^p\times\mathbb{R}^q)^2} k_\lambda(x,x')l_\mu(y,y')\psi(x,y)\psi(x',y')\mathrm{d}x\mathrm{d}y\mathrm{d}x'\mathrm{d}y',\end{aligned}$$

where $\psi(x,y) = f(x,y) - f_1(x)f_2(y)$.

Thereafter, we reformulate this equation by replacing $k_\lambda(x,x')$ with $\varphi_\lambda(x-x')$ and replacing $l_\lambda(y,y')$ with $\phi_\mu(y-y')$, where φ_λ and ϕ_μ are respectively the functions defined in Equations (5.2) and (5.3):

$$\begin{aligned}\mathrm{HSIC}_{\lambda,\mu}(f) &= \int_{\mathbb{R}^p\times\mathbb{R}^q} \psi(x,y)\left[\int_{\mathbb{R}^p\times\mathbb{R}^q}\psi(x',y')\varphi_\lambda(x-x')\phi_\mu(y-y')\mathrm{d}x'\mathrm{d}y'\right]\mathrm{d}x\mathrm{d}y \\ &= \int_{\mathbb{R}^p\times\mathbb{R}^q} \psi(x,y)\left[\psi*(\varphi_\lambda\otimes\phi_\mu)\right](x,y)\,\mathrm{d}x\mathrm{d}y \\ &= \langle\psi,\psi*(\varphi_\lambda\otimes\phi_\mu)\rangle_2.\end{aligned}$$

5.7.6 Proof of Proposition 5.3

First notice that according to Equations (5.36) and (5.47), one can write:

$$\mathrm{Var}_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu}) \le \frac{C}{n}\mathrm{Var}_f\left(\mathbb{E}[h_{1,2,3,4}\mid Z_1]\right) + \frac{C\left(M_f,p,q\right)}{\lambda_1\ldots\lambda_p\mu_1\ldots\mu_q n^2}, \tag{5.74}$$

where $h_{1,2,3,4}$ is defined in Equation (5.35).

To prove the intended result from the last equation, we aim now to upper bound $\mathrm{Var}_f\left(\mathbb{E}[h_{1,2,3,4}\mid Z_1]\right)$ by $\|\psi*(\varphi_\lambda\otimes\phi_\mu)\|^2_{\mathbb{L}_2}$ up to a positive constant which depends only on M_f. The following lemma gives such an upper bound.

Lemma 5.9. *For all λ in $(0,+\infty)^p$ and μ in $(0,+\infty)^q$, we have*

$$\mathrm{Var}_f\left(\mathbb{E}[h_{1,2,3,4}\mid Z_1]\right) \le C(M_f)\,\|\psi*(\varphi_\lambda\otimes\phi_\mu)\|^2_{\mathbb{L}_2}.$$

Proof. The first step to upper bound $\mathrm{Var}_f\left(\mathbb{E}[h_{1,2,3,4} \mid Z_1]\right)$ is to rewrite $h_{1,2,3,4}$ by isolating all the terms depending on Z_1.

$$\begin{aligned} h_{1,2,3,4} =& \frac{1}{4!} \sum_{(t,u,v,w)}^{(1,2,3,4)} [k_{t,u} l_{t,u} + k_{t,u} l_{v,w} - 2k_{t,u} l_{t,v}] \\ =& \frac{2}{4!} \sum_{(u,v,w)}^{(2,3,4)} [k_{1,u} l_{1,u} + k_{1,u} l_{v,w} + k_{u,v} l_{1,w} - k_{w,v} l_{w,1} - k_{u,1} l_{u,v} - k_{1,u} l_{1,v}] + R(Z_2, Z_3, Z_4), \end{aligned}$$

where the last sum represents all triplets (u, v, w) drawn without replacement from $(2, 3, 4)$ and $R(Z_2, Z_3, Z_4)$ is a random variable depending only on Z_2, Z_3 and Z_4.

Then,

$$h_{1,2,3,4} = R(Z_2, Z_3, Z_4) + \frac{1}{12} \sum_{(u,v,w)}^{(2,3,4)} [k_{1,u}(l_{1,u} - l_{1,v}) - k_{u,1}(l_{u,v} - l_{v,w}) - (k_{w,v} - k_{u,v}) l_{1,w}] .$$

The random variable $R(Z_2, Z_3, Z_4)$ being independent from Z_1, the variance of its expectation conditionally to Z_1 is equal to 0. It is then easy to see that $\mathrm{Var}_f\left(\mathbb{E}[h_{1,2,3,4} \mid Z_1]\right)$ can be upper bounded as follows:

$$\begin{aligned} \mathrm{Var}_f\left(\mathbb{E}[h_{1,2,3,4} \mid Z_1]\right) \leq C \big[& \mathrm{Var}_f\left(\mathbb{E}[k_{1,2}(l_{1,2} - l_{1,3}) \mid Z_1]\right) + \mathrm{Var}_f\left(\mathbb{E}[k_{2,1}(l_{2,3} - l_{3,4}) \mid X_1]\right) \\ & + \mathrm{Var}_f\left(\mathbb{E}[(k_{2,3} - k_{4,3}) l_{1,2} \mid Y_1]\right) \big], \end{aligned} \tag{5.75}$$

By now, we reformulate the function $\psi * (\varphi_\lambda \otimes \phi_\mu)$ in a simpler form in order to link its $\mathbb{L}_2$-norm with the upper bound given in Equation (5.75). For notational convenience, we denote $G_{\lambda,\mu} = \psi * (\varphi_\lambda \otimes \phi_\mu)$. We then write

$$\begin{aligned} G_{\lambda,\mu}(x, y) &= \int_{\mathbb{R}^p \times \mathbb{R}^q} \psi(x', y') k_\lambda(x, x') l_\mu(y, y') \, \mathrm{d}x' \mathrm{d}y' \\ &= \int_{\mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^q} k_\lambda(x, x') \Big(l_\mu(y, y') - l_\mu(y, y'') \Big) \, f(x', y') f_2(y'') \mathrm{d}x' \mathrm{d}y' \mathrm{d}y'' \\ &= \mathbb{E}\Big[k_\lambda(x, X') \Big(l_\mu(y, Y') - l_\mu(y, Y'') \Big) \Big] , \end{aligned}$$

where (X', Y') and Y'' are independent random variables with respective densities f and f_2.

Thereafter, the conditional expectations in Equation (5.75) can all be expressed as follows:

$$\begin{aligned} \mathbb{E}[k_{1,2}(l_{1,2} - l_{1,3}) \mid Z_1] &= G_{\lambda,\mu}(X_1, Y_1), \\ \mathbb{E}[k_{2,1}(l_{2,3} - l_{3,4}) \mid X_1] &= \mathbb{E}\left[G_{\lambda,\mu}(X_1, Y_3) \mid X_1 \right], \\ \mathbb{E}[(k_{2,3} - k_{4,3}) l_{1,2} \mid Y_1] &= \mathbb{E}\left[G_{\lambda,\mu}(X_3, Y_1) \mid Y_1 \right]. \end{aligned}$$

Thus, using the law of total variance Weiss (2006), we have the following upper bound for $\mathrm{Var}_f\left(\mathbb{E}[h_{1,2,3,4} \mid Z_1]\right)$:

$$\mathrm{Var}_f\left(\mathbb{E}[h_{1,2,3,4} \mid Z_1]\right) \leq C \Big[\mathrm{Var}_f\left(G_{\lambda,\mu}(X_1, Y_1)\right) + \mathrm{Var}_f\left(G_{\lambda,\mu}(X_1, Y_3)\right) + \mathrm{Var}_f\left(G_{\lambda,\mu}(X_3, Y_1)\right) \Big].$$

On the other hand, it is straightforward to upper bound the three variances in the last equation as

$$\begin{aligned}
\mathrm{Var}_f\left(G_{\lambda,\mu}(X_1,Y_1)\right) &\leq \|f\|_\infty \|G_{\lambda,\mu}\|_{\mathbb{L}_2}^2 ,\\
\mathrm{Var}_f\left(G_{\lambda,\mu}(X_1,Y_3)\right) &\leq \|f_1\otimes f_2\|_\infty \|G_{\lambda,\mu}\|_{\mathbb{L}_2}^2 ,\\
\mathrm{Var}_f\left(G_{\lambda,\mu}(X_3,Y_1)\right) &\leq \|f_1\otimes f_2\|_\infty \|G_{\lambda,\mu}\|_{\mathbb{L}_2}^2 .
\end{aligned}$$

Consequently, combining the three last Equations with Equation (5.75) gives us the following upper bound of $\mathrm{Var}_f\left(\mathbb{E}[h_{1,2,3,4} \mid Z_1]\right)$:

$$\mathrm{Var}_f\left(\mathbb{E}[h_{1,2,3,4} \mid Z_1]\right) \leq C(M_f)\,\|\psi * (\varphi_\lambda \otimes \phi_\mu)\|_{\mathbb{L}_2}^2 .$$

□

We then obtain as a result of Equation (5.74) and Lemma 5.9:

$$\mathrm{Var}_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu}) \leq \frac{C(M_f)\,\|\psi * (\varphi_\lambda \otimes \phi_\mu)\|_{\mathbb{L}_2}^2}{n} + \frac{C\left(M_f,p,q\right)}{\lambda_1 \ldots \lambda_p \mu_1 \ldots \mu_q n^2}.$$

5.7.7 Proof of Lemma 5.3

The objective here is the provide an upper bound of the bias term $\|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_{\mathbb{L}_2}^2$ w.r.t λ and μ, when $\psi \in \mathcal{S}_{p+q}^{\delta}(R)$, where $\delta \in (0,2]$. We first set $b = \psi * (\varphi_\lambda \otimes \phi_\mu) - \psi$, using that $b \in \mathbb{L}^1(\mathbb{R}^{p+q}) \cap \mathbb{L}^2(\mathbb{R}^{p+q})$, Plancherel's theorem gives that

$$\begin{aligned}
(2\pi)^{p+q}\,\|b\|_{\mathbb{L}_2}^2 &= \|\hat{b}\|_{\mathbb{L}_2}^2 \\
&= \|(1 - \widehat{\varphi_\lambda \otimes \phi_\mu})\hat{\psi}\|_{\mathbb{L}_2}^2 . \qquad (5.76)
\end{aligned}$$

Let us denote g_1 as in Equation (5.1), the real function defined for all $z \in \mathbb{R}$ as $g_1(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$. We then obviously have the following equation

$$\varphi_\lambda \otimes \phi_\mu(x,y) = \prod_{i=1}^{p} \left[\frac{1}{\lambda_i} g_1\left(\frac{x_i}{\lambda_i}\right)\right] \prod_{j=1}^{q} \left[\frac{1}{\mu_j} g_1\left(\frac{y_j}{\mu_j}\right)\right].$$

Moreover, it is known that $\hat{g}_1 = \sqrt{2\pi} g_1$, and that the Fourier transform of a tensor product of functions is the product of Fourier transform of each of these functions. We also recall that if G is a real function and $a > 0$ then, the Fourier transform of $z \mapsto 1/a \cdot G\left(z/a\right)$ is $u \mapsto \widehat{G}\left(au\right)$. We then obtain

$$\begin{aligned}
\widehat{\varphi_\lambda \otimes \phi_\mu}(\xi,\zeta) &= (2\pi)^{\frac{p+q}{2}} \prod_{i=1}^{p} g_1(\lambda_i \xi_i) \prod_{j=1}^{q} g_1(\mu_j \zeta_j) \\
&= \exp\left(-(\lambda_1^2\xi_1^2 + \ldots + \lambda_p^2\xi_p^2 + \mu_1^2\zeta_1^2 + \ldots + \mu_q^2\zeta_q^2)/2\right).
\end{aligned}$$

Thereafter, using Equation (5.76), the bias term $\|b\|_{\mathbb{L}_2}^2$ can then expressed as follows

$$\begin{aligned}\|b\|_{\mathbb{L}_2}^2 =& \frac{1}{(2\pi)^{p+q}} \int \left(1 - \exp\left(-(\lambda_1^2\xi_1^2 + \ldots + \lambda_p^2\xi_p^2 + \mu_1^2\zeta_1^2 + \ldots + \mu_q^2\zeta_q^2)/2\right)\right)^2 |\hat{\psi}(\xi,\zeta)|^2 \, d\xi d\zeta \\ =& \frac{1}{(2\pi)^{p+q}} \int \left[\frac{\left(1 - \exp\left(-(\lambda_1^2\xi_1^2 + \ldots + \lambda_p^2\xi_p^2 + \mu_1^2\zeta_1^2 + \ldots + \mu_q^2\zeta_q^2)/2\right)\right)}{\|(\xi,\zeta)\|^\delta}\right]^2 \|(\xi,\zeta)\|^{2\delta} |\hat{\psi}(\xi,\zeta)|^2 \, d\xi d\zeta.\end{aligned} \tag{5.77}$$

In order to upper bound the last integral, one can first notice that for all λ, ξ in $(0,+\infty)^p$ and μ, ζ in $(0,+\infty)^p$, we have: $\lambda_1^2\xi_1^2 + \ldots + \lambda_p^2\xi_p^2 + \mu_1^2\zeta_1^2 + \ldots + \mu_q^2\zeta_q^2 \leq \|(\lambda,\mu)\|^2\|(\xi,\zeta)\|^2$. We then obtain for all (ξ,ζ) in $\mathbb{R}^{p+q} \setminus \{0\}$,

$$\begin{aligned}\frac{1 - \exp\left(-(\lambda_1^2\xi_1^2 + \ldots + \lambda_p^2\xi_p^2 + \mu_1^2\zeta_1^2 + \ldots + \mu_q^2\zeta_q^2)/2\right)}{\|(\xi,\zeta)\|^\delta} &\leq \frac{1 - \exp\left(-\|(\lambda,\mu)\|^2\|(\xi,\zeta)\|^2/2\right)}{\|(\xi,\zeta)\|^\delta} \\ &\leq \|(\lambda,\mu)\|^\delta \sup_{H>0} \frac{1 - \exp(-H/2)}{H^{\delta/2}}.\end{aligned}$$

For δ in $(0,2]$, the function $H \mapsto \dfrac{1 - \exp(-H/2)}{H^{\delta/2}}$ is bounded in $(0,+\infty)$. Indeed, it is continuous on $(0,+\infty)$, tends to 0 in $+\infty$ and has a finite limit at 0 (1/2 if $\delta = 2$ and 0 otherwise). Hence, we thus obtain for all (ξ,ζ) in $\mathbb{R}^{p+q} \setminus \{0\}$,

$$\left[\frac{\left(1 - \exp\left(-(\lambda_1^2\xi_1^2 + \ldots + \lambda_p^2\xi_p^2 + \mu_1^2\zeta_1^2 + \ldots + \mu_q^2\zeta_q^2)/2\right)\right)}{\|(\xi,\zeta)\|^\delta}\right]^2 \leq C(\delta)\|(\lambda,\mu)\|^{2\delta}.$$

Thereafter, using Hölder's inequality if $\delta \geq 1$ and the concavity of $t \mapsto t^\delta$ on $\mathbb{R}_+$ if $\delta < 1$, it is straightforward to see that

$$\|(\lambda,\mu)\|^{2\delta} \leq C(p,q,\delta) \left[\sum_{i=1}^p \lambda_i^{2\delta} + \sum_{j=1}^q \mu_j^{2\delta}\right]. \tag{5.78}$$

Hence, combining the two last inequalities gives

$$\|b\|_{\mathbb{L}_2}^2 \leq C(p,q,\delta) \left[\sum_{i=1}^p \lambda_i^{2\delta} + \sum_{j=1}^q \mu_j^{2\delta}\right] \int \|(\xi,\zeta)\|^{2\delta} |\hat{\psi}(\xi,\zeta)|^2 \, d\xi d\zeta.$$

Recalling that ψ belongs to the Sobolev ball $\mathcal{S}_{p+q}^\delta(R)$, we obtain

$$\|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_{\mathbb{L}_2}^2 \leq C(p,q,\delta,R) \left[\sum_{i=1}^p \lambda_i^{2\delta} + \sum_{j=1}^q \mu_j^{2\delta}\right].$$

5.7.8 Proof of Theorem 5.2

We easily deduce from Theorem 5.1 and Lemma 5.3 that if ψ belongs to the Sobolev balls $\mathcal{S}^{\delta}_{p+q}(R)$ with δ in $(0,2]$, $P_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu} \leq q^{\lambda,\mu}_{1-\alpha}) \leq \beta$ as soon as

$$\|\psi\|^2_{\mathbb{L}_2} > C(p,q,\delta,R)\left[\sum_{i=1}^{p}\lambda_i^{2\delta}+\sum_{j=1}^{q}\mu_j^{2\delta}\right]+\frac{C\left(M_f,p,q,\beta\right)}{n\sqrt{\lambda_1\ldots\lambda_p\mu_1\ldots\mu_q}}\log\left(\frac{1}{\alpha}\right).$$

It now follows from the definition (3.16) of the uniform separation rate that

$$\left[\rho\left(\Delta^{\lambda,\mu}_{\alpha},\mathcal{S}^{\delta}_{p+q}(R),\beta\right)\right]^2 \leq C(p,q,\delta,R)\left[\sum_{i=1}^{p}\lambda_i^{2\delta}+\sum_{j=1}^{q}\mu_j^{2\delta}\right]+\frac{C\left(M_f,p,q,\beta\right)}{n\sqrt{\lambda_1\ldots\lambda_p\mu_1\ldots\mu_q}}\log\left(\frac{1}{\alpha}\right).$$

5.7.9 Proof of Corollary 5.2

The objective here is to give the uniform separation rate having the smallest upper bound w.r.t. the sample-size n, when ψ belongs to a Sobolev ball $\mathcal{S}^{\delta}_{p+q}(R)$ with δ in $(0,2]$. For this, we recall that according to Theorem 5.2, we have

$$\left[\rho\left(\Delta^{\lambda,\mu}_{\alpha},\mathcal{S}^{\delta}_{p+q}(R),\beta\right)\right]^2 \leq C(p,q,\delta,R)\left[\sum_{i=1}^{p}\lambda_i^{2\delta}+\sum_{j=1}^{q}\mu_j^{2\delta}\right]+\frac{C\left(M_f,p,q,\beta\right)}{n\sqrt{\lambda_1\ldots\lambda_p\mu_1\ldots\mu_q}}\log\left(\frac{1}{\alpha}\right).$$

In order to have the smallest behavior of the right side of the last inequality w.r.t. n, one has then to choose bandwidths $\lambda^* = (\lambda_1^*,\ldots,\lambda_p^*)$ and $\mu^* = (\mu_1^*,\ldots,\mu_q^*)$ w.r.t. n in such a way that

$$\sum_{i=1}^{p}\lambda_i^{*2\delta}+\sum_{j=1}^{q}\mu_j^{*2\delta} \quad \text{and} \quad \frac{1}{n\sqrt{\lambda_1^*\ldots\lambda_p^*\mu_1^*\ldots\mu_q^*}}$$

have the same behavior in n. Thereafter, it is clear that all λ_i^*'s and μ_j^*'s have the same behavior w.r.t. n. It obviously follows that for all i in $\{1,\ldots,p\}$ and all j in $\{1,\ldots,q\}$, we have

$$\lambda_i^* = \mu_j^* = n^{-2/(4\delta+p+q)}.$$

Consequently, the separation rate $\rho\left(\Delta^{\lambda^*,\mu^*}_{\alpha},\mathcal{S}^{\delta}_{p+q}(R),\beta\right)$ can be upper bounded as

$$\rho\left(\Delta^{\lambda^*,\mu^*}_{\alpha},\mathcal{S}^{\delta}_{p+q}(R),\beta\right) \leq C\left(M_f,p,q,\alpha,\beta,\delta\right)n^{-2\delta/(4\delta+p+q)}.$$

5.7.10 Proof of Lemma 5.4

The objective here is to give an upper bound of the bias term $\|\psi-\psi*(\varphi_\lambda\otimes\phi_\mu)\|^2_{\mathbb{L}_2}$ w.r.t. λ and μ, when ψ belongs to a Nikol'skii-Besov ball $\mathcal{N}^{\delta}_{2,p+q}(R)$, with $\delta=(\nu_1,\ldots,\nu_p,\gamma_1,\ldots,\gamma_q)$ in $(0,2]^{p+q}$. We first set $b=\psi*(\varphi_\lambda\otimes\phi_\mu)-\psi$ and we write

$$\begin{aligned} b(x,y) =& \psi*(\varphi_\lambda\otimes\phi_\mu)(x,y)-\psi(x,y)\\ =& \int\psi(x',y')\varphi_\lambda(x-x')\phi_\mu(y-y')\mathrm{d}x'\mathrm{d}y'-\psi(x,y). \end{aligned}$$

Moreover, using Equations (5.2) and (5.3), the fonction b can be written in terms of the functions g_p and g_q defined in Equation (5.1) as

$$\begin{aligned} b(x,y) =& \frac{1}{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q} \int \psi(x',y') g_p\left(\frac{x_1 - x_1'}{\lambda_1}, \dots, \frac{x_p - x_p'}{\lambda_p}\right) g_q\left(\frac{y_1 - y_1'}{\mu_1}, \dots, \frac{y_q - y_q'}{\mu_p}\right) \mathrm{d}x'\mathrm{d}y' - \psi(x,y) \\ =& \int \psi(x_1 + \lambda_1 u_1, \dots, x_p + \lambda_p u_p, y_1 + \mu_1 v_1, \dots, y_q + \mu_q v_q) g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) \, \mathrm{d}u\mathrm{d}v - \psi(x,y). \end{aligned}$$

Thereafter, using that $\int_{\mathbb{R}^p} g_p = \int_{\mathbb{R}^q} g_q = 1$, the function b can be expressed as

$$b(x,y) = \int g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) \Big[\psi(x_1 + \lambda_1 u_1, \dots, x_p + \lambda_p u_p, y_1 + \mu_1 v_1, \dots, y_q + \mu_q v_q) - \psi(x,y) \Big] \, \mathrm{d}u\mathrm{d}v.$$

Let us from now define for all i in $\{1, \dots, p\}$ and j in $\{1, \dots, q\}$, the functions $b_{1,i}$ and $b_{2,j}$ by

$$\begin{aligned} b_{1,i}(x,y) &= \int g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) \omega_{1,i}(x, y, u_1, \dots, u_i) \, \mathrm{d}u\mathrm{d}v, \\ b_{2,j}(x,y) &= \int g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) \omega_{2,j}(x, y, u_1, \dots, u_p, v_1, \dots, v_j) \, \mathrm{d}u\mathrm{d}v, \end{aligned}$$

where the function $\omega_{1,i}$ is defined as

$$\omega_{1,i}(x, y, u_1, \dots, u_i) = \psi(x_1 + \lambda_1 u_1, \dots, x_i + \lambda_i u_i, x_{i+1}, \dots, x_p, y) - \psi(x_1 + \lambda_1 u_1, \dots, x_{i-1} + \lambda_{i-1} u_{i-1}, x_i, \dots, x_p, y),$$

while the function $\omega_{2,j}$ is defined as

$$\begin{aligned} \omega_{2,j}(x, y, u_1, \dots, u_p, v_1, \dots, v_j) = & \psi(x_1 + \lambda_1 u_1, \dots, x_p + \lambda_p u_p, y_1 + \mu_1 v_1, \dots, y_j + \mu_j v_j, y_{j+1}, \dots, y_q) \\ & - \psi(x_1 + \lambda_1 u_1, \dots, x_p + \lambda_p u_p, y_1 + \mu_1 v_1, \dots, y_{j-1} + \mu_{j-1} v_{j-1}, y_j, \dots, y_q). \end{aligned}$$

It is then easy to see that the function b is the sum of all the functions $b_{1,i}$ and $b_{2,j}$

$$b(x,y) = \sum_{i=1}^{p} b_{1,i}(x,y) + \sum_{j=1}^{q} b_{2,j}(x,y).$$

One can then deduce that it would be sufficient for the control of the $\mathbb{L}_2$-norm of b, to control the $\mathbb{L}_2$-normes of all the functions $b_{1,i}$ and $b_{2,j}$. Using the triangular inequality, we have

$$\|b\|_{\mathbb{L}_2} \leq \sum_{i=1}^{p} \|b_{1,i}\|_{\mathbb{L}_2} + \sum_{j=1}^{q} \|b_{2,j}\|_{\mathbb{L}_2}. \tag{5.79}$$

By now, let us upper bound $\|b_{1,i}\|_{\mathbb{L}_2}^2$ and $\|b_{2,j}\|_{\mathbb{L}_2}^2$ for all i in $\{1, \dots, p\}$ and j in $\{1, \dots, q\}$. We distinguish two cases

Case 1. $0 < \nu_i \leq 1$

We first recall that $\|b_{1,i}\|_{\mathbb{L}_2}^2$ can be written as

$$\|b_{1,i}\|_{\mathbb{L}_2}^2 = \int \left[\int g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) \omega_{1,i}(x, y, u_1, \dots, u_i) \, \mathrm{d}u\mathrm{d}v \right]^2 \mathrm{d}x\mathrm{d}y.$$

We use the following lemma from page 13 of Tsybakov (2009).

Lemma 5.10. *Let $\rho : \mathbb{R}^d \times \mathbb{R}^{d'} \to \mathbb{R}$ be a Borel function, then we have the following inequality:*

$$\int \left(\int \rho(\theta, z) d\theta \right)^2 dz \leq \left[\int \left(\int \rho^2(\theta, z) dz \right)^{1/2} d\theta \right]^2.$$

By applying Lemma 5.10 to $((u, v), (x, y)) \mapsto g_p(u_1, \ldots, u_p) g_q(v_1, \ldots, v_q) \omega_{1,i}(x, y, u_1, \ldots, u_i)$, we obtain:

$$\begin{aligned} \|b_{1,i}\|_{\mathbb{L}_2}^2 &\leq \left[\int \left(\int g_p^2(u_1, \ldots, u_p) g_q^2(v_1, \ldots, v_q) \omega_{1,i}^2(x, y, u_1, \ldots, u_i) \, \mathrm{d}x\mathrm{d}y \right)^{1/2} \mathrm{d}u\mathrm{d}v \right]^2 \\ &= \left[\int g_p(u_1, \ldots, u_p) g_q(v_1, \ldots, v_q) \left(\int \omega_{1,i}^2(x, y, u_1, \ldots, u_i) \, \mathrm{d}x\mathrm{d}y \right)^{1/2} \mathrm{d}u\mathrm{d}v \right]^2. \end{aligned} \tag{5.80}$$

On the other hand, since ψ belongs to the Nikol'skii-Besov ball $\mathcal{N}_{2,p+q}^{\delta}(R)$, we have:

$$\left(\int \omega_{1,i}^2(x, y, u_1, \ldots, u_i) \, \mathrm{d}x\mathrm{d}y \right)^{1/2} \leq R \lambda_i^{\nu_i} |u_i|^{\nu_i}.$$

We then have by injecting this last inequation in Equation (5.80), that

$$\|b_{1,i}\|_{\mathbb{L}_2}^2 \leq C(R, \nu_i) \lambda_i^{2\nu_i}.$$

Case 2. $1 < \nu_i \leq 2$
In this case the function ψ has continuous first-order partial derivatives. Using Taylor expansion with integral form of the remainder w.r.t. the i^{th} variable of ψ, we have:

$$\omega_{1,i}(x, y, u_1, \ldots, u_i) = \lambda_i u_i \int_0^1 (1 - \tau) D_i^1 \psi(x_1 + \lambda_1 u_1, \ldots, x_i + \tau \lambda_i u_i, x_{i+1}, \ldots, y) d\tau.$$

where D_i^1 denotes the first-order partial derivative of ψ w.r.t. the i^{th} variable.

Thereafter, by injecting the last equation in the expression of $b_{1,i}$, we obtain:

$$b_{1,i}(x, y) = \int \lambda_i u_i g_p(u_1, \ldots, u_p) g_q(v_1, \ldots, v_q) \left[\int_0^1 (1-\tau) D_i^1 \psi(x_1 + \lambda_1 u_1, \ldots, x_i + \tau \lambda_i u_i, x_{i+1}, \ldots, y) d\tau \right] \mathrm{d}u\mathrm{d}v.$$

Furthermore, using the fact that g_p is the density function of the multivariate normal distribution with mean 0 and covariance matrix equals identity, we have that $\int u_i g_p(u_1, \ldots, u_p) \mathrm{d}u_i = 0$. The function $b_{1,i}$ can then be written as

$$b_{1,i}(x, y) = \int \lambda_i u_i g_p(u_1, \ldots, u_p) g_q(v_1, \ldots, v_q) \left[\int_0^1 (1 - \tau) D_i^1 \omega_{1,i}(x, y, u_1, \ldots, \tau u_i) \, d\tau \right] \mathrm{d}u\mathrm{d}v.$$

We have then the following equation for the $\mathbb{L}_2$-norm of $b_{1,i}$:

$$\|b_{1,i}\|_{\mathbb{L}_2}^2 = \int \left[\int \lambda_i u_i g_p(u_1, \ldots, u_p) g_q(v_1, \ldots, v_q) \left(\int_0^1 (1-\tau) D_i^1 \omega_{1,i}(x, y, u_1, \ldots, \tau u_i) \, d\tau \right) \mathrm{d}u\mathrm{d}v \right]^2 \mathrm{d}x\mathrm{d}y.$$

By now, we use as in Case 1 of Lemma 5.10 in order to upper bound $\|b_{1,i}\|^2_{\mathbb{L}_2}$. We then obtain:

$$\|b_{1,i}\|^2_{\mathbb{L}_2} \leq \left(\int \left[\int \left(\lambda_i u_i g_p(u_1,\dots,u_p) g_q(v_1,\dots,v_q) \int_0^1 (1-\tau) D_i^1 \omega_{1,i}(x,y,u_1,\dots,\tau u_i)\, d\tau\right)^2 \mathrm{d}x\mathrm{d}y\right]^{1/2} \mathrm{d}u\mathrm{d}v\right)^2$$
$$= \left(\int \lambda_i u_i g_p(u_1,\dots,u_p) g_q(v_1,\dots,v_q) \left[\int \left(\int_0^1 (1-\tau) D_i^1 \omega_{1,i}(x,y,u_1,\dots,\tau u_i)\, d\tau\right)^2 \mathrm{d}x\mathrm{d}y\right]^{1/2} \mathrm{d}u\mathrm{d}v\right)^2.$$

We apply a second time Lemma 5.10. For this, consider the function $\rho((x,y),\tau) = (1-\tau) D_i^1 \omega_{1,i}(x,y,u_1,\dots,\tau u_i)$, we then have:

$$\|b_{1,i}\|^2_{\mathbb{L}_2} \leq \left(\int \lambda_i u_i g_p(u_1,\dots,u_p) g_q(v_1,\dots,v_q) \left[\int_0^1 (1-\tau) \left(\int \left(D_i^1 \omega_{1,i}(x,y,u_1,\dots,\tau u_i)\right)^2 \mathrm{d}x\mathrm{d}y\right)^{1/2} d\tau\right] \mathrm{d}u\mathrm{d}v\right)^2. \tag{5.81}$$

On the other hand, using that ψ belongs to the Nikol'skii-Besov ball $\mathcal{N}^{\delta}_{2,p+q}(R)$:

$$\left(\int \left(D_i^1 \omega_{1,i}(x,y,u_1,\dots,\tau u_i)\right)^2 \mathrm{d}x\mathrm{d}y\right)^{1/2} \leq R \lambda_i^{\nu_i - 1} |\tau u_i|^{\nu_i - 1}.$$

We then obtain by injecting this last inequation in Equation (5.81), that

$$\|b_{1,i}\|^2_{\mathbb{L}_2} \leq C(R,\nu_i) \lambda_i^{2\nu_i}.$$

Besides, for all j in $\{1,\dots,q\}$, by similar arguments:

$$\|b_{2,j}\|^2_{\mathbb{L}_2} \leq C(R,\gamma_j) \mu_j^{2\gamma_j}.$$

Consequently, according to Equation (5.79), we have the following upper bound of $\|b\|^2_{\mathbb{L}_2}$

$$\|b\|^2_{\mathbb{L}_2} \leq C(R,\delta) \left[\sum_{i=1}^{p} \lambda_i^{2\nu_i} + \sum_{j=1}^{q} \mu_j^{2\gamma_j}\right].$$

5.7.11 Proof of Theorem 5.3

The proof of this theorem is similar to that of Theorem 5.2. Indeed, assuming the conditions of Theorem 5.1, we have according to this theorem and Lemma 5.4 that if ψ belongs to $\mathcal{N}^{\delta}_{2,p+q}(R)$, with $\delta = (\nu_1,\dots,\nu_p,\gamma_1,\dots,\gamma_q)$ in $(0,2]^{p+q}$, then $P_f(\widehat{\mathrm{HSIC}}_{\lambda,\mu} \leq q^{\lambda,\mu}_{1-\alpha}) \leq \beta$ as soon as

$$\|\psi\|^2_{\mathbb{L}_2} > C(R,\delta) \left[\sum_{i=1}^{p} \lambda_i^{2\nu_i} + \sum_{j=1}^{q} \mu_j^{2\gamma_j}\right] + \frac{C(M_f,p,q,\beta)}{n\sqrt{\lambda_1\dots\lambda_p\mu_1\dots\mu_q}} \log\left(\frac{1}{\alpha}\right).$$

One can then conclude from the definition (3.16) of the uniform separation rate that

$$\left[\rho\left(\Delta^{\lambda,\mu}_{\alpha}, \mathcal{N}^{\delta}_{2,p+q}(R), \beta\right)\right]^2 \leq C(R,\delta) \left[\sum_{i=1}^{p} \lambda_i^{2\nu_i} + \sum_{j=1}^{q} \mu_j^{2\gamma_j}\right] + \frac{C(M_f,p,q,\beta)}{n\sqrt{\lambda_1\dots\lambda_p\mu_1\dots\mu_q}} \log\left(\frac{1}{\alpha}\right).$$

5.7.12 Proof of Corollary 5.3

We aim here to give the uniform separation rate having the smallest upper bound w.r.t. the sample-size n, when ψ belongs to a Nikol'skii-Besov ball $\mathcal{N}^{\delta}_{2,p+q}(R)$, with $\delta = (\nu_1, \ldots, \nu_p, \gamma_1, \ldots, \gamma_q)$ in $(0,2]^{p+q}$. We first recall that Theorem 5.3 shows that:

$$\left[\rho\left(\Delta_{\alpha}^{\lambda,\mu}, \mathcal{N}^{\delta}_{2,p+q}(R), \beta\right)\right]^2 \leq C(R,\delta)\left[\sum_{i=1}^{p}\lambda_i^{2\nu_i} + \sum_{j=1}^{q}\mu_j^{2\gamma_j}\right] + \frac{C\left(M_f, p, q, \beta\right)}{n\sqrt{\lambda_1 \ldots \lambda_p \mu_1 \ldots \mu_q}}\log\left(\frac{1}{\alpha}\right).$$

So as to minimize the right side of the last inequality w.r.t. n, we have to choose bandwidths $\lambda^* = (\lambda_1^*, \ldots, \lambda_p^*)$ and $\mu^* = (\mu_1^*, \ldots, \mu_q^*)$ w.r.t. n such as

$$\left[\sum_{i=1}^{p}\lambda_i^{*2\nu_i} + \sum_{j=1}^{q}\mu_j^{*2\gamma_j}\right] \quad \text{and} \quad \frac{1}{n\sqrt{\lambda_1^* \ldots \lambda_p^* \mu_1^* \ldots \mu_q^*}}$$

have the same behavior in n. Let us set for all i in $\{1, \ldots, p\}$ and all j in $\{1, \ldots, q\}$, $\lambda_i^* = n^{a_i}$ and $\mu_j^* = n^{b_j}$. It is than clear that for all i and all j:

$$2a_i\nu_i = 2b_j\gamma_j = -\frac{1}{2}\left[\sum_{r=1}^{p}a_r + \sum_{s=1}^{q}b_s\right] - 1. \tag{5.82}$$

One can first express all a_i's and all b_j's w.r.t a_1 as

$$a_i = a_1\frac{\nu_1}{\nu_i} \quad \text{and} \quad b_j = a_1\frac{\nu_1}{\gamma_j}.$$

Thereafter, using Equation (5.82) we have the following:

$$2a_1\nu_1 = \frac{-a_1\nu_1}{2\eta} - 1.$$

We then first write that $a_1 = \dfrac{-2\eta}{\nu_1(4\eta+1)}$. We next obtain for all i and for all j that:

$$a_i = \frac{-2\eta}{\nu_i(4\eta+1)} \quad \text{and} \quad b_j = \frac{-2\eta}{\gamma_j(4\eta+1)}.$$

Consequently, the separation rate $\rho\left(\Delta_{\alpha}^{\lambda^*,\mu^*}, \mathcal{N}^{\delta}_{2,p+q}(R), \beta\right)$ can be upper bound as

$$\rho\left(\Delta_{\alpha}^{\lambda^*,\mu^*}, \mathcal{N}^{\delta}_{2,p+q}(R), \beta\right) \leq C\left(M_f, p, q, \alpha, \beta, \delta\right) n^{-\frac{2\eta}{(1+4\eta)}}.$$

5.7.13 Proof of Lemma 5.5

Let α be in $(0,1)$, we first prove that $u_\alpha \geq \alpha$. For this, we apply Bonferroni's Inequality:

$$\begin{aligned}
&P_{f_1\otimes f_2}\left(\sup_{(\lambda,\mu)\in\Lambda\times U}\left(\widehat{\mathrm{HSIC}}_{\lambda,\mu} - q^{\lambda,\mu}_{1-\alpha e^{-\omega_{\lambda,\mu}}}\right) > 0\right)\\
=&P_{f_1\otimes f_2}\left(\bigcup_{(\lambda,\mu)\in\Lambda\times U}\left\{\widehat{\mathrm{HSIC}}_{\lambda,\mu} > q^{\lambda,\mu}_{1-\alpha e^{-\omega_{\lambda,\mu}}}\right\}\right)\\
\leq&\sum_{(\lambda,\mu)\in\Lambda\times U} P_{f_1\otimes f_2}\left(\widehat{\mathrm{HSIC}}_{\lambda,\mu} > q^{\lambda,\mu}_{1-\alpha e^{-\omega_{\lambda,\mu}}}\right)\\
\leq&\sum_{(\lambda,\mu)\in\Lambda\times U} \alpha e^{-\omega_{\lambda,\mu}}\\
\leq&\ \alpha.
\end{aligned}$$

Then, by definition of u_α we have: $u_\alpha \geq \alpha$. Thereafter, we obtain:

$$\begin{aligned}
P_f\left(\Delta_\alpha = 0\right) = P_f\left(\bigcap_{(\lambda,\mu)\in\Lambda\times U}\left\{\widehat{\mathrm{HSIC}}_{\lambda,\mu} \leq q^{\lambda,\mu}_{1-u_\alpha e^{-\omega_{\lambda,\mu}}}\right\}\right)&\\
\leq \inf_{(\lambda,\mu)\in\Lambda\times U} P_f\left(\widehat{\mathrm{HSIC}}_{\lambda,\mu} \leq q^{\lambda,\mu}_{1-u_\alpha e^{-\omega_{\lambda,\mu}}}\right)&\\
\leq \inf_{(\lambda,\mu)\in\Lambda\times U} P_f\left(\widehat{\mathrm{HSIC}}_{\lambda,\mu} \leq q^{\lambda,\mu}_{1-\alpha e^{-\omega_{\lambda,\mu}}}\right)&\\
= \inf_{(\lambda,\mu)\in\Lambda\times U}\left\{P_f\left(\Delta^{\lambda,\mu}_{\alpha e^{-\omega_{\lambda,\mu}}} = 0\right)\right\},&
\end{aligned}$$

which concludes the proof.

5.7.14 Proof of Theorem 5.4

Let α and β be in $(0,1)$. According to Lemma 5.5, $P_f\left(\Delta_\alpha = 0\right) \leq \beta$ as soon as there exists (λ,μ) in $\Lambda\times U$ such that

$$P_f\left(\Delta^{\lambda,\mu}_{\alpha e^{-\omega_{\lambda,\mu}}} = 0\right) \leq \beta.$$

Then, according to Theorem 5.2 (resp. Theorem 5.3) if ψ belongs to $\mathcal{N}^\delta_{2,p+q}(R)$ (resp. ψ belongs to $\mathcal{S}^\delta_{p+q}(R)$): we take the infimum of the upper bounds for the uniform separation rates of the single tests over $\Lambda\times U$ while replacing $\log\left(\frac{1}{\alpha}\right)$ by $\log\left(\frac{1}{\alpha}\right) + \omega_{\lambda,\mu}$.

5.7.15 Proof of Corollary 5.4

Let us start with the case where ψ belongs to $\mathcal{N}^\delta_{2,p+q}(R)$. In this case, using Theorem 5.4, we have the following inequality for $\rho\left(\Delta_\alpha, \mathcal{N}^\delta_{2,p+q}(R), \beta\right)$,

$$\begin{aligned}
\left[\rho\left(\Delta_\alpha, \mathcal{N}^\delta_{2,p+q}(R), \beta\right)\right]^2 \leq C\left(M_f, p, q, \beta, \delta\right) \inf_{(\lambda,\mu)\in\Lambda\times U}&\left\{\frac{1}{\sqrt{\lambda_1\ldots\lambda_p\mu_1\ldots\mu_q n}}\left(\log\left(\frac{1}{\alpha}\right) + \omega_{\lambda,\mu}\right)\right.\\
&\left. + \left[\sum_{i=1}^p \lambda_i^{2\nu_i} + \sum_{j=1}^q \mu_j^{2\gamma_j}\right]\right\}.
\end{aligned}$$

Let us take $\lambda^* = (2^{-m^*_{1,1}}, \dots, 2^{-m^*_{1,p}})$ and $\mu^* = (2^{-m^*_{2,1}}, \dots, 2^{-m^*_{2,q}})$, where the integers $m^*_{1,1}, \dots, m^*_{1,p}$, $m^*_{2,1}, \dots, m^*_{2,q}$ are defined as follows

$$m^*_{1,i} = \left\lfloor \log_2\left(\left(\frac{n}{\log\log(n)}\right)^{\frac{2\eta}{\nu_i(1+4\eta)}}\right)\right\rfloor \quad \text{and} \quad m^*_{2,j} = \left\lfloor \log_2\left(\left(\frac{n}{\log\log(n)}\right)^{\frac{2\eta}{\gamma_j(1+4\eta)}}\right)\right\rfloor.$$

where $\dfrac{1}{\eta} = \displaystyle\sum_{i=1}^{p} \frac{1}{\nu_i} + \sum_{j=1}^{q} \frac{1}{\gamma_j}$.

Then, we obviously have

$$\left[\rho\left(\Delta_\alpha, \mathcal{N}^\delta_{2,p+q}(R), \beta\right)\right]^2 \leq C(M_f, p, q, \beta, \delta)\left[\frac{1}{\sqrt{\lambda^*_1 \dots \lambda^*_p \mu^*_1 \dots \mu^*_q}\, n}\left(\log\left(\frac{1}{\alpha}\right) + \omega_{\lambda^*,\mu^*}\right) + \sum_{i=1}^{p} (\lambda^*_i)^{2\nu_i} + \sum_{j=1}^{q} (\mu^*_j)^{2\gamma_j}\right].$$

Besides, using the inequalities

$$m^*_{1,i} \leq \log_2\left(\left(\frac{n}{\log\log(n)}\right)^{\frac{2\eta}{\nu_i(1+4\eta)}}\right) \quad \text{and} \quad m^*_{2,j} \leq \log_2\left(\left(\frac{n}{\log\log(n)}\right)^{\frac{2\eta}{\gamma_j(1+4\eta)}}\right),$$

we upper bound $(\lambda^*_i)^{-1/2}$ and $(\mu^*_j)^{-1/2}$ by

$$(\lambda^*_i)^{-1/2} = 2^{m^*_{1,i}/2} \leq \left(\frac{n}{\log\log(n)}\right)^{\frac{\eta}{\nu_i(1+4\eta)}} \quad \text{and} \quad \mu^*_j = 2^{m^*_{2,j}/2} \leq \left(\frac{n}{\log\log(n)}\right)^{\frac{\eta}{\gamma_j(1+4\eta)}}.$$

Therefore, we obtain

$$(\lambda^*_1 \dots \lambda^*_p \mu^*_1 \dots \mu^*_q)^{-1/2} \leq \left(\frac{n}{\log\log(n)}\right)^{\frac{1}{(1+4\eta)}}. \tag{5.83}$$

Let us now upper bound ω_{λ^*,μ^*}, we first write

$$\begin{aligned}\omega_{\lambda^*,\mu^*} &= 2\sum_{i=1}^{p} \log(m^*_{1,i} \times \pi/\sqrt{6}) + 2\sum_{j=1}^{q} \log(m^*_{2,j} \times \pi/\sqrt{6}) \\ &= 2\log\left(m^*_{1,1} \dots m^*_{1,p} m^*_{2,1} \dots m^*_{2,q}\right) + 2(p+q)\log(\pi/\sqrt{6}).\end{aligned}$$

Moreover, it is easy to see that

$$m^*_{1,i} \leq \frac{2\eta}{\nu_i(1+4\eta)}\log(n) \quad \text{and} \quad \mu^*_j \leq \frac{2\eta}{\gamma_j(1+4\eta)}\log(n).$$

Then,

$$\log(m^*_{1,1} \dots m^*_{1,p} m^*_{2,1} \dots m^*_{2,q}) \leq C(\delta)\log\log(n).$$

Thereafter, ω_{λ^*,μ^*} can be upper bound as

$$\omega_{\lambda^*,\mu^*} \leq C(\delta)\log\log(n). \tag{5.84}$$

From Equations (5.83) and (5.84), we have

$$\frac{1}{n\sqrt{\lambda_1^* \dots \lambda_p^* \mu_1^* \dots \mu_q^*}} \left(\log\left(\frac{1}{\alpha}\right) + \omega_{\lambda^*,\mu^*} \right) \leq C(\alpha,\delta) \left(\frac{\log\log(n)}{n} \right)^{\frac{4\eta}{(1+4\eta)}}. \tag{5.85}$$

We aim now to upper bound $\sum_{i=1}^{p} (\lambda_i^*)^{2\nu_i} + \sum_{j=1}^{q} (\mu_j^*)^{2\gamma_j}$. For this, we first write

$$m_{1,i}^* \geq \log_2\left(\left(\frac{n}{\log\log(n)} \right)^{\frac{2\eta}{\nu_i(1+4\eta)}} \right) - 1 \quad \text{and} \quad m_{2,j}^* \geq \log_2\left(\left(\frac{n}{\log\log(n)} \right)^{\frac{2\eta}{\gamma_j(1+4\eta)}} \right) - 1.$$

We then have the following inequalities for $(\lambda_i^*)^{2\nu_i}$ and $(\mu_j^*)^{2\gamma_j}$,

$$(\lambda_i^*)^{2\nu_i} \leq 2^{2\nu_i} \left(\frac{\log\log(n)}{n} \right)^{\frac{4\eta}{(1+4\eta)}} \quad \text{and} \quad (\mu_j^*)^{2\gamma_j} \leq 2^{2\gamma_j} \left(\frac{\log\log(n)}{n} \right)^{\frac{4\eta}{(1+4\eta)}}.$$

Therefore, we obtain

$$\sum_{i=1}^{p} (\lambda_i^*)^{2\nu_i} + \sum_{j=1}^{q} (\mu_j^*)^{2\gamma_j} \leq C(\delta) \left(\frac{\log\log(n)}{n} \right)^{\frac{4\eta}{(1+4\eta)}}. \tag{5.86}$$

Consequently, from Equations (5.85) and (5.86),

$$\rho\left(\Delta_\alpha, \mathcal{N}_{2,p+q}^{\delta}(R), \beta\right) \leq C\left(M_f, p, q, \alpha, \beta, \delta\right) \left(\frac{\log\log(n)}{n} \right)^{\frac{2\eta}{(1+4\eta)}}.$$

In the case where ψ belongs to $\mathcal{S}_{p+q}^{\delta}(R)$, the same arguments above is applied by taking $\nu_1 = \ldots = \nu_p = \gamma_1 = \ldots = \gamma_q = \delta$, lead to

$$\rho\left(\Delta_\alpha, \mathcal{S}_{p+q}^{\delta}(R), \beta\right) \leq C\left(M_f, p, q, \alpha, \beta, \delta\right) \left(\frac{\log\log(n)}{n} \right)^{\frac{2\eta}{(1+4\eta)}},$$

where $\frac{1}{\eta} = (p+q)\frac{1}{\delta}$.

5.7.16 Proof of Lemma 5.6

Assume there exists a distribution f_0 that satisfies $(\mathcal{H}_0)$ such that the probability measure $P_{\nu_{\rho_*}}$ is absolutely continuous w.r.t. P_{f_0} and verifies Equation (5.22).

Let us first lower bound $\beta\left[\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)\right]$ w.r.t. the distributions $P_{\nu_{\rho_*}}$ and P_{f_0},

$$\begin{aligned}
\beta\left[\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)\right] &\geq \inf_{\Delta_\alpha} P_{\nu_{\rho_*}}\left(\Delta_\alpha = 0\right) \\
&= 1 - \sup_{\Delta_\alpha} P_{\nu_{\rho_*}}\left(\Delta_\alpha = 1\right) \\
&\geq 1 - \alpha - \sup_{\Delta_\alpha} \left| P_{\nu_{\rho_*}}\left(\Delta_\alpha = 1\right) - P_{f_0}\left(\Delta_\alpha = 1\right) \right|.
\end{aligned}$$

We denote by $\left\| P_{\nu_{\rho_*}} - P_{f_0} \right\|_{TV}$ the total variation distance between the distributions $P_{\nu_{\rho_*}}$ and P_{f_0}. We recall that,

$$\left\| P_{\nu_{\rho_*}} - P_{f_0} \right\|_{TV} = \sup_{E \in \mathcal{E}} \left| P_{\nu_{\rho_*}}(E) - P_{f_0}(E) \right|,$$

where $\mathcal{E}$ is the space of measurable sets. We then obtain

$$\beta\left[\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)\right] \geq 1 - \alpha - \left\|P_{\nu_{\rho_*}} - P_{f_0}\right\|_{TV}.$$

Notice that,

$$\left\|P_{\nu_{\rho_*}} - P_{f_0}\right\|_{TV} = \sup_{E\in\mathcal{E}}\left[P_{\nu_{\rho_*}}(E) - P_{f_0}(E)\right] = \sup_{E\in\mathcal{E}}\left[P_{f_0}(E) - P_{\nu_{\rho_*}}(E)\right].$$

It is then straightforward to show that

$$\begin{aligned}\left\|P_{\nu_{\rho_*}} - P_{f_0}\right\|_{TV} &= \frac{1}{2}\int \left|L_{\nu_{\rho_*}} - 1\right| \mathrm{d}P_{f_0} \\ &= \frac{1}{2}\mathbb{E}_{P_{f_0}}\left[\left|L_{\nu_{\rho_*}}(\mathbb{Z}_n) - 1\right|\right] \\ &\leq \frac{1}{2}\left(\mathbb{E}_{P_{f_0}}\left[L^2_{\nu_{\rho_*}}(\mathbb{Z}_n)\right] - 1\right)^{1/2},\end{aligned}$$

where the last inequality holds by applying Cauchy-Schwarz and the fact that $\mathbb{E}_{P_{f_0}}\left[L_{\nu_{\rho_*}}(\mathbb{Z}_n)\right] = 1$. Thus,

$$\beta\left[\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)\right] \geq 1 - \alpha - \frac{1}{2}\left(\mathbb{E}_{P_{f_0}}\left[L^2_{\nu_{\rho_*}}(\mathbb{Z}_n)\right] - 1\right)^{1/2}.$$

If the condition (5.22) holds, we then obtain

$$\beta\left[\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)\right] > \beta.$$

Furthermore, using that $\mathcal{F}_{\rho_*}(\mathcal{C}_\delta) \subset \mathcal{F}_\rho(\mathcal{C}_\delta)$ for all $\rho \leq \rho_*$, we have

$$\beta\left[\mathcal{F}_{\rho}(\mathcal{C}_\delta)\right] > \beta.$$

Let us now prove that this implies the lower bound

$$\rho\left(\mathcal{C}_\delta, \alpha, \beta\right) = \inf_{\Delta_\alpha} \rho\left(\Delta_\alpha, \mathcal{C}_\delta, \beta\right) \geq \rho_*. \tag{5.87}$$

Assume $\beta\left[\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)\right] > \beta$, then

$$\forall \Delta_\alpha, \qquad \sup_{f\in\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)} P_f\left(\Delta_\alpha = 0\right) > \beta.$$

In particular, since the family $\left\{\mathcal{F}_\rho\left(\mathcal{C}_\delta\right)\right\}_{\rho>0}$ is non increasing for the inclusion,

$$\forall \Delta_\alpha, \quad \rho\left(\Delta_\alpha, \mathcal{C}_\delta, \beta\right) = \inf\left\{\rho > 0\ ;\ \sup_{f\in\mathcal{F}_\rho(\mathcal{C}_\delta)} P_f\left(\Delta_\alpha = 0\right) \leq \beta\right\} > \rho_*,$$

which directly implies (5.87).

5.7.17 Proof of Proposition 5.4

Proof of 1. Let us prove that the functions f_θ are probability density functions for n large enough.
First, it is obvious from Equation (5.24) that

$$\int_{\mathbb{R}^{p+q}} f_\theta(x, y)\,\mathrm{d}x\,\mathrm{d}y = 1,$$

since $f_1 \otimes f_2$ is a probability density function and that $\int_{\mathbb{R}} G(x)\,\mathrm{d}x = 0$. It remains to check that f_θ is a non-negative function for n large enough.

Let $j = (j_1, \ldots, j_p)$ in $\{1, \ldots, M_n\}^p$ and $l = (l_1, \ldots, l_1)$ in $\{1, \ldots, M_n\}^q$. Knowing that for all $1 \le r \le p$ and all $1 \le s \le q$, the supports of the functions $G_{h_n}(\cdot - j_r h_n)$ and $G_{h_n}(\cdot - l_s h_n)$ are respectively the intervals $\Big((j_r - 1)h_n, j_r h_n\Big]$ and $\Big((l_s - 1)h_n, l_s h_n\Big]$, the support of the function

$$g_{n,j,l} : (x, y) \mapsto \prod_{r=1}^{p} G_{h_n}(x_r - j_r h_n) \prod_{s=1}^{q} G_{h_n}(y_s - l_s h_n) \tag{5.88}$$

is the set

$$D_{(j,l)} = \prod_{r=1}^{p} \Big((j_r - 1)h_n, j_r h_n\Big] \times \prod_{s=1}^{q} \Big((l_s - 1)h_n, l_s h_n\Big]. \tag{5.89}$$

These supports are then disjoint for different multi-indexes (j, l) in $I_{n,p,q}$ and have as union set $(0, 1]^{p+q}$ (since $M_n h_n = 1$). In particular, for all (x, y) in $(0, 1]^{p+q}$,

$$\begin{aligned}\left| \sum_{(j,l)\in I_{n,p,q}} \theta_{(j,l)} \prod_{r=1}^{p} G_{h_n}(x_r - j_r h_n) \prod_{s=1}^{q} G_{h_n}(y_s - l_s h_n) \right| &\le \frac{1}{h_n^{p+q}} \left(\sup_{t\in[-1,0]} |G(t)| \right)^{p+q} \\ &= \frac{1}{(e h_n)^{p+q}}.\end{aligned}$$

Hence, if (x, y) belongs to $[0, 1]^{p+q}$, then

$$f_\theta(x, y) \ge 1 - \frac{h_n^\delta}{e^{p+q}},$$

which is non negative for h_n small enough. Otherwise, $f_\theta(x, y) = 0$. In particular, for all (x, y) in $\mathbb{R}^{p+q}$, $f_\theta(x, y) \ge 0$ which ends the proof of this first point.

Proof of 2. Let us prove that the functions $f_\theta - f_1 \otimes f_2$ belong to the Sobolev ball $\mathcal{S}^\delta_{p+q}(R)$ for n large enough. This point relies on Lemma (Butucea, 2007, Lemma 2) recalled bellow.

Lemma 5.11 (Butucea (2007))**.** *Let G be the function defined in Equation (5.23). Then G is an infinitely differentiable function such that $\int_{\mathbb{R}} G(x)\,\mathrm{d}x = 0$. Its Fourier transform verifies*

$$|\widehat{G}(u)| \le C \exp(-a\sqrt{|u|}) \quad \text{as} \quad |u| \to \infty,$$

for some positive constants C and a. Moreover, $\widehat{G}$ is an infinitely differentiable and bounded function.

According to the Fourier transform properties, we write

$$\widehat{f}_\theta(u, v) = \widehat{f}_1 \otimes \widehat{f}_2(u, v) + h_n^{\delta+(p+q)} \sum_{(j,l)\in I_{n,p,q}} \theta_{j,l} \prod_{r=1}^{p} \exp(i u_r j_r h_n) \widehat{G}(h_n u_r) \prod_{s=1}^{q} \exp(i v_s l_s h_n) \widehat{G}(h_n v_s).$$

Then,

$$\left| \widehat{f}_\theta(u, v) - \widehat{f}_1 \otimes \widehat{f}_2(u, v) \right|^2 = H_{1,n}(u, v) + H_{2,n}(u, v), \tag{5.90}$$

where the functions $H_{1,n}$ and $H_{2,n}$ are respectively defined by

$$H_{1,n}(u,v) = M_n^{p+q} h_n^{2\delta+2(p+q)} \prod_{r=1}^{p} |\widehat{G}(h_n u_r)|^2 \prod_{s=1}^{q} |\widehat{G}(h_n v_s)|^2, \tag{5.91}$$

$$H_{2,n}(u,v) = h_n^{2\delta+2(p+q)} \sum_{\substack{(j_1,l_1)\in I_{n,p,q} \\ (j_2,l_2)\in I_{n,p,q} \\ (j_1,l_1)\neq(j_2,l_2)}} \theta_{j_1,l_1}\theta_{j_2,l_2}\mathcal{G}_{j_1,l_1,j_2,l_2}(h_n u, h_n v), \tag{5.92}$$

where the vectors j_1, j_2, l_1 and l_2 in the last sum are defined as $j_1 = (j_{1,1},\ldots,j_{1,p})$, $j_2 = (j_{2,1},\ldots,j_{2,p})$, $l_1 = (l_{1,1},\ldots,l_{1,q})$ and $l_2 = (l_{2,1},\ldots,l_{2,q})$. In addition, the function $\mathcal{G}_{j_1,l_1,j_2,l_2}$ is defined as

$$\mathcal{G}_{j_1,l_1,j_2,l_2} : (u,v) \mapsto \prod_{r=1}^{p} \exp\left(iu_r(j_{1,r}-j_{2,r})\right)|\widehat{G}(u_r)|^2 \prod_{s=1}^{q} \exp\left(iv_s(l_{1,s}-l_{1,s})\right)|\widehat{G}(v_s)|^2. \tag{5.93}$$

By now, we aim to upper bound the integral of the function $(u,v) \mapsto \|(u,v)\|^{2\delta}\left|\widehat{f}_\theta(u,v) - \widehat{f}_1 \otimes \widehat{f}_2(u,v)\right|^2$. First, one easily show, as in Equation (5.78), that

$$\|(u,v)\|^{2\delta} \leq C(p,q,\delta)\left[\sum_{i=1}^{p}|u_i|^{2\delta} + \sum_{j=1}^{q}|v_j|^{2\delta}\right]. \tag{5.94}$$

We then obtain from Equations (5.91) and (5.94) the following result,

$$\begin{aligned}\int_{\mathbb{R}^{p+q}} \|(u,v)\|^{2\delta} H_{1,n}(u,v)\,\mathrm{d}u\,\mathrm{d}v &\leq C(p,q,\delta)M_n^{p+q}h_n^{2\delta+2(p+q)}\left(\int_{\mathbb{R}}|t|^{2\delta}|\widehat{G}(h_n t)|^2\,\mathrm{d}t\right)\left(\int_{\mathbb{R}}|\widehat{G}(h_n z)|^2\,\mathrm{d}z\right)^{p+q-1} \\ &= C(p,q,\delta)(M_n h_n)^{p+q}\left(\int_{\mathbb{R}}|t|^{2\delta}|\widehat{G}(t)|^2\,\mathrm{d}t\right)\left(\int_{\mathbb{R}}|\widehat{G}(z)|^2\,\mathrm{d}z\right)^{p+q-1}.\end{aligned}$$

The functions $t \mapsto |t|^{2\delta}|\widehat{G}(t)|^2$ and $z \mapsto |\widehat{G}(z)|^2$ being integrable according to Lemma 5.11, we have

$$\begin{aligned}\int_{\mathbb{R}^{p+q}} \|(u,v)\|^{2\delta} H_{1,n}(u,v)\,\mathrm{d}u\,\mathrm{d}v &\leq C(p,q,\delta)(M_n h_n)^{p+q} \\ &\leq C(p,q,\delta).\end{aligned} \tag{5.95}$$

To complete this point of the proof, we demonstrate a similar equation for $H_{2,n}$. Starting from the expression of $H_{2,n}$ in (5.92), we write

$$\int_{\mathbb{R}^{p+q}} \|(u,v)\|^{2\delta} H_{2,n}(u,v)\,\mathrm{d}u\,\mathrm{d}v = h_n^{p+q} \sum_{\substack{(j_1,l_1)\in I_{n,p,q} \\ (j_2,l_2)\in I_{n,p,q} \\ (j_1,l_1)\neq(j_2,l_2)}} \theta_{j_1,l_1}\theta_{j_2,l_2} \int_{\mathbb{R}^{p+q}} \|(u,v)\|^{2\delta}\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v.$$

Therefore, according to the triangular inequality, we obtain

$$\left|\int_{\mathbb{R}^{p+q}} \|(u,v)\|^{2\delta} H_{2,n}(u,v)\,\mathrm{d}u\,\mathrm{d}v\right| \leq h_n^{p+q} \sum_{\substack{(j_1,l_1)\in I_{n,p,q} \\ (j_2,l_2)\in I_{n,p,q} \\ (j_1,l_1)\neq(j_2,l_2)}} \left|\int_{\mathbb{R}^{p+q}} \|(u,v)\|^{2\delta}\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v\right|. \tag{5.96}$$

We first assume that all the components of the vectors (j_1, l_1) and (j_2, l_2) are different. Knowing that $\widehat{G}$ is differentiable, the function $|\widehat{G}|^2 = \widehat{G}\overline{\widehat{G}}$ is also differentiable. We apply an integration by parts to the function $\mathcal{G}_{j_1,l_1,j_2,l_2}$ w.r.t. the variable u_1,

$$\int_{\mathbb{R}^{p+q}} \|(u,v)\|^{2\delta}\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v = \frac{1}{i(j_{1,1}-j_{2,1})}\int_{\mathbb{R}^{p+q-1}}\left[\|(u,v)\|^{2\delta}\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)\right]_{u_1=-\infty}^{u_1=+\infty}\frac{\mathrm{d}u}{\mathrm{d}u_1}\,\mathrm{d}v - \frac{1}{i(j_{1,1}-j_{2,1})}\int_{\mathbb{R}^{p+q}}\frac{\partial}{\partial u_1}\left\{\|(u,v)\|^{2\delta}|\widehat{G}(u_1)|^2\right\}\frac{\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)}{|\widehat{G}(u_1)|^2}\,\mathrm{d}u\,\mathrm{d}v. \tag{5.97}$$

Moreover, it is straightforward to see that the first term on the right side of Equation (5.97) is equal to zero. Indeed, for all a in $\mathbb{R}$ and $(u_2,\ldots,u_p,v)$ in $\mathbb{R}^{p+q-1}$, we have the following

$$\left|\left[\|(u,v)\|^{2\delta}\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)\right]_{u_1=-a}^{u_1=+a}\right| \leq \|(a,u_2,\ldots,u_p,v)\|^{2\delta}\left(|\widehat{G}(a)|^2+|\widehat{G}(-a)|^2\right)\prod_{r=2}^{p}|\widehat{G}(u_r)|^2\prod_{s=1}^{q}|\widehat{G}(v_s)|^2.$$

In addition, according to Lemma 5.11 we obviously obtain for all $u_2,\ldots,u_p,v$ the following

$$\lim_{|a|\to+\infty}\|(a,u_2,\ldots,u_p,v)\|^{2\delta}|\widehat{G}(a)|^2 = \lim_{|a|\to+\infty}\|(a,u_2,\ldots,u_p,v)\|^{2\delta}|\widehat{G}(-a)|^2 = 0. \tag{5.98}$$

It follows from Equation (5.97) that

$$\int_{\mathbb{R}^{p+q}} \|(u,v)\|^{2\delta}\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v = \frac{i}{(j_{1,1}-j_{2,1})}\int_{\mathbb{R}^{p+q}}\frac{\partial}{\partial u_1}\left\{\|(u,v)\|^{2\delta}|\widehat{G}(u_1)|^2\right\}\frac{\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)}{|\widehat{G}(u_1)|^2}\,\mathrm{d}u\,\mathrm{d}v. \tag{5.99}$$

From Equation (5.99), we perform an integration by parts w.r.t. u_2,

$$\begin{aligned}(j_{1,1}-j_{2,1})(j_{1,2}-j_{2,2})&\int_{\mathbb{R}^{p+q}}\|(u,v)\|^{2\delta}\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v\\ &= \int_{\mathbb{R}^{p+q-1}}\left[\frac{\partial}{\partial u_1}\left\{\|(u,v)\|^{2\delta}|\widehat{G}(u_1)|^2\right\}\frac{\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)}{|\widehat{G}(u_1)|^2}\right]_{u_2=-\infty}^{u_2=+\infty}\frac{\mathrm{d}u}{\mathrm{d}u_2}\,\mathrm{d}v\\ &\quad-\int_{\mathbb{R}^{p+q}}\frac{\partial^2}{\partial u_1\partial u_2}\left\{\|(u,v)\|^{2\delta}|\widehat{G}(u_1)|^2|\widehat{G}(u_2)|^2\right\}\frac{\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)}{|\widehat{G}(u_1)|^2|\widehat{G}(u_2)|^2}\,\mathrm{d}u\,\mathrm{d}v.\end{aligned}$$

Moreover, by analogy with Equation (5.98), we can easily show using Lemma 5.11 that

$$\left[\frac{\partial}{\partial u_1}\left\{\|(u,v)\|^{2\delta}|\widehat{G}(u_1)|^2\right\}\frac{\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)}{|\widehat{G}(u_1)|^2}\right]_{u_2=-\infty}^{u_2=+\infty} = 0.$$

This leads to

$$\begin{aligned}(j_{1,1}-j_{2,1})(j_{1,2}-j_{2,2})&\int_{\mathbb{R}^{p+q}}\|(u,v)\|^{2\delta}\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v\\ &= -\int_{\mathbb{R}^{p+q}}\frac{\partial^2}{\partial u_1\partial u_2}\left\{\|(u,v)\|^{2\delta}|\widehat{G}(u_1)|^2|\widehat{G}(u_2)|^2\right\}\frac{\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)}{|\widehat{G}(u_1)|^2|\widehat{G}(u_2)|^2}\,\mathrm{d}u\,\mathrm{d}v.\end{aligned}$$

By repeating this last process for the variables $u_3, \dots, u_p, v_1, \dots, v_q$, we get

$$\prod_{r=1}^{p}(j_{1,r}-j_{2,r})\prod_{s=1}^{q}(l_{1,s}-l_{1,s})\int_{\mathbb{R}^{p+q}}\|(u,v)\|_2^{2\delta}\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v$$
$$=(i)^{p+q}\int_{\mathbb{R}^{p+q}}\frac{\partial^{p+q}}{\partial u_1\dots\partial v_q}\left\{\|(u,v)\|^{2\delta}\prod_{r=1}^{p}|\widehat{G}(u_r)|^2\prod_{s=1}^{q}|\widehat{G}(v_s)|^2\right\}\frac{\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)}{\prod_{r=1}^{p}|\widehat{G}(u_r)|^2\prod_{s=1}^{q}|\widehat{G}(v_s)|^2}\,\mathrm{d}u\,\mathrm{d}v. \tag{5.100}$$

Starting from Equation (5.100), we perform a second integration by part w.r.t. u_1. We write

$$\int_{\mathbb{R}^{p+q}}\frac{\partial^{p+q}W}{\partial u_1\dots\partial v_q}(u,v)R_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v$$
$$=\frac{1}{i(j_{1,1}-j_{2,1})}\int_{\mathbb{R}^{p+q-1}}\left[\frac{\partial^{p+q}W}{\partial u_1\dots\partial v_q}(u,v)R_{j_1,l_1,j_2,l_2}(u,v)\right]_{u_1=-\infty}^{u_1=+\infty}\frac{\mathrm{d}u}{\mathrm{d}u_1}\,\mathrm{d}v$$
$$-\frac{1}{i(j_{1,1}-j_{2,1})}\int_{\mathbb{R}^{p+q}}\frac{\partial^{p+q+1}W}{\partial^2 u_1\partial u_2\dots\partial v_q}(u,v)R_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v, \tag{5.101}$$

where the functions W and R_{j_1,l_1,j_2,l_2} are defined as

$$W:(u,v)\mapsto\|(u,v)\|^{2\delta}\prod_{r=1}^{p}|\widehat{G}(u_r)|^2\prod_{s=1}^{q}|\widehat{G}(v_s)|^2,$$
$$R_{j_1,l_1,j_2,l_2}:(u,v)\mapsto\frac{\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)}{\prod_{r=1}^{p}|\widehat{G}(u_r)|^2\prod_{s=1}^{q}|\widehat{G}(v_s)|^2}.$$

Let us now show that the first term in the right side of Equation (5.101) is equal to zero. Using the differentiability of $\widehat{G}$, we write

$$\frac{\partial^{p+q}}{\partial u_1\dots\partial v_q}\left\{\|(u,v)\|^{2\delta}\prod_{r=1}^{p}|\widehat{G}(u_r)|^2\prod_{s=1}^{q}|\widehat{G}(v_s)|^2\right\}$$
$$=|\widehat{G}(u_1)|^2\frac{\partial^{p+q}}{\partial u_1\dots\partial v_q}\left\{\|(u,v)\|^{2\delta}\prod_{r=2}^{p}|\widehat{G}(u_r)|^2\prod_{s=1}^{q}|\widehat{G}(v_s)|^2\right\}$$
$$+\frac{\partial}{\partial u_1}|\widehat{G}(u_1)|^2\frac{\partial^{p+q}}{\partial u_2\dots\partial v_q}\left\{\|(u,v)\|^{2\delta}\prod_{r=2}^{p}|\widehat{G}(u_r)|^2\prod_{s=1}^{q}|\widehat{G}(v_s)|^2\right\}. \tag{5.102}$$

Furthermore, according to Lemma 5.11 we obviously have

$$\lim_{|u_1|\to+\infty}|\widehat{G}(u_1)|^2\frac{\partial^{p+q}}{\partial u_1\dots\partial v_q}\left\{\|(u,v)\|^{2\delta}\prod_{r=2}^{p}|\widehat{G}(u_r)|^2\prod_{s=1}^{q}|\widehat{G}(v_s)|^2\right\}=0.$$

In addition,

$$\frac{\partial}{\partial u_1}|\widehat{G}(u_1)|^2\frac{\partial^{p+q}}{\partial u_2\dots\partial v_q}\left\{\|(u,v)\|^{2\delta}\prod_{r=2}^{p}|\widehat{G}(u_r)|^2\prod_{s=1}^{q}|\widehat{G}(v_s)|^2\right\}$$
$$=\left\{\widehat{G}(u_1)\frac{\partial\overline{\widehat{G}}(u_1)}{\partial u_1}+\overline{\widehat{G}}(u_1)\frac{\partial\widehat{G}(u_1)}{\partial u_1}\right\}\frac{\partial^{p+q}}{\partial u_2\dots\partial v_q}\left\{\|(u,v)\|^{2\delta}\prod_{r=2}^{p}|\widehat{G}(u_r)|^2\prod_{s=1}^{q}|\widehat{G}(v_s)|^2\right\}.$$

Moreover, $\dfrac{\partial \widehat{G}(u_1)}{\partial u_1}$ is the Fourier transform of $x \mapsto ixG(x)$ which is $\mathbb{L}_1$-integrable as it is continuous with a bounded support. Thus, we deduce according to Riemann–Lebesgue lemma Bochner and Chandrasekharan (1949) that

$$\lim_{|u_1|\to+\infty} \frac{\partial \widehat{G}(u_1)}{\partial u_1} = 0.$$

It follows that

$$\lim_{|u_1|\to+\infty} \left|\frac{\partial \widehat{G}(u_1)}{\partial u_1}\right| |\widehat{G}(u_1)| \left|\frac{\partial^{p+q}}{\partial u_2 \dots \partial v_q}\left\{\|(u,v)\|^{2\delta} \prod_{r=2}^{p} |\widehat{G}(u_r)|^2 \prod_{s=1}^{q} |\widehat{G}(v_s)|^2\right\}\right| = 0.$$

We then obtain from Equation (5.102) that

$$\left[\frac{\partial^{p+q}}{\partial u_1 \dots \partial v_q}\left\{\|(u,v)\|^{2\delta} \prod_{r=1}^{p} |\widehat{G}(u_r)|^2 \prod_{s=1}^{q} |\widehat{G}(v_s)|^2\right\} \frac{\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)}{\prod_{r=1}^{p} |\widehat{G}(u_r)|^2 \prod_{s=1}^{q} |\widehat{G}(v_s)|^2}\right]_{u_1=-\infty}^{u_1=+\infty} = 0.$$

Therefore, we have according to Equation (5.101) that

$$\begin{aligned}
&\int_{\mathbb{R}^{p+q}} \frac{\partial^{p+q} W}{\partial u_1 \dots \partial v_q}(u,v) R_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v \\
&= -\frac{1}{i(j_{1,1}-j_{2,1})} \int_{\mathbb{R}^{p+q}} \frac{\partial^{p+q+1} W}{\partial^2 u_1 \partial u_2 \dots \partial v_q}(u,v) R_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v.
\end{aligned}$$

Repeating this last process for the variables $u_2, \dots, u_p, v_1, \dots, v_q$ gives

$$\begin{aligned}
&\int_{\mathbb{R}^{p+q}} \frac{\partial^{p+q} W}{\partial u_1 \dots \partial v_q}(u,v) R_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v \\
&= \frac{(i)^{p+q}}{\prod_{r=1}^{p}(j_{1,r}-j_{2,r}) \prod_{s=1}^{q}(l_{1,s}-l_{1,s})} \int_{\mathbb{R}^{p+q}} \frac{\partial^{2(p+q)} W}{\partial^2 u_1 \dots \partial^2 v_q}(u,v) R_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v. \qquad (5.103)
\end{aligned}$$

By injecting this last Equation (5.103) in Equation (5.100), we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^{p+q}} \|(u,v)\|^{2\delta} \mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v \\
&= \frac{(-1)^{p+q}}{\prod_{r=1}^{p}(j_{1,r}-j_{2,r})^2 \prod_{s=1}^{q}(l_{1,s}-l_{1,s})^2} \int_{\mathbb{R}^{p+q}} \frac{\partial^{2(p+q)} W}{\partial^2 u_1 \dots \partial^2 v_q}(u,v) R_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v.
\end{aligned}$$

By analogy with the last equation, we can show that in the general case for two different vectors (j_1, l_1) and (j_2, l_2) that

$$\begin{aligned}
&\int_{\mathbb{R}^{p+q}} \|(u,v)\|^{2\delta} \mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v \\
&= \frac{(-1)^{|\mathcal{S}_1|+|\mathcal{S}_2|}}{\prod_{r\in\mathcal{S}_1}(j_{1,r}-j_{2,r})^2 \prod_{s\in\mathcal{S}_2}(l_{1,s}-l_{1,s})^2} \int_{\mathbb{R}^{p+q}} \frac{\partial^{2(p+q)} W}{\prod_{r\in\mathcal{S}_1} \partial^2 u_r \prod_{s\in\mathcal{S}_2} \partial^2 v_s}(u,v) T_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v,
\end{aligned}$$

where $\mathcal{S}_1$ (resp. $\mathcal{S}_2$) is the set of indices r (resp. s) such that $j_{1,r} \neq j_{2,r}$ (resp. $l_{1,s} \neq l_{1,s}$), while the notation $|\cdot|$ designates the cardinal. In addition, the function T is defined as

$$T_{j_1,l_1,j_2,l_2} : (u,v) \mapsto \frac{\mathcal{G}_{j_1,l_1,j_2,l_2}(u,v)}{\prod_{r\in\mathcal{S}_1} |\widehat{G}(u_r)|^2 \prod_{s\in\mathcal{S}_2} |\widehat{G}(v_s)|^2}.$$

Moreover, one can easily show using Lemma 5.11 that independently from the values of j_1, l_1 and j_2, l_2, we have

$$\left| \int_{\mathbb{R}^{p+q}} \frac{\partial^{2(p+q)} W}{\prod_{r\in\mathcal{S}_1} \partial^2 u_r \prod_{s\in\mathcal{S}_2} \partial^2 v_s}(u,v) T_{j_1,l_1,j_2,l_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v \right| \leq C(p,q).$$

We deduce using Equation (5.96) the following

$$\left| \int_{\mathbb{R}^{p+q}} \|(u,v)\|^{2\delta} H_{2,n}(u,v)\,\mathrm{d}u\,\mathrm{d}v \right| \leq C(p,q,\delta) h_n^{p+q} \sum_{\substack{(j_1,l_1)\in I_{n,p,q}\\(j_2,l_2)\in I_{n,p,q}\\(j_1,l_1)\neq(j_2,l_2)}} \frac{1}{\prod_{r\in\mathcal{S}_1}(j_{1,r}-j_{2,r})^2 \prod_{s\in\mathcal{S}_2}(l_{1,s}-l_{1,s})^2}. \tag{5.104}$$

Furthermore,

$$\begin{aligned}
&\sum_{\substack{(j_1,l_1)\in I_{n,p,q}\\(j_2,l_2)\in I_{n,p,q}\\(j_1,l_1)\neq(j_2,l_2)}} \frac{1}{\prod_{r\in\mathcal{S}_1}(j_{1,r}-j_{2,r})^2 \prod_{s\in\mathcal{S}_2}(l_{1,s}-l_{1,s})^2} \\
&\qquad = \sum_{\vartheta=1}^{p+q} \sum_{\substack{(j_1,l_1)\in I_{n,p,q}\\(j_2,l_2)\in I_{n,p,q}\\(j_1,l_1)\neq(j_2,l_2)}} \frac{1}{\prod_{r\in\mathcal{S}_1}(j_{1,r}-j_{2,r})^2 \prod_{s\in\mathcal{S}_2}(l_{1,s}-l_{1,s})^2} \\
&\qquad\qquad = \sum_{\vartheta=1}^{p+q} \binom{p+q}{\vartheta} M_n^{p+q-\vartheta} \left[\sum_{j\neq k=1}^{M_n} \frac{1}{(j-k)^2} \right]^{\vartheta}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\sum_{\substack{(j_1,l_1)\in I_{n,p,q}\\(j_2,l_2)\in I_{n,p,q}\\(j_1,l_1)\neq(j_2,l_2)}} \frac{1}{\prod_{r\in\mathcal{S}_1}(j_{1,r}-j_{2,r})^2 \prod_{s\in\mathcal{S}_2}(l_{1,s}-l_{1,s})^2} &= \sum_{\vartheta=1}^{p+q} \binom{p+q}{\vartheta} M_n^{p+q-\vartheta} \left[\sum_{l=1}^{M_n} \frac{M_n - l}{l^2} \right]^{\vartheta} \\
&\leq \sum_{\vartheta=1}^{p+q} \binom{p+q}{\vartheta} M_n^{p+q} \left[\sum_{l=1}^{M_n} \frac{1}{l^2} \right]^{\vartheta}.
\end{aligned}$$

Thus, using the convergence of the sum $\sum_{l>0} 1/l^2$, we obtain

$$\sum_{\substack{(j_1,l_1)\in I_{n,p,q}\\(j_2,l_2)\in I_{n,p,q}\\(j_1,l_1)\neq(j_2,l_2)}} \frac{1}{\prod_{r\in\mathcal{S}_1}(j_{1,r}-j_{2,r})^2 \prod_{s\in\mathcal{S}_2}(l_{1,s}-l_{1,s})^2} \quad \leq \quad C(p,q) M_n^{p+q}.$$

We deduce from Equation (5.104) that

$$\left| \int_{\mathbb{R}^{p+q}} \|(u,v)\|^{2\delta} H_{2,n}(u,v)\,\mathrm{d}u\,\mathrm{d}v \right| \leq C(p,q,\delta). \tag{5.105}$$

Consequently, combining Equations (5.90), (5.95) and (5.105) we have the following

$$\int_{\mathbb{R}^{p+q}} \|(u,v)\|^{2\delta} \left| \widehat{f}_\theta(u,v) - \widehat{f}_1 \otimes \widehat{f}_2(u,v) \right|^2 \mathrm{d}u\,\mathrm{d}v \leq C(p,q,\delta).$$

By well choosing the constant of the sequence $(h_n)_n$ w.r.t. p, q and δ, the constant $C(p,q,\delta)$ in the last equation can be replaced by R, which achieves this point of the proof.

Proof of 3. Let us prove that, for all $\theta \in \{-1,1\}^{M_n^{p+q}}$, f_θ satisfies $\|f_\theta - f_{\theta,1} \otimes f_{\theta,2}\|_{\mathbb{L}_2} = Cn^{-2\delta/(4\delta+p+q)}$.
Since, $\int_{\mathbb{R}} G(t)\mathrm{d}t = 0$, we know that $f_{\theta,1} = \mathbb{1}_{[0,1]^p}$ and $f_{\theta,2} = \mathbb{1}_{[0,1]^q}$, thus $f_{\theta,1} \otimes f_{\theta,2} = \mathbb{1}_{[0,1]^{p+q}}$ and

$$f_\theta - f_{\theta,1} \otimes f_{\theta,2} = h_n^{\delta+(p+q)} \sum_{(j,l)\in I_{n,p,q}} \theta_{(j,l)} g_{n,j,l}(x,y),$$

where the functions $g_{n,j,l}$ are defined in (5.88), with disjoint supports.
In particular,

$$\|f_\theta - f_{\theta,1} \otimes f_{\theta,2}\|_{\mathbb{L}_2}^2 = h_n^{2\delta+2(p+q)} \sum_{(j,l)\in I_{n,p,q}} \|g_{n,j,l}\|_{\mathbb{L}_2}^2 .$$

Moreover, for all $(j,l) \in I_{n,p,q}$,

$$\begin{aligned}
\|g_{n,j,l}\|_{\mathbb{L}_2}^2 &= \int_{\mathbb{R}^{p+q}} \left[\prod_{r=1}^{p} G_{h_n}^2(x_r - j_r h_n) \prod_{s=1}^{q} G_{h_n}^2(y_s - l_s h_n)\right] \mathrm{d}x_1 \dots \mathrm{d}x_p \mathrm{d}y_1 \dots \mathrm{d}y_q \\
&= \left[\prod_{r=1}^{p} \left(\int_{\mathbb{R}} G_{h_n}^2(x_r - j_r h_n)\mathrm{d}x_r\right)\right] \times \left[\prod_{s=1}^{q} \left(\int_{\mathbb{R}} G_{h_n}^2(y_s - l_s h_n)\mathrm{d}y_s\right)\right],
\end{aligned}$$

and for all k in $\{1,\dots,M_n\}$, a simple change of variables implies that

$$\int_{\mathbb{R}} G_{h_n}^2(t - kh_n)\mathrm{d}t = \frac{1}{h_n^2}\int_{\mathbb{R}} G^2\left(\frac{t-kh_n}{h_n}\right)\mathrm{d}t = \frac{1}{h_n}\int_{\mathbb{R}} G^2(t)\mathrm{d}t = \frac{C}{h_n},$$

since G belongs to $\mathbb{L}_2(\mathbb{R})$.
We thus deduce that

$$\|g_{n,j,l}\|_{\mathbb{L}_2}^2 = \frac{C(p,q)}{h_n^{p+q}} \tag{5.106}$$

and that, since the cardinality of $I_{n,p,q}$ equals M_n^{p+q},

$$\|f_\theta - f_{\theta,1} \otimes f_{\theta,2}\|_{\mathbb{L}_2}^2 = h_n^{2\delta+2(p+q)} \times M_n^{p+q} \times \frac{C(p,q)}{h_n^{p+q}} = C(p,q)h_n^{2\delta}.$$

5.7.18 Proof of Proposition 5.5

Let $\mathbb{Z}_n = (X_i, Y_i)_{1\le i\le n}$ be an i.i.d sample with common uniform distribution P_{f_0} on $[0,1]^{p+q}$.

For simplicity, denote for all $1 \le i \le n$ and all (j,l) in $I_{n,p,q}$,

$$a_{i,j,l} = h_n^{\delta+(p+q)} g_{n,j,l}(X_i, Y_i) = h_n^{\delta+(p+q)} \prod_{r=1}^{p} G_{h_n}(X_i^{(r)} - j_r h_n) \prod_{s=1}^{q} G_{h_n}(Y_i^{(s)} - l_s h_n),$$

where $g_{n,j,l}$ is defined in Equation (5.88), such that $f_\theta(X_i,Y_i) = 1 + \sum_{(j,l)\in I_{n,p,q}} \theta_{(j,l)} a_{i,j,l}$. Note that $a_{i,j,l} \neq 0$ if and only if (X_i, Y_i) belongs to the set $D_{(j,l)}$ defined in Equation (5.89).

Then, since $f_0 = \mathbb{1}_{[0,1]^{p+q}}$, the likelihood ratio equals

$$\begin{aligned} L_\nu(\mathbb{Z}_n) = \frac{\mathrm{d}P_\nu}{\mathrm{d}P_{f_0}}(\mathbb{Z}_n) &= \int \prod_{i=1}^{n} \frac{f_\theta}{f_0}(X_i, Y_i)\pi(\mathrm{d}\theta) \\ &= \mathbb{E}_\Theta\left[\prod_{i=1}^{n}\left(1 + \sum_{(j,l)\in I_{n,p,q}} \Theta_{(j,l)} a_{i,j,l}\right)\right], \end{aligned}$$

where $\Theta = (\Theta_{(j,l)})_{(j,l)\in I_{n,p,q}}$ has i.i.d. Rademacher components $\Theta_{(j,l)}$, and $\mathbb{E}_\Theta[\cdot]$ denotes the expectation w.r.t. Θ.

Noticing that for all $1 \leq i \leq n$, there exists a unique (j,l) in $I_{n,p,q}$ such that $a_{i,j,l} \neq 0$, we obtain

$$1 + \sum_{(j,l)\in I_{n,p,q}} \Theta_{(j,l)} a_{i,j,l} = \prod_{(j,l)\in I_{n,p,q}} \left(1 + \Theta_{(j,l)} a_{i,j,l}\right).$$

Thus,

$$\begin{aligned} L_\nu(\mathbb{Z}_n) &= \mathbb{E}_\Theta\left[\prod_{(j,l)\in I_{n,p,q}} \prod_{i=1}^{n} \left(1 + \Theta_{(j,l)} a_{i,j,l}\right)\right] \\ &= \prod_{(j,l)\in I_{n,p,q}} \left[\frac{1}{2}\prod_{i=1}^{n}(1 - a_{i,j,l}) + \frac{1}{2}\prod_{i=1}^{n}(1 + a_{i,j,l})\right]. \end{aligned}$$

Moreover, for ε in $\{-1, 1\}$,

$$\prod_{i=1}^{n}(1 + \varepsilon a_{i,j,l}) = 1 + \sum_{k=1}^{n} \varepsilon^k \left(\sum_{1\leq i_1<\ldots<i_k\leq n} a_{i_1,j,l} \ldots a_{i_k,j,l}\right).$$

Hence, by cancelling the odd terms, we obtain

$$\begin{aligned} \frac{1}{2}\prod_{i=1}^{n}(1 - a_{i,j,l}) + \frac{1}{2}\prod_{i=1}^{n}(1 + a_{i,j,l}) &= 1 + \sum_{k=1}^{[n/2]} \sum_{1\leq i_1<\ldots<i_{2k}\leq n} a_{i_1,j,l} \ldots a_{i_{2k},j,l} \\ &= 1 + \sum_{k=1}^{[n/2]} A_{k,j,l}, \end{aligned}$$

where $[\cdot]$ denotes the integer part, and

$$A_{k,j,l} = \sum_{1\leq i_1<\ldots<i_{2k}\leq n} a_{i_1,j,l} \ldots a_{i_{2k},j,l}. \tag{5.107}$$

Thus,

$$[L_\nu(\mathbb{Z}_n)]^2 = \prod_{(j,l)\in I_{n,p,q}} \left(1 + \sum_{k=1}^{[n/2]} A_{k,j,l}\right)^2 = \prod_{(j,l)\in I_{n,p,q}} (1 + B_{j,l}),$$

where

$$B_{j,l} = 2\sum_{k=1}^{[n/2]} A_{k,j,l} + \sum_{k,k'=1}^{[n/2]} A_{k,j,l} A_{k',j,l}. \tag{5.108}$$

Then,

$$[L_\nu(\mathbb{Z}_n)]^2 = 1 + \sum_{m=1}^{M_n^{p+q}} \frac{1}{m!} \sum_{(j_1,l_1),\dots,(j_m,l_m)}^{\neq} B_{j_1,l_1} \dots B_{j_m,l_m}, \tag{5.109}$$

where $\sum^{\neq}$ means that the indexes are all distinct.

After tedious computations, up to a possible permutation of the indexes $(j_1,l_1),\dots,(j_m,l_m)$, we can express the product $B_{j_1,l_1}\dots B_{j_m,l_m}$ as a sum of terms of the form

$$2^P \sum_{k_1=1}^{[n/2]} \sum_{k_2'',\dots,k_M''=1}^{[n/2]} 2A_{k_1,j_1,l_1} \times A_{k_2'',j_2',l_2'} \times \dots \times A_{k_M'',j_M',l_M'} \tag{5.110}$$

or

$$2^Q \sum_{k_1,k_1'=1}^{[n/2]} \sum_{k_2'',\dots,k_M''=1}^{[n/2]} A_{k_1,j_1,l_1} A_{k_1',j_1,l_1} \times A_{k_2'',j_2',l_2'} \times \dots \times A_{k_M'',j_M',l_M'} \tag{5.111}$$

where P and Q are integers, $M \in \{m-1,\dots,2m-2\}$ and $(j_2',l_2'),\dots,(j_M',l_M')$ are drawn in $(j_2,l_2),\dots,(j_m,l_m)$ such that each (j_r,l_r) for $2 \le r \le m$ appears exactly once or twice. To be more precise, P and Q count the number of indexes $(j_r,l_r)_r$, appearing exactly once in the product.

First note that in Equation (5.110), the index (j_1,l_1) appears only once. Moreover

$$\begin{aligned}
&\mathbb{E}_{f_0}\left[A_{k_1,j_1,l_1} \times A_{k_2'',j_2',l_2'} \times \dots \times A_{k_M'',j_M',l_M'}\right] \\
&\quad = \sum_{i_{1,1}<\dots<i_{1,2k_1}} \sum_{i_{2,1}<\dots<i_{2,2k_2''}} \dots \sum_{i_{M,1}<\dots<i_{M,2k_M''}} \mathbb{E}_{f_0}\left[a_{i_{1,1},j_1,l_1} \times \dots \times a_{i_{1,2k_1},j_1,l_1} \times \right. \\
&\qquad \left. \times a_{i_{2,1},j_2',l_2'} \times \dots \times a_{i_{2,2k_2''},j_2',l_2'} \times \dots \times a_{i_{M,1},j_M',l_M'} \times \dots \times a_{i_{M,2k_M''},j_M',l_M'}\right]
\end{aligned}$$

If $i_{1,1}$ appears at least twice in the sums, that is there exists $2 \le r \le M$ and $1 \le s \le 2k_r''$ such that $i_{1,1} = i_{r,s}$, then, $a_{i_{1,1},j_1,l_1} a_{i_{1,1},j_r,l_r} = 0$ since $D_{j_1,l_1} \cap D_{j_r,l_r} = \emptyset$. Otherwise, if $i_{1,1}$ appears only once, by independence between the $(X_i,Y_i)_i$, we obtain that

$$\begin{aligned}
&\mathbb{E}_{f_0}\left[a_{i_{1,1},j_1,l_1} \times \dots \times a_{i_{1,2k_1},j_1,l_1} \times a_{i_{2,1},j_2',l_2'} \times \dots \times a_{i_{M,2k_M''},j_M',l_M'}\right] \\
&\quad = \mathbb{E}_{f_0}\left[a_{i_{1,1},j_1,l_1}\right] \times \mathbb{E}_{f_0}\left[a_{i_{1,2},j_1,l_1} \dots \times a_{i_{1,2k_1},j_1,l_1} \times a_{i_{2,1},j_2',l_2'} \times \dots \times a_{i_{M,2k_M''},j_M',l_M'}\right] = 0
\end{aligned}$$

Hence,

$$\mathbb{E}_{f_0}[A_{k_1,j_1,l_1} \times A_{k_2'',j_2',l_2'} \times \dots \times A_{k_M'',j_M',l_M'}] = 0.$$

and thus, all the terms of the form (5.110) have a null expectation.

Let us now consider Equation (5.111) (where the index (j_1,l_1) appears twice).

$$\begin{aligned}
&\mathbb{E}_{f_0}[A_{k_1,j_1,l_1} \times A_{k_1',j_1,l_1} \times A_{k_2'',j_2',l_2'} \times \dots \times A_{k_M'',j_M',l_M'}] = \\
&\sum_{i_{1,1}<\dots<i_{1,2k_1}} \sum_{i'_{1,1}<\dots<i'_{1,2k_1'}} \sum_{i_{2,1}<\dots<i_{2,2k_2''}} \dots \sum_{i_{M,1}<\dots<i_{M,2k_M''}} \mathbb{E}_{f_0}\left[a_{i_{1,1},j_1,l_1} \times \dots \times a_{i_{1,2k_1},j_1,l_1} \times a_{i'_{1,1},j_1,l_1} \times \right. \\
&\qquad \left. \times \dots \times a_{i'_{1,2k_1'},j_1,l_1} \times a_{i_{2,1},j_2',l_2'} \times \dots \times a_{i_{2,2k_2''},j_2',l_2'} \times \dots \times a_{i_{M,1},j_M',l_M'} \times \dots \times a_{i_{M,2k_M''},j_M',l_M'}\right]
\end{aligned}$$

If there exists at least one index $i_{1,\cdot}$ or $i'_{1,\cdot}$ that can be isolated, then by independence,

$$\mathbb{E}_{f_0}\big[a_{i_{1,1},j_1,l_1}\times\ldots\times a_{i_{1,2k_1},j_1,l_1}\times a_{i'_{1,1},j_1,l_1}\times\ldots\times a_{i'_{1,2k'_1},j_1,l_1}\times \\ \times a_{i_{2,1},j'_2,l'_2}\times\ldots\times a_{i_{2,2k''_2},j'_2,l'_2}\times\ldots\times a_{i_{M,1},j'_M,l'_M}\times\ldots\times a_{i_{M,2k''_M},j'_M,l'_M}\big]=0.$$

Hence, the remaining terms are obtained for $k_1=k'_1$ and $i_{1,s}=i'_{1,s}$ for all $1\leq s\leq 2k_1$. These arguments are being valid for any index (j_r,l_r), we obtain that $Q=0$ and

$$\begin{aligned}\mathbb{E}_{f_0}\big[B_{j_1,l_1}\ldots B_{j_m,l_m}\big] &= \sum_{k_1,\ldots,k_m=1}^{[n/2]}\sum_{i_{1,1}<\ldots<i_{1,2k_1}}\cdots\sum_{i_{m,1}<\ldots<i_{m,2k_m}}\mathbb{E}_{f_0}\Big[a^2_{i_{1,1},j_1,l_1}\times\ldots\times a^2_{i_{1,2k_1},j_1,l_1}\times\ldots\\ &\qquad\times a^2_{i_{m,1},j_m,l_m}\times\ldots\times a^2_{i_{m,2k_m},j_m,l_m}\Big]\\ &= \sum_{k_1,\ldots,k_m=1}^{[n/2]}\sum_{\substack{I_1,\ldots,I_m\subset\{1,\ldots n\}\\ \mathrm{Card}(I_r)=2k_r}}\mathbb{E}_{f_0}\left[\prod_{i_1\in I_1}a^2_{i_1,j_1,l_1}\times\ldots\times\prod_{i_m\in I_m}a^2_{i_m,j_m,l_m}\right].\end{aligned}$$

If the subsets I_r are not pairwise disjoints, the product $\prod_{i_1\in I_1}a^2_{i_1,j_1,l_1}\times\ldots\times\prod_{i_m\in I_m}a^2_{i_m,j_m,l_m}=0$, since the supports $D_{(j_r,l_r)}$ are disjoint.

Thus,

$$\begin{aligned}\mathbb{E}_{f_0}[B_{j_1,l_1}\ldots B_{j_m,l_m}] &= \sum_{k_1,\ldots,k_m=1}^{[n/2]}\sum_{\substack{I_1,\ldots,I_m\subset\{1,\ldots n\}\\ I_r\cap I_s=\emptyset,\forall r\neq s\\ \mathrm{Card}(I_r)=2k_r}}\mathbb{E}_{f_0}\left[\prod_{i_1\in I_1}a^2_{i_1,j_1,l_1}\times\ldots\times\prod_{i_m\in I_m}a^2_{i_m,j_m,l_m}\right]\\ &= \sum_{k_1,\ldots,k_m=1}^{[n/2]}\sum_{\substack{I_1,\ldots,I_m\subset\{1,\ldots n\}\\ I_r\cap I_s=\emptyset,\forall r\neq s\\ \mathrm{Card}(I_r)=2k_r}}\prod_{i_1\in I_1}\mathbb{E}_{f_0}\big[a^2_{i_1,j_1,l_1}\big]\times\ldots\times\prod_{i_m\in I_m}\mathbb{E}_{f_0}\big[a^2_{i_m,j_m,l_m}\big],\end{aligned}$$

by independence of the $(X_i,Y_i)_{1\leq i\leq n}$.

Besides, for all $1\leq i\leq n$, $\mathbb{E}_{f_0}\Big[a^2_{i,j,l}\Big]=h_n^{2\delta+2p+2q}\left\|g_{n,j,l}\right\|^2_{\mathbb{L}_2}=C(p,q)h_n^{2\delta+p+q}$ by Equation (5.106). Thus,

$$\begin{aligned}\mathbb{E}_{f_0}[B_{j_1,l_1}\ldots B_{j_m,l_m}] &= \sum_{k_1,\ldots,k_m=1}^{[n/2]}\sum_{\substack{I_1,\ldots,I_m\subset\{1,\ldots n\}\\ I_r\cap I_s=\emptyset,\forall r\neq s\\ \mathrm{Card}(I_r)=2k_r}}\left(C(p,q)h_n^{2\delta+p+q}\right)^{2k_1+\ldots+2k_m}\\ &= \sum_{k_1,\ldots,k_m=1}^{[n/2]}\binom{n}{2k_1,\ldots,2k_m,n-\sum_{r=1}^m 2k_r}\left(C(p,q)h_n^{2\delta+p+q}\right)^{2k_1+\ldots+2k_m}.\end{aligned}$$

Moreover, the multinomial coefficient $\binom{n}{2k_1,\ldots,2k_m,n-\sum_{r=1}^m 2k_r}$ can be upper-bounded by $n^{2k_1+\ldots+2k_m}$. Hence,

$$\begin{aligned}\mathbb{E}_{f_0}[B_{j_1,l_1}\ldots B_{j_m,l_m}] &\leq \sum_{k_1=1}^{[n/2]}\cdots\sum_{k_m=1}^{[n/2]}\left(C(p,q)n\times h_n^{2\delta+p+q}\right)^{2k_1+\ldots+2k_m}\\ &= \left[\sum_{k=1}^{[n/2]}\left(C(p,q)n\times h_n^{2\delta+p+q}\right)^{2k}\right]^m.\end{aligned}$$

Furthermore, for h_n defined in (5.26) (for any constant $C(p,q,\alpha,\beta,\delta)$),

$$C(p,q)\ n\times h_n^{2\delta+p+q} = C(p,q,\alpha,\beta,\delta)n^{-(p+q)/(4\delta+p+q)} < 1/2$$

for n large enough, and thus, by property of geometric series, we get

$$\sum_{k=1}^{[n/2]}\left\{\left[C(p,q)\ n\times h_n^{2\delta+p+q}\right]^2\right\}^k \leq \frac{\left[C(p,q)\ n\times h_n^{2\delta+p+q}\right]^2}{1-\left[C(p,q)\ n\times h_n^{2\delta+p+q}\right]^2} \leq \frac{4}{3}\left[C(p,q)\ n\times h_n^{2\delta+p+q}\right]^2.$$

We recall that the constant $C(p,q)$ may vary from line to line. This being true for all $(j,l)\in I_{n,p,q}$, from Equation (5.109), we deduce that

$$\begin{aligned}\mathbb{E}_{f_0}\left[\{L_\nu(\mathbb{Z}_n)\}^2\right] &\leq 1+\sum_{m=1}^{M_n^{p+q}}\binom{M_n^{p+q}}{m}\left[C(p,q)\ n\times h_n^{2\delta+p+q}\right]^{2m}\\ &\leq 1+\sum_{m=1}^{M_n^{p+q}}\left\{M_n^{p+q}\left[C(p,q)\ n\times h_n^{2\delta+p+q}\right]^2\right\}^m\\ &\leq 1+\sum_{m=1}^{M_n^{p+q}}\left[C(p,q)\ n^2\times h_n^{4\delta+p+q}\right]^m,\end{aligned}$$

since $\binom{M_n^{p+q}}{m}\leq [M_n^{p+q}]^m$ and $M_n h_n = 1$.

Finally, for h_n defined in (5.26), with

$$C(p,q,\alpha,\beta,\delta) = \left(\frac{1}{C(p,q)}\times\frac{2(1-\alpha-\beta)^2}{1+4(1-\alpha-\beta)^2}\right)^{1/(4\delta+p+q)},$$

we directly obtain that

$$C(p,q)\ n^2\times h_n^{4\delta+p+q} = \frac{2(1-\alpha-\beta)^2}{1+4(1-\alpha-\beta)^2} < 1.$$

Hence, by property of the geometric series we obtain,

$$\begin{aligned}\mathbb{E}_{f_0}\left[(L_\nu(\mathbb{Z}_n))^2\right] &\leq 1+\frac{\left[C(p,q)\ n^2\times h_n^{4\delta+p+q}\right]}{1-\left[C(p,q)\ n^2\times h_n^{4\delta+p+q}\right]}\\ &< 1+4(1-\alpha-\beta)^2,\end{aligned}$$

which ends the proof of Proposition 5.5.

Chapter 6

Conclusion and Prospects (english)

Summary of the main contributions

The general scope of this manuscript is the Global Sensitivity Analysis (GSA) of numerical simulators, which consists in studying the impact of input uncertainties on the output. One possible and efficient method to perform GSA is based on the use of dependence measures, which quantify the probabilistic dependence between each input and the output. Indeed, under some assumptions, the measure of dependence between a given input and the output equal zero if and only if the two random variables are independent. Many statistical methods then evaluate the degree of dependence according to the estimated value of these measures. In this work, we focus in particular on the measure of dependence called HSIC (for Hilbert-Schmidt Independence Criterion). Indeed, HSIC measures are attractive tools to perform GSA, in view of their theoretical properties as well as the characteristics of the associated estimators. The objective of this thesis is to propose new theoretical and methodological tools, to face some limitations in the practical use of HSIC measures for GSA.

Statistical inference around HSIC measures

A first contribution of this Phd thesis are some theoretical developments around the estimation of HSIC. An estimation of HSIC using a sample from an alternative law of inputs, different from the prior one is proposed. To do this, we proposed weighted estimators based on importance sampling techniques. We also demonstrate the properties of these new estimators (bias, variance and asymptotic law), which are similar to the properties of the usual estimators. Moreover, asymptotic and non-asymptotic tests of independence built with these new estimators are also presented. These theoretical developments are a key step for the methodology suggested below, for the sensitivity analysis of a model whose input distributions are uncertain.

Methodology of GSA for uncertain probabilistic distributions of inputs

As a continuation of the work on the first contribution, the case of *second-level* Global Sensitivity Analysis (denoted here GSA2) is then discussed. Indeed, when the probability laws of the inputs are uncertain (e.g. due to lack of of knowledge), the purpose of GSA2 is to assess the impact of these uncertainties on the results of the usual GSA of the code output (denoted here GSA1 for *first-level* GSA). To perform GSA2, we proposed a new "single loop" Monte-Carlo methodology to answer various issues raised by GSA2 (characterization of GSA1 results, definition of 2^{nd}-level dependence measures and reduction of the calculation budget).

This methodology is based on a unique sample of inputs and associated outputs, drawn according to a well-chosen reference law. Three options have been proposed for this law: mixture distribution and barycentric distributions in the sense of Wasserstein and symmetric Kullback-Leibler distances. The construction of these distributions has been detailed and supported by illustrative examples. This new methodology seems to have good performances comparing to the straightforward "double loop" methodology and significantly reduces the simulation budget. In addition, the methodology is efficiently applied to the simulator MACARENa (French: *Modélisation de l'ACcident d'Arrêt des pompes d'un Réacteur refroidi au sodium*), developed by CEA to simulate the accidental scenario "Unprotected Loss Of Flow" (ULOF) for a sodium-cooled fast reactor, in the framework of safety demonstration.

These two contributions are the subject of two oral communications and a paper under revision for the scientific journal *SIAM/ASA Journal on Uncertainty Quantification* and currently available on the platform HALL.

- Meynaoui, A., Marrel, A., and Laurent-Bonneau, B. (2018). Méthodologie basée sur les mesures de dépendance HSIC pour l'analyse de sensibilité de second niveau. *Actes des 50ème Journées de Statistique (JDS 2018).(in French). Saclay, France.*
- Meynaoui, A., Marrel, A., and Laurent-Bonneau, B. (2018). Statistical methodology for second-level sensitivity analysis with dependence measures for numerical simulators. *SIAM conference on Uncertainty Quantification. Garden Grove, California, USA.*
- Meynaoui, A., Marrel, A., and Laurent, B. (2019). New statistical methodology for second-level global sensitivity analysis. Submitted to SIAM/ASA JUQ, under revision and available at *arXiv preprint arXiv:1902.07030.*

Theoretical development of an adaptive test of independence based on HSIC measures

This contribution involves the development of an HSIC-based testing procedure which aggregates several possible Gaussian kernel parametrizations. The objective is to overcome the prior choice of the kernels defining the HSIC and to give some theoretical guarantees (in terms of the non-asymptotic uniform separation rate) of the single and aggregated HSIC tests. For this, we first upper bound the uniform separation rate of the single tests over the regularity spaces of Sobolev and Nikol'skii balls. Then, we aggregate several single tests, and obtain similar upper bounds for the uniform separation rate of the aggregated procedure over the same regularity spaces. One other main result is the lower bound for the separation rate over Sobolev balls.

This lower bound allows to deduce that the aggregated procedure is adaptive in the minimax sense over Sobolev balls. In addition, given that the obtained sharp bounds over Nikol'skii balls are compared to the "classical" optimal separation rates over these spaces obtained for other statistical tests, the aggregated procedure seems to be optimal over Nikol'skii balls spaces as well.

Methodology of screening based on aggregated HSIC tests

In the continuity of the preceding works, the last contribution consist in using the procedure of HSIC aggregated test and the associated theoretical developments for screening purpose. For this, we propose a statistical methodology having a better ability to detect the dependence between the inputs and the output. We first propose a practical implementation of a non-asymptotic HSIC-test of level α, for any chosen parametrization. Then, analytical evaluation of the smallest uniform separation rate of theoretical tests over all parametrizations is demonstrated. The methodology is implemented on some analytical examples to investigate different ways of aggregating tests and to assess its efficiency . Lastly, the proposed methodology is illustrated on the test case of ULOF scenario with MACARENa simulator.

These last two contributions are the subject of an oral communication and a submitted paper which is currently available on the platform HALL.

- Meynaoui, A., Albert, M., Laurent, B., and Marrel, A. (2019). Aggregated tests of independence based on HSIC measures. European Meeting of Statisticians, Palermo, Italy.
- Meynaoui, A., Albert, M., Laurent, B., and Marrel, A. (2019). Adaptive test of independence based on HSIC measures. *arXiv preprint arXiv:1902.06441.*

Prospects

Some methodological and theoretical points can be improved and better investigated, to further extend the application scope of the proposed work. Moreover, the methodological developments proposed here could be compared to some existing methodologies in other frameworks.

A first possible improvement is to extend the two proposed methodologies (GSA2 and aggregated procedure of testing) to samples which are not generated according to pure Monte Carlo designs. Indeed, *Space-Filling* or *Quasi-Monte Carlo* designs which ensure a good coverage or repartition of points in the input space are often used in practice. This designs have the advantage of converging the estimators (of an expectation) more rapidly, especially in high dimension (large number of inputs), but do not have all the basic properties of Monte Carlo designs such as the independence of the realizations. In particular, the estimation of the quantiles involved in our statistical tests to compute the critical value should be corrected.

Moreover, it would be interesting to propose screening independence tests in the framework of GSA2. Meaning to perform statistical tests of independence between input distributions and GSA1 results. These independence tests will provide a more rigorous framework for GSA2 and allows to assess the dependence between input distributions and GSA1 results with more theoretical guarantees.

Another potential extension of this methodology is to combine the conclusions drawn from the GSA1 and GSA2. Once GSA2 is performed, the input distributions which have no impact

on GSA1 results can be set as a reference distribution, the GSA1 results can then be considered as the average results w.r.t. the GSA2 most influential input distributions. By analogy with the GSA2, this study can also be performed using the unique sample and importance sampling techniques.

The GSA2 methodology can also be compared to other methodologies such as Perturbed-Law based Indices (Lemaître et al., 2015) or the approach proposed by Chabridon et al. (2017) dealing with the estimation of failure probabilities in the presence of second-level uncertainties.

Furthermore, some theoretical and methodological improvements or extensions can be conducted for the aggregated testing procedure. In particular, we can attempt to show that the aggregated procedure is also adaptive over Nikol'skii balls. Indeed, the sharp upper bounds of the uniform separation rate obtained over these spaces support the intuition that the aggregated procedure is adaptive over these spaces.

Further works can also aim to extend the methodological scope of aggregated procedure. We may focus on taking into account other types of kernels than the Gaussian one. In addition, the choice of the collections of parameter bandwidths (or kernels), their relationships with the types of the inputs and the output (discrete, functional, etc.) and with the nature of the dependence (local, global) need to be more fully explored. A first possible solution for the selection of these parameters can be to choose a group of parameters containing both small and large values to cover more dependence forms. Another solution can be inspired from the cross-validation methods. A part of the available sample is set aside to identify the "good" bandwidths, while the other part is used to perform the aggregated test. In order to make this last procedure more robust, we can repeat the evaluation of the best bandwidths for different partitions of the initial sample and keep the bandwidths having in average the better performance (in terms of rate of selection of inputs).

Some other improvements such as the computation of the *p-value* associated to the aggregated procedure as well as the definition of an asymptotic framework can theoretically and practically be proposed.

Chapter 7

Conclusion et Perspectives (français)

Résumé des principales contributions

Le sujet de ce manuscrit est l'Analyse de Sensibilité Globale (ASG) des simulateurs numériques, qui consiste à étudier l'impact des incertitudes des variables d'entrée du simulateur sur la (ou les) sorties. Une méthode possible et efficace pour effectuer l'ASG est basée sur l'utilisation des mesures de dépendance, quantifiant d'un point de vue probabiliste la dépendance entre chaque entrée et la sortie. En effet, sous certaines hypothèses, la mesure de dépendance entre une entrée et la sortie est égale à zéro si et seulement si les deux variables aléatoires sont indépendantes. Des méthodes statistiques permettent alors de quantifier le degré de dépendance en fonction de la valeur estimée des mesures. Plus particulièrement, nous nous intéressons ici à la mesure de dépendance dénommée HSIC pour *Hilbert-Schmidt Independence Criterion*. Les mesures HSIC sont des outils pertinents et efficaces pour réaliser une ASG, compte tenu de leurs propriétés théoriques ainsi que des caractéristiques des estimateurs statistiques associés. Dans ce contexte, l'objectif de cette thèse est de proposer des nouveaux outils théoriques et méthodologiques autour de l'utilisation des mesures HSIC pour l'ASG, afin d'étendre encore davantage leur spectre applicatif et permettre une utilisation plus robuste.

Inférence statistique autour des mesures HSIC

Une première contribution de cette thèse est de proposer des développements théoriques autour de l'estimation des HSIC. Plus précisément, une estimation des HSIC à partir d'un échantillon issu d'une loi alternative des entrées, différente de la loi *a priori*, est proposée. Pour ce faire, des estimateurs pondérés basés sur des techniques d'échantillonnage préférentiel sont proposés. Les propriétés de ces nouveaux estimateurs (biais, variance et loi asymptotique), qui sont similaires aux propriétés des estimateurs usuels sont démontrées. De plus, des tests d'indépendance asymptotiques et non asymptotiques construits avec ces nouveaux estimateurs sont également proposés. Ces développements théoriques constituent une étape clé pour la méthodologie développée ensuite pour l'analyse de sensibilité d'un simulateur dont les distributions des entrées sont incertaines.

Méthodologie d'ASG pour les distributions incertaines des entrées

Dans le prolongement des travaux relatifs à la première contribution, le cas de l'Analyse de Sensibilité Globale de *second niveau* (en abrégé ASG2) est ensuite abordé. En effet, lorsque les lois de probabilité des entrées sont incertaines (par exemple en raison d'un manque de connaissance), l'objectif de l'ASG2 est d'évaluer l'impact de ces incertitudes sur les résultats de l'ASG usuelle de la sortie du simulateur (dénommée ici ASG1 pour ASG de *premier niveau*). Pour réaliser l'ASG2, une nouvelle méthodologie Monte-Carlo "simple boucle" est développée et permet de répondre aux diverses questions soulevées par l'ASG2, à savoir la caractérisation des résultats de l'ASG1, la définition de mesures de dépendance de 2^{nd}-niveau et la réduction du budget de simulations.

Cette méthodologie est basée sur un échantillon unique des entrées et sorties associées, tiré suivant une loi de référence bien choisie. Trois options ont été proposées pour cette loi : la distribution de mélange et les distributions barycentriques au sens des distances de Wasserstein et de Kullback-Leibler symétrisée. La construction de ces distributions a été détaillée et accompagnée par des exemples illustratifs. Cette nouvelle méthodologie présente de bonnes performances par rapport à la méthodologie classique "double boucle" et permet de considérablement réduire le budget de simulations. De plus, la méthodologie a été efficacement appliquée au simulateur MACARENa (Modélisation de l'ACcident d'Arrêt des pompes d'un Réacteur refroidi au sodium), développé par le CEA pour simuler le scénario accidentel de "Perte de débit non protégée" (ULOF, pour *Unprotected Loss Of Flow*) pour un réacteur rapide refroidi au sodium, dans le cadre d'une démonstration de sûreté.

Ces deux contributions ont fait l'objet de deux communications orales et d'un article en cours de révision pour la revue scientifique *SIAM/ASA Journal on Uncertainty Quantification*. Cet article est actuellement disponible sur la plateforme HALL.

- Meynaoui, A., Marrel, A., and Laurent-Bonneau, B. (2018). Méthodologie basée sur les mesures de dépendance HSIC pour l'analyse de sensibilité de second niveau. *Actes des 50ème Journées de Statistique (JDS 2018).(in French). Saclay, France.*
- Meynaoui, A., Marrel, A., and Laurent-Bonneau, B. (2018). Statistical methodology for second-level sensitivity analysis with dependence measures for numerical simulators. *SIAM conference on Uncertainty Quantification. Garden Grove, California, USA.*
- Meynaoui, A., Marrel, A., and Laurent, B. (2019). New statistical methodology for second-level global sensitivity analysis. Submitted to SIAM/ASA JUQ, under revision and available at *arXiv preprint arXiv:1902.07030.*

Développement théorique d'un test d'indépendance adaptatif basé sur les mesures HSIC

Cette contribution porte sur le développement d'une procédure de test basée sur les HSIC qui agrège plusieurs paramétrisations possibles des noyaux (noyaux gaussiens en l'occurrence). L'objectif est d'éviter le choix subjectif des paramètres des noyaux intervenant dans la définition du HSIC et de fournir certaines garanties théoriques des tests simples et agrégés (en termes de vitesse de séparation uniforme non asymptotique). Pour cela, des majorations fines des

vitesses de séparation uniforme des tests simples sur des espaces de régularité de types Sobolev et Nikol'skii sont d'abord obtenues. Ensuite, plusieurs tests simples sont ensuite agrégés et des bornes supérieures similaires aux précédentes sont obtenues pour la vitesse de séparation uniforme de la procédure agrégée, sur les espaces de Sobolev et Nikol'skii. Un autre résultat principal est la démonstration d'une borne inférieure pour la vitesse de séparation sur les espaces de Sobolev. Cette borne inférieure permet de déduire que la procédure agrégée est adaptative au sens minimax sur les espaces de Sobolev. En outre, les vitesses obtenues sur les espaces de Nikol'skii sont comparées aux vitesses de séparation optimales "classiques" sur ces mêmes espaces obtenues pour d'autres tests statistiques : la procédure agrégée semble également être optimale sur les espaces de Nikol'skii.

Méthodologie de criblage basée sur les tests HSIC agrégés

Dans la continuité des travaux précédents, la dernière contribution de cette thèse consiste à utiliser la procédure agrégée et les développements théoriques associés afin de réaliser un criblage des variables d'entrées. L'objectif est ainsi d'améliorer la capacité de détection de la dépendance entre une entrée et la sortie et de permettre un criblage plus robuste au choix des paramètres des noyaux. Pour cela, nous proposons une méthodologie pratique permettant tout d'abord de construire un test HSIC agrégé de niveau α dans un cadre non asymptotique et quelle que soit la collection de paramètres choisie pour les noyaux. Des collections et poids associés sont d'ailleurs proposés. L'évaluation de la plus petite vitesse de séparation uniforme des tests théoriques sur l'ensemble des paramétrisations est ensuite démontrée. La méthodologie est mise en œuvre sur plusieurs exemples analytiques pour étudier et comparer les différentes options d'agrégation des tests et pour évaluer l'efficacité générale de la méthode. Enfin, la méthodologie proposée est illustrée sur le cas test du scénario ULOF avec le simulateur MACARENa.

Ces deux dernières contributions font l'objet d'une communication orale et d'un document soumis et actuellement disponible sur la plateforme HALL.

- Meynaoui, A., Albert, M., Laurent, B., and Marrel, A. (2019). Aggregated tests of independence based on HSIC measures. European Meeting of Statisticians, Palermo, Italy.
- Meynaoui, A., Albert, M., Laurent, B., and Marrel, A. (2019). Adaptive test of independence based on HSIC measures. *arXiv preprint arXiv:1902.06441.*

Perspectives

Certains points méthodologiques et théoriques peuvent être améliorés et mériteraient d'être approfondis, pour étendre encore davantage la portée applicative des travaux proposés. En outre, les développements méthodologiques proposés ici pourraient être comparés à certaines méthodologies existantes dans d'autres cadres applicatifs.

Une première amélioration possible est d'étendre les deux méthodologies proposées (ASG2 et procédure de test agrégée) aux échantillons qui ne sont pas générés suivant des tirages Monte-Carlo purs. En effet, les plans de types *Space-Filling* ou *Quasi-Monte-Carlo* qui assurent une bonne couverture ou répartition des points dans l'espace des entrées sont très souvent utilisés en pratique dans la prise en compte des incertitudes en simulation numérique. Ces plans présentent l'avantage d'obtenir une convergence plus rapide des estimateurs (d'une espérance), en particulier

en grande dimension (grand nombre des entrées), mais ne possèdent pas toutes les propriétés des plans purs Monte-Carlo telles que l'indépendance des réalisations. En particulier, l'estimation des quantiles intervenant dans nos procédures de tests statistiques nécessite d'être corrigée et adaptée.

Par ailleurs, il serait intéressant de proposer des tests d'indépendance pour le criblage dans le cadre de l'ASG2, *i.e.* des tests statistiques d'indépendance entre les distributions des entrées et les résultats d'ASG1. Ces tests d'indépendance apporteront un cadre statistique plus rigoureux pour l'ASG2 et permettront d'évaluer la dépendance entre les distributions des entrées et les résultats d'ASG1 de manière plus robuste et avec davantage de garanties théoriques.

Une autre extension envisageable consiste à combiner les conclusions tirées de l'ASG1 et de l'ASG2. A partir des résultats d'ASG2, les distributions des entrées qui n'ont pas d'impact sur les résultats d'ASG1 pourront être fixées à une distribution de référence et les résultats d'ASG1 pourront être considérés comme les résultats moyens par rapport aux distributions des entrées les plus influentes de l'ASG2. Comme pour l'ASG2, cette ASG1 "moyenne" pourra être réalisée à partir d'un échantillon unique en utilisant les estimateurs pondérés inspirés de l'échantillonnage préférentiel.

La méthodologie de l'ASG2 peut également être comparée à d'autres approches telles que celle basée sur des indices obtenus par perturbation des lois (Lemaître et al., 2015) ou celle proposée par Chabridon et al. (2017) portant sur l'estimation des probabilités de défaillance en présence d'incertitudes de second niveau.

Concernant la procédure de test agrégée, certaines améliorations ou extensions théoriques et méthodologiques peuvent être apportées. En particulier, il pourrait être intéressant de démontrer que la procédure agrégée est également adaptative sur les espaces de Nikol'skii. En effet, les bornes supérieures fines des vitesses de séparation uniforme obtenues sur ces espaces renforcent l'intuition que la procédure agrégée est aussi adaptative sur ces espaces.

D'autres travaux peuvent également viser à étendre le champ applicatif de la procédure agrégée. On pourra par exemple envisager de prendre en compte d'autres types de noyaux. Plus généralement, le choix des collections des paramétrisations (ou noyaux), leurs relations avec les types d'entrées et de sorties (discrètes, fonctionnelles, etc.) et avec la nature de la dépendance (locale, globale) méritent d'être explorés plus en profondeur. Une première solution possible pour une sélection plus optimale des paramètres de la collection pourrait être de choisir un groupe de paramètres contenant à la fois des petites et grandes valeurs pour couvrir davantage de formes de dépendance. Une autre solution serait de s'inspirer des méthodes de validation croisée. Une partie de l'échantillon disponible est mise de côté pour identifier les "bons" paramètres, tandis que l'autre partie est utilisée pour effectuer le test agrégé. Afin de rendre cette dernière procédure plus robuste, elle pourra être répétée pour différentes partitions de l'échantillon initial.

Enfin, d'autres améliorations telles que le calcul de la *p-valeur* associée à la procédure agrégée ainsi que la définition d'un cadre asymptotique pourront être développées d'un point de vue théorique et pratique.

Bibliography

Agueh, M. and Carlier, G. (2011). Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924.

Alam, F. M., McNaught, K. R., and Ringrose, T. J. (2004). Using morris' randomized oat design as a factor screening method for developing simulation metamodels. In *Proceedings of the 36th conference on Winter simulation*, pages 949–957. Winter Simulation Conference.

Albert, M. (2015). *Tests of independence by bootstrap and permutation: an asymptotic and non-asymptotic study. Application to neurosciences.* PhD thesis, Université Nice Sophia Antipolis.

Alvarez, D. A. (2009). Reduction of uncertainty using sensitivity analysis methods for infinite random sets of indexable type. *International journal of approximate reasoning*, 50(5):750–762.

Arcones, M. A. and Gine, E. (1993). Limit theorems for u-processes. *The Annals of Probability*, pages 1494–1542.

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3):337–404.

Bach, F. R. and Jordan, M. I. (2002). Kernel independent component analysis. *Journal of machine learning research*, 3(Jul):1–48.

Bae, H.-R., Grandhi, R. V., and Canfield, R. A. (2004). Epistemic uncertainty quantification techniques including evidence theory for large-scale structures. *Computers & Structures*, 82(13-14):1101–1112.

Baker, C. R. (1973). Joint measures and cross-covariance operators. *Transactions of the American Mathematical Society*, 186:273–289.

Baraud, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8(5):577–606.

Baraud, Y., Huet, S., and Laurent, B. (2003). Adaptive tests of linear hypotheses by model selection. *The Annals of Statistics*, 31(1):225–251.

Baucells, M. and Borgonovo, E. (2013). Invariant probabilistic sensitivity analysis. *Management Science*, 59(11):2536–2549.

Bergsma, W. and Dassios, A. (2014). A consistent test of independence based on a sign covariance related to kendall's tau. *Bernoulli*, 20(2):1006–1028.

Berlinet, A. and Thomas-Agnan, C. (2011). *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media.

Berrett, T. B. and Samworth, R. J. (2017). Nonparametric independence testing via mutual information. *arXiv preprint arXiv:1711.06642*.

Blaschko, M. and Gretton, A. (2009). Learning taxonomies by dependence maximization. In *Advances in neural information processing systems*, pages 153–160.

Bochner, S. and Chandrasekharan, K. (1949). *Fourier transforms*. Princeton University Press.

Butucea, C. (2007). Goodness-of-fit testing and quadratic functional estimation from indirect observations. *The Annals of Statistics*, 35(5):1907–1930.

Butucea, C. and Tribouley, K. (2006). Nonparametric homogeneity tests. *Journal of statistical planning and inference*, 136(3):597–639.

Cacuci, D. G. (1981). Sensitivity theory for nonlinear systems. i. nonlinear functional analysis approach. *Journal of Mathematical Physics*, 22(12):2794–2802.

Cacuci, D. G. (2003). *Sensitivity & uncertainty analysis, volume 1: Theory*. Chapman and Hall/CRC.

Caflisch, R. E. (1998). Monte carlo and quasi-monte carlo methods. *Acta numerica*, 7:1–49.

Cannamela, C. (2007). *Apport des méthodes probabilistes dans la simulation du comportement sous irradiation du combustible à particules*. PhD thesis, Paris 7.

Chabridon, V. (2018). *Reliability-oriented sensitivity analysis under probabilistic model uncertainty-Application to aerospace systems*. PhD thesis, UNIVERSITE CLERMONT AUVERGNE.

Chabridon, V., Balesdent, M., Bourinet, J.-M., Morio, J., and Gayton, N. (2017). Evaluation of failure probability under parameter epistemic uncertainty: application to aerospace system reliability assessment. *Aerospace Science and Technology*, 69:526–537.

Christmann, A. and Steinwart, I. (2010). Universal kernels on non-standard input spaces. In *Advances in neural information processing systems*, pages 406–414.

Chwialkowski, K. P., Sejdinovic, D., and Gretton, A. (2014). A wild bootstrap for degenerate kernel tests. In *Advances in neural information processing systems*, pages 3608–3616.

Cioppa, T. M. (2002). *Efficient nearly orthogonal and space-filling experimental designs for high-dimensional complex models*. PhD thesis.

Cortes, C., Mohri, M., and Rostamizadeh, A. (2012). Algorithms for learning kernels based on centered alignment. *Journal of Machine Learning Research*, 13(Mar):795–828.

Csiszár, I. (1972). A class of measures of informativity of observation channels. *Periodica Mathematica Hungarica*, 2(1-4):191–213.

Cukier, R., Fortuin, C., Shuler, K. E., Petschek, A., and Schaibly, J. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. i theory. *The Journal of chemical physics*, 59(8):3873–3878.

Da Veiga, S. (2015). Global sensitivity analysis with dependence measures. *Journal of Statistical Computation and Simulation*, 85(7):1283–1305.

Da Veiga, S. and Gamboa, F. (2013). Efficient estimation of sensitivity indices. *Journal of Nonparametric Statistics*, 25(3):573–595.

Da Veiga, S., Wahl, F., and Gamboa, F. (2009). Local polynomial estimation for sensitivity analysis on models with correlated inputs. *Technometrics*, 51(4):452–463.

Dantan, J.-Y., Gayton, N., Qureshi, A. J., Lemaire, M., and Etienne, A. (2013). Tolerance analysis approach based on the classification of uncertainty (aleatory/epistemic). *Procedia CIRP*, 10:287–293.

De Lozzo, M. and Marrel, A. (2016a). Estimation of the derivative-based global sensitivity measures using a gaussian process metamodel. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):708–738.

De Lozzo, M. and Marrel, A. (2016b). New improvements in the use of dependence measures for sensitivity analysis and screening. *Journal of Statistical Computation and Simulation*, 86(15):3038–3058.

De Rocquigny, E., Devictor, N., and Tarantola, S. (2008). *Uncertainty in industrial practice*. Wiley.

Dempster, A. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38(2):325–339.

Droin, J.-B. (2016). *Modélisation d'un transitoire de perte de débit primaire non protégé dans un RNR-Na*. PhD thesis, Université Grenoble Alpes.

Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Everitt, B. and Hand, D. (1981). *Finite Mixture Distributions*. Monographs on applied probability and statistics. Chapman and Hall.

Fromont, M., Laurent, B., and Reynaud-Bouret, P. (2013). The two-sample problem for poisson processes: Adaptive tests with a nonasymptotic wild bootstrap approach. *The Annals of Statistics*, 41(3):1431–1461.

Fukumizu, K., Bach, F. R., and Jordan, M. I. (2004). Dimensionality reduction for supervised learning with reproducing kernel hilbert spaces. *Journal of Machine Learning Research*, 5(Jan):73–99.

Fukumizu, K., Gretton, A., Lanckriet, G. R., Schölkopf, B., and Sriperumbudur, B. K. (2009). Kernel choice and classifiability for rkhs embeddings of probability distributions. In *Advances in neural information processing systems*, pages 1750–1758.

Fukumizu, K., Gretton, A., Sun, X., and Schölkopf, B. (2008). Kernel measures of conditional dependence. In *Advances in neural information processing systems*, pages 489–496.

Ghanem, R., Higdon, D., and Owhadi, H. (2017). *Handbook of uncertainty quantification*. Springer.

Giné, E., Latała, R., and Zinn, J. (2000). Exponential and moment inequalities for u-statistics. In *High Dimensional Probability II*, pages 13–38. Springer.

Givens, C. R. and Shortt, R. M. (1984). A class of wasserstein metrics for probability distributions. *The Michigan Mathematical Journal*, 31(2):231–240.

Gretton, A. (2015). A simpler condition for consistency of a kernel independence test. *arXiv preprint arXiv:1501.06103*.

Gretton, A. (2019). Notes on mean embeddings and covariance operators.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. (2012). A kernel two-sample test. *Journal of Machine Learning Research*, 13:723–773.

Gretton, A., Bousquet, O., Smola, A., and Scholkopf, B. (2005a). Measuring statistical dependence with hilbert-schmidt norms. In *International conference on algorithmic learning theory*, volume 16, pages 63–78. Springer.

Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B., and Smola, A. J. (2008). A kernel statistical test of independence. In *Advances in neural information processing systems*, pages 585–592.

Gretton, A. and Györfi, L. (2010). Consistent nonparametric tests of independence. *Journal of Machine Learning Research*, 11(Apr):1391–1423.

Gretton, A., Herbrich, R., Smola, A., Bousquet, O., and Schölkopf, B. (2005b). Kernel methods for measuring independence. *Journal of Machine Learning Research*, 6(Dec):2075–2129.

Gretton, A., Herbrich, R., and Smola, A. J. (2003). The kernel mutual information. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 4, pages IV–880. IEEE.

Gretton, A., Smola, A. J., Bousquet, O., Herbrich, R., Belitski, A., Augath, M., Murayama, Y., Pauls, J., Schölkopf, B., and Logothetis, N. K. (2005c). Kernel constrained covariance for dependence measurement. In *AISTATS*, volume 10, pages 112–119.

Hall, M. C., Cacuci, D. G., and Schlesinger, M. E. (1982). Sensitivity analysis of a radiative-convective model by the adjoint method. *Journal of the Atmospheric Sciences*, 39(9):2038–2050.

Hellinger, E. (1909). Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. *Journal für die reine und angewandte Mathematik*, 136:210–271.

Helton, J. C. (1997). Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. *Journal of Statistical Computation and Simulation*, 57(1-4):3–76.

Hoeffding, W. (1948a). A class of statistics with asymptotically normal distributions. *Annals of Mathematical Statiatics*, 19:293–325.

Hoeffding, W. (1948b). A non-parametric test of independence. *The annals of mathematical statistics*, pages 546–557.

Hoeffding, W. (1992). A class of statistics with asymptotically normal distribution. In *Breakthroughs in Statistics*, pages 308–334. Springer.

Hora, S. C. (1996). Aleatory and epistemic uncertainty in probability elicitation with an example from hazardous waste management. *Reliability Engineering & System Safety*, 54(2-3):217–223.

Houdré, C. and Reynaud-Bouret, P. (2003). Exponential inequalities, with constants, for u-statistics of order two. In *Stochastic inequalities and applications*, pages 55–69. Springer.

Husslage, B. G., Rennen, G., Van Dam, E. R., and Den Hertog, D. (2011). Space-filling latin hypercube designs for computer experiments. *Optimization and Engineering*, 12(4):611–630.

Ingster, Y. I. (1989). An asymptotically minimax test of the hypothesis of independence. *J. Soviet Math*, 44:466–476.

Ingster, Y. I. (1993a). Asymptotically minimax hypothesis testing for nonparametric alternatives. i, ii, iii. *Math. Methods Statist*, 2(2):85–114.

Ingster, Y. I. (1993b). Minimax testing of the hypothesis of independence for ellipsoids in l_p. *Zapiski Nauchnykh Seminarov POMI*, 207:77–97.

Ingster, Y. I. and Suslina, I. A. (1998). Minimax detection of a signal for besov bodies and balls. *Problemy Peredachi Informatsii*, 34(1):56–68.

Iooss, B. (2011). Revue sur l'analyse de sensibilité globale de modèles numériques. *Journal de la Société Française de Statistique*, 152(1):1–23.

Iooss, B. and Lemaître, P. (2015). A review on global sensitivity analysis methods. In *Uncertainty management in simulation-optimization of complex systems*, pages 101–122. Springer.

Ishigami, T. and Homma, T. (1990). An importance quantification technique in uncertainty analysis for computer models. In *Uncertainty Modeling and Analysis, 1990. Proceedings., First International Symposium on*, pages 398–403. IEEE.

Janon, A., Klein, T., Lagnoux, A., Nodet, M., and Prieur, C. (2014). Asymptotic normality and efficiency of two sobol index estimators. *ESAIM: Probability and Statistics*, 18:342–364.

Jiao, Y. and Vert, J.-P. (2018). The kendall and mallows kernels for permutations. *IEEE transactions on pattern analysis and machine intelligence*, 40(7):1755–1769.

Johnson, D. and Sinanovic, S. (2001). Symmetrizing the kullback-leibler distance. *IEEE Transactions on Information Theory*.

Joseph, V. R., Gul, E., and Ba, S. (2015). Maximum projection designs for computer experiments. *Biometrika*, 102(2):371–380.

Kojadinovic, I. and Holmes, M. (2009). Tests of independence among continuous random vectors based on cramér–von mises functionals of the empirical copula process. *Journal of Multivariate Analysis*, 100(6):1137–1154.

Kucherenko, S. and Iooss, B. (2017). Derivative-based global sensitivity measures. *Handbook of uncertainty quantification*, pages 1241–1263.

Kucherenko, S., Rodriguez-Fernandez, M., Pantelides, C., and Shah, N. (2009). Monte carlo evaluation of derivative-based global sensitivity measures. *Reliability Engineering & System Safety*, 94(7):1135–1148.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86.

Laurent, B., Loubes, J.-M., and Marteau, C. (2012). Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. *Electronic Journal of Statistics*, 6:91–122.

Lehmann, E. L. and Romano, J. P. (2006). *Testing statistical hypotheses*. Springer Science & Business Media.

Lemaître, P. (2014). *Analyse de sensibilité en fiabilité des structures*. PhD thesis, Bordeaux.

Lemaître, P., Sergienko, E., Arnaud, A., Bousquet, N., Gamboa, F., and Iooss, B. (2015). Density modification-based reliability sensitivity analysis. *Journal of Statistical Computation and Simulation*, 85(6):1200–1223.

Mania, H., Ramdas, A., Wainwright, M. J., Jordan, M. I., and Recht, B. (2018). On kernel methods for covariates that are rankings. *Electronic Journal of Statistics*, 12(2):2537–2577.

Marrel, A., Iooss, B., Laurent, B., and Roustant, O. (2009). Calculations of sobol indices for the gaussian process metamodel. *Reliability Engineering & System Safety*, 94(3):742–751.

Meyer, M. A. and Booker, J. M. (2001). *Eliciting and analyzing expert judgment: a practical guide*, volume 7. Siam.

Meynaoui, A., Albert, M., Laurent, B., and Marrel, A. (2019). Aggregated test of independence based on hsic measures. *arXiv preprint arXiv:1902.06441*.

Micchelli, C. A., Xu, Y., and Zhang, H. (2006). Universal kernels. *Journal of Machine Learning Research*, 7(Dec):2651–2667.

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174.

Niederreiter, H. (1992). *Random number generation and quasi-Monte Carlo methods*, volume 63. Siam.

Oakley, J. E. and O'Hagan, A. (2004). Probabilistic sensitivity analysis of complex models: a bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):751–769.

Oberkampf, W., Helton, J., and Sentz, K. (2001). Mathematical representation of uncertainty. In *19th AIAA Applied Aerodynamics Conference*, page 1645.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T. (2006). *Uncertain judgements: eliciting experts' probabilities*. John Wiley & Sons.

Owen, A. B. (2013). Better estimation of small sobol'sensitivity indices. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 23(2):11.

Pfister, N., Bühlmann, P., Schölkopf, B., and Peters, J. (2018). Kernel-based tests for joint independence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):5–31.

Póczos, B., Ghahramani, Z., and Schneider, J. (2012). Copula-based kernel dependency measures. *arXiv preprint arXiv:1206.4682*.

Pronzato, L. and Müller, W. G. (2012). Design of computer experiments: space filling and beyond. *Statistics and Computing*, 22(3):681–701.

Pujol, G. (2009). Simplex-based screening designs for estimating metamodels. *Reliability Engineering & System Safety*, 94(7):1156–1160.

Romano, J. P. and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108.

Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. (2010). Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270.

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. (2004). Sensitivity analysis in practice: a guide to assessing scientific models. *Chichester, England*.

Saltelli, A., Tarantola, S., and Chan, K.-S. (1999). A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41(1):39–56.

Schölkopf, B., Tsuda, K., and Vert, J.-P. (2004). *Kernel methods in computational biology*. MIT press.

Sejdinovic, D., Sriperumbudur, B., Gretton, A., Fukumizu, K., et al. (2013). Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *The Annals of Statistics*, 41(5):2263–2291.

Serfling, R. J. (2009). *Approximation theorems of mathematical statistics*, volume 162. John Wiley & Sons.

Shafer, G. (1976). *A mathematical theory of evidence*, volume 42. Princeton university press.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423.

Siegmund, D. (2013). *Sequential analysis: tests and confidence intervals*. Springer Science & Business Media.

Smets, P. (1994). What is dempster-shafer's model. *Advances in the Dempster-Shafer theory of evidence*, pages 5–34.

Sobol, I. and Kucherenko, S. (2010). Derivative based global sensitivity measures. *Procedia-Social and Behavioral Sciences*, 2(6):7745–7746.

Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical modelling and computational experiments*, 1(4):407–414.

Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and computers in simulation*, 55(1-3):271–280.

Song, L., Smola, A., Gretton, A., Bedo, J., and Borgwardt, K. (2012). Feature selection via dependence maximization. *Journal of Machine Learning Research*, 13(May):1393–1434.

Spokoiny, V. G. (1996). Adaptive hypothesis testing using wavelets. *The Annals of Statistics*, 24(6):2477–2498.

Sriperumbudur, B. K., Gretton, A., Fukumizu, K., Schölkopf, B., and Lanckriet, G. R. (2010). Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research*, 11(Apr):1517–1561.

Steinwart, I. (2001). On the influence of the kernel on the consistency of support vector machines. *Journal of machine learning research*, 2(Nov):67–93.

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability engineering & system safety*, 93(7):964–979.

Sudret, B. and Mai, C. V. (2015). Computing derivative-based global sensitivity measures using polynomial chaos expansions. *Reliability Engineering & System Safety*, 134:241–250.

Sugiyama, M. and Suzuki, T. (2011). Least-squares independence test. *IEICE TRANSACTIONS on Information and Systems*, 94(6):1333–1336.

Sugiyama, M. and Yamada, M. (2012). On kernel parameter selection in hilbert-schmidt independence criterion. *IEICE TRANSACTIONS on Information and Systems*, 95(10):2564–2567.

Suzuki, T., Sugiyama, M., Kanamori, T., and Sese, J. (2009). Mutual information estimation reveals global associations between stimuli and biological processes. *BMC bioinformatics*, 10(1):S52.

Swiler, L. P., Paez, T. L., and Mayes, R. L. (2009). Epistemic uncertainty quantification tutorial. In *Proceedings of the 27th International Modal Analysis Conference*.

Szabó, Z. and Sriperumbudur, B. K. (2018). Characteristic and universal tensor product kernels. *Journal of Machine Learning Research*, 18(233):1–29.

Székely, G. J. and Rizzo, M. L. (2013). The distance correlation t-test of independence in high dimension. *Journal of Multivariate Analysis*, 117:193–213.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The annals of statistics*, 35(6):2769–2794.

Székely, G. J., Rizzo, M. L., et al. (2009). Brownian distance covariance. *The annals of applied statistics*, 3(4):1236–1265.

Tarantola, S., Gatelli, D., and Mara, T. A. (2006). Random balance designs for the estimation of first order global sensitivity indices. *Reliability Engineering & System Safety*, 91(6):717–727.

Titterington, D. M., Smith, A. F., and Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. Wiley,.

Tsybakov, A. B. (2009). Introduction to nonparametric estimation. revised and extended from the 2004 french original. translated by vladimir zaiats.

Veldhuis, R. (2002). The centroid of the symmetrical kullback-leibler distance. *IEEE signal processing letters*, 9(3):96–99.

Villani, C. (2003). *Topics in optimal transportation*. Number 58. American Mathematical Soc.

Wang, G. G. and Shan, S. (2007). Review of metamodeling techniques in support of engineering design optimization. *Journal of Mechanical design*, 129(4):370–380.

Weihs, L., Drton, M., and Meinshausen, N. (2018). Symmetric rank covariances: a generalized framework for nonparametric measures of dependence. *Biometrika*, 105(3):547–562.

Weiss, N. A. (2006). *A course in probability*. Addison-Wesley.

Yao, S., Zhang, X., and Shao, X. (2018). Testing mutual independence in high dimension via distance covariance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(3):455–480.

Yodé, A. (2004). Asymptotically minimax test of independence. *Mathematical Methods of Statistics*, 13(2):201–234.

Yodé, A. (2011). Adaptive minimax test of independence. *Mathematical Methods of Statistics*, 20(3):246.

Zhang, K., Peters, J., Janzing, D., and Schölkopf, B. (2012). Kernel-based conditional independence test and application in causal discovery. *arXiv preprint arXiv:1202.3775*.

Zhang, Q., Filippi, S., Gretton, A., and Sejdinovic, D. (2018). Large-scale kernel methods for independence testing. *Statistics and Computing*, 28(1):113–130.

Résumé

Dans le cadre des études de sûreté pour les réacteurs nucléaires, les simulateurs numériques sont essentiels pour comprendre, modéliser et prévoir des phénomènes physiques. Les informations relatives à certaines entrées des simulateurs sont souvent limitées ou incertaines. L'Analyse de Sensibilité Globale (ASG) vise alors à déterminer comment la variabilité des paramètres en entrée influe sur la valeur de la sortie ou de la quantité d'intérêt. Les travaux réalisés dans cette thèse ont pour objectif de proposer des nouvelles méthodes statistiques basées sur les mesures de dépendance pour l'ASG des simulateurs numériques. On s'intéresse plus particulièrement aux mesures de dépendance de type HSIC (*Hilbert-Schmidt Independence Criterion*). Après les Chapitres 1 et 2 introduisant le contexte général et les motivations de la thèse respectivement en versions française et anglaise, le Chapitre 3 présente d'abord une revue générale des mesures HSIC, dans un cadre théorique et méthodologique. Ensuite, des nouveaux développements autour de l'estimation des mesures HSIC à partir d'un échantillon alternatif et s'inspirant des techniques d'échantillonnage préférentiel sont proposés. Grâce à ces développements théoriques, une méthodologie efficace pour l'ASG en présence d'incertitudes sur les distributions de probabilité des entrées est développée dans le Chapitre 4. La pertinence de la méthodologie proposée est démontrée d'abord sur un cas analytique avant d'être appliquée au simulateur MACARENa modélisant un scénario accidentel de type ULOF (*Unprotected Loss Of Flow*), sur un réacteur à neutrons rapides refroidi au sodium. Le Chapitre 5 porte ensuite sur le développement d'un test d'indépendance agrégeant plusieurs paramétrisations des noyaux intervenant dans les HSIC. La méthodologie proposée permet ainsi de capturer un plus large spectre de dépendance entre les entrées et la sortie. L'optimalité de cette méthodologie est tout d'abord démontrée d'un point de vue théorique. Ses performances et son intérêt applicatif sont ensuite illustrés sur plusieurs exemples analytiques ainsi que sur le cas du simulateur MACARENa.

Abstract

As part of safety studies for nuclear reactors, numerical simulators are essential for understanding, modelling and predicting physical phenomena. However, the information on some of the input variables of the simulator is often limited or uncertain. In this framework, Global Sensitivity Analysis (GSA) aims at determining how the variability of the input parameters affects the value of the output or the quantity of interest. The work carried out in this thesis aims at proposing new statistical methods based on dependence measures for GSA of numerical simulators. We are particularly interested in HSIC-type dependence measures (Hilbert-Schmidt Independence Criterion). After Chapters 1 and 2 introducing the general context and motivations of the thesis in French and English versions respectively, Chapter 3 first presents a general review of HSIC measures, in a theoretical and methodological framework. Subsequently, new developments around the estimation of HSIC measures from an alternative sample and inspired by importance sampling techniques are proposed. As a result of these theoretical developments, an efficient methodology for GSA in the presence of uncertainties of input probability distributions is developed in Chapter 4. The relevance of the proposed methodology is first demonstrated on an analytical case before being applied to the MACARENa simulator modeling a ULOF (*Unprotected Loss Of Flow*) accidental scenario on a sodium-cooled fast neutron reactor. Finally, Chapter 5 deals with the development of an independence test aggregating several parametrizations of HSIC kernels and allowing to capture a wider spectrum of dependencies between the inputs and the output. The optimality of this methodology is first demonstrated from a theoretical point of view. Then, its performance and practical interest are illustrated on several analytical examples as well as on the test case of the MACARENa simulator.