

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le 25/06/2019 **par :**

Pierre DERENNES

Mesures de sensibilité de Borgonovo: estimation des indices d'ordre un et supérieur, et application à l'analyse de fiabilité.

JURY

Fabrice Gamboa	Université Paul Sabatier	Examinateur
Jean-Michel Marin	Université de Montpellier	Président du jury
Amandine Marrel	CEA/DER	Rapporteur
Jérôme Morio	ONERA Toulouse	Co-directeur de thèse
Clémentine Prieur	Université Grenoble Alpes	Rapporteur
Florian Simatos	ISAE-SUPAERO	Co-directeur de thèse

École doctorale et spécialité :
MITT : Domaine Mathématiques : Mathématiques appliquées
Unité de Recherche :
ISAE-ONERA MOIS MOdélisation et Ingénierie des Systèmes
Directeur(s) de Thèse :
Jérôme Morio (ONERA Toulouse) et *Florian Simatos (ISAE-SUPAERO)*
Rapporteurs :
Amandine Marrel (CEA/DER) et *Clémentine Prieur (Université Grenoble Alpes)*

Remerciements

Je tiens à remercier toutes les personnes qui m'ont soutenu pendant ces trois années de thèse et pendant la rédaction du présent manuscrit.

En premier lieu, mes remerciements vont à mes deux directeurs Jérôme Morio et Florian Simatos qui ont accepté d'encadrer mes travaux de thèse. Jérôme, merci de m'avoir fait confiance et de m'avoir donner la possibilité de travailler à l'ONERA pendant ces trois années (et demie en comptant le stage de Master 2). Ta passion contagieuse et ta bienveillance m'ont permis de ressortir confiant et reboosté après chacune de nos discussions. Florian, merci pour tes conseils sur ma méthode de travail et pour ta pédagogie (j'aurais bien aimé avoir eu un professeur de Proba/stats expliquant les choses avec autant d'intuition que toi !). Merci à tous les deux pour votre patience et votre soutient pendant mes moments doute et de démotivation (par ailleurs, je suis pas toujours très expansif mais je tenais à vous dire que vos quelques mots pendant la soutenance m'ont beaucoup ému).

J'adresse mes sincères remerciements à Clémentine Prieur et Amandine Marrel pour avoir accepté d'être rapporteur de ces travaux et de participer au jury. Je remercie également Fabrice Gamboa pour sa participation en tant qu'examinateur et Jean-Michel Marrin en tant que président du Jury. Je vous remercie pour vos commentaires bienveillants et enthousiastes ainsi que pour l'ensemble de vos suggestions.

Sur le plan plus personnel, un gros bisous à ma famille pour leur soutient inconditionnel ! Merci encore à mon papa et à ma moman pour s'être déplacé depuis la Bretagne pour assister à ma soutenance (*j'espère que vous avez compris quelques trucs ... ! Je vous demanderais un résumé à l'occasion aha*). Une pensée également à mon grand-père Rémi parti trop tôt et qui était très fier que son petit-fils fasse une thèse (*Selon lui j'ai permis de redorer le blason familial, je suis sûr que ça en fera rire certains de l'ONERA*). En parlant de l'ONERA tient ! Merci à toutes les personnes de l'ONERA que j'ai pu côtoyer : travailler avec vous à été une aventure humaine géniale, et je m'excuse par avance pour tous ceux que je vais oublier de citer. Merci à tous mes ami(e)s thésard(e)s pour (*Oui ? Hein ? Attendez on me parle dans l'oreillette... Ah oui et les stagiaires bien sûr aha... Hum... Bref*) les moments de détente le midi sous forme de parties de tarot endiablées (au passage désolé d'avoir été **un peu** ~~râgeux~~ râleur). Merci à Gabriel pour nos discussions et pour ton énorme contribution à la réalisation de certains codes et du chapitre 6 de ce manuscrit (si si j'insiste, t'es trop modeste). Merci à Sébastien pour nos nombreuses discussions sur les séries/films et jeux vidéos ! Merci à Rémi P. pour toutes ces pauses clopes et les discussions passionnées qui allaient avec. Ces années à l'ONERA auraient été bien fades sans toi ! Merci à tous mes partenaires de squash : Julien (qui m'a fait découvrir ce sport), mon poto Nicolas F. (*qui est nul au passage ! Une sombre prophétie*

raconte qu'un jour il parviendra à me battre. Mais étant donné que je déménage à Paris je doute sincèrement que cela arrive), Maxime, la miss Morgane, Gaspard et Valentin (et j'en oublie peut être d'autres). Merci à mes deux collègues "CGTistes" Armand et Frank (*pour les stagiaires ou thésard qui me succéderont, demandez à ce dernier comment il va le lundi matin, vous verrez il adore ça*) pour nos longues discussions sur la politique (entre autres) et pour leur "bienveillance" envers moi et plus généralement les "stagiaires de thèse". Merci à Nicolas S. et Rémy (alias mimi parait-il) qui se sont succédé dans mon bureau, et à tous ceux que je n'ai pas encore cité : Damien, François, Lynda (merci pour tes gâteaux :p), Jérémy, Benoit, Heidi, Ludivine, Marie, Pablo, Alex, Alessandro, Nathalie, Sylvain, Antoine, Rémi, Peter, Claire etc. Un dernier coucou à Rémi P. (alias Doremus0) et tous les Last-Mohicans (RIP le TS) pour nos nombreuses parties de R6S et leur "amour indéfectible" envers les Bretons. Enfin, un dernier coucou également à Nicolas F. avec qui j'ai pu partagé plein de moments géniaux : à Hossegor, à Lacanau, au ski, sur Portal 2 (merci en passant pour ton amabilité légendaire envers mes choix de maps), à la Baraka jeu etc.

Pour terminer, un petit coucou à mes deux sœurettes adorées que je rejoindrai à Paris pour la rentrée 2019, ayant vendu mon âme à l'Éducation Nationale ;)

Acronymes

CDF	fonction de répartition (cumulative density function)
PDF	densité de probabilité (probability density function)
i.i.d	indépendantes et identiquement distribuées
ME	maximum d'entropie
CV	coefficient de variation
SMC	Monte-Carlo séquentielle (Sequential Monte Carlo)
TCL	théorème central limite

Table des matières

1 Introduction **14**

2 Propagation des incertitudes dans les modèles boîte noire entrée-sortie **18**
2.1 Modèles boîte noire entrée-sortie 18
2.2 Modélisation des incertitudes 20
2.2.1 Éléments de théorie des probabilités 20
2.2.2 Estimation non-paramétrique de densité 25
2.2.3 Méthodes d'échantillonnage et estimation des moments d'une variable aléatoire 32
2.3 Propagation des incertitudes dans les modèles boîte noire 36
2.4 Analyse de fiabilité des modèles boîte noire 37
2.4.1 L'approximation FORM/SORM 38
2.4.2 Techniques d'échantillonnage préférentiel 40
2.4.3 Méthodes de Monte-Carlo séquentielles 44
2.5 Techniques de décompositions fonctionnelles 46
2.5.1 Décomposition de Hoeffding-Sobol et ANOVA. 47
2.5.2 Cut-HDMR 47

3 Analyse de sensibilité globale basée sur les distributions **49**
3.1 Introduction et motivations 49
3.2 Décomposition fonctionnelle de la variance et indices de Sobol 51
3.3 Indices de sensibilité basés sur des mesures de dissimilarité 53
3.3.1 Indices de sensibilité basés sur des fonctions de contraste 54
3.3.2 Indices de sensibilité basés sur la distance de Cramér Von Mises . . 55
3.3.3 Les mesures de dépendance de Csizár 55
3.4 Les mesures d'importance δ_i de Borgonovo : un état de l'art 57
3.4.1 δ_i vue comme une espérance 58
3.4.2 δ_i vue comme une mesure de dépendance 62
3.4.3 Représentation de δ_i en terme de densité de copule 68
3.5 Conclusion 69

4 Contribution à l'estimation des indices de Borgonovo d'ordre 1 **71**
4.1 Introduction 71
4.1.1 Motivations 71
4.1.2 Présentation des cas tests analytiques 72
4.2 Estimateur $\delta_i^{\mathrm{IS},g}$ combinant échantillonnage préférentiel et approximation par noyau gaussien 74

4.2.1 Retour sur la méthode *simple-boucle* . . . 74
4.2.2 Définition de l'estimateur $\delta_i^{\mathrm{IS},g}$ . . . 76
4.2.3 Étude des performances de $\delta_i^{\mathrm{IS},g}$ . . . 77
4.2.4 Application à des cas test analytiques . . . 86
4.3 Estimateur δ_i^{ME} basé sur la représentation de δ_i en terme de densité de copule et le principe du maximum d'entropie . . . 88
4.3.1 Choix des contraintes . . . 88
4.3.2 Définition de δ_i^{ME} . . . 90
4.3.3 Application à des cas test analytiques . . . 92
4.4 Conclusion : application à un modèle d'estimation de la zone de retombée d'un étage de lanceur . . . 95
4.4.1 Description du cas test . . . 95
4.4.2 Résultats numériques . . . 96
4.4.3 Synthèse . . . 98

5 Application à l'analyse de sensibilité fiabiliste d'un modèle boîte noire 99
5.1 Introduction et motivations . . . 99
5.2 Deux indices de sensibilité complémentaires : $\bar{\eta}_i$ et δ_i^f . . . 101
5.3 Estimation des indices $\bar{\eta}_i$ et δ_i^f . . . 102
5.3.1 Comment échantillonner dans la région de défaillance ? . . . 103
5.3.2 Schéma d'estimation de $\bar{\eta}_i$ et δ_i^f . . . 103
5.4 Applications numériques . . . 105
5.4.1 Retour sur le cas jouet . . . 105
5.4.2 Un cas test analytique . . . 105
5.4.3 Un oscillateur non linéaire à un degré de liberté . . . 106
5.4.4 Le modèle de retombée d'un étage de lanceur . . . 108
5.5 Généralisation : analyse de sensibilité régionale et conditionnelle . . . 109
5.5.1 Une mesure de dissimilarité plus générale . . . 109
5.5.2 Impact sur une fonction de la sortie Y . . . 110
5.5.3 Généralisation . . . 110
5.6 Conclusion . . . 112

6 Étude des indices de Borgonovo d'ordres supérieurs 113
6.1 Introduction et motivations . . . 113
6.2 Modèle *vigne* d'une densité de copule multivariée . . . 115
6.2.1 Introduction aux distributions *vignes* . . . 115
6.2.2 Modèles de R-*vignes* et formulation générale des distributions *vignes* 119
6.2.3 Sélection de la R-*vigne* . . . 122
6.2.4 Application : estimation des indices de Borgonovo d'ordres supérieurs . . . 124
6.3 Valeur de Shapley . . . 125
6.3.1 Définition et propriétés de la valeur de Shapley . . . 125
6.3.2 Application en analyse de sensibilité . . . 126
6.4 Applications numériques . . . 127
6.4.1 Le modèle gaussien (4.2) . . . 128
6.4.2 L'oscillateur non linéaire à un degré de liberté . . . 129
6.4.3 Le modèle de retombée d'un étage de lanceur . . . 130

6.5 Conclusion 132

7 Conclusion et perspectives 133

Résumé

Dans de nombreuses disciplines, un système complexe est modélisé par une fonction *boîte noire* dont le but est de simuler le comportement du système réel. Le système est donc représenté par un modèle entrée-sortie, i.e, une relation entre la sortie Y (ce que l'on observe sur le système) et un ensemble de paramètres extérieurs X_i (représentant typiquement des variables physiques). Ces paramètres sont usuellement supposés aléatoires pour prendre en compte les incertitudes phénoménologiques inhérentes au système. L'*analyse de sensibilité globale* joue alors un rôle majeur dans la gestion de ces incertitudes et dans la compréhension du comportement du système. Cette étude repose sur l'estimation de *mesures d'importance* dont le rôle est d'identifier et de classifier les différentes entrées en fonction de leur influence sur la sortie du modèle.

Les indices de Sobol, dont l'objectif est de quantifier la contribution d'une variable d'entrée (ou d'un groupe de variables) à la variance de la sortie, figurent parmi les mesures d'importance les plus considérées. Néanmoins, la variance est une représentation potentiellement restrictive de la variabilité du modèle de sortie. Le sujet central de cette thèse porte sur une méthode alternative, introduite par Emanuele Borgonovo, et qui est basée sur l'analyse de l'ensemble de la distribution de sortie. Les mesures d'importance de Borgonovo admettent des propriétés très utiles en pratique qui justifient leur récent gain d'intérêt, mais leur estimation constitue un problème complexe. En effet, la définition initiale des indices de Borgonovo fait intervenir les densités inconditionnelle et conditionnelles de la sortie du modèle, malheureusement inconnues en pratique. Dès lors, les premières méthodes proposées menaient à un budget de simulation élevé, la fonction boite noire pouvant être très coûteuse à évaluer.

La première contribution de cette thèse consiste à proposer de nouvelles méthodologies pour estimer les mesures d'importance de Borgonovo du premier ordre, i.e, les indices mesurant l'influence de la sortie Y relativement à une entrée X_i scalaire. Dans un premier temps, nous choisissons d'adopter la réinterprétation des indices de Borgonovo en terme de mesure de dépendance, i.e, comme une distance entre la densité jointe de X_i et Y et la distribution produit. En outre, nous développons une procédure d'estimation combinant échantillonnage préférentiel et approximation par noyau gaussien de la densité de sortie et de la densité jointe. Cette approche permet de calculer l'ensemble des indices de Borgonovo d'ordre 1, et ce, avec un faible budget de simulation indépendant de la dimension du modèle. Cependant, l'utilisation de l'estimation par noyau gaussien peut fournir des estimations imprécises dans le cas des distributions à queue lourde. Pour pallier ce problème, nous nous appuyons dans un second temps sur une autre définition des indices de Borgonovo reposant sur le formalisme des copules. Cette démarche permet de se soustraire au cas des distributions à queue lourde puisqu'elle repose sur l'estimation d'une densité de copule bivariée à support borné. Par ailleurs, l'estimation par noyau gaussien pouvant

souffrir d'effets de bord, nous choisissons d'estimer cette densité de copule via la méthode du maximum d'entropie. Les deux méthodologies proposées sont appliquées et comparées sur un exemple concret aérospatial.

La seconde contribution de cette thèse s'inscrit dans le contexte de l'analyse de sensibilité fiabiliste. Là où l'analyse de sensibilité classique se focalise sur la sortie du modèle, l'approche fiabiliste s'intéresse au système dans un mode de fonctionnement critique, typiquement une défaillance associée à un état non sécurisé et donc indésirable. On distingue deux méthodes complémentaires d'analyse de sensibilité fiabilistes : celles visant à étudier l'impact d'une variable d'entrée sur une fonction de la sortie, telle que la fonction indicatrice du domaine critique, et celles visant à étudier l'influence d'une entrée conditionnellement à l'évènement de défaillance. Dans cette partie, nous définissons deux indices de sensibilité fiabilistes liés à ces méthodes et intrinsèquement liés aux indices de Borgonovo traditionnels. Nous montrons que ces deux mesures peuvent être simultanément estimées à partir d'un échantillon commun donné par exemple par un algorithme de Monte-Carlo séquentiel, un des plus utilisés dans le cadre de l'analyse de fiabilité.

Enfin, la dernière contribution de cette thèse concerne l'estimation des indices de Borgonovo d'ordres supérieurs, i.e, affiliés à un groupe de paramètres d'entrée, ce qui rentre dans le cadre de l'estimation de densité multidimensionnelle. Cette estimation peut théoriquement être réalisée avec les méthodes mentionnées précédemment. Toutefois, celles-ci souffrent de problèmes numériques lorsque la dimension du problème augmente. Cet effet néfaste de la dimension est connu sous le nom de *malédiction de la dimension*. Pour autant, nous montrons que ce problème peut être surmonté grâce à la théorie des copules *vignes*. Par ailleurs, il semble nécessaire de synthétiser l'information fournie par l'ensemble des indices de Borgonovo, leur nombre grandissant rapidement avec la dimension du modèle. Pour répondre à ce besoin, nous considérons la notion de *valeur de Shapley* initialement introduite en théorie des jeux. Nous montrons que cela permet de définir des indices de sensibilité qui facilitent l'interprétation des indices de Borgonovo d'ordres supérieurs et qui peuvent être calculés par simple post-traitement une fois l'ensemble des indices de Borgonovo estimés.

Mots-clés : Analyse de sensibilité globale, Mesures d'importance de Borgonovo, Estimation de densité, Analyse de fiabilité, Théorie des copules.

Abstract

In many disciplines, a complex system is modeled by a black box function whose purpose is to mimic the behavior of the real system. Then, the system is represented by an input-output model, i.e, a relationship between the output Y (the observation made on the system) and a set of external parameters X_i (typically representing physical variables). These parameters are usually assumed to be random in order to take phenomenological uncertainties into account. Global sensitivity analysis (GSA) plays a crucial role in the handling of these uncertainties and in the understanding of the system behavior. This study is based on the estimation of importance measures which aim at identifying and ranking the different inputs with respect to their influence on the model output. Variance-based sensitivity indices are one of the most widely used GSA measures. They are based on Sobol's indices which express the share of variance of the output that is due to a given input or input combination. However, by definition they only study the impact on the second-order moment of the output which may be a restrictive representation of the whole output distribution. The central subject of this thesis is an alternative method, introduced by Emanuele Borgonovo, which is based on the analysis of the whole output distribution. Borgonovo's importance measures present very convenient properties that justify their recent gain of interest, but their estimation is a challenging task. Indeed, the initial definition of the Borgonovo's indices involves the unconditional and conditional densities of the model output, which are unfortunately unknown in practice. Thus, the first proposed methods led to a high computational burden especially since the black box function may be very costly-to-evaluate.

The first contribution of this thesis consists in proposing new methodologies for estimating first order Borgonovo's importance measures which quantify the influence of the output Y relatively to a scalar input X_i. First, we choose to adopt the reinterpretation of the Borgonovo's indices in term of dependence measure, i.e, as a distance between the joint density of X_i and Y and the product distribution. In addition, we develop an estimation procedure combining an Importance Sampling procedure and Gaussian kernel approximation of the output density and the joint density. This approach allows the computation of all first order Borgonovo's indices with a low budget simulation, independent of the model dimension. However, the use of Gaussian kernel estimation may provide inaccurate estimates for heavy tail distributions. To overcome this problem, we consider an alternative definition of the Borgonovo's indices based on the copula formalism. This approach makes it possible to avoid the case of heavy-tailed distributions since it relies on the estimation of a bounded bivariate copula density. Moreover, since Gaussian kernel estimation may suffer from bound effects, we choose to estimate this copula density using a maximum entropy method under fractional moment constraints. The two proposed methodologies are applied and compared on a concrete aerospace example.

The second contribution of this thesis lies in the reliability sensitivity analysis context. Classical sensitivity analysis focuses on the model output whereas the reliability approach is interested in the system behavior under a critical operating mode, typically a failure associated with an unsafe and so undesirable state of the system. Two complementary reliability sensitivity analysis methods can be distinguished : those which aim at studying the impact of an input variable on a function of the model output (typically the indicator function of the critical domain), and those which aim at studying the influence of an input conditionally on the failure event. We define two distinct reliability indices intrinsically linked to traditional Borgonovo's indices. We show that these two measures are complementary and may be simultaneously estimated from a common sample distributed with respect to the law of the input conditioned on the system failure.

Finally, the last contribution of the thesis concerns the estimation of high order Borgonovo's indices associated to a group of input parameters, which belongs the multidimensional density estimation framework. This estimation can theoretically be performed with the previously mentioned methods. However, they suffer from numerical problems when the dimension of the problem increases. This deleterious effect of the dimension is known as the curse of the dimension. Nevertheless, we show that this issue may be overcome by using the theory of vine copulas. Moreover, it seems necessary to synthesize the information provided by all the Borgonovo's indices since their number rapidly grow with the model dimension. To answer this need, we consider the notion of Shapley value initially introduced in game theory. We show that this concept allows the introduction of new sensitivity indices which facilitate the interpretation of Borgonovo's indices and which can be computed by a simple post-processing once the set of Borgonovo's indices are estimated.

Keywords : Global sensitivity analysis, Borgonovo's importance measures, Density estimation, Reliability analysis, Copula theory.

Chapitre 1

Introduction

Contexte

Dans de nombreuses disciplines, que ce soit en physique des particules, en biologie cellulaire, en économie ou encore dans l'étude des mouvements collectifs, un système est usuellement modélisé par une fonction *boîte noire* $\mathcal{M}(\cdot)$, représentant typiquement un code de calcul dont le but est de simuler le comportement du système réel. Ce dernier est donc représenté par un modèle entrée-sortie, i.e, une relation entre la sortie Y (ce que l'on observe sur le système) et un ensemble de paramètres extérieurs X_i (représentant typiquement des variables physiques).

L'étude d'un tel modèle passe par la prise en compte des incertitudes qui peuvent affecter l'appréciation du comportement du système réel, celles-ci pouvant être dues à un aléa intrinsèque au système, à des erreurs de mesure ou encore au modèle lui-même (précision limitée des calculs numériques intervenant dans le code $\mathcal{M}(\cdot)$). Ceci rentre dans le cadre de la *quantification des incertitudes* dont le but est de modéliser les sources d'incertitudes et étudier leur propagation vers la sortie du modèle. La méthodologie de la quantification des incertitudes peut être résumée en trois étapes, comme illustré dans la figure [1.1].

- **Modélisation des incertitudes.** La première étape consiste à identifier et quantifier les sources d'incertitudes qui affectent les paramètres d'entrée. Il peut s'agir de variables stochastiques présentant une variabilité naturelle résultant de phénomènes aléatoires, ou de variables épistémiques liées à un manque de connaissance. Cette incertitude peut être modélisée par exemple à l'aide de la théorie probabiliste moderne ou le formalisme des probabilités imprécises. Le choix de l'approche dépend d'un avis d'expert et de la qualité des données disponibles.

- **Propagation des incertitudes.** Par propagation des incertitudes au travers du code $\mathcal{M}(\cdot)$, la sortie Y est également entachée d'incertitudes. Plusieurs quantités d'intérêt peuvent alors être considérées pour étudier le comportement statistique de Y ou réaliser une étude de risque associée à un état non sécurisé du système, souvent défini comme le dépassement par la sortie Y d'un seuil critique S.

- **Analyse de sensibilité.** La dernière étape consiste à expliquer la variabilité de la sortie Y en fonction des X_i. Elle joue un rôle crucial dans la compréhension

du comportement du système et dans la réduction de l'incertitude du modèle en identifiant les variables les plus influentes. Il existe dans la littérature de nombreuses méthodes d'analyse de sensibilité, chacune étant fondée sur un critère d'influence qui lui est propre. Ces méthodes peuvent également s'inscrire dans un contexte fiabiliste pour lequel l'étude du modèle de sortie est remplacée par celle d'une mesure de fiabilité, typiquement la probabilité de défaillance du système, ce qui rentre dans le cadre de l'étude des évènements rares.

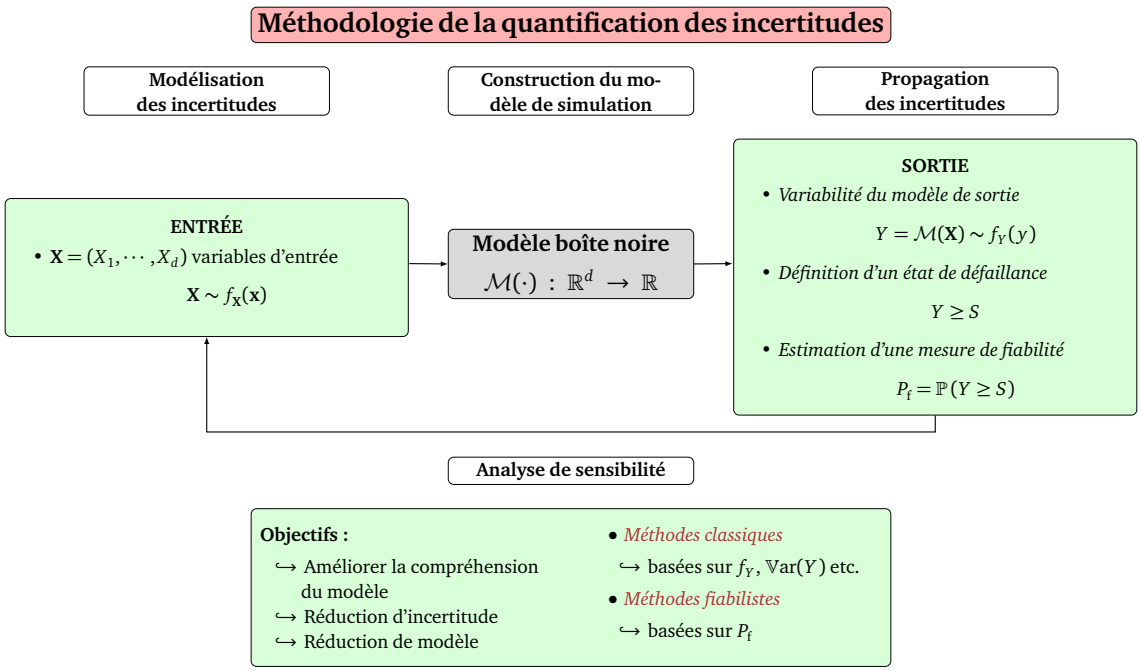


FIGURE 1.1. SCHÉMA D'ILLUSTRATION DE LA MÉTHODOLOGIE DE LA QUANTIFICATION DES INCERTITUDES D'UN MODÈLE BOÎTE NOIRE ENTRÉE-SORTIE [CHABRIDON, 2018].

Problématique et objectifs de la thèse

Le sujet qui est au cœur de cette thèse est l'étape d'analyse de sensibilité, et plus particulièrement l'analyse de sensibilité globale qui repose sur l'estimation de mesures d'importance dont le rôle est d'identifier et de classifier les différentes entrées en fonction de leur influence sur la sortie du modèle. Les indices de Sobol [Sobol, 1993], dont l'objectif est de quantifier la contribution d'une variable d'entrée (ou d'un groupe de variables) à la variance de la sortie, figurent parmi les mesures d'importance les plus considérées. Néanmoins, la variance est une représentation potentiellement restrictive de la variabilité du modèle de sortie. Ces travaux de thèse se concentrent sur une méthode alternative, introduite par Emanuele Borgonovo [Borgonovo, 2007], basée sur l'analyse de l'ensemble de la distribution de sortie. Les mesures d'importance de Borgonovo admettent des propriétés très utiles en pratique qui justifient leur récent gain d'intérêt, mais leur estimation constitue un problème complexe. En effet, la définition initiale des indices de Borgonovo fait

intervenir la densité inconditionnelle et les densités conditionnelles de la sortie du modèle, malheureusement inconnues en pratique. Ainsi, les premières méthodes proposées pour estimer les indices de Borgonovo menaient à un budget de simulation élevé, la fonction boîte noire $\mathcal{M}(\cdot)$ pouvant être très coûteuse à évaluer. Dès lors, ce problème d'estimation a donné lieu à d'abondants travaux de recherche, le but étant d'optimiser le nombre d'appels à la boîte noire $\mathcal{M}(\cdot)$ tout en assurant une estimation suffisamment précise des indices pour obtenir une classification reflétant correctement la réalité du système étudié. En outre, l'objectif principal de cette thèse est de **proposer de nouvelles méthodes d'estimation des indices d'ordre un et supérieur** ainsi que de **clarifier l'information apportée par l'ensemble des mesures d'importance de Borgonovo**. Le second objectif concerne l'**adaptation de la méthode d'analyse de sensibilité de Borgonovo dans un contexte fiabiliste** dans le but d'étudier l'influence d'un paramètre d'entrée spécifiquement dans un régime jugé critique du système.

Structure du mémoire

Ce manuscrit de thèse est composé des cinq chapitres suivants :

Le chapitre 2 introduit les notations et les outils de la théorie des probabilités utilisés tout au long du manuscrit pour modéliser les incertitudes affectant les paramètres d'entrées d'un modèle boîte noire. Il aborde également les différentes techniques relatives à la phase de propagation des incertitudes, notamment l'estimation de densités de probabilité et de probabilités d'évènements rares.

Le chapitre 3 est une introduction à l'analyse de sensibilité. Une attention toute particulière est portée sur la présentation des méthodes globales quantitatives, et plus spécifiquement sur la classe des indices de sensibilité basés sur une mesure de dissimilarité à laquelle appartiennent les mesures d'importance de Borgonovo. Les différentes propriétés et définitions de ces indices de sensibilité sont présentées. Enfin, un état de l'art des différentes méthodes d'estimation introduites à ce jour est effectué.

Le chapitre 4 présente deux nouvelles approches d'estimation des indices de Borgonovo d'ordre un, i.e, les indices mesurant l'influence de la sortie Y relativement à une entrée X_i scalaire. La première est basée sur la réinterprétation des indices de Borgonovo en terme de mesure de dépendance, i.e, comme une distance entre la densité jointe de X_i et Y et la distribution produit. Elle est définie par une procédure d'estimation combinant échantillonnage préférentiel et approximation par noyau gaussien de la densité de sortie et de la densité jointe. La seconde méthode repose sur le formalisme des copules et sur l'estimation de la densité de copule via le principe du maximum d'entropie. Ces deux approches permettent de calculer l'ensemble des indices de Borgonovo d'ordre un, et ce, avec un faible budget de simulation indépendant de la dimension du modèle $\mathcal{M}$. Les deux méthodologies proposées sont appliquées et comparées sur un exemple concret aérospatial.

Le chapitre 5 s'inscrit dans le contexte de l'analyse de sensibilité fiabiliste. Là où l'analyse de sensibilité classique se focalise sur la sortie du modèle, l'approche fiabiliste

s'intéresse au système dans un mode de fonctionnement critique, typiquement une défaillance associée à un état non sécurisé et donc indésirable. Deux méthodes complémentaires d'analyse de sensibilité fiabilistes peuvent être distinguées : celles visant à étudier l'impact d'une variable d'entrée sur une fonction de la sortie, telle que la fonction indicatrice du domaine critique, et celles visant à étudier l'influence d'une entrée, conditionnellement à l'évènement de défaillance. En outre, ce chapitre définit deux indices de sensibilité fiabilistes liés à ces méthodes et intrinsèquement liés aux indices de Borgonovo traditionnels. Il est montré que ces deux mesures peuvent être simultanément estimées à partir d'un échantillon commun donné par exemple par un algorithme de Monte-Carlo séquentiel, un des plus utilisés dans le cadre de l'analyse de fiabilité.

Le chapitre 6 est constitué de deux parties. La première aborde la question de l'estimation des indices de Borgonovo d'ordres supérieurs, i.e, affiliés à un groupe de paramètres d'entrée, ce qui rentre dans le cadre de l'estimation de densité multidimensionnelle. Cette estimation peut théoriquement être réalisée avec les méthodes présentées dans les chapitres précédents. Toutefois, celles-ci souffrent de problèmes numériques lorsque la dimension du problème augmente. Pour surmonter cette difficulté, il est proposé d'utiliser la théorie des copules *vignes*. La seconde partie s'intéresse au problème de l'interprétation de l'information fournie par l'ensemble des indices de Borgonovo, leur nombre grandissant rapidement avec la dimension du modèle. La notion de *valeur de Shapley*, initialement introduite en théorie des jeux, est considérée dans ce chapitre pour répondre à ce besoin. De cela découle la définition d' indices de sensibilité qui permettent de clarifier l'interprétation des indices de Borgonovo d'ordre supérieur et qui peuvent être calculés par simple post-traitement une fois l'ensemble des indices de Borgonovo estimés.

Chapitre 2

Propagation des incertitudes dans les modèles boîte noire entrée-sortie

Contents

2.1 Modèles boîte noire entrée-sortie 18
2.2 Modélisation des incertitudes 20
2.2.1 Éléments de théorie des probabilités 20
2.2.2 Estimation non-paramétrique de densité 25
2.2.3 Méthodes d'échantillonnage et estimation des moments d'une variable aléatoire 32
2.3 Propagation des incertitudes dans les modèles boîte noire . 36
2.4 Analyse de fiabilité des modèles boîte noire 37
2.4.1 L'approximation FORM/SORM 38
2.4.2 Techniques d'échantillonnage préférentiel 40
2.4.3 Méthodes de Monte-Carlo séquentielles 44
2.5 Techniques de décompositions fonctionnelles 46
2.5.1 Décomposition de Hoeffding-Sobol et ANOVA. 47
2.5.2 Cut-HDMR 47

2.1 Modèles boîte noire entrée-sortie

Dans de nombreuses disciplines, la modélisation d'un système complexe est réalisée en construisant une fonction $\mathcal{M}(\cdot)$ dont le but est de simuler le comportement du système réel. Le système est donc représenté par un modèle entrée-sortie, i.e, une relation (définie par $\mathcal{M}(\cdot)$) entre la sortie (ce que l'on observe sur le système) et un ensemble de paramètres extérieurs (représentant typiquement des variables physiques). La fonction $\mathcal{M}(\cdot)$ peut être donnée par une formule analytique ou par un code de calcul potentiellement coûteux à évaluer. On peut notamment penser à un code de type éléments finis, dont la complexité rend impossible toute étude analytique de la fonction et donc de la sortie. Dans ce manuscrit, $\mathcal{M}(\cdot)$ sera supposée déterministe et considérée comme une boîte noire.

En outre, les différentes études mentionnées dans la suites sont *non intrusives*, dans le sens où aucune connaissance approfondie du code de calcul n'est requise.

En guise d'exemple introductif, considérons l'étude de la zone de retombée d'un étage de lanceur. La complexité des lanceurs spatiaux découle du couplage entre plusieurs sous systèmes, tels que les propulseurs ou la structure des différents étages qui l'équipent. Pendant le vol, des incertitudes aléatoires (par exemple dues à des perturbations météorologiques) peuvent affecter le bon déroulement des différentes phases du vol et compromettre la mission. En effet, ces incertitudes combinées peuvent mener à une trajectoire indésirée car potentiellement dramatique en terme de sécurité et d'impact environnemental (voir figure [2.1]).

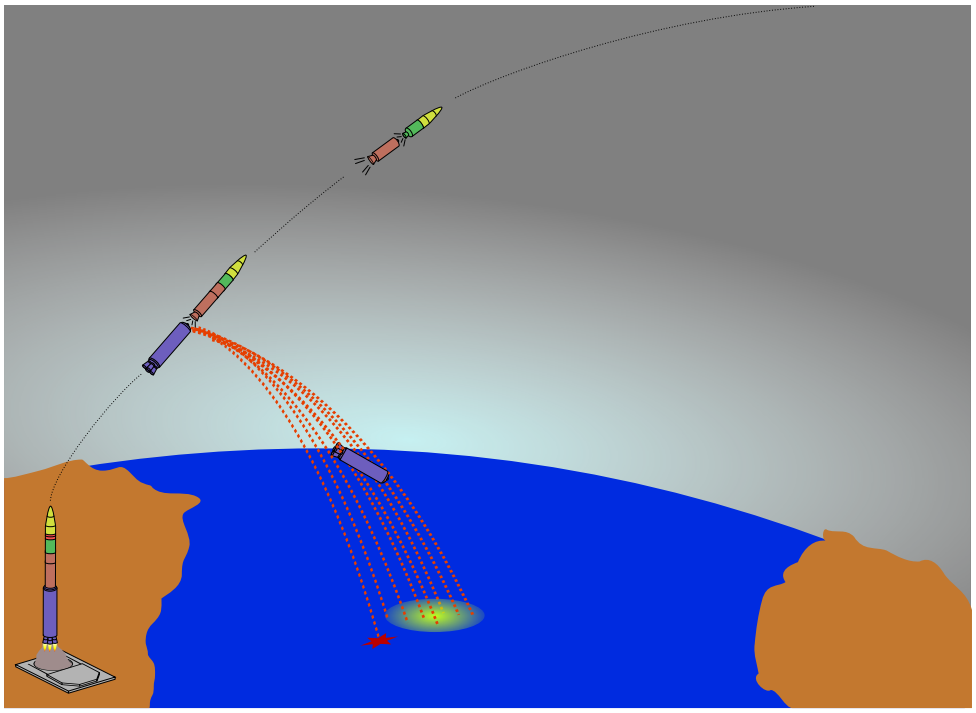


Figure 2.1. Schéma d'illustration de l'étape de retombée d'un étage de lanceur dans l'Océan Atlantique. Plusieurs trajectoire sont tracées (lignes rouges en pointillés), menant à une zone de sécurité (surface jaune). À cause des incertitudes, le lanceur peut retomber hors de la zone souhaitée (étoile rouge) [Derennes et al., 2019a].

L'étude d'un modèle boîte noire entrée-sortie comporte ainsi plusieurs étapes majeures. Tout d'abord, il convient de prendre en compte les incertitudes affectant les paramètres d'entrée, celles-ci pouvant être dues à des erreurs de mesure ou à un aléa intrinsèque au système réel. La modélisation de ces incertitudes est discutée dans la section 2.2. Vient ensuite la phase de propagation des incertitudes dont le but est d'étudier les répercussions des incertitudes sur le comportement de la sortie du modèle. Ceci inclut en particulier l'analyse de fiabilité du modèle, i.e, l'étude du régime de défaillance du système, correspondant à un état non sécurisé et donc indésirable. Cette question sera abordée dans la section 2.4. Enfin, une autre question fondamentale qui se pose en pratique est l'identification et le classement des différentes entrées en fonction de leur influence sur la sortie du modèle. Cette étude est communément appelée analyse de sensibilité et sera détaillée dans le chapitre 3.

2.2 Modélisation des incertitudes

Les incertitudes sont traditionnellement modélisées par une distribution de probabilité, i.e, via le cadre probabiliste classique dont la théorie moderne à été axiomatisée par Andreï Kolmogorov. La théorie probabiliste suppose que tout état d'incertitude peut se modéliser par une probabilité. Cependant, en pratique il n'est pas toujours possible d'attribuer une probabilité à tous les états d'entrée, notamment lorsque les données disponibles sont lacunaires, non fiables ou en faible nombre. Dans ce cas, des alternatives existent comme le formalisme des probabilités imprécises (voir par exemple [Augustin et al., 2014]) qui offre de nouveaux outils permettant de surmonter ce manque d'information.

Dans cette thèse, on suppose que l'information disponible est suffisante pour construire un modèle de probabilité pour les entrées. En outre, les prochaines sous-sections introduisent les notations et les outils de la théorie des probabilités utilisés tout au long de ce manuscrit.

2.2.1 Éléments de théorie des probabilités

Considérons un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, i.e, un triplet constitué d'un ensemble Ω muni d'une tribu $\mathcal{A}$ et d'une mesure de probabilité $\mathbb{P} : \mathcal{A} \to [0, 1]$. On appelle vecteur aléatoire de dimension d toute application (mesurable) de $(\Omega, \mathcal{A}, \mathbb{P})$ dans l'espace mesurable $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ où $\mathcal{B}(\mathbb{R}^d)$ désigne la tribu borélienne de $\mathbb{R}^d$. Dans toute la suite, on adoptera la convention typographique suivante : les variables à une dimension ($d = 1$) seront écrites en caractère normal, et les variables de dimensions supérieures ($d \geq 2$) seront écrites en caractère gras.

2.2.1.1 Distribution d'une variable aléatoire

Soit $\mathbf{X} = (X_1, \ldots, X_d)$ un vecteur aléatoire de dimension d. On appelle loi de $\mathbf{X}$ la mesure $\mathbb{P}_{\mathbf{X}}$ sur $\mathbb{R}^d$ définie par

$$\forall A \in \mathcal{B}(\mathbb{R}^d),\ \mathbb{P}_{\mathbf{X}}(A) = \mathbb{P}(\mathbf{X} \in A)\ .$$

On dit que la mesure $\mathbb{P}_{\mathbf{X}}$ est *absolument continue* par rapport à une mesure η si et seulement si pour tout borélien A tel que $\eta(A) = 0$, alors $\mathbb{P}_{\mathbf{X}}(A) = 0$. Rappelons que lorsque la mesure dominante η est σ-finie, i.e, lorsqu'il existe un recouvrement dénombrable de $\mathbb{R}^d$ par des sous-ensembles A tels que $\eta(A) < +\infty$, le théorème de Radon-Nikodym entraine l'existence d'une fonction $\dfrac{d\mathbb{P}_{\mathbf{X}}}{d\eta}$ positive et η-intégrable, appelée *dérivée de Radon-Nikodym*, telle que

$$\forall A \in \mathcal{B}(\mathbb{R}^d),\ \mathbb{P}_X(A) = \int_A \frac{d\mathbb{P}_{\mathbf{X}}}{d\eta} d\eta\ . \tag{2.1}$$

Dans cette thèse, les variables rencontrées seront exclusivement discrètes ou à densité, i.e, absolument continues par rapport à la mesure de comptage ou la mesure de Lebesgue sur $\mathbb{R}^d$ respectivement. Par ailleurs, le cas à densité sera appelé abusivement *cas absolument continu* dans le reste du manuscrit. La variable $\mathbf{X}$ peut être caractérisée de différentes façons :

- par sa loi $\mathbb{P}_{\mathbf{X}}$,
- par sa fonction de répartition (cumulative density function) (CDF) $F_{\mathbf{X}} : \mathbb{R}^d \to [0,1]$ définie par

$$F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{X}}(x_1, \ldots, x_d) = \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d) \ ,$$

- ou encore (dans le cas absolument continu) par sa densité de probabilité (probability density function) (PDF) $f_{\mathbf{X}} : \mathbb{R}^d \to [0, +\infty[$ définie par

$$\forall A \in \mathcal{B}(\mathbb{R}^d), \ \mathbb{P}_{\mathbf{X}}(A) = \int_A f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \ .$$

  La fonction $f_{\mathbf{X}}$ vérifie en particulier $\int_{\mathbb{R}^d} f_{\mathbf{X}} = 1 < +\infty$ ce que l'on note $f_{\mathbf{X}} \in L^1(\mathbb{R}^d, \mathbb{R}^+)$. D'une manière générale, l'ensemble $L^p(\mathbb{R}^d, \mathbb{R})$ désigne dans la suite l'ensemble des fonctions $f : \mathbb{R}^d \to \mathbb{R}$ vérifiant $\int_{\mathbb{R}^d} \|f\|^p < +\infty$ où $\|\cdot\|$ est une mesure quelconque sur $\mathbb{R}^d$.

Pour terminer, rappelons que la distribution d'une variable aléatoire X est dite *à queue lourde* si

$$\forall \lambda > 0, \ \lim_{x \to +\infty} e^{\lambda x}(1 - F_X(x)) = \infty \ .$$

2.2.1.2 Distance statistique

La notion de *distance statistique* sert à mesurer l'écart entre deux lois ou mesures de probabilité. Cette section ne prétend pas à l'exhaustivité et a seulement pour but de rappeler le formalisme de distance statistique et donner quelques exemples qui interviendront tout au long du manuscrit. Notons $\mathcal{P}^d$ l'ensemble des lois de probabilité définies sur $\mathbb{R}^d$. On adopte dans la suite de ce manuscrit le formalisme suivant :

Définition : Soit $d(\cdot, \cdot)$ une application définie sur $\mathcal{P}^d \times \mathcal{P}^d$ et à valeur dans $\mathbb{R}$. On dit que :

- $d(\cdot, \cdot)$ est une **mesure de dissimilarité** si :

$$\forall P, Q \in \mathcal{P}^d, \ P = Q \Rightarrow d(P, Q) = 0 \ .$$

- $d(\cdot, \cdot)$ est une **divergence** si elle vérifie de plus que

$$\forall P, Q \in \mathcal{P}^d, \ d(P, Q) = 0 \Leftrightarrow P = Q \ .$$

- $d(\cdot, \cdot)$ est une **distance** si elle vérifie de plus les propriétés de symétrie et d'inégalité triangulaire.

$$\forall P, Q \in \mathcal{P}^d, \ d(P, Q) = d(Q, P) \ ,$$
$$\forall P, Q, Z \in \mathcal{P}^d, \ d(P, Q) \leq d(P, Z) + d(Z, Q) \ .$$

Exemples : Soient $P, Q \in \mathcal{P}^d$. Voici deux exemples classiques de distances statistiques entre P et Q. D'autres exemples seront rencontrés dans le chapitre 3.

- **Distance en variation totale**

$$d_{\mathrm{TV}}(P,Q) = \operatorname*{Sup}_{A\in\mathcal{B}(\mathbb{R}^d)} |P(A) - Q(A)| \ . \tag{2.2}$$

Lorsque P et Q sont absolument continues par rapports à la mesure de Lebesgue de PDF respective f et g, la distance en variation totale d_{TV} peut se réécrire comme suit en vertu du Lemme de Scheffé [Ouvrard, 2000] :

$$d_{\mathrm{TV}}(P,Q) = \frac{1}{2}\|f-g\|_{L^1(\mathbb{R}^d)} \ .$$

- **Divergence de Kullback-Leibler**

Supposons quoi P et Q sont absolument continues par rapports à la mesure de Lebesgue de PDF respective f et g. La divergence de Kullback-Leibler [Kullback and Leibler, 1951] de P par rapport à Q est définie par

$$d_{\mathrm{KL}}(P|Q) := d_{\mathrm{KL}}(f|g) = \int f \log\left(\frac{f}{g}\right) \ . \tag{2.3}$$

Remarque : L'*inégalité de Pinsker* établit un lien entre la distance en variation totale et la divergence de Kullback-Leibler :

$$\|f-g\|_{L^1} \leq \sqrt{2 d_{\mathrm{KL}}(f|g)} \ . \tag{2.4}$$

2.2.1.3 Lois conditionnelles

Une application ν de $\mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d)$ dans $[0,1]$ est appelée *noyau de transition* (ou *noyau de Markov*) de $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ vers $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ si elle satisfait les deux propriétés suivantes :

- pour tout $\mathbf{x} \in \mathbb{R}^d$, l'application $\nu(\mathbf{x}, \cdot)$ est une probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$,
- pour tout $B \in \mathcal{B}(\mathbb{R}^d)$, l'application $\nu(\cdot, B)$ est $\mathcal{B}(\mathbb{R}^d)$-mesurable.

Soient $\mathbf{X}$ et $\mathbf{X}'$ deux variables aléatoires. On appelle *loi conditionnelle de* $\mathbf{X}$ *sachant* $\mathbf{X}'$ le noyau de probabilité ν tel que la loi jointe de $\mathbf{X}$ et $\mathbf{X}'$ vérifie

$$\mathbb{P}_{\mathbf{X},\mathbf{X}'} = \mathbb{P}_{\mathbf{X}} \cdot \nu \ ,$$

avec

$$\forall A, B \in \mathcal{B}(\mathbb{R}^d), \ (\mathbb{P}_{\mathbf{X}} \cdot \nu)(A \times B) = \int_A \nu(\mathbf{x}, B) d\mathbf{x} \ .$$

L'existence d'une loi conditionnelle est assurée par le théorème de Jirina dans le cadre de la définition précédente (voir par exemple [Ouvrard, 2000]). Dans toute la suite, le noyau de transition ν sera noté $\mathbb{P}_{\mathbf{X}|\mathbf{X}'}$ et $\mathbf{X}|\mathbf{X}'$ désignera une variable aléatoire dont la loi (aléatoire) est une loi conditionnelle de $\mathbf{X}$ sachant $\mathbf{X}'$.

2.2.1.4 Moments d'une variable aléatoire

Les moments d'une variable aléatoire X, quand ils existent, sont des paramètres qui donnent des renseignements sur la loi de cette variable aléatoire. La quantité

$$\mathbb{E}[X^\alpha] := \int_{\mathbb{R}} x^\alpha d\mathbb{P}_X(x),\ \alpha \in \mathbb{N}\ ,$$

est appelée *moment d'ordre* α de X et la quantité $\mathbb{E}[(X - \mathbb{E}[X])^\alpha]$ est appelée moment centré d'ordre α de X. En particulier, le moment d'ordre 1 (parfois noté μ_X) est appelé moyenne (ou espérance) de X et le moment centré d'ordre 2 est appelé *variance* de X et noté $\mathbb{V}\mathrm{ar}(X)$. On définit également l'*écart-type* par $\sigma_X = \sqrt{\mathbb{V}\mathrm{ar}(X)}$ et le coefficient de variation (CV) par le ratio entre la moyenne de X et son écart-type :

$$\mathrm{CV}(X) = \frac{\sigma_X}{\mu_X}\ .$$

Lorsque $X \geq 0$, on peut également définir $\mathbb{E}[X^\alpha]$ lorsque α n'est pas un entier naturel, on parle alors de *moment fractionnaire*.

Dans le cas absolument continu, le théorème de transfert [Ouvrard, 2000] donne une expression plus calculatoire des moments :

$$\mathbb{E}[X^\alpha] = \int_{\mathbb{R}} x^\alpha f_X(x)dx\ . \tag{2.5}$$

On rappelle également les définitions de la moyenne et de la matrice de covariance d'un vecteur aléatoire $\mathbf{X}$:

$$\mathbb{E}[\mathbf{X}] := (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d]) \in \mathbb{R}^d\ ,$$

$$\mathbf{\Sigma_X} = \begin{pmatrix} \mathrm{cov}(X_1, X_1) & \cdots & \mathrm{cov}(X_1, X_d) \\ \vdots & & \vdots \\ \mathrm{cov}(X_d, X_1) & \cdots & \mathrm{cov}(X_d, X_d) \end{pmatrix} \in M_d(\mathbb{R})\ ,$$

où $\mathrm{cov}(X_i, X_j) = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]$ désigne la *covariance* entre les variables marginales X_i et X_j et où M_d désigne l'ensemble des matrices carrées de taille $d \times d$.

2.2.1.5 Exemples de lois de probabilité :

- **Loi uniforme $\mathcal{U}(A)$, $A \in \mathcal{B}(\mathbb{R}^d)$.**

Soit A une partie borélienne de $\mathbb{R}^d$ de mesure de Lebesgue $\lambda(A)$ finie et strictement positive. Un vecteur aléatoire $\mathbf{X} \sim \mathcal{U}(A)$ suit la loi uniforme sur A si sa PDF est définie par

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\lambda(A)}\mathbb{1}_A(\mathbf{x})\ .$$

Un cas particulier largement utilisé est la loi uniforme $\mathcal{U}([0,1])$. Cette distribution intervient dans le résultat fondamental suivant (utilisé notamment pour la simulation de lois de probabilités) :

Proposition 2.1. *Soit X une variable aléatoire et $U \sim \mathcal{U}([0,1])$.*

1. *$F_X(X) \sim \mathcal{U}([0,1])$.*
2. *Soit $G(u) = inf\ \{t \in \mathbb{R}, F_X(t) \geq u\}$ la fonction quantile de X. Alors $G(U)$ a pour fonction de répartition F_X. En particulier, lorsque F_X est continue et strictement croissante, $F_X^{-1}(U)$ suit la même loi que X.*

• **Loi de Gauss (ou normale)** $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$.

Une variable aléatoire réelle $X \sim N(\mu, \sigma^2)$ suit une loi de Gauss (ou normale) de moyenne μ et de variance σ^2 si sa PDF est définie par

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) .$$

Un vecteur aléatoire $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ suit une loi de Gauss (ou normale) multidimensionnelle de moyenne $\boldsymbol{\mu} \in \mathbb{R}^d$ et de matrice de covariance $\boldsymbol{\Sigma} \in M_d(\mathbb{R})$ (supposée inversible) si sa PDF est définie par

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{\det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})^T\right) .$$

• **Loi lognormale** $L(\mu, \sigma^2)$**.**

Une variable aléatoire réelle strictement positive $X \sim L(\mu, \sigma^2)$ suit une loi lognormale de paramètres μ et σ^2 si $\ln(X) \sim N(\mu, \sigma^2)$.

2.2.1.6 Indépendance de variables aléatoires

On termine cette section par quelques rappels sur la notion d'indépendance de variables aléatoires. Lorsque des variables aléatoires $X_1, X_2, \ldots, X_d$ sont indépendantes, la distribution du vecteur joint $\mathbf{X} = (X_1, \ldots, X_d)$ est donnée par le produit des marginales, i.e,

$$F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{X}}(x_1, \ldots, x_d) = \prod_{i=1}^{d} F_{X_i}(x_i) ,$$

ce qui se traduit également dans le cas absolument continu par l'égalité entre la densité jointe et le produit des densités marginales :

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(x_1, \ldots, x_d) = \prod_{i=1}^{d} f_{X_i}(x_i) .$$

Dans le cas non indépendant, les lois marginales ne suffisent plus à décrire intégralement la loi jointe. En revanche, celle-ci est entièrement caractérisée par la donnée des lois marginales et d'une *copule*. La notion de copule est un outil permettant de modéliser la structure de dépendance entre des variables aléatoires. Plus formellement, une copule d-dimensionnelle est une CDF d'une variable aléatoire dont les marginales unidimensionnelles sont uniformément distribuées sur $[0,1]$. L'utilité de cette notion prend tout son sens avec le théorème suivant (voir par exemple [Jaworski et al., 2010]) :

Théorème 2.1 (Sklar). *Soit $\boldsymbol{X} = (X_1, \ldots, X_d)$ un vecteur aléatoire. Il existe une copule d-dimensionnelle $C_{\boldsymbol{X}}$ vérifiant*

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = C_{\boldsymbol{X}}(F_{X_1}(x_1), \ldots, F_{X_d}(x_d)) \ . \tag{2.6}$$

De plus, lorsque $F_{X_1}, \ldots, F_{X_d}$ sont continues, C est unique et donnée par

$$\forall \boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d, \ C_{\boldsymbol{X}}(\boldsymbol{u}) = F_{\boldsymbol{X}}(F_{X_1}^{-1}(u_1), \ldots, F_{X_d}^{-1}(u_d)) \ .$$

Le théorème de Sklar est un résultat central de la théorie des copules puisqu'il stipule que la loi d'un vecteur aléatoire est entièrement caractérisée par la donnée des lois marginales unidimensionnelles et d'une copule. De plus, dans le cas absolument continu, en dérivant l'expression donnée par Eq.(2.6), on obtient l'égalité suivante :

$$f_{\mathbf{X}}(x_1, \ldots, x_d) = c_{\mathbf{X}}(F_{X_1}(x_1), \ldots, F_{X_d}(x_d)) \times \prod_{i=1}^{d} f_{X_i}(x_i) \ , \tag{2.7}$$

où $c_{\mathbf{X}}$ est définie pour pour tout élément $\mathbf{u} = (u_1, \ldots, u_d)$ de l'hypercube $[0,1]^d$ par

$$c_{\mathbf{X}}(\mathbf{u}) = \frac{\partial^d C_{\mathbf{X}}(\mathbf{u})}{\partial u_1 \cdots \partial u_d} \ . \tag{2.8}$$

La fonction c définie par Eq.(2.8) est appelée *densité de copule* de $\mathbf{X} = (X_1, \ldots, X_d)$ et correspond à la PDF de $(F_{X_1}(X_1), \ldots, F_{X_d}(X_d))$. Pour plus de résultats sur la théorie des copules, le lecteur peut se référer par exemple à [Jaworski et al., 2010].

2.2.2 Estimation non-paramétrique de densité

Soit un ensemble $(\mathbf{X}^1, \ldots, \mathbf{X}^N)$ de réalisations indépendantes et identiquement distribuées (i.i.d) selon la loi $\mathbb{P}_{\mathbf{X}}$ supposée absolument continue. Le problème discuté dans cette section est celui de l'estimation de la densité $f_{\mathbf{X}}$, i.e, la construction d'un estimateur de $f_{\mathbf{X}}$ à partir de l'échantillon disponible $\{\mathbf{X}^n\}_{n=1}^N$. Il existe principalement deux familles de méthodes, à savoir les méthodes paramétriques et les méthodes non-paramétriques.

La première approche consiste à se donner un modèle statistique paramétrique, i.e, de supposer que la densité $f_{\mathbf{X}}$ appartient à une famille connue de distributions paramétrées, par exemple la famille des distributions normales $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ de moyenne $\boldsymbol{\mu}$ et de matrice de covariance $\boldsymbol{\Sigma}$. La densité $f_{\mathbf{X}}$ peut alors être approximée en considérant la PDF de la distribution $N(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ où $\hat{\boldsymbol{\mu}}$ et $\hat{\boldsymbol{\Sigma}}$ sont des estimateurs de $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$ respectivement. Pour cela plusieurs techniques existent, les plus classiques étant la méthode des moments et la méthode du maximum de vraisemblance [Rivoirard and Stoltz, 2012].

L'approche non-paramétrique quant à elle est plus générale dans le sens où les hypothèses faites sur la distribution des observations $(\mathbf{X}^1, \ldots, \mathbf{X}^N)$ sont moins rigides, la densité $f_{\mathbf{X}}$ n'étant plus contrainte à appartenir à une famille donnée. Dans les deux prochaines sous-sections, les deux plus grandes méthodes non-paramétriques d'estimation de densité sont présentées, à savoir l'estimation par noyau et l'estimation par maximisation d'entropie.

2.2.2.1 Estimation par noyau

Formulation. Pour simplifier les notations, plaçons nous dans le cas unidimensionnel. On suppose donc que l'on dispose d'un échantillon $(X^1, \dots, X^N)$ i.i.d selon une PDF f_X à support dans $\mathbb{R}$. On définit alors l'estimateur par noyau K et de *fenêtre* (ou *largeur de bande*) h comme suit :

$$\hat{f}_X(x) = \frac{1}{Nh} \sum_{n=1}^{N} K\left(\frac{x - X^n}{h}\right) , \tag{2.9}$$

où $K : \mathbb{R} \to \mathbb{R}$ est une fonction telle que

$$\int_{\mathbb{R}} K = 1 .$$

En pratique, les noyaux les plus couramment utilisés sont

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \qquad \text{(noyau gaussien)} , \tag{2.10}$$

$$K(x) = \frac{3}{4}(1 - x^2)\mathbb{1}_{[-1,1]}(x) \qquad \text{(noyau d'Epanechnikov)} . \tag{2.11}$$

La définition (2.9) peut être étendue au cas multidimensionnel. Si l'on dispose d'un ensemble $(\mathbf{X}^1, \dots, \mathbf{X}^N)$ i.i.d de réalisations d'un vecteur aléatoire $\mathbf{X}$ à valeur dans $\mathbb{R}^d$, on peut définir l'estimateur par noyau

$$\hat{f}_{\mathbf{X}}(\mathbf{x}) = \frac{\det(\mathbf{H})^{-1/2}}{N} \sum_{n=1}^{N} K_d\left(\mathbf{H}^{-1/2}(\mathbf{x} - \mathbf{X}^n)\right) , \tag{2.12}$$

où K_d est un noyau d-dimensionnel et $\mathbf{H} \in \mathcal{M}_d(\mathbb{R})$ une matrice symétrique définie positive également appelée fenêtre.

La performance de l'estimateur (2.9) repose essentiellement sur le choix de la fenêtre [Sheather and Jones, 1991]. En effet, la fenêtre h est un paramètre qui contrôle le degré de lissage de l'estimation, comme on peut l'observer sur la figure [2.2] où l'on compare la PDF de la distribution normale centrée réduite avec deux estimations par noyau différentes. Une fenêtre trop petite conduit à l'apparition d'oscillations artificielles sur le graphe de l'estimateur (courbe bleue). À l'inverse, une fenêtre trop grande entraîne un effet de *surlissage* (courbe noire).

On peut noter que la méthode d'estimation par noyau est en fait une généralisation de la *méthode des histogrammes*. En effet, la PDF f_X correspond à la dérivée de la CDF F_X sous l'hypothèse de continuité de cette dernière. Un estimateur naïf peut alors être obtenu par approximation par différences finies :

$$\hat{f}_X(x) := \frac{\hat{F}_X(x + h) - \hat{F}_X(x - h)}{2h} ,$$

où $\hat{F}_X$ est l'estimateur empirique de F_X :

$$\hat{F}_X(x) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}\left\{X^n \leq x\right\} .$$

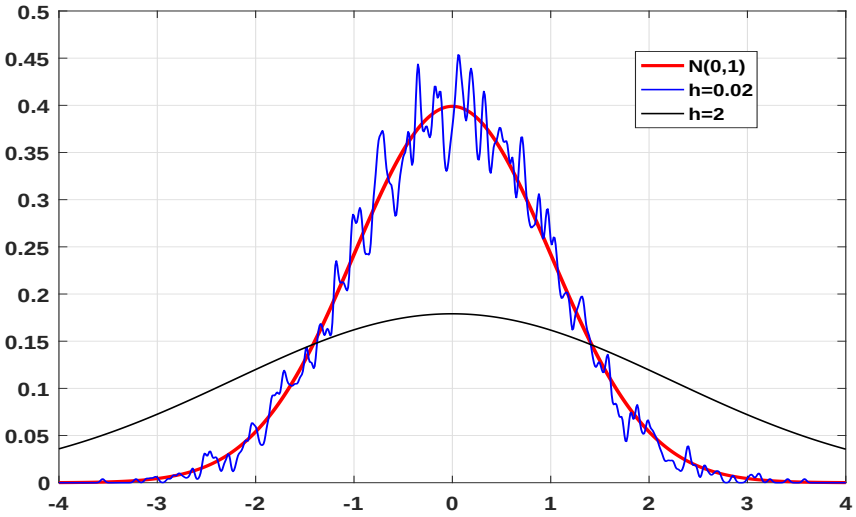


FIGURE 2.2. REPRÉSENTATION DE LA PDF DE LA DISTRIBUTION $N(0,1)$ (COURBE ROUGE) ET DE DEUX ESTIMATIONS PAR NOYAU GAUSSIEN POUR DIFFÉRENTES VALEURS DU PARAMÈTRE h (OBTENUES À PARTIR D'UN ÉCHANTILLON DE TAILLE $N = 5000$).

Cet estimateur peut être réécrit de la façon suivante :

$$\hat{f}_X(x) = \frac{1}{2Nh} \sum_{n=1}^{N} \mathbb{1}\left\{x - h \leq X^n \leq x + h\right\} = \frac{1}{Nh} \sum_{n=1}^{N} K\left(\frac{x - X^n}{h}\right) ,$$

où

$$K(x) = \begin{cases} \frac{1}{2} \text{ si } |x| \leq 1 , \\ 0 \text{ sinon } . \end{cases}$$

est la PDF de la loi de distribution uniforme sur $[-1, 1]$. L'estimateur naïf peut être vu comme une somme de "boîtes" centrées aux observations X^k. Heuristiquement, si les boîtes sont trop étriquées, certaines d'entre elles risquent de contenir très peu d'observations. À l'inverse, choisir une largeur h trop grande relativement au nombre N d'observations risque de "gommer" certaines caractéristiques de la distribution f_X, ce qui corrobore l'effet de lissage du paramètre h précédemment observé sur la figure [2.2].

Étude des performances. Les critères de performance de $\hat{f}_X$ les plus étudiés sont l'*erreur quadratique intégrée*

$$\mathrm{ISE}[\hat{f}_X] = \int (\hat{f}_X(x) - f_X(x))^2 dx , \tag{2.13}$$

qui est une mesure de précision aléatoire locale, l'*erreur quadratique moyenne*

$$\mathrm{MSE}[\hat{f}_X(x)] = \mathbb{E}[(\hat{f}_X(x) - f_X(x))^2] , \tag{2.14}$$

qui est une mesure de précision déterministe locale, et l'*erreur quadratique moyenne intégrée*

$$\mathrm{MISE}[\hat{f}_X] = \mathbb{E}\left[\int (\hat{f}_X - f_X)^2\right] , \tag{2.15}$$

qui est une mesure de précision globale. Si K est un noyau positif symétrique satisfaisant les conditions suivantes :

$$\kappa_2(K) := \int x^2 K(x)dx < +\infty \quad \text{et} \quad R(K) := \int K(x)^2 dx < +\infty \ , \tag{2.16}$$

alors l'erreur quadratique moyenne et l'erreur quadratique moyenne intégrée admettent les approximations asymptotiques suivantes lorsque N tend vers $+\infty$ [Wand and Jones, 1994] :

$$\text{AMSE}[\hat{f}_X(x)] := \frac{1}{4}\kappa_2(K)^2 (f_X^{(2)}(x))^2 h^4 + \frac{f_X(x)R(K)}{Nh} \ , \tag{2.17}$$

$$\text{AMISE}[\hat{f}_X] := \frac{1}{4}\kappa_2(K)^2 R(f_X^{(2)}) h^4 + \frac{R(K)}{Nh} \ , \tag{2.18}$$

l'argument reposant essentiellement sur le développement de Taylor de f_X à l'ordre 2 (ce qui sous-entend que f_X soit suffisamment régulière, i.e, que la dérivée seconde $f_X^{(2)}$ de f_X existe).

Résultats de convergence. Il découle de l'approximation (2.17) que l'estimateur par noyau (2.9) est ponctuellement consistant lorsque les conditions suivantes sont vérifiées :

$$h \underset{N\to+\infty}{\longrightarrow} 0 \quad \text{et} \quad Nh \underset{N\to+\infty}{\longrightarrow} +\infty \ . \tag{2.19}$$

Par ailleurs, on peut trouver dans [Devroye and Gyorfi, 1985] un résultat de convergence encore plus remarquable, dont la démonstration repose sur des techniques de concentration de la mesure, appliquées à des lois multinomiales. Avec comme unique hypothèse que K soit une fonction mesurable satisfaisant $K \geq 0$ et $\int K = 1$, il établit que les conditions (2.19) sont nécessaires et suffisantes pour assurer la convergence presque sûre de $\hat{f}_X$ vers f_X dans l'espace $L^1(\mathbb{R})$:

$$\|\hat{f}_X - f_X\|_{L^1(\mathbb{R})} \longrightarrow 0 \ \text{ presque sûrement} \ . \tag{2.20}$$

Remarques. La méthode d'estimation par noyau présentée précédemment admet certaines limites. Tout d'abord, l'optimisation de la fenêtre h est un problème difficile qui n'admet pas de solution générale. De nombreuses méthodes de sélection de fenêtre existent et deux grandes familles peuvent être distinguées : les méthodes de validation croisée qui se focalisent sur l'optimisation de $\text{ISE}[\hat{f}_X]$, et les méthodes *plug-in* qui cherchent à estimer le minimiseur h_{opt} de $\text{MISE}[\hat{f}_X]$. Sous les hypothèses précédentes, la valeur optimale asymptotique de h, i.e celle qui minimise $\text{AMISE}(\hat{f}_X)$ est donnée par

$$h^*_{\text{opt}} = \frac{R(K)^{1/5}}{\kappa_2(K)^{2/5} R(f_X^{(2)})^{1/5} N^{1/5}} \ . \tag{2.21}$$

Dans le cas du noyau gaussien, la méthode la plus populaire est la règle de Silverman [Silverman, 2018] qui consiste à faire une approximation de h^*_{opt} en supposant que f_X est gaussienne. Pour plus d'information sur les différentes méthodes de sélection de fenêtre, le lecteur intéressé peut se référer à la bibliographie [Heidenreich et al., 2013] et aux références associées.

D'autre part, la précision de l'estimation par noyau peut être impactée par des effets de bord, comme illustré par la figure [2.3(b)]. Par exemple lorsque les données sont positives, l'estimateur par noyau gaussien est non consistant pour la borne $x = 0$ [Wand and Jones, 1994] :

$$\mathbb{E}[\hat{f}_X(0)] = \frac{1}{2} f_X(0) + O(h), \quad N \to +\infty \ .$$

Plusieurs méthodes ont été proposées pour surmonter ce problème. Cela inclut par exemple les estimateurs à noyau à fenêtre variable [Terrell and Scott, 1992] et l'estimateur de [Botev et al., 2010] dont la construction est basée sur la théorie des processus de diffusion. L'approche développée par [Botev et al., 2010] assure la consistance de l'estimateur aux bornes du support [Botev et al., 2010], comme on peut l'observer sur la figure [2.3(a)]. Par ailleurs, [Botev et al., 2010] introduit également un estimateur *plug-in* qui ne requiert aucune information a priori sur f_X.

La précision de la méthode classique d'estimation par noyau (avec le noyau gaussien par exemple) peut également être affectée dans le cas des distributions à queue lourde. Ce problème peut être résolu par transformation des données [Wand et al., 1991, Yang and Marron, 1999] ou en employant des noyaux plus sophistiqués. On peut citer par exemple l'estimateur de [Buch-Larsen et al., 2005] basé sur la distribution de Champernowne modifiée, ou la méthode basée sur la distribution béta inverse [Bolancé et al., 2008]. Une autre alternative pour surmonter ces deux problèmes est de recourir au principe du maximum d'entropie présenté dans la prochaine sous-section. Pour plus d'information sur la méthode d'estimation par noyau et les avantages et inconvénients associés, le lecteur peut se référer à [Wand and Jones, 1994].

2.2.2.2 Principe du maximum d'entropie

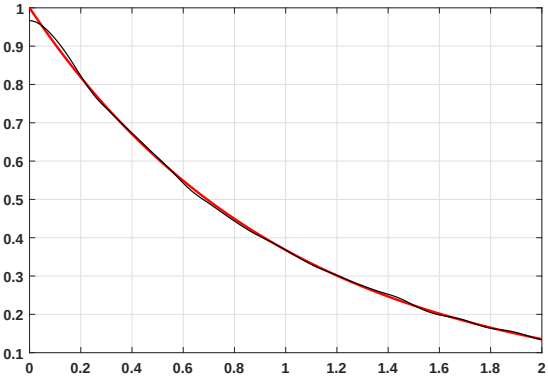
Formulation. Le principe du maximum d'entropie a été initialement introduit par Jaynes [Jaynes, 1957], et le lecteur peut se référer à [Soize, 2017, Kapur and Kesavan, 1992] pour plus de détails. L'idée de cette approche repose sur la théorie de l'information [Shannon, 1948]. Étant donné un vecteur aléatoire $\mathbf{X}$ à valeur dans un sous-ensemble S de $\mathbb{R}^d$ dont on souhaite estimer la PDF $f_{\mathbf{X}}$, le principe du maximum d'entropie préconise de considérer la distribution de probabilité qui maximise l'incertitude sous contrainte de l'information disponible sur la distribution de $\mathbf{X}$. La mesure de l'incertitude est définie par l'*entropie de Shannon* (ou *entropie différentielle*) de la PDF $f_{\mathbf{X}}$:

$$H(f_{\mathbf{X}}) = -\int_S f_{\mathbf{X}}(\mathbf{x}) \log(f_{\mathbf{X}}(\mathbf{x})) d\mathbf{x}. \tag{2.22}$$

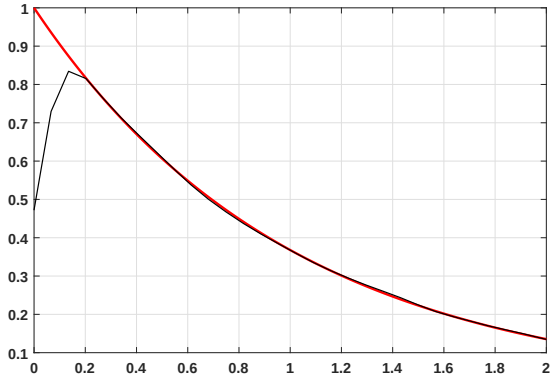
On suppose ici que l'information disponible sur la distribution de $\mathbf{X}$ s'écrit sous la forme $\mathbb{E}[g(\mathbf{X})] = \mu_{\mathbf{X}} \in \mathbb{R}^n$, où $g : \mathbb{R}^d \to \mathbb{R}^n$ est une application et où n désigne le nombre de contraintes. L'estimateur du maximum d'entropie (ME) $\hat{f}_{\mathbf{X}}$ est alors défini comme la solution du problème d'optimisation suivant

$$\begin{cases} \hat{f}_{\mathbf{X}} = \underset{f \in L^1(S, \mathbb{R}^+)}{\text{Argmax}} \ H(f) \\ \text{tel que } \int g(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \mu_{\mathbf{X}} \end{cases} \tag{2.23}$$

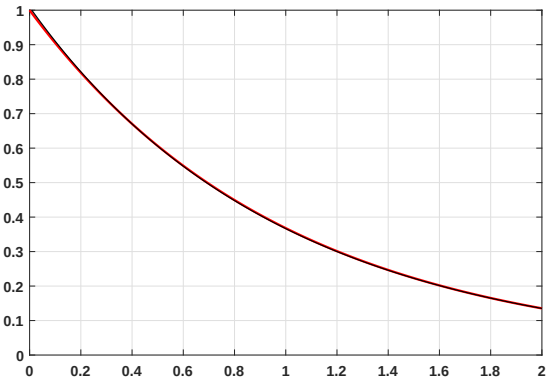
Le problème d'optimisation (2.23) est convexe et peut être reformulé à l'aide des *multiplicateurs de Lagrange*, voir par exemple [Kapur and Kesavan, 1992]. En effet, on



(a) Sélection de la fenêtre avec l'estimateur *plug-in* de [Botev et al., 2010] (résultat obtenu avec la toolbox développée par Botev et al.).



(b) Sélection de la fenêtre via la règle de Silverman.



(c) Estimation par maximum d'entropie (paramètres : $\alpha_k \in \{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}\}$).

FIGURE 2.3. REPRÉSENTATION DE DIFFÉRENTES ESTIMATIONS (COURBES NOIRES) DE LA PDF DE LA LOI EXPONENTIELLE $\mathcal{E}(1)$ (COURBES ROUGES). LES FIGURES [2.3(A)] ET [2.3(B)] CORRESPONDENT À DES ESTIMATIONS PAR NOYAU AVEC DEUX MÉTHODES DIFFÉRENTES DE SÉLECTION DE FENÊTRE. LA FIGURE [2.3(C)] CORRESPOND À UN ESTIMATEUR ME.

peut montrer que la dualité forte est vérifiée, de sorte qu'une solution du problème (2.23) est donnée par la distribution

$$\mathbf{x} \mapsto c_0(\boldsymbol{\lambda}^*)\mathbb{1}_S(\mathbf{z})e^{-\langle \boldsymbol{\lambda}^*, g(\mathbf{z})\rangle} \ ,$$

où $c_0(\boldsymbol{\lambda}^*)$ est une constante de normalisation et où $\boldsymbol{\lambda}^*$ est l'unique[1] solution du problème dual associé ($\langle \cdot, \cdot \rangle$ désigne le produit scalaire canonique de $\mathbb{R}^n$)

$$\boldsymbol{\lambda}^* = \underset{\boldsymbol{\lambda} \in \mathbb{R}^n}{\text{Argmin}} \ \langle \Lambda, \mu_{\mathbf{X}} \rangle + \log \int e^{-\langle \boldsymbol{\lambda}, g(\mathbf{x}) \rangle} d\mathbf{x} \ . \tag{2.24}$$

Dans le cas de la dimension $d = 1$ et lorsque la variable X est positive, le principe du maximum d'entropie est souvent appliqué en fixant des moments fractionnaires de X comme contraintes (un exemple est donné en figure [2.3(c)]), en vertu du résultat

[1] On peut montrer que la fonction duale apparaissant dans Eq.(2.24) est strictement convexe.

suivant qui donne une caractérisation de la loi d'une variable aléatoire positive en terme de moments :

Théorème 2.2 ([Lin, 1992]). *La distribution d'une variable aléatoire réelle positive X est entièrement caractérisée par la donnée d'une collection infinie $\{\mathbb{E}[X^{\alpha_i}]\}_{i=1}^{\infty}$ de moments fractionnaires avec $\alpha_i \in [0, \alpha^*]$ où $\mathbb{E}[X^{\alpha^*}] < +\infty$ pour un certain $\alpha^* > 0$.*

Résultat de convergence. Le choix de considérer des moments fractionnaires est motivé par plusieurs raisons. D'une part, [Zhang and Pandey, 2013] illustre numériquement que l'emploi de moments fractionnaires peut produire de meilleures estimations que celles obtenues avec des moments entiers. De plus, [Novi Inverardi and Tagliani, 2003] montre que cela assure également la convergence de l'estimateur du maximum d'entropie au sens de la divergence de Kullback-Leibler (2.3).

Théorème 2.3 ([Novi Inverardi and Tagliani, 2003]). *Soit X une variable aléatoire réelle positive, et une collection de moments fractionnaires $\{\mathbb{E}[X_k^{\alpha}]\}_{k=1}^{n}$ avec $\alpha_k = \frac{k\alpha^*}{n}$. Alors, l'estimateur du maximum d'entropie associé à ces contraintes, i.e,*

$$\hat{f}_X^n(x) \propto \exp\left(-\sum_{k=1}^{n} \lambda_k x^{\alpha_k}\right) ,$$

converge en entropie vers f_X lorsque le nombre de contraintes n tend vers $+\infty$. Plus précisément, on a la convergence suivante de la divergence de Kullback-Leibler :

$$d_{KL}(\hat{f}_X^n | f_X) = -\int_0^{+\infty} f_X \log\left(\frac{\hat{f}_X^n}{f_X}\right) \xrightarrow[n\to+\infty]{} 0 .$$

Remarques. La conclusion du théorème 2.3 entraine également la convergence L^1 d'après l'inégalité de Pinsker (2.4). La méthode du maximum d'entropie est facile à mettre en oeuvre mais le choix des contraintes reste une difficulté en pratique : considérons une variable aléatoire réelle X et une collection de moments fractionnaires $\{\mu_X(\alpha_k) := \mathbb{E}[X^{\alpha_k}]\}_{k=1}^{n}$. L'estimateur ME $\hat{f}_X^n$ associé est alors donné par

$$\hat{f}_X^n(x) \propto \exp\left(-\sum_{k=1}^{n} \lambda_k^* x^{\alpha_k}\right) ,$$

où $\boldsymbol{\lambda}^* = (\lambda_1^*, \dots, \lambda_n^*)$ est solution du problème d'optimisation convexe donné par Eq.(2.24). Ici, ce problème s'écrit comme suit :

$$\begin{aligned}\boldsymbol{\lambda}^* &= \underset{\lambda_1,\dots,\lambda_n}{\text{Argmin}} \left\{ \sum_{k=1}^{n} \lambda_k \mu_X(\alpha_k) + \log\left(\int_{\mathbb{R}} \exp\left(-\sum_{k=1}^{n} \lambda_k x^{\alpha_k}\right) dx \right) \right\} , \\ &:= \underset{\boldsymbol{\lambda}}{\text{Argmin}}\ \Gamma(\boldsymbol{\lambda}, \boldsymbol{\alpha}) .\end{aligned}$$

La question qui se pose en pratique est *quels* et *combien de* moments fractionnaires il faut considérer pour assurer une bonne estimation de la PDF f_X tout en minimisant le coût de simulation. Ceci sous-entend l'étude du problème d'optimisation suivant :

$$\underset{n}{\text{Min}} \left\{ \underset{\boldsymbol{\alpha}}{\text{Min}} \left\{ \underset{\boldsymbol{\lambda}}{\text{Min}}\ \Gamma(\boldsymbol{\lambda}, \boldsymbol{\alpha}) \right\} \right\} , \tag{2.25}$$

dont la résolution est très coûteuse puisque $\boldsymbol{\lambda}$ est fonction de $\boldsymbol{\alpha}$. On peut trouver dans [Tagliani, 2011] une procédure simplifiée pour résoudre le problème $\underset{\boldsymbol{\alpha}}{\text{Min}} \left\{ \underset{\boldsymbol{\lambda}}{\text{Min}}\, \Gamma(\boldsymbol{\lambda}, \boldsymbol{\alpha}) \right\}$. À noter toutefois que cette approche dépend de la résolution de systèmes linéaires mal conditionnés, ce qui peut occasionner des problèmes numériques.

Terminons cette section par un aspect non mentionné jusqu'à présent, connu sous le nom de *malédiction de la dimension* (*curse of dimensionality* en anglais). Lorsque la dimension augmente, le volume de l'espace augmente de telle sorte que les données disponibles deviennent éparses. En conséquence, de nombreuses méthodes numériques deviennent inefficaces à mesure que la dimension augmente pour un budget fixé. Ceci inclut les deux méthodes d'estimation de densité vues précédemment, celles-ci devenant moins robustes à partir de la dimension 3 pour un budget de simulation raisonnable. On verra dans le chapitre 6 que ce fléau de la dimension peut être surpassé en ayant recours aux modèles de copules *vignes* (*vines* en anglais) [Bedford and Cooke, 2001, Bedford and Cooke, 2002] dans le cas de l'estimation d'une densité de copule.

2.2.3 Méthodes d'échantillonnage et estimation des moments d'une variable aléatoire

L'objet des prochaines sous-sections est l'évaluation d'un moment de la forme $\mathbb{E}[\varphi(\mathbf{X})]$ où $\mathbf{X}$ est vecteur aléatoire de dimension d et $\varphi : \mathbb{R}^d \to \mathbb{R}$ une application mesurable telle que $\varphi(\mathbf{X})$ soit de variance finie. Rappelons que lorsque $\mathbf{X}$ admet une PDF $f_{\mathbf{X}}$, l'espérance de $\varphi(\mathbf{X})$ se réécrit comme l'intégrale suivante :

$$\mu_{\varphi(\mathbf{X})} := \mathbb{E}[\varphi(\mathbf{X})] = \int_{\mathbb{R}^d} \varphi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \ . \tag{2.26}$$

2.2.3.1 La méthode de Monte-Carlo

Soit $\{\mathbf{X}^n\}_{n=1}^N$ un échantillon de réalisations i.i.d de $\mathbf{X}$. On définit les estimateurs Monte-Carlo de l'espérance et de l'écart-type de $\varphi(\mathbf{X})$ par

$$\hat{\mu}^{\text{MC}}_{\varphi(\mathbf{X})} = \frac{1}{N} \sum_{n=1}^{N} \varphi(\mathbf{X}^n) \ , \tag{2.27}$$

$$\hat{\sigma}^{\text{MC}}_{\varphi(\mathbf{X})} = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} \left(\varphi(\mathbf{X}^n) - \hat{\mu}^{\text{MC}}_{\varphi(\mathbf{X})} \right)^2} \ . \tag{2.28}$$

Ces estimateurs sont non biaisés et fortement consistants en vertu de la *loi forte des grands nombres*. Par ailleurs, le *théorème central limite* permet d'obtenir des intervalles de confiance asymptotique de l'espérance (2.26) de niveau $\alpha \in]0,1[$:

$$\left[\hat{\mu}^{\text{MC}}_{\varphi(\mathbf{X})} \pm \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{N}} \hat{\sigma}^{\text{MC}}_{\varphi(\mathbf{X})} \right] \ , \tag{2.29}$$

où $z_{1-\frac{\alpha}{2}}$ désigne le $(1-\frac{\alpha}{2})$-quantile de la distribution $N(0,1)$.

Remarques. La méthode de Monte-Carlo présente une vitesse de convergence en $\frac{1}{\sqrt{N}}$ ce qui s'avère souvent plus lent que les méthodes numériques. En effet, dans le cas absolument continu, l'espérance de $\varphi(\mathbf{X})$ peut être évaluée par intégration numérique de l'intégrale définie par Eq.(2.27), par exemple avec les méthodes des rectangles, des trapèzes ou encore de Simpson qui admettent respectivement, dans le cas unidimensionnel $d = 1$ et si l'intégrande est suffisamment régulière, les vitesses de convergence $\frac{1}{N}$, $\frac{1}{N^2}$ et $\frac{1}{N^4}$. Cependant, la méthode de Monte-Carlo présente l'avantage d'être insensible à la régularité de l'intégrande $\varphi(\cdot)f_{\mathbf{X}}(\cdot)$ et à la dimension du problème, contrairement aux méthodes de quadratures classiques. En effet, la *malédiction de la dimension* impacte fortement la précision des formules de quadrature en grande dimension. Il existe certains résultats théoriques qui quantifient cet effet de "malédiction", voir par exemple le théorème de Bakhvalov dont une démonstration figure dans [Dimov, 2008]. Dans la section suivante, la question de la réduction de la variance de l'estimateur Monte-Carlo est abordée. Diverses méthodes existent comme l'échantillonnage stratifié, les variables antithétiques, les variables de contrôles etc. (voir par exemple [Hammersley, 2013]). Cependant, par soucis de concision, seule la méthode d'échantillonnage préférentiel est détaillée.

2.2.3.2 Réduction de la variance : l'échantillonnage préférentiel

Dans cette section, on se place dans le cas où $\mathbf{X}$ est absolument continue. La méthode d'échantillonnage préférentiel (*Importance Sampling* (IS) en anglais) à été introduite dans l'objectif de réduire la variance de l'estimateur Monte-Carlo défini par Eq.(2.27). L'idée principale est de générer des échantillons suivant une distribution auxiliaire g au lieu de la densité initiale $f_{\mathbf{X}}$. En outre, lorsque f est absolument continue par rapport à g, l'espérance de $\mathbf{X}$ admet la représentation suivante :

$$\begin{aligned}
\mu_{\varphi(\mathbf{X})} &= \int_{\mathbb{R}^d} \varphi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \ , \\
&= \int_{\mathbb{R}^d} \frac{\varphi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x}) d\mathbf{x} \ , \\
&= \mathbb{E}\left[\frac{\varphi(\tilde{\mathbf{X}}) f_{\mathbf{X}}(\tilde{\mathbf{X}})}{g(\tilde{\mathbf{X}})} \right], \text{ où } \tilde{\mathbf{X}} \sim g \ .
\end{aligned}$$

L'estimateur par tirages préférentiels est alors donné par

$$\hat{\mu}^{\mathrm{IS}}_{\varphi(\mathbf{X})} = \frac{1}{N} \sum_{n=1}^{N} \frac{\varphi(\tilde{\mathbf{X}}^n) f_{\mathbf{X}}(\tilde{\mathbf{X}}^n)}{g(\tilde{\mathbf{X}}^n)}, \ \{\tilde{\mathbf{X}}^n\}_{n=1}^{N} \overset{\text{i.i.d}}{\sim} g \ . \tag{2.30}$$

La variance de cet estimateur s'écrit :

$$\mathbb{V}\mathrm{ar}(\hat{\mu}^{\mathrm{IS}}_{\varphi(\mathbf{X})}) = \frac{1}{N} \left(\mathbb{E}\left[\left(\frac{\varphi(\tilde{\mathbf{X}}) f_{\mathbf{X}}(\tilde{\mathbf{X}})}{g(\tilde{\mathbf{X}})} \right)^2 \right] - \mu^2_{\varphi(\mathbf{X})} \right) \ . \tag{2.31}$$

L'avantage de cette méthode est que l'on a un contrôle de l'erreur sur l'estimation de la moyenne de $\varphi(\mathbf{X})$ puisque la variance $\mathbb{V}\mathrm{ar}(\hat{\mu}^{\mathrm{IS}}_{\varphi(\mathbf{X})})$ peut être approximée à l'aide de la méthode de Monte-Carlo (cf. section 2.2.3.1). La principale difficulté de l'approche par

tirages préférentiels est le choix de la distribution auxiliaire. En effet, on peut noter que la variance de $\hat{\mu}^{\mathrm{IS}}_{\varphi(\mathbf{X})}$ dépend du ratio $\varphi(\cdot)f_{\mathbf{X}}(\cdot)/g(\cdot)$ qui peut exploser si g est mal choisie. Pour que $\hat{\mu}^{\mathrm{IS}}_{\varphi(\mathbf{X})}$ soit plus performant que $\hat{\mu}^{\mathrm{MC}}_{\varphi(\mathbf{X})}$ il convient donc de choisir g dans l'objectif de minimiser la variance définie par Eq.(2.31). La distribution optimale, (i.e celle qui annule la variance) est définie par

$$g_{\mathrm{opt}}(\cdot) = \frac{\varphi(\cdot)f_{\mathbf{X}}(\cdot)}{\mu_{\varphi(\mathbf{X})}} \ . \tag{2.32}$$

La distribution optimale est cependant inutilisable en pratique puisqu'elle dépend de la quantité que l'on cherche précisément à estimer. Le challenge est donc de parvenir à déterminer une distribution auxiliaire la plus proche possible de la distribution optimale, et avec laquelle on sait échantillonner. Pour une bibliographie de ces méthodes, le lecteur intéressé peut se référer par exemple à [Smith et al., 1997, Tokdar and Kass, 2010]. On termine cette section par une présentation de l'échantillonnage préférentiel non-paramétrique (*Nonparametric Importance Sampling* (NIS) en anglais) introduit par [Zhang, 1996] qui sera employé dans le chapitre 4. Le principe de la méthode NIS est de construire un estimateur non-paramétrique de la distribution optimale g_{opt} à partir d'un échantillon généré suivant une distribution initiale g_0. Plus précisément, étant donné un N-échantillon $\{\mathbf{X}^n\}_{n=1}^N \overset{\text{i.i.d}}{\sim} g_0$ et un noyau K_d, g_{opt} est estimé par

$$\hat{g}_{\mathrm{opt}}(\mathbf{x}) = \frac{\det(\mathbf{H})^{-1/2}}{N\bar{\omega}} \sum_{n=1}^{N} \omega(\mathbf{X}^n) K_d\left(\mathbf{H}^{-1/2}(\mathbf{x}-\mathbf{X}^n)\right) \ , \tag{2.33}$$

où les poids sont égaux à

$$\bar{\omega} = \frac{1}{N}\sum_{n=1}^{N} \omega(\mathbf{X}^n) \quad \text{avec} \quad \omega(\mathbf{X}^n) = \frac{\varphi(\mathbf{X}^n)f_{\mathbf{X}}(\mathbf{X}^n)}{g_0(\mathbf{X}^n)} \ .$$

L'espérance de $\varphi(\mathbf{X})$ est alors approximée par tirages préférentiels avec la distribution auxiliaire $\hat{g}_{\mathrm{opt}}$:

$$\hat{\mu}^{\mathrm{NIS}}_{\varphi(\mathbf{X})} = \frac{1}{N'}\sum_{k=1}^{N'} \frac{\varphi(\tilde{\mathbf{X}})f_{\mathbf{X}}(\tilde{\mathbf{X}}^k)}{\hat{g}_{\mathrm{opt}}(\tilde{\mathbf{X}}^k)}, \ \{\tilde{\mathbf{X}}^k\}_{k=1}^{N'} \overset{\text{i.i.d}}{\sim} \hat{g}_{\mathrm{opt}} \ . \tag{2.34}$$

L'estimateur (2.33) est obtenu par une méthode d'estimation par noyau plus élaborée que celle détaillée en section 2.2.2.1 puisque des poids aléatoires interviennent dans sa définition. La convergence de ces estimateurs est étudiée dans [Hansen, 2008]. La convergence de l'estimateur $\hat{\mu}^{\mathrm{NIS}}_{\varphi(\mathbf{X})}$ est étudiée dans [Zhang, 1996]. Celle-ci est assurée lorsque K_d est un noyau produit $K_d(\mathbf{x}) = K(x_1)\cdots K(x_d)$ où K est un noyau unidimensionnel borné, symétrique et admettant un moment fini d'ordre 2 et lorsque $f_{\mathbf{X}}$ et φ sont de classe C^2 et à support compact. D'autres stratégies sont présentées dans la section 2.4 dans un contexte fiabiliste.

2.2.3.3 Méthodes de Monte-Carlo par chaînes de Markov

Supposons que $\mathbf{X}$ admette une densité par rapport à une mesure dominante λ. Pour simplifier les notations, la densité de Radon-Nikodym est notée $f_{\mathbf{X}}$ dans cette section.

Algorithme 1 : Algorithme de Métropolis-Hastings

Initialisation :
- Considérer une distribution initiale η_0 et un noyau d'exploration $\mathcal{Q}$.
- Tirer $\mathbf{X}_0 \sim \eta_0$.

pour k allant de 0 à n **faire**

Exploration : générer un candidat $\mathbf{X}'_k \sim Q(\mathbf{X}_k, \cdot)$.

Probabilité d'acceptation : calculer

$$\alpha(\mathbf{X}_k, \mathbf{X}'_k) = \min\left(1, \frac{f_{\mathbf{X}}(\mathbf{X}'_k) q(\mathbf{X}'_k, \mathbf{X}_k)}{f_{\mathbf{X}}(\mathbf{X}_k) q(\mathbf{X}_k, \mathbf{X}'_k)}\right) .$$

Acceptation-rejet : Tirer $u \sim \mathcal{U}([0,1])$.

si $u \leq \alpha(\mathbf{X}_k, \mathbf{X}'_k)$ **alors**
$\mathbf{X}_{k+1} = \mathbf{X}'_k$.

sinon
$\mathbf{X}_{k+1} = \mathbf{X}_k$.

fin

fin

Résultat : $(\mathbf{X}_k)_{0 \leq k \leq n}$.

Dans de nombreux cas, la densité $f_{\mathbf{X}}$ n'est connue qu'à une constante près. Les méthodes de Monte-Carlo par chaînes de Markov (Markov Chain Monte Carlo (MCMC) en anglais) permettent précisément d'éviter le calcul de la constante $\int f_{\mathbf{X}}(\mathbf{x}) d\lambda(\mathbf{x})$. Elles consistent à approcher la loi $f_{\mathbf{X}}$ en construisant une chaîne de Markov ergodique $(\mathbf{X}_k)$ admettant $f_{\mathbf{X}}$ comme loi stationnaire. Les deux méthodes les plus utilisées sont l'algorithme de Métropolis-Hastings (MH) et l'échantillonneur de Gibbs [Geman and Geman, 1984] qui est un cas particulier du premier, comme souligné par [Gelman et al., 1992]. L'algorithme MH à été introduit pour la première fois par [Metropolis et al., 1953] et généralisé ensuite par [Hastings, 1970]. Le principe est d'initialiser la chaîne à l'aide d'une loi initiale arbitraire η_0 puis de la faire évoluer selon des phases successives de proposition et d'acceptation-rejet. Partant d'un état $\mathbf{x}$, l'étape de proposition vise à explorer l'espace, en proposant un état $\mathbf{x}'$ suivant un noyau de transition $\mathcal{Q}$. Pour tout $\mathbf{x} \in \mathbb{R}^d$, on notera $q(\mathbf{x}, \cdot)$ la densité de la loi $\mathcal{Q}(\mathbf{x}, \cdot)$ par rapport à la mesure dominante λ. L'étape d'acceptation-rejet consiste quant à elle à accepter la proposition $\mathbf{x}'$ avec la probabilité suivante, appelée *taux d'acceptation*,

$$\alpha(\mathbf{x}, \mathbf{x}') = \min\left(1, \frac{f_{\mathbf{X}}(\mathbf{x}') q(\mathbf{x}', \mathbf{x})}{f_{\mathbf{X}}(\mathbf{x}) q(\mathbf{x}, \mathbf{x}')}\right), \quad \mathbf{x} \neq \mathbf{x}' , \tag{2.35}$$

la chaîne conservant l'état précédent $\mathbf{x}$ si la proposition est rejetée. Les étapes de l'algorithme MH sont résumées dans l'algorithme 1.

À l'issue de l'algorithme MH, on dispose d'une chaine de Markov $(\mathbf{X}_k)$ de transition

$$P(\mathbf{x}, d\mathbf{x}') = \alpha(\mathbf{x}, \mathbf{x}') q(\mathbf{x}, \mathbf{x}') d\mathbf{x}' + \left(1 - \int_{\mathbb{R}^d} \alpha(\mathbf{x}, \mathbf{x}') q(\mathbf{x}, \mathbf{x}') d\mathbf{x}'\right) \delta_{\mathbf{x}}(d\mathbf{x}') . \tag{2.36}$$

Par construction, P admet $f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}$ comme mesure invariante [Chib and Greenberg, 1995]. La convergence en loi de la chaine, et donc la convergence de l'algorithme, est alors

assurée si P vérifie[2] les hypothèses d'irréductibilité et d'apériodicité du théorème ergodique [Roberts and Smith, 1994]. Sous ces hypothèses, on peut alors approcher l'espérance de $\varphi(\mathbf{X})$ comme suit :

$$\frac{1}{N}\sum_{k=n_0+1}^{n_0+N}\varphi(\mathbf{X}_k)\xrightarrow[N\to+\infty]{\text{p.s}}\mathbb{E}[\varphi(\mathbf{X})]\ , \tag{2.37}$$

où $(\mathbf{X}_{n_0+1},\ldots,\mathbf{X}_{n_0+N})$ est une trajectoire de la chaine considérée à partir d'un *temps de chauffe* n_0, souvent introduit en pratique afin que la chaîne "oublie" sa loi initiale η_0.

Remarques. L'algorithme MH est très utilisé car il est simple à mettre en oeuvre. Cependant, un problème inhérent à cette approche est qu'il n'existe pas de moyen universel permettant d'estimer à partir de quand la loi de la chaîne est suffisamment proche de la loi cible $f_{\mathbf{X}}$. Dans le cas où $\mathbf{X}$ est à valeurs dans un espace fini, [Propp and Wilson, 1996] propose un algorithme de simulation permettant de pallier ce problème, au sens où la distribution simulée est exacte et l'algorithme s'arrête en temps (aléatoire) fini. Cependant le coût de simulation associé est en général non raisonnable pour les problèmes de grande taille et sans structure particulière, par exemple de monotonie (voir par exemple [Benaïm and El Karoui, 2005]). Pour finir, la performance de l'algorithme MH dépend des paramètres $q(\mathbf{x},d\mathbf{x}')$ et α. Pour plus d'information à propos du choix de noyau $\mathcal{Q}$ et sur les valeurs recommandées du taux d'acceptation α, le lecteur peut se référer à [Chib and Greenberg, 1995] et aux références associées.

2.3 Propagation des incertitudes dans les modèles boîte noire

Dans cette section, on considère un système complexe modélisé par un modèle boîte noire entrée sortie, dont la notion a été introduite dans la section 2.1. Compte tenu de la discussion menée dans l'introduction de la section 2.2, un tel modèle peut être représenté par une relation de la forme

$$Y=\mathcal{M}(\mathbf{X})\ , \tag{2.38}$$

où Y désigne la sortie du modèle (supposée scalaire dans cette thèse), $\mathcal{M}$ la boîte noire et $\mathbf{X}=(X_1,\ldots,X_d)$ un ensemble de d variables aléatoires correspondant aux différents paramètres d'entrée. Par propagation des incertitudes au travers du code $\mathcal{M}$, la sortie Y est une variable aléatoire réelle. Plusieurs quantités d'intérêt peuvent alors être considérées pour étudier le comportement stochastique de Y. On peut s'intéresser à :

- **Reconstruire** (si elle existe) **la PDF** f_Y de la sortie pour caractériser la variabilité de sortie. Dans ce cas, les techniques d'estimation non-paramétriques de densité abordées dans la section 2.2.2 peuvent être considérées.
- **Estimer certains moments de** Y pour étudier son comportement statistique. Lorsque la densité est connue et de faible dimension, les moments peuvent être

[2]Une condition suffisante est que le noyau d'exploration Q soit lui-même irréductible apériodique et vérifie la condition $q(x,y)>0\Rightarrow q(y,x)>0$ [Roberts and Smith, 1994].

obtenus par intégration numérique en vertu de la relation donnée par Eq.(2.5). Dans le cas contraire, il est souvent plus approprié de recourir à des méthodes stochastiques (cf. section 2.2.3).

- **Estimer la probabilité que Y dépasse un certain seuil critique S.** Pour un système robuste, une telle probabilité est très faible ($\ll 10^{-4}$ typiquement) ce qui rend difficile son estimation. L'étude des évènements rares est un sujet riche et complexe qui rentre dans le cadre de l'analyse de fiabilité. La section 2.4 passe en revue les méthodes d'estimation de probabilité d'évènements rares les plus utilisées.

2.4 Analyse de fiabilité des modèles boîte noire

L'étude des systèmes complexes représentés par un modèle entrée-sortie (2.38) comporte deux étapes majeures, à savoir l'*analyse de fiabilité* et l'*analyse de sensibilité* du modèle de sortie qui sera abordée dans le chapitre 3. Un problème fondamental en fiabilité est l'évaluation de la *probabilité de défaillance du système*, correspondant à un état non sécurisé et donc indésirable du système. Dans la suite, il est supposé sans perte de généralité que l'état de défaillance est de la forme $\{Y > S\}$, i.e, correspond au dépassement d'un certain seuil critique S par la sortie Y et la *région de défaillance* désignera l'ensemble $\mathcal{M}^{-1}([S, +\infty[)$ des réalisations de $\mathbf{X}$ qui mènent à la défaillance du système.

L'approche la plus simple pour estimer la probabilité de défaillance $\mathbb{P}(Y > S)$ (renotée parfois p_f dans ce manuscrit dans le but d'alléger les notations) est l'estimation Monte-Carlo présentée en section 2.2.3.1. En remarquant que $p_f = \mathbb{E}[\mathbb{1}_{\mathcal{M}(\mathbf{X})>S}]$, i.e, $p_f = \mathbb{E}[\varphi(\mathbf{X})]$ avec $\varphi(\cdot) = \mathbb{1}_{\mathcal{M}(\cdot)>S}$, l'estimateur Monte-Carlo de p_f s'écrit

$$\hat{p}_f^{\mathrm{MC}} := \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\mathcal{M}(\mathbf{X}^n)>S} \quad \text{où} \quad \{\mathbf{X}^n\}_{n=1}^{N} \overset{\text{i.i.d}}{\sim} f_{\mathbf{X}} \ . \tag{2.39}$$

Une statistique très utilisée pour évaluer la précision d'un estimateur est le coefficient de variation. Ici, le coefficient de variation $\mathrm{CV}(\hat{p}_f^{\mathrm{MC}})$ de l'estimateur Monte-Carlo (2.39) s'écrit comme suit :

$$\mathrm{CV}(\hat{p}_f^{\mathrm{MC}}) = \sqrt{\frac{1-p_f}{Np_f}} \ . \tag{2.40}$$

Lorsque la probabilité p_f prend de faibles valeurs on obtient

$$\mathrm{CV}(\hat{p}_f^{\mathrm{MC}}) \sim \frac{1}{\sqrt{Np_f}} \xrightarrow[p_f \to 0]{} +\infty \ , \tag{2.41}$$

ce qui implique que le coefficient de variation est non borné. En outre, cela entraine que l'estimation Monte-Carlo d'une probabilité de défaillance d'ordre 10^{-k} requiert $N = 10^{k+2}$ échantillons pour assurer une erreur relative de 10%. Ainsi, l'estimation Monte-Carlo classique n'est pas adaptée dans le contexte de l'estimation de probabilités d'évènements rares, d'autant plus que l'évaluation du code de calcul $\mathcal{M}$ peut s'avérer très coûteuse en pratique lorsque l'on considère des systèmes complexes. Ceci est dû au fait que la majorité des échantillons générés avec la distribution $f_{\mathbf{X}}$ sont situés en dehors de la zone d'intérêt $\mathcal{M}^{-1}([S, +\infty[)$.

Les prochaines sections décrivent différentes alternatives à la méthode de Monte-Carlo, le but étant d'assurer une bonne précision des estimations tout en minimisant le coût de simulation associé au nombre d'appels à la boîte noire $\mathcal{M}$. Par soucis de concision, seules les techniques d'estimation d'évènements rares les plus utilisées sont présentées. Le lecteur peut se référer par exemple à la bibliographie [Morio et al., 2014] pour plus d'exhaustivité.

2.4.1 L'approximation FORM/SORM

Les méthodes FORM (acronyme de *First Order Reliability Method*, méthode de fiabilité du premier ordre) et SORM (acronyme de *Second Order Reliability Method*, méthode de fiabilité du second ordre) [Madsen et al., 2006, Bjerager, 1991, Zhao and Ono, 1999] sont des méthodes très utilisées en fiabilité pour déterminer la probabilité de défaillance. L'approche consiste dans les deux cas à faire une approximation du modèle $\mathcal{M}$ à partir du *point de conception*, i.e, le point de défaillance le plus probable. Plus précisément, ces deux méthodes requièrent les étapes suivantes :

Passage à l'espace gaussien standard. La variable $\mathbf{X}$ est transformée en une variable normale centrée réduite

$$\mathbf{U} := T(\mathbf{X}) \sim N(0, I_d) \ ,$$

où T est un difféomorphisme appelé *transformation isoprobabiliste* [Bourinet, 2018]. Différentes transformations peuvent être considérées, les plus connues étant la transformation de Nataf (introduite par [Nataf, 1962] puis généralisée par [Lebrun and Dutfoy, 2009]) et la transformation de Rosenblatt [Rosenblatt, 1952]. La région de défaillance se réécrit donc sous la forme suivante :

$$\left\{\mathbf{u} \in \mathbb{R}^d \mid \mathcal{M}^*(\mathbf{u}) := \mathcal{M}(T^{-1}(\mathbf{u})) - S \geq 0\right\} \ . \tag{2.42}$$

Calcul du point de conception. Dans l'espace gaussien standard, le point de conception correspond au d-uplet $\mathbf{u}^*$ de l'hypersurface $\{\mathcal{M}^*(\mathbf{u}) = 0\}$ le plus proche de l'origine. Le point $\mathbf{u}^*$ est donc défini comme la solution du problème d'optimisation suivant :

$$\begin{cases} \mathbf{u}^* = \underset{\mathbf{u}\in\mathbb{R}^d}{\text{Argmin}} \ (\mathbf{u}^T\mathbf{u})^{1/2} \\ \text{tel que } \mathcal{M}^*(\mathbf{u}) = 0 \ . \end{cases} \tag{2.43}$$

Différents algorithmes ont été proposés pour résoudre ce problème d'optimisation quadratique sous contrainte non linéaire [Liu and Der Kiureghian, 1991].

Calcul de l'indice de fiabilité. En supposant que le point de conception est unique, on définit l'*indice de fiabilité* β par [Hasofer and Lind, 1974] :

$$\beta = (\mathbf{u}^{*T}\mathbf{u}^*)^{1/2} \ , \tag{2.44}$$

i.e, la distance entre l'origine et la surface limite $\{\mathcal{M}^*(\mathbf{u}) = 0\}$. Par ailleurs, β peut se réécrire de la façon suivante :

$$\beta = \boldsymbol{\alpha}^T\mathbf{u}^* \ ,$$

où $\boldsymbol{\alpha}$ est défini par l'opposé du vecteur gradient renormalisé de $\mathcal{M}^*$ évalué en $\mathbf{u}^*$:

$$\boldsymbol{\alpha} = -\frac{\nabla \mathcal{M}^*(\mathbf{u}^*)}{\|\nabla \mathcal{M}^*(\mathbf{u}^*)\|_2} \ .$$

Approximation FORM. La méthode FORM consiste alors à approcher la surface limite $\{\mathcal{M}^*(\mathbf{u}) = 0\}$ par un hyperplan affine de $\mathbb{R}^d$, ce qui revient à approcher g par son développement de Taylor à l'ordre 1 :

$$\begin{aligned}
\mathcal{M}^*(\mathbf{u}) &= \mathcal{M}^*(\mathbf{u}^*) + (\nabla \mathcal{M}^* \cdot \mathbf{u}^*)^T(\mathbf{u} - \mathbf{u}^*) + o(\|\mathbf{u} - \mathbf{u}^*\|_2) \ , \\
&= \|\nabla \mathcal{M}^* \cdot \mathbf{u}^*\|_2 \times (\beta - \boldsymbol{\alpha}^T \mathbf{u}) + o(\|\mathbf{u} - \mathbf{u}^*\|_2) \ , \\
&\approx \|\nabla \mathcal{M}^* \cdot \mathbf{u}^*\|_2 \times (\beta - \boldsymbol{\alpha}^T \mathbf{u}) \ .
\end{aligned}$$

La probabilité de défaillance peut alors être approchée comme suit :

$$p_f = \mathbb{P}(\mathcal{M}^*(\mathbf{U}) \geq 0) \approx \mathbb{P}(\beta - \boldsymbol{\alpha}^T \mathbf{U} \geq 0) = \mathbb{P}(\boldsymbol{\alpha}^T \mathbf{U} \leq \beta) \ .$$

Or,

$$\boldsymbol{\alpha}^T \mathbf{U} \sim N(0, \boldsymbol{\alpha}^T I_d \boldsymbol{\alpha}) = N(0,1) \ ,$$

ce qui donne l'estimateur de p_f suivant :

$$\hat{p}_f^{\mathrm{FORM}} = \Phi(\beta) \ , \tag{2.45}$$

où Φ désigne la fonction de répartition de la distribution normale centrée réduite $N(0,1)$.

Approximation SORM. La méthode FORM est une approximation linéaire du modèle transformé $\mathcal{M}^*$ et peut donc produire des estimations peu précises lorsque le code de calcul $\mathcal{M}$ est hautement non linéaire. La méthode SORM a été introduite afin d'améliorer les performances de FORM, et ce, en considérant le développement de Taylor à l'ordre 2 de la fonction $\mathcal{M}^*$ (ce qui requiert le calcul de la Hessienne). Différentes approximations de la probabilité de défaillance ont alors été proposées [Lemaire, 2013]. À titre d'exemple, [Breitung, 1984] a introduit la formule asymptotique suivante, qui approche la valeur exacte de la probabilité de défaillance lorsque $\beta \to +\infty$:

$$\hat{p}_f^{\mathrm{SORM}} = \Phi(\beta) \prod_{i=1}^{d-1} \frac{1}{\sqrt{1 + \beta \kappa_i}} \ ,$$

où les κ_i désignent les courbures principales[3] de $\mathcal{M}^*$ au point de conception.

Remarques. Les méthodes FORM-SORM sont couramment utilisées puisqu'elles permettent d'évaluer des probabilités de défaillance avec un faible budget de simulation. Cependant, les hypothèses inhérentes à l'application de ces deux méthodes sont très restrictives. Premièrement, il est difficile en pratique de vérifier la pertinence des approximations portant sur la fonction $\mathcal{M}^*$ définie par Eq.(2.42). De plus, aucun contrôle de l'erreur sur l'estimation de p_f n'est disponible contrairement aux méthodes d'échantillonnage préférentiel pour lesquelles on dispose d'un estimateur de la variance (cf. section 2.2.3.2). Enfin,

[3] La notion de courbure principale est utilisée en géométrie différentielle pour caractériser la géométrie locale des surfaces à l'ordre 2.

ces deux approches sont construites sur l'hypothèse d'unicité du point de conception, ce qui peut compromettre leur utilisation lorsque la région de défaillance n'est pas connexe. Le cas multimodal peut néanmoins être géré en ayant recours à des techniques plus élaborées, comme la méthode proposée par [Der Kiureghian and Dakessian, 1998] qui repose sur des applications successives d'approximations FORM-SORM.

2.4.2 Techniques d'échantillonnage préférentiel

L'échantillonnage préférentiel, dont le principe a été présenté en section 2.2.3.2, est une méthode de réduction de la variance qui peut être utilisée dans la méthode de Monte-Carlo. Le principe de cette méthode est également très utile dans le cadre de l'analyse de fiabilité. En effet, comme constaté dans l'introduction de la section 2.4, l'estimation Monte-Carlo échoue à estimer de faibles probabilités. Ceci est dû au fait que la majorité des échantillons générés avec la distribution initiale $f_{\mathbf{X}}$ sont situés en dehors de la région critique $\mathcal{M}^{-1}([S, +\infty[)$. L'idée de l'échantillonnage préférentiel est de considérer une distribution auxiliaire qui concentre l'échantillonnage dans cette région d'intérêt. Dans cette sous-section, trois différentes techniques d'échantillonnage préférentiel sont présentées.

2.4.2.1 Méthodes basées sur la recherche du point de conception

L'idée de ces méthodes est d'utiliser l'information obtenue par l'approximation FORM. Avec les notations de la section précédente, rappelons que la probabilité de défaillance peut se réécrire

$$p_f = \mathbb{P}(\mathcal{M}^*(\mathbf{U}) \geq 0), \ \mathbf{U} \sim N(0, I_d) \ ,$$

après le passage dans l'espace gaussien standard. Sous l'hypothèse que le point de conception $\mathbf{u}^*$ soit connue, [Schuëller and Stix, 1987] et [Melchers, 1989] proposent d'estimer p_f en échantillonnant via la distribution de $\mathbf{U}$ recentrée au point de conception $\mathbf{u}^*$. La distribution auxiliaire considérée s'écrit donc

$$g(\mathbf{u}) = \varphi_d(\mathbf{u} - \mathbf{u}^*) \ , \tag{2.46}$$

où φ_d désigne la PDF de la distribution $N(0, I_d)$, ce qui procure l'estimateur suivant de la probabilité de défaillance :

$$\hat{p}_f^{\text{FORM-Shift}} = \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{\mathcal{M}^*(\mathbf{U}^n) \geq 0} \frac{\varphi_d(\mathbf{U}^n)}{g(\mathbf{U}^n)}, \ \{\mathbf{U}^n\}_{n=1}^N \overset{\text{i.i.d}}{\sim} g \ , \tag{2.47}$$

où g est définit par Eq.(2.46). Une autre approche développée par [Harbitz, 1986] consiste à exclure la β-sphère $\{\|\mathbf{u}\|_2 \leq \beta\}$ du domaine d'échantillonnage, où β est définie par Eq.(2.44). Cette méthode est appelée *échantillonnage préférentiel tronqué* et propose d'estimer la probabilité de défaillance par la formule suivante :

$$\hat{p}_f^{\text{FORM-T}} = (1 - F_{\chi_d^2}(\beta^2)) \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{\|\mathbf{u}\|_2 > \beta}(\mathbf{U}^n), \ \{\mathbf{U}^n\}_{n=1}^N \overset{\text{i.i.d}}{\sim} \varphi_d \ , \tag{2.48}$$

où $F_{\chi_d^2}$ désigne la CDF de la distribution du chi-deux à d degrés de liberté.

Remarques. Les deux techniques précédentes héritent des inconvénients de la méthode FORM mentionnés dans la section 2.45. En outre, elles sont qualifiées de *non adaptatives* puisqu'elles supposent l'existence d'un unique point de conception. Ainsi, lorsque la région de défaillance est non connexe, certaines zones peuvent être non prises en compte ce qui a un impact sur la précision de l'estimation. Une extension adaptative de ces méthodes a été proposée par [Grooteman, 2008, Grooteman, 2011], combinant une procédure de tirages préférentiels avec la technique de *simulation directionnelle* [Bjerager, 1988, Ditlevsen et al., 1990].

2.4.2.2 La méthode de *Cross-Entropy*

Dans cette section, la distribution auxiliaire g est supposée appartenir à une famille de densités $\mathcal{H} = \{g_{\boldsymbol{\lambda}}; \boldsymbol{\lambda} \in \Delta\}$, où Δ désigne l'espace des valeurs du paramètre $\boldsymbol{\lambda}$. Un exemple classique est la famille des densités gaussiennes de dimension d paramétrée par la moyenne $\boldsymbol{\mu}$ et la matrice de covariance $\boldsymbol{\Sigma}$. L'objectif de la méthode de *Cross-Entropy* (CE) [Rubinstein and Kroese, 2004] est de déterminer le paramètre $\boldsymbol{\lambda}_{\text{opt}}$ tel que $g_{\boldsymbol{\lambda}_{\text{opt}}}$ soit le plus "proche" possible de la distribution optimale théorique

$$g_{\text{opt}}(\cdot) = \mathbb{1}_{\mathcal{M}(\mathbf{x})>S} \frac{f_{\mathbf{X}}(\cdot)}{p_f} \ . \tag{2.49}$$

Cette "proximité" peut être mesurée par la divergence de Kullback-Leibler (2.3). La valeur de $\boldsymbol{\lambda}_{\text{opt}}$ est donc définie par

$$\lambda_{\text{opt}} = \underset{\boldsymbol{\lambda}\in\Delta}{\text{Argmin}}\ d_{\text{KL}}(g_{\text{opt}}|g_{\boldsymbol{\lambda}}) \ . \tag{2.50}$$

Résoudre le problème (2.50) n'est pas chose aisée puisque celui-ci dépend de la densité optimale g_{opt} qui est inconnue en pratique. En réinjectant le fait que $g_{\text{opt}} \propto \mathbb{1}_{\mathcal{M}(\cdot)>S} f_{\mathbf{X}}$, [Rubinstein and Kroese, 2004] montre que Eq.(2.50) est équivalent au problème suivant :

$$\lambda_{\text{opt}} = \underset{\boldsymbol{\lambda}\in\Delta}{\text{Argmax}}\ \mathbb{E}\left[\mathbb{1}_{\mathcal{M}(\mathbf{X})>S} \log(g_{\boldsymbol{\lambda}}(\mathbf{X}))\right] \ . \tag{2.52}$$

En pratique on considère le problème approché suivant, obtenu en estimant l'espérance intervenant dans Eq.(2.52) par tirage préférentiel :

$$\lambda_{\text{opt}} = \underset{\boldsymbol{\lambda}\in\Delta}{\text{Argmax}}\ \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\mathcal{M}(\mathbf{X}^n)>S} \frac{f_{\mathbf{X}}(\mathbf{X}^n)}{\tilde{g}(\mathbf{X}^n)} \log\left(g_{\boldsymbol{\lambda}}(\mathbf{X}^n)\right) \ , \tag{2.53}$$

où les échantillons $\mathbf{X}^1, \ldots, \mathbf{X}^N$ sont tirés selon une distribution auxiliaire $\tilde{g}$. Cependant, le problème (2.53) ne peut être résolu directement dans un contexte d'estimation d'estimation d'évènements rares puisque l'échantillon $\{\mathbf{X}^n\} \sim \tilde{g}$ risque de ne pas contenir suffisamment de points vérifiant $\mathcal{M}(\mathbf{X}^n) > S$. Cette difficulté peut être résolue itérativement en faisant intervenir une suite croissante de seuils intermédiaires

$$S_0 < S_1 < \cdots < S_k < \cdots \leq S \ ,$$

choisis de manière adaptative à partir de ρ-quantiles d'un échantillon de sortie $\{Y^n\} \overset{\text{i.i.d}}{\sim} f_Y$ [Rubinstein and Kroese, 2004]. Les différentes étapes de la procédure de *Cross-Entropy* sont résumées par l'algorithme 2.

Algorithme 2 : Algorithme de *Cross Entropy*

Initialisation : k=1 ;
- Considérer une distribution initiale $g_{\boldsymbol{\lambda}_0} := g_0$;
- Définir un niveau de quantile $\rho \in]0,1[$;
- Générer $\{\mathbf{X}_n^{[1]}\}_{n=1}^N \overset{\text{i.i.d}}{\sim} g_0$ puis évaluer $Y_1^{[1]} = \mathcal{M}(\mathbf{X}_1^{[1]}), \dots, Y_N^{[1]} = \mathcal{M}(\mathbf{X}_N^{[1]})$;
- Calculer le ρ-quantile empirique γ_1 de $Y_1^{[1]}, \dots, Y_N^{[1]}$.

tant que $\gamma_k < S$ **faire**

Calculer

$$\boldsymbol{\lambda}_k = \underset{\boldsymbol{\lambda} \in \Delta}{\text{Argmax}} \sum_{n=1}^{N} \mathbb{1}_{\mathcal{M}(\mathbf{X}_n^{[k]}) > \gamma_k} \frac{f_{\mathbf{X}}(\mathbf{X}_n^{[k]})}{g_{\boldsymbol{\lambda}_{k-1}}(\mathbf{X}_n^{[k]})} \log\left(g_{\boldsymbol{\lambda}}(\mathbf{X}_n^{[k]})\right) , \tag{2.51}$$

où $\{\mathbf{X}_n^{[k]}\}_{n=1}^N \overset{\text{i.i.d}}{\sim} g_{\boldsymbol{\lambda}_{k-1}}$ et où il y a exactement ρN termes non nuls par construction.

- Fixer $k \to k+1$;
- Générer $\{\mathbf{X}_n^{[k]}\}_{n=1}^N \overset{\text{i.i.d}}{\sim} g_{\boldsymbol{\lambda}_{k-1}}$ puis évaluer $Y_1^{[k]} = \mathcal{M}(\mathbf{X}_1^{[k]}), \dots, Y_N^{[k]} = \mathcal{M}(\mathbf{X}_N^{[k]})$;
- Calculer le ρ-quantile empirique γ_k de $Y_1^{[k]}, \dots, Y_N^{[k]}$.

fin

Résultat : Fixer $k_{\text{end}} = k$ puis estimer la probabilité de défaillance par

$$\hat{p}_f^{\text{CE}} = \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{\mathcal{M}(\mathbf{X}_n^{[k_{\text{end}}]}) > S} \frac{f_{\mathbf{X}}(\mathbf{X}_n^{[k_{\text{end}}]})}{g_{\boldsymbol{\lambda}_{k_{\text{end}}-1}}(\mathbf{X}_n^{[k_{\text{end}}]})} .$$

Remarque. L'algorithme de *Cross-Entropy* est simplifié dans le cas où $\mathcal{H}$ est la famille exponentielle, i.e, la classe des densités dont la forme générale est

$$g_{\boldsymbol{\lambda}}(\cdot) = \phi(\cdot) \exp(a(\boldsymbol{\lambda})\eta(\cdot) - b(\boldsymbol{\lambda})) .$$

En effet, le problème (2.51) admet alors une solution analytique [Rubinstein and Kroese, 2004], ce qui accélère fortement le temps de calcul. La limite de la *Cross-Entropy* réside dans l'hypothèse que la distribution optimale g_{opt} soit "proche" de la famille de référence $\mathcal{H}$, ce qui ne peut être vérifié en pratique. En outre, lorsque $\mathcal{H}$ est composée de distributions unimodales, la *Cross-Entropy* peut perdre en précision lorsque la région de défaillance présente plusieurs modes. De récents travaux proposent de gérer ce problème en combinant l'algorithme de *Cross-Entropy* avec un modèle de mélange, i.e, considérer des distributions de référence s'exprimant comme une somme de densités. On peut citer par exemple [Kurtz and Song, 2013] qui utilise un modèle de mélange gaussien pour lequel les distributions de $\mathcal{H}$ s'expriment comme une somme de plusieurs gaussiennes. À noter toutefois que dans ce dernier cas, le problème (2.51) ne peut plus être résolu de manière analytique.

2.4.2.3 Échantillonnage préférentiel adaptatif non-paramétrique

Algorithme 3 : Algorithme NAIS

Initialisation : Considérer un noyau K_d et un niveau de quantile $\rho \in]0,1[$;
- Générer $\{\mathbf{X}_n^{[0]}\}_{n=1}^N \overset{\text{i.i.d}}{\sim} f_{\mathbf{X}}$ puis évaluer $Y_1^{[0]} = \mathcal{M}(\mathbf{X}_1^{[0]}), \ldots, Y_N^{[0]} = \mathcal{M}(\mathbf{X}_N^{[0]})$;
- calculer le ρ-quantile empirique γ_0 de $Y_1^{[0]}, \ldots, Y_N^{[0]}$;
- Calculer

$$\hat{g}_{\text{opt}}^{[0]} = \frac{\det(\mathbf{H}^{[0]})^{-1/2}}{N\bar{\omega}^{[0]}} \sum_{n=1}^{N} \omega(\mathbf{X}_n^{[0]}) K_d\left((\mathbf{H}^{[0]})^{-1/2}(\mathbf{x} - \mathbf{X}_n^{[0]})\right) ,$$

où $\bar{\omega}^{[0]} = \frac{1}{N}\sum_{n=1}^{N} \omega(\mathbf{X}_n^{[0]})$ avec $\omega(\mathbf{X}_n^{[0]}) = \mathbb{1}_{\mathcal{M}(\mathbf{X}_n^{[0]})>\gamma_0}$;

- Générer $\{\mathbf{X}_n^{[1]}\}_{n=1}^N \overset{\text{i.i.d}}{\sim} \hat{g}_{\text{opt}}^{[0]}$ puis évaluer $Y_1^{[1]} = \mathcal{M}(\mathbf{X}_1^{[1]}), \ldots, Y_N^{[1]} = \mathcal{M}(\mathbf{X}_N^{[1]})$;
- Calculer le ρ-quantile empirique γ_1 de $Y_1^{[1]}, \ldots, Y_N^{[1]}$;
- Fixer k=1 ;

tant que $\gamma_k < S$ **faire**

- Calculer

$$\bar{\omega}^{[k]} = \frac{1}{N}\sum_{n=1}^{N} \omega(\mathbf{X}_n^{[k]}) \quad \text{avec} \quad \omega(\mathbf{X}_n^{[k]}) = \mathbb{1}_{\mathcal{M}(\mathbf{X}_n^{[k]})>\gamma_k} \frac{f_{\mathbf{X}}(\mathbf{X}_n^{[k]})}{\hat{g}_{\text{opt}}^{[k-1]}(\mathbf{X}_n^{[k]})} ,$$

- Considérer l'estimateur NIS associé à la distribution initiale $\hat{g}_{\text{opt}}^{[k-1]}$:

$$\hat{g}_{\text{opt}}^{[k]} = \frac{\det(\mathbf{H}^{[k]})^{-1/2}}{N\bar{\omega}^{[k]}} \sum_{n=1}^{N} \omega(\mathbf{X}_n^{[k]}) K_d\left((\mathbf{H}^{[k]})^{-1/2}(\mathbf{x} - \mathbf{X}_n^{[k]})\right) .$$

- Fixer $k \to k+1$;
- Générer $\{\mathbf{X}_n^{[k]}\}_{n=1}^N \overset{\text{i.i.d}}{\sim} \hat{g}_{\text{opt}}^{[k-1]}$ puis évaluer $Y_1^{[k]} = \mathcal{M}(\mathbf{X}_1^{[k]}), \ldots, Y_N^{[k]} = \mathcal{M}(\mathbf{X}_N^{[k]})$;
- Calculer le ρ-quantile empirique γ_k de $Y_1^{[k]}, \ldots, Y_N^{[k]}$.

fin

Résultat : Fixer $k_{\text{end}} = k$ puis estimer la probabilité de défaillance par

$$\hat{p}_f^{\text{NAIS}} = \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\mathcal{M}(\mathbf{X}_n^{[k_{\text{end}}]})>S} \frac{f_{\mathbf{X}}(\mathbf{X}_n^{[k_{\text{end}}]})}{g^{[k_{\text{end}}-1]}(\mathbf{X}_n^{[k_{\text{end}}]})} . \tag{2.54}$$

La méthode NIS présentée en section 2.2.3.2 admet une version adaptative appelée échantillonnage préférentiel adaptatif non-paramétrique (Nonparametric Adaptive Importance Sampling (NAIS) en anglais). La méthode NAIS a été introduite par [Zhang, 1996] et adaptée au cadre de l'estimation de probabilités d'évènements rares dans divers travaux [Neddermeyer, 2009, Morio, 2012]. Le principe de NAIS est similaire à la procédure de

Cross-Entropy, excepté que la densité d'échantillonnage est remplacée par un estimateur à noyaux pondérés (cf. algorithme 3).

Remarques. La méthode de NAIS est particulièrement efficace pour gérer les cas où la région de défaillance a une géométrie très complexe. Cependant, elle souffre des limitations inhérentes à l'estimation par noyau. En outre, l'optimisation des fenêtres $\mathbf{H}^{[k]}$ peut être très complexe en pratique et la précision de l'estimateur $\hat{p}_f^{\mathrm{NAIS}}$ est fortement diminuée en grande dimension.

2.4.3 Méthodes de Monte-Carlo séquentielles

Le principe de la méthode de Monte-Carlo séquentielle (Sequential Monte Carlo) (SMC) est de décomposer la probabilité de défaillance $p_f = \mathbb{P}(Y > S)$ en un produit de probabilités conditionnelles moins rares et donc estimables à moindre coût. Parmi les précurseurs de cette approche figurent [Kahn and Harris, 1951] et [Rosenbluth and Rosenbluth, 1955] dans le cadre de la physique des particules et de la chimie moléculaire respectivement. Différentes variantes on été développées, et ce avec différents formalismes mathématiques [Au and Beck, 2001, Glasserman et al., 1999, Cérou et al., 2006, Cérou and Guyader, 2007, Del Moral and Lezaud, 2006]. On adopte dans cette section le point de vue de [Cérou et al., 2012] qui repose sur la théorie des *modèles de Feynman-Kac* [Del Moral, 2004] et leur approximation par *algorithmes particulaires*.

Cas non adaptatif (NA-SMC). Considérons une suite de seuils intermédiaires

$$S_0 := -\infty < \cdots < S_m := S \ .$$

La probabilité de défaillance peut alors se réécrire comme suit

$$p_f = \mathbb{P}(Y > S) = \prod_{p=0}^{m-1} \mathbb{P}(Y > S_{p+1} | Y > S_p) \ , \tag{2.55}$$

et peut donc être évaluée en estimant chacune des probabilités intermédiaires sous réserve de savoir échantillonner selon les lois conditionnelles $f_{Y|Y>S_p}$. Dans [Cérou et al., 2012], il est montré que la théorie des modèles de Feynman-Kac [Del Moral, 2004] s'applique dans ce contexte. En outre, p_f peut être approximée via un filtre particulaire évoluant au travers d'étapes de *sélection* et de *mutation* successives. À chaque itération p, cela consiste à sélectionner les particules qui ont atteint le seuil S_{p+1} et de mettre à jour les particules ainsi sélectionnées avec un noyau de Markov qui laisse invariante la loi conditionnelle $\mathbf{X}|Y > S_p$. Plus précisément l'algorithme SMC non adaptatif se déroule comme suit :

Initialisation : générer $\{\mathbf{X}_0^n\}_{n=1}^{N_x} \overset{\text{i.i.d}}{\sim} f_{\mathbf{X}}$ où N_x désigne la taille de la population de particules, puis calculer $Y_0^n = \mathcal{M}(\mathbf{X}_0^n)$ pour $n = 1, \ldots, N_x$.

Sélection : étant donné une configuration $\{\mathbf{X}_p^n, Y_p^n\}_{n=1}^{N_x}$, choisir aléatoirement et indépendamment N_x particules vérifiant $Y_p^n > S_{p+1}$.

Mutation : appliquer l'algorithme de Metropolis-Hastings 1 de noyau d'exploration $\mathcal{Q}$ et de loi cible $f_{\mathbf{X}|Y>S_{p+1}}$ à chacune des particules $\tilde{\mathbf{X}}_p^n$ précédemment sélectionnées. On obtient alors une nouvelle configuration $\{\mathbf{X}_{p+1}^n, Y_{p+1}^n\}_{n=1}^{N_x}$.

Algorithme 4 : Algorithme SMC adaptatif

Initialisation : Considérer un niveau de quantile $\rho \in]0,1[$ et un noyau d'exploration $\mathcal{Q}$;
- Générer $\{\mathbf{X}_0^n\}_{n=1}^{N_x} \overset{\text{i.i.d}}{\sim} f_{\mathbf{X}}$;
- Évaluer $Y_0^n = \mathcal{M}(\mathbf{X}_0^n)$ pour $n = 1, \ldots, N_x$;
- Calculer le ρ-quantile γ_0 de $Y_0^1, \ldots, Y_0^{N_x}$;
- Fixer $p = 0$;

tant que $\gamma_p < S$ **faire**
- *Sélection*

Dupliquer les ρN_x particules vérifiant $Y_p^n > \gamma_p$ en N_x particules $\{\tilde{\mathbf{X}}_p^n\}_{n=1}^{N_x}$;
- *Mutation*

pour n allant de 1 à N_x **faire**
- Générer $\mathbf{X}_{p+1}^n \sim M_p(\tilde{\mathbf{X}}_p^n, \cdot)$ où M_p est un noyau de transition de Metropolis-Hastings ciblant la loi $\mathbf{X}|Y > \gamma_p$ et de noyau d'exploration $\mathcal{Q}$;
- Évaluer $Y_{p+1}^n = \mathcal{M}(\mathbf{X}_{p+1}^n)$;

fin
- Calculer le ρ-quantile γ_{p+1} de $Y_{p+1}^1, \ldots, Y_{p+1}^n$;
- Fixer $p \to p+1$;

fin

Résultat : Le nombre (aléatoire) m d'étapes de sélection-mutation et estimer la probabilité de défaillance p_f par

$$\hat{p}_f^{\text{A-SMC}} = (1-\rho)^{m-1} \times \left(\frac{1}{N_x} \sum_{n=1}^{N_x} \mathbb{1}_{\mathcal{M}(\mathbf{X}_m^n)>S} \right) . \tag{2.56}$$

Les échantillons $\{\mathbf{X}_p^n\}_{n=1}^{N_x}$ ainsi générés sont approximativement distribués selon $\mathbf{X}|Y > S_{p+1}$ mais ne sont pas indépendants [Cérou et al., 2012]. L'étape de mutation peut être effectuée A_p fois pour diminuer cette corrélation. La théorie de Feynman-Kac assure alors que l'estimateur suivant

$$\hat{p}_f^{\text{NA-SMC}} = \prod_{p=0}^{m-1} \left(\frac{1}{N_x} \sum_{n=1}^{N_x} \mathbb{1}_{\mathcal{M}(\mathbf{X}_p^n)>S_{p+1}} \right) , \tag{2.57}$$

est un estimateur de p_f non biaisé et fortement consistant. L'estimateur (2.57) est obtenu à partir de (2.55) en approchant chacune des probabilités conditionnelles $\mathbb{P}(Y > S_{p+1}|Y > S_p)$ par son approximation Monte-Carlo associée aux variables $\{\mathbf{X}_p^n\}_{n=1}^{N_x}$ (qui sont approximativement indépendantes et distribuées selon $f_{\mathbf{X}}^{Y>S_p}$ par construction). Une démonstration du théorème central limite (TCL) pour la méthode SMC non adaptative est proposée dans [Cérou et al., 2012].

Cas adaptatif (A-SMC). Au lieu de fixer les seuils intermédiaires à l'avance, ce qui peut être est difficile en pratique, ceux-ci peuvent être choisis de manière adaptative. Par ailleurs, la variance de l'estimateur $\hat{p}_f^{\text{NA-SMC}}$ défini par Eq.(2.57) est minimisée lorsque

chacune des probabilités intermédiaires sont égales [Cérou et al., 2006]. En ce sens, l'approche SMC adaptative (résumée dans l'algorithme 4) consiste à appliquer la procédure SMC vue précédemment en imposant un un taux de succès $(1-\rho)$ entre deux seuils consécutifs. L'algorithme s'arrête lorsque le ρ-quantile de la population de particule dépasse le seuil critique S.

Remarques. La méthode SMC adaptative est robuste[4] dans le cas des systèmes hautement non linéaires. En revanche, le nombre d'appels à la boîte noire $\mathcal{M}$ est égal à

$$N_x \left(1 + mA_x\right) \ ,$$

où A_x est le nombre de mutations effectuées à chaque étape et où m est le nombre (aléatoire et potentiellement infini) d'étapes, ce qui est supérieur au budget de simulation requis par les techniques de fiabilité rencontrées dans les sections précédentes. Par ailleurs, la précision de l'estimation est dans certains cas très sensible au réglage des paramètres N_x et A_x. Néanmoins, à l'instar de la simulation de Monte-Carlo et des méthodes d'échantillonnage préférentiels, la méthode SMC adaptative peut être couplée à l'utilisation de métamodèles [Bourinet, 2016, Bourinet et al., 2011, Huang et al., 2016] dont le but est de produire une version simplifiée du modèle $\mathcal{M}$ pour laquelle le coût d'évaluation est fortement réduit. Enfin, l'estimateur $\hat{p}_f^{\text{A-SMC}}$ défini par (2.56) présente un biais, qui est néanmoins négligeable relativement à son écart relatif [Cérou et al., 2012].

2.5 Techniques de décompositions fonctionnelles

Considérons $Y = \mathcal{M}(\mathbf{X})$ un modèle boîte noire entrée-sortie général, où $\mathbf{X} = (X_1, ..., X_d)$ est un vecteur aléatoire de dimension d et $\mathcal{M} : \mathbb{R}^d \to \mathbb{R}$ une fonction déterministe. Pour tout sous-ensemble $\mathbf{u}$ de $\{1, \ldots, d\}$, on note $\mathbf{X_u}$ le vecteur marginal $(X_i, i \in \mathbf{u})$, $|\mathbf{u}|$ le cardinal de $\mathbf{u}$ et $-\mathbf{u}$ son complémentaire.

La résolution de problèmes de fiabilité comptant un grand nombre de variables d'entrée (typiquement $d > 20$) est un véritable challenge du fait de la malédiction de la dimension. Les méthodes présentées dans la section précédente permettent souvent d'estimer des probabilités de défaillance en présence d'un grand nombre de variables aléatoires mais requièrent un grand nombre d'appels à la boîte noire $\mathcal{M}(\cdot)$ potentiellement coûteuse à évaluer. L'essence des techniques de décompositions fonctionnelles est de proposer une représentation de la boîte noire $\mathcal{M}(\cdot)$ en une combinaison de fonctions plus simples. L'une des plus célèbres est le développement en série de Taylor permettant d'approcher localement $\mathcal{M}(\cdot)$ par une fonction polynômiale dont les coefficients dépendent des dérivées de $\mathcal{M}(\cdot)$ (sous réserve que celle-ci soit suffisamment régulière). La technique de HDMR (pour *High Dimensional Model Reduction*) [Rabitz et al., 1999] a pour objectif quant à elle d'exprimer $\mathcal{M}(\cdot)$ comme somme de fonctions de dimensions inférieures :

$$\mathcal{M}(\mathbf{x}) = \mathcal{M}_0 + \sum_{i=1}^{d} \mathcal{M}_i(x_i) + \sum_{1 \leq i < j \leq d} \mathcal{M}_{ij}(x_i, x_j) + \cdots + \mathcal{M}_{12..d}(\mathbf{x}) \ , \tag{2.58}$$

où $\mathcal{M}_0$ est une constante, $\mathcal{M}_i$, $i \in \{1, \ldots, d\}$ sont les effets principaux, les fonctions $\mathcal{M}_{ij}$, $i < j \in \{1, \ldots, d\}$ représentent les effets d'interactions d'ordre 2, etc. La décomposition

[4] Cet algorithme est utilisé pour l'estimation de probabilité faibles (« 10^{-4}). Dans le cas contraire, on préfèrera utiliser la méthode de Monte-Carlo classique.

(2.58) n'est pas unique et il existe principalement deux techniques dans la littérature permettant d'identifier les composantes de la décomposition HDMR : ANOVA et Cut-HDMR.

2.5.1 Décomposition de Hoeffding-Sobol et ANOVA.

La *décomposition de Hoeffding-Sobol* [Sobol, 1993] correspond à l'identification des composantes de Eq.(2.58) après transformation de l'espace des variables en l'hypercube unité $[0,1]^d$ et ajout des contraintes d'orthogonalité suivantes :

$$\forall \mathbf{u} \subset \{1,\dots,d\},\ \mathbf{u} \neq \emptyset,\ \forall j \notin \mathbf{u},\ \int_0^1 \mathcal{M}_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}})\mathrm{d}x_j = 0\ . \tag{2.59}$$

Lorsque la fonction $\mathcal{M}$ est de carré intégrable sur $[0,1]^d$, i.e, $\mathcal{M} \in L^2([0,1]^d, \mathbb{R})$, les composantes de la décomposition (2.58) sont alors uniquement déterminées et s'expriment récursivement comme suit :

$$\begin{aligned}
\mathcal{M}_0 &= \int_{[0,1]^d} \mathcal{M}(\mathbf{x})\mathrm{d}\mathbf{x}\ , \\
\mathcal{M}_i(x_i) &= \int_{[0,1]^{d-1}} \mathcal{M}(\mathbf{x})\mathrm{d}\mathbf{x}_{-i} - \mathcal{M}_0\ , \\
\mathcal{M}_{ij}(x_i,x_j) &= \int_{[0,1]^{d-2}} \mathcal{M}(\mathbf{x})\mathrm{d}\mathbf{x}_{-\{i,j\}} - \mathcal{M}_i(x_i) - \mathcal{M}_j(x_j) - \mathcal{M}_0\ . \\
&\vdots
\end{aligned}$$

Une conséquence de la condition (2.59) est que les termes de la décomposition ANOVA sont deux à deux orthogonaux, i.e,

$$\forall \mathbf{u},\mathbf{v} \subset \{1,\dots,d\},\ \mathbf{u} \neq \mathbf{v},\ \int_{[0,1]^d} \mathcal{M}_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}})\mathcal{M}_{\mathbf{v}}(\mathbf{x}_{\mathbf{v}})\mathrm{d}\mathbf{x} = 0\ . \tag{2.60}$$

Supposons que l'entrée $\mathbf{X} \sim \mathcal{U}([0,1]^d)$ soit uniformément distribuée sur l'hypercube unité. La décomposition de Hoeffding-Sobol couplée avec la propriétés d'orthogonalité (2.60) assure la décomposition de la variance globale $\mathbb{V}\mathrm{ar}(Y)$ en somme de variances partielles, appelée *décomposition ANOVA* :

$$\mathbb{V}\mathrm{ar}(Y) = \mathbb{V}\mathrm{ar}(\mathcal{M}(\mathbf{X})) = \sum_{\mathbf{u}\neq\emptyset} \mathbb{V}\mathrm{ar}(\mathcal{M}_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}))\ , \tag{2.61}$$

avec

$$\mathbb{V}\mathrm{ar}(\mathcal{M}_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}})) = \mathbb{V}\mathrm{ar}(\mathbb{E}[Y|\mathbf{X}_{\mathbf{u}}]) - \sum_{\mathbf{v}\subsetneq\mathbf{u}} (-1)^{|\mathbf{u}|-|\mathbf{v}|}\mathbb{V}\mathrm{ar}(\mathbb{E}[Y|\mathbf{X}_{\mathbf{v}}])\ .$$

2.5.2 Cut-HDMR

La technique d'identification cut-HDMR des composantes de la décomposition (2.58) s'appuie sur une projection de l'espace sur des plans affines passant par un unique point

$\mathbf{c} = (c_1, \ldots, c_d)$, appelé dénoté *point d'ancrage*. Les composantes sont alors définies de la manière suivante :

$$\begin{aligned}
\mathcal{M}_0 &= \mathcal{M}(\mathbf{c}) \ , \\
\mathcal{M}_i(x_i) &= \mathcal{M}(c_1, .., c_{i-1}, x_i, c_{i+1}, .., c_d)) - \mathcal{M}_0 \ , \\
\mathcal{M}_{ij}(x_i, x_j) &= \mathcal{M}(c_1, .., c_{i-1}, x_i, c_{i+1}, .., c_{j-1}, x_j, c_{j+1}, .., c_d) - \mathcal{M}_i(x_i) - \mathcal{M}_j(x_j) - \mathcal{M}_0 \ , \\
&\vdots
\end{aligned}$$

Cette méthode comparée à la précédente requiert beaucoup moins d'appels à la boîte noire pour calculer les composantes, mais repose cependant sur le choix d'un point d'ancrage.

Supposons désormais que la sortie Y est strictement positive. Le modèle $\mathcal{M}(\cdot)$ peut alors être transformé comme suit :

$$\varphi(\mathbf{X}) = \log(Y) = \log(\mathcal{M}(\mathbf{X})) \ . \tag{2.62}$$

[Zhang and Pandey, 2013] propose alors d'approcher le modèle $\varphi(\mathbf{X})$ en conservant uniquement les termes d'ordre 1 de sa décomposition cut-HDMR :

$$\begin{aligned}
\varphi(\mathbf{X}) &\approx \sum_{i=1}^{d} \log(\mathcal{M}(c_1, .., c_{i-1}, X_i, c_{i+1}, .., c_d)) - (d-1)\log(\mathcal{M}(\mathbf{c})) \ , \\
&:= \sum_{i=1}^{d} \log(\mathcal{M}(X_i, \mathbf{c})) - (d-1)\log(\mathcal{M}(\mathbf{c})) \ ,
\end{aligned}$$

où le point d'ancrage $\mathbf{c}$ est défini par la moyenne de l'entrée $\mathbf{X}$. Dès lors, le modèle initial $Y = \mathcal{M}(\mathbf{X})$ peut être approximé comme un produit de fonctions unidimensionnelles :

$$Y \approx (\mathcal{M}(\mathbf{c}))^{(1-d)} \prod_{i=1}^{d} \mathcal{M}(X_i, \mathbf{c}) \ . \tag{2.63}$$

Pour plus d'exhaustivité sur les méthodes HDMR, le lecteur intéressé peut consulter l'état de l'art effectué dans la thèse [Lelièvre, 2018].

Chapitre 3

Analyse de sensibilité globale basée sur les distributions

Contents

3.1 Introduction et motivations 49

3.2 Décomposition fonctionnelle de la variance et indices de Sobol 51

3.3 Indices de sensibilité basés sur des mesures de dissimilarité 53

3.3.1 Indices de sensibilité basés sur des fonctions de contraste . . . 54

3.3.2 Indices de sensibilité basés sur la distance de Cramér Von Mises 55

3.3.3 Les mesures de dépendance de Csizár 55

3.4 Les mesures d'importance δ_i de Borgonovo : un état de l'art 57

3.4.1 δ_i vue comme une espérance 58

3.4.2 δ_i vue comme une mesure de dépendance 62

3.4.3 Représentation de δ_i en terme de densité de copule 68

3.5 Conclusion 69

3.1 Introduction et motivations

Cette section se focalise sur l'étude d'un modèle boîte noire entrée-sortie général, notion précédemment introduite en section 2.3. On considère donc une relation de la forme $Y = \mathcal{M}(\mathbf{X})$ où $\mathbf{X} = (X_1, \ldots, X_d)$ est un vecteur aléatoire et $\mathcal{M} : \mathbb{R}^d \to \mathbb{R}$ une fonction déterministe appelée *boîte noire*. Pour tout sous-ensemble $\mathbf{u}$ de $\{1, \ldots, d\}$, on note $\mathbf{X_u}$ le vecteur marginal $(X_i, i \in \mathbf{u})$, $|\mathbf{u}|$ le cardinal de $\mathbf{u}$ et $-\mathbf{u}$ son complémentaire.

La prise en compte des incertitudes est une étape majeure de l'étude du comportement des modèles boîte noire. Cette étape, connue sous le nom de *propagation des incertitudes*, à été abordée dans le chapitre 2 et consiste à étudier le comportement stochastique de la sortie Y.

Dans cette section, on s'intéresse à une autre question fondamentale formulée par Saltelli [Saltelli, 2002], à savoir *"comment l'incertitude au niveau de la sortie Y peut être*

répartie entre les différentes sources d'incertitudes associées aux variables d'entrée". Ceci rentre dans le cadre de l'*analyse de sensibilité* qui vise à étudier l'influence de la loi de l'entrée $\mathbf{X}$ sur la sortie Y. En outre, l'analyse de sensibilité présente plusieurs objectifs :

- **Réduction d'incertitude :** diminuer (par ajout de connaissance) les incertitudes affectant les entrées détectées comme étant les plus influentes afin de réduire l'incertitude au niveau du modèle de sortie.
- **Réduction de modèle :** identifier les variables d'entrée les moins influentes pour simplifier le modèle en diminuant la dimension d de l'entrée, ces dernières pouvant être fixées à des valeurs nominales sans perte d'information sur la sortie.
- **Compréhension de modèle :** améliorer la connaissance du modèle en étudiant les interactions entre les variables d'entrée.

Il existe deux grandes approches en analyse de sensibilité, à savoir les méthodes *locales* et les méthodes *globales* (désignées dans la suite par l'acronyme anglais GSA, pour *global sensitivity analysis*). Le reste de cette section ne prétend pas à l'exhaustivité et pose uniquement les bases des méthodes mentionnées. Toutefois, une attention toute particulière sera portée sur la description des méthodes globales dites *quantitatives* dans les prochaines sections. Pour plus de détails sur les méthodes d'analyse de sensibilité, le lecteur intéressé peut se référer aux bibliographies [Borgonovo and Plischke, 2016, Iooss and Lemaître, 2015] et aux références associées.

Historiquement, les méthodes locales sont les premières à avoir été appliquées pour étudier l'impact de petites perturbations de l'entrée sur la sortie du modèle. En outre, l'étude se focalise sur des valeurs nominales de l'entrée $\mathbf{X}$. Cet impact local est souvent mesuré à partir de dérivées partielles du modèle $\mathcal{M}$ à ces valeurs nominales $\mathbf{x}^*$, approche connue sous le nom de *One-At-a-Time design* (OAT) dans la littérature anglaise [Saltelli and Annoni, 2010] :

$$\frac{\partial \mathcal{M}}{\partial x_i}(\mathbf{x}^*) \ . \tag{3.1}$$

Les indices de sensibilité locaux, comme l'indice OAT défini par Eq.(3.1), ont pour avantage d'être simples à calculer et à interpréter. Toutefois, la linéarité du modèle est nécessaire pour quantifier de manière globale l'influence des entrées à partir de la seule connaissance des indices de sensibilité locaux.

Contrairement aux approches locales, les méthodes GSA considèrent l'ensemble des valeurs prises par les entrées du modèle. Elles peuvent être classifiées en deux familles distinctes, les méthodes *qualitatives* et les méthodes *quantitatives*.

Tout d'abord, il y a les techniques de *screening* dont le principe est de discrétiser l'espace d'entrée puis d'effectuer des analyses de sensibilité locales autour de points de la grille choisis aléatoirement. Dans le cas de la méthode de Morris [Morris, 1991] qui figure parmi les méthodes de *screening* les plus utilisées, ces effets locaux sont évalués par l'approche OAT définie précédemment. En outre, ces méthodes adoptent une stratégie qualifiée de globale dans le sens où elles considèrent une moyennisation des effets élémentaires. Ce type d'approche a pour avantage de fournir une manière *qualitative* d'explorer rapidement le

comportement du code de calcul $\mathcal{M}$, et ce, avec un faible budget de simulation. De plus ces méthodes sont particulièrement adaptées dans le cas où la dimension du modèle est grande. Cependant, elles ne prennent pas réellement en compte toute la distribution de l'entrée $\mathbf{X}$. Les méthodes de *screening* se distinguent ainsi des méthodes *quantitatives* qui constituent le sujet central des prochaines sections de ce chapitre. Récemment introduite, la méthode DGSM (acronyme de *Derivative-based global sensitivity measures*) [Sobol and Gresham, 1995] est une méthode qualitative consistant à calculer une statistique globale des dérivées locales (3.1). Par exemple, [Kucherenko et al., 2009] définit l'indice suivant :

$$\nu_i = \mathbb{E}\left[\left(\frac{\partial \mathcal{M}}{\partial x_i}(\mathbf{X})\right)^2\right] ,$$

dans le cadre où les entrées X_i sont gaussiennes et indépendantes. À l'instar des méthodes de screening, la méthode DGSM nécessite beaucoup moins d'appels au code de calcul que les méthodes globales quantitatives. Par ailleurs, une relation entre les indices ν_i et les indices de Sobol totaux (présentés en section 3.2) est établie dans [Lamboni et al., 2013]. Pour plus de détails sur la méthode DGSM, le lecteur intéressé peut se référer à [Kucherenko and Iooss, 2017].

Les méthodes globales quantitatives sont associées à une *mesure d'importance*, i.e, une quantité permettant de hiérarchiser les différentes entrées selon un critère d'influence donné. La section 3.2 porte sur l'une des mesures d'importance les plus connues, à savoir les indices de Sobol basés sur la décomposition fonctionnelle de la variance. La section 3.3 s'intéresse aux indices de sensibilité définis à partir d'une mesure de *dissimilarité*. Il s'agit d'une grande famille d'indice qui inclut en particulier les indices de Sobol et les indices de Borgonovo dont le problème d'estimation constitue le coeur du propos de ce manuscrit. En outre, la dernière section 3.4 présente un état de l'art sur l'estimation des indices de Borgonovo du premier ordre.

3.2 Décomposition fonctionnelle de la variance et indices de Sobol

Supposons que l'entrée $\mathbf{X}$ soit de loi uniforme[1] sur $[0,1]^d$ et que la fonction $\mathcal{M}$ soit de carré intégrable sur l'hypercube $[0,1]^d$, i.e, $\mathcal{M} \in L^2([0,1]^d, \mathbb{R})$. Sous ces hypothèses, la *décomposition ANOVA* (2.61) (cf. section 2.5.1) s'applique et assure alors la décomposition suivante de la variance globale en somme de variances partielles :

$$\mathbb{V}\mathrm{ar}(Y) = \sum_{\mathbf{u}\neq\emptyset} \mathbb{V}\mathrm{ar}(\mathcal{M}_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}})) , \tag{3.2}$$

avec

$$\mathbb{V}\mathrm{ar}(\mathcal{M}_{\mathbf{u}}(\mathbf{X})) = \mathbb{V}\mathrm{ar}(\mathbb{E}[Y|\mathbf{X}_{\mathbf{u}}]) - \sum_{\mathbf{v}\subsetneq\mathbf{u}} (-1)^{|\mathbf{u}|-|\mathbf{v}|} \mathbb{V}\mathrm{ar}(\mathbb{E}[Y|\mathbf{X}_{\mathbf{v}}]) .$$

Il en résulte la définition suivante des *indices de Sobol* explicitée par [Sobol, 1993]

$$S_{\mathbf{u}} = \frac{\mathbb{V}\mathrm{ar}(\mathcal{M}_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}))}{\mathbb{V}\mathrm{ar}(Y)} , \tag{3.3}$$

[1] Cette hypothèse peut être généralisée à toute distribution produit par simple transformation (en appliquant les fonctions de répartition inverse par exemple).

où l'indice de Sobol $S_{\mathbf{u}}$ d'ordre $|\mathbf{u}|$ quantifie la contribution du groupe de variable $\mathbf{X}_{\mathbf{u}}$ à la variance du modèle de sortie $Y = \mathcal{M}(\mathbf{X})$. Par construction, ces indices vérifient les relations suivantes

$$0 \leq S_{\mathbf{u}} \leq 1 \ ,$$

et

$$\sum_{\mathbf{u} \neq \emptyset} S_{\mathbf{u}} = 1 \ ,$$

ce qui rend leur interprétation aisée. En outre, si $S_{\mathbf{u}}$ est proche de 1, cela signifie que la part de variabilité induite par $\mathbf{X}_{\mathbf{u}}$ est prédominante. En pratique, il peut s'avérer laborieux de considérer l'ensemble des $2^d - 1$ indices de Sobol, notamment lorsque la dimension d du problème est grande. On préfère alors regarder les *indices de Sobol totaux* introduits par [Homma and Saltelli, 1996] et définis par

$$S_{T_i} = \frac{\mathbb{V}\mathrm{ar}(\mathbb{E}[Y|\mathbf{X}_{-i}])}{\mathbb{V}\mathrm{ar}(Y)} \ , \tag{3.4}$$

où $\mathbf{X}_{-i}$ désigne l'entrée $\mathbf{X}$ privée de sa i-ème composante X_i. D'après (3.2) et (**??**) l'indice S_{T_i} peut se réécrire comme suit :

$$S_{T_i} = \sum_{i \in \mathbf{u}} S_{\mathbf{u}} \ ,$$

où la somme porte sur tous les groupes de variables $\mathbf{X}_{\mathbf{u}}$ incluant X_i. L'indice de Sobol total S_{T_i} quantifie ainsi la contribution de X_i à la variance de sortie en prenant en compte les interactions à tous les ordres avec les autres variables .

L'estimation des indices de Sobol peut être effectuée en utilisant la méthode *pick-and-freeze* [Sobol, 2001]. Pour une étude détaillée des propriétés statistiques de l'estimateur *pick-freeze*, le lecteur intéressé peut se référer à [Gamboa et al., 2016]. Cependant, le nombre d'appels au modèle nécessaires à l'application de cette méthode est conséquent. À titre d'exemple l'estimation de l'ensemble des indices d'ordre un et d'ordre deux requiert respectivement $N(d+1)$ et $(\binom{d}{2}+1)N$ évaluations de modèle. Dès lors diverses méthodes reposant sur la construction d'un méta-modèle ont été proposées pour réduire ce coût de simulation. Peuvent être cités entre autre, les travaux [Sudret, 2008] et [Marrel et al., 2009] qui utilisent respectivement des modèles basés sur des *développements en polynômes de chaos* et des processus gaussiens (modèles de *krigeage*).

Remarques. Les indices de Sobol d'ordre 1 S_i et les indices totaux S_{T_i} sont très souvent considérés conjointement pour faire de l'analyse de sensibilité globale. En effet, lorsque les entrées sont indépendantes on a

$$\sum_{i=1}^{d} S_i \leq 1 \leq \sum_{i=1}^{d} S_{T_i} \ . \tag{3.5}$$

Ainsi, la première inégalité suggère de considérer les indices totaux au lieu des indices d'ordre 1 qui ont tendance à sous-estimer l'effet d'une variable d'entrée sur la variance de sortie en omettant les effets d'interaction avec les autres variables. Cependant la seconde inégalité montre que les indices totaux ont tendance au contraire à surestimer l'impact

des entrées. Cette observation ne tient plus lorsqu'il existe une structure de dépendance entre les variables d'entrée. En effet, [Song et al., 2016] montre que lorsque les entrées sont corrélées, la somme des indices totaux peut être inférieure à 1 et la somme des indices d'ordre 1 peut être supérieure à 1, ce qui rend leur interprétation difficile. Ceci est dû au fait que la décomposition ANOVA (3.2) n'est plus valable dans le cas corrélé. L'extension de l'utilisation des indices de Sobol dans le cas corrélé est un problème qui a donné lieu à d'abondants travaux dans la littérature. Dans le cas dépendent, [Mara et al., 2015] introduit de nouveaux indices de sensibilité basés sur la variance et définis à partir de la transformation de Rosenblatt [Rosenblatt, 1952] de l'entrée $\mathbf{X}$, permettant de se ramener dans le cadre où l'entrée est uniformément distribuée sur l'hypercube $[0,1]^d$. Par ailleurs, la recherche d'une décomposition de Hoeffding-Sobol généralisée a été effectuée par [Hooker, 2007] puis [Chastaing et al., 2012]. En outre, [Chastaing et al., 2012] montre que de nouveaux indices de sensibilités peuvent être définis sous certaines conditions générales portant sur la distribution de l'entrée. D'autre part, [Owen, 2014, Song et al., 2016, Iooss and Prieur, 2017] définissent des mesures de sensibilités basées sur le concept de *valeur de Shapley* [Shapley, 1953] utilisé en théorie des jeux. L'interprétation de ces nouveaux indices est intéressante puisqu'ils vérifient une décomposition similaire à la décomposition ANOVA (3.2). Le chapitre 6 sera l'occasion de revenir plus en détail sur la définition, les propriétés et l'interprétation de la valeur de Shapley.

Pour finir et motiver la prochaine section, notons que les indices de Sobol prennent uniquement en compte la variance de Y qui n'est pas toujours suffisante pour représenter toute la variabilité de la distribution de sortie. Plusieurs indices de sensibilité ont été proposés pour dépasser cette restriction et sont présentés dans la prochaine section.

3.3 Indices de sensibilité basés sur des mesures de dissimilarité

De nombreuses mesures de sensibilité globales ont été proposées en alternative aux indices de Sobol qui sont extrêmement populaires mais dont l'utilisation est inadéquate lorsque la variance ne suffit pas à caractériser toute la distribution de la sortie du modèle. Celles-ci ont été introduites sous des hypothèses différentes et avec des formalismes variés. Dans cette section, on adopte le point de vue proposé par [Da Veiga, 2015] qui permet de définir de nombreux indices de sensibilité sous un formalisme commun. L'idée est de mesurer l'impact d'une entrée X_i via une mesure de dissimilarité $d(\cdot,\cdot)$ entre la distribution de sortie $\mathbb{P}_Y$ et la distribution conditionnelle $\mathbb{P}_{Y|X_i}$:

$$\mathbb{E}[d(\mathbb{P}_Y, \mathbb{P}_{Y|X_i})] \ .$$

Le choix de l'application $d(\cdot,\cdot)$ détermine alors le type d'indice de sensibilité. À titre d'exemple, considérer la mesure suivante

$$d(\mathbb{P}_Y, \mathbb{P}_{Y|X_i}) = (\mathbb{E}[Y] - \mathbb{E}[Y|X_i])^2 \ ,$$

permet de retrouver la définition de l'indice de Sobol d'ordre 1 non normalisé :

$$\tilde{S}_i = \mathbb{V}\mathrm{ar}(\mathbb{E}[Y|X_i]) \ .$$

Dans la suite, trois autres choix de mesures de dissimilarité et les indices de sensibilité associés sont présentés. Les sections 3.3.1 et 3.3.2 portent respectivement sur l'utilisation des fonctions de contraste et de la distance de Cramér Von Mises. La section 3.3.3 quant à elle s'intéresse aux f-divergences de Csiszár dont découle la définition des indices de Borgonovo [Borgonovo, 2007]. Pour plus d'exemples de mesures de dissimilarité, le lecteur intéressé peut se référer à [Da Veiga, 2015].

3.3.1 Indices de sensibilité basés sur des fonctions de contraste

Dans le contexte où le but est d'estimer un paramètre θ de la distribution de la sortie Y, [Fort et al., 2016] propose une approche dite "*goal-oriented*" qui vise à quantifier l'impact de chaque entrée X_i sur ce paramètre d'intérêt. [Fort et al., 2016] adopte pour cela le formalisme des *fonctions de contraste*. Notons Θ l'ensemble des paramètres possibles. Une fonction de contraste associée au paramètre $\theta \in \Theta$ est une application définie par

$$\psi : \left| \begin{array}{ccc} \Theta & \longrightarrow & L^1(\mathbb{P}_Y) \\ \theta & \longmapsto & \psi(\cdot, \theta) \end{array} \right. , \tag{3.6}$$

et telle que le problème d'optimisation suivant admette une unique solution :

$$\underset{\theta \in \Theta}{\text{Min}} \ \mathbb{E}[\psi(\theta, Y)] := \underset{\theta \in \Theta}{\text{Min}} \ \Psi(\theta) \ .$$

Sous l'hypothèse que la fonction de contraste ψ vérifie

$$\mathbb{E}\left[\underset{\theta \in \Theta}{\text{Min}} \ \psi(\theta, Y)\right] < +\infty \ ,$$

[Fort et al., 2016] propose de considérer l'indice de sensibilité suivant

$$S_i^{\psi} \propto \mathbb{E}[d(\mathbb{P}_Y, \mathbb{P}_{Y|X_i})] \in [0, 1] \ , \tag{3.7}$$

où le symbole $\propto$ désigne l'égalité à une constante multiplicative près et où la mesure de dissimilarité $d(\cdot, \cdot)$ est définie par

$$d(\mathbb{P}_Y, \mathbb{P}_{Y|X_i}) = \underset{\theta \in \Theta}{\text{Min}} \ \Psi(\theta) - \underset{\theta \in \Theta}{\text{Min}} \ \mathbb{E}[\psi(Y, \theta)|X_i] \ , \tag{3.8}$$

et où la constante de normalisation est

$$\underset{\theta \in \Theta}{\text{Min}} \ \Psi(\theta) - \mathbb{E}\left[\underset{\theta \in \Theta}{\text{Min}} \ \psi(\theta, Y)\right] \ .$$

Les indices S_i^{ψ} sont compris entre 0 et 1, la valeur 0 étant atteinte lorsque Y et X_i sont indépendants. Par ailleurs, leur définition (3.7) est très permissive et permet d'estimer divers paramètres associés à la distribution de sortie. Par exemple, choisir la fonction de contraste $\psi(y, \theta) = (y-\theta)^2$ permet de retrouver la définition des indices de Sobol d'ordre 1 [Fort et al., 2016]. Cependant l'estimation de ces indices reste un problème difficile. Dans le cas du contraste associé à l'α-quantile, i.e,

$$\psi(\theta) = (y - \theta)(\alpha - \mathbb{1}_{y \geq \theta}) \ ,$$

certains estimateurs ont été proposés par [Maume-Deschamps and Niang, 2018] et [Browne et al., 2017]. D'autres exemples classiques de fonctions de contrastes sont donnés dans [Fort et al., 2016].

3.3.2 Indices de sensibilité basés sur la distance de Cramér Von Mises

Pour quantifier l'impact d'un groupe de paramètre $\mathbf{X_u}$, [Gamboa et al., 2018] définit l'indice de sensibilité suivant

$$S_{2,\mathbf{u}}^{\mathrm{CVM}} \propto \mathbb{E}[d(\mathbb{P}_Y, \mathbb{P}_{Y|\mathbf{X_u}})] \in [0,1] \ , \tag{3.9}$$

où la mesure de dissimilarité $d(\cdot,\cdot)$ est définie par la distance de Cramér Von Mises entre la distribution de sortie $\mathbb{P}_Y$ et la distribution de sortie conditionnelle $\mathbb{P}_{Y|\mathbf{X_u}}$:

$$d(\mathbb{P}_Y, \mathbb{P}_{Y|\mathbf{X_u}}) = \int |F_Y(t) - F_{Y|\mathbf{X_u}}(t)|^2 \mathrm{d}\mathbb{P}_Y(t) \ , \tag{3.10}$$

et où la constante de normalisation est

$$\int F_Y(t)(1 - F_Y(t))\mathrm{d}\mathbb{P}_Y(t) \ .$$

La définition (3.9) des indices de Cramér est très permissive puisqu'elle est licite quelque soit la distribution de sortie. Une valeur nulle pour $S_{2,\mathbf{u}}^{\mathrm{CVM}}$ indique que $\mathbf{X_u}$ n'a aucune influence, et inversement, une valeur proche de 1 signifie que $\mathbf{X_u}$ a un fort impact sur la distrubution de sortie. Les auteurs montrent que les indices d'ordre un $S_{2,i}^{\mathrm{CVM}}$ peuvent être estimés par une méthode de type *Pick-and-freeze* et necessitant seulement $3N$ évaluations du modèle de sortie (indépendamment de la dimension du modèle). Une étude théorique des propriétés statistiques de l'estimateur *pick-freeze* est effectuée dans [Gamboa et al., 2018]. En particulier, sa forte consistance et un théorème central limite sont établis. Plus récemment, [Gamboa et al., 2019] généralise cette méthodologie dans le cadre plus général des espaces métriques séparables. En outre, une autre méthode d'estimation basée sur l'utilisation de U-statistiques et V-statistiques est proposée et un nouveau théorème central limite est démontré. Cette nouvelle approche est un raffinement de la précédente puisqu'elle requiert seulement $2N$ évaluations du modèle.

3.3.3 Les mesures de dépendance de Csizár

Formulation générale. Une grande classe de mesures de dissimilarité est donnée par la famille des divergences de Csizár. Considérons $\phi : [0, +\infty[\to \mathbb{R}$ une fonction convexe telle que $\phi(1) = 0$. Pour tout couple (P, Q) de mesures de probabilité, on définit la divergence de Csizár de P et Q par

$$d_\phi(P|Q) = \int \phi\left(\frac{dP}{dQ}\right) dQ \ , \tag{3.11}$$

où P est supposé absolument continue par rapport à Q. À partir de cette divergence, on peut alors définir un ensemble d'indices de sensibilité de la sortie Y relativement à un groupe d'entrée $\mathbf{X_u}$:

$$\mathbb{E}[d_\phi(\mathbb{P}_Y|\mathbb{P}_{Y|\mathbf{X_u}})] \ . \tag{3.12}$$

Cas absolument continu. Dans le cas particulier où $(Y, \mathbf{X})$ est absolument continu, l'indice (3.12) s'écrit de la façon suivante :

$$\begin{aligned}
\mathrm{CDM}_{\mathbf{u}}^{\phi} &:= \int_{\mathbb{R}^{|\mathbf{u}|}} d_{\phi}(\mathbb{P}_Y | \mathbb{P}_{Y|\mathbf{X}_\mathbf{u}}) f_{\mathbf{X}_\mathbf{u}}(\mathbf{x}_\mathbf{u}) d\mathbf{x}_\mathbf{u} \ , \\
&= \int_{\mathbb{R}^{|\mathbf{u}|+1}} \phi\left(\frac{f_Y(y)}{f_{Y|\mathbf{X}_\mathbf{u}}(y)}\right) f_{Y|\mathbf{X}_\mathbf{u}}(y) f_{\mathbf{X}_\mathbf{u}}(\mathbf{x}_\mathbf{u}) d\mathbf{x}_\mathbf{u} dy \ , \\
&= \int_{\mathbb{R}^{|\mathbf{u}|+1}} \phi\left(\frac{f_Y(y) f_{\mathbf{X}_\mathbf{u}}(\mathbf{x}_\mathbf{u})}{f_{Y,\mathbf{X}_\mathbf{u}}(y, \mathbf{x}_\mathbf{u})}\right) f_{Y,\mathbf{X}_\mathbf{u}}(y, \mathbf{x}_\mathbf{u}) d\mathbf{x}_\mathbf{u} dy \ , \\
&= d_{\phi}(\mathbb{P}_Y \otimes \mathbb{P}_{\mathbf{X}_\mathbf{u}} | \mathbb{P}_{Y,\mathbf{X}_\mathbf{u}}) \ , \qquad (3.13)
\end{aligned}$$

où $|\mathbf{u}|$ désigne le cardinal du sous-ensemble $\mathbf{u}$ de $\{1, \ldots, d\}$. L'indice $\mathrm{CDM}_{\mathbf{u}}^{\phi}$ ainsi construit s'interprète donc comme une mesure de dépendance, i.e, comme une mesure de dissimilarité entre la loi jointe $\mathbb{P}_{Y,\mathbf{X}_\mathbf{u}}$ de Y et $\mathbf{X}_\mathbf{u}$ et la loi produit $\mathbb{P}_Y \otimes \mathbb{P}_{\mathbf{X}_\mathbf{u}}$. On parle alors de *mesure de dépendance de Csizár*. Le théorème 3.1, dont une preuve est disponible dans [Rahman, 2016], réunit les propriétés fondamentales vérifiées par les mesures $\mathrm{CDM}_{\mathbf{u}}^{\phi}$.

Théorème 3.1 (Propriétés fondamentales de $\mathrm{CDM}_{\mathbf{u}}^{\phi}$)**.**

- *L'indice* $CDM_{\boldsymbol{u}}^{\phi}$ *admet les bornes suivantes*

$$0 \leq CDM_{\boldsymbol{u}}^{\phi} \leq \phi(0) + \phi^*(0) \ , \qquad (3.14)$$

où ϕ^* *est définie par* $\phi^*(t) = t\phi(\frac{1}{t})$ *pour* $t > 0$ *et* $\phi^*(0) = \lim_{t \to 0} t\phi(\frac{1}{t})$*. De plus la borne inférieure est atteinte lorsque* Y *et* $\boldsymbol{X}_{\boldsymbol{u}}$ *sont indépendants.*

- *L'indice de sensibilité* $CDM_{\{1,\ldots,d\}}^{\phi}$ *de* Y *relativement à toutes les entrées* $\boldsymbol{X} = (X_1, \ldots, X_d)$ *est*

$$CDM_{\{1,\ldots,d\}}^{\phi} = \phi(0) + \phi^*(0) \ . \qquad (3.15)$$

- *(**Monotonie.**) Soit* $\boldsymbol{v}$ *un sous-ensemble de* $\{1, \ldots, d\}$ *tel que* $\boldsymbol{v} \subset \boldsymbol{u}$*. Alors*

$$CDM_{\boldsymbol{v}}^{\phi} \leq CDM_{\boldsymbol{u}}^{\phi} \ . \qquad (3.16)$$

- *(**Invariance.**) L'indice* $CDM_{\boldsymbol{u}}^{\phi}$ *est invariant par application d'un* C^1*-difféomorphisme à* Y *et* $\boldsymbol{X}_{\boldsymbol{u}}$*.*

Remarques. L'intérêt majeur de ces indices est qu'ils rendent compte de l'influence de la loi de l'entrée $\mathbf{X}$ sur le modèle de sortie de manière très complète puisque la distribution de Y est prise en compte dans son ensemble. Ils surpassent donc les méthodes d'analyse de sensibilité basées sur la variance lorsque celle-ci n'est pas suffisamment représentative de la variabilité de la sortie du modèle. De plus, leur propriété d'invariance par transformation régulière des données est très utile en pratique. Par ailleurs, leur définition est permissive et pour certains choix particuliers de la fonction ϕ, on retrouve l'expression de plusieurs indices de sensibilité proposés dans la littérature dans différents contextes [Rahman, 2016, Da Veiga, 2015]. On peut mentionner par exemple l'*indice d'entropie* (ou *information mutuelle*) [Krzykacz-Hausmann, 2001] qui correspond au cas $\phi(\cdot) = -\log(\cdot)$, ainsi que l'*indice de Borgonovo* [Borgonovo, 2007] qui repose sur la distance en variation totale

$\phi(\cdot) = \frac{1}{2}|\cdot - 1|$. Les indices de Borgonovo [Borgonovo, 2007] gagnent de plus en plus d'attention [Borgonovo and Plischke, 2016] et ont déjà commencé à être utilisé dans des cas pratiques [Rajabi et al., 2015, Pianosi and Wagener, 2016, Xie et al., 2018]. En revanche, leur estimation est un problème complexe, notamment parce que les densités de sortie qui interviennent dans leurs définitions sont inconnues lorsque l'on considère un modèle réaliste. L'objet de la prochaine section est de proposer un état de l'art des différentes méthodes d'estimation des indices de Borgonovo du premier ordre introduites à ce jour.

3.4 Les mesures d'importance δ_i de Borgonovo : un état de l'art

Dans toute cette section on se place dans le cas où $(Y, \mathbf{X})$ est absolument continu. On suppose par ailleurs que la distribution de $\mathbf{X}$ est entièrement connue. Pour illustrer certaines définitions qui vont suivre, on considère le modèle jouet suivant :

$$Y = X_1 + X_2 \ , \tag{3.17}$$

où les variables d'entrée sont indépendantes et distribuées respectivement selon les distributions $N(0,1)$ et $N(0,5)$, le second argument correspondant à la variance.

L'idée de la méthode d'analyse de sensibilité globale introduite par Emanuele Borgonovo [Borgonovo, 2007] est de mesurer à quel point fixer un groupe de variables $\mathbf{X_u}$ à une certaine valeur $\mathbf{x_u}$ modifie la distribution de sortie sur l'ensemble de son support. Dans [Borgonovo, 2007], cette perturbation est quantifiée par le *shift* $s(\mathbf{x_u})$ défini comme la norme L^1 entre la PDF de sortie f_Y la PDF de sortie conditionnée $f_{Y|\mathbf{X_u}=\mathbf{x_u}}$, i.e, l'aire entre leur courbes représentatives respectives (voir figure [3.1])

$$s(\mathbf{x_u}) = \left\| f_Y - f_{Y|\mathbf{X_u}=\mathbf{x_u}} \right\|_{L^1(\mathbb{R})} \ . \tag{3.18}$$

La sensibilité de Y relativement à $\mathbf{X_u}$ est alors définie par la moyenne renormalisée du shift, i.e, l'indice de Borgonovo de $\mathbf{X_u}$ est défini par

$$\delta_{\mathbf{u}} = \frac{1}{2}\mathbb{E}[s(\mathbf{X_u})] \ . \tag{3.19}$$

Comme mentionné dans la section précédente, l'indice de sensibilité $\delta_{\mathbf{u}}$ s'interprète comme mesure dépendance de Csiszár $\mathrm{CDM}^{\phi}_{\mathbf{u}}$ (3.13), avec la fonction $\phi(\cdot) = \frac{1}{2}|\cdot - 1|$. Les indices de Borgonovo héritent donc des propriétés vérifiées par les indices de sensibilité $\mathrm{CDM}^{\phi}_{\mathbf{u}}$. On obtient donc le théorème 3.2 qui découle directement du théorème 3.1.

Théorème 3.2 (Propriétés des indices $\delta_{\mathbf{u}}$).
Soit $\boldsymbol{u}$ un sous-ensemble de $\{1, \dots, d\}$.

- $0 \leq \delta_{\boldsymbol{u}} \leq 1$.
- *L'indice de Borgonovo de toutes les entrées $\boldsymbol{X} = (X_1, \dots, X_d)$ est*

$$\delta_{\{1,\dots,d\}} = 1 \ . \tag{3.20}$$

- ***(Monotonie.)** Soit $\boldsymbol{v} \subset \{1, \dots, d\}$ tel que $\boldsymbol{v} \subset \boldsymbol{u}$. Alors $\delta_{\boldsymbol{v}} \leq \delta_{\boldsymbol{u}}$.*
- ***(Invariance.)** Les indices de Borgonovo sont invariants par application d'un C^1-difféomorphisme sur la sortie et les entrées du modèle.*

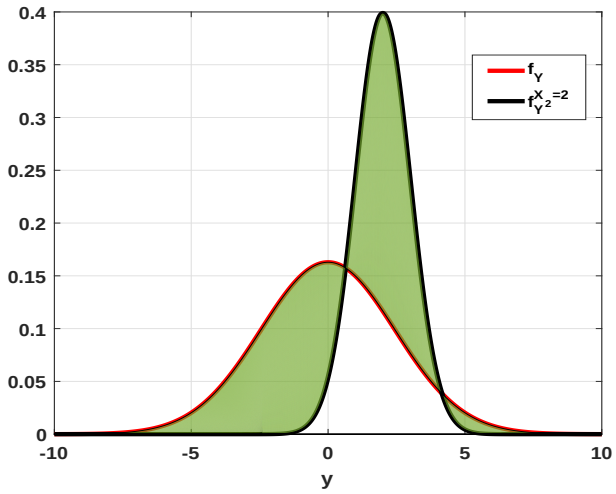


FIGURE 3.1. REPRÉSENTATION LA DENSITÉ f_Y (COURBE ROUGE) DE LA SORTIE DU MODÈLE DÉFINI PAR EQ.(3.17), LA DENSITÉ CONDITIONNÉE $f_Y^{X_2=2}$ (COURBE NOIRE) ET LE SHIFT ASSOCIÉ (AIRE VERTE) ; ICI $s(2) = 1.091$.

L'estimation des indices de Borgonovo $\delta_{\mathbf{u}}$ est une tâche compliquée, et ce, à cause des PDFs f_Y et $f_{Y|\mathbf{X_u}}$ qui interviennent dans leur définitions (3.18) et (3.19). Les prochaines sous-sections se focalisent sur les différentes méthodes d'estimation des indices de Borgonovo du premier ordre δ_i introduites à ce jour, classées en fonction de la définition de δ_i considérée.

3.4.1 δ_i vue comme une espérance

Dans cette section, le point de vue adopté est celui de l'article original [Borgonovo, 2007], i.e, l'indice de Borgonovo de l'entrée X_i est défini par l'espérance renormalisée du shift (3.18) :

$$\delta_i = \frac{1}{2}\mathbb{E}[s(X_i)] \ . \tag{3.21}$$

Les premières méthodologies qui ont été proposées pour estimer l'indice δ_i reposent sur une idée commune, à savoir l'approximation Monte-Carlo de l'espérance (3.21) :

$$\delta_i \approx \frac{1}{2N}\sum_{n=1}^{N} s(X_i^n), \ \{X_i^n\}_{n=1}^N \overset{\text{i.i.d}}{\sim} f_{X_i} \ . \tag{3.22}$$

La principale difficulté réside alors dans l'évaluation des différentes intégrales (aléatoires) $s(X_i^n)$ provenant de la définition (3.18) du shift. Pour cela, plusieurs approches existent et sont listées dans les sous-sections suivantes.

3.4.1.1 La méthode PDF

Formulation. Dans l'article pionnier [Borgonovo, 2007], il est proposé d'estimer les PDFs qui interviennent dans la somme (3.22) en utilisant la méthode du maximum de vraisemblance couplée à un test de Kolmogorov-Smirnov. [Borgonovo et al., 2011] adopte par la suite une approche non-paramétrique, basée sur la procédure d'estimation par

noyau détaillée en section 2.2.2.1. Cette approche (que l'on désignera par méthode PDF dans la suite de ce manuscrit) peut alors être résumée comme suit :

(i) Générer $\{\mathbf{X}^n = (X_1^n, \dots, X_d^n)\}_{n=1}^N \overset{\text{i.i.d}}{\sim} f_{\mathbf{X}}$ puis évaluer $Y^n = \mathcal{M}(\mathbf{X}^n)$ pour $n = 1, \dots, N$.

Ensuite, générer $\{Y^{(i),n}\}_{n=1}^N \overset{\text{i.i.d}}{\sim} f_{Y|X_i=X_i^n}$ pour $n = 1, \dots, N$[2].

(ii) Considérer les estimateurs à noyau de la densité f_Y et des N densités conditionnelles $f_{Y|X_i=X_i^n}$:

$$\hat{f}_Y(y) := \frac{1}{Nh} \sum_{n=1}^{N} K\left(\frac{y - Y^n}{h}\right), \ y \in \mathbb{R}\ , \tag{3.23}$$

$$\hat{f}_{Y|X_i=X_i^n}(y) := \frac{1}{Nh^{(i),n}} \sum_{n=1}^{N} K\left(\frac{y - Y^{(i),n}}{h^{(i),n}}\right), \ y \in \mathbb{R}\ . \tag{3.24}$$

où K est un noyau unidimensionnel et h, $\{h^{(i),n}\}_{n=1}^N$ les fenêtres respectives.

(iii) Estimer δ_i par

$$\hat{\delta}_i^{\text{PDF}} := \frac{1}{2N} \sum_{n=1}^{N} \int \left| \hat{f}_Y - \hat{f}_Y^{X_i=X_i^n} \right| \ , \tag{3.25}$$

où les intégrales sont évaluées par intégration numérique.

Remarques. La méthode PDF nécessite $N(1 + dN)$ appels au modèle de sortie pour obtenir une estimation des indices de Borgonovo d'ordre 1. Elle n'est donc pas adaptée aux systèmes complexes réalistes pour lesquels l'évaluation du code de calcul $\mathcal{M}$ est coûteuse. D'autre part, lorsque les entrées sont corrélées, l'étape **(i)** peut être difficile en pratique puisqu'elle requiert d'être capable d'échantillonner selon $\mathbf{X}|X_i = X_i^n$ ce qui peut s'avérer délicat lorsque l'entrée $\mathbf{X}$ n'est pas gaussienne. Une variante de la méthode PDF, reposant sur le partitionnement de l'ensemble E_i des valeurs prises par l'entrée X_i, est développée par [Plischke et al., 2013]. Étant donnée une partition $\bigsqcup_{m=1}^M E^{(i),m}$ de E_i, l'idée est de remplacer les densités de Y sachant que X_i prenne une valeur ponctuelle, par les densités $f_{Y|E^{(i),m}}$ de Y sachant que X_i appartienne à une des classes $E^{(i),m}$, i.e,

$$f_{Y|E^{(i),m}}(y) = \left(\int_{E^{(i),m}} f_{X_i}(x) dx \right)^{-1} \cdot \int_{E^{(i),m}} f_{X_i,Y}(x, y) dx\ . \tag{3.26}$$

[Plischke et al., 2013] obtient alors l'estimateur suivant :

$$\hat{\delta}_i^{\text{Pseudo}} := \frac{1}{2N} \sum_{m=1}^{M} N_m \int \left| \hat{f}_Y - \hat{f}_{Y|E^{(i),m}} \right| \ , \tag{3.27}$$

[2]Dans le cas où les entrées sont indépendantes, il suffit pour cela de fixer la i-ème composante de $\mathbf{X}^n$ à X_i^n.

où N_m désigne le nombre de réalisations X_i^n appartenant à la classe $E^{(i),m}$ et où $\hat{f}_{Y|E^{(i),m}}$ est l'estimateur à noyau de la densité définie par Eq.(3.26) :

$$\hat{f}_{Y|E^{(i),m}}(y) := \frac{1}{N_m h_m} \cdot \sum_{X_i^n \in E^{(i),m}} K\left(\frac{y - Y^n}{h_m}\right), \ y \in \mathbb{R} \ , \tag{3.28}$$

Cette approche nécessite seulement N évaluations de modèle et des résultats de convergence sont discutés dans [Plischke et al., 2013]. En particulier, les auteurs démontrent que l'estimateur $\hat{\delta}_i^{\text{Pseudo}}$ converge presque sûrement vers δ_i lorsque le nombre d'échantillons N et la taille de la partition M tendent vers $+\infty$. En revanche, cette convergence peut s'avérer très lente. De plus l'application de cette méthode repose sur le choix d'une partition (plusieurs stratégies de partitionnement sont présentées dans [Plischke, 2012]) de E_i ce qui peut être fastidieux.

3.4.1.2 La méthode CDF

Dans [Liu and Homma, 2009], Liu and Homma propose une nouvelle méthode (intitulée "méthode CDF" dans la suite de cette section) pour estimer la mesure δ_i définie par Eq.(3.21). Le principe de la méthode CDF est de réécrire le shift de densité $s(X_i)$ en fonction des CDFs F_Y et $F_{Y|X_i}$. Pour toute réalisation x_i de X_i, le shift $s(x_i)$ peut se réécrire comme suit :

$$\begin{aligned} s(x_i) &= \int_{\Omega(x_i)} |f_Y - f_{Y|X_i=x_i}| + \int_{\Omega(x_i)^c} |f_Y - f_{Y|X_i=x_i}| \ , \\ &= \int_{\Omega(x_i)} (f_Y - f_{Y|X_i=x_i}) - \int_{\Omega(x_i)^c} (f_Y - f_{Y|X_i=x_i}) \ , \end{aligned} \tag{3.29}$$

où $\Omega(x_i)$ désigne l'ensemble $\left\{y \in \mathbb{R} | f_Y(y) \geq f_{Y|X_i=x_i}(y)\right\}$. Pour fixer les idées, supposons que $\Omega(x_i) =]-\infty, a]$, i.e, que f_Y et $f_{Y|X_i=x_i}$ s'intersectent en un unique point $y = a$ et que $f_Y - f_{Y|X_i=x_i}$ soit positive sur $]-\infty, a]$. On obtient alors

$$\begin{aligned} s(x_i) &= \int_{-\infty}^{a} (f_Y - f_{Y|X_i=x_i}) - \int_{a}^{+\infty} (f_Y - f_{Y|X_i=x_i}) \ , \\ &= (F_Y(a) - F_{Y|X_i=x_i}(a)) - ((1 - F_Y(a)) - (1 - F_{Y|X_i=x_i}(a))) \ , \\ &= 2(F_Y(a) - F_{Y|X_i=x_i}(a)) \ . \end{aligned}$$

Plus généralement, supposons que les densités f_Y et $f_{Y|X_i=x_i}$ vérifient les hypothèses suivantes :

(A.1) f_Y et $f_{Y|X_i=x_i}$ admettent un nombre fini $m(x_i)$ de points d'intersection

$$y_1^{(i)} < \cdots < y_{m(x_i)}^{(i)} \ .$$

(A.2) $f_Y - f_{Y|X_i=x_i}$ change de signe en chacun des points $y_k^{(i)}$ et est positive sur $]-\infty, y_1^{(i)}]$.

En réinjectant ces hypothèses dans Eq.(3.29) et en procédant comme dans le cas d'un unique point d'intersection, on obtient[3]

$$s(x_i) = 2 \sum_{k=1}^{m(x_i)} (-1)^{k-1} \times \left(F_Y(y_k^{(i)}) - F_Y^{X_i=x_i}(y_k^{(i)}) \right) \ . \tag{3.30}$$

Un point d'intersection entre f_Y et $f_{Y|X_i}$ est une solution de l'équation suivante

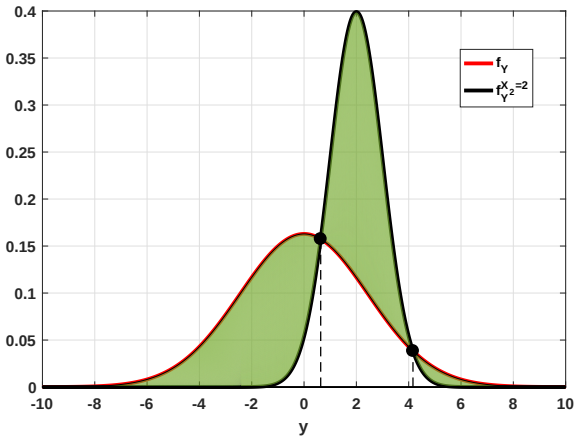
$$f_Y(y) - f_{Y|X_i}(y) = 0 \ ,$$

qui peut se réécrire comme suit

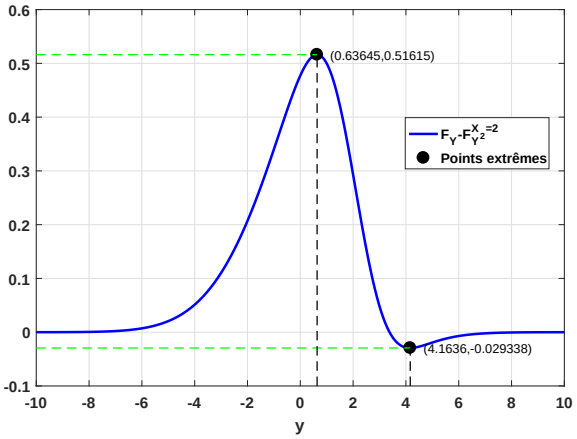
$$\frac{d(F_Y - F_{Y|X_i})}{dy}(y) = 0 \ .$$

Ainsi, les points d'intersection de f_Y et $f_{Y|X_i=x_i}$ correspondent aux points extrémaux de la fonction $F_Y - F_{Y|X_i=x_i}$ et peuvent donc être déterminés en utilisant les méthodes de dérivation numérique classiques (schéma aux différences finies etc.). Dans le cas du modèle jouet (3.17), on peut observer sur la figure [3.2] que les PDFs f_Y et $f_{Y|X_2=2}$ présentent deux points d'intersection $y_1^{(2)} = 0.63645$ et $y_2^{(2)} = 4.1636$ tels que $f_Y(y_1^{(2)}) - f_{Y|X_2=2}(y_1^{(2)}) \geq 0$. Le shift associé $s(2)$ est donc égal à

$$s(2) = 2 \times (0.51615 - (-0.029338)) = 1.091 \ .$$



(a) Représentation la densité f_Y (courbe rouge) de la sortie du modèle défini par Eq.(3.17)et la densité conditionnée $f_Y^{X_2=2}$ (courbe noire).



(b) Représentation de la différence entre la CDF F_Y et la CDF conditionnée $F_Y^{X_2=2}$.

FIGURE 3.2. REPRÉSENTATIONS GÉOMÉTRIQUES DU SHIFT $s(2)$ DU MODÈLE DÉFINI PAR EQ.(3.17).

Pour résumer, les différentes étapes de la méthode CDF sont

(i) Générer $\{\mathbf{X}^n = (X_1^i \ldots, X_d^n)\}_{n=1}^N \overset{\text{i.i.d}}{\sim} f_{\mathbf{X}}$ puis évaluer $Y^n = \mathcal{M}(\mathbf{X}^n)$ pour $n = 1, \ldots, N$.

[3] Si $f_Y - f_{Y|X_i=x_i}$ est négative sur $]-\infty, y_1^{(i)}]$, l'égalité (3.30) diffère d'un facteur -1.

(ii) Considérer les estimateurs à noyau respectifs $\hat{F}_Y$ et $\hat{F}_{Y|X_i=X_i^n}$ de la CDF F_Y et des N CDFs conditionnelles $F_{Y|X_i=X_i^n}$.

(iii) Pour $n = 1, \ldots, N$, déterminer les points extrémaux $\{y_k^{(i)}\}_{k=1}^{m(X_i^n)}$ de la fonction $\hat{F}_Y - \hat{F}_Y^{X_i=X_i^n}$.

(iv) Considérer l'estimateur suivant :

$$\hat{\delta}_i^{\text{CDF}} := \frac{1}{N}\sum_{n=1}^{N} \epsilon_n \times \sum_{k=1}^{m(X_i^n)} (-1)^{k-1} \times \left(\hat{F}_Y(y_k^{(i)}) - \hat{F}_{Y|X_i=X_i^n}(y_k^{(i)})\right) , \tag{3.31}$$

où ϵ_n est le signe de $\hat{F}_Y(y_1^{(i)}) - \hat{F}_{Y|X_i=X_i^n}(y_1^{(i)})$.

Remarques. La méthode CDF à pour avantage de s'épargner l'estimation des densités f_Y et $f_{Y|X_i=X_i^n}$. Néanmoins, elle reste équivalente à la méthode PDF en terme de coût de simulation puisque $N(1 + dN)$ appels au code de calcul $\mathcal{M}$ sont requis pour obtenir une estimation des indices δ_i. De plus, la charge de calcul inhérente à l'estimation des densités est remplacée par la recherche des points extrémaux des fonctions $F_Y - F_{Y|X_i=X_i^n}$ qui n'est pas exempt d'erreurs d'approximation. En effet, les applications numériques effectuées dans [Liu and Homma, 2009] illustrent que l'estimateur δ_i^{CDF} peut présenter une variance plus grande que celle de δ_i^{PDF}.

Par la suite, le problème d'estimation des mesures d'importance δ_i a été abordé dans de nombreux travaux de recherche. En outre, l'objectif est de diminuer le budget de simulation exigé par les méthodes PDF et CDF tout en conservant une estimation suffisamment précise des indices pour obtenir le *ranking* correct, i.e, le classement des entrées en fonction de leur influence sur la sortie Y. La section suivante présente les méthodologies d'estimation basées sur la définition de δ_i en terme de mesure de dépendance de Csiszár.

3.4.2 δ_i vue comme une mesure de dépendance

Dans cette section, on considère la définition de δ_i en terme de mesure de dépendance de Csiszár. Eq.(3.13) s'écrit ici

$$\delta_i = \frac{1}{2}\left\| f_{X_i} f_Y - f_{X_i,Y} \right\|_{L_1(\mathbb{R}^2)} . \tag{3.32}$$

Cette interprétation ouvre la voie à diverses procédures d'estimation qui n'impliquent plus l'estimation de densités conditionnelles puisque seules la densité de sortie f_Y et la densité jointe $f_{X_i,Y}$ doivent être évaluées. Ces méthodes sont décrites dans les sections qui suivent.

3.4.2.1 La méthode *simple-boucle*

Les méthodes PDF et CDF sont des méthodes dites *double-boucle*, dans le sens où elles traitent le problème d'estimation de δ_i comme l'évaluation de deux intégrales. C'est pourquoi elles nécessitent un budget de simulation non raisonnable en pratique. Dans le but de réduire cette charge de calcul, Wei et al. [Wei et al., 2013] proposent une approche dîtes *simple-boucle*. Les différentes étapes de cette méthodes sont les suivantes :

(i) Générer $\{\mathbf{X}^n = (X_1^n, \ldots, X_d^n)\}_{n=1}^N \overset{\text{i.i.d}}{\sim} f_{\mathbf{X}}$ puis évaluer $Y^n = \mathcal{M}(\mathbf{X}^n)$ pour $n = 1, \ldots, N$.

(ii) Utiliser l'échantillon $\{\mathbf{X}^n, Y^n)\}$ pour estimer les PDFs f_Y et $f_{X_i,Y}$ via la procédure d'estimation par noyau :

$$\hat{f}_Y(y) := \frac{1}{Nh} \sum_{n=1}^{N} K\left(\frac{y - Y^n}{h}\right), \; y \in \mathbb{R}, \tag{3.33}$$

$$\hat{f}_{X_i,Y}(x,y) := \frac{1}{Nh_1h_2} \sum_{n=1}^{N} K\left(\frac{x - X_i^n}{h_1}\right) K\left(\frac{y - Y^n}{h_2}\right), \; (x,y) \in \mathbb{R}^2 \,, \tag{3.34}$$

où K est le noyau gaussien $K(t) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2)$ et où les fenêtres h, h_1 et h_2 sont calculées à l'aide de la toolbox développée par Botev et al. [Botev et al., 2010].

(iii) Estimer δ_i par

$$\hat{\delta}_i^{\text{SL}} = \frac{1}{2N} \sum_{n=1}^{N} \left| \frac{f_{X_i}(X_i^n)\hat{f}_Y(Y^n)}{\hat{f}_{X_i,Y}(X_i^n, Y^n)} - 1 \right| . \tag{3.35}$$

Remarques. Cette méthode requiert seulement N appels à la fonction boîte noire $\mathcal{M}$ pour estimer l'ensemble des indices de Borgonovo du premier ordre, contre $N(1 + dN)$ pour les méthodes *double-boucle*. Cependant, l'estimateur souffre de certains problèmes théoriques.

Premièrement, on peut noter que l'échantillon $\{\mathbf{X}^n, Y^n\}$ est utilisé à la fois pour la simulation Monte-Carlo et pour la procédure d'estimation par noyau. En conséquence, la loi forte des grands nombres ne s'applique plus puisque les variables de la somme donnée par Eq.(3.35) ne sont pas i.i.d.

D'autre part, quand bien même un nouvel échantillon $\{\mathbf{U}^k = (U_1^k, U_2^k)\}_{k=1}^{N'} \sim f_{X_i,Y}$ indépendant de $\{\mathbf{X}^n, Y^n\}_{n=1}^N$ serait généré, on obtiendrait

$$\begin{aligned} \frac{1}{N'} \sum_{k=1}^{N'} \left| \frac{f_{X_i}(U_1^k)\hat{f}_Y(U_2^k)}{\hat{f}_{X_i,Y}(\mathbf{U}^k)} - 1 \right| &\approx \int \frac{|f_{X_i} \times \hat{f}_Y - \hat{f}_{X_i,Y}|}{\hat{f}_{X_i,Y}} \times f_{X_i,Y} \\ &\neq \int |f_{X_i} \times \hat{f}_Y - \hat{f}_{X_i,Y}| \,. \end{aligned}$$

La méthode *simple-boucle* sort donc du cadre classique de l'échantillonnage préférentiel présenté en section 2.2.3.2 et il n'est donc pas clair que l'estimateur $\hat{\delta}_i^{\text{SL}}$ soit consistant. L'introduction de la section 4.2 permettra d'étayer cette affirmation en appliquant la méthode *simple-boucle* au modèle jouet (3.17).

3.4.2.2 Méthode basée sur la transformation de Nataf

Cette section détaille les différentes étapes de la méthode d'estimation introduite par [Zhang et al., 2014].

Estimation de f_Y. En vertu de la propriété d'invariance de l'indice δ_i, il peut être supposé sans perte de généralité que Y est positive. La PDF f_Y est alors estimée en

utilisant le principe du maximum d'entropie (cf. section 2.2.2.2) associé à une collection de n moments fractionnaires de Y :

$$\left\{\mu_Y^t := \mathbb{E}[Y^{\beta_t}]\right\}_{t=1}^n \ , \tag{3.36}$$

où $\beta_1 < \cdots < \beta_n$. La distribution de sortie étant inconnue, il en va de même pour les contraintes (3.36). Afin d'estimer les moments μ_Y^t, [Zhang et al., 2014] impose les hypothèses suivantes :

(A.1) Les variables d'entrées X_i sont indépendantes.

(A.2) Le modèle $\mathcal{M}$ est approché en utilisant la technique de réduction de modèle cut-HDMR (cf. section 2.5.2).

Sous **(A.2)**, la sortie Y peut être approchée par un produit de fonctions unidimensionnelles :

$$Y \approx (\mathcal{M}(\boldsymbol{\mu}_{\mathbf{X}}))^{(1-d)} \prod_{i=1}^{d} \mathcal{M}(X_i, \boldsymbol{\mu}_{\mathbf{X}}) \ , \tag{3.37}$$

où $\boldsymbol{\mu}_{\mathbf{X}}$ désigne la moyenne de l'entrée $\mathbf{X}$ et où $\mathcal{M}(X_i, \boldsymbol{\mu}_{\mathbf{X}})$ est l'évaluation de $\mathcal{M}$ en $\boldsymbol{\mu}_{\mathbf{X}}$ excepté la i-ème composante qui est remplacée par X_i. Grâce à l'hypothèse **(A.1)**, l'estimation des moments μ_Y^t est alors réduite à l'évaluation d'intégrales unidimensionnelles :

$$\hat{\mu}_Y^t := (\mathcal{M}(\boldsymbol{\mu}_{\mathbf{X}}))^{(1-d)\beta_t} \prod_{i=1}^{d} \left(\int \mathcal{M}(x_i, \boldsymbol{\mu}_{\mathbf{X}})^{\beta_t} f_{X_i}(x_i) dx_i \right) \ . \tag{3.38}$$

Dans [Zhang et al., 2014], le calcul de ces intégrales est effectué via une méthode de quadrature de Gauss à N_q points, de sorte que seulement $dN_q + 1$ appels au code $\mathcal{M}$ sont nécessaires pour calculer l'ensemble des contraintes (3.36).

Estimation de la densité jointe $f_{X_i,Y}$. La transformation de Nataf de la paire (X_i, Y) est définie par

$$\mathbf{R}_i = (r_i(X_i), r_y(Y)) = \left(\Phi^{-1}(F_{X_i}(X_i)), \Phi^{-1}(F_Y(Y))\right) \ ,$$

où Φ est la CDF de la distribution normale centrée réduite $N(0,1)$. Par construction, les marginales $r_i(X_i)$ et $r_y(Y)$ suivent la loi $N(0,1)$ mais sans hypothèses supplémentaires, $\mathbf{R}_i$ n'est pas nécessairement un vecteur gaussien. En outre, d'après le théorème de changement de variable, la distribution de $\mathbf{R}_i$ est donnée par

$$f_{\mathbf{R}_i}(r_i(x), r_y(y)) = f_{X_i,Y}(x,y) \frac{\phi(r_i(x))\phi(r_y(y))}{f_{X_i}(x) f_Y(y)} \ , \tag{3.39}$$

où ϕ désigne la PDF de la distribution $N(0,1)$. Néanmoins, [Zhang et al., 2014] font l'hypothèse suivante :

(A.3) La relation suivante est vérifiée :

$$f_{X_i,Y}(x,y) = f_{X_i}(x) f_Y(y) \frac{\phi_2(r_i(x), r_y(y))}{\phi(r_i(x))\phi(r_y(y))} \ , \tag{3.40}$$

où ϕ_2 désigne la PDF de la distribution $N_2(0_{\mathbb{R}^2}, \boldsymbol{\rho}_{0,i})$ où $\boldsymbol{\rho}_{0,i}$ est la matrice de corrélation du couple $(r_i(X_i), r_y(Y))$. En vertu de l'égalité (3.39), l'hypothèse **(A.3)** est équivalente au

fait d'exiger que $\mathbf{R}_i$ soit un vecteur gaussien de matrice de corrélation $\boldsymbol{\rho}_{0,i}$. Par construction, $\boldsymbol{\rho}_{0,i}$ est donnée par

$$\boldsymbol{\rho}_{0,i} = \begin{pmatrix} 1 & \rho_{0,i} \\ \rho_{0,i} & 1 \end{pmatrix} ,$$

où $\rho_{0,i}$ est la covariance de $(r_i(X_i), r_y(Y))$. Ainsi, la construction d'un estimateur de la densité jointe se résume à l'estimation du coefficient $\rho_{0,i}$. Par ailleurs, le théorème de changement de variable permet d'établir un lien entre $\rho_{0,i}$ et le coefficient de variation ρ_i de (X_i, Y)

$$\rho_i = \frac{\mathrm{Cov}(X_i, Y)}{\sigma_i \sigma_Y} = \frac{\mathbb{E}[X_i Y] - \mathbb{E}[X_i]\mathbb{E}[Y]}{\sqrt{\mathbb{E}[Y^2] - (\mathbb{E}[Y])^2} \cdot \sigma_i} , \tag{3.41}$$

où σ_i, σ_Y sont les écarts-type respectifs de X_i et Y. Considérons la décomposition de Cholesky de la matrice définie positive $\boldsymbol{\rho}_{0,i}$:

$$\boldsymbol{\rho}_{0,i} = \boldsymbol{L}_0 \boldsymbol{L}_0^T ,$$

où $\boldsymbol{L}_0$ est la matrice triangulaire inférieure, de sorte que $\boldsymbol{L}_0^{-1}\mathbf{R}_i \sim N_2(0, I_2)$. On obtient alors

$$\begin{aligned} \rho_i &= \int\int \left(\frac{x_i - \mathbb{E}[X_i]}{\sigma_i}\right)\left(\frac{y - \mathbb{E}[Y]}{\sigma_Y}\right) f_{X_i,Y}(x_i, y) dx_i dy , \\ &= \int\int \left(\frac{F_{X_i}^{-1}(\Phi(r_i)) - \mathbb{E}[X_i]}{\sigma_i}\right)\left(\frac{F_Y^{-1}(\Phi(r_y)) - \mathbb{E}[Y]}{\sigma_Y}\right) \phi_2(r_i, r_y) dr_i dr_y , \\ &= \int\int \left(\frac{F_{X_i}^{-1}(\Phi(u_i)) - \mathbb{E}[X_i]}{\sigma_i}\right)\left(\frac{F_Y^{-1}(\Phi(u_y)) - \mathbb{E}[Y]}{\sigma_Y}\right) \phi(u_i)\phi(u_y) du_y du_y , \end{aligned} \tag{3.42}$$

où $(u_i, u_y)^T = \boldsymbol{L}_0^{-1}(r_i, r_y)^T$. Ainsi, $\rho_{0,i}$ peut être calculé à partir d'un algorithme de recherche des zéros d'une fonction (non linéaire), comme la méthode de Newton-Raphson.

Formulation. Sous l'hypothèse **(A.3)**, le théorème de changement de variable permet de réinterpréter l'indice δ_i comme l'espérance suivante [Zhang et al., 2014]

$$\delta_i = \mathbb{E}\left[\left|\frac{\phi(r_i(X_i))\phi(r_y(Y))}{\phi_2(r_i(X_i), r_y(Y))} - 1\right|\right] . \tag{3.43}$$

Pour conclure, la procédure complète introduite par [Zhang et al., 2014] peut être résumée comme suit :

(i) Calculer les poids $\{\{z_{i,k}\}_{k=1}^{N_q}\}_{i=1}^{d}$ correspondant à la formule de quadrature de Gauss considérée pour estimer les intégrales intervenant dans Eq.(3.38). Ensuite, calculer $\mathcal{M}(\mu_{\mathbf{X}})$ et les $d \times N_q$ termes $\{\{\mathcal{M}(z_{i,k}, \mu_{\mathbf{X}})\}_{k=1}^{N_q}\}_{i=1}^{d}$.

(ii) **Estimation de f_Y.** Considérer n nombres réels $\beta_1 < \cdots < \beta_n$ et les moments fractionnaires associés $\{\mu_Y^t\}$. Calculer les contraintes approchées $\hat{\mu}_Y^t$ données par

Eq.(3.38) en utilisant la grille d'intégration établie à l'étape **(i)**. Ensuite, déterminer le minimum (λ_t^*) de la fonction suivante :

$$(\lambda_t)_{t=1}^n \longmapsto \sum_{t=1}^n \lambda_t \hat{\mu}_Y^t + \log\left(\int \exp\left(-\sum_{t=1}^n \lambda_t y^{\beta_t} dy\right)\right) .$$

Considérer l'estimateur ME de f_Y :

$$\hat{f}_Y(y) \propto \exp\left(-\sum_{t=1}^n \lambda_t^* y^{\beta_t}\right) .$$

(iii) Calcul du coefficient ρ_i. Calculer ρ_i à partir de Eq.(3.41). Pour cela, calculer $\mathbb{E}[Y]$ et $\mathbb{E}[Y^2]$ en considérant $\beta_t = 1$ et $\beta_t = 2$ dans Eq.(3.38). De même, calculer $\mathbb{E}[X_i Y]$ à l'aide de l'approximation de modèle donnée par Eq.(3.37) :

$$\mathbb{E}[X_i Y] \approx (\mathcal{M}(\boldsymbol{\mu}_{\mathbf{X}}))^{(1-d)} \int x_i \mathcal{M}(x_i, \boldsymbol{\mu}_{\mathbf{X}}) f_{X_i}(x_i) dx_i \times \prod_{j \neq i}^{d} \left(\int \mathcal{M}(x_j, \boldsymbol{\mu}_{\mathbf{X}}) f_{X_j}(x_j) dx_j\right) . \tag{3.44}$$

(iv) Calcul de $\boldsymbol{\rho}_{0,i}$. Considérer $\{\theta_k\}_{k=1}^{N_q}$ et $\{\omega_k\}_{k=1}^{N_q}$ les poids et les nœuds d'intégration associés à la méthode de Gauss-Hermite. Déterminer la racine de la fonction suivante (approximation de Eq.(3.42))

$$\rho_{0,i} \longmapsto \rho_i - \sum_{k=1}^{N_q}\sum_{l=1}^{N_q} \omega_k \omega_l \left(\frac{x_{i,k} - \mathbb{E}[X_i]}{\sigma_i}\right)\left(\frac{y_l - \mathbb{E}[Y]}{\sigma_Y}\right) ,$$

où

$$(x_{i,k}, y_l) = \left(F_{X_i}^{-1}(\Phi(r_{i,k})), \hat{F}_Y^{-1}(\Phi(r_{y,l}))\right) , (r_{i,k}, r_{y,l})^T = \boldsymbol{L}_0(\theta_k, \theta_l)^T ,$$

avec $\hat{F}_Y(y) = \int_{-\infty}^{y} \hat{f}_Y(u)du$ et $\boldsymbol{L}_0$ la matrice triangulaire inférieure de la décomposition de Choslesky de la matrice $\begin{pmatrix} 1 & \rho_{0,i} \\ \rho_{0,i} & 1 \end{pmatrix}$.

(v) Générer $\{(R_{i,1}^k, R_{i,2}^k)\}_{k=1}^N \overset{\text{i.i.d}}{\sim} N_2(0, \boldsymbol{\rho}_{0,i})$ puis estimer δ_i par

$$\hat{\delta}_i^{\text{Nataf}} = \frac{1}{2N}\sum_{k=1}^{N} \left| \frac{\phi(R_{i,1}^k)\phi(R_{2,1}^k)}{\phi_2(R_{i,1}^k, R_{i,2}^k)} - 1 \right| . \tag{3.45}$$

Remarques. Cette approche d'estimation est très compétitive puisqu'elle fait intervenir très peu d'appels au modèle $\mathcal{M}$. Ceci est dû au fait que les hypothèses **(A.1)** et **(A.2)** permettent de réduire le calcul des contraintes (3.36) à l'évaluation déterministe d'intégrales unidimensionnelles (voir Eq.(3.38)). Le nombre d'appels au code $\mathcal{M}$ est donc égal à $dN_q + 1$ où N_q est l'ordre de la méthode de quadrature de Gauss utilisée pour estimer ces intégrales. Néanmoins, les hypothèses sur lesquelles repose cette approche sont restrictives.

L'hypothèse **(A.3)** peut être évitée en appliquant la méthode proposée par [Yun et al., 2018] qui repose sur des techniques similaires. Il s'agit d'une méthode *double-boucle* basée sur la définition de δ_i en terme d'espérance. En outre, la densité de sortie et les PDFs de sortie conditionnelles sont simultanément estimées à l'aide du principe du maximum d'entropie. Le calcul des contraintes associées est effectué sous les hypothèses **(A.1)** et **(A.2)**, ce qui assure un budget de simulation identique à la méthode reposant sur l'utilisation de la transformation de Nataf.

On peut également se soustraire à l'hypothèse d'indépendance **(A.1)** en calculant différemment les contraintes

$$\mu_Y^t = \int_{\mathbb{R}^d} \mathcal{M}(\mathbf{x})^{\beta_t} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \ t = 1, \ldots, n$$

La malédiction de la dimension qui affecte les méthodes de quadratures classiques rend difficile l'évaluation de ces intégrales d-dimensionnelles. Ceci peut être surmonté dans une certaine mesure par la technique d'intégration SGI (acronyme anglais de *Sparse Grid integration*) qui est basée sur le produit tensoriel de Smolyak [Gerstner and Griebel, 1998, Zhang et al., 2015] et dont le but est de diminuer le nombre de nœuds d'intégration considérés. Considérant une suite de formules de quadrature unidimensionnelle (non nécessairement imbriquées) pour une fonction g à une dimension

$$Q_l^1 g = \sum_{i=1}^{N_l^1} \omega_{i,l} g(z_{i,l}) \ ,$$

on définit la formule de quadrature suivante

$$\Delta_k^1 g := (Q_k^1 - Q_{k-1}^1) g \ \text{ avec } \ Q_0^1 g := 0 \ ,$$

et le produit tensoriel de d formules de quadrature $\left(Q_{l_1}^1 \otimes \cdots \otimes Q_{l_d}^1\right)$ par

$$\left(Q_{l_1}^1 \otimes \cdots \otimes Q_{l_d}^1\right) f = \sum_{i_1=1}^{N_{l_1}^1} \cdots \sum_{i_d=1}^{N_{l_d}^1} \omega_{i_1,l_1} \cdots \omega_{i_d,l_d} f(z_{i_1,l_1}, \ldots, z_{i_d,l_d}) \ ,$$

où f est une fonction à d dimensions. Alors, la formule de Smolyak avec un l niveaux de précision est définie par

$$Q_l^d f := \sum_{|\mathbf{k}|_1 \leq l+d-1} \left(\Delta_{k_1}^1 \otimes \cdots \otimes \Delta_{k_d}^1\right) f \ , \tag{3.46}$$

où la somme porte sur le simplexe

$$\left\{ \sum_{i=1}^{d} k_i \leq l + d - 1; \mathbf{k} = (k_1, ..., k_d) \in \mathbb{N}^d \right\} \ .$$

L'étude de la convergence et de l'erreur de la méthode SGI est disponible dans [Gerstner and Griebel, 1998]. En outre, estimer les moments fractionnaires μ_Y^t avec cette approche requiert

$$\sum_{|\mathbf{k}|_1 \leq l+d-1} N_{k_1}^1 \cdots N_{k_d}^1 \ , \tag{3.47}$$

appels au code de calcul $\mathcal{M}$. À titre d'exemple, pour un modèle comportant $d = 6$ entrées et pour une suite (Q_k^1) de formules de quadrature imbriquées d'ordres $N_k^1 = k$, la technique SGI avec $l = 2$ (respectivement $l = 3$ et $l = 4$) niveaux de précision nécessite seulement 13 (respectivement 31 and 91) évaluations de la fonction $\mathcal{M}$. En revanche, cette technique est complexe à implémenter en pratique puisqu'il n'existe pas de formule générale explicite pour calculer les poids d'intégration qui interviennent dans la formule de Smolyak donnée par Eq.(3.46) [Gerstner and Griebel, 1998].

3.4.3 Représentation de δ_i en terme de densité de copule

La mesure de sensibilité δ_i peut également être réinterprétée en utilisant le formalisme de la théorie des copules. En effet, le vecteur (X_i, Y) étant supposé absolument continu, le théorème de Sklar (2.1) assure qu'il existe une unique copule $C_i(\cdot,\cdot)$ de dimension 2 telle que

$$F_{X_i,Y}(x_i, y) = C_i(F_{X_i}(x_i), F_Y(y)) \ . \tag{3.48}$$

En posant

$$(u_i, v) = (F_{X_i}(x_i), F_Y(y)) \ , \tag{3.49}$$

et en dérivant la relation (3.48), on obtient

$$f_{X_i,Y}(x_i, y) = c_i(u_i, v) f_{X_i}(x_i) f_Y(y) \ , \tag{3.50}$$

où $c_i(\cdot,\cdot)$ désigne la densité de copule du couple (X_i, Y)

$$c_i(u_i, v) = \frac{\partial^2 C_i(u_i, v)}{\partial u_i \partial v} \ ,$$

i.e, la densité du couple $(U_i, V) := (F_{X_i}(X_i), F_Y(Y))$ (voir figure [3.3] pour un exemple illustré).

Ainsi, en effectuant le changement de variable (3.49) on obtient la représentation suivante de l'indice de Borgonovo :

$$\delta_i = \frac{1}{2} \int_{[0,1]^2} |c_i(u_i, v) - 1| du_i dv \ . \tag{3.51}$$

Remarque. La définition de δ_i en terme de densité de copule (3.51) a été introduite pour la première fois par [Wei et al., 2014]. Il n'existe pas méthode d'estimation basée sur cette définition, mise à part celle proposée par [Wei et al., 2014] qui repose essentiellement sur l'estimation par noyau gaussien de la densité de copule c_i. Cependant, le domaine d'étude étant le carré unité $[0, 1]^2$, des effets de bord peuvent survenir (cf. section 2.2.2.1) et impacter fortement la précision de cette estimation. Cette remarque motive la section 4.3 qui décrit une méthode d'estimation robuste basée sur la représentation de δ_i en terme de densité de copule.

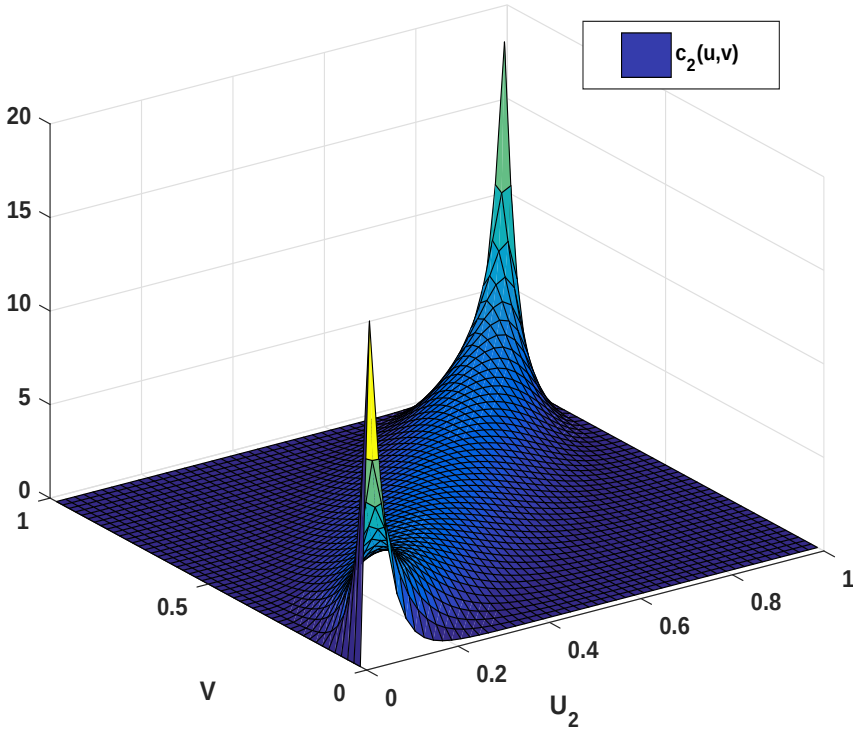


FIGURE 3.3. REPRÉSENTATION DE LA DENSITÉ DE COPULE DU COUPLE (X_2, Y) DU MODÈLE JOUET (3.17).

3.5 Conclusion

L'objet de ce chapitre était d'introduire les principes de l'analyse de sensibilité des modèles entrée-sortie. Une attention toute particulière a été portée à la présentation des méthodes d'analyse de sensibilité globales quantitatives. L'une des méthodes les plus utilisées par les praticiens est celle reposant sur l'estimation des indices de Sobol. Cette approche présente cependant l'inconvénient de concentrer l'étude sur la variance de la sortie. En outre, diverses alternatives ont été proposées pour gérer le cas où la variance n'est plus suffisamment représentative de la distribution de sortie. Celles-ci peuvent être regroupées dans la famille des indices de sensibilité construits à partir d'une mesure de dissimilarité. La fin du chapitre s'est focalisé sur la classe des mesures de dépendance de Csiszár, et plus particulièrement sur les indices de Borgonovo δ_i. De par leurs propriétés, ces indices ont suscité ces derniers temps une attention croissante dans la communauté de l'analyse de sensibilité. En particulier, le problème de leur estimation à été abordé dans de nombreux travaux de recherche. Un état de l'art des différentes méthodologies d'estimation à été présenté et a permis de souligner les avantages et les inconvénients de chacune. Initialement, les indices de Borgonovo étaient très peu utilisés puisque les premières méthodes introduites, qualifiées de *double-boucle*, exigeaient un nombre d'appel à la boîte noire non raisonnable en pratique. La méthode GSA de Borgonovo a commencé à être appliquée à des cas d'application concrets avec l'arrivée des nouvelles approches d'estimation basées sur la réinterprétation de δ_i en terme de mesure de dépendance. Celles-ci ont permis de réduire le budget de simulation des méthodes *double-boucle* mais souffrent cependant de problèmes théoriques qui peuvent affecter leur champ d'application

ou la précision de leur estimations. L'objectif de cette thèse est de proposer de nouvelles manières d'estimer les mesures d'importance de Borgonovo, avec comme contraintes de surmonter les inconvénients précités et de respecter un compromis entre l'optimisation du nombre d'appel au modèle et l'obtention d'estimations suffisamment précises pour obtenir le bon ranking. Ainsi, dans le prochain chapitre, deux nouvelles méthodes d'estimations des indices de Borgonovo du premier ordre sont présentées.

Chapitre 4

Contribution à l'estimation des indices de Borgonovo d'ordre 1

Contents

4.1 Introduction 71
4.1.1 Motivations 71
4.1.2 Présentation des cas tests analytiques 72
4.2 Estimateur $\delta_i^{\mathrm{IS},g}$ combinant échantillonnage préférentiel et approximation par noyau gaussien 74
4.2.1 Retour sur la méthode *simple-boucle* 74
4.2.2 Définition de l'estimateur $\delta_i^{\mathrm{IS},g}$ 76
4.2.3 Étude des performances de $\delta_i^{\mathrm{IS},g}$ 77
4.2.4 Application à des cas test analytiques 86
4.3 Estimateur δ_i^{ME} basé sur la représentation de δ_i en terme de densité de copule et le principe du maximum d'entropie . . 88
4.3.1 Choix des contraintes 88
4.3.2 Définition de δ_i^{ME} 90
4.3.3 Application à des cas test analytiques 92
4.4 Conclusion : application à un modèle d'estimation de la zone de retombée d'un étage de lanceur 95
4.4.1 Description du cas test 95
4.4.2 Résultats numériques 96
4.4.3 Synthèse 98

4.1 Introduction

4.1.1 Motivations

Ce chapitre se focalise sur la question de l'estimation des indices de Borgonovo du premier ordre (dont l'état de l'art a été effectué en section 3.4) et s'organise comme suit.

La section 4.2 présente un schéma d'estimation général combinant des procédures d'échantillonnage préférentiel et d'estimation par noyau gaussien et est basée sur la publication suivante :

Derennes, P., Morio, J., and Simatos, F. (2018) *A nonparametric importance sampling estimator for moment independent importance measures*. Accepted in Reliability Engineering and System Safety, https://doi.org/10.1016/j.ress.2018.02.009

La section 4.3 présente une méthode d'estimation basée sur la représentation de δ_i en terme de densité de copule (3.51) et sur l'estimation de cette copule par le principe du maximum d'entropie (voir section 2.2.2.2). Cette partie est basée sur l'article de conférence suivant :

Derennes, P., Morio, J., and Simatos, F. (9-12 December 2018) *Estimation of the moment independent importance sampling measures using copula and maximum entropy framework*, Winter Simulation Conference, Gothenburg, Sweden, 1623-1634.

Dans le but d'étudier la performance de ces deux nouvelles méthodes, celles-ci seront appliquées aux différents cas test analytiques décrits dans la section 4.1.2. Enfin, la section 4.4 fournit une conclusion résumant les avantages et les inconvénients des deux méthodes, ainsi qu'un cas d'application concret portant sur l'estimation de la zone de retombée d'un étage de lanceur et provenant du chapitre de livre suivant :

P. Derennes, V. Chabridon, J. Morio, M. Balesdent, F. Simatos, J-M. Bourinet and M. Gayton (2019) *Nonparametric importance sampling techniques for sensitivity analysis and reliability assessment of a launcher stage fallout*, Optimization in Space Engineering, G. Fasano and J. Pinter, Springer.

4.1.2 Présentation des cas tests analytiques

Tout au long de ce chapitre, les méthodes introduites seront appliquées à divers cas test analytiques, définis par les modèles boîte noire entrée-sortie suivants :

Modèle jouet.

$$\mathcal{M}_1 : \left| \begin{array}{rcl} \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ (x_1, x_2) & \longmapsto & x_1 + x_2 \end{array} \right. \tag{4.1}$$

$$Y = X_1 + X_2 \quad \text{avec} \quad \mathbf{X} = (X_1, X_2) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix} \right) .$$

Ainsi,

$$Y \sim N(0, 6), \ (Y \mid X_1 = x_1) \sim N(x_1, 5) \ \text{ et } \ (Y \mid X_2 = x_2) \sim N(x_2, 1) .$$

Modèle linéaire gaussien.

$$\mathcal{M}_2 : \left| \begin{array}{rcl} \mathbb{R}^4 & \longrightarrow & \mathbb{R} \\ \mathbf{x} & \longmapsto & \langle \mathbf{A}, \mathbf{x} \rangle_{\mathbb{R}^4} \end{array} \right. \tag{4.2}$$

$$Y = \langle \mathbf{A}, \mathbf{X} \rangle_{\mathbb{R}^4} ,$$

où $\mathbf{A} = [1,7 \quad 1,8 \quad 1,9 \quad 2]$ et où l'entrée $\mathbf{X}$ suit la distribution gaussienne centrée et de matrice de covariance

$$\Sigma = \begin{pmatrix} 1 & 1/2 & 1/3 & 1/4 \\ 1/2 & 1 & 1/2 & 1/3 \\ 1/3 & 1/2 & 1 & 1/2 \\ 1/4 & 1/3 & 1/2 & 1 \end{pmatrix} .$$

Les résultats classiques de la théorie des vecteurs gaussiens (voir par exemple [Ouvrard, 2000]) permettent de déterminer les distributions de sortie inconditionnelle et conditionnelles :

$$Y \sim N(0, \mathbf{A}\Sigma\mathbf{A}^T) \quad \text{et} \quad Y|X_i = x_i \sim N\left(m_i, \sigma_i^2\right) ,$$

où la moyenne m_i et la variance σ_i^2 sont définies par

$$m_i = A_i x_i + \mathbf{A}_{-i}\mathbf{C}_i\Sigma_{ii}^{-1}x_i ,$$

$$\sigma_i^2 = \mathbf{A}_{-i}(\Sigma_{-i} - \mathbf{C}_i\mathbf{C}_i^T\Sigma_{ii}^{-1})\mathbf{A}_{-i}^T ,$$

où $\mathbf{A}_{-i}$ désigne le vecteur $\mathbf{A}$ privé de la i-ème composante, Σ_{-i} la matrice Σ privée de ses i-ème ligne et colonne et $\mathbf{C}_i$ le vecteur colonne $[\Sigma_{ij}]_{j\neq i}$.

Modèle multiplicatif.

$$\mathcal{M}_3^d : \left| \begin{array}{ccc} \mathbb{R}^d & \longrightarrow & \mathbb{R} \\ \mathbf{x} & \longmapsto & \prod_{i=1}^d x_i \end{array} \right. \tag{4.3}$$

$$Y = \prod_{i=1}^{d} X_i, \text{ avec } X_i \overset{\text{i.i.d}}{\sim} L(0,1) .$$

Ainsi,

$$Y \sim L(0,d) \quad \text{et} \quad (Y|X_i = x_i) \sim L(\ln(x_i), d-1) .$$

Ces trois modèles entrée-sortie sont définis par des fonctions analytiques et des distributions d'entrée pour lesquelles il est possible de déterminer les distributions de sortie. A fortiori, les valeurs théoriques des indices de Borgonovo δ_i sont accessibles par intégration numérique. Dès lors, un indicateur de précision supplémentaire pour un estimateur $\hat{\delta}_i$ est l'*écart relatif* (*Relative Difference* (RD) en anglais)

$$\mathrm{RD}(\hat{\delta}_i) = \frac{\hat{\delta}_i - \delta_i}{\delta_i} , \tag{4.4}$$

en plus de la moyenne et de l'écart-type traditionnellement utilisés. En outre, chacune des futures applications numériques se déroulera selon la procédure suivante :

1. Calcul de M estimations $(\hat{\delta}_i^1, ..., \hat{\delta}_i^M)$ indépendantes.

2. Calcul des estimateurs non biaisés respectifs de la moyenne et de l'écart-type (voir section 2.2.3) :

$$\bar{\delta}_i = \frac{1}{M}\sum_{k=1}^{M}\hat{\delta}_i^k \quad \text{et} \quad \bar{\sigma}_{\hat{\delta}_i} := \sqrt{\frac{1}{M-1}\sum_{k=1}^{M}(\hat{\delta}_i^k - \bar{\delta}_i)^2},$$

ce qui procure l'approximation suivante du coefficient de variation

$$\mathrm{CV}(\hat{\delta}_i) \approx \frac{\bar{\delta}_i}{\bar{\sigma}_{\hat{\delta}_i}} \ ,$$

3. Calcul de l'écart relatif moyen

$$\mathrm{RD}(\bar{\delta}_i) = \frac{\bar{\delta}_i - \delta_i}{\delta_i} \ ,$$

lorsque la valeur théorique de δ_i est connue.

4.2 Estimateur $\delta_i^{\mathbf{IS},g}$ combinant échantillonnage préférentiel et approximation par noyau gaussien

Cette section présente une nouvelle méthode d'estimation des indices de Borgonovo d'ordre 1 δ_i, inspirée de celle de la *simple-boucle* détaillée dans la section 3.4.2.1 et repose donc sur la définition de δ_i en terme de mesure de dépendance :

$$\delta_i = \frac{1}{2}\left\| f_{X_i} f_Y - f_{X_i,Y} \right\|_{L_1(\mathbb{R}^2)} \ . \tag{4.5}$$

Cette nouvelle approche est motivée par la section 4.2.1 qui montre, par le biais de simulations numériques sur le modèle jouet (4.1), que l'estimateur *simple-boucle* (3.35) peut fournir de mauvaises estimations, et propose des explications pour cette imprécision. Partant du postulat que ce défaut de précision est dû à un problème de ré-échantillonnage, la section 4.2.2 décrit un nouvel estimateur δ_i^{IS} basé sur les procédures d'échantillonnage préférentiel et d'estimation par noyau gaussien. Les propriétés théoriques de δ_i^{IS} ainsi que l'influence de la distribution d'échantillonnage sur les performances de δ_i^{IS} sont étudiées dans la section 4.2.3. Enfin, la section 4.2.4 présente plusieurs applications numériques dans le but d'illustrer le gain de précision apporté par cette nouvelle approche d'estimation.

4.2.1 Retour sur la méthode *simple-boucle*

Dans cette section, la méthode *simple-boucle* est appliquée au modèle jouet $\mathcal{M}_1$ défini par Eq.(4.1). La table 4.1 répertorie les résultats obtenus après $M = 100$ simulations, et ce, avec $N = 5000$ échantillons à chaque exécution de l'algorithme *simple-boucle* décrit dans la section 3.4.2.1. Avec des coefficients de variation estimés autour de 1% et 5%, il semble licite de penser que les estimateurs $\hat{\delta}_i^{\mathrm{SL}}$ ont convergé et devraient donc fournir une estimation précise de δ_i. Cette intuition est renforcée par la figure [4.1] illustrant une

TABLE 4.1. ESTIMATIONS DES MESURES δ_i PAR LA MÉTHODE *simple-boucle* DANS LE CAS DU MODÈLE JOUET $\mathcal{M}_1$ (4.1) ; [PARAMÈTRES : $N = 5000$ ET $M = 100$].

Entrée	**Valeur théorique** δ_i	$\bar{\delta}_i^{\mathrm{SL}}$	$\mathrm{RD}(\bar{\delta}_i^{\mathrm{SL}})$	$\bar{\sigma}_i^{\mathrm{SL}}$	$\mathrm{CV}(\hat{\delta}_i^{\mathrm{SL}})$
X_1	0.1436	0.1217	-0.1522	0.0055	0.0449
X_2	0.5382	0.3297	-0.3873	0.0046	0.0138

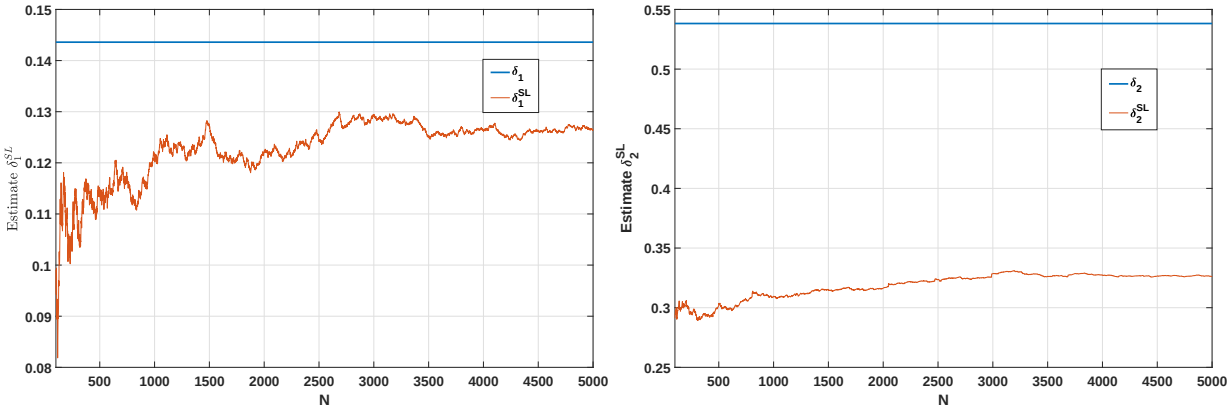
trajectoire des deux estimateurs *simple-boucle* à mesure que la taille N de l'échantillon augmente : dans les deux cas, la trajectoire s'aplatit et semble avoir convergé.

Toutefois, ces estimateurs présentent un biais significatif. En effet, puisque les valeurs théoriques des mesures δ_i sont disponibles, l'intervalle de confiance de niveau $(1-\alpha)$ (2.29) de $\mathbb{E}(\hat{\delta}_i^{\mathrm{SL}})$ et l'écart relatif (4.4) peuvent être calculés dans le but d'apprécier le biais de l'estimateur $\hat{\delta}_i^{SL}$. En outre, il apparait que les deux estimateurs admettent de fortes incertitudes relatives de l'ordre de 15% et 40%, notamment $\bar{\delta}_2^{\mathrm{SL}}$ qui sous-estime fortement l'indice δ_2. De plus, les valeurs de référence $[\delta_1, \delta_2] = [0.1436\ 0.5382]$ n'appartiennent pas aux intervalles de confiance à 95%

$$\delta_1 \notin \left[0.1217 \pm \frac{1.96 \times 0.0055}{\sqrt{100}}\right] = [0.1206, 0.1228] \ ,$$

et

$$\delta_2 \notin \left[0.3297 \pm \frac{1.96 \times 0.0046}{\sqrt{100}}\right] = [0.3288, 0.3306] \ .$$



(a) Une trajectoire de l'estimateur δ_1^{SL} de δ_1. (b) Une trajectoire de l'estimateur δ_2^{SL} de δ_2.

FIGURE 4.1. ESTIMATIONS DES INDICES δ_i VIA LA MÉTHODE *simple-boucle* EN FONCTION DU NOMBRE N D'ÉCHANTILLONS.

Ainsi, la méthode *simple-boucle* peut produire de mauvaises estimations dans un cas simple où la procédure d'estimation par noyau est parfaitement adaptée puisque les distributions de sortie sont gaussiennes. Au vu des présents résultats numériques et des problèmes théoriques soulignés dans la section 3.4.2.1, il n'est pas clair que δ_i^{SL} soit convergeant. En particulier, le postulat motivant la suite de cette section est que l'utilisation d'un échantillon commun pour la réalisation des procédures d'estimation par noyau et

de Monte-Carlo est une des causes de cette non consistance. À cet égard, la prochaine section introduit une nouvelle approche d'estimation impliquant un ré-échantillonnage via une distribution auxiliaire.

4.2.2 Définition de l'estimateur $\delta_i^{\mathbf{IS},g}$

Dans cette section, un nouveau schéma d'estimation des indices δ_i est présenté. Celui-ci est sensiblement identique à celui de la méthode *simple-boucle*, mis à part que la simulation Monte-Carlo est remplacée par une étape d'échantillonnage préférentiel.

Étape IS1. Générer $\{\mathbf{X}^n = (X_1^n, \dots, X_d^n)\}_{n=1}^N \overset{\text{i.i.d}}{\sim} f_{\mathbf{X}}$ puis évaluer $Y^n = \mathcal{M}(\mathbf{X}^n)$ pour $n = 1, \dots, N$.

Étape IS2. Utiliser l'échantillon $\{\mathbf{X}^n, Y^n\}$ pour estimer les PDFs f_Y et $f_{X_i,Y}$ via la procédure d'estimation par noyau :

$$\hat{f}_Y(y) := \frac{1}{Nh} \sum_{n=1}^{N} K\left(\frac{y - Y^n}{h}\right), \ y \in \mathbb{R}, \tag{4.6}$$

$$\hat{f}_{X_i,Y}(x,y) := \frac{1}{Nh_1h_2} \sum_{n=1}^{N} K\left(\frac{x - X_i^n}{h_1}\right) K\left(\frac{y - Y^n}{h_2}\right), \ (x,y) \in \mathbb{R}^2 \ , \tag{4.7}$$

où K est le noyau gaussien $K(t) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2)$ et où les fenêtres h, h_1 et h_2 sont calculées à l'aide de l'estimateur *plug-in* de [Botev et al., 2010].

Étape IS3. Calculer l'estimateur par tirages préférentiels

$$\delta_i^{\mathrm{IS},g} = \frac{1}{2N'} \sum_{k=1}^{N'} \frac{\left|\hat{f}_Y(U_2^k) f_{X_i}(U_1^k) - \hat{f}_{X_i,Y}(\mathbf{U}^k)\right|}{g(\mathbf{U}^k)}, \ \{\mathbf{U}^k = (U_1^k, U_2^k)\}_{k=1}^{N'} \overset{\text{i.i.d}}{\sim} g \ , \tag{4.8}$$

où g est une distribution d'échantillonnage définie sur $\mathbb{R}^2$ et pouvant dépendre de l'échantillon $\{X_i^n, Y^n\}$.

Remarques. L'approche proposée conserve les avantages de la méthode *simple-boucle*. D'une part, seulement N appels à la fonction boîte noire $\mathcal{M}$ sont requis pour estimer l'ensemble des indices de Borgonovo du premier ordre. D'autre part, cette procédure est applicable aux modèles présentant des entrées corrélées. Les propriétés théoriques de ce schéma d'estimation et la question du choix de g seront discutées dans la prochaine section. En outre, pour motiver la prochaine section, notons que dans le cas où $g = \hat{f}_{X_i,Y}$ l'estimateur $\delta_i^{\mathrm{IS},g}$ est donné par

$$\hat{\delta}_i^{\mathrm{IS},\hat{f}_{X_i,Y}} = \frac{1}{2N'} \sum_{k=1}^{N'} \frac{|f_{X_i}(U_1^k)\hat{f}_Y(U_2^k) - \hat{f}_{X_i,Y}(\mathbf{U}^k)|}{\hat{f}_{X_i,Y}(\mathbf{U}^k)} \ .$$

Cette expression est similaire à la définition (3.35) de l'estimateur *simple-boucle* $\hat{\delta}_i^{\mathrm{SL}}$. Néanmoins, une différence notable est que deux échantillons différents, à savoir $\{\mathbf{U}^k\}$ et $\{\mathbf{X}^n, Y^n\}$, sont utilisés dans l'étape d'estimation par noyau **IS2** et l'étape de tirage préférentiel **IS3**. Cette différence clef est au cœur de la démonstration de la consistance de l'estimateur $\hat{\delta}_i^{\mathrm{IS},g}$ établie dans la section suivante.

4.2.3 Étude des performances de $\delta_i^{\mathbf{IS},g}$

4.2.3.1 Propriétés théoriques de $\delta_i^{\mathbf{IS},g}$

La performance de l'estimateur $\hat{\delta}_i^{\mathrm{IS},g}$ repose en partie sur le choix de la distribution auxiliaire g. En particulier, une hypothèse essentielle est que le support de g contienne celui de l'intégrande :

$$\mathrm{Supp}(|f_{X_i}f_Y - f_{X_i,Y}|) \subset \mathrm{Supp}(g) \ , \tag{A1}$$

Lorsque l'hypothèse (A1) est vérifiée conjointement aux hypothèses classiques (2.19) portant sur les fenêtres h, h_1 et h_2 qui assurent la convergence des estimateurs à noyau $\hat{f}_Y$ et $\hat{f}_{X_i,Y}$, cela assure que $\hat{\delta}_i^{\mathrm{IS},g}$ soit asymptotiquement non biaisé et convergeant lorsque $N' \to \infty$ puis $N \to \infty$.

Proposition 4.1. *Supposons que l'hypothèse* (A1) *soit vérifiée et que* $Nh, Nh_1h_2 \to +\infty$ *avec* $h, h_1h_2 \to 0$. *Alors on a les convergence suivantes :*

$$\delta_i^{IS,g} \xrightarrow[N'\to+\infty]{\mathbb{P}-p.s} \hat{\Delta} \xrightarrow[N\to+\infty]{\mathbb{P}-p.s} \delta_i \ , \tag{4.9}$$

et

$$\sup_{N'} \mathbb{E}(\hat{\delta}_i^{IS,g}) \xrightarrow[N\to+\infty]{} \delta_i \ .$$

Démonstration. Soit

$$\hat{\Delta} = \frac{1}{2}\left\| f_{X_i}\hat{f}_Y - \hat{f}_{X_i,Y} \right\|_{L_1(\mathrm{Supp}(g))} .$$

Alors $\mathbb{E}(\hat{\delta}_i^{\mathrm{IS},g}) = \mathbb{E}(\hat{\Delta})$, indépendamment de N', et la loi forte des grand nombres implique que

$$\delta_i^{\mathrm{IS},g} \xrightarrow[N'\to+\infty]{\mathbb{P}(\cdot|\{X^n,Y^n\})-\mathrm{p.s}} \mathbb{E}[\delta_i^{\mathrm{IS},g}|\{X^n,Y^n\}] = \hat{\Delta} \ .$$

En outre, on a

$$\mathbb{P}\left(\omega \in \Omega \ ; \ \delta_i^{\mathrm{IS},g}(\omega) \xrightarrow[N'\to+\infty]{} \hat{\Delta}(\omega)|\{X^n,Y^n\}\right) = 1 \ .$$

Dès lors,

$$\mathbb{P}\left(\omega \in \Omega \ ; \ \delta_i^{\mathrm{IS},g}(\omega) \xrightarrow[N'\to+\infty]{} \hat{\Delta}(\omega)\right) = \mathbb{E}\left[\mathbb{P}\left(\omega \in \Omega \ ; \ \delta_i^{\mathrm{IS},g}(\omega) \xrightarrow[N'\to+\infty]{} \hat{\Delta}(\omega)|\{X^n,Y^n\}\right)\right] = 1 \ .$$

Ainsi, on obtient la convergence suivante

$$\delta_i^{\mathrm{IS},g} \xrightarrow[N'\to+\infty]{\mathbb{P}-\mathrm{p.s}} \hat{\Delta} \ .$$

De plus, les hypothèses sur les fenêtres impliquent que

$$\hat{\Delta} \xrightarrow[N\to+\infty]{\mathrm{p.s}} \delta_i \ ,$$

en vertu du résultat (2.20) démontré dans [Devroye and Gyorfi, 1985, Théorème 1]. Cela montre le premier point du théorème, à partir duquel le deuxième découle par application du théorème de convergence dominé puisque $\hat{\Delta} \leq 1$. □

Remarques. Puisque les variables $\mathbf{U}^k$ sont i.i.d conditionnellement aux variables $\{\mathbf{X}^n, \mathbf{Y}^n\}$, le théorème de la variance totale donne la décomposition suivante de la variance de l'estimateur $\hat{\delta}_i^{\mathrm{IS},g}$

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}\left(\hat{\delta}_i^{\mathrm{IS},g}\right) \\
&= \mathbb{V}\mathrm{ar}\left(\mathbb{E}\left[\hat{\delta}_i^{\mathrm{IS},g}|\{\mathbf{X}^n,\mathbf{Y}^n\}\right]\right) + \mathbb{E}\left[\mathbb{V}\mathrm{ar}\left(\hat{\delta}_i^{IS,g}|\{\mathbf{X}^n,\mathbf{Y}^n\}\right)\right] \ , \\
&= \mathbb{V}\mathrm{ar}\left(\mathbb{E}\left[\frac{1}{2N'}\sum_{k=1}^{N'}\hat{h}(\mathbf{U}^k)|\{\mathbf{X}^n,\mathbf{Y}^n\}\right]\right) + \mathbb{E}\left[\mathbb{V}\mathrm{ar}\left(\frac{1}{2N'}\sum_{k=1}^{N'}\hat{h}(\mathbf{U}^k)|\{\mathbf{X}^n,\mathbf{Y}^n\}\right)\right] \ , \\
&= \mathbb{V}\mathrm{ar}\left(\frac{1}{2N'}\sum_{k=1}^{N'}\mathbb{E}\left[\hat{h}(\mathbf{U}^k)|\{\mathbf{X}^n,\mathbf{Y}^n\}\right]\right) + \frac{1}{4(N')^2}\mathbb{E}\left[\sum_{k=1}^{N'}\mathbb{V}\mathrm{ar}\left(\hat{h}(\mathbf{U}^k)|\{\mathbf{X}^n,\mathbf{Y}^n\}\right)\right] \ , \\
&= \mathbb{V}\mathrm{ar}\left(\hat{\Delta}\right) + \frac{1}{4N'}\mathbb{E}\left[\mathbb{V}\mathrm{ar}\left(\hat{h}(\mathbf{U}) \mid \{\mathbf{X}^n,\mathbf{Y}^n\}\right)\right] \ , \qquad (4.10)
\end{aligned}$$

avec

$$\hat{h}(x,y) = 1_{g(x,y)>0}\frac{\left|\hat{f}_Y(y)f_{X_i}(x) - \hat{f}_{X_i,Y}(x,y)\right|}{g(x,y)} \ .$$

Cette décomposition met en évidence les deux types d'erreurs commises lors de l'estimation de $\hat{\delta}_i^{IS,g}$: le premier terme $\mathbb{V}\mathrm{ar}(\hat{\Delta})$ correspond à l'erreur induite par la procédure d'estimation par noyau (étape **IS2**) et le second terme à l'erreur induite par l'échantillonnage préférentiel (étape **IS3**).

D'après [Devroye and Gyorfi, 1985, Théorème 1] et le théorème de convergence dominé de Lebesgue, le premier terme tend vers 0 lorsque N tend vers $+\infty$. En ce qui concerne le second terme, la variance suivante

$$\mathbb{V}\mathrm{ar}\left(\hat{h}(\mathbf{U}) \mid \{\mathbf{X}^n,\mathbf{Y}^n\}\right) = \int \frac{\left|\hat{f}_Y f_{X_i} - \hat{f}_{X_i,Y}\right|^2}{g} - \left(\int \left|\hat{f}_Y f_{X_i} - \hat{f}_{X_i,Y}\right|\right)^2 \ ,$$

peut être infinie si la distribution g est mal choisie. La question du choix de la distribution auxiliaire est abordée dans la section suivante. Néanmoins, il peut d'ores et déjà être mentionné que si g est presque proportionnelle à l'intégrande $\left|\hat{f}_Y f_{X_i} - \hat{f}_{X_i,Y}\right|$, alors le second terme d'erreur peut être négligé grâce au facteur $\frac{1}{N'}$, et ce, sans appels supplémentaires à la boîte noire $\mathcal{M}$.

Nous terminons cette section théorique avec une étude plus détaillée de la consistance de $\hat{\delta}_i^{IS,g}$. Le résultat de convergence donné par la proposition 4.1 n'est pas complètement satisfaisant puisque en pratique, pour un budget de simulation donné, celui-ci doit être réparti entre l'étape d'estimation par noyau **IS2** et l'étape de tirage préférentiel **IS3**. En d'autres termes, les paramètres N et N' peuvent être amenés à augmenter conjointement, tandis que la convergence $\hat{\delta}_i^{IS,g} \xrightarrow[N',N\to+\infty]{\text{p.s}} \delta_i$ sous-entend que la majorité du budget est attribué à l'étape **IS3**. En supposant que la distribution g ne dépend pas de l'échantillon $\{\mathbf{X}^n,\mathbf{Y}^n\}$, les résultats suivants fournissent des résultats partiels dans le cas où N' est fonction de N.

Proposition 4.2. *Supposons que :*

1. *f_{X_i} est bornée ;*
2. *f_Y et $f_{X_i,Y}$ sont bornées et de classe C^2.*
3. *La condition (A1) est vérifiée et g est à support borné ;*
4. *Les fenêtres vérifient $h_1 = h_2 = h$, $h \to 0$ et $h \gg \ln\ln N\sqrt{\ln N/N}$, i.e.,*
$$\frac{(\ln\ln N)^2 \ln N}{Nh^2} \xrightarrow[N\to\infty]{} 0. \tag{4.11}$$
5. *$N' = N'(N)$ dépend de N de telle sorte que $N' \to \infty$ lorsque $N \to \infty$.*

Alors $\hat{\delta}_i^{IS,g} \xrightarrow[N\to+\infty]{p.s} \delta_i$.

Démonstration. La preuve de la proposition 4.2 repose sur le théorème 7 de [Hansen, 2008], qui dans notre contexte se réécrit comme suit :

Théorème 4.1 (Théorème 7 de [Hansen, 2008]). *Supposons que f_{X_i} est bornée, que f_Y et $f_{X_i,Y}$ sont bornées et de classe C^2 et que la condition (4.11) est vérifiée avec $h = h_1 = h_2$. Alors, pour tout $q > 0$ on a*

$$\sup_{|y|\le c(N)} \left|\hat{f}_Y(y) - f_Y(y)\right| = O\left(\left(\frac{\ln N}{Nh}\right)^{1/2} + h^2\right) := O(\epsilon_1(N)) \ ,$$

$$\sup_{\|(x,y)\|\le c(N)} \left|\hat{f}_{X_i,Y}(x,y) - f_{X_i,Y}(x,y)\right| = O\left(\left(\frac{\ln N}{Nh^2}\right)^{1/2} + h^2\right) := O(\epsilon_2(N)) \ .$$

presque sûrement, où

$$c(N) = O\left(\ln\ln N\sqrt{\ln N}N^{1/(2q)}\right) \ .$$

Posons

$$h(x,y) = 1_{g(x,y)>0}\frac{|f_Y(y)f_{X_i}(x) - f_{X_i,Y}(x,y)|}{g(x,y)}, \ x,y \in \mathbb{R}^2.$$

Puisque

$$\frac{1}{2N'}\sum_{k=1}^{N'} h(\mathbf{U}^k) \xrightarrow[N\to+\infty]{\text{p.s}} \delta_i$$

en vertu de la loi forte des grands nombres, il suffit de montrer que

$$\hat{\delta}_i^{IS,g} - \frac{1}{2N'}\sum_{k=1}^{N'} h(\mathbf{U}^k) \xrightarrow[N\to+\infty]{\text{p.s}} 0.$$

L'inégalité triangulaire donne

$$\left|\hat{\delta}_i^{IS,g} - \frac{1}{2N'}\sum_{k=1}^{N'} h(\mathbf{U}^k)\right|$$
$$\le \frac{1}{2N'}\sum_{k=1}^{N'}\frac{1}{g(\mathbf{U}^k)}\left|\left|f_{X_i}(U_1^k)\hat{f}_Y(U_2^k) - \hat{f}_{X_i,Y}(\mathbf{U}^k)\right| - \left|f_{X_i}(U_1^k)f_Y(U_2^k) - f_{X_i,Y}(\mathbf{U}^k)\right|\right| \ ,$$

et en utilisant la deuxième inégalité triangulaire $||a|-|b|| \leq |b-a|$ et une nouvelle fois l'inégalité triangulaire on obtient

$$\left|\hat{\delta}_i^{IS,g} - \frac{1}{2N'}\sum_{k=1}^{N'} h(\mathbf{U}^k)\right| \leq \frac{1}{2N'}\sum_{k=1}^{N'} \frac{f_{X_i}(U_1^k)}{g(\mathbf{U}^k)}\left|\hat{f}_Y(U_2^k) - f_Y(U_2^k)\right| + \frac{1}{2N'}\sum_{k=1}^{N'} \frac{1}{g(\mathbf{U}^k)}\left|\hat{f}_{X_i,Y}(\mathbf{U}^k) - f_{X_i,Y}(\mathbf{U}^k)\right|. \quad (4.12)$$

Puisque g à un support borné, les variables aléatoires $\mathbf{U}^k$ sont bornées et le théorème 4.1 s'applique. En outre,il existe des constantes $C_1, C_2 < \infty$ telles que

$$\left|\hat{f}_Y(U_2^k) - f_Y(U_2^k)\right| \leq C_1\epsilon_1(N), \ \left|\hat{f}_{X_i,Y}(\mathbf{U}^k) - f_{X_i,Y}(\mathbf{U}^k)\right| \leq C_2\epsilon_2(N)$$

ce qui donne

$$\left|\hat{\delta}_i^{IS,g} - \frac{1}{2N'}\sum_{k=1}^{N'} h(\mathbf{U}^k)\right| \leq \frac{C_1\epsilon_1(N)}{2N'}\sum_{k=1}^{N'} \frac{f_{X_i}(U_1^k)}{g(\mathbf{U}^k)} + \frac{C_2\epsilon_2(N)}{2N'}\sum_{k=1}^{N'} \frac{1}{g(\mathbf{U}^k)}.$$

D'autre part, on a

$$\frac{1}{N'}\sum_{k=1}^{N'} \frac{f_{X_i}(U_1^k)+1}{g(\mathbf{U}^k)} \xrightarrow[N\to+\infty]{\text{p.s}} \int_{\text{Supp}(g)} (f_{X_i}(x)+1)\mathrm{d}x\mathrm{d}y \ .$$

Comme Supp(g) est borné et $\epsilon_1(N), \epsilon_2(N) \to 0$, le résultat est démontré.

□

Proposition 4.3. *Supposons que :*

1. f_{X_i} *est bornée ;*
2. f_Y *et* $f_{X_i,Y}$ *sont bornées et de classe* C^2.
3. *Il existe* $q > 2$ *tel que*

$$\int |y|^{2q} f_Y(y)\mathrm{d}y < \infty \quad \textit{and} \quad \int \|(x,y)\|^{2q} f_{X_i,Y}(x,y)\mathrm{d}x\mathrm{d}y < \infty \ ;$$

4. g *a un support borné et il existe* $\alpha < 1$ *tel que* $\int g^\alpha < \infty$ *;*
5. *Les fenêtres vérifient* $h_1 = h_2 = h$, $h \to 0$ *et* (4.11).
6. $N' = N'(N)$ *dépend de* N *de telle sorte qu'il existe* $\beta > \frac{2}{1-\alpha}$ *vérifiant*

$$(N')^\beta \frac{\sqrt{\ln N}}{\sqrt{N}h} \to 0.$$

Alors $\hat{\delta}_i^{IS,g} \xrightarrow[N\to+\infty]{p.s} \delta_i$.

Démonstration. Pour démontrer la proposition 4.3 on a besoin de l'extension suivante du théorème 7 de [Hansen, 2008]. La preuve de ce résultat est la même que [Hansen, 2008, Théorème 7] mais utilise [Hansen, 2008, Théorème 5] au lieu de [Hansen, 2008, Théorème 3].

Théorème 4.2 (Extension du théorème 7 de [Hansen, 2008]). *Supposons que f_{X_i} est bornée, que f_Y et $f_{X_i,Y}$ sont bornées et de classe C^2 et qu'il existe $q > 2$ tel que*

$$\int \|y\|^{2q} f_Y(y)\mathrm{d}y < \infty \quad \textit{and} \quad \int \|(x,y)\|^{2q} f_{X_i,Y}(x,y)\mathrm{d}x\mathrm{d}y < \infty \ ,$$

et que la condition (4.11) soit vérifiée avec $h = h_1 = h_2$. Alors

$$\sup_{y\in\mathbb{R}} \left|\hat{f}_Y(y) - f_Y(y)\right| \leq O\left(\epsilon_1(N)\right) \ ,$$

$$\sup_{(x,y)\in\mathbb{R}^2} \left|\hat{f}_{X_i,Y}(x,y) - f_{X_i,Y}(x,y)\right| \leq O\left(\epsilon_2(N)\right) \ ,$$

presque sûrement.

À partir de (4.12), on obtient

$$\begin{aligned}\left|\hat{\delta}_i^{IS,g} - \frac{1}{2N'}\sum_{k=1}^{N'} h(\mathbf{U}^k)\right| &\leq \frac{C_1\epsilon_1(N)}{2N'}\sum_{k=1}^{N'}\frac{f_{X_i}(U_1^k)}{g(\mathbf{U}^k)} + \frac{C_2\epsilon_2(N)}{2N'}\sum_{k=1}^{N'}\frac{1}{g(\mathbf{U}^k)} \ , \\ &\leq \frac{C(\epsilon_1(N)+\epsilon_2(N))}{2N'}\sum_{k=1}^{N'}\frac{1}{g(\mathbf{U}^k)} \ .\end{aligned}$$

Ici on a

$$\frac{1}{N'}\sum_{k=1}^{N'}\frac{1}{g(\mathbf{U}^k)} \xrightarrow[N\to+\infty]{\text{p.s}} \int_{\mathrm{Supp}(g)} \mathrm{d}x\mathrm{d}y \ ,$$

qui est infinie puisque $\mathrm{Supp}(g)$ est non borné. Toutefois, pour $\beta > \frac{2}{1-\alpha}$ on a

$$\sup_{n\geq 1}\frac{1}{n^{1+\beta}}\sum_{k=1}^{n}\frac{1}{g(\mathbf{U}^k)} < \infty \text{ p.s.} \tag{4.13}$$

En effet, pour tout $\eta > 0$ on a

$$\mathbb{P}\left(\frac{1}{n^{1+\beta}}\sum_{k=1}^{n}\frac{1}{g(\mathbf{U}^k)} \geq \eta\right) \leq n\mathbb{P}\left(\frac{1}{g(U)} \geq \eta n^{\beta}\right) \ ,$$

et l'inégalité de Markov implique

$$\mathbb{P}\left(\frac{1}{n^{\beta}}\sum_{k=1}^{n}\frac{1}{g(\mathbf{U}^k)} \geq \eta\right) \leq n\frac{1}{(\eta n^{\beta})^{1-\alpha}}\mathbb{E}\left[\left(\frac{1}{g(U)}\right)^{1-\alpha}\right] = \frac{1}{\eta^{1-\alpha}n^{\beta(1-\alpha)-1}}\int g^{\alpha}.$$

Par hypothèse on a $\int g^{\alpha} < \infty$ and $\beta(1-\alpha) - 1 > 1$, de sorte que

$$\sum_{n\geq 1}\mathbb{P}\left(\frac{1}{n^{1+\beta}}\sum_{k=1}^{n}\frac{1}{g(\mathbf{U}^k)} \geq \eta\right) < \infty$$

ce qui entraine, par le lemme de Borel–Cantelli, que $\frac{1}{n^{1+\beta}}\sum_{k=1}^{n}\frac{1}{g(\mathbf{U}^k)} \xrightarrow[n\to+\infty]{\text{p.s}} 0$, et en particulier (4.13) est vérifiée. À une constante multiplicative près, on obtient alors

$$\left|\hat{\delta}_i^{IS,g} - \frac{1}{2N'}\sum_{k=1}^{N'} h(\mathbf{U}^k)\right| \leq C(\epsilon_1(N)+\epsilon_2(N))(N')^{\beta}$$

et l'hypothèse sur N' entraine que cette borne tend vers 0. □

4.2.3.2 Influence de la distribution g

Cette section porte sur l'étude de l'influence de la distribution d'échantillonnage g intervenant à l'étape **IS3**. Pour cela, les étapes **IS1** et **IS2** sont appliquées au modèle jouet (4.1) avec un échantillon $\{\mathbf{X}^n, Y^n\}$ de taille $N = 5000$, procurant ainsi une approximation $\frac{1}{2}\int |f_{X_i}\hat{f}_Y - \hat{f}_{X_i,Y}|$ de l'indice δ_i. Ensuite, $M = 100$ réalisations de l'étape **IS3** sont effectuées avec $N' = 2\times 10^4$ et à partir du panel de distributions auxiliaires suivant :

1. L'estimateur à noyau
$$g_1 = \hat{f}_{X_i,Y} \ ,$$
de la distribution jointe de X_i et Y.

2. La distribution produit
$$g_2(x,y) = f_{X_i}(x)\times \hat{f}_Y(y) \ ,$$
où $\hat{f}_Y$ est l'estimateur à noyau de la PDF de sortie.

3. La PDF g_3 de la distribution gaussienne $N(\hat{\mu}_i, \hat{\Sigma}_i)$ où
$$\hat{\mu}_i = (\hat{\mu}_{X_i}^{\text{MC}}, \hat{\mu}_Y^{\text{MC}}) = \left(\frac{1}{N}\sum_{n=1}^{N} X_i^n, \frac{1}{N}\sum_{k=1}^{N} Y^n\right) \ ,$$
et
$$\hat{\Sigma}_i = \frac{1}{N-1}\begin{pmatrix} \sum_{n=1}^{N}(X_i^n - \hat{\mu}_{X_i}^{\text{MC}})^2 & \sum_{n=1}^{N}(X_i^n - \hat{\mu}_{X_i}^{\text{MC}})(Y^n - \hat{\mu}_Y^{\text{MC}}) \\ \sum_{n=1}^{N}(X_i^n - \hat{\mu}_{X_i}^{\text{MC}})(Y^n - \hat{\mu}_Y^{\text{MC}}) & \sum_{n=1}^{N}(Y^n - \hat{\mu}_Y^{\text{MC}})^2 \end{pmatrix} \ ,$$
sont les estimateurs Monte-Carlo classiques de la moyenne et de la matrice de covariance respectivement.

4. La distribution uniforme g_4 sur le domaine fourni par la toolbox développée par Botev et al. [Botev et al., 2010] utilisée à l'étape **IS2**. À noter toutefois que ce domaine contient l'échantillon $\{\mathbf{X}^n, Y^n\}$ mais que g_4 ne vérifie pas la condition (A1) si l'entrée X_i n'est pas bornée.

La précision d'une distribution d'échantillonnage g_j peut être mesurée avec la moyenne $\Delta(g_j)$ et le coefficient de variation $\text{CV}(g_j)$ définis comme suit

$$\Delta(g_j) = \mathbb{E}\left(\hat{\delta}_i^{\text{IS},g_j}|\{X_i^n, Y^n\}\right) \quad \text{et} \quad \text{CV}(g_j) = \frac{\sqrt{\mathbb{V}\text{ar}(\hat{\delta}_i^{\text{IS},g_j}|\{X_i^n, Y^n\})}}{\Delta(g_j)} \ .$$

Pour chaque tirage préférentiel réalisé avec les distributions g_j définies plus haut, la table 4.2 indique les estimations correspondantes du coefficient de variation $\mathrm{CV}(g_j)$ et de la moyenne $\Delta(g_j)$. Les résultats mettent en évidence l'impact du choix de la distribution auxiliaire. Les estimations de δ_1 obtenues avec les différentes fonction g_j sont similaires avec $1\%-2\%$ d'écart type relatif. Cependant, il apparait que les approximations associées à g_1 et g_3 sous-estiment fortement la valeur théorique de δ_2 et présentent une forte variabilité.

Ce manque de précision est dû à un problème d'échantillonnage. En effet, la performance de l'estimateur $\hat{\delta}_2^{\mathrm{IS},g_j}$ dépend en partie de la répartition des échantillons générés par la distribution g_j sur l'ensemble du support de l'intégrande $|f_{X_2}\hat{f}_Y - \hat{f}_{X_2,Y}|$.

Table 4.2. Estimations des indices de Borgonovo associés au modèle jouet (4.1) ; résultats obtenus après $M = 100$ exécutions de l'étape **IS3** avec différentes distributions d'échantillonnage.

Entrée	δ_i	g_1		g_2		g_3		g_4		$\hat{g}_{opt}$	
		$\Delta(g_1)$	$\mathrm{CV}(g_1)$	$\Delta(g_2)$	$\mathrm{CV}(g_2)$	$\Delta(g_3)$	$\mathrm{CV}(g_3)$	$\Delta(g_4)$	$\mathrm{CV}(g_4)$	$\Delta(\hat{g}_{opt})$	$\mathrm{CV}(\hat{g}_{opt})$
X_1	0.1436	0.1465	0.0278	0.1474	0.0083	0.1470	0.0172	0.1471	0.0191	0.1473	0.0038
X_2	0.5382	0.4685	1.1882	0.5175	0.0097	0.4644	0.4822	0.5166	0.0237	0.5174	0.0050

Ceci est confirmé par la figure [4.2] qui illustre la répartition imparfaite des échantillons générés avec g_1 (figure [4.2(c)]) et g_3 (figure [4.2(e)]) compte tenu de la région où la fonction $|f_{X_2}\hat{f}_Y - \hat{f}_{X_2,Y}|$ prend des valeurs significatives (figure [4.2(a)]). A contrario, les figures [4.2(d)] et [4.2(f)] semblent indiquer que les distributions g_2 et g_4 sont plus appropriées que g_1 and g_3, ce qui corrobore les meilleurs résultats obtenus avec g_4 et par dessus tout g_2 qui présente le plus faible coefficient de variation.

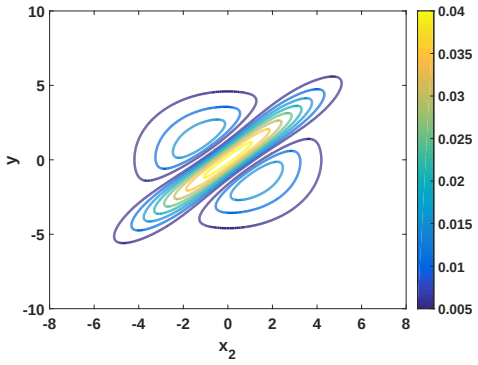
Ainsi, ce modèle simple montre que la fonction $\hat{f}_{X_i,Y}$ ne figure pas parmi les distributions d'échantillonnage les plus appropriées pour estimer l'indice de Borgonovo δ_i par tirage préférentiel. En outre, cela peut expliquer l'imprécision de l'estimateur *simple-boucle* constatée à la section 4.2.1 puisque, comme indiqué précédemment, ces deux estimateurs présentent quelques similarités. Par ailleurs, cette erreur d'estimation ne peut pas être attribuée à l'estimation par noyau puisque celle-ci est très efficace sur ce modèle jouet, les PDFs impliquées étant gaussiennes (voir figure [4.2(a)] et [4.2(b)]). Ce cas test met en évidence la précaution à avoir quant au choix de la distribution auxiliaire. À cet égard, le recherche de la distribution optimale est discutée dans la prochaine section.

4.2.3.3 La distribution d'échantillonnage optimale

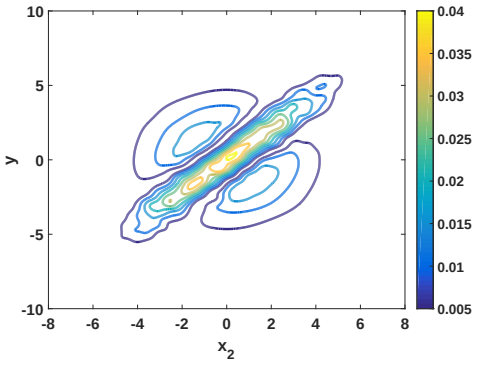
Considérons le second terme $\mathbb{V}\mathrm{ar}\left(\hat{h}(\mathbf{U}) \mid \{\mathbf{X}^n, Y^n\}\right)$ de la décomposition de la variance de $\hat{\delta}_i^{\mathrm{IS},g}$ donnée par Eq.(4.10). Ce terme d'erreur correspond à l'erreur induite par l'étape d'échantillonnage préférentiel **IS3** et peut donc être optimisé en considérant la distribution optimale (cf. section 2.2.3.2)

$$g_{opt} = \frac{|f_{X_i}\hat{f}_Y - \hat{f}_{X_i,Y}|}{\|f_{X_i}\hat{f}_Y - \hat{f}_{X_i,Y}\|_{L^1(\mathbb{R}^2)}} \ .$$

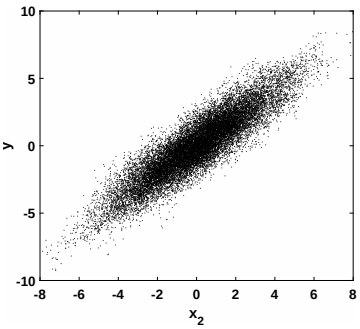
Comment mentionné dans la section 2.2.3.2, g_{opt} n'est pas utilisable directement en pratique mais peut être approximée en utilisant l'algorithme NIS de [Zhang, 1996]. À



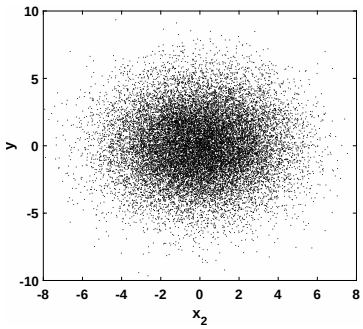
(a) Lignes de niveau $|f_{X_1} f_Y - f_{X_1,Y}|$.



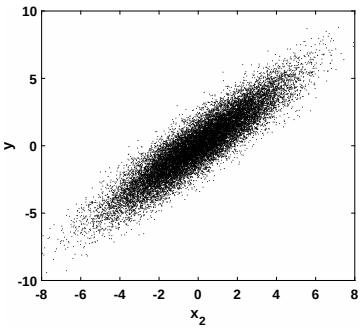
(b) Lignes de niveau $\left|f_{X_1} \hat{f}_Y - \hat{f}_{X_1,Y}\right|$.



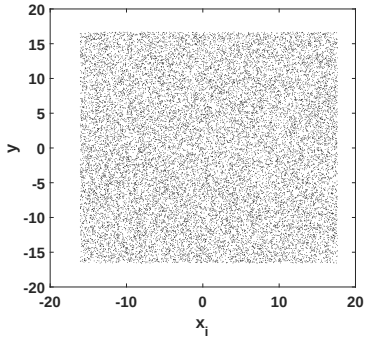
(c) $\{\mathbf{U}^k\} \overset{\text{i.i.d}}{\sim} g_1$.



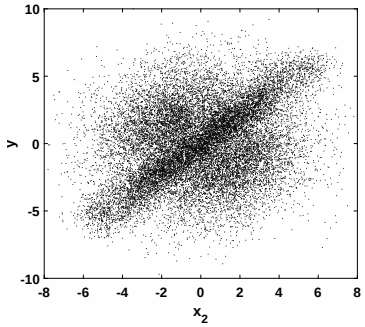
(d) $\{\mathbf{U}^k\} \overset{\text{i.i.d}}{\sim} g_2$.



(e) $\{\mathbf{U}^k\} \overset{\text{i.i.d}}{\sim} g_3$.



(f) $\{\mathbf{U}^k\} \overset{\text{i.i.d}}{\sim} g_4$.



(g) $\{\mathbf{U}^k\} \overset{\text{i.i.d}}{\sim} \hat{g}_{opt}$.

FIGURE 4.2. VISUALISATION DES DIFFÉRENTS ÉCHANTILLONS $\{\mathbf{U}^k\}$ UTILISÉS POUR L'ESTIMATION DE LA MEASURE δ_1 DU MODÈLE JOUET (4.1) EN COMPARAISON AVEC LES LIGNES DE NIVEAUX DE L'INTÉGRANDE $|f_{X_1} f_Y - f_{X_1,Y}|$ ET DE SON ESTIMATION $\left|f_{X_1} \hat{f}_Y - \hat{f}_{X_1,Y}\right|$.

partir d'un échantillon $\{(\mathbf{X}^n, \mathbf{Y}^n)\}_{n=1}^N \overset{\text{i.i.d}}{\sim} f_{X_i,Y}$ et des estimateurs à noyau $\hat{f}_Y$, $\hat{f}_{X_i,Y}$ associés, l'estimateur NIS $\hat{g}_{opt}$ peut être obtenu en respectant les étapes suivantes :

- Générer un échantillon i.i.d $\{(\tilde{X}_i^k, \tilde{Y}^k)\}_{k=1}^{N''}$ selon une distribution initiale g_0 arbitraire. Au vu des applications numériques effectuées dans la section précédente, cette étape peut être accomplie avec la fonction $g_0 = f_{X_i} \times \hat{f}_Y = g_2$ qui présente de bons résultats sur le modèle jouet.

- Calculer les poids

$$\omega(\tilde{X}_i^k, \tilde{Y}^k) = \frac{\left| f_{X_i}(\tilde{X}_i^k)\hat{f}_Y(\tilde{Y}^k) - \hat{f}_{X_i,Y}(\tilde{X}_i^k, \tilde{Y}^k) \right|}{g_0(\tilde{X}_i^k, \tilde{Y}^k)}, \ k = 1, \ldots, N''.$$

- Estimer g_{opt} par l'estimateur à noyaux pondérés

$$\hat{g}_{opt}(x, y) = \frac{1}{N''\tilde{h}_1\tilde{h}_2\bar{\omega}} \sum_{k=1}^{N''} \omega(\tilde{X}_i^k, \tilde{Y}^k) K\left(\frac{x - \tilde{X}_i^k}{\tilde{h}_1}\right) K\left(\frac{y - \tilde{Y}^k}{\tilde{h}_2}\right), \tag{4.14}$$

avec $\bar{\omega} = \frac{1}{N''}\sum_{k=1}^{N''} \omega(\tilde{X}_i^k, \tilde{Y}^k)$, K est un noyau de dimension 1 et où les fenêtres $\tilde{h}_1$ et $\tilde{h}_2$ sont calculées avec la toolbox développée par Alex Ihler et Mike Mandel[1] adaptée aux estimateurs à noyaux pondérés.

On obtient alors l'estimateur suivant :

$$\hat{\delta}_i^{\mathrm{Opt}} := \hat{\delta}_i^{\mathrm{IS},\hat{g}_{opt}} = \frac{1}{2N'} \sum_{k=1}^{N'} \frac{\left| \hat{f}_Y(U_2^k) f_{X_i}(U_1^k) - \hat{f}_{X_i,Y}(\mathbf{U}^k) \right|}{\hat{g}_{opt}(\mathbf{U}^k)}, \tag{4.15}$$

où $\{\mathbf{U}^k = (U_1^k, U_2^k)\}_{k=1}^{N'} \overset{\text{i.i.d}}{\sim} \hat{g}_{opt}$. La procédure suivante peut être utilisée pour générer une réalisation $\mathbf{U}$ selon $\hat{g}_{opt}$:

1. Générer $Z \sim \sum_{k=1}^{N''} \frac{\omega(\tilde{X}_i^k, \tilde{Y}^k)}{\bar{\omega}} \delta_k$, où δ_k désigne ici la distribution de Dirac de support $\{k\}$.
2. Générer $\mathbf{U} \sim N\left(\begin{pmatrix} \tilde{X}_i^Z \\ \tilde{Y}^Z \end{pmatrix}, \begin{pmatrix} \tilde{h}_1 & 0 \\ 0 & \tilde{h}_2 \end{pmatrix}\right)$.

Pour terminer cette section, examinons l'influence de la distribution $\hat{g}_{opt}$ avec la même méthodologie mise en œuvre dans la section 4.2.3.2. La table 4.2 montre que $\hat{\delta}^{\mathrm{Opt}}$ présente le plus faible coefficient de variation, à savoir $0,5\%$, soit environ 2 fois moins que celui de $\hat{\delta}^{\mathrm{IS},g_2}$. Ainsi, conditionnellement à l'étape d'estimation par noyau, l'utilisation de l'approximation $\hat{g}_{opt}$ de la distribution optimale semble être le choix le plus judicieux en terme d'erreur d'estimation (ce qui est conforme à la théorie) bien que les moyennes respectives de $\hat{\delta}^{\mathrm{Opt}}$ et $\hat{\delta}^{\mathrm{IS},g_2}$ soient similaires. Cette similarité est due au fait que l'erreur induite par l'estimation par noyau est prédominante dans l'erreur globale. Ainsi, la distribution g_2 peut être un bon candidat étant donné que le calcul de $\hat{g}_{opt}$ est une procédure plus fastidieuse que la simple considération de g_2 et qui rajoute du temps de simulation ($3s$ sur le modèle jouet). Néanmoins, cette différence de charge de calcul devient négligeable lorsque la boîte noire $\mathcal{M}$ est coûteuse à évaluer. Dans la prochaine section, plusieurs exemples numériques sont considérés pour illustrer l'application de l'estimateur $\hat{\delta}_i^{\mathrm{Opt}}$.

[1] https://www.ics.uci.edu/~ihler/code/kde.html.

Table 4.3. Estimations des indices de Borgonovo du modèle gaussien $\mathcal{M}_2$(4.2).

Entrée	δ_i	simple-boucle $\hat{\delta}_i^{\mathbf{SL}}$			Méthode proposée $\hat{\delta}_i^{\mathbf{Opt}}$		
		Moyenne	CV	RD	Moyenne	CV	RD
X_1	0.2857	0.2201	0.0280	-0.2297	0.2707	0.0218	-0.0526
X_2	0.3620	0.2620	0.0226	-0.2761	0.3444	0.0157	-0.0486
X_3	0.3792	0.2690	0.0218	-0.2907	0.3607	0.0157	-0.0487
X_4	0.3176	0.2383	0.0219	-0.2497	0.3010	0.0199	-0.0522

4.2.4 Application à des cas test analytiques

Dans cette section, les modèles (4.2) et (4.3) définis plus haut sont considérés pour illustrer l'application de la procédure d'estimation par échantillonnage préférentiel associée à l'approximation $\hat{g}_{opt}$ (4.14) de la distribution optimale g_{opt}. Les résultats obtenus sont comparés avec ceux effectués avec la méthode *simple-boucle*.

4.2.4.1 Exemple 1 : la transformation affine d'un vecteur gaussien (4.2)

La table 4.3 réunit les résultats obtenus avec les paramètres $N'' = N = 5000$ et $N' = 5 \times 10^4$ (le nombre d'appels à $\mathcal{M}_2$ est le même pour les deux méthodes considérées). Il est important de noter que seulement un échantillon de taille $N = 5000$ permet dans les deux cas de calculer l'ensemble des quatre mesures d'importance de Borgonovo. Les temps de simulations associés à la méthode *simple-boucle* et à la méthode proposée sont respectivement de 4 s et 54 s, dont $3,5$ s pour le calcul de $\hat{g}_{opt}$. Cette différence de temps de calcul, notamment due au fait que $N' \gg N$, devient négligeable lorsque la boite noire $\mathcal{M}$ est coûteuse à évaluer.

Il peut être noté que les deux méthodes produisent des estimations respectant le *ranking* exact, à savoir

$$X_3 > X_2 > X_4 > X_1 \quad ,$$

et convergent puisque ces estimations sont obtenues avec des coefficients de variation de l'ordre de 2%. Cependant, les estimations de $(\hat{\delta}_i^{\mathrm{Opt}})_{1\leq i\leq 4}$ sont approximativement égales aux valeurs théoriques avec des écarts relatifs de l'ordre de 5% tandis que les estimations *simple-boucle* témoignent d'une moindre précision avec des écarts relatifs supérieurs à 20%.

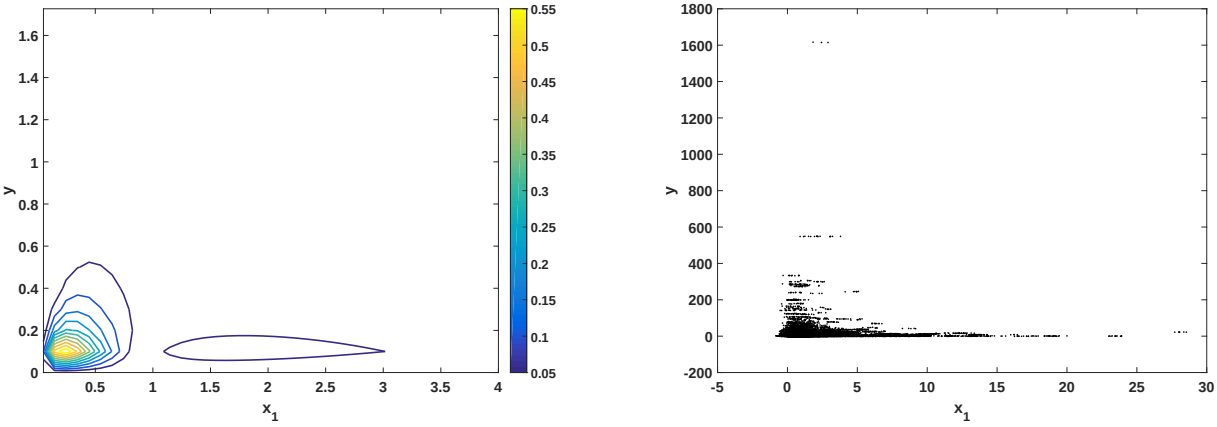
4.2.4.2 Exemple 2 : le modèle multiplicatif (4.3)

La table 4.4 réunit les résultats obtenus avec les paramètres $N'' = N = 5000$ et $N' = 2 \times 10^4$ (à noter que, par symétrie du modèle $\mathcal{M}_3^d$, toutes les entrées ont une influence identique) et en faisant varier la paramètre de dimension d.

Bien que les résultats ne soient pas complètement satisfaisants, la méthode proposée permet d'augmenter la précision de la méthode *simple-boucle* d'un facteur 3 pour $d = 4$ et par un facteur 10 pour $d = 5$ et $d = 6$. De plus, l'estimateur *simple-boucle* peut fournir des estimations supérieures à 1 avec de forts coefficients de variation. Il convient de noter que le présent problème est difficile puisque les distributions de sortie sont à

Table 4.4. Estimations des indices de Borgonovo du modèle $\mathcal{M}_3^d$ (4.3).

Entrée	δ_i	**simple-boucle** $\hat{\delta}_i^{\mathbf{SL}}$			**Méthode proposée** $\hat{\delta}_i^{\mathbf{Opt}}$		
		Moyenne	CV	RD	Moyenne	CV	RD
$d=4$	0.1846	0.8757	0.5219	3.7440	0.3972	0.0997	1.1517
$d=5$	0.1604	2.0302	1.3814	11.6574	0.4570	0.0719	1.8492
$d=6$	0.1436	3.6376	0.5324	24.3318	0.5036	0.2515	2.5068



(a) Lignes de niveau de $|f_{X_1} f_Y - f_{X_1,Y}|$.

(b) $\{\mathbf{U}^k\} \overset{\text{i.i.d}}{\sim} \hat{g}_{opt}$.

Figure 4.3. Visualisation de l'échantillon généré via la distribution $\hat{g}_{opt}$ en comparaison avec l'intégrande $|f_{X_1} f_Y - f_{X_1,Y}|$ de l'exemple 2.

queue lourde ce qui complexifie l'estimation de leurs PDFs respectives. En outre, le biais observé sur cet exemple est dû à l'estimation par noyau gaussien qui affecte la qualité de l'estimateur $|f_{X_1}\hat{f}_Y - \hat{f}_{X_1,Y}|$. En effet, dans le cas $d=4$, on peut voir sur la figure [4.3(b)] que l'échantillon généré selon $\hat{g}_{opt}$ n'est pas correctement réparti dans la zone d'intérêt, i.e, la région où l'intégrande $|f_{X_1} f_Y - f_{X_1,Y}|$ prend de grandes valeurs (figure [4.3(a)]), notamment en ce qui concerne l'échelle des coordonnées de sortie.

Pour conclure, la méthode d'estimation par échantillonnage préférentiel présentée dans cette section apparait être une bonne alternative à la méthode *simple-boucle* qui présente des problèmes théoriques pouvant affecter sa précision. Elle n'est cependant pas complètement satisfaisante puisque celle-ci peut être mise en défaut lorsque les distributions sous-jacentes sont à queue lourde, comme constaté sur le modèle multiplicatif. Cette dernière remarque motive la section suivante dont le but est de proposer une approche alternative permettant de surmonter cette difficulté.

4.3 Estimateur δ_i^{ME} basé sur la représentation de δ_i en terme de densité de copule et le principe du maximum d'entropie

Cette section introduit un nouvel estimateur δ_i^{ME} des mesures de sensibilité de Borgonovo du premier ordre. Celui-ci est construit à partir de la représentation de δ_i en terme de densité de copule (cf. section 3.4.3)

$$\delta_i = \frac{1}{2}\int_{[0,1]^2} |c_i(u_i, v) - 1| \mathrm{d}u_i \mathrm{d}v \ , \tag{4.16}$$

combinée à l'estimation de la PDF c_i du couple $(F_{X_i}(X_i), F_Y(Y))$ via le principe du maximum d'entropie rencontré à la section 2.2.2.2. L'objet de cette nouvelle méthode est de pallier les limites de l'approche par échantillonnage préférentiel détaillée dans la précédente section. D'une part, adopter le formalisme des copules permet de se soustraire au cas des distributions à queue lourde dont l'estimation peut être ardue. D'autre part, la définition (4.16) est tout à fait adaptée au cadre d'application de la méthode du maximum d'entropie décrite à la section 2.2.2.2 puisque c_i est à support borné et admet donc des moments de tous ordres. En outre, le choix des contraintes, i.e, les moments de c_i considérés, est discuté dans la section 4.3.1. Ensuite, la section 4.3.2 décrit les différentes étapes du schéma d'estimation associé à δ_i^{ME}. Enfin, plusieurs applications numériques sont effectuées en section 4.3.3.

4.3.1 Choix des contraintes

Dans la suite, les marginales $F_{X_i}(X_i)$ et $F_Y(Y)$ sont notées comme suit :

$$U_i := F_{X_i}(X_i) \quad \text{et} \quad V := F_Y(Y) \ .$$

En tant que PDF du couple (U_i, V), la densité c_i peut être approchée par un estimateur $\hat{c}_i$ par maximum d'entropie. Conformément à la procédure décrite à la section 2.2.2.2, la construction d'un tel estimateur repose sur un choix de contraintes s'exprimant comme des moments de la distribution de (U_i, V). Ce choix est une question difficile puisque les résultats de convergence explicités dans la section 2.2.2.2 concernent uniquement le cas de la dimension $d = 1$. [AghaKouchak, 2014], [Hao and Singh, 2015] et [Hao and Singh, 2013] proposent de considérer les contraintes associées à l'application

$$g(u, v) = (u, u^2, \ldots, u^m, v, v^2, \ldots, v^m, uv). \tag{4.17}$$

Cela revient à spécifier les m premiers moments des marginales U_i et V et le moment d'ordre 1 de U_iV. Ici, il est proposé de considérer les contraintes suivantes :

$$g(u, v) = (u^{\alpha_k} v^{\alpha_l}, k, l = 1, \ldots, m) \, , \tag{4.18}$$

où $\alpha_1 < \cdots < \alpha_m$: cela correspond à une collection de m^2 moments fractionnaires du produit U_iY notés dans la suite

$$\mu_{k,l} = \mathbb{E}[U_i^{\alpha_k} V^{\alpha_l}] \ . \tag{4.19}$$

Ce choix repose sur les deux arguments suivants :

1. D'une part, comme mentionné dans la section 2.2.2.2, [Zhang and Pandey, 2013] montre que préférer l'utilisation des moments fractionnaires au lieu de moments entiers permet d'améliorer la précision des estimations.

2. Le second postulat est que la caractérisation de la loi d'une variable aléatoire positive donnée par le théorème 2.2 peut être étendue en dimension 2 en considérant une collection infinie de moments de la forme $\mu_{k,l}$.

Dans le présent problème, la subtilité est que la distribution de la variable U_iV est inconnue et il en va donc de même pour ses moments $\mu_{k,l}$. En conséquence, ces contraintes doivent êtres estimées. Cela mène donc au problème d'optimisation suivant :

$$\begin{cases} \hat{c}_i = \underset{f \in L^1([0,1]^2)}{\text{Argmin}} \; H(f) \\ \\ \text{tel que } \int_{[0,1]^2} u^{\alpha_r} v^{\alpha_s} f(u,v) \mathrm{d}u\mathrm{d}v = \hat{\mu}_{k,l}, \; r,s = 1,\dots,m. \end{cases} \tag{4.20}$$

où $\hat{\mu}_{k,l}$ est un estimateur de $\mu_{k,l}$. Le problème dual associé à (4.20) correspond alors à la minimisation de la fonction suivante :

$$(\lambda_{k,l}) \mapsto \sum_{k=1}^{m}\sum_{l=1}^{m} \lambda_{k,l}\hat{\mu}_{k,l} + \log\left(\int_0^1\int_0^1 \exp\left(-\sum_{k=1}^{m}\sum_{l=1}^{m}\lambda_{k,l}u^{\alpha_k}v^{\alpha_l}\right) dudv\right) . \tag{4.21}$$

Cette fonction est strictement convexe sur l'ensemble des points faisables et admet donc unique minimum global $(\lambda^*_{k,l}) \in \mathbb{R}^{m^2}$ qui peut être déterminé en utilisant des techniques classiques d'optimisation convexe [Boyd and Vandenberghe, 2004]. Ainsi, l'estimateur du maximum d'entropie de c_i s'exprime comme suit :

$$\hat{c}_i(u,v) \propto \mathbb{1}_{[0,1]^2}(u,v) \exp\left(-\sum_{k=1}^{m}\sum_{l=1}^{m}\lambda^*_{k,l}u^{\alpha_k}v^{\alpha_l}\right). \tag{4.22}$$

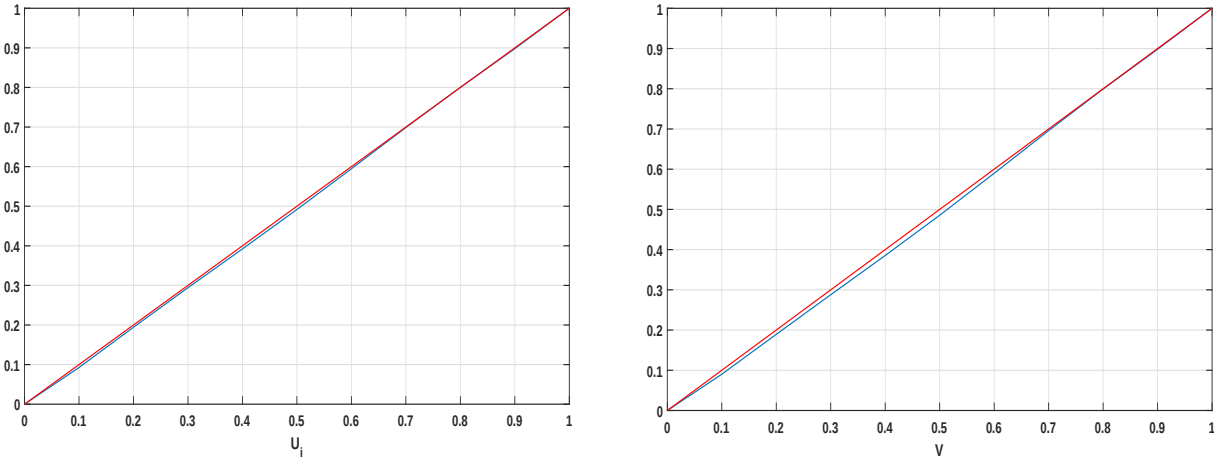
Remarque. Il convient de souligner que rien n'assure que l'estimateur $\hat{c}_i$ donné par Eq.(4.22) soit une densité de copule, i.e, que les distributions marginales associées soient uniformes. Pour s'en assurer, il est nécessaire d'ajouter les contraintes suivantes :

$$\forall u \in [0,1], \; \int_0^u\int_0^1 f(u,v)dudv = u \; ,$$

$$\forall v \in [0,1], \; \int_0^1\int_0^v f(u,v)dudv = v \; ,$$

ce qui dépasse le cadre des résultats théoriques de la section 2.2.2.2. Heuristiquement et en vertu du théorème 2.2, ces conditions peuvent néanmoins être remplacées par la donnée d'une collection infinie de moments fractionnaires des marginales. De plus, il peut être constaté numériquement sur le cas du modèle gaussien (4.2) que les distributions marginales de l'estimateur $\hat{c}_1$ sont approximativement uniformes. En effet, la figure [4.4] illustre que les fonctions de répartition marginales

$$\hat{F}_{U_1}(t) = \int_0^t\int_0^1 \hat{c}_1(u,v)\mathrm{d}u\mathrm{d}v \quad \text{et} \quad \hat{F}_V(t) = \int_0^1\int_0^t \hat{c}_1(u,v)\mathrm{d}u\mathrm{d}v \; ,$$



(a) (b)

FIGURE 4.4. REPRÉSENTATION DES FONCTIONS DE RÉPARTITION MARGINALES (COURBES BLEUES) ASSOCIÉE À L'APPROXIMATION (4.22) DE LA DENSITÉ DE COPULE c_1 DU MODÈLE GAUSSIEN (4.2), EN COMPARAISON AVEC LA CDF DE LA DISTRIBUTION UNIFORME $U([0,1])$ (COURBES ROUGES) [PARAMÈTRES : $N = 10^4$ ET $\alpha_k \in \{\frac{2}{3}, \frac{4}{3}, 2\}$].

sont proches de la CDF de $U([0,1])$. En outre, ceci corrobore le second postulat portant sur la caractérisation de la loi de (U_i, V) via une collection infinie de moments fractionnaires $\mu_{k,l}$.

4.3.2 Définition de $\delta_i^{\mathbf{ME}}$

Cette section résume les différentes étapes menant à la construction de l'estimateur δ_i^{ME} de δ_i basé sur la procédure décrite dans la section précédente.

Étape ME1. Générer $\{\mathbf{X}^n = (X_1^n, \ldots, X_d^n)\}_{n=1}^N \overset{\text{i.i.d}}{\sim} f_{\mathbf{X}}$ puis évaluer $Y^n = \mathcal{M}(\mathbf{X}^n)$ pour $n = 1, \ldots, N$.

Étape ME2. Estimation des contraintes. Choisir m nombres réels $\alpha_1 < \cdots < \alpha_m$ et considérer les estimateurs Monte-Carlo des moments $\mu_{k,l} = \mathbb{E}[U_i^{\alpha_k} V^{\alpha_l}]$:

$$\hat{\mu}_{k,l} := \frac{1}{N} \sum_{n=1}^{N} (F_{X_i}(X_i^n))^{\alpha_k} (\hat{F}_Y(Y^n))^{\alpha_l}, \ k, l = 1, \ldots, m,$$

où $\hat{F}_Y$ désigne l'estimateur empirique de la CDF F_Y, i.e,

$$\hat{F}_Y(t) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{Y^n \leq t} \ .$$

Étape ME3. Calcul des multiplicateurs de Lagrange. En utilisant l'algorithme du point intérieur, déterminer le minimum $(\lambda_{k,l}^*)$ de la fonction définie par Eq.(4.21).

Étape ME4. Générer $\{(U_1^k, U_2^k)\}_{k=1}^{N'} \overset{\text{i.i.d.}}{\sim} U([0,1]^2)$ puis estimer δ_i par :

$$\hat{\delta}_i^{\mathrm{ME}} = \frac{1}{2N'} \sum_{k=1}^{N'} \left| \hat{c}_i(U_1^k, U_2^k) - 1 \right| , \tag{4.23}$$

où $\hat{c}_i$ est l'estimateur du maximum d'entropie défini par Eq.(4.22).

Discussion sur le biais de l'estimateur $\hat{\delta}_i^{\mathbf{ME}}$. L'estimateur $\hat{\delta}_i^{\mathrm{ME}}$ présente deux biais distincts inhérents à l'estimation de la PDF c_i de (U_i, V) par maximisation de l'entropie :

1. Un biais associé au remplacement des contraintes $\mu_{k,l}$ par les contraintes approchées $\hat{\mu}_{k,l}$ obtenues à partir de l'échantillon
$$\{(U_i^n, V^n) = (F_{X_i}(X_i^n), \hat{F}_Y(Y^n))\}_{n=1}^N .$$
2. Un biais dû au nombre fini m^2 de contraintes considérées (pour des raisons numériques évidentes).

En pratique, un compromis entre ces deux biais est nécessaire. En effet, le problème d'optimisation (4.20) associé aux contraintes approchées $\hat{\mu}_{k,l}$ correspond en fait au problème d'estimation par maximum d'entropie de la distribution empirique

$$\frac{1}{N} \sum_{k=1}^{N} \delta_{(U_i^k, V^k)} . \tag{4.24}$$

En outre, en partant du postulat que la convergence de l'estimateur du maximum d'entropie est vérifiée en dimension 2, un compromis entre les paramètres N et m est donc obligatoire pour éviter un phénomène de surapprentissage. En effet, pour une taille d'échantillon N fixée, l'estimateur $\hat{c}_i$ devrait tendre vers la distribution empirique (4.24) lorsque le nombre m^2 de contraintes tend vers l'infini, au lieu de la vraie PDF c_i.

Par ailleurs, contrairement à (4.17), aucunes contraintes sur les marginales U_i et V ne sont imposées. Pour réduire le second biais, il serait effectivement naturel de rajouter les contraintes suivantes

$$\int_{[0,1]^2} u^{\alpha_k} f(u,v) du dv = \frac{1}{\alpha_k + 1} , \tag{4.25}$$

compte tenu du fait que $U_i \sim U([0,1])$. Cependant, il peut être constaté numériquement sur le modèle gaussien $\mathcal{M}_2$ (4.2) que cela fournit des estimations moins précises et montrant une grande variabilité. En effet, le tableau 4.5 rassemble les résultats obtenus avec l'ajout des contraintes (4.25). On peut constater que les écarts relatifs sont supérieurs à 10% ce qui est moins satisfaisant que les résultats obtenus avec la méthode d'échantillonnage préférentiel détaillée précédemment. Ce défaut de précision est une conséquence directe du premier biais, les contraintes données par Eq.(4.25) étant incompatibles avec la distribution empirique $\frac{1}{N}\sum_{k=1}^N \delta_{U_i^k}$ qui n'est pas uniformément distribuée.

Remarques. À l'instar de la méthode d'échantillonnage préférentiel, le calcul des estimateurs $\{\hat{\delta}_i^{\mathrm{ME}}\}_{i=1}^d$ s'effectue à partir d'un échantillon commun et donc indépendamment

TABLE 4.5. ESTIMATION DES INDICES DE BORGONOVO DU MODÈLE $\mathcal{M}_2$ (4.2) AVEC AJOUT DES CONTRAINTES MARGINALES DONNÉES PAR (4.25) [PARAMÈTRES : $N = 10^4$; $\alpha_k \in \{\frac{2}{3}, \frac{4}{3}, 2\}$].

Entrée	δ_i	$\hat{\delta}_i^{\mathrm{ME}}$		
		Moyenne	CV	RD
X_1	0.2857	0.3139	0.1555	0.0988
X_2	0.3620	0.3980	0.1510	0.0994
X_3	0.3792	0.4261	0.1842	0.1238
X_4	0.3176	0.3489	0.1714	0.0986

de la dimension d du modèle. De plus, l'étape **ME4.** ne requiert aucune évaluation supplémentaire de la boîte noire $\mathcal{M}(\cdot)$ de sorte que la taille N' de l'échantillon Monte-Carlo peut être choisie aussi grande que souhaité. Pour conclure, la définition de l'estimateur $\hat{\delta}_i^{\mathrm{ME}}$ est proche de celle de $\hat{\delta}_i^{\mathrm{Nataf}}$ dont la construction a été détaillée dans la section 3.4.2.2. En effet, l'estimateur $\hat{\delta}_i^{\mathrm{Nataf}}$ coïncide avec l'estimateur Monte-Carlo de l'espérance suivante

$$\begin{aligned}\frac{1}{2}\int |\tilde{c}_i(u,v) - 1| \mathrm{d}u\mathrm{d}v &= \frac{1}{2}\int \left|\frac{1}{\tilde{c}_i(u,v)} - 1\right| \tilde{c}_i(u,v)\mathrm{d}u\mathrm{d}v \ , \\ &= \frac{1}{2}\mathbb{E}\left[\left|\frac{1}{\tilde{c}_i(U_i,V)} - 1\right|\right] \ ,\end{aligned}$$

où c_i est approximée par la densité de copule gaussienne

$$c_i(u,v) \approx \tilde{c}_i(u,v) := \frac{\phi_2(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}, \tag{4.26}$$

avec ϕ et ϕ_2, les PDFs respectives des distributions $N(0,1)$ et $N_2(0_{\mathbb{R}^2}, \rho_{0,i})$ où $\rho_{0,i}$ est la matrice de corrélation de $(\Phi^{-1}(U_i), \Phi^{-1}(V))$ (cf. section 3.4.2.2 pour plus de détails). Cependant, la définition de $\hat{\delta}_i^{\mathrm{ME}}$ est beaucoup plus permissive que celle de $\hat{\delta}_i^{\mathrm{Nataf}}$ qui repose sur les hypothèses et approximations suivantes (cf. section 3.4.2.2) :

(A1) Les variables d'entrées X_i sont indépendantes.

(A2) Le modèle $\mathcal{M}$ est approché en utilisant la technique de réduction de modèle cut-HDMR (cf. section 2.5.2).

(A3) La densité de copule c_i est approximée par Eq.(4.26).

Dans la prochaine section, plusieurs exemples numériques sont considérés pour illustrer et comparer l'application de ces deux estimateurs.

4.3.3 Application à des cas test analytiques

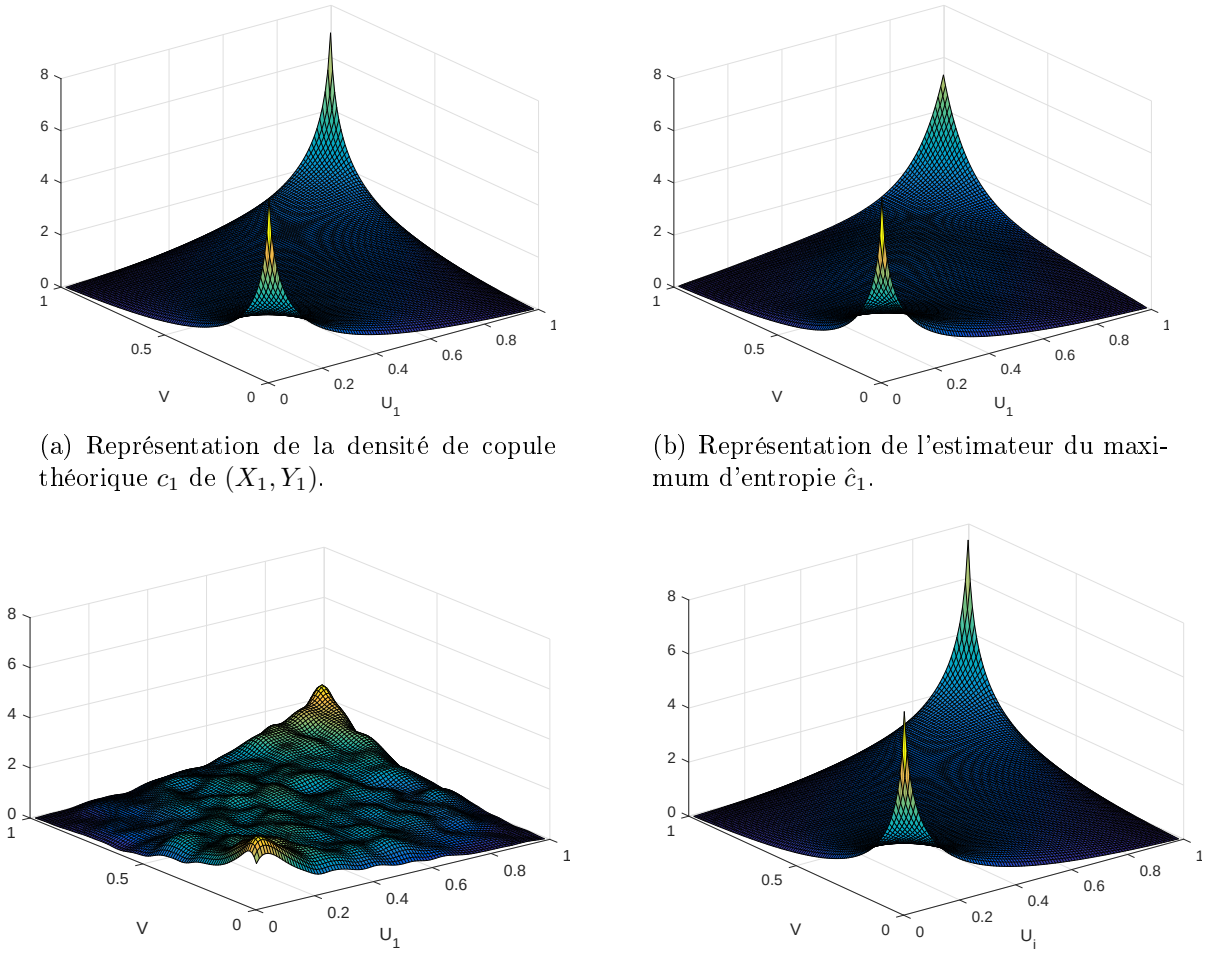
Dans cette section, les modèles analytiques (4.2) et (4.3) sont considérés pour illustrer l'application de l'estimateur $\hat{\delta}_i^{\mathrm{ME}}$ et comparer ce dernier avec l'estimateur $\hat{\delta}_i^{\mathrm{Nataf}}$ basé sur la transformation de Nataf.

4.3.3.1 Exemple 1 : le modèle multiplicatif (4.3)

Dans cette partie, les deux méthodes considérées sont appliquées $M = 100$ fois au modèle multiplicatif (4.3) dans le cas de la dimension $d = 4$. Les paramètres et les temps de simulations respectifs pour calculer δ_1 sont de $N = 21$ et 62 s pour $\hat{\delta}_1^{\text{Nataf}}$ et $N = 5000$, $\alpha_k \in \{\frac{2}{3}, \frac{4}{3}, 2\}$ $(m = 3)$ et 8 s pour $\hat{\delta}_1^{\text{ME}}$.

TABLE 4.6. ESTIMATIONS DE L'INDICE δ_1 DU MODÈLE MULTIPLICATIF $\mathcal{M}_3^d$ (4.3) POUR $d = 4$.

Entrée	δ_1	$\hat{\delta}_1^{\text{Nataf}}$			$\hat{\delta}_1^{\text{ME}}$		
		Moyenne	CV	RD	Moyenne	CV	RD
X_1	0.1846	0.1946	0.0032	0.0541	0.1864	0.0277	0.0096



(a) Représentation de la densité de copule théorique c_1 de (X_1, Y_1).

(b) Représentation de l'estimateur du maximum d'entropie $\hat{c}_1$.

(c) Représentation de l'estimateur à noyau gaussien.

(d) Représentation de l'approximation de c_1 définie par Eq.(4.26) (cette figure est obtenue en supposant f_Y connue).

FIGURE 4.5. VISUALISATION DE LA DENSITÉ DE COPULE c_1 DE (X_1, Y_1) ET DE SES DIFFÉRENTES ESTIMATIONS .

Les résultats sont réunis dans la table 4.6. Il peut être constaté que chacune des deux

méthodes a convergé puisque les estimations sont obtenues avec un coefficient de variation inférieur à 2% et fournissent des estimations bien plus précises que celles obtenues avec la méthode d'échantillonnage préférentiel (cf. section 4.2.4.2). Les estimations de $\hat{\delta}_1^{\mathrm{ME}}$ sont très proches de la valeur théorique de δ_1 avec un écart relatif de 1% environ. En outre, on peut observer sur la figure [4.5(b)] que la densité de copule de (X_1, Y_1) est bien approximée par l'estimateur du maximum d'entropie défini par Eq.(4.22) en comparaison avec l'estimateur à noyau gaussien (figure [4.5(c)]). Les estimations de $\hat{\delta}_1^{\mathrm{Nataf}}$ présentent une bonne précision avec 5% d'erreur relative associé à une très faible variabilité et sont obtenues avec seulement $N = 21$ évaluations de modèle. Cette excellente performance est néanmoins à relativiser puisqu'il s'agit d'un cas idéal pour lequel toutes les conditions **(A1)**, **(A2)** et **(A3)** sont vérifiées. En effet, les entrées sont indépendantes et il peut être vérifié aisément que le modèle de sortie coïncide avec son approximation cut-HDMR. De plus, la densité de copule théorique c_1 (figure [4.5(a)]) est proche de son approximation gaussienne donnée par Eq.(4.26) (figure [4.5(d)]) ce qui semble indiquer que l'hypothèse **(A3)** est vérifiée. Un exemple pour lequel l'hypothèse d'indépendance **(A1)** est mise en défaut est étudié dans la prochaine section.

4.3.3.2 Exemple 2 : la transformation affine d'un vecteur gaussien (4.2)

Dans cette section, les deux méthodes considérées sont appliquées $M = 100$ fois au modèle linéaire gaussien (4.2). Les paramètres et les temps de simulations respectifs pour calculer l'ensemble des indices δ_i sont de $N = 21$ et 65 s pour $\hat{\delta}_i^{\mathrm{Nataf}}$ et $N = 5000$, $\alpha_k \in \{\frac{2}{3}, \frac{4}{3}, 2\}$ $(m = 3)$ et 57 s pour $\hat{\delta}_i^{\mathrm{ME}}$.

Pour tout $i \in \{1, \ldots, d\}$ la variable (X_i, Y) est gaussienne. Dès lors, la densité de copule c_i associée est gaussienne, i.e, coïncide avec l'approximation donnée par Eq. (4.26), et l'hypothèse **(A3)** est donc licite. En revanche, l'hypothèse **(A1)** n'est pas vérifiée puisque les entrées sont corrélées. En ce qui concerne l'hypothèse **(A2)**, le modèle ne coïncide pas avec son approximation cut-HDMR, mais il est difficile de trancher quant à la pertinence de cette approximation.

Table 4.7. Estimations des indices de Borgonovo du premier ordre du modèle $\mathcal{M}_2$ (4.2).

Entrée	δ_i	$\hat{\delta}_i^{\mathrm{Nataf}}$			$\hat{\delta}_i^{\mathrm{ME}}$		
		Moyenne	CV	RD	Moyenne	CV	RD
X_1	0.2857	0.1654	0.0026	-0.4209	0.2840	0.0174	-0.0058
X_2	0.3620	0.1777	0.0028	-0.5090	0.3538	0.0138	-0.0225
X_3	0.3792	0.1906	0.0037	-0.4973	0.3688	0.0121	-0.0274
X_4	0.3176	0.2042	0.0058	-0.3571	0.3141	0.0180	-0.0110

Les résultats sont rassemblés dans la table 4.7. Ceux-ci montrent que la méthode proposée respecte le bon ranking, à savoir

$$X_3 > X_2 > X_4 > X_1 \ ,$$

et présente des écarts relatifs légèrement plus faibles que ceux obtenus avec les estimateurs $(\delta_i^{\mathrm{IS}})_{1 \leq i \leq 4}$ à la section 4.2.4.1. En revanche, les estimations obtenues avec la transformation

de Nataf sont fortement impactées par la corrélation entre les variables d'entrée. En effet, celles-ci présentent des écart-relatifs de l'ordre de 50% et ne respectent pas le ranking correct.

4.4 Conclusion : application à un modèle d'estimation de la zone de retombée d'un étage de lanceur

Cette section est organisée comme suit. Tout d'abord, une brève présentation du cas test d'estimation de la zone de retombée d'un étage de lanceur est proposé dans la section 4.4.1, suivie dans la section 4.4.2 par une application des deux nouvelles méthodes introduites dans ce chapitre. Les valeurs théoriques des indices de Borgonovo δ_i étant inconnues, les résultats obtenus sont comparés avec les estimations fournies par la méthode *double-boucle* décrite à la section 3.4.1.1, qui feront office de valeurs de référence. Une comparaison avec les informations apportées par les indices de Sobol est également effectuée. Enfin, la section 4.4.3 propose une synthèse des avantages et inconvénients des principales méthodes d'estimations rencontrées dans la section 3.4 et le présent chapitre.

4.4.1 Description du cas test

Cette section revient sur l'exemple introductif portant sur l'estimation de la zone de retombée d'un étage de lanceur [Derennes et al., 2019a]. Est considéré ici un code de simulation de trajectoire simplifié [Morio and Balesdent, 2015] prenant en compte $d = 6$ paramètres d'entrée représentant certaines conditions initiales, variables environnementales et caractéristiques du lanceur au moment de sa séparation :

X_1 : altitude de l'étage de lanceur(Δa (m)) ;

X_2 : vitesse de l'étage de lanceur (Δv (m.s^{-1})) ;

X_3 : angle de trajectoire de vol ($\Delta\gamma$ (rad)) ;

X_4 : l'azimuth ($\Delta\psi$ (rad)) ;

X_5 : masse résiduelle de propulseur (Δm (kg)) ;

X_6 : l'erreur de force de traînée (ΔC_d dimensionless).

Ces variables sont supposées indépendantes et distribuées selon des lois normales dont les paramètres sont donnés dans le tableau 4.8. La sortie du modèle $Y = \mathcal{M}(X_1, \ldots, X_6)$ représente alors la distance entre la zone de retombée théorique dans l'océan et celle estimée par le code $\mathcal{M}$.

Table 4.8. Distributions des paramètres d'entrée du modèle de simulation de trajactoire de lanceur..

Entrée	Distribution	Moyenne μ_{X_i}	Écart type σ_{X_i}
$X_1 = \Delta a$ (m)	Normale	0	165
$X_2 = \Delta v$ (m.s^{-1})	Normale	0	3.7
$X_3 = \Delta\gamma$ (rad)	Normale	0	0.001
$X_4 = \Delta\psi$ (rad)	Normale	0	0.0018
$X_5 = \Delta m$ (kg)	Normale	0	70
$X_6 = \Delta C_d$ (1)	Normale	0	0.1

4.4.2 Résultats numériques

L'ensemble des méthodes d'analyse de sensibilité considérées sont appliquées $M = 100$ fois au modèle de lanceur décrit dans la section 4.4.1, et ce, avec des échantillons de taille $N = 5000$. Les tables 4.9 et 4.10 réunissent respectivement les résultats d'estimation des indices de Borgonovo du premier ordre δ_i et les indices de Sobol d'ordre 1 S_i et totaux S_{T_i}.

Table 4.9. Estimations des indices de Borgonovo du premier ordre du modèle de lanceur.

Entrée	$\hat{\boldsymbol{\delta}}_{\boldsymbol{i}}^{\mathrm{PDF}}$		$\hat{\delta}_i^{\mathrm{IS}}$		$\hat{\delta}_i^{\mathrm{ME}}$	
	Moyenne	CV	Moyenne	CV	Moyenne	CV
X_1	**0.0182**	0.2214	0.0358	0.0998	0.0156	0.2926
X_2	**0.1630**	0.0621	0.1546	0.0366	0.1535	0.0362
X_3	**0.0848**	0.0936	0.0742	0.0992	0.0683	0.0855
X_4	**0.3288**	0.0488	0.2998	0.0616	0.1832	0.0273
X_5	**0.0189**	0.2092	0.0364	0.1000	0.0153	0.2992
X_6	**0.0550**	0.1172	0.0491	0.1030	0.0399	0.1465

Le cadre présent est bien adapté à l'application de l'estimateur $\hat{\delta}_i^{\mathrm{IS}}$. En effet, les distributions d'entrée sont gaussiennes et la distribution de sortie est très bien apprise avec l'estimation par noyau gaussien (voir figure [4.6]). En outre, les estimations de $\hat{\delta}_i^{\mathrm{IS}}$ sont proches des valeurs de références assimilées à l'estimateur *double-boucle* $\hat{\delta}_i^{\mathrm{PDF}}$. En revanche, l'estimateur $\hat{\delta}_4^{\mathrm{ME}}$ de l'indice de Borgonovo de X_4 présente un biais notable. L'origine de cette imprécision est le choix des contraintes inhérent à l'application du principe du maximum d'entropie. Ici, les contraintes fixées sont associées aux moments fractionnaires d'ordre $\alpha_k \in \{\frac{2k}{m}\}_{k=1}^{m}$ avec $m = 3$. Comme mentionné dans la section 2.2.2.2, l'optimisation du nombre et des valeurs des paramètres α_k contrainte est un problème difficile. Néanmoins, l'erreur de précision peut être réduite en changeant de manière naïve ces paramètres. En effet, fixer le paramètre m à 5, i.e, passer de 9 à 25 contraintes permet d'obtenir une approximation de l'ordre de $\hat{\delta}_4^{\mathrm{ME}} \approx 0.24$, ce qui se rapproche de la valeur de référence.

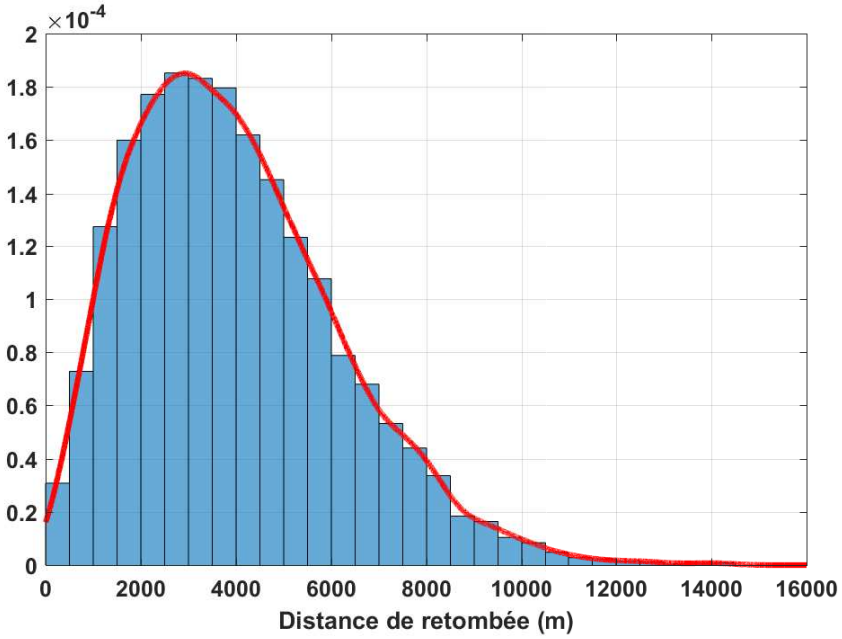


FIGURE 4.6. REPRÉSENTATION DE L'ESTIMATEUR PAR NOYAU GAUSSIEN (COURBE ROUGE) DE LA DISTRIBUTION DE SORTIE DU MODÈLE DE LANCEUR, OBTENU À PARTIR D'UN ÉCHANTILLON DE TAILLE $N = 5000$ (HISTOGRAMME BLEU).

TABLE 4.10. ESTIMATIONS DES INDICES DE SOBOL DU PREMIER ORDRE ET TOTAUX DU MODÈLE DE LANCEUR.

Entrée	S_i		S_{T_i}	
	Moyenne	CV	Moyenne	CV
X_1	0.0046	16.8374	0.0153	0.1154
X_2	0.2576	0.2436	0.6687	0.0888
X_3	0.0682	1.0210	0.3892	0.1049
X_4	0.1924	0.3643	0.2126	0.1428
X_5	0.0043	17.8161	0.0169	0.1055
X_6	0.0122	6.5117	0.2229	0.1003

D'autre part, il apparait que les indices de Borgonovo δ_i et les indices de Sobol d'ordre 1 S_i mènent à la même conclusion, à savoir la prédominance des entrées X_2 et X_4. Toutefois, on remarque que $\delta_4 > \delta_2$ tandis que $S_4 < S_2$ ce qui laisse penser que l'impact de X_4 est sous-estimé relativement à X_2 lorsque seule la variance est considérée.

Enfin, les indices de Sobol totaux S_{T_i} apportent une information supplémentaire, à savoir que les entrées X_3 et X_6 possèdent également un impact notable sur le modèle de sortie si leurs interactions avec les autres variables sont prises en compte. Ceci n'a pu être détecté avec les indices de Borgonovo d'ordre 1. En outre, il serait intéressant de comparer les indices S_{T_i} aux indices de Borgonovo d'ordres supérieurs. Malheureusement, les méthodes présentement utilisées ne sont pas suffisamment robustes pour estimer ces indices, du fait de la malédiction de la dimension.

4.4.3 Synthèse

Nous terminons ce chapitre par une synthèse des avantages et inconvénients des principales méthodes d'estimations des indices de Borgonovo du premier ordre rencontrées antérieurement.

♦ **Dans le cas idéal où le budget est illimité**, il est approprié d'utiliser les méthodes *double-boucle* (cf. section 3.4.1) puisque celles-ci sont particulièrement robustes et précises.

Cependant, dans les cas d'applications concrets, la quantité de données sur l'entrée $\mathbf{X}$ est bien évidemment restreinte et l'obtention d'observations de sortie $Y = \mathcal{M}(\mathbf{X})$ est également limitée puisque l'évaluation du code de calcul $\mathcal{M}$ est potentiellement très coûteuse. Il convient alors de recourir à des méthodes alternatives.

♦ **Dans le cas où les entrées sont indépendantes** la méthode basée sur la transformation de Nataf est particulièrement compétitive puisqu'elle requiert très peu d'appels au code de calcul $\mathcal{M}$. Néanmoins, cette approche est fondée sur plusieurs hypothèses/approximations concernant le modèle $Y = \mathcal{M}(\mathbf{X})$ (cf. section 3.4.2.2) dont la pertinence vis à vis du système réel est laissée à l'appréciation de l'utilisateur.

♦ **Dans le cas où les entrées sont corrélées**, les deux approches introduites dans ce chapitre peuvent être considérées. Ces dernières sont par ailleurs très permissives puisqu'elles n'imposent aucunes conditions sur le modèle, outre le fait de connaitre la distribution de l'entrée. De plus, elles permettent d'obtenir une estimation de tous les indices de Borgonovo d'ordre 1 à partir d'un seul jeu de données $\{(\mathbf{X}^n, Y^n)\}_{n=1}^N$, et donc N appels au modèle de calcul (indépendamment de la dimension d du modèle). Elles se distinguent ainsi de la procédure d'estimation classique des indices de Sobol qui nécessite de subdiviser l'échantillon $\{\mathbf{X}^n\}$ et requiert en outre $N(d+1)$ exécutions du code $\mathcal{M}$ [Sobol, 2001]. La méthode d'échantillonnage préférentiel présentée dans la section 4.2 peut cependant révéler des imprécisions **dans le cas où les distributions impliquées sont à queue lourde et/ou à support borné.** Dans ce dernier cas, la seconde méthode basée sur le principe du maximum d'entropie détaillée dans la section 4.3 est particulièrement adaptée et montre de bons résultats sur des cas analytiques. De ce fait, elle sera utilisée dans le prochain chapitre portant sur l'estimation des indices de Borgonovo dans un contexte fiabiliste, plus précisément, conditionnellement au dépassement d'un seuil S par la sortie Y. En revanche, elle repose sur le choix de paramètres α_k dont les valeurs et le nombre sont difficiles à optimiser (voir section 2.2.2.2). En outre, nous avons pu observer sur l'exemple du modèle de lanceur (cf. section 4.4.2) que la précision de cette technique peut être affectée par un mauvais ajustement de ces paramètres.

♦ **Extension en dimension supérieure ?** L'ensemble de ces méthodologies peuvent théoriquement être étendues en grande dimension pour estimer la totalité des mesures de Borgonovo. Malheureusement, ces méthodes échouent en pratique en raison de la malédiction de la dimension. L'examen des indices de Borgonovo d'ordre supérieur sera approfondi au chapitre 6.

Chapitre 5

Application à l'analyse de sensibilité fiabiliste d'un modèle boîte noire

Contents

5.1 Introduction et motivations 99

5.2 Deux indices de sensibilité complémentaires : $\bar{\eta}_i$ et δ_i^f 101

5.3 Estimation des indices $\bar{\eta}_i$ et δ_i^f 102

5.3.1 Comment échantillonner dans la région de défaillance ? 103

5.3.2 Schéma d'estimation de $\bar{\eta}_i$ et δ_i^f 103

5.4 Applications numériques 105

5.4.1 Retour sur le cas jouet 105

5.4.2 Un cas test analytique 105

5.4.3 Un oscillateur non linéaire à un degré de liberté 106

5.4.4 Le modèle de retombée d'un étage de lanceur 108

5.5 Généralisation : analyse de sensibilité régionale et conditionnelle 109

5.5.1 Une mesure de dissimilarité plus générale 109

5.5.2 Impact sur une fonction de la sortie Y 110

5.5.3 Généralisation 110

5.6 Conclusion 112

5.1 Introduction et motivations

Dans le cadre de l'analyse de fiabilité d'un système complexe modélisé par une relation entrée-sortie $Y = \mathcal{M}(\mathbf{X})$, l'étude directe du modèle de sortie Y inhérente à l'analyse de sensibilité classique est remplacée par des quantités d'intérêt de natures différentes. En outre, l'*analyse de sensibilité fiabiliste* désigne l'ensemble des méthodes affiliées à une mesure de fiabilité comme la probabilité de défaillance p_f ou l'indice de fiabilité β (utilisé

dans la méthode FORM-SORM). Récemment, il a été proposé par [Raguet and Marrel, 2018] de classifier ces méthodes en deux catégories :

- Premièrement, l'analyse de sensibilité *ciblée* ou *régionale* considère l'ensemble du support de l'entrée $\mathbf{X}$ mais focalise l'étude sur une fonction de la sortie Y. Il peut s'agir par exemple de la fonction indicatrice d'une région restreinte du support de sortie comme l'intervalle $[S, +\infty[$ relatif à la probabilité de défaillance p_f, ou plus généralement un domaine critique C_Y. De nombreuses méthode d'analyse de sensibilité, locales ou globales, peuvent êtres adaptées à des mesures de fiabilité comme p_f avec des méthodes dédiées comme celles présentées à la section 2.4. Le lecteur intéressé peut se référer à la thèse [Chabridon, 2018] pour une bibliographie des principales méthodes d'analyse de sensibilité régionales.

- D'autre part, l'analyse de sensibilité *conditionnelle* vise à étudier l'impact des entrées $\mathbf{X}_i$ exclusivement dans le domaine critique C_Y, i.e, conditionnellement à celui-ci.

Ces deux points de vue peuvent mener à des conclusions drastiquement différentes lors de l'étude du comportement du système à la défaillance. Pour illustrer cette affirmation, considérons par exemple le modèle jouet suivant :

$$Y = X_1 + \mathbb{1}_{X_1>3}|X_2| \tag{5.1}$$

où X_1 et X_2 sont deux variables gaussiennes centrées de variance respective 1 et 5. La défaillance de ce système est supposée survenir lorsque lorsque $Y > 3$, i.e, lorsque $X_1 > 3$ par construction. Si l'on souhaite déterminer quelle est l'entrée la plus influence selon une perspective fiabiliste, la réponse dépend du point de vue considéré :

1 ♦ Si on s'intéresse à l'**impact d'une entrée sur la probabilité de défaillance**, alors X_1 est évidemment hautement influente tandis X_2 qui rentre en jeu seulement à la défaillance, ne joue aucun rôle.

2 ♦ Si on s'intéresse maintenant à l'**influence d'une entrée sous le régime de défaillance**, X_2 est heuristiquement plus importante que X_1 puisqu'elle présente une plus grande variance.

Ce chapitre se concentre sur deux indices de sensibilité fiabilistes, $\bar{\eta}_i$ et δ_i^f. La section 5.2 vise à présenter ces deux indices, dont la définition est intrinsèquement liée à la méthode d'analyse de sensibilité de Borgonovo [Borgonovo, 2007], et à mettre en évidence leur complémentarité au sens du paradigme de [Raguet and Marrel, 2018]. Nous verrons en effet, au travers du modèle jouet (5.1) que les indices $\bar{\eta}_i$ et δ_i^f permettent de satisfaire conjointement les deux objectifs **1** ♦ et **2** ♦. La question de l'estimation de ces mesures d'importance est abordée dans la section 5.3, suivie ensuite par des applications numériques présentées en section 5.4. Enfin, la section 5.5 explique comment étendre l'approche détaillée dans les sections 5.2 et 5.3 à un contexte plus général.

5.2 Deux indices de sensibilité complémentaires : $\bar{\eta}_i$ et δ_i^f

L'objet de cette section est de définir deux indices de sensibilité fiabilistes distincts en s'appuyant sur la méthode d'analyse de sensibilité globale de Borgonovo [Borgonovo, 2007]. Rappelons que les mesures d'importance de Borgonovo admettent la définition générale suivante introduite dans la section 3.4 :

$$\delta_i = \mathbb{E}\left[d_{\mathrm{TV}}\left(Y, Y \mid X_i\right)\right] \tag{5.2}$$

où d_{TV} désigne la distance en variation totale. Ce formalisme permet également de définir un indice de Borgonovo pour des distributions discrètes. Lorsque (X_i, Y) est absolument continu par rapport à une mesure produit $\lambda(\mathrm{d}x) \otimes \mu(\mathrm{d}y)$, la mesure δ_i peut s'interpréter comme une mesure de dépendance

$$\delta_i = d_{\mathrm{TV}}\left((X_i, Y), (X_i, Y')\right) \ , \tag{5.3}$$

où Y' est une copie de Y.

Un problème typique en fiabilité est l'étude de la probabilité de défaillance p_f. En outre, l'étude de l'influence d'une entrée X_i sur p_f (cf. objectif **1** ⧫) est une question qui a été largement abordée et qui a donné lieu à diverses méthodes d'analyse de sensibilité régionales, voir par exemple [Cui et al., 2010], [Li et al., 2012] ou encore [Lemaître et al., 2015]. Ce problème revient à regarder l'influence de X_i sur une fonction de la sortie, à savoir la variable aléatoire $\mathbb{1}_{Y>S}$. L'indice de Borgonovo généralisé est alors donné par l'indice de sensibilité régional suivant

$$\eta_i = \mathbb{E}\left[d_{\mathrm{TV}}(\mathbb{1}_{Y>S}, \mathbb{1}_{Y>S} \mid X_i)\right] = \mathbb{E}\left[\left|\mathbb{P}(Y > S) - \mathbb{P}(Y > S \mid X_i)\right|\right] \tag{5.4}$$

qui coïncide en fait, à un facteur multiplicatif $\frac{1}{2}$ près, avec l'indice proposé par Cui et al. [Cui et al., 2010]. L'inconvénient de cet indice est qu'il est non normalisé et majoré par le double de la probabilité de défaillance $2\mathbb{P}(Y > S)$. Pour obtenir un indice à valeur dans $[0, 1]$, le théorème de Bayes implique la relation suivante

$$\eta_i = 2\mathbb{P}(Y > S) \times d_{\mathrm{TV}}(X_i, X_i \mid Y > S) \tag{5.5}$$

obtenue par [Wang et al., 2018], ce qui permet de définir l'indice renormalisé suivant

$$\bar{\eta}_i = d_{\mathrm{TV}}(X_i, X_i \mid Y > S) = \frac{1}{2}\left\|f_{X_i} - f_{X_i|Y>S}\right\|_{L^1(\mathbb{R})}. \tag{5.6}$$

En complément de cette approche, il peut être pertinent de s'intéresser à l'influence de X_i conditionnellement à la défaillance du système (cf. objectif **2** ⧫). Pour cela, l'indice de sensibilité conditionnel suivant peut être considéré :

$$\delta_i^f := \mathbb{E}\left[d_{\mathrm{TV}}((Y \mid Y > S), (Y \mid Y > S, X_i))\right]. \tag{5.7}$$

Désignons par $(\tilde{X}_i, \tilde{Y})$ une variable aléatoire distribuée selon (X_i, Y) conditionné par $Y > S$. Lorsque (X_i, Y) est absolument continu,$(\tilde{X}_i, \tilde{Y})$ l'est également et (5.7) s'avère être un cas particulier de (5.3) :

$$\delta_i^f = \frac{1}{2}\left\|f_{\tilde{X}_i,\tilde{Y}} - f_{\tilde{X}_i} f_{\tilde{Y}}\right\|_{L^1(\mathbb{R}^2)}. \tag{5.8}$$

En ce qui concerne le modèle jouet (5.1), on a $Y > S$ si et seulement si $X_1 > S$: ceci implique directement $\mathbb{P}(Y > S \mid X_1) = \mathbb{1}_{X_1>S}$ et $\mathbb{P}(Y > S \mid X_2) = \mathbb{P}(Y > S)$ et donc

$$\bar{\eta}_1 = 1 - \mathbb{P}(X_1 > S) \approx 0.9987 \quad \text{and} \quad \bar{\eta}_2 = 0.$$

Cela confirme l'intuition que, en ce qui concerne l'atteinte ou non du régime de défaillance, X_1 est extrêmement influent alors que X_2 n'a aucun impact. D'autre part, sur ce cas simple, les indices δ_i^f peuvent être directement calculés par intégration numérique :

$$\delta_1^f \approx 0.0781 \text{ and } \delta_2^f \approx 0.7686.$$

Ainsi, lorsque la défaillance a lieu, X_2 devient beaucoup plus influente que X_1. Cet exemple simple illustre la complémentarité des indices $\bar{\eta}_i$ et δ_i^f selon une perspective fiabiliste. Le but de la prochaine section est de montrer que ces mesures peuvent être estimées simultanément à partir à partir d'un échantillon commun distribué selon $\tilde{\mathbf{X}}$, i.e, un échantillon $\{\mathbf{X}^k\}$ vérifiant $\mathcal{M}(\mathbf{X}^k) > S$. Dans ce chapitre, cet échantillon est généré à l'aide de l'algorithme SMC adaptatif 4, usuellement utilisé pour estimer la probabilité de défaillance p_f (cf. section 2.4). Plus généralement, nous montrons que sous réserve d'avoir estimé cette probabilité avec une méthode d'échantillonnage, l'estimation de $\bar{\eta}_i$ et δ_i^f peut s'effectuer par simple post-traitement.

5.3 Estimation des indices $\bar{\eta}_i$ et δ_i^f

Dans toute la suite, $\tilde{\mathbf{X}} = (\tilde{X}_1, \ldots, \tilde{X}_d)$ désigne une variable aléatoire distribuée selon $\mathbf{X}$ conditionnée par $Y > S$ et $\tilde{Y} = \mathcal{M}(\tilde{\mathbf{X}})$ la sortie correspondante. De plus, il est supposé que pour tout i, (X_i, Y) est absolument continu, ce qui implique que $(\tilde{X}_i, \tilde{Y})$ l'est également. La question centrale de cette section est l'estimation de l'indice conditionnel δ_i^f défini par Eq.(5.7) et de l'indice régional $\bar{\eta}_i$ défini par

$$\bar{\eta}_i = \frac{1}{2} \left\| f_{X_i} - f_{\tilde{X}_i} \right\|_{L^1(\mathbb{R})}. \tag{5.9}$$

En ce qui concerne l'estimation de δ_i^f, deux points sont à souligner :

1. La distribution de sortie $f_{\tilde{Y}}$ est à support dans $[S, +\infty[$. La méthode d'estimation par maximum d'entropie semble donc plus adaptée que l'estimation par noyau gaussien qui peut souffrir d'effets de bord.
2. Les composantes de l'entrée $\tilde{\mathbf{X}}$ sont corrélées du fait du conditionnement.

Ainsi, au regard de la synthèse établie à la section 4.4.3, l'estimation de δ_i^f sera effectuée dans cette section avec la méthode détaillée dans la section 4.3 et fondée sur la représentation des mesures d'importance de Borgonovo en terme de densité de copule :

$$\delta_i^f = \frac{1}{2} \int_{[0,1]^2} \left| c_i^f(u, v) - 1 \right| \mathrm{d}u\mathrm{d}v, \tag{5.10}$$

où c_i^f est la densité de copule de $(\tilde{X}_i, \tilde{Y})$, i.e., la PDF de $(F_{\tilde{X}_i}(\tilde{X}_i), F_{\tilde{Y}}(\tilde{Y}))$.

Pour résumer, l'estimation de $\bar{\eta}_i$ et δ_i^f repose respectivement sur l'estimation des PDFs $f_{\tilde{X}_i}$ et c_i^f. Or, un point essentiel est que les techniques d'estimations présentées dans les chapitres précédents s'appuient sur la connaissance de la loi de l'entrée, et plus précisément sur le fait de savoir échantillonner selon cette distribution. Dans le cadre présent, la loi de $\tilde{\mathbf{X}}$ est inconnue, et son inférence nécessite donc les méthodes dédiées présentées dans la section 2.4. Le choix de la méthode est discuté dans la prochaine section.

5.3.1 Comment échantillonner dans la région de défaillance ?

Le présent problème, à savoir comment échantillonner selon $\tilde{\mathbf{X}}$, est intrinsèquement lié au problème d'estimation de la probabilité d'évènements rares précédemment abordé dans la section 2.4. La méthode la plus naïve pour échantillonner selon $\tilde{\mathbf{X}}$, i.e, générer des échantillons à la défaillance, est la méthode du rejet. Celle-ci consiste, pour un échantillon $\{\mathbf{X}^n\} \overset{\text{i.i.d}}{\sim} f_{\mathbf{X}}$ donné, à conserver les échantillons vérifiant $\mathcal{M}(\mathbf{X}^n) > S$. Toutefois, cette procédure de rejet implique un coût de calcul considérable dans le cas où la probabilité de défaillance est faible. Dès lors, les méthodes basées sur la recherche du point de conception ou la méthode de *Cross-Entropy* peuvent être considérées. Par exemple, [Yun et al., 2018] propose une méthode d'échantillonnage se déroulant en deux étapes : la première vise à construire une distribution d'échantillonnage auxiliaire à l'aide d'approximations FORM successives, la seconde consiste à utiliser cette distribution comme la loi initiale d'une chaine de Metropolis-Hastings ciblant la loi $f_{\tilde{\mathbf{X}}}$. Néanmoins, déterminer une distribution auxiliaire efficace est difficile lorsque le domaine de défaillance est non connexe ou lorsque la dimension est grande. C'est pourquoi cette section privilégie l'algorithme SMC adaptatif (cf. section 2.4.3) qui figure parmi les méthodes de Monte-Carlo par chaînes de Markov les plus performantes, bien que le budget de simulation inhérent à son utilisation peut être plus important que celui des méthodes précédemment citées.

Les paramètres de l'algorithme SMC 4 sont N_x, ρ, A_x et $\mathcal{Q}$ et correspondent respectivement au nombre de particules, au niveau de quantile, au nombre d'étapes pour l'échantillonneur de Metropolis-Hastings et au noyau d'exploration utilisé. Au terme de l'algorithme 4, un échantillon de particules $(\mathbf{X}_m^1, \dots, \mathbf{X}_m^{N_x})$ est disponible. Ces particules sont non indépendantes et approximativement distribuées selon $\mathbf{X} \mid Y > \gamma_m$, où γ_m est un ρ-quantile de la distribution de sortie tel que $\gamma_m > S$ et où m désigne le nombre (aléatoire) d'étapes de l'algorithme [Cérou et al., 2012]. Par ailleurs, à ce stade le nombre d'appels au modèle est égal à $N_x(1 + mA_x)$.

Dans le but de diminuer la corrélation d'une part, et d'obtenir de nouvelles particules distribuées selon $\mathbf{X}|Y > S$ d'autre part, une étape additionnelle est considérée. Tout d'abord, les N_x particules $\mathbf{X}_m^k$ sont dupliquées en N particules, puis transformées en appliquant A fois l'algorithme de Metropolis-Hastings ciblant la distribution $\mathbf{X}|\mathcal{M}(\mathbf{X}) > S$. Cela fournit un échantillon $(\tilde{\mathbf{X}}^1, \dots, \tilde{\mathbf{X}}^N)$ approximativement i.i.d selon $\tilde{\mathbf{X}}$ ainsi que les sorties correspondantes $\tilde{Y}^k = \mathcal{M}(\tilde{\mathbf{X}}^k)$. In fine, cette procédure nécessite $N \times A + N_x(1 + mA_x)$ appels au code de calcul $\mathcal{M}$.

5.3.2 Schéma d'estimation de $\bar{\eta}_i$ et δ_i^f

À l'issue de la procédure d'échantillonnage détaillée dans la section précédente, on dispose d'une collection de pseudo réalisations de $\tilde{\mathbf{X}}$. Cet échantillon peut alors être utilisé

pour inférer les deux PDFs $f_{\tilde{\mathbf{X}}_i}$ et c_i^f avec le principe du maximum d'entropie :

1 - Génération de réalisations d'entrée. En utilisant la procédure d'échantillonnage décrite dans la section 5.3.1, obtenir $(\tilde{\mathbf{X}}^1, \dots, \tilde{\mathbf{X}}^N)$ approximativement i.i.d selon $f_{\tilde{\mathbf{X}}}$ et les observations de sortie correspondantes $\tilde{Y}^k = \mathcal{M}(\tilde{\mathbf{X}}^k)$.

2 - Estimation des contraintes. Considérer $\tilde{n}, n \in \mathbb{N}$ et des nombres réels $\alpha_1 < \cdots < \alpha_{\tilde{n}}$ et $\beta_1 < \cdots < \beta_n$. Utiliser l'échantillon $\{(\tilde{X}_i^k, \tilde{Y}^k)\}_{k=1}^N$ pour estimer les moments fractionnaires associés aux paramètres α, β :

$$\hat{\mu}_{r,s} := \frac{1}{N} \sum_{k=1}^{N} \left(\hat{F}_{\tilde{X}_i}(\tilde{X}_i^k)\right)^{\alpha_r} \left(\hat{F}_{\tilde{Y}}(\tilde{Y}^k)\right)^{\alpha_s}, \quad r, s = 1, \dots, \tilde{n} \ ,$$

où $\tilde{F}_{\tilde{X}_i}$ et $\tilde{F}_{\tilde{Y}}$ sont les estimateurs empiriques des CDFs de $\tilde{X}_i$ et $\tilde{Y}$, associés à l'échantillon $\{(\tilde{X}_i^k, \tilde{Y}^k)\}_{k=1}^N$, et

$$\hat{\mu}_t^i := \frac{1}{N} \sum_{k=1}^{N} (\tilde{X}_i^k)^{\beta_t}, \ t = 1, \dots, n \ .$$

3 - Estimation de densité par maximisation de l'entropie. Estimer $f_{\tilde{X}_i}$ et c_i^f respectivement par

$$\begin{cases} \hat{f}_{\tilde{X}_i} = \underset{f \in L^1(\mathrm{Supp}(\tilde{X}_i))}{\mathrm{Argmin}} \ H(f) \\ \text{tel que } \displaystyle\int_{\mathrm{Supp}(\tilde{X}_i)} x^{\beta_t} f(x) \mathrm{d}x = \hat{\mu}_t^i, \ t = 1, \dots, n. \end{cases}$$

$$\begin{cases} \hat{c}_i^f = \underset{f \in L^1([0,1]^2)}{\mathrm{Argmin}} \ H(f) \\ \text{tel que } \displaystyle\int_{[0,1]^2} x^{\alpha_r} y^{\alpha_s} f(x,y) \mathrm{d}x\mathrm{d}y = \hat{\mu}_{r,s}, \ r, s = 1, \dots, \tilde{n}. \end{cases}$$

4 - Estimation des indices. Utiliser les estimateurs $\hat{f}_{\tilde{X}_i}$ et $\hat{c}_i^f$ pour obtenir les estimations de $\bar{\eta}_i$ et δ_i^f définies comme suit :

- Pour $\bar{\eta}_i$, estimer l'intégrale unidimensionnelle $\|f_{X_i} - \hat{f}_{\tilde{X}_i}\|_{L^1(\mathbb{R})}$ par intégration numérique.
- Pour δ_i^f, générer $\{(U_1^k, U_2^k)\}_{k=1}^{N'} \overset{\text{i.i.d}}{\sim} U([0,1]^2)$ et considérer l'estimateur Monte-Carlo

$$\hat{\delta}_i^f = \frac{1}{2N'} \sum_{k=1}^{N'} |\hat{c}_i^f(U_1^k, U_2^k) - 1| \ . \tag{5.11}$$

Ce schéma d'estimation permet une estimation simultanée de $\bar{\eta}_i$ et δ_i^f (ainsi que de la probabilité de défaillance p_f) à partir d'une unique procédure SMC : en effet, après l'étape d'initialisation aucune évaluation supplémentaire du modèle de calcul n'est nécessaire. En outre, le nombre (aléatoire) d'appels au modèle $\mathcal{M}$ est $N_x + mA_xN_x + AN$, comme stipulé dans la section 5.3.1.

5.4 Applications numériques

Dans cette section, le schéma d'estimation proposé est appliqué à différents modèles de sortie : deux exemples analytiques pour lesquels les valeurs théoriques des mesures d'importance $\bar{\eta}_i$ et δ_i^f sont connues, un modèle d'oscillateur non linéaire à un degré de liberté avec $d = 6$ entrées distribuées selon des lois lognormales, et enfin, le modèle de lanceur introduit à la section 4.4.1.

Lorsque l'entrée $\mathbf{X} \sim N(\boldsymbol{\nu}, \boldsymbol{\Sigma})$ est gaussienne, un choix naturel de noyau d'exploration $\mathcal{Q}$ est le noyau suivant, appelé *mélangeur* de Crank Nicholson

$$T(\mathbf{x}, \cdot) \sim \mathbf{L}\left(\sqrt{1-a} \times \mathbf{L}^{-1}(\mathbf{x} - \boldsymbol{\nu}) + \sqrt{a}Z\right) + \boldsymbol{\nu} \ ,$$

avec $a \in]0, 1[$, $Z \sim N(0, I_d)$ et $\mathbf{L}$ la matrice triangulaire inférieur de la décomposition de Cholesky de $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^T$.

Comme dans les précédentes applications numériques, les écarts type des estimateurs sont estimés en exécutant à 100 reprises le schéma proposé et la précision d'un estimateur $\hat{\theta}$ est mesurée par la moyenne de l'écart relatif $\frac{\theta - \hat{\theta}}{\theta}$ lorsque les valeurs théoriques sont disponibles

5.4.1 Retour sur le cas jouet

Revenons pour commencer sur le cas jouet (5.1) de l'introduction, i.e., $Y = X_1 + \mathbb{1}_{X_1 > S}|X_2|$ où $S = 3$, X_1 et X_2 sont indépendantes, $X_1 \sim N(0, 1)$ et $X_2 \sim N(0, 5)$. La table 5.1 établit une comparaison entre les valeurs théoriques et les estimations obtenues avec la méthode proposée. En moyenne, chaque exécution dure 317 secondes et fait 34,640 appels au code de calcul. Au vu des différents écarts relatifs, nous pouvons affirmer que les estimations de δ_i^f et $\bar{\eta}_i$ sont proches de leur valeurs de référence respectives et présentent une variabilité raisonnable au regard du budget fixé.

5.4.2 Un cas test analytique

Considérons le modèle entre-sortie suivant :

$$Y = X_1 + X_2^2 \tag{5.12}$$

où $X_1, X_2 \overset{\text{i.i.d}}{\sim} N(0, 1)$ et pour lequel l'évènement de défaillance est défini par $\{Y > 15\}$. Les distributions de sortie inconditionnelle et conditionnelles sont connues : $Y \mid X_1$ suit une distribution du χ^2-distribution décalée de X_1 et $Y \mid X_2$ est distribuée selon une variable gaussienne de moyenne X_2 et de variance 1. Les expressions des PDFs sont donc connues et données par

$$f_Y(y) = \int_0^\infty \frac{e^{-\frac{(y-t)^2}{2} - \frac{t}{2}}}{2\pi\sqrt{t}} \mathrm{d}t \ ,$$

et

$$f_{Y|X_1}(y) = \frac{e^{-\frac{(y-X_1)}{2}}}{\sqrt{2\pi(y - X_1)}} \mathbb{1}_{y \geq X_1} \quad \text{et} \quad f_{Y|X_2}(y) = \frac{e^{-\frac{(y-X_2)^2}{2}}}{\sqrt{2\pi}} \ .$$

Dès lors, les valeurs théoriques des mesures d'importance (δ_1, δ_2), (δ_1^f, δ_2^f) et $(\bar{\eta}_1, \bar{\eta}_2)$ sont accessibles par intégration numérique. La probabilité de défaillance peut également

Table 5.1. estimations de δ_i^f et $\bar{\eta}_i$ du cas jouet (5.1). Ensemble de paramètres pour l'algorithme SMC : $N_x = 500$, $A_x = 3$, $\rho = 0.3935$, $a = 0.5$, $A = 5$ et $N = 3000$.

Entrée	δ_i	**Estimation** $\hat{\delta}_i^f$		
	(rang)	Moyenne (rang)	Sd	RD
X_1	**0.0781 (2)**	0.0930 (2)	0.0101	-0.1908
X_2	**0.7686 (1)**	0.7200 (1)	0.0077	0.0632
Entrée	$\bar{\eta}_i$	**Estimation** $\hat{\bar{\eta}}_i$		
	(rang)	Moyenne (rang)	Sd	RD
X_1	**0.9987 (1)**	0.9997 (1)	0.0095	-0.001
X_2	**0 (2)**	0.0315 (2)	0.0103	-

être évaluée et vaut 1.2387×10^{-4}. Ces valeurs de références sont comparées avec les estimations réunies dans la table 5.2. En moyenne, chaque application du schéma d'estimation proposé prend 350 secondes pour calculer l'ensemble des indices, et ce, avec 25,200 appels à la boîte noire.

Il apparait que les estimations de $\{\hat{\delta}_i^f\}$ respectent le bon ranking, à savoir $X_2 > X_1$. En revanche, l'estimation de δ_1^f montre une différence notable avec la valeur de référence. Cette imprécision est due au fait que les échantillons $\{\mathbf{X}^k\}$ obtenu avec la procédure SMC ne sont pas complètement indépendants et distribués selon $f_{\tilde{\mathbf{X}}}$ puisque seulement $A = 3$ étapes de de Metropolis-Hastings sont effectuées l'étape finale d'échantillonnage. En effet, en passant la valeur du paramètre A de 3 à 30, nous obtenons pour $\hat{\delta}_1^f$ une valeur moyenne 0.0206 avec un écart type de 0.0091 (ce résultat ne figure pas dans la table 5.2).

Dans cet exemple, les indices de sensibilité conditionnels $\{\delta_i^f\}$ permettent de détecter un changement drastique dans la classification de l'influence. En effet, la contribution de la première entrée X_1 devient négligeable sous le régime de défaillance alors qu'elle est la plus influente lorsque l'on regarde le comportement global du système (cf. indices $\{\delta_i\}$). Cela illustre également l'importance de l'analyse de sensibilité fiabiliste, puisqu'une entrée du modèle peut apparaitre comme négligeable de manière globale et s'avérer fortement impactante lorsque l'on se focalise sur une partie restreinte du support de sortie. Les indices ciblés $\{\bar{\eta}_i\}$ mènent ici à la même conclusion puisque l'influence de X_2 sur la probabilité de défaillance est prédominante avec un indice $\bar{\eta}_2$ proche de 1.

5.4.3 Un oscillateur non linéaire à un degré de liberté

Le troisième exemple est un oscillateur non linéaire à un degré de liberté [Bucher et al., 1989] constitué d'une masse m et de deux ressorts de longueur à vide r et de rigidités respectives c_1 et c_2. Il est soumis à une force d'amplitude F pendant une durée aléatoire t. La sortie du modèle s'exprime de la façon suivante

$$Y = -3r + \left| \frac{2F}{c_1 + c_2} \sin\left(\sqrt{\frac{c_1 + c_2}{m}} \frac{t}{2} \right) \right| , \tag{5.13}$$

TABLE 5.2. ESTIMATIONS DE δ_i^f ET $\bar{\eta}_i$ DU MODÈLE (5.12). ENSEMBLE DE PARAMÈTRES POUR L'ALGORITHME SMC : $N_x = 300$, $A_x = 3$, $\rho = 0.5507$, $a = 0.5$, $A = 3$ ET $N = 5000$.

Entrée	δ_i **(rang)**	δ_i^f **(rang)**	**Estimation** $\hat{\delta}_i^f$ Moyenne (rang)	Sd	RD
X_1	**0.4930 (1)**	**0.001 (2)**	0.0721 (2)	0.0266	-71.1
X_2	**0.3049 (2)**	**0.4136 (1)**	0.3998 (1)	0.0343	0.0334
Entrée	\ \	$\bar{\eta}_i$ **(rang)**	**Estimation** $\hat{\bar{\eta}}_i$ Moyenne (rang)	Sd	RD
X_1	\	**0.2093 (2)**	0.2066 (1)	0.0605	0.0129
X_2	\	**0.9969 (1)**	0.9723 (2)	0.0567	0.0247

TABLE 5.3. DISTRIBUTION DES VARIABLES D'ENTRÉE DU MODÈLE D'OSCILLATEUR (5.13) (LA MOYENNE ET L'ÉCART TYPE SONT CEUX DE LA DISTRIBUTION NORMALE ASSOCIÉE).

Entrée	**Distribution**	**Moyenne**	**Écart type**
c_1	Lognormale	2	0.2
c_2	Lognormale	0.2	0.02
m	Lognormale	0.6	0.05
r	Lognormale	1	0.05
t	Lognormale	1	0.2
F	Lognormale	1	0.2

i.e., l'écart entre le déplacement maximal du système et $3r$. Les six paramètres d'entrée c_1, c_2, r, m, t et F sont supposées être des variables aléatoires indépendantes et distribuées selon des variables lognormales dont les paramètres sont donnés dans la table 5.3. La défaillance du système est atteinte lorsque la sortie Y dépasse le seuil 0 et la probabilité de défaillance associée vaut environ 9×10^{-5}.

La table 5.4 réunit les estimations de δ_i^f et $\bar{\eta}_i$ obtenues avec la méthodes proposée. En moyenne, chaque application du schéma d'estimation nécessite 960 secondes de calcul et 51,725 appels au modèle boîte noire. Il apparait que le ranking global donné par les mesures δ_i, à savoir $X_4 < X_2 < X_5 < X_1 < X_6 < X_3$, diffère drastiquement de la classification fournie par les indices de sensibilité conditionnels δ_i^f. En particulier, l'entrée la plus influente $X_3 = r$ devient négligeable conditionnellement à la défaillance. En ce qui concerne les indices ciblés $\bar{\eta}_i$ les changements sont plus nuancés puisque ceux-ci donnent un ranking similaire au ranking global, la prédominance de X_1 et X_2 mise à part.

À l'instar du deuxième exemple (5.12), la variabilité des estimations est non négligeable. Ici, les entrées sont distribuées selon des lois lognormales et il n'y a pas de noyau d'exploration $\mathcal{Q}$ naturel comme le *mélangeur* de Crank Nicholson dans le cas gaussien. Dans le cas présent, un candidat est proposé en ajoutant un bruit gaussien d'écart type identique à ceux des entrées. Avec ce choix, nous obtenons un taux d'acceptation de

Table 5.4. estimations de δ_i^f et $\bar{\eta}_i$ du modèle d'oscillateur (5.13). Ensemble de paramètres pour l'algorithme SMC : $N_x = 500$, $A_x = 10$, $\rho = 0.1813$, $A = 10$ et $N = 3000$.

Entrée	**Estimation** $\hat{\delta}_i$		**Estimation** $\hat{\delta}_i^f$		**Estimation** $\hat{\bar{\eta}}_i$	
	Moyenne (rang)	Sd	Moyenne (rang)	Sd	Moyenne (rang)	Sd
$X_1 = c_1$	0.0769 (3)	0.0066	0.0995 (1)	0.0210	0.8332 (2)	0.0653
$X_2 = c_2$	0.0231 (5)	0.0050	0.0322 (6)	0.0090	0.1352 (5)	0.0340
$X_3 = r$	0.4441 (1)	0.0063	0.0329 (5)	0.0117	0.6494 (3)	0.0690
$X_4 = m$	0.0219 (6)	0.0051	0.0343 (4)	0.0101	0.1306 (6)	0.0874
$X_5 = t$	0.0751 (4)	0.0075	0.0474 (3)	0.0150	0.3312 (4)	0.0710
$X_6 = F$	0.1554 (2)	0.0074	0.0871 (2)	0.0191	0.9078 (1)	0.0317

Table 5.5. estimations de δ_i^f et $\bar{\eta}_i$ du modèle d'oscillateur (5.13). Nouvel ensemble de paramètres pour l'algorithme SMC : $N_x = 500$, $A_x = 3$, $\rho = 0.4866$, $A = 10$ et $N = 3000$.

Entrée	**Estimation** $\hat{\delta}_i^f$		**Estimation** $\hat{\bar{\eta}}_i$	
	Moyenne (rang)	Sd	Moyenne (rang)	Sd
c_1	0.0674 (2)	0.0150	0.7949 (2)	0.0325
c_2	0.0275 (5)	0.0064	0.1131 (5)	0.0173
r	0.0346 (4)	0.0089	0.5651 (3)	0.0375
m	0.0267 (6)	0.0056	0.0459 (6)	0.0196
t	0.0366 (3)	0.0074	0.2812 (4)	0.0200
F	0.1147 (1)	0.0164	0.9205 (1)	0.0149

l'ordre de 20%–25% ce qui est en accord avec la valeur usuellement préconisée [Sherlock and Roberts, 2009] Ainsi, l'unique moyen d'améliorer les résultats actuels est d'augmenter le budget alloué à l'algorithme de Metropolis–Hastings en augmentant le paramètre A et en diminuant le paramètre ρ qui régule le nombre de seuils intermédiaires impliqués dans la procédure SMC. Au regard de la table 5.5 qui expose les nouveaux résultats obtenus, il apparait que la variabilité a effectivement été réduite. Le nouveau budget de simulation est d'environ 262,500 appels au modèle, ce qui est non négligeable. Néanmoins, cela reste substantiellement moins coûteux que le coût de simulation associé à une procédure de Monte-Carlo classique. Par ailleurs, le coût de simulation peut être diminué en ayant recours à des métamodèles permettant de remplacer la boîte noire $\mathcal{M}$ par une fonction moins coûteuse à évaluer. On peut citer entre autres la méthode AK-SS [Huang et al., 2016] qui combine l'algorithme SMC avec une procédure de Krigeage [Jones et al., 1998] dont le principe est de voir $\mathcal{M}$ comme la réalisation d'un processus gaussien.

5.4.4 Le modèle de retombée d'un étage de lanceur

Le dernier exemple est le modèle de retombée d'un étage de lanceur introduit à la section 4.4.1. Les résultats d'estimation des indices ciblés et conditionnels sont présentés

TABLE 5.6. ESTIMATIONS DE δ_i^f ET $\bar{\eta}_i$ DU MODÈLE DE LANCEUR. ENSEMBLE DE PARAMÈTRES POUR L'ALGORITHME SMC : $N_x = 800$, $A_x = 10$, $\rho = 0.4$, $A = 10$ ET $N = 5000$.

Entrée	**Estimation** $\hat{\delta}_i^{\mathbf{PDF}}$	**Estimation** $\hat{\delta}_i^f$		**Estimation** $\hat{\bar{\eta}}_i$	
	Moyenne (rang)	Moyenne (rang)	Sd	Moyenne (rang)	Sd
X_1	**0.0182 (6)**	0.0149 (5)	0.0060	0.1235 (4)	0.0367
X_2	**0.1630 (2)**	0.0848 (1)	0.0095	0.8218 (1)	0.0518
X_3	**0.0848 (3)**	0.0406 (2)	0.0084	0.5768 (2)	0.0514
X_4	**0.3288 (1)**	0.0143 (6)	0.0052	0.0722 (6)	0.0478
X_5	**0.0189 (5)**	0.0160 (4)	0.0059	0.1230 (5)	0.0330
X_6	**0.0550 (4)**	0.0284 (3)	0.0083	0.3809 (3)	0.0660

dans la table 5.6. Ceux-ci sont obtenus avec une moyenne de 750 s et 193,520 appels au modèle boîte noire et comparés avec les estimés des indices de Borgonovo traditionnels obtenus à la section 4.4.2. Le ranking global donné par l'estimateur *double-boucle* $\hat{\delta}_i^{\mathrm{PDF}}$ place X_4 comme la variable d'entrée la plus influence, suivie par X_2 puis X_3. En revanche, les approches ciblées et conditionnelles montrent que X_4 devient négligeable selon une perspective fiabiliste. Le reste du ranking reste relativement inchangé et semble indiquer que les entrées X_2 et X_3 sont les plus influentes au régime de défaillance du lanceur. À l'instar du cas test analytique (5.12), cet exemple illustre l'importance de mener une étude de sensibilité fiabiliste. En effet, l'impact d'une entrée du modèle (X_2 dans le cas présent) peut dépendre fortement du support de la distribution de sortie sur lequel l'étude se focalise. Enfin, comme pour les exemples précédents il convient de noter que les résultats présentent une variabilité non négligeable, mais qui peut être diminuée en augmentant le budget attribué à l'algorithme SMC.

5.5 Généralisation : analyse de sensibilité régionale et conditionnelle

Dans cette section, on explique comment étendre le schéma d'estimation proposé à la famille des divergences de Csiszár (cf. section 3.3.3), d'une part, et à l'étude de l'influence de X_i sur une fonction quelconque de Y d'autre part.

5.5.1 Une mesure de dissimilarité plus générale

Les mesures d'importance δ_i^f et $\bar{\eta}_i$ sont toutes deux définies à partir de la distance en variation total d_{TV}. Comme mentionné dans la section 3.3.3, d'autre mesures de dissimilarité existent, par exemple les divergences de Csiszár

Considérons $\phi : [0, +\infty[\to \mathbb{R}$ une fonction convexe telle que $\phi(1) = 0$. Rappelons que la divergence de Csiszár entre deux mesures de probabilité P et Q est définie par

$$d_\phi(P|Q) = \int \phi\left(\frac{dP}{dQ}\right) dQ \ ,$$

où P est supposée absolument continue de dérivée de Radon-Nikodym $\frac{dP}{dQ}$ par rapport à Q. À partir de cette divergence, nous pouvons définir la *mesure de dépendance de Csiszár* (CDM_f) entre deux variables aléatoires Z_1 et Z_2 comme

$$\mathrm{CDM}_\phi(Z_1, Z_2) = d_\phi((Z_1, Z_2)|(Z_1, Z_2')) \ , \tag{5.14}$$

où Z_2' est une copie de Z_2 indépendante de Z_1. À l'instar de l'indice de Borgonovo, cette mesure de dépendance peut être exprimée avec le formalisme des copules, i.e, en fonction de la densité de copule c de (Z_1, Z_2) (lorsque celle-ci existe) :

$$\mathrm{CDM}_\phi(Z_1, Z_2) = \int \phi(c(u,v))\mathrm{d}u\mathrm{d}v \ . \tag{5.15}$$

5.5.2 Impact sur une fonction de la sortie Y

Considérons une fonction $w : \mathcal{M}(\mathbb{R}^d) \to [0, +\infty[$ telle que $w(Y)$ soit intégrable et $\tilde{\mathbb{P}}^w$ la mesure de probabilité qui est absolument continue par rapport à $\mathbb{P}$ de dérivée de Radon-Nikodym $w(Y)$, i.e, l'unique mesure de probabilité définie sur $(\Omega, \mathcal{A})$ telle que

$$\tilde{\mathbb{P}}^w(A) = \frac{\mathbb{E}(w(Y)\mathbb{1}_A)}{\mathbb{E}(w(Y))}$$

pour tout mesurable $A \in \mathcal{A}$.

En adoptant la termonologie de [Raguet and Marrel, 2018], on peut généraliser les deux objectifs **1** ♦ et **2** ♦ comme suit :

Analyse de sensibilité régionale : quelle est l'influence de X_i sur $w(Y)$?

Analyse de sensibilité conditionnelle : quelle est l'influence de X_i sur Y sous $\tilde{\mathbb{P}}^w$?

Ce que nous avons traité précédemment correspond au cas où $w(y) = \mathbb{1}_{y>S}$. En effet, avec ce choix la mesure $\tilde{\mathbb{P}}^w \circ \mathbf{X}^{-1}$ coïncide avec la loi de la variable $\tilde{\mathbf{X}}$ définie précédemment :

$$\tilde{\mathbb{P}}^w(\mathbf{X} \in A) = \mathbb{P}(\mathbf{X} \in A \mid Y > S) = \mathbb{P}(\tilde{\mathbf{X}} \in A).$$

On généralise alors $\tilde{\mathbf{X}}$ en définissant la variable $\tilde{\mathbf{X}}^w = (X_1^w, \dots, X_d^w)$ de loi $\tilde{\mathbb{P}}^w \circ \mathbf{X}^{-1}$, et on définit $\tilde{Y}^w = \mathcal{M}(\tilde{\mathbf{X}}^w)$.

5.5.3 Généralisation

Au vu des équations (5.6) et (5.7) définissant $\bar{\eta}_i$ et δ_i^f, les précédentes extensions suggèrent de considérer les généralisations suivantes :

$$\eta_i^{\phi,w} = \mathrm{CDM}_\phi\left(X_i, w(Y)\right), \ \bar{\eta}_i^{\phi,w} = d_\phi\left(\tilde{X}_i^w, X_i\right) \quad \text{et} \quad \delta_i^{\phi,w} = \mathrm{CDM}_\phi\left(\tilde{X}_i^w, \tilde{Y}^w\right).$$

On suppose dans la suite que (X_i, Y) est absolument continu de densité $f_{X_i,Y}$, et que $(X_i, w(Y))$ est absolument continu par rapport à la mesure produit $\mathrm{d}x \otimes \mu(\mathrm{d}a)$ où μ est une mesure sur $\mathcal{M}(\mathbb{R}^d)$ de dérivée de Radon-Nikodym notée $f_{X_i,w(Y)}$. Lorsque $w(Y)$ est à

valeur dans $\mathbb{R}$, il faut voir μ comme la mesure de Lebesgue, mais ce formalisme général permet d'englober le cas où $w(Y)$ suit une distribution discrète : dans ce cas, il faut voir μ comme la mesure de comptage. Sous ces hypothèses, on a

$$\eta_i^{\phi,w} = \mathbb{E}\left[d_\phi(w(Y), w(Y) \mid X_i)\right] \ ,$$

et $(\tilde{X}^w, \tilde{Y}^w)$ est absolument continu de densité

$$f_{\tilde{X}_i^w, \tilde{Y}^w}(x,y) = \frac{w(y) f_{X_i,Y}(x,y)}{\mathbb{E}(w(Y))} \ .$$

Pour $w(y) = \mathbb{1}_{y>s}$ et $\phi(x) = |1-x|$, on obtient une relation entre $\eta_i^{\phi,w}$ et $\bar{\eta}_i^{\phi,w}$ similaire à (5.5) :

$$\eta_i^{\phi,w} = \mathbb{E}(w(Y)) \times \bar{\eta}_i^{\phi,w}.$$

En revanche, il n'est pas évident que cette relation soit licite en dehors de ce cas particulier. En outre, l'existence d'un lien entre $\eta_i^{\phi,w}$ et $\bar{\eta}_i^{\phi,w}$ n'est pas clair. Guidé par le choix effectué dans le cas $w(y) = \mathbb{1}_{y>S}$, on considère dans la suite l'indice $\bar{\eta}_i^{\phi,w}$ bien qu'il soit de prime abord moins naturel que $\eta_i^{\phi,w}$.

Pour échantillonner selon la distribution $\tilde{\mathbb{P}}^w$, l'algorithme de Metropolis–Hastings classique ciblant la densité $w(\mathcal{M}(\cdot)) f_{\mathbf{X}}(\cdot)/\mathbb{E}(w(Y))$ peut être utilisé. À l'instar du cas $w(y) = \mathbb{1}_{y>S}$, il est cependant difficile d'échantillonner directement de cette façon, et des distributions intermédiaires sont nécessaires. C'est pourquoi on a besoin d'une généralisation de l'algorithme SMC adaptatif 4 pour généraliser le schéma d'estimation établi à la section 5.3.2. Ceci peut être fait à l'aide de la méthodologie décrite dans [Moral et al., 2006].

Supposons désormais que l'on dispose d'un échantillon $(\tilde{\mathbf{X}}^{w,1}, \ldots, \tilde{\mathbf{X}}^{w,N}) \sim \mathbf{X}^w$ approximativement i.i.d. ainsi que des valeurs associées $\tilde{Y}^{w,k} = \mathcal{M}(\tilde{\mathbf{X}}^{w,k})$. Alors, les étapes 2 et 3 du précédent schéma d'estimation demeurent inchangées et fournissent un estimateur $\hat{f}_{\tilde{X}_i^w}$ de la densité $f_{\tilde{X}_i^w}$ de $\tilde{X}_i^w$, ainsi qu'un estimateur $\hat{c}^w$ de la densité de copule c^w de $(\tilde{X}_i^w, \tilde{Y}^w)$.

Une estimation $\hat{\bar{\eta}}_i^{\phi,w}$ de $\bar{\eta}_i^{\phi,w}$ peut être alors obtenue par intégration numérique de l'intégrale unidimensionnelle

$$d_\phi(\tilde{X}_i^w, X_i) = \int \phi\left(\frac{f_{\tilde{X}_i^w}(x)}{f_{X_i}(x)}\right) f_{X_i}(x)\mathrm{d}x \ \ ,$$

ou bien par approximation Monte-Carlo :

$$\hat{\bar{\eta}}_i^{\phi,w} = \frac{1}{N'} \sum_{k=1}^{N'} \phi\left(\frac{f_{\tilde{X}_i^w}(X_i^k)}{f_{X_i}(X_i^k)}\right) \ \ , \ \{X_i^k\}_{k=1}^{N'} \overset{\text{i.i.d}}{\sim} f_{X_i} \ .$$

Enfin, en ce qui concerne l'estimation de $\delta_i^{\phi,w}$, on peut considérer l'estimateur Monte-Carlo

$$\hat{\delta}_i^{\phi,w} = \frac{1}{N'} \sum_{k=1}^{N'} \phi\left(\hat{c}^w(U_1^k, U_2^k)\right) \ \ , \ \{(U_1^k, U_2^k)\}_{k=1}^{N'} \overset{\text{i.i.d}}{\sim} U([0,1]^2) \ .$$

5.6 Conclusion

L'objectif de ce chapitre était double :

♦ Étendre, puis appliquer la méthode d'analyse de sensibilité introduite par Emanuele Borgonovo dans un cadre fiabiliste. Cela à fait l'objet de l'introduction des indices δ_i^f et $\bar{\eta}_i$, ensuite étendus dans un cadre plus général dans la section 5.5. Nous avons illustré la complémentarité de ces deux indices au sens du formalisme proposée par [Raguet and Marrel, 2018]. En effet, la mesure d'importance $\bar{\eta}_i$ est un indice de sensibilité qualifié de *régional*, qui vise à étudier l'influence de X_i sur la probabilité de défaillance $p_f = \mathbb{P}(Y > S)$, et plus précisément sur la fonction $\mathbb{1}_{Y>S}$. En d'autres termes, l'étude est restreinte à la partie $[S, +\infty[$ du support de la sortie Y. La mesure δ_i^f quant à elle est un indice de sensibilité dit *conditionnel*, qui quantifie l'influence de X_i sur la sortie du modèle sous le régime critique du système, i.e, conditionnellement à l'évènement de défaillance.

♦ Appliquer la procédure d'estimation proposée à la section 4.3 à l'estimation de l'indice δ_i^f. Par ailleurs, la section 5.3 a permis de montrer que les indices δ_i^f et $\bar{\eta}_i$ peuvent être simultanément estimés à partir d'un échantillon commun $\{\mathbf{X}^k, Y^k\}$ vérifiant $\mathcal{M}(\mathbf{X}^k) = Y^k > S$. L'obtention d'un tel échantillon est ici effectué à partir d'une procédure SMC (cf. section 2.4.3). Le schéma d'estimation proposé à été appliqué sur plusieurs cas tests dans la section 5.4. Les résultats obtenus sont satisfaisant compte tenu du faible budget de simulation fixé dans ce contexte d'estimation d'évènements rares, mais présentent cependant une variabilité non négligeable. Celle-ci peut être réduite en augmentant le nombre d'appels à la boîte noire $\mathcal{M}$. En revanche, lorsque cette dernière est coûteuse à évaluer, cela peut donner lieu à un coût de simulation non raisonnable en pratique. Néanmoins, cette contrainte pourrait être surmontée via l'utilisation d'un métamodèle.

Chapitre 6

Étude des indices de Borgonovo d'ordres supérieurs

Contents

6.1 Introduction et motivations 113

6.2 Modèle *vigne* d'une densité de copule multivariée 115

6.2.1 Introduction aux distributions *vignes* 115

6.2.2 Modèles de R-*vignes* et formulation générale des distributions *vignes* 119

6.2.3 Sélection de la R-*vigne* 122

6.2.4 Application : estimation des indices de Borgonovo d'ordres supérieurs 124

6.3 Valeur de Shapley 125

6.3.1 Définition et propriétés de la valeur de Shapley 125

6.3.2 Application en analyse de sensibilité 126

6.4 Applications numériques 127

6.4.1 Le modèle gaussien (4.2) 128

6.4.2 L'oscillateur non linéaire à un degré de liberté 129

6.4.3 Le modèle de retombée d'un étage de lanceur 130

6.5 Conclusion 132

6.1 Introduction et motivations

Dans ce chapitre, on considère toujours un modèle entrée-sortie $Y = \mathcal{M}(\mathbf{X})$ où l'entrée $\mathbf{X} = (X_1, ..., X_d)$ est une variable aléatoire de dimension d telle que $(\mathbf{X}, Y)$ soit absolument continu. Dans le précédent chapitre 4, le but était de proposer de nouvelles approches pour estimer les indices de Borgonovo du premier ordre $\{\delta_i\}_{i=1}^d$ dont le rôle est de quantifier la sensibilité de la sortie Y vis à vis des différentes entrées X_i considérées individuellement. L'objet du présent chapitre est d'étudier les indices de Borgonovo d'ordres supérieurs $\delta_{\mathbf{u}}$

où $\mathbf{u} = (i_1, \ldots, i_r) \subset \{1, ..., d\}$, i.e, la sensibilité de Y relativement à un groupe de r entrée $\mathbf{X_u} := (X_{i_1}, \ldots, X_{i_r})$.

Dans la continuité de l'approche présentée dans la section 4.3, la représentation suivante de $\delta_{\mathbf{u}}$ en terme de copule peut être considérée :

$$\delta_{\mathbf{u}} = \frac{1}{2} \int_{[0,1]^{r+1}} \left| c_{\mathbf{u}}(u_1, \ldots, u_r, v) - \frac{f_{\mathbf{X_u}}\left(F^{-1}_{X_{i_1}}(u_1), ..., F^{-1}_{X_{i_r}}(u_r)\right)}{\prod_{k=1}^{r} f_{X_{i_k}}(F^{-1}_{X_{i_k}}(u_k))} \right| dv \prod_{k=1}^{r} du_k \ , \tag{6.1}$$

où r désigne le cardinal $|\mathbf{u}|$ de $\mathbf{u}$ et $c_{\mathbf{u}}$ la densité de copule de $(\mathbf{X_u}, Y)$, i.e, la densité du vecteur

$$(U_1, \ldots, U_r, V) := (F_{X_{i_1}}(X_{i_1}), \ldots, F_{X_{i_r}}(X_{i_r}), F_Y(Y)) \ .$$

L'estimation de $\delta_{\mathbf{u}}$ repose donc sur celle de la densité $c_{\mathbf{u}}$ qui est de dimension $|\mathbf{u}| + 1$. Le cas $|\mathbf{u}| = 1$, i.e, lorsque $\mathbf{u}$ est un singleton $\{i\}$, a déjà été étudié dans la section 4.3 où il est proposé d'estimer la densité de copule bivariée c_i au moyen du principe du maximum d'entropie associé à une collection de m^2 contraintes portant sur les moments du produit $U_i V$:

$$\mu_{k,l} := \mathbb{E}[U_i^{\alpha_k} V^{\alpha_l}] \ ,$$

où $\alpha_1 < \cdots < \alpha_m$. Adopter une démarche similaire pour un groupe quelconque d'entrée nécessite donc de considérer $m^{|\mathbf{u}|+1}$ contraintes, ce qui est difficilement praticable lorsque la dimension du problème augmente. Plus généralement, les techniques d'estimation non-paramétriques de densité présentées à la section 2.2.2 résistent mal à la *malédiction de la dimension*. Une alternative possible serait de recourir à une approche paramétrique. Malheureusement, malgré l'existence dans le cas bivarié une multitude de copules paramétriques, la généralisation en plus grande dimension de ces familles paramétriques n'est plus suffisamment exhaustive pour pouvoir prétendre représenter correctement des structures de dépendance complexes. Une alternative plus pertinente serait de recourir à la théorie des copules *vignes*. En effet, il a été démontré par [Bedford and Cooke, 2002] que toute copule de dimension n peut se réécrire comme un produit de $n(n-1)/2$ copules bidimensionnelles. La procédure permettant d'aboutir à la construction de cette décomposition est bien maîtrisée [Dissmann et al., 2013] et a déjà été employée pour estimer les indices de Borgonovo dans le cas où la sortie Y est multivariée [Zhou et al., 2018].

D'autre part, en supposant qu'il soit possible d'estimer $\delta_{\mathbf{u}}$ à un ordre $|\mathbf{u}|$ quelconque, il y a conséquemment $2^d - 1$ indices de Borgonovo à prendre en compte. En outre, à l'instar des indices de Sobol totaux S_{T_i} qui ont été introduit par [Homma and Saltelli, 1996], il semble nécessaire de chercher à synthétiser l'information fournie par ces $2^d - 1$ indices de Borgonovo. Plus précisément, une perspective intéressante serait de construire un indice permettant d'évaluer l'influence d'une entrée X_i (au sens de Borgonovo) en prenant en compte l'ensemble des interactions avec les autres variables. Pour cela, une idée est de considérer la notion de *valeur de Shapley*, initialement introduite en théorie des jeux par [Shapley, 1953] pour donner une répartition équitable des gains aux différents joueurs, et ce, en prenant en compte l'ensemble des collaborations possibles avec les autres joueurs.

Ce chapitre s'organise comme suit. La section 6.2 décrit la procédure de construction des modèles de copules et son application à l'estimation des indices de Borgonovo d'ordres supérieurs. Ensuite, la définition de la valeur de Shapley et son application dans le cadre

des indices de Borgonovo est discutée dans la section 6.3. Les différents résultats développés dans ce chapitre seront illustrés via des applications numériques dans la section 6.4. Enfin, une conclusion est donnée dans la section 6.5.

6.2 Modèle *vigne* d'une densité de copule multivariée

Un estimateur naturel de l'indice de Borgonovo $\delta_{\mathbf{u}}$ défini par (6.1) est l'estimateur Monte-Carlo

$$\frac{1}{2N'}\sum_{k=1}^{N'}\left|\hat{c}_{\mathbf{u}}(\mathbf{U}^k)-\frac{f_{\mathbf{X_u}}\left(F^{-1}_{X_{i_1}}(U^k_1),...,F^{-1}_{X_{i_r}}(U^k_r)\right)}{\prod_{l=1}^{r}f_{X_{i_l}}(F^{-1}_{X_{i_l}}(U^k_l))}\right|\ , \tag{6.2}$$

où $\{\mathbf{U}^k=(U^k_1,\dots,U^k_{r+1})\}_{k=1^{N'}}\sim\mathcal{U}([0,1]^{|\mathbf{u}|+1})$ et où $\hat{c}_{\mathbf{u}}$ désigne un estimateur de la densité de copule $c_{\mathbf{u}}$ du vecteur $(\mathbf{X_u},Y)$. L'objet de cette section est d'expliquer comment la théorie des copules *vignes* permet de construire un tel estimateur. Pour simplifier les notations, nous nous plaçons dans le cadre général d'un vecteur aléatoire $\tilde{\mathbf{X}}=(\tilde{X}_1,\dots,\tilde{X}_n)$ absolument continu de dimension $n\geq 3$. De plus, nous adoptons les notations suivantes :

- f_i et F_i, $i\in\{1,\dots,n\}$ désignent les PDFs et CDFs marginales.
- Pour tout sous-ensemble $D\subset\{1,\dots,n\}\backslash\{i\}$, $f_{i|D}$ et $F_{i|D}$ désignent les PDFs et CDFs conditionnelles de $\tilde{X}_i$ sachant $\tilde{\mathbf{X}}_D=(\tilde{X}_k,k\in D)$.

Par continuité des CDFs marginales F_i, le théorème de Sklar 2.6 assure l'existence d'une unique densité de copule $c_{1,..,n}:[0,1]^n\to\mathbb{R}^+$ telle que

$$f_{\tilde{\mathbf{X}}}(x_1,\dots,x_n)=c_{1,..,n}(F_1(x_1),\dots,F_n(x_n))\times f_1(x_1)\times\cdots\times f_n(x_n)\ , \tag{6.3}$$

où, rappelons le, $c_{1,..,n}$ correspond à la PDF du vecteur aléatoire

$$\mathbf{U}=(U_1,\dots,U_n):=(F_1(\tilde{X}_1),\dots,F_n(\tilde{X}_n))\ .$$

6.2.1 Introduction aux distributions *vignes*

L'objet de cette section est d'introduire les principaux ingrédients et outils de la théorie des copules *vigne* et de donner une intuition sur un exemple en dimension $n=4$. Le point de départ de la construction d'un modèle *vigne* est le théorème de Bayes qui permet de factoriser la PDF $f_{\tilde{\mathbf{X}}}$ de dimension n en un produit de n PDFs conditionnelles $f_{i|D}$ de dimension 1.

Exemple. Considérons le cas de la dimension $n=4$. Par conditionnements successifs, $f_{\tilde{\mathbf{X}}}$ peut s'écrire de la façon suivante :

$$\begin{aligned}f_{\tilde{\mathbf{X}}}(x_1,x_2,x_3,x_4)&=f_{4|123}(x_4|x_1,x_2,x_3)\times f_{123}(x_1,x_2,x_3)\ ,\\&=f_{4|123}(x_4|x_1,x_2,x_3)\times f_{3|12}(x_3|x_1,x_2)\times f_{12}(x_1,x_2)\ ,\\&=f_{4|123}(x_4|x_1,x_2,x_3)\times f_{3|12}(x_3|x_1,x_2)\times f_{2|1}(x_2|x_1)\times f_1(x_1)\ .\end{aligned}$$

La seconde étape consiste à exprimer chaque PDF conditionnelle $f_{i|D}$ en fonction de la PDF marginale f_i et de densités de copules bivariées conditionnelles. Cela repose sur la version conditionnelle du théorème de Sklar 2.6 :

$$\begin{aligned} &\forall i \neq j \in \{1,\dots,n\},\ D \subset \{1,\dots,n\}\backslash\{i,j\}\ , \\ f_{ij|D}(x_i,x_j|\mathbf{x}_D) &= c_{ij|D}(F_{i|D}(x_i|\mathbf{x}_D), F_{j|D}(x_j|\mathbf{x}_D);\mathbf{x}_D) \times f_{i|D}(x_i|\mathbf{x}_D) f_{j|D}(x_j|\mathbf{x}_D)\ . \end{aligned} \tag{6.4}$$

À ce stade, deux points importants sont à noter :

1. $c_{ij|D}$ est la copule de $(\tilde{X}_i, \tilde{X}_j)$ sous la mesure conditionnelle $\mathbb{P}(\cdot|\tilde{\mathbf{X}}_D = \mathbf{x}_D)$. Elle caractérise donc la loi du couple

$$(F_{i|D}(\tilde{X}_i|\tilde{\mathbf{X}}_D = \mathbf{x}_D), F_{j|D}(\tilde{X}_j|\tilde{\mathbf{X}}_D = \mathbf{x}_D))\ ,$$

   et non celle de (U_i, U_j) sachant $\mathbf{U}_D = (F_k(X_k), k \in D)$.

2. La densité de copule $c_{ij|D}$ prend $\mathbf{x}_D$ comme argument puisqu'elle peut être différente pour chaque réalisation $\mathbf{x}_D$ de $\mathbf{X}_D$. Néanmoins, il est couramment supposé que cette dépendance peut être ignorée [Czado, 2010, Haff et al., 2010], i.e,

$$c_{ij|D}(F_{i|D}(x_i|\mathbf{x}_D), F_{j|D}(x_j|\mathbf{x}_D);\mathbf{x}_D) = c_{ij|D}(F_{i|D}(x_i|\mathbf{x}_D), F_{j|D}(x_j|\mathbf{x}_D))\ ,$$

   pour toute réalisation $\mathbf{x}_D$ de $\mathbf{X}_D$. En d'autres termes, la densité de copule $c_{ij|D}$ dépend du conditionnement D seulement au travers des CDFs marginales conditionnelles $F_{i|D}$ et $F_{j|D}$. Cette hypothèse simplificatrice est effectuée pour des raisons purement pratiques, notamment pour simplifier l'inférence de la densité de copule $c_{ij|D}$. Cette hypothèse standard est adoptée dans la suite de ce chapitre.

Dans la suite, nous adoptons la notation suivante :

$$u_{i|D} := F_{i|D}(x_i|\mathbf{x}_D)\ . \tag{6.5}$$

Dès lors, il découle de Eq.(6.4) que chaque PDF $f_{i|D}$ peut se réécrire en fonction d'une PDF de la forme $f_{i|D'}$ où D' désigne l'ensemble D réduit d'un élément :

$$\begin{aligned} &\forall D \subset \{1,\dots,n\}\backslash\{i\},\ \forall j \in D,\ D' := D\backslash\{j\}\ , \\ f_{i|D}(x_i|\mathbf{x}_D) &= \frac{f_{ij|D'}(x_i,x_j|\mathbf{x}_{D'})}{f_{j|D'}(x_j|\mathbf{x}_D)} = c_{ij|D'}(u_{i|D'}, u_{j|D'}) \times f_{i|D'}(x_i|\mathbf{x}_{D'})\ . \end{aligned} \tag{6.6}$$

Exemple. Reprenons l'exemple prédécent. Pour alléger les prochaines équations, nous considérons temporairement les abréviations $f_{i|D} := f_{i|D}(x_i|\mathbf{x}_D)$ et $c_{ij|D} = c_{ij|D}(u_{i|D}, u_{j|D})$. En appliquant Eq.(6.6) de façon répétée, les PDFs $f_{2|1}$, $f_{3|12}$ et $f_{4|123}$ peuvent se réécrire comme suit :

$$\begin{aligned} f_{2|1} &= \frac{f_{12}}{f_1} = c_{12} \times f_2\ , \\ f_{3|12} &= \frac{f_{13|2}}{f_{1|2}} = \frac{c_{13|2} \times f_{1|2} \times f_{3|2}}{f_{1|2}} = c_{13|2} \times c_{23} \times f_3\ , \\ f_{4|123} &= \frac{f_{14|23}}{f_{1|23}} = \frac{c_{14|23} \times f_{1|23} \times f_{4|23}}{f_{1|23}} = c_{14|23} \times \frac{f_{24|3}}{f_{2|3}} = c_{14|23} \times c_{24|3} \times c_{34} \times f_4\ . \end{aligned}$$

Par unicité de la densité de copule, on obtient donc

$$\begin{aligned} c_{1234}(u_1,u_2,u_3,u_4) &= c_{1,2}(u_1,u_2)\times c_{2,3}(u_2,u_3)\times c_{3,4}(u_3,u_4) \\ &\quad\times c_{1,3|2}(u_{1|2},u_{3|2})\times c_{2,4|3}(u_{2|3},u_{4|3}) \\ &\quad\times c_{1,4|2,3}(u_{1|2,3},u_{4|2,3})\ . \end{aligned} \tag{6.7}$$

L'Eq.(6.7) procure une décomposition explicite de la densité de copule c_{1234} en un produit de densités de copules bidimensionnelles. Cette décomposition est appelée distribution *vigne*. À première vue il semble cependant très compliqué d'évaluer cette quantité puisque chaque densité de copule $c_{ij|D}$ dépend des CDFs conditionnelles $u_{i|D}$ et $u_{j|D}$ définies par (6.5). Néanmoins, les quantités $u_{i|D}$ en argument des copules conditionnelles peuvent toujours être réécrites en fonction des quantités $u_1,\dots,u_n$ initiales. Pour cela, on fait l'hypothèse que toutes les CDFs $F_{i|D}$ sont inversibles. Supposons que l'on cherche à réécrire la quantité $u_{i|D}$ qui intervient dans la factorisation comme le premier argument d'une copule $c_{ij|D}(\cdot,\cdot)$ ou comme le second argument d'une copule $c_{ji|D}(\cdot,\cdot)$. Le fonctionnement du mécanisme de factorisation implique l'existence (parmi les autres termes de la décomposition) d'une copule de type $c_{il|D'}$ ou d'une copule du type $c_{li|D'}$ avec $D'=D\setminus\{l\}$. Dans le premier cas, on peut écrire que :

$$\begin{aligned} u_{i|D} = F_{i|D}(x_i \mid \mathbf{x}_D) &= \int_{-\infty}^{x_i} f_{i|D}(x \mid \mathbf{x}_D)\,\mathrm{d}x\ , \\ &= \int_{-\infty}^{x_i} c_{il|D'}\Big(F_{i|D'}(x \mid \mathbf{x}_{D'}), F_{l|D'}(x_l \mid \mathbf{x}_{D'})\Big)\times f_{i|D'}(x \mid \mathbf{x}_{D'})\,\mathrm{d}x\ , \\ &= \int_0^{u_{i|D'}} c_{il|D'}\Big(t, u_{l|D'}\Big)\,\mathrm{d}t \text{ avec } t = F_{i|D'}(x \mid \mathbf{x}_{D'})\ , \\ &= h^{(g)}_{il|D'}(u_{i|D'}, u_{l|D'})\ , \end{aligned}$$

avec la fonction $h^{(g)}_{il|D'}$ qui est définie par :

$$h^{(g)}_{il|D'} : (u,v) \mapsto \int_0^u c_{il|D'}\Big(t,v\Big)\,\mathrm{d}t\ .$$

Cette fonction $h^{(g)}_{il|D'}$, appelée *h-fonction* à gauche de la copule $c_{il|D'}$, se calcule comme l'intégrale par rapport à la variable de gauche de la densité de copule.

Dans le second cas, on peut écrire que :

$$\begin{aligned} u_{i|D} = F_{i|D}(x_i \mid \mathbf{x}_D) &= \int_{-\infty}^{x_i} f_{i|D}(x \mid \mathbf{x}_D)\,\mathrm{d}x\ , \\ &= \int_{-\infty}^{x_i} c_{li|D'}\Big(F_{l|D'}(x_l \mid \mathbf{x}_{D'}), F_{i|D'}(x \mid \mathbf{x}_{D'})\Big)\times f_{i|D'}(x \mid \mathbf{x}_{D'})\,\mathrm{d}x\ , \\ &= \int_0^{u_{i|D'}} c_{li|D'}\Big(u_{l|D'}, t\Big)\,\mathrm{d}t \text{ avec } t = F_{i|D'}(x \mid \mathbf{x}_{D'})\ , \\ &= h^{(d)}_{li|D'}(u_{l|D'}, u_{i|D'})\ , \end{aligned}$$

avec la fonction $h^{(d)}_{li|D'}$ qui est définie par :

$$h^{(d)}_{li|D'} : (u, v) \mapsto \int_0^v c_{li|D'}\Big(u, t\Big)\,\mathrm{d}t \ .$$

Cette fonction $h^{(d)}_{li|D'}$, appelée *h-fonction* à droite de la copule $c_{li|D'}$, se calcule comme l'intégrale par rapport à la variable de droite de la densité de copule.

Utiliser une *h-fonction* (qu'elle soit à gauche ou à droite) permet d'exprimer $u_{i|D}$ en fonction des quantités $u_{i|D'}$ et $u_{l|D'}$ présentant un élément de moins dans leur ensemble de conditionnement. On peut alors à nouveau utiliser des *h-fonctions* pour décomposer $u_{i|D'}$ et $u_{l|D'}$, et ainsi de suite jusqu'à exprimer $u_{i|D}$ à partir des seules quantités initiales $u_1, \ldots, u_n$, moyennant l'imbrication de différentes couches de *h-fonctions*.

Exemple. Reprenons de nouveau l'exemple précédent. On a

$$\begin{aligned} u_{1|2,3} &= h^{(g)}_{1,2|3}(u_{1|3}; u_{2|3}) \ , \\ &= h^{(g)}_{1,2|3}(h^{(g)}_{1,3}(u_1; u_3); h^{(g)}_{2,3}(u_2; u_3)) \ , \end{aligned}$$

avec

$$\begin{aligned} h^{(g)}_{1,2|3}(u_{1|3}; u_{2|3}) &= \int_0^{u_{1|3}} c_{1,2|3}(t, u_{2|3})\mathrm{d}t \ , \\ u_{1|3} = h^{(g)}_{1,3}(u_1; u_3) &= \int_0^{u_1} c_{1,3}(t, u_3)\mathrm{d}t \ , \\ u_{2|3} = h^{(g)}_{2,3}(u_2; u_3) &= \int_0^{u_2} c_{2,3}(t, u_3)\mathrm{d}t \ . \end{aligned}$$

De même,

$$\begin{aligned} u_{4|2,3} &= h^{(d)}_{2,4|3}(u_{2|3}; u_{4|3}) \ , \\ &= h^{(g)}_{1,2|3}(h^{(g)}_{2,3}(u_2; u_3); h^{(d)}_{3,4}(u_3; u_4)) \ , \end{aligned}$$

avec

$$\begin{aligned} h^{(d)}_{2,4|3}(u_{2|3}; u_{4|3}) &= \int_0^{u_{4|3}} c_{2,4|3}(u_{2|3}, t)\mathrm{d}t \ , \\ u_{2|3} = h^{(g)}_{2,3}(u_2; u_3) &= \int_0^{u_2} c_{2,3}(t, u_3)\mathrm{d}t \ , \\ u_{4|3} = h^{(d)}_{3,4}(u_3; u_4) &= \int_0^{u_4} c_{3,4}(u_3, t)\mathrm{d}t \ . \end{aligned}$$

Remarques. Deux points peuvent être relevés au regard de l'exemple précédent en dimension $n = 4$. Tout d'abord, il n'y a pas unicité de la décomposition (6.7) étant donné que les ensembles de conditionnements dépendent des choix effectués par l'utilisateur lors des applications successives du théorème de Bayes et du théorème de Sklar conditionnel (6.6). Par ailleurs, même dans le cas de la dimension $n = 4$, le déroulement des calculs est relativement fastidieux. Il parait donc nécessaire de recourir à un outil graphique pour résumer ce procédé de factorisation. En outre, la prochaine section a pour but de présenter la définition générale des distributions vignes, d'une part, et d'expliquer l'approche graphique introduite par [Bedford and Cooke, 2001] permettant d'obtenir des expressions comme (6.7) de manière plus efficace.

6.2.2 Modèles de R-*vignes* et formulation générale des distributions *vignes*

D'une manière générale, en utilisant Eq.(6.6) de façon répétée, la densité de copule $c_{1,..,n}$ peut être décomposée en un produit de $n(n-1)/2$ copules bivariées. Cette décomposition porte le nom de distribution *vigne* et contient plus précisément :

- $(n-1)$ densités de copule bivariées non conditionnelles,
- $(n-2)$ densités de copule bivariées avec une variable de conditionnement.

$\vdots$

- 2 densités de copule bivariées avec $(n-3)$ variables de conditionnement.
- 1 densité de copule bivariée avec $(n-2)$ variables de conditionnement.

Une telle décomposition n'est pas unique et le nombre de distributions *vignes* possibles devient considérable lorsque la dimension n augmente. En pratique, la factorisation d'une copule peut être résumée par une méthode graphique établie par [Bedford and Cooke, 2001]. Cette représentation s'appuie sur une suite d'arbres non orientés $\{T_m\}_{m=1}^{n-1}$. Pour chaque m fixé, l'arbre $T_m = (N_m, E_m)$ est caractérisé par un ensemble N_m de sommets connectés entre eux par des arêtes rassemblées dans un ensemble E_m. Plus précisément, N_m est une collection de parties de $\{1,\ldots,n\}$, toutes de même cardinal (à m fixé). Il en est de même pour E_m, dont les éléments ont cependant un cardinal commun différent de celui de N_m. Nous nous proposons de commencer par énoncer (en des termes purement ensembliste) les règles de construction de cette suite d'arbres, notamment le processus permettant d'obtenir un arbre T_m à partir de son prédécesseur T_{m-1}. Ensuite, nous chercherons à faire comprendre comment de tels graphes permettent de résumer efficacement l'information inhérente à la structure d'une factorisation.

On appelle *vigne régulière* (*regular vine* ou R-*vigne*), une suite d'arbres $T_m = (N_m, E_m)$ vérifiant les spécifications suivantes :

(i) L'arbre T_1 est défini par $N_1 = \Big\{\{1\},\ldots,\{n\}\Big\}$ et E_1, un ensemble de $|E_1| = n-1$ arêtes permettant de constituer un *arbre couvrant*. Autrement dit, les $n-1$ arêtes contenues dans E_1 sont disposées de telle sorte à ce que tous les sommets de N_1 soient accessibles via les arêtes. Une conséquence directe est l'absence de cycle dans la représentation graphique de T_1. Une arête de T_1 est notée $e_{AB} = \{ij\}$ si elle rassemble par exemple les sommets $A = \{i\}$ et $B = \{j\}$ de N_1. L'union $A \cup B = \{ij\}$ des deux singletons permet d'annoter l'arête.

(ii) Pour m entre 2 et $n-1$, l'arbre T_m vérifie $N_m = E_{m-1}$. Ainsi, l'arbre T_m prend comme sommets les arêtes de l'arbre précédent.

(iii) Pour m entre 2 et $n-1$, l'arbre T_m vérifie également $|E_m| = n-m$ et, contrairement à l'arbre T_1 où tout ensemble d'arêtes conduisant à un arbre couvrant est envisageable, l'ensemble des arêtes d'un arbre T_m $(m \geq 2)$ est soumis à de plus

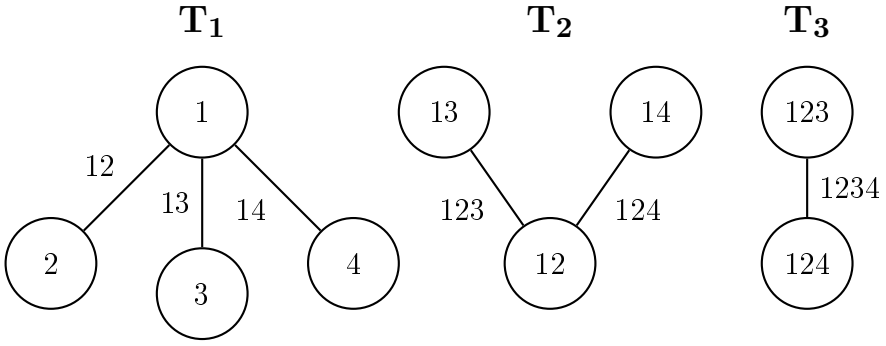
grandes restrictions. Une arête potentielle entre deux sommets $A = \{j_1, \ldots, j_m\}$ et $B = \{k_1, \ldots, k_m\}$ de N_m ne peut être rattachée à E_m que si elle vérifie une condition dite de "proximité". Il faut que $|A \cap B| = m - 1$, c'est-à-dire que ces deux sommets (qui, mathématiquement, ne sont rien d'autre que des ensembles de même cardinal égal à m) ne différent l'un de l'autre que par un seul élément. On note l_A l'élément spécifique à A, l_B celui spécifique à B et $l_1, \ldots, l_{m-1}$ les éléments communs constituant l'intersection de A et B. Si une telle condition est respectée, l'arête entre A et B est acceptée dans E_m et elle est notée $e_{AB} = A \cup B = \{l_A, l_B, l_1, ..., l_{m-1}\}$. Ainsi, l'ensemble E_m va rassembler $n - m$ arêtes, toutes de cardinal $m + 1$.

La condition de proximité est là pour imposer une cohérence entre la représentation sous forme de graphes (qui laisse beaucoup d'amplitude) et le mécanisme de factorisation (qui ne peut pas être conduit aussi librement). Cependant, il convient de noter que respecter la condition de proximité ne doit surtout pas être vu comme une application permettant d'obtenir T_m à partir de T_{m-1}. Pour un prédécesseur T_{m-1} donné, dans le cas général, et notamment lorsqu'on monte en dimension ($n \geq 5$), plusieurs arbres T_m peuvent convenir. La condition de proximité permet simplement de restreindre l'ensemble des arbres T_m pouvant découler de T_{m-1}.

Exemple. Considérons le cas de la dimension $n = 4$. Au vu de ce qui précède, on peut écrire que

$$T_1 : \quad N_1 = \Big\{\{1\},\{2\},\{3\},\{4\}\Big\} \qquad E_1 = \Big\{\{12\},\{13\},\{14\}\Big\}$$
$$T_2 : \quad N_2 = E_1 = \Big\{\{12\},\{13\},\{14\}\Big\} \qquad E_2 = \Big\{\{123\},\{124\}\Big\}$$
$$T_3 : \quad N_3 = E_2 = \Big\{\{123\},\{124\}\Big\} \qquad E_3 = \Big\{\{1234\}\Big\}$$

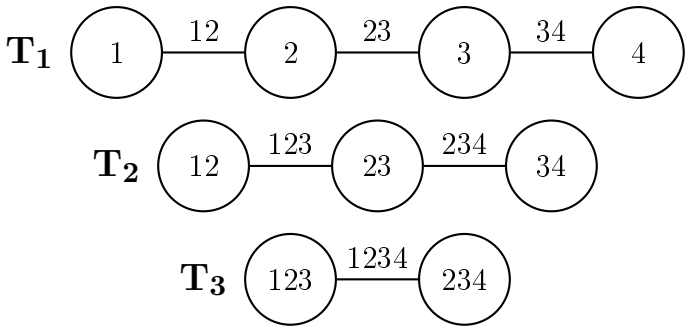
Ce qui donne la représentation graphique suivante :



Cette configuration particulière de *vigne régulière* où les arbres sont disposés en étoile avec un sommet central "pilote" (pour chaque arbre) est appelée *vigne canonique* (*canonical vine* ou C-*vigne*). Mais il aurait été tout à fait envisageable d'écrire que :

$$T_1 : \quad N_1 = \Big\{\{1\},\{2\},\{3\},\{4\}\Big\} \qquad E_1 = \Big\{\{12\},\{23\},\{34\}\Big\}$$
$$T_2 : \quad N_2 = E_1 = \Big\{\{12\},\{23\},\{34\}\Big\} \qquad E_2 = \Big\{\{123\},\{234\}\Big\}$$
$$T_3 : \quad N_3 = E_2 = \Big\{\{123\},\{234\}\Big\} \qquad E_3 = \Big\{\{1234\}\Big\}$$

On aurait alors obtenu une représentation graphique d'une toute autre forme, organisée cette fois-ci comme une chaîne de sommets. On parle de *vigne étirable* (*drawable vine* ou D-*vigne*) :



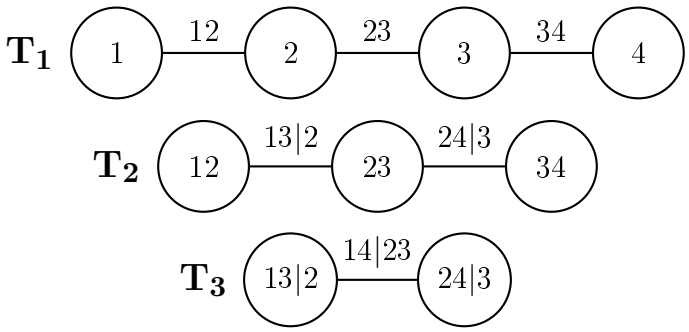
Il existe des R-vignes qui ne sont ni des C-vignes ni des D-vignes. Un exemple illustré en dimension $n = 7$ est fourni dans [Dissmann et al., 2013]. Examinons à présent en quoi la séquence $\{T_m\}_{m=1}^{n-1}$ est un résumé de la factorisation d'une copule multidimensionnelle.

D'après ce que nous avons expliqué plus haut, quel que soit l'arbre T_m $(m \geq 2)$ considéré, toute arête e provient de la connexion de deux sommets A et B et peut s'écrire $e \equiv e_{AB} = \{l_A, l_B, l_1, ..., l_{m-1}\}$, l'ensemble $\{l_1, ..., l_{m-1}\}$ étant l'ensemble vide pour T_2. On a défini l_A (resp. l_B) comme l'élément de $\{1, \ldots, n\}$ spécifique au sommet A (resp. B). Pour chaque arête e, on peut alors définir les quantités (ensembles) suivantes :

- $j(e) = \min(l_A, l_B)$ (première variable),
- $k(e) = \max(l_A, l_B)$ (seconde variable),
- $D(e) = \{l_1, \ldots, l_{m-1}\}$ (variables de conditionnement),

et remplacer la notation $\{l_A, l_B, l_1, ..., l_{m-1}\}$ par la notation $j(e), k(e) \mid D(e)$.

Exemple. Reprenons le cas de la dimension $n = 4$. Avec les nouvelles notations, on obtient la représentation graphique de D-*vigne* suivante :



Si chaque arête e est identifiée à la copule $c_{j(e),k(e)|D(e)}$, collecter les copules bidimensionnelles de tous les arbres permet de disposer de :

- $(n-1) = 3$ copules bivariées non-conditionnelles,
- $(n-2) = 2$ copules bivariées avec une seule variable de conditionnement,

- 1 copule bivariée avec $(n-2) = 2$ variables de conditionnement,

ce qui correspond à la forme attendue. Dès lors, on obtiendrait la densité de copule suivante :

$$\begin{aligned} c_{1234}(u_1,u_2,u_3,u_4) = &\underbrace{c_{12}(u_1,u_2) \times c_{23}(u_2,u_3) \times c_{34}(u_3,u_4)}_{T_1} \\ &\cdots \times \underbrace{c_{13|2}(u_{1|2},u_{3|2}) \times c_{24|3}(u_{2|3},u_{2|4})}_{T_2} \\ &\cdots \times \underbrace{c_{14|23}(u_{1|23},u_{3|23})}_{T_3} , \end{aligned}$$

ce qui correspond à la distribution *vigne* (6.7) obtenue dans la section précédente.

L'exemple en dimension $n = 4$ illustre l'efficacité des R-*vignes* quant à la représentation de la factorisation d'une densité de copule. La démonstration rigoureuse de l'équivalence entre la factorisation d'une copule de dimension n et la représentation en R-*vigne* a été effectuée par [Bedford and Cooke, 2001]. En outre, étant donnée une une R-*vigne* $\{T_m = (N_m, E_m)\}_{m=1}^{n-1}$, la densité de copule $c_{1..n}$ de $\tilde{\mathbf{X}}$ est uniquement déterminée et définie pour tout $\mathbf{u} \in [0,1]^n$ par

$$c_{1,..,n}(\mathbf{u}) = \prod_{e \in E_1} c_{j(e),k(e)}(u_{j(e)}, u_{k(e)}) \times \prod_{m=2}^{n-1} \prod_{e \in E_m} c_{j(e),k(e)|D(e)}\left(u_{j(e)|D(e)}, u_{k(e)|D(e)}\right) , \quad (6.8)$$

où les quantités $u_{i|D}$ sont définies à partir des *h-fonctions*. Grâce à ces *h-fonctions*, l'estimation d'une distribution *vigne* peut être effectuée de proche en proche, i.e, en estimant récursivement les densités de copules du premier arbre T_1 jusqu'au dernier T_{n-1}. Dès lors, plusieurs approches d'estimation sont disponibles. [Nagler and Czado, 2016] propose d'estimer les densités de copules bivariées à l'aide d'une procédure d'estimation par noyau. Les auteurs introduisent ainsi un estimateur non-paramétrique de la densité de copule $c_{1,..,n}$ et étudient ses propriétés, à savoir sa consistance, sa vitesse de convergence et sa normalité asymptotique. Le lecteur intéressé peut se référer à [Nagler et al., 2017] pour une bibliographie exhaustive des différentes méthodes d'estimation non-paramétriques. Une autre approche consiste à apprendre chaque densité de copule bivariée en choisissant son meilleur représentant parmi une large bibliothèque de copules paramétriques pour lesquelles les *h-fonctions* sont connues.

6.2.3 Sélection de la R-*vigne*

La précédente section portait sur l'estimation d'une distribution *vigne* associée à une structure de R-*vigne* donnée. Une question non abordée jusque là est celle de la sélection de R-*vigne*. L'approche naïve pour déterminer le "meilleur" modèle de copule *vigne* est de tester toutes les structures possibles, en estimant la distribution *vigne* associée. Adopter cette stratégie est cependant illusoire en pratique puisque le nombre de R-*vigne* de dimension n

$$\frac{n!}{2} \times 2^{\binom{n-2}{2}} ,$$

augmente très vite avec n [Morales Napoles et al., 2010]. Une approche très populaire consiste à effectuer la sélection de proche en proche, conjointement avec l'estimation [Dissmann et al., 2013]. En d'autres termes, il s'agit de choisir la structure de l'arbre T_1 et d'estimer les $(n-1)$ densités de copule bivariées non conditionnelles. Une fois que toute l'inférence a été faite sur l'arbre T_1, on est alors en mesure de choisir l'arbre T_2 puis d'estimer les $(n-2)$ densités de copule bivariées à 1 degré de conditionnement, et ainsi de suite jusqu'au dernier arbre T_{n-1}.

Sélection de l'arbre T_1. [Dissmann et al., 2013] propose de construire l'arbre T_1 de telle sorte que les densités de copules bivariées correspondantes soient associées aux paires de variables présentant les plus fortes dépendances. Pour $i < j \in \{1, \dots, n\}$, la dépendance entre les variables $\tilde{X}_i$ et $\tilde{X}_j$ est mesurée par le τ de Kendall [Kendall, 1948] qui est une statistique définie par la différence entre la probabilité de concordance et de discordance entre $\tilde{X}_i$ et $\tilde{X}_j$, i.e,

$$\tau_{i,j} = \mathbb{P}((\tilde{X}_i - \tilde{X}_i')(\tilde{X}_j - \tilde{X}_j') > 0) - \mathbb{P}((\tilde{X}_i - \tilde{X}_i')(\tilde{X}_j - \tilde{X}_j') < 0) \ , \tag{6.9}$$

où $\tilde{X}_i'$ et $\tilde{X}_j'$ sont des copies indépendantes de $\tilde{X}_i$ et $\tilde{X}_j$ respectivement. Considérons un échantillon $\{\tilde{\mathbf{X}}^k = (\tilde{X}_1^k, .., \tilde{X}_n^k)\}_{k=1}^N$. Deux paires d'observations $(\tilde{X}_i^k, \tilde{X}_j^k)$ et $(\tilde{X}_i^l, \tilde{X}_j^l)$ sont dites concordantes si $\tilde{X}_i^k < \tilde{X}_i^l$ et $\tilde{X}_j^k < \tilde{X}_j^l$ ou si $\tilde{X}_i^k > \tilde{X}_i^l$ et $\tilde{X}_j^k > \tilde{X}_j^l$. À l'inverse, elles sont dites discordantes lorsque $\tilde{X}_i^k < \tilde{X}_i^l$ et $\tilde{X}_j^k > \tilde{X}_j^l$ ou $\tilde{X}_i^k > \tilde{X}_i^l$ et $\tilde{X}_j^k < \tilde{X}_j^l$. Le τ de Kendall $\tau_{i,j}$ peut alors être approximé par

$$\hat{\tau}_{i,j} = \frac{(\text{nombre de paires concordantes}) - (\text{nombre de paires discordantes})}{N(N-1)/2} \ . \tag{6.10}$$

L'arbre T_1 est alors construit en sélectionnant l'arbre qui maximise la somme des valeurs absolues des τ de Kendall des paires impliquées, i.e, en résolvant le problème d'optimisation suivant :

$$\underset{T \in \mathcal{A}}{\text{Argmax}} \sum_{(i,j) \in T} |\hat{\tau}_{i,j}| \ , \tag{6.11}$$

où $\mathcal{A}$ désigne l'ensemble des arbres T couvrants de l'ensemble des sommets $N_1 = \{1, \dots, n\}$. Le problème (6.11) est bien connu en recherche opérationnelle. Il s'agit d'un *problème de l'arbre couvrant de poids maximal* et peut être résolu avec des algorithmes dédiés, comme l'algorithme de Prim ou l'algorithme de Kruskal [Cormen et al., 2009].

Sélection de l'arbre T_{i+1} à partir de T_i. Les observations conditionnelles nécessaires à l'inférence sur l'arbre T_{i+1} ne sont pas directement disponibles mais peuvent être obtenues en utilisant les *h-fonctions*. À noter toutefois que celles-ci peuvent être entachées des erreurs d'approximations commises lors de l'inférence sur l'arbre T_i. Supposons construit l'arbre T_i. Une part de la structure de T_{i+1} est alors fixée par la condition de proximité. Les sous-arbres de T_{i+1} pour lesquels plusieurs possibilités existent sont construits comme précédemment, i.e, en se ramenant à un problème de l'arbre couvrant de poids maximal. Ici, les poids qui sont donnés par les τ de Kendall conditionnels. Ces derniers sont calculés à partir des pseudo-réalisations conditionnelles obtenues grâce aux *h-fonctions* (6.2.1) et (6.2.1).

La méthode de sélection introduite par [Dissmann et al., 2013] a déjà été implémentée dans diverses bibliothèques, comme le package **VineCopula** de **R** pour lequel l'estimation

des copules bivariées est effectuée avec la procédure paramétrique décrite dans la section précédente. L'objet de la prochaine section est d'appliquer cette méthode pour l'estimation des indices de Borgonovo d'ordres supérieurs définis par Eq.(6.1).

6.2.4 Application : estimation des indices de Borgonovo d'ordres supérieurs

Le but de cette section est d'appliquer la théorie des modèles de copules *vignes* pour estimer les indices de Borgonovo d'ordres supérieurs $\delta_{\mathbf{u}}$ où $\mathbf{u} = (i_1, \ldots, i_r) \subset \{1, ..., d\}$. Adopter la définition (6.1) permet en effet de ramener l'estimation de $\delta_{\mathbf{u}}$ à un problème d'apprentissage d'une copule de dimension $|\mathbf{u}| + 1$. La méthode détaillée dans les précédentes sections n'a en revanche d'intérêt que pour les copules de dimension supérieure à 3. De plus, l'estimation des indices de Borgonovo d'ordre 1 à déjà été étudiée dans le chapitre 4. Il est donc supposé dans cette section que le sous-ensemble d'indice $\mathbf{u}$ comporte au moins 2 éléments.

Étape VINE1. Générer $\{\mathbf{X}^n = (X_1^n \ldots, X_d^n)\}_{n=1}^N \overset{\text{i.i.d}}{\sim} f_{\mathbf{X}}$ puis évaluer $Y^n = \mathcal{M}(\mathbf{X}^n)$ pour $n = 1, \ldots, N$.

Étape VINE2. Construire, à l'aide du package **VineCopula** de **R**, une R-*vigne* associée au vecteur aléatoire $(\mathbf{X}_{\mathbf{u}}, Y)$ de dimension $|\mathbf{u}| + 1 = r + 1$. La densité de copule associée $c_{\mathbf{u}}$ peut alors être estimée par la distribution *vigne* $c_{\mathbf{u}}^{\text{vine}}$ définie par

$$c_{\mathbf{u}}^{\text{vine}} = \prod_{e \in E_1} \hat{c}_{j(e),k(e)}(u_{j(e)}, u_{k(e)}) \times \prod_{m=2}^{r} \prod_{e \in E_m} \hat{c}_{j(e),k(e)|D(e)} \left(\hat{u}_{j(e)|D(e)}, \hat{u}_{k(e)|D(e)}\right) \ , \quad (6.12)$$

où $\hat{c}_{j(e),k(e)}$ est un estimateur (paramétrique) de $c_{j(e),k(e)}$ et $\hat{c}_{j(e),k(e)|D(e)}$ un estimateur (paramétrique) de $c_{j(e),k(e)|D(e)}$ obtenu récursivement grâce aux *h-fonctions*.

Étape VINE3. Générer $\{\mathbf{U}^k = (U_1^k, \ldots, U_{r+1}^k)\}_{k=1}^{N'} \overset{\text{i.i.d.}}{\sim} U([0,1]^{r+1})$ puis estimer $\delta_{\mathbf{u}}$ par :

$$\delta_{\mathbf{u}}^{\text{vine}} = \frac{1}{2N'} \sum_{k=1}^{N'} \left| c_{\mathbf{u}}^{\text{vine}}(\mathbf{U}^k) - \frac{f_{\mathbf{X}_{\mathbf{u}}}\left(F_{X_{i_1}}^{-1}(U_1^k), ..., F_{X_{i_r}}^{-1}(U_r^k)\right)}{\prod_{l=1}^{r} f_{X_{i_l}}(F_{X_{i_l}}^{-1}(U_l^k))} \right| . \quad (6.13)$$

Remarque. À l'instar des deux méthodes d'estimation présentées au chapitre 4, le budget de simulation requis par la présente approche est indépendant de la dimension d du modèle. En effet l'échantillon généré à l'étape **VINE1.** avec N appels à $\mathcal{M}$, permet à lui seul de calculer l'estimateur $\delta_{\mathbf{u}}^{\text{vine}}$ pour tout groupe d'indice $\mathbf{u} \subset \{1, .., d\}$ de cardinal supérieur à 2.

6.3 Valeur de Shapley

6.3.1 Définition et propriétés de la valeur de Shapley

En théorie des jeux, la valeur de Shapley [Shapley, 1953] est un concept de solution de répartition des gains dans le cas d'un jeu coopératif. Un jeu coopératif est la donnée d'un ensemble fini de joueurs $\{1, ..., d\}$ et d'une *fonction caractéristique* v, qui à toute *coalition* de joueurs $\mathbf{u} \subset \{1, ..., d\}$, associe $v(\mathbf{u})$, le gain que les joueurs de $\mathbf{u}$ obtiennent en coopérant. Plus formellement, v est une fonction

$$v : \left| \begin{array}{rcl} \mathcal{P}(\{1, \ldots, d\}) & \longrightarrow & \mathbb{R} \\ \mathbf{u} & \longmapsto & v(\mathbf{u}) \end{array} \right. \quad \text{avec} \quad v(\emptyset) = 0 \ , \tag{6.14}$$

où $\mathcal{P}(\{1, \ldots, d\})$ désigne l'ensemble des coalitions possibles. La valeur de Shapley pour le joueur i est alors définie par :

$$\phi_i(v) = \frac{1}{d} \sum_{\mathbf{u} \subset \{1,...,d\} \setminus \{i\}} \binom{d-1}{|\mathbf{u}|}^{-1} (v(\mathbf{u} \cup \{i\}) - v(\mathbf{u})), \ . \tag{6.15}$$

Il faut interpréter la valeur de Shapley en supposant que chaque joueur entre de manière aléatoire dans chaque coalition possible. Le terme $\binom{d-1}{|\mathbf{u}|}^{-1}$ correspond en effet à la probabilité que le joueur rejoigne une coalition particulière $\mathbf{u} \subset \{1, ..., d\} \setminus \{i\}$ de taille $|\mathbf{u}|$. En outre, la quantité $(v(\mathbf{u} \cup \{i\}) - v(\mathbf{u}))$ correspond quant à elle à la contribution marginale du joueur i dans la coalition $\mathbf{u} \cup \{i\}$. Ainsi, la valeur de Shapley ϕ_i s'interprète comme le gain moyen obtenu par le joueur i, en prenant en compte toutes les coopérations possibles.

[Shapley, 1953] énonce une série d'axiomes qui admettent comme unique solution les valeurs de Shapley définies par Eq.(6.15) :

Théorème 6.1. *Soit $(\{1, \ldots, d\}, v)$ un jeu coopératif quelconque. Il existe une unique fonction $\boldsymbol{\phi} = (\phi_1, ..., \phi_d)$ satisfaisant les axiomes suivants :*

1. ***Efficacité.*** *la somme des gains attribués aux joueurs doit être égale à ce que la coalition de tous les joueurs peut obtenir :*

$$\sum_{i=1}^{d} \phi_i(v) = v(\{1, ..., d\}) \ . \tag{6.16}$$

2. ***Symétrie.*** *Si $v(\boldsymbol{u} \cup \{i\}) = v(\boldsymbol{u} \cup \{j\})$ pour toute coalition $\boldsymbol{u} \subset \{1, ..., d\}$, alors $\phi_i(v) = \phi_j(v)$.*

3. ***Linéarité.*** *Pour toutes fonctions caractéristiques v et w,*

$$\boldsymbol{\phi}(v + w) = \boldsymbol{\phi}(v) + \boldsymbol{\phi}(w) \ .$$

De plus, $\boldsymbol{\phi}$ se calcule explicitement et est donnée par la formule définie par Eq.(6.15).

6.3.2 Application en analyse de sensibilité

Dans le contexte de l'analyse de sensibilité, l'ensemble de joueurs $\{1,\dots,d\}$ peut être pensé comme l'ensemble des entrées $X_1,\dots,X_d$ du modèle $Y = \mathcal{M}(\mathbf{X})$ et la fonction caractéristique v comme un mesure d'importance. En outre, ce point de vue a été étudié dans le cadre des méthodes d'analyse de sensibilité basées sur la variance dans plusieurs travaux, [Owen, 2014] dans le contexte d'entrées indépendantes et [Song et al., 2016, Iooss and Prieur, 2017] dans le cas dépendant.

6.3.2.1 Cas des indices de Sobol

[Owen, 2014] propose de considérer la fonction caractéristique

$$\tilde{c}(\mathbf{u}) = \mathbb{V}\mathrm{ar}(\mathbb{E}[Y|\mathbf{X}_\mathbf{u}]) \ . \tag{6.17}$$

En vertu de l'axiome d'efficacité (6.16), ceci entraîne la relation suivante.

$$\sum_{i=1}^{d} \phi_i(\tilde{c}) = \mathbb{V}\mathrm{ar}(Y) \ , \tag{6.18}$$

où $\phi_i(\tilde{c})$ représente la contribution de X_i à la variance de la sortie Y, prenant en compte les interactions ou corrélations possibles avec les autres variables X_j, $j \neq i$. Dans [Song et al., 2016], il est démontré qu'il est équivalent de considérer la fonction caractéristique

$$c(\mathbf{u}) = \mathbb{E}[\mathbb{V}\mathrm{ar}(Y|\mathbf{X}_{-\mathbf{u}})] \ , \tag{6.19}$$

où $-\mathbf{u}$ désigne le complémentaire de $\mathbf{u}$, i.e, que $\phi_i(\tilde{c}) = \phi_i(c)$ pour tout $i \in \{1,\dots,d\}$. À l'instar de la décomposition de Hoeffding-Sobol rencontrée à la section 3.2, Eq.(6.18) procure une décomposition de la variance de sortie $\mathbb{V}\mathrm{ar}(Y)$ et présente en plus l'avantage d'être toujours valide dans le cas où les entrées X_i sont corrélées. Lorsque les entrées sont indépendantes, la relation suivante est vérifiée [Owen, 2014]

$$S_i \leq \frac{\phi_i(c)}{\mathbb{V}\mathrm{ar}(Y)} \leq S_{T_i} \ . \tag{6.20}$$

La valeur de Shapley $\phi_i(c)/\mathbb{V}\mathrm{ar}(Y)$ peut donc être vue comme une "valeur moyenne" des indices de Sobol du premier ordre S_i et totaux S_{T_i}. Dans le cas où les entrées présentent une structure de dépendance, l'interprétation des valeurs de Shapley vis à vis des indices de Sobol est moins aisée puisque l'encadrement (6.20) n'est plus vérifié [Iooss and Prieur, 2017]. Néanmoins, une valeur de $\phi_i(c)$ proche de 0 permet toujours de conclure quant à la faible contribution de X_i à la variance de sortie.

6.3.2.2 Cas des indices de Borgonovo

Similairement à l'approche de [Owen, 2014, Song et al., 2016, Iooss and Prieur, 2017], une idée naturelle est considérer la fonction suivante

$$v(\mathbf{u}) = \delta_\mathbf{u} \ , \tag{6.21}$$

qui associe à toute coalition $\mathbf{u}$ l'indice de Borgonovo de Y relativement au groupe de paramètre $\mathbf{X}_\mathbf{u}$. Par ailleurs, adopter la définition générale $\delta_\mathbf{u} = \mathbb{E}[d_{\mathrm{TV}}(Y, Y|\mathbf{X}_\mathbf{u})]$ introduite à la section 3.3.3 permet de donner un sens à la valeur $v(\emptyset)$:

$$v(\emptyset) = \delta_\emptyset = 0 \ .$$

Par conséquent, l'application v (6.21) est une fonction caractéristique et il est donc possible d'après le théorème 6.1 de définir les valeurs de Shapley suivantes relatives aux indices de Borgonovo

$$\mathrm{Sh}_i := \phi_i(v) = \frac{1}{d} \sum_{\mathbf{u} \subset \{1,\dots,d\} \setminus \{i\}} \binom{d-1}{|\mathbf{u}|}^{-1} (\delta_{\mathbf{u} \cup \{i\}} - \delta_\mathbf{u}) \ . \tag{6.22}$$

De plus, l'axiome d'efficacité (6.16) implique la décomposition suivante :

$$\sum_{i=1}^{d} \mathrm{Sh}_i = 1 \ , \tag{6.23}$$

où Sh_i représente l'influence (au sens de Borgonovo) de l'entrée X_i prenant en compte tous les effets d'interaction avec les autres variables.

Remarque. Les valeurs de Shapley Sh_i permettent de simplifier l'interprétation des indices de Borgonovo. D'une part, l'information apportée par les $2^d - 1$ indices de Borgonovo est résumée en seulement d mesures d'importance. D'autre part, la décomposition (6.23) facilite la comparaison des indices Sh_i. Enfin, il convient de noter que le calcul de ces indices peut s'effectuer par simple post-traitement, à savoir, après l'estimation de l'ensemble des indices de Borgonovo au moyen de la méthode détaillée à la section 6.2.4.

6.4 Applications numériques

Dans cette section, deux exemples numériques sont considérés pour illustrer l'application de l'approche discutée dans ce chapitre. En outre, les deux applications numériques mises en œuvre se déroulent selon la procédure suivante :

1. Calcul de M estimations $\{\hat{\delta}^k_\mathbf{u}\}_{k=1}^M$ indépendantes des indices de Borgonovo. Les indices d'ordres 1 sont obtenus avec l'estimateur δ_i^{ME} basé sur le principe du maximum d'entropie. Les indices d'ordres supérieurs sont calculés à l'aide du schéma d'estimation décrit dans la section 6.2.4 basé sur la théorie des copules *vignes*.

2. Calcul par post-traitement de M estimations indépendantes $\{\widehat{\mathrm{Sh}}^k_i\}_{k=1}^M$ des valeurs de Shapley, où
$$\widehat{\mathrm{Sh}}^k_i = \frac{1}{d} \sum_{\mathbf{u} \subset \{1,\dots,d\} \setminus \{i\}} \binom{d-1}{|\mathbf{u}|}^{-1} (\hat{\delta}^k_{\mathbf{u} \cup \{i\}} - \hat{\delta}^k_\mathbf{u}) \ .$$

3. Calcul des estimateurs non biaisés respectifs de la moyenne et de l'écart-type :
$$\bar{\mathrm{Sh}}_i := \frac{1}{M} \sum_{k=1}^{M} \widehat{\mathrm{Sh}}^k_i \quad \text{et} \quad \bar{\sigma}_{\widehat{\mathrm{Sh}}_i} := \sqrt{\frac{1}{M-1} \sum_{k=1}^{M} (\widehat{\mathrm{Sh}}^k_i - \bar{\mathrm{Sh}}_i)^2},$$
et du coefficient de variation correspondant.

4. Calcul de l'écart relatif moyen

$$\mathrm{RD} = \frac{\bar{\mathrm{Sh}}_i - \mathrm{Sh}_i}{\mathrm{Sh}_i} \ ,$$

lorsque la valeur théorique de Sh_i est connue, i.e, lorsque les indices de Borgonovo sont connus à tous les ordres.

6.4.1 Le modèle gaussien (4.2)

Considérons le modèle linéaire gaussien (4.2) défini par

$$Y = \langle A, \mathbf{X} \rangle \ , \ \mathbf{X} \sim N(0_{\mathbb{R}^4}, \Sigma) \ ,$$

où A est le vecteur colonne $[1,7 \ 1,8 \ 1,9 \ 2]$ et $\Sigma \in M_4(\mathbb{R})$. Pour tout sous-ensemble $\mathbf{u} = \{i_1, .., i_r\} \subset \{1, \dots, 4\}$, des calculs élémentaires sur les vecteurs gaussiens montrent que la variable $(\mathbf{X_u}, Y)$ est gaussienne, centrée et de matrice de covariance

$$\begin{pmatrix} \Lambda_{\mathbf{X_u}} & \Lambda_{\mathbf{X_u},Y} \\ \Lambda_{\mathbf{X_u},Y} & \mathbb{V}\mathrm{ar}(Y) \end{pmatrix} \in M_{|\mathbf{u}|+1}(\mathbb{R}) \ ,$$

où $\Lambda_{\mathbf{X_u}} = (\Sigma_{ij})_{i,j\in\mathbf{u}}$ est l'opérateur de covariance de $\mathbf{X_u}$ et $\Lambda_{\mathbf{X_u},Y}$ le vecteur colonne $(\langle A, \Sigma_i \rangle)_{i\in\mathbf{u}}$ où Σ_i désigne la $i^{\text{ème}}$ colonne de Σ. En particulier, la densité de copule $c_\mathbf{u}$ de $(\mathbf{X_u}, Y)$ est connue ainsi que l'indice de Borgonovo $\delta_\mathbf{u}$ correspondant. En outre, l'estimateur Monte-Carlo [1]

$$\delta_\mathbf{u} := \frac{1}{2N'} \sum_{k=1}^{N'} \left| c_\mathbf{u}(\mathbf{U}^k) - \frac{f_{\mathbf{X_u}}\left(F^{-1}_{X_{i_1}}(U_1^k), ..., F^{-1}_{X_{i_r}}(U_r^k)\right)}{\prod_{l=1}^r f_{X_{i_l}}(F^{-1}_{X_{i_l}}(U_l^k))} \right| \ , \tag{6.24}$$

est considéré dans la suite comme valeur de référence pour la mesure $\delta_\mathbf{u}$. La procédure décrite dans la section 6.2.4 est appliquée $M = 100$ fois sur ce modèle, et ce, avec des échantillons de taille $N = 5000$. La table 6.1 rassemble les résultats ainsi obtenus et les compare avec les valeurs de Shapley de référence (6.24). Ces résultats illustrent la parfaite estimation des différentes densités de copule $c_\mathbf{u}$ par les distributions *vigne* $c_\mathbf{u}^{\text{vine}}$. En effet, pour tout $i \in \{1, \dots, 4\}$, l'écart relatif moyen entre $\widehat{\mathrm{Sh}}_i$ et Sh_i est inférieur à $0,5\%$, où, rappelons le, Sh_i désigne la valeur de Shapley de référence pour laquelle les densités de copules sont supposées connues. Cet exemple montre donc que l'application des modèles de R-*vigne* permet d'estimer efficacement les indices de Borgonovo d'ordre supérieur. Il convient cependant de noter qu'il s'agit là d'un exemple idéal d'application pour lequel les structures de dépendance des variables $(\mathbf{X_u}, Y)$ sont toutes gaussiennes et donc parfaitement adaptées à une estimation paramétrique. Des exemples présentant des structures de dépendance plus complexes sont étudiés dans les prochaines sections.

[1] $\{\mathbf{U}^k\}_{k=1}^{N'} \overset{\text{i.i.d.}}{\sim} U([0,1]^{|\mathbf{u}|+1})$ est l'échantillon Monte-Carlo intervenant à l'étape **VINE3.** .

Table 6.1. Estimations des indices de Shapley Sh_i du modèle gaussien (4.2).

Entrée	Sh_i		$\widehat{\mathrm{Sh}_i}$		
	Moyenne	CV	Moyenne	CV	RD
X_1	**0.2213**	0.0057	0.2227	0.0090	0.0062
X_2	**0.2534**	0.0041	0.2526	0.0065	-0.0031
X_3	**0.2673**	0.0054	0.2660	0.0070	-0.0051
X_4	**0.2579**	0.0046	0.2587	0.0072	0.0030

6.4.2 L'oscillateur non linéaire à un degré de liberté

Dans cette section, le schéma d'estimation proposé est appliqué $M = 100$ fois au modèle d'oscillateur non linéaire (5.13), et ce, avec des échantillons de taille $N = 5000$. La table 6.3 rassemble les résultats d'estimations des indices de Borgonovo du premier ordre δ_i et des valeurs de Shapley Sh_i. La table 6.2 donne quant à elle les estimations obtenues avec une méthode *double-boucle* et considérées dans la suite comme valeurs de référence. Pour les entrées X_1, X_3, X_5 et X_6, les résultats obtenus témoignent d'une bonne estimation des indices de Shapley avec des écarts relatif moyens variant entre 0.5% et 15%. En revanche, au vu des résultats obtenus pour X_2 et X_4, il apparait que les indices de Shapley ont tendance a être surévalués lorsque ceux-ci sont faibles.

Table 6.2. Valeurs de Shapley Sh_i du modèle d'oscillateur. Les indices de Borgonovo correspondants sont estimés à l'aide de la méthode du *double-boucle* et avec des échantillons de taille 100. Les résultats affichés sont obtenus après 100 itérations de ce procédé, chaque itération requérant 630,000 évaluations du modèle.

Entrée :	X_1	X_2	X_3	X_4	X_5	X_6
Sh_i (rang)	0.1129 (3)	0.0174 (6)	0.5261 (1)	0.0379 (5)	0.0929 (4)	0.2128 (2)
CV	0.1076	0.1324	0.0549	0.1381	0.1120	0.0943

Table 6.3. Estimations des indices de Shapley Sh_i du modèle d'oscillateur.

Entrée	$\hat{\delta}_i^{\mathrm{ME}}$		$\widehat{\mathrm{Sh}_i}$		
	Moyenne (rang)	CV	Moyenne (rang)	CV	RD
X_1	0.0384 (4)	0.0770	0.1161 (3)	0.0168	0.0286
X_2	0.0100 (5)	0.3299	0.0734 (5)	0.0208	3.2115
X_3	0.1717 (1)	0.0277	0.4478 (1)	0.0073	-0.1488
X_4	0.0083 (6)	0.4160	0.0728 (6)	0.0195	0.9197
X_5	0.0394 (3)	0.0992	0.0924 (4)	0.0241	-0.0050
X_6	0.0985 (2)	0.0380	0.1975 (2)	0.0190	-0.0719

6.4.3 Le modèle de retombée d'un étage de lanceur

Dans cette section, le schéma d'estimation proposé est appliqué $M = 100$ fois au modèle de lanceur présenté dans la section 4.4.1, et ce, avec des échantillons de taille $N = 5000$. La table 6.5 rassemble les résultats d'estimations des indices de Borgonovo du premier ordre δ_i et des valeurs de Shapley Sh_i. La table 6.4 rappelle quant à elle les estimations des indices de Sobol d'ordre 1 S_i et totaux S_{T_i} obtenus à la section 4.4.2. Enfin, la table 6.6 donne les résultats obtenus avec une méthode *double-boucle* et assimilés dans la suite comme valeurs de référence.

Table 6.4. Estimations des indices de Sobol d'ordre 1 et totaux du modèle de lanceur.

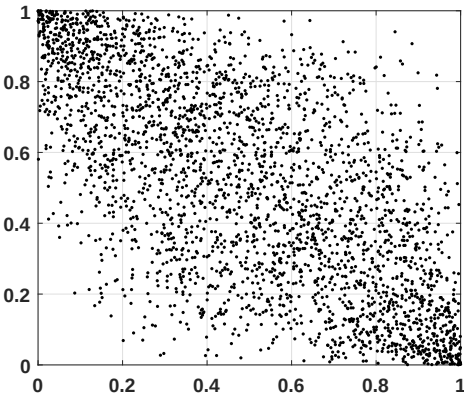
Entrée	S_i		S_{T_i}	
	Moyenne (rang)	CV	Moyenne (rang)	CV
X_1	0.0046 (5)	16.8374	0.0153 (6)	0.1154
X_2	0.2576 (1)	0.2436	0.6687 (1)	0.0888
X_3	0.0682 (3)	1.0210	0.3892 (2)	0.1049
X_4	0.1924 (2)	0.3643	0.2126 (4)	0.1428
X_5	0.0043 (6)	17.8161	0.0169 (5)	0.1055
X_6	0.0122 (4)	6.5117	0.2229 (3)	0.1003

Table 6.5. Estimations des indices de Shapley Sh_i du modèle de lanceur.

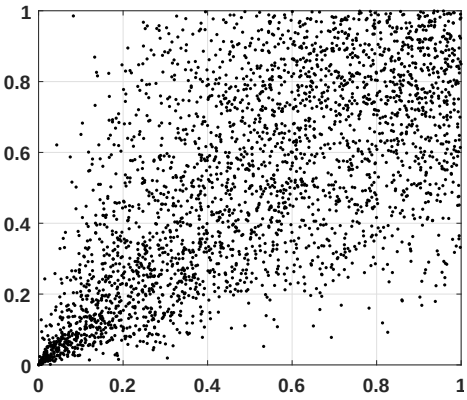
Entrée	$\hat{\delta}_i^{\mathrm{ME}}$		$\widehat{\mathrm{Sh}}_i$	
	Moyenne (rang)	CV	Moyenne (rang)	CV
X_1	0.0154 (6)	0.2795	0.1248 (5)	0.0142
X_2	0.1553 (2)	0.0346	0.2258 (1)	0.0181
X_3	0.0686 (3)	0.0838	0.1720 (3)	0.0185
X_4	0.1821 (1)	0.0286	0.2088 (2)	0.0127
X_5	0.0153 (5)	0.2629	0.1247 (6)	0.0120
X_6	0.0403 (4)	0.1389	0.1438 (4)	0.0175

Table 6.6. Valeurs de Shapley Sh_i du modèle de lanceur. Les indices de Borgonovo correspondants sont estimés à l'aide de la méthode du *double-boucle* et avec des échantillons de taille 100. Les résultats affichés sont obtenus après 20 itérations de ce procédé, chaque itération requérant 630,000 évaluations du modèle.

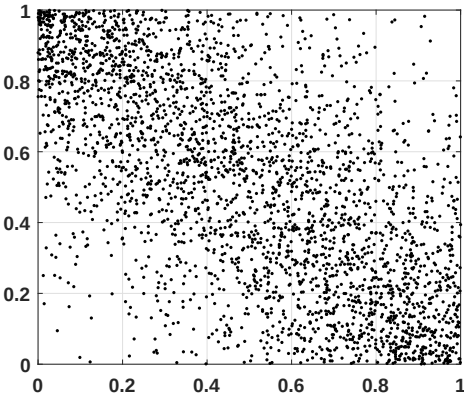
Entrée :	X_1	X_2	X_3	X_4	X_5	X_6
Sh_i (rang)	0.0448 (5)	0.2789 (1)	0.2014 (3)	0.2786 (2)	0.0444 (6)	0.1489 (4)
CV	0.1333	0.0849	0.0669	0.1170	0.1039	0.0745



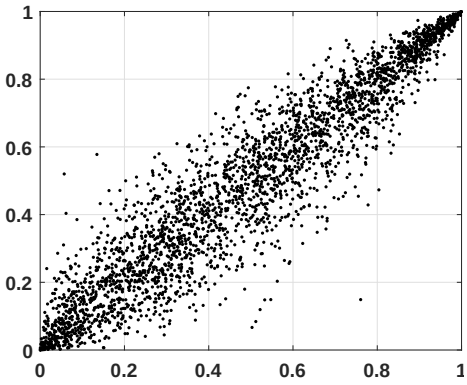
(a) Échantillon de taille $3,000$ généré selon une copule gaussienne de paramètre $\rho = -0.7$.



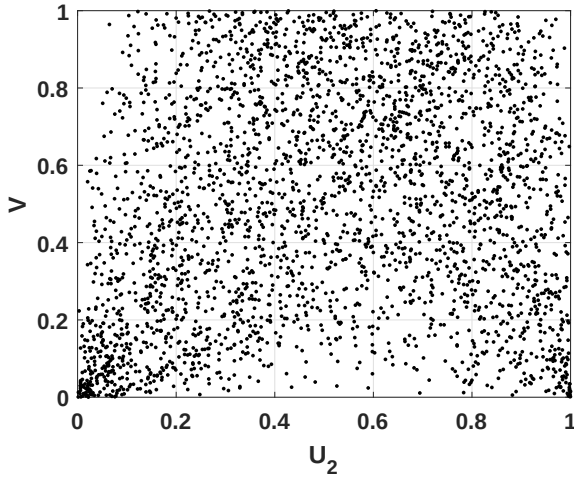
(b) Échantillon de taille $3,000$ généré selon une copule de Clayton de paramètre $\alpha = 2$.



(c) Échantillon de taille $3,000$ généré selon une copule de Frank de paramètre $\alpha = -5$.



(d) Échantillon de taille $3,000$ généré selon une copule de Gumbel de paramètre $\alpha = 5$.



(e) Échantillon de taille $3,000$ généré selon la copule de (X_2, Y).

FIGURE 6.1. COMPARAISON DE LA COPULE DU COUPLE (X_2, Y) DU MODÈLE DE LANCEUR AVEC PLUSIEURS COPULES PARAMÉTRIQUES CLASSIQUES.

Deux points peuvent être soulignés. Premièrement, les indices de Sobol totaux S_{T_i} conduisent à un classement analogue à celui des valeurs de Shapley Sh_i (cf. table 6.6). Il en ressort en effet l'influence non négligeable des variables X_3 et X_6, ce qui n'a pas pu être détecté avec la seule connaissance des indices d'ordre un. D'autre part, il peut être constaté que les estimations des valeurs de Shapley (cf. table 6.5) respectent le ranking correct mais sont très imprécises. À titre d'exemple, l'indice de Borgonovo du 5-uplet $(X_2, \ldots, X_6)$ a pour valeur de référence $\bar{\delta}^{\mathrm{PDF}}_{2,..,6} \approx 0.8$ et est estimé par $\bar{\delta}^{\mathrm{Vine}}_{2,..,6} \approx 0.2$. Nous pensons que ce défaut de précision est dû aux structures de dépendance très complexes inhérentes au modèle de lanceur et pour lesquelles il n'y a pas de famille paramétrique suffisamment représentative. La figure [6.1] vient corroborer ce postulat et montre en effet une différence notable entre la copule de (X_2, Y) et des copules paramétriques classiques. Au regard de la répartition des échantillons générés selon c_2 (cf. figure [6.1(e)]), une idée serait de recourir à une méthode GMCM (pour *Gaussian Mixture Copula Models*) [Tewari et al., 2011]. Une autre alternative serait d'estimer les copules bivariées par une approche non-paramétrique.Pour étayer la pertinence de cette perspective, revenons sur le cas de l'indice $\delta_{2,..,6}$. En tronquant la distribution *vigne* $c^{\mathrm{Vine}}_{2,..,6}$ au premier niveau d'arbre, i.e, en ne considérant que les densités de copules bivariées non conditionnelles :

$$c^{\mathrm{Vine}}_{2,..,6} \approx \prod_{e \in E_1} c_{j(e),k(e)}(u_{j(e)}, u_{k(e)}) \ , \tag{6.25}$$

nous obtenons une approximation $\delta^{\mathrm{Vine}}_{2,..,6} \approx 0.5$ en estimant les densités de copule intervenant dans Eq.(6.25) par les estimateurs du maximum d'entropie utilisés pour calculer les indices d'ordre un.

6.5 Conclusion

L'objectif de ce chapitre était double :

♦ Estimer les indices de Borgonovo d'ordres supérieurs. Pour surmonter la malédiction de la dimension nous avons proposé d'utiliser de la théorie des copules *vignes*. Cette démarche permet d'estimer l'ensemble des mesures d'importance de Borgonovo à partir d'un unique jeu de données $\{\mathbf{X}^k, Y\}$ de l'entrée et des valeurs de sortie correspondantes. Le schéma d'estimation proposé montre de très bonnes performance sur un cas test analytique. En revanche, sa précision peut être fortement affecté lorsque des structures de dépendance complexes interviennent, comme il a pu être constaté sur modèle de lanceur.

♦ Synthétiser l'information fournie par l'ensemble des 2^d-1 indices de Borgonovo. Pour simplifier l'interprétation de ces indices, nous proposons d'utiliser la notion de valeur de Shapley, et nous montrons que cela permet de définir d indices de sensibilité qui quantifient l'influence des entrées X_i en tenant compte de tous les effets d'interaction avec les autres variables. Nous montrons également que ces indices peuvent être calculés par simple post-traitement, i.e, sans appels supplémentaires au modèle $\mathcal{M}(\cdot)$.

Chapitre 7

Conclusion et perspectives

Résumé des principales contributions

L'analyse de sensibilité globale est une étape clef de la quantification des incertitudes des modèles boîte noire entrée-sortie. Elle a pour but d'identifier et de classifier les différentes entrées en fonction de leur influence sur la sortie du modèle. Ces travaux de thèse contribuent au développement de méthodes numériques permettant de caractériser l'influence de ces entrées, et ce, à l'aide des mesures d'importance de Borgonovo. Les indices de Borgonovo prennent en compte l'ensemble de la distribution de sortie et constituent ainsi une alternative pertinente aux indices de Sobol qui se focalisent uniquement sur la variance de la sortie.

Estimation des indices de Borgonovo d'ordre un

La première contribution de cette thèse réside dans l'exploration de nouvelles méthodes d'estimation des indices δ_i de Borgonovo d'ordre un. Initialement, ces indices étaient très peu utilisés car les premières méthodes d'estimation introduites, qualifiées de *double-boucle*, exigeaient un nombre non raisonnable d'appels à la fonction boîte noire. Leur utilisation dans des cas d'application concrets a débuté avec l'arrivée de nouvelles approches d'estimation basées sur la réinterprétation de δ_i en terme de mesure de dépendance, notamment la méthode *simple-boucle* et la méthode basée sur la transformation de Nataf. Celles-ci ont permis de réduire fortement le budget de simulation des méthodes *double-boucle* mais souffrent cependant de problèmes théoriques qui peuvent affecter leur champ d'application ou la précision de leur estimations.

Deux nouvelles méthodes d'estimation sont proposées (Chapitre 4). La première est une amélioration de la méthode *simple-boucle* et est basée sur une procédure d'échantillonnage préférentiel combinée à la technique d'estimation par noyau Gaussien. Une comparaison entre cette méthode et la méthode *simple-boucle* a été effectuée sur plusieurs exemples numériques de complexité croissante. Cette contribution a donné lieu à la publication suivante :

Derennes, P., Morio, J., and Simatos, F. (2018) *A nonparametric importance sampling estimator for moment independent importance measures*. Accepted in Reliability Engineering and System Safety, https://doi.org/10.1016/j.ress.2018.02.009

La seconde méthode introduite se base sur la représentation de δ_i en terme de densité de copule et sur l'estimation de cette copule au moyen du principe du maximum d'entropie. Cette contribution est liée à l'article de conférence suivant :

Derennes, P., Morio, J., and Simatos, F. (9-12 December 2018) *Estimation of the moment independent importance sampling measures using copula and maximum entropy framework*, Winter Simulation Conference, Gothenburg, Sweden, 1623-1634.

Les deux méthodologies ainsi proposées sont très permissives puisqu'elles peuvent être appliquées sans hypothèses préalables sur le modèle étudié. En particulier, la boîte noire $\mathcal{M}$ peut être hautement non linéaire et les entrées du modèles peuvent être non indépendantes. En outre, elles ont été appliquées sur un modèle concret portant sur l'estimation de la zone de retombée d'un étage de lanceur et provenant du chapitre de livre suivant :

P. Derennes, V. Chabridon, J. Morio, M. Balesdent, F. Simatos, J-M. Bourinet and M. Gayton (2019) *Nonparametric importance sampling techniques for sensitivity analysis and reliability assessment of a launcher stage fallout*, Optimization in Space Engineering, G. Fasano and J. Pinter, Springer.

Pour finir, le budget de simulation inhérent à leur utilisation est faible puisque l'ensemble des indices d'ordre un peuvent être estimés à partir d'un jeu de donnée commun, et ce, indépendamment de la dimension du modèle.

Application en analyse de sensibilité fiabiliste

La seconde contribution de cette thèse s'inscrit dans le contexte de l'analyse de sensibilité fiabiliste. Là où l'analyse de sensibilité classique se focalise sur la sortie du modèle, l'approche fiabiliste s'intéresse au système dans un mode de fonctionnement critique, typiquement une défaillance associée à un état non sécurisé et donc indésirable. Deux méthodes complémentaires d'analyse de sensibilité fiabilistes peuvent être distinguées : celles visant à étudier l'impact d'une variable d'entrée sur une fonction de la sortie, telle que la fonction indicatrice du domaine critique, et celles visant à étudier l'influence d'une entrée conditionnellement à l'évènement de défaillance. Nous avons défini deux indices de sensibilité fiabilistes liés à ces méthodes et intrinsèquement liés aux indices de Borgonovo traditionnels (Chapitre 5). Nous avons montré que ces deux mesures peuvent être simultanément estimées à partir d'un échantillon commun donné par exemple par un algorithme de Monte-Carlo séquentiel, un des plus utilisés dans le cadre de l'analyse de fiabilité. Par ailleurs, plusieurs applications numériques ont été conduites et ont permis d'illustrer l'importance de l'approche fiabiliste en analyse de sensibilité. En effet, une entrée du modèle peut apparaitre comme négligeable de manière globale et s'avérer fortement influente lorsque l'on se focalise sur le régime de défaillance du système, et inversement. Il est intéressant de pouvoir détecter un tel phénomène puisque les conséquences peuvent être dramatiques en terme de sécurité et d'impact environnemental. Cette contribution a donné lieu à l'article suivant, actuellement en révision après demandes de modifications :

P. Derennes, J. Morio and F. Simatos (2019) *Simultaneous estimation of complementary moment independent sensitivity measures for reliability analysis*, (under revision).

Étude des indices d'ordre supérieur

La dernière contribution de la thèse concerne l'estimation des indices de Borgonovo d'ordre supérieur, i.e, affiliés à un groupe de paramètres d'entrée, ce qui rentre dans le cadre de l'estimation de densité multidimensionnelle. Cette estimation peut théoriquement être réalisée avec les méthodologies introduites au chapitre 4. Toutefois, celles-ci souffrent de problèmes numériques lorsque la dimension du problème augmente. Pour surmonter cette malédiction de la dimension, nous avons proposé d'utiliser la théorie des modèles R-*vigne* (Chapitre 6). Cette approche fournit de très bonnes estimations sur un cas analytique. En revanche, de grandes imprécisions sont obtenues sur le modèle de lanceur. Par conséquent, la question de l'estimation des indices d'ordre supérieur reste ouverte. Néanmoins, une perspective d'amélioration est proposée dans la prochaine section.

Enfin, le dernier objectif de la thèse à été de clarifier l'interprétation des indices de Borgonovo. En effet, leur nombre grandit rapidement avec la dimension du modèle, ce qui rend laborieux le traitement de l'information fournie par chacun des indices. Pour synthétiser cette information, nous avons proposer d'utiliser la notion de *valeur de Shapley* initialement introduite en théorie des jeux. Cela a permit de définir des indices de sensibilité qui quantifiant l'influence (au sens de Borgonovo) de chacune des entrées avec prise en compte de tous les effets d'interaction. De plus, ces indices peuvent être obtenus par simple post-traitement une fois l'ensemble des indices de Borgonovo estimés.

Perspectives

Plusieurs investigations peuvent encore être menées pour améliorer l'estimation des indices de Borgonovo d'ordre un au moyen du principe du maximum d'entropie. Tout d'abord, il serait intéressant de faire l'étude théorique des performances de l'estimateur $\hat{\delta}_i^{\mathrm{ME}}$ comme la démonstration de sa convergence. Ceci repose sur la convergence de l'estimateur du maximum d'entropie en dimension deux, ce qui constitue un problème ouvert. Des considérations d'ordre numériques peuvent également être abordées. En effet, la procédure de maximisation de l'entropie utilisée dans ce manuscrit repose sur le choix de paramètres α et m, et nous avons pu observer sur l'exemple du modèle de lanceur qu'un mauvais ajustement de ces paramètres peut affecter la précision des estimations. Une perspective intéressante serait donc de développer une stratégie permettant d'optimiser efficacement les valeurs de ces paramètres. Ce problème a été abordé dans plusieurs travaux, comme [Tagliani, 2011, Novi Inverardi and Tagliani, 2003].

Une seconde piste de réflexion est de recourir à un modèle de substitution pour alléger la charge de calcul inhérente à l'estimation des indices de sensibilité fiabilistes considérés dans ces travaux de thèse. Il convient cependant de s'assurer que la réduction du nombre d'appels à la boîte noire se soit pas entachée par des ajouts excessifs d'erreurs induites par l'utilisation du métamodèle.

Une autre perspective concerne le schéma d'estimation des indices de Borgonovo d'ordre supérieur fondé sur la construction d'un modèle de R-*vigne*. Une piste pour réduire l'imprécision constatée sur le modèle de lanceur serait d'estimer les densités de copules bivariées de manière non-paramétrique. Pour cela, une idée serait de combiner la procédure

de sélection de la R-*vigne* fournie par le package **RVineCopula** de **R** avec la méthode d'estimation non-paramétrique développée dans [Nagler and Czado, 2016].

Une dernière perspective serait d'explorer la possibilité d'étendre la méthode d'analyse de sensibilité globale de Borgonovo dans le cas du double aléa, i.e, lorsque l'entrée $\mathbf{X} \sim f_{\mathbf{X}}(\cdot, \mathbf{\Theta})$ admet une loi de distribution définie par un paramètre aléatoire $\mathbf{\Theta}$. Ce double niveau d'aléa est souvent utilisé pour distinguer les incertitudes phénoménologiques des incertitudes liées à une méconnaissance du modèle. En outre, effectuer cette étude permettrait d'étudier l'impact du choix du modèle de probabilité de l'entrée sur la sortie du modèle.

Bibliographie

[Aas et al., 2009] Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. Insurance : Mathematics and economics, 44(2) :182–198.

[AghaKouchak, 2014] AghaKouchak, A. (2014). Entropy–copula in hydrology and climatology. Journal of Hydrometeorology, 15(6) :2176–2189.

[Au and Beck, 2001] Au, S.-K. and Beck, J. L. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. Probabilistic engineering mechanics, 16(4) :263–277.

[Augustin et al., 2014] Augustin, T., Coolen, F. P., de Cooman, G., and Troffaes, M. C. (2014). Introduction to imprecise probabilities. John Wiley & Sons.

[Bedford and Cooke, 2001] Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. Annals of Mathematics and Artificial intelligence, 32(1-4) :245–268.

[Bedford and Cooke, 2002] Bedford, T. and Cooke, R. M. (2002). Vines : A new graphical model for dependent random variables. The Annals of Statistics, pages 1031–1068.

[Benaïm and El Karoui, 2005] Benaïm, M. and El Karoui, N. (2005). Promenade aléatoire : chaînes de Markov et simulations : martingales et stratégies. Editions Ecole Polytechnique.

[Bjerager, 1988] Bjerager, P. (1988). Probability integration by directional simulation. Journal of Engineering Mechanics, 114(8) :1285–1302.

[Bjerager, 1991] Bjerager, P. (1991). Methods for structural reliability computations. In Reliability problems : general principles and applications in mechanics of solids and structures, pages 89–135. Springer.

[Bolancé et al., 2008] Bolancé, C., Guillén, M., and Nielsen, J. P. (2008). Inverse beta transformation in kernel density estimation. Statistics & Probability Letters, 78(13) :1757–1764.

[Borgonovo, 2007] Borgonovo, E. (2007). A new uncertainty importance measure. Reliability Engineering & System Safety, 92(6) :771–784.

[Borgonovo et al., 2011] Borgonovo, E., Castaings, W., and Tarantola, S. (2011). Moment independent importance measures : New results and analytical test cases. Risk Analysis, 31(3) :404–428.

[Borgonovo and Plischke, 2016] Borgonovo, E. and Plischke, E. (2016). Sensitivity analysis : a review of recent advances. European Journal of Operational Research, 248(3) :869–887.

[Botev et al., 2010] Botev, Z. I., Grotowski, J. F., Kroese, D. P., et al. (2010). Kernel density estimation via diffusion. The Annals of Statistics, 38(5) :2916–2957.

[Bourinet, 2016] Bourinet, J.-M. (2016). Rare-event probability estimation with adaptive support vector regression surrogates. Reliability Engineering & System Safety, 150 :210–221.

[Bourinet, 2018] Bourinet, J.-M. (2018). Reliability analysis and optimal design under uncertainty - Fo Habilitation à diriger des recherches, Université Clermont Auvergne.

[Bourinet et al., 2011] Bourinet, J.-M., Deheeger, F., and Lemaire, M. (2011). Assessing small failure probabilities by combined subset simulation and support vector machines. Structural Safety, 33(6) :343–353.

[Boyd and Vandenberghe, 2004] Boyd, S. and Vandenberghe, L. (2004). Convex optimization. Cambridge University Press, Cambridge.

[Breitung, 1984] Breitung, K. (1984). Asymptotic approximations for multinormal integrals. Journal of Engineering Mechanics, 110(3) :357–366.

[Browne et al., 2017] Browne, T., Fort, J.-C., Iooss, B., and Le Gratiet, L. (2017). Estimate of quantile-oriented sensitivity indices. https://hal.archives-ouvertes.fr/hal-01450891.

[Buch-Larsen et al., 2005] Buch-Larsen, T., Nielsen, J. P., Guillén, M., and Bolancé, C. (2005). Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. Statistics, 39(6) :503–516.

[Bucher et al., 1989] Bucher, C., Chen, Y., and Schuëller, G. (1989). Time variant reliability analysis utilizing response surface approach. In Reliability and Optimization of Structural Systems 88, pages 1–14. Springer.

[Cérou et al., 2012] Cérou, F., Del Moral, P., Furon, T., and Guyader, A. (2012). Sequential Monte Carlo for rare event estimation. Statistics and computing, 22(3) :795–808.

[Cérou et al., 2006] Cérou, F., Del Moral, P., Le Gland, F., and Lezaud, P. (2006). Genetic genealogical models in rare event analysis. PhD thesis, INRIA.

[Cérou and Guyader, 2007] Cérou, F. and Guyader, A. (2007). Adaptive multilevel splitting for rare event analysis. Stochastic Analysis and Applications, 25(2) :417–443.

[Chabridon, 2018] Chabridon, V. (2018). Analyse de sensibilité fiabiliste avec prise en compte d'incertitudes sur le modèle probabiliste–Application aux systèmes aérospatiaux. PhD thesis, Université Clermont Auvergne.

[Chastaing et al., 2012] Chastaing, G., Gamboa, F., Prieur, C., et al. (2012). Generalized hoeffding-sobol decomposition for dependent variables-application to sensitivity analysis. Electronic Journal of Statistics, 6 :2420–2448.

[Chib and Greenberg, 1995] Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. The american statistician, 49(4) :327–335.

[Cormen et al., 2009] Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2009). Introduction to algorithms. MIT press.

[Cui et al., 2010] Cui, L., Lü, Z., and Zhao, X. (2010). Moment-independent importance measure of basic random variable and its probability density evolution solution. Science China Technological Sciences, 53(4) :1138–1145.

[Czado, 2010] Czado, C. (2010). Pair-copula constructions of multivariate copulas. In Copula theory and its applications, pages 93–109. Springer.

[Da Veiga, 2015] Da Veiga, S. (2015). Global sensitivity analysis with dependence measures. Journal of Statistical Computation and Simulation, 85(7) :1283–1305.

[Del Moral, 2004] Del Moral, P. (2004). Feynman-Kac formulae. Springer.

[Del Moral and Lezaud, 2006] Del Moral, P. and Lezaud, P. (2006). Branching and interacting particle interpretations of rare event probabilities. Springer.

[Der Kiureghian and Dakessian, 1998] Der Kiureghian, A. and Dakessian, T. (1998). Multiple design points in first and second-order reliability. Structural Safety, 20(1) :37–49.

[Derennes et al., 2019a] Derennes, P., Chabridon, V., Morio, J., Balesdent, M., Simatos, F., Bourinet, J.-M., and Gayton, M. (2019a). Nonparametric importance sampling techniques for sensitivity analysis and reliability assessment of a launcher stage fallout, Optimization in Space Engineering, G. Fasano and J. Pinte. Springer.

[Derennes et al., 2018a] Derennes, P., Morio, J., and Simatos, F. (2018a). Estimation of the moment independent importance sampling measures using copula and maximum entropy framework. Winter Simulation Conference, pages 1623–1634.

[Derennes et al., 2018b] Derennes, P., Morio, J., and Simatos, F. (2018b). A nonparametric importance sampling estimator for moment independent importance measures. Reliability Engineering & System Safety, (In press).

[Derennes et al., 2019b] Derennes, P., Morio, J., and Simatos, F. (2019b). Simultaneous estimation of complementary moment independent sensitivity measures for reliability analysis. (Under revision).

[Devroye and Gyorfi, 1985] Devroye, L. and Gyorfi, L. (1985). Nonparametric density estimation : the L1 view. John Wiley & Sons Incorporated.

[Dimov, 2008] Dimov, I. T. (2008). Monte Carlo methods for applied scientists. World Scientific.

[Dissmann et al., 2013] Dissmann, J., Brechmann, E. C., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. Computational Statistics & Data Analysis, 59 :52–69.

[Ditlevsen et al., 1990] Ditlevsen, O., Melchers, R. E., and Gluver, H. (1990). General multi-dimensional probability integration by directional simulation. Computers & Structures, 36(2) :355–368.

[Dudik et al., 2004] Dudik, M., Phillips, S. J., and Schapire, R. E. (2004). Performance guarantees for regularized maximum entropy density estimation. In International Conference on Computational Learning Theory, pages 472–486. Springer.

[Echard et al., 2011] Echard, B., Gayton, N., and Lemaire, M. (2011). Ak-mcs : an active learning reliability method combining Kriging and Monte Carlo simulation. Structural Safety, 33(2) :145–154.

[Fort et al., 2016] Fort, J.-C., Klein, T., and Rachdi, N. (2016). New sensitivity analysis subordinated to a contrast. Communications in Statistics-Theory and Methods, 45(15) :4349–4364.

[Gamboa et al., 2016] Gamboa, F., Janon, A., Klein, T., Lagnoux, A., and Prieur, C. (2016). Statistical inference for Sobol pick-freeze Monte Carlo method. Statistics, 50(4) :881–902.

[Gamboa et al., 2018] Gamboa, F., Klein, T., and Lagnoux, A. (2018). Sensitivity analysis based on Cramér–Von Mises distance. SIAM/ASA Journal on Uncertainty Quantification, 6(2) :522–548.

[Gamboa et al., 2019] Gamboa, F., Klein, T., Lagnoux, A., and Moreno, L. (2019). Sensitivity analysis in general metric spaces. working paper or preprint.

[Gelman et al., 1992] Gelman, A., Rubin, D. B., et al. (1992). Inference from iterative simulation using multiple sequences. Statistical science, 7(4) :457–472.

[Geman and Geman, 1984] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on pattern analysis and machine intelligence, (6) :721–741.

[Georgii, 2011] Georgii, H.-O. (2011). Gibbs measures and phase transitions, volume 9. Walter de Gruyter.

[Gerstner and Griebel, 1998] Gerstner, T. and Griebel, M. (1998). Numerical integration using sparse grids. Numerical algorithms, 18(3-4) :209.

[Geyer et al., 2019] Geyer, S., Papaioannou, I., and Straub, D. (2019). Cross entropy-based importance sampling using gaussian densities revisited. Structural Safety, 76 :15–27.

[Glasserman et al., 1999] Glasserman, P., Heidelberger, P., Shahabuddin, P., and Zajic, T. (1999). Multilevel splitting for estimating rare event probabilities. Operations Research, 47(4) :585–600.

[Grooteman, 2008] Grooteman, F. (2008). Adaptive radial-based importance sampling method for structural reliability. Structural Safety, 30(6) :533–542.

[Grooteman, 2011] Grooteman, F. (2011). An adaptive directional importance sampling method for structural reliability. Probabilistic Engineering Mechanics, 26(2) :134–141.

[Haff et al., 2010] Haff, I. H., Aas, K., and Frigessi, A. (2010). On the simplified pair-copula construction—simply useful or too simplistic ? Journal of Multivariate Analysis, 101(5) :1296–1310.

[Hammersley, 2013] Hammersley, J. (2013). Monte Carlo methods. Springer Science & Business Media.

[Hammersley and Morton, 1956] Hammersley, J. and Morton, K. (1956). A new Monte Carlo technique : antithetic variates. In Mathematical proceedings of the Cambridge philosophical society, volume 52, pages 449–475. Cambridge University Press.

[Hansen, 2008] Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. Econometric Theory, 24(3) :726–748.

[Hao and Singh, 2013] Hao, Z. and Singh, V. P. (2013). Modeling multisite streamflow dependence with maximum entropy copula. Water Resources Research, 49(10) :7139–7143.

[Hao and Singh, 2015] Hao, Z. and Singh, V. P. (2015). Integrating entropy and copula theories for hydrologic modeling and analysis. Entropy, 17(4) :2253–2280.

[Harbitz, 1986] Harbitz, A. (1986). An efficient sampling method for probability of failure calculation. Structural Safety, 3(2) :109–115.

[Hasofer and Lind, 1974] Hasofer, A. M. and Lind, N. C. (1974). Exact and invariant second-moment code format. Journal of the Engineering Mechanics division, 100(1) :111–121.

[Hastings, 1970] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications.

[Heidenreich et al., 2013] Heidenreich, N.-B., Schindler, A., and Sperlich, S. (2013). Bandwidth selection for kernel density estimation : a review of fully automatic selectors. AStA Advances in Statistical Analysis, 97(4) :403–433.

[Homem-de Mello and Rubinstein, 2002] Homem-de Mello, T. and Rubinstein, R. Y. (2002). Rare event estimation for static models via cross-entropy and importance sampling.

[Homma and Saltelli, 1996] Homma, T. and Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. Reliability Engineering & System Safety, 52(1) :1–17.

[Hooker, 2007] Hooker, G. (2007). Generalized functional anova diagnostics for high-dimensional functions of dependent variables. Journal of Computational and Graphical Statistics, 16(3) :709–732.

[Huang et al., 2016] Huang, X., Chen, J., and Zhu, H. (2016). Assessing small failure probabilities by AK–SS : an active learning method combining kriging and subset simulation. Structural Safety, 59 :86–95.

[Imbens and Lancaster, 1996] Imbens, G. W. and Lancaster, T. (1996). Efficient estimation and stratified sampling. Journal of Econometrics, 74(2) :289–318.

[Inverardi et al., 2005] Inverardi, P. N., Petri, A., Pontuale, G., and Tagliani, A. (2005). Stieltjes moment problem via fractional moments. Applied mathematics and computation, 166(3) :664–677.

[Iooss and Lemaître, 2015] 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.

[Iooss and Prieur, 2017] Iooss, B. and Prieur, C. (2017). Shapley effects for sensitivity analysis with dependent inputs : comparisons with sobol'indices, numerical estimation and applications. arXiv preprint arXiv :1707.01334.

[Jaworski et al., 2010] Jaworski, P., Durante, F., Hardle, W. K., and Rychlik, T. (2010). Copula theory and its applications, volume 198. Springer.

[Jaynes, 1957] Jaynes, E. T. (1957). Information theory and statistical mechanics. Physical review, 106(4) :620.

[Jones et al., 1998] Jones, D. R., Schonlau, M., and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. Journal of Global optimization, 13(4) :455–492.

[Kahn and Harris, 1951] Kahn, H. and Harris, T. E. (1951). Estimation of particle transmission by random sampling. National Bureau of Standards applied mathematics series, 12 :27–30.

[Kapur and Kesavan, 1992] Kapur, J. N. and Kesavan, H. K. (1992). Entropy optimization principles with applications. Academic Press, Inc., Boston, MA.

[Kendall, 1948] Kendall, M. G. (1948). Rank correlation methods.

[Kroese and Rubinstein, 2012] Kroese, D. P. and Rubinstein, R. Y. (2012). Monte Carlo methods. Wiley Interdisciplinary Reviews : Computational Statistics, 4(1) :48–58.

[Krzykacz-Hausmann, 2001] Krzykacz-Hausmann, B. (2001). Epistemic sensitivity analysis based on the concept of entropy. Proceedings of SAMO, pages 31–35.

[Kucherenko and Iooss, 2017] Kucherenko, S. and Iooss, B. (2017). Derivative-based global sensitivity measures. Handbook of uncertainty quantification, pages 1241–1263.

[Kucherenko et al., 2009] 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 and Leibler, 1951] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. The Annals of Mathematical Statistics, 22(1) :79–86.

[Kurtz and Song, 2013] Kurtz, N. and Song, J. (2013). Cross-entropy-based adaptive importance sampling using gaussian mixture. Structural Safety, 42 :35–44.

[Lamboni et al., 2013] Lamboni, M., Iooss, B., Popelin, A.-L., and Gamboa, F. (2013). Derivative-based global sensitivity measures : general links with sobol' indices and numerical tests. Mathematics and Computers in Simulation, 87 :45–54.

[Lebrun and Dutfoy, 2009] Lebrun, R. and Dutfoy, A. (2009). A generalization of the nataf transformation to distributions with elliptical copula. Probabilistic Engineering Mechanics, 24(2) :172–178.

[Lelièvre, 2018] Lelièvre, N. (2018). Développement des méthodes AK pour l'analyse de fiabilité - Focus sur les évènements rares et la grande dimension. PhD thesis, Université Clermont Auvergne.

[Lemaire, 2013] Lemaire, M. (2013). Structural reliability. John Wiley & Sons.

[Lemaître et al., 2015] 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.

[Li et al., 2012] Li, L., Lu, Z., and Feng, Jun andWang, B. (2012). Moment-independent importance measure of basic variable and its state dependent parameter solution. Structural Safety, 38 :40–47.

[Lin, 1992] Lin, G. D. (1992). Characterizations of distributions via moments. Sankhyā : The Indian Journal of Statistics, Series A, pages 128–132.

[Liu and Der Kiureghian, 1991] Liu, P.-L. and Der Kiureghian, A. (1991). Optimization algorithms for structural reliability. Structural Safety, 9(3) :161–177.

[Liu and Homma, 2009] Liu, Q. and Homma, T. (2009). A new computational method of a moment-independent uncertainty importance measure. Reliability Engineering & System Safety, 94(7) :1205–1211.

[Madsen et al., 2006] Madsen, H. O., Krenk, S., and Lind, N. C. (2006). Methods of structural safety. Courier Corporation.

[Mara et al., 2015] Mara, T. A., Tarantola, S., and Annoni, P. (2015). Non-parametric methods for global sensitivity analysis of model output with dependent inputs. Environmental modelling & software, 72 :173–183.

[Marrel et al., 2009] 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.

[Maume-Deschamps and Niang, 2018] Maume-Deschamps, V. and Niang, I. (2018). Estimation of quantile oriented sensitivity indices. Statistics & Probability Letters, 134 :122–127.

[Melchers, 1989] Melchers, R. (1989). Importance sampling in structural systems. Structural Safety, 6(1) :3–10.

[Metropolis et al., 1953] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. The journal of chemical physics, 21(6) :1087–1092.

[Moral et al., 2006] Moral, P. D., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo samplers. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 68(3) :411–436.

[Morales Napoles et al., 2010] Morales Napoles, O., Cooke, R. M., and Kurowicka, D. (2010). About the number of vines and regular vines on n nodes. https://doi.org/10.13140/RG.2.1.4400.8083.

[Morio, 2012] Morio, J. (2012). Extreme quantile estimation with nonparametric adaptive importance sampling. Simulation Modelling Practice and Theory, 27 :76–89.

[Morio and Balesdent, 2015] Morio, J. and Balesdent, M. (2015). Estimation of Rare Event Probabilities in Complex Aerospace and Other Systems : A Practical App Woodhead Publishing, Elsevier.

[Morio et al., 2014] Morio, J., Balesdent, M., Jacquemart, D., and Vergé, C. (2014). A survey of rare event simulation methods for static input–output models. Simulation Modelling Practice and Theory, 49 :287–304.

[Morris, 1991] Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. Technometrics, 33(2) :161–174.

[Nagler and Czado, 2016] Nagler, T. and Czado, C. (2016). Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. Journal of Multivariate Analysis, 151 :69–89.

[Nagler et al., 2017] Nagler, T., Schellhase, C., and Czado, C. (2017). Nonparametric estimation of simplified vine copula models : comparison of methods. Dependence Modeling, 5(1) :99–120.

[Nataf, 1962] Nataf, A. (1962). Détermination des distributions dont les marges sont données. Comptes Rendus de l'Academie des Sciences, 225 :42–43.

[Neddermeyer, 2009] Neddermeyer, J. C. (2009). Computationally efficient nonparametric importance sampling. Journal of the American Statistical Association, 104(486) :788–802.

[Novi Inverardi and Tagliani, 2003] Novi Inverardi, P. and Tagliani, A. (2003). Maximum entropy density estimation from fractional moments. Communications in Statistics-Theory and Methods, 32(2) :327–345.

[Ouvrard, 2000] Ouvrard, J.-Y. (2000). Probabilités tome 2, Master–Agrégation. Cassini.

[Owen, 2014] Owen, A. B. (2014). Sobol' indices and Shapley value. SIAM/ASA Journal on Uncertainty Quantification, 2(1) :245–251.

[Pianosi and Wagener, 2016] Pianosi, F. and Wagener, T. (2016). Understanding the time-varying importance of different uncertainty sources in hydrological modelling using global sensitivity analysis. Hydrological Processes, 30(22) :3991–4003.

[Plischke, 2012] Plischke, E. (2012). An adaptive correlation ratio method using the cumulative sum of the reordered output. Reliability Engineering & System Safety, 107 :149–156.

[Plischke et al., 2013] Plischke, E., Borgonovo, E., and Smith, C. L. (2013). Global sensitivity measures from given data. European Journal of Operational Research, 226(3) :536–550.

[Propp and Wilson, 1996] Propp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. Random Structures & Algorithms, 9(1-2) :223–252.

[Rabitz et al., 1999] Rabitz, H., Aliş, Ö. F., Shorter, J., and Shim, K. (1999). Efficient input-output model representations. Computer Physics Communications, 117(1-2) :11–20.

[Raguet and Marrel, 2018] Raguet, H. and Marrel, A. (2018). Target and conditional sensitivity analysis with emphasis on dependence measures. arXiv preprint arXiv :1801.10047.

[Rahman, 2016] Rahman, S. (2016). The f-sensitivity index. SIAM/ASA Journal on Uncertainty Quantification, 4(1) :130–162.

[Rajabi et al., 2015] Rajabi, M. M., Ataie-Ashtiani, B., and Simmons, C. T. (2015). Polynomial chaos expansions for uncertainty propagation and moment independent sensitivity analysis of seawater intrusion simulations. Journal of Hydrology, 520 :101–122.

[Rivoirard and Stoltz, 2012] Rivoirard, V. and Stoltz, G. (2012). Statistique mathématique en action. Vuibert.

[Roberts and Smith, 1994] Roberts, G. O. and Smith, A. F. (1994). Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. Stochastic processes and their applications, 49(2) :207–216.

[Rosenblatt, 1952] Rosenblatt, M. (1952). Remarks on a multivariate transformation. The annals of mathematical statistics, 23(3) :470–472.

[Rosenbluth and Rosenbluth, 1955] Rosenbluth, M. N. and Rosenbluth, A. W. (1955). Monte Carlo calculation of the average extension of molecular chains. The Journal of Chemical Physics, 23(2) :356–359.

[Rubinstein and Kroese, 2004] Rubinstein, R. Y. and Kroese, D. P. (2004). The cross-entropy method : a unified approach to combinatorial optimization, monte-carlo simulation and machine learning. In Information Science and Statistics. Springer.

[Saltelli, 2002] Saltelli, A. (2002). Sensitivity analysis for importance assessment. Risk Analysis, 22(3) :579–590.

[Saltelli and Annoni, 2010] Saltelli, A. and Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. Environmental Modelling & Software, 25(12) :1508–1517.

[Schöbi et al., 2016] Schöbi, R., Sudret, B., and Marelli, S. (2016). Rare event estimation using polynomial-chaos Kriging. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A : Civil Engineering, 3(2) :D4016002.

[Schuëller and Stix, 1987] Schuëller, G. I. and Stix, R. (1987). A critical appraisal of methods to determine failure probabilities. Structural Safety, 4(4) :293–309.

[Shannon, 1948] Shannon, C. E. (1948). A mathematical theory of communication. Bell System Tech. J., 27 :379–423, 623–656.

[Shapley, 1953] Shapley, L. S. (1953). A value for n-person games. Contributions to the Theory of Games, 2(28) :307–317.

[Sheather and Jones, 1991] Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society : Series B (Methodological), 53(3) :683–690.

[Sherlock and Roberts, 2009] Sherlock, C. and Roberts, G. (2009). Optimal scaling of the random walk metropolis on elliptically symmetric unimodal targets. Bernoulli, pages 774–798.

[Silverman, 2018] Silverman, B. W. (2018). Density estimation for statistics and data analysis. Routledge.

[Smith et al., 1997] Smith, P. J., Shafi, M., and Gao, H. (1997). Quick simulation : A review of importance sampling techniques in communications systems. IEEE Journal on Selected Areas in Communications, 15(4) :597–613.

[Sobol and Gresham, 1995] Sobol, I. and Gresham, A. (1995). On an alternative global sensitivity estimators. Proceedings of SAMO, pages 40–42.

[Sobol, 1993] Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. Mathematical modelling and computational experiments, 1(4) :407–414.

[Sobol, 2001] 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.

[Soize, 2017] Soize, C. (2017). Uncertainty Quantification : An Accelerated Course with Advanced Applications in Computational Engineering. Springer.

[Song et al., 2016] Song, E., Nelson, B. L., and Staum, J. (2016). Shapley effects for global sensitivity analysis : Theory and computation. SIAM/ASA Journal on Uncertainty Quantification, 4(1) :1060–1083.

[Sudret, 2008] Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. Reliability engineering & system safety, 93(7) :964–979.

[Tagliani, 2011] Tagliani, A. (2011). Hausdorff moment problem and fractional moments : a simplified procedure. Applied Mathematics and Computation, 218(8) :4423–4432.

[Terrell and Scott, 1992] Terrell, G. R. and Scott, D. W. (1992). Variable kernel density estimation. The Annals of Statistics, pages 1236–1265.

[Tewari et al., 2011] Tewari, A., Giering, M. J., and Raghunathan, A. (2011). Parametric characterization of multimodal distributions with non-gaussian modes. In 2011 IEEE 11th International Conference on Data Mining Workshops, pages 286–292. IEEE.

[Tokdar and Kass, 2010] Tokdar, S. T. and Kass, R. E. (2010). Importance sampling : a review. Wiley Interdisciplinary Reviews : Computational Statistics, 2(1) :54–60.

[Wand and Jones, 1994] Wand, M. P. and Jones, M. C. (1994). Kernel smoothing. Chapman and Hall/CRC.

[Wand et al., 1991] Wand, M. P., Marron, J. S., and Ruppert, D. (1991). Transformations in density estimation. Journal of the American Statistical Association, 86(414) :343–353.

[Wang et al., 2018] Wang, Y., Xiao, S., and Lu, Z. (2018). A new efficient simulation method based on Bayes' theorem and importance sampling Markov chain simulation to estimate the failure-probability-based global sensitivity measure. Aerospace Science and Technology, 79 :364 – 372.

[Wei et al., 2014] Wei, P., Lu, Z., and Song, J. (2014). Moment-independent sensitivity analysis using copula. Risk Analysis, 34(2) :210–222.

[Wei et al., 2013] Wei, P., Lu, Z., and Yuan, X. (2013). Monte Carlo simulation for moment-independent sensitivity analysis. Reliability Engineering & System Safety, 110 :60–67.

[Xie et al., 2018] Xie, X., Schenkendorf, R., and Krewer, U. (2018). Efficient sensitivity analysis and interpretation of parameter correlations in chemical engineering. Reliability Engineering & System Safety.

[Yang and Marron, 1999] Yang, L. and Marron, J. S. (1999). Iterated transformation–kernel density estimation. Journal of the American Statistical Association, 94(446) :580–589.

[Yun et al., 2018] Yun, W., Lu, Z., and Jiang, X. (2018). An efficient method for moment-independent global sensitivity analysis by dimensional reduction technique and principle of maximum entropy. Reliability Engineering & System Safety.

[Zhang et al., 2014] Zhang, L., Lu, Z., Cheng, L., and Fan, C. (2014). A new method for evaluating Borgonovo moment-independent importance measure with its application in an aircraft structure. Reliability Engineering & System Safety, 132 :163–175.

[Zhang et al., 2015] Zhang, L., Lu, Z., Cheng, L., and Hong, D. (2015). Moment-independent regional sensitivity analysis of complicated models with great efficiency. International Journal for Numerical Methods in Engineering, 103(13) :996–1014.

[Zhang, 1996] Zhang, P. (1996). Nonparametric importance sampling. Journal of the American Statistical Association, 91(435) :1245–1253.

[Zhang and Pandey, 2013] Zhang, X. and Pandey, M. D. (2013). Structural reliability analysis based on the concepts of entropy, fractional moment and dimensional reduction method. Structural Safety, 43 :28–40.

[Zhao and Ono, 1999] Zhao, Y.-G. and Ono, T. (1999). A general procedure for first/second-order reliability method (FORM/SORM). Structural Safety, 21(2) :95–112.

[Zhou et al., 2018] Zhou, Y., Lu, Z., Shi, Y., and Cheng, K. (2018). A vine copula–based method for analyzing the moment-independent importance measure of the multivariate output. Proceedings of the Institution of Mechanical Engineers, Part O : Journal of Risk and Reliability, page 1748006X18781121.