Méthodes probabilistes pour l'estimation de probabilités de défaillance

Lucie Bernard

UNIVERSITÉ DE TOURS

École Doctorale Mathématiques, Informatique, Physique Théorique et Ingénierie des Systèmes

Institut Denis Poisson

THÈSE présenté par :

Lucie Bernard

soutenue le : 28 juin 2019

pour obtenir le grade de : Docteur de l'université de Tours

Discipline/ Spécialité : Mathématiques

Méthodes probabilistes pour l'estimation de probabilités de défaillance

THÈSE DIRIGÉE PAR :

MALRIEU Florent	Professeur, Université de Tours
GUYADER Arnaud	Professeur, Sorbonne Université

RAPPORTEURS :

BIERMÉ Hermine	Professeure, Université de Poitiers
NOUY Anthony	Professeur, École Centrale Nantes

JURY :

BIERMÉ Hermine	Professeure, Université de Poitiers
CHAUVEAU Didier	Professeur, Université d'Orléans
GUYADER Arnaud	Professeur, Sorbonne Université
LEDUC Philippe	Docteur ingénieur, STMicroelectronics
MALRIEU Florent	Professeur, Université de Tours
NOUY Anthony	Professeur, École Centrale Nantes
SCIAUVEAU Marion	Docteure PRAG, Université de Tours

MEMBRE INVITÉ :

REMY Emmanuel	Ingénieur, EDF R&D

Remerciements

En premier lieu, je remercie très sincèrement mon encadrant industriel Philippe Leduc, pour avoir eu l'idée et la volonté de monter ce projet de thèse Cifre, ainsi que pour m'avoir fait confiance quant à son aboutissement. Tout en me laissant une grande liberté de travail, Philippe a été présent pour me guider lors des étapes inévitablement laborieuses. Nos échanges, ses nombreuses idées sur le sujet et ses relectures attentives de ce manuscrit m'ont en effet beaucoup aidé. Je lui suis également reconnaissante de m'avoir offert plusieurs opportunités sympathiques au sein de STMicroelectronics, et notamment celle de la responsabilité du stage de Raphaël Smith.

Je tiens ensuite à exprimer toute ma gratitude à mes directeurs de thèse, Arnaud Guyader et Florent Malrieu, pour leur investissement qui s'est manifesté, entre autre, par de nombreux déplacements entre Paris et Tours, ainsi que par des heures et des heures passées à me partager leurs multiples connaissances. Leurs conseils, souvent complémentaires, ont largement contribué à me construire et à m'affirmer en tant que jeune chercheuse. Je remercie notamment Florent de m'avoir toujours laissé repartir d'une réunion de travail avec l'information ou la petite astuce qui me manquait, sa capacité à comprendre immédiatement n'importe quel problème de mathématiques étant redoutable. Je tiens également à mentionner son soutien lors de la période de rédaction à l'Université de Tours : sa bienveillance et ses encouragements quotidiens m'auront considérablement aidé à aller au bout de cet exercice difficile. Je remercie aussi particulièrement Arnaud pour ses relectures minutieuses et systématiques de mes diverses productions, ainsi que pour son aide et ses précieuses suggestions. En effet, les sujets de recherche, originaux et stimulants, vers lesquels Arnaud m'a continuellement orienté, m'ont permis non seulement d'étendre et de renouveler régulièrement mes idées, mais aussi de surmonter les moments de doute et de découragement.

Je tiens aussi à remercier Hermine Biermé et Anthony Nouy d'avoir accepté de rapporter ma thèse et, par conséquent, pour le temps consacré à la lecture attentive et aux commentaires constructifs de ces travaux. J'exprime également toute ma gratitude à Didier Chauveau et Marion Sciauveau pour leur présence dans le jury de ma soutenance, avec un petit clin d'œil à cette dernière pour son coup de pouce sur la rédaction du chapitre 4 de ce manuscrit. Enfin, je remercie également Emmanuel Remy pour avoir répondu à l'invitation et accepté de porter un regard pragmatique sur mes travaux.

Je souhaite également profiter de cet exercice des remerciements pour saluer amicalement les différentes personnes qui m'ont aidé à avancer, d'une façon ou d'une autre, tout au

long de cette thèse. Je pense notamment à Albert Cohen qui m'a accordé de son temps pour discuter lors de la conférence annuelle du GdR MascotNum, à Nantes, en 2018. Cet échange a été d'autant plus intéressant que l'algorithme développé dans le chapitre 4 de ce manuscrit s'inspire largement de ses travaux sur les méthodes d'optimisation de fonctions. J'ai notamment pris connaissance de ces derniers par l'intermédiaire de Virginie Ehrlacher, qui avait accepté de venir discuter de ma problématique à l'UPMC, au printemps 2017. Je tiens à mentionner aussi Julien Bect, qui a accepté de me rencontrer en mars 2017 afin de répondre à toutes mes questions sur les stratégies SUR utilisées dans le chapitre 3 de ce manuscrit. Enfin, je salue Pierre Barbillon qui a aussi pris de son temps, en juillet 2017, pour m'expliquer ses travaux sur l'estimation de probabilités d'évènements rares. C'est souvent grâce aux conférences que ce type d'échanges peut avoir lieu, et j'en profite pour saluer tous les organisateurs, chercheurs et doctorants, que j'ai rencontrés lors de ces évènements et avec qui j'ai souvenir d'avoir passé de très bons moments.

J'ai débuté ma thèse au LPSM et, pour l'accueil bienveillant qui m'y a été réservé, je tiens à remercier Roxane, Matthieu, Erwan, Assia, Mokhtar et Quyen. Plus tard, mon quotidien s'étant aussi déroulé aux côtés de Félix, Nazih, Yohann, Émilie, Taieb et Vincent, je tiens à les saluer amicalement et je leur souhaite le meilleur pour la suite. Enfin, je remercie les différents membres de l'Institut Denis Poisson, pour leur réelle sympathie à mon arrivée courant 2018, et notamment Marion et Amélie, que je ne remercierai jamais assez pour leur amitié instantanée et leur soutien quotidien durant les fastidieux mois de rédaction.

Afin que je puisse travailler à plein temps chez STMicroelectronics, je me suis officiellement installée à Tours en septembre 2017. Je salue donc les membres de l'équipe CAD que j'ai côtoyé quotidiennement à partir de cette date-là : Arnaud, Fabrice, Fabien, Armelle, WeiZhen, Vincent, Olivier, Véronique et Gwenaël. De mars à septembre 2018, j'ai eu la chance de travailler avec Raphaël Smith pour son stage de master 2. Non seulement la présence et les échanges quotidiens avec Raphaël ont été très agréables, mais je lui suis vraiment reconnaissante de son investissement et de l'aide qu'il m'a apporté sur le développement et la mise à jour des outils informatiques de STMicroelectronics.

Ces années de thèse se sont accompagnées d'un nombre considérable d'allers-retours entre Tours et Paris, au point où je considère la gare Montparnasse comme une résidence secondaire. Je tiens donc à remercier tous mes amis qui ne m'ont absolument jamais reproché mon indisponibilité et mon organisation de dernière minute. Je pense évidemment en premier lieu à mon CRI, parisien et breton, avec qui j'évolue depuis 7 ans déjà et dont l'amitié m'est très précieuse : Anne, Julie, Florian, Noémie, Adime et Coralie. Les nombreux weekends en dehors de Paris (et les soirées cocktails pâte à crêpes) ont été une réelle bouffée d'air frais pour oublier un peu la thèse et, pour autant, je les remercie d'avoir toujours patiemment écouté mes petites angoisses vis-à-vis de celle-ci (et pour les encouragements que cela a nécessairement engendré). J'en profite pour exprimer toute mon affection à Julie, pour toutes les fois où l'on s'est retrouvées pour déjeuner dans Paris lors de cette période plutôt difficile qu'a été l'automne 2016. Je remercie également Jeanne, Valentin Brdd, Matthieu et Valentin Blld pour toutes les soirées, à Croissy ou ailleurs, qui ont rythmé mes années de thèse à Paris. Bien que je les confonde un peu toutes, il est certain qu'elles se sont accompagnées de hurlements fan-zonesques, de fuites dans le métro à toute berzingue, d'escape games malaisants et de dégustations de burgers au goût mitigé. C'était donc avec

le cœur franchement serré que je m'étais préparée à déménager à Tours à la fin de l'été 2017. Mais, c'était sans savoir que, 1 an et demi plus tard, je serais toute aussi émue à l'idée d'en repartir définitivement. J'adresse donc un grand merci, rempli d'une tendre affection, à tous les tourangeaux avec qui j'ai partagé un footing sur les bords de Loire, une pinte de blanche à la Guinguette ou encore un café place Plum. Si, en écrivant ces mots, je pense fortement à mes rencontres avec Chloé & Dan, ou encore Anthony-Matthieu, j'admets que la *médaille d'honneur* revient à Florencja, aka Marie la plage, à qui je suis infiniment redevable pour son amitié, sa présence (c'est quand même pratique d'être voisines) et son soutien quotidien durant la totalité de ce séjour tourangeau. Chaque moment passé ensemble a abouti à tellement de souvenirs et de réparties mémorables, que j'en ris encore aujourd'hui et pour longtemps. Je la remercie notamment pour les challenges sportifs (le semi à Paris et le passage à la buvette des 10km de Tours), les apéros à rallonge rue Colbert et les cafés du vendredi matin (suivis du Vouvray en fin de journée). Je tiens aussi à saluer amicalement Clément, Laurianne et Anastasia, car ils ont aussi contribué à leur façon à mon arrivée jusqu'ici : Clément en organisant une colocation Rochelaise déjantée à base de produits périmés, Laurianne en m'incitant à papoter pendant littéralement chaque cours (pour mon plus grand plaisir) et Anna en me prouvant que vodka et titre de majeure de promo ne sont pas incompatibles. J'ai aussi une petite pensée chaleureuse pour mes amies des Sables, à savoir Sophie, Meg-Anne et Claudia. Enfin, j'adresse toute mon affection à Valentin, dont l'ambition démesurée et l'amitié inébranlable depuis 10 ans m'ont toujours énormément apporté et inspiré.

Je souhaite à présent remercier l'ensemble de ma famille pour les bons moments passés tous ensemble en Vendée. J'ai une pensée particulièrement émue pour ma grand-mère, Marie-Joseph, qui a toujours été très présente pour mon frère, mes sœurs, et pour moi-même, et qui nous a quitté au cours de ces années de doctorat. Je tiens évidemment à remercier mes parents, Chrystel et Jean-François, pour leur soutien sans égal, leur confiance concernant mon choix de poursuite d'étude, et pour n'avoir jamais douté de ma capacité d'y parvenir. C'est grâce au nombre incalculable d'encouragements qu'ils m'ont prodigués, ainsi qu'à leur réconfort, que j'ai réussi à aller au bout de ce projet. Je pourrais en dire autant du soutien apporté par mon frère, Raphaël, et par mes sœurs, Jeanne et Angèle. Puisque chaque instant passé tous ensemble s'accompagne nécessairement des blagues hilarantes de Raph (dont je suis la plus grande fan, bien que la plupart d'entre elles soient suffisamment subtiles pour que je ne les comprenne en fait pas du tout), de l'empathie démesurée de Jeanne (qui a outrepassé depuis longtemps le statut de petite sœur et qui m'inspire un peu plus chaque jour qui passe) et de la bienveillance incomparable de Gégèle (qui performe dans tous les domaines, aussi bien dans le jeté de carottes cuites que dans la plaidoirie féministe), il n'est pas difficile de comprendre pourquoi leur présence à mes côtés est ce que j'ai de plus précieux. Tous ces moments de bonheur en famille n'auraient d'ailleurs pas été les mêmes sans la présence de Noémie, Valentin, Benjamin et de Maxence, mon petit neveu adoré.

Enfin, il est grand temps de préciser que toutes ces petites anecdotes et moments de vie ont majoritairement été vécus aux côtés d'Arthur, qui me pousse à persévérer et à ne pas me décourager depuis plusieurs années maintenant. Son soutien permanent, ses conseils rassurants et sa patience indéfectible face aux monologues interminables et répétitifs que

j'ai pu tenir à propos de la thèse, m'ont inévitablement aidé à me lancer dans ce projet et à m'y tenir. Je ne le remercierai jamais assez pour cela, mais aussi pour les heures passées ensemble (bien installés dans Alphonse) à discuter, souvent de tout et rarement de rien, et à rire inlassablement de *Carabistouille*. Je lui adresse tout mon amour (ainsi que toute mon affection aux membres de sa plus proche famille), et c'est sans aucun doute que je lui dédie ce travail de thèse.

Résumé

Pour évaluer la rentabilité d'une production en amont du lancement de son processus de fabrication, la plupart des entreprises industrielles ont recours à la simulation numérique. Cela permet de tester virtuellement plusieurs configurations des paramètres d'un produit donné et de statuer quant à ses performances (i.e. les spécifications imposées par le cahier des charges). Afin de mesurer l'impact des fluctuations des procédés industriels sur les performances du produit, nous nous intéressons en particulier à l'estimation de sa probabilité de défaillance. Chaque simulation exigeant l'exécution d'un code de calcul complexe et coûteux, il n'est pas possible d'effectuer un nombre de tests suffisant pour estimer cette probabilité via, par exemple, une méthode Monte-Carlo. Sous la contrainte d'un nombre limité d'appels au code, nous proposons deux méthodes d'estimation très différentes.

La première s'appuie sur les principes de l'estimation bayésienne. Nos observations sont les résultats de simulation numérique. La probabilité de défaillance est vue comme une variable aléatoire, dont la construction repose sur celle d'un processus aléatoire destiné à modéliser le code de calcul coûteux. Pour définir correctement ce modèle, on utilise la méthode de krigeage. Conditionnellement aux observations, la loi a posteriori de la variable aléatoire, qui modélise la probabilité de défaillance, est inaccessible. Pour apprendre sur cette loi, nous construisons des approximations des caractéristiques suivantes : espérance, variance, quantiles... On utilise pour cela la théorie des ordres stochastiques pour la comparaison de variables aléatoires et, plus particulièrement, l'ordre convexe. La construction d'un plan d'expériences optimal est assurée par la mise en place d'une procédure de planification d'expériences séquentielle, basée sur le principe des stratégies SUR (« Stepwise Uncertainty Reduction »).

La seconde méthode est une procédure itérative, particulièrement adaptée au cas où la probabilité de défaillance est très petite, i.e. l'événement redouté est rare. Le code de calcul coûteux est représenté par une fonction que l'on suppose lipschitzienne. À chaque itération, cette hypothèse est utilisée pour construire des approximations, par défaut et par excès, de la probabilité de défaillance. Nous montrons que ces approximations convergent vers la valeur vraie avec le nombre d'itérations. En pratique, on les estime grâce à la méthode Monte-Carlo dite de « splitting ».

Les méthodes que l'on propose sont relativement simples à mettre en œuvre et les résultats qu'elles fournissent peuvent être interprétés sans difficulté. Nous les testons sur divers exemples, ainsi que sur un cas réel provenant de la société STMicroelectronics.

Mots clés : Probabilité de défaillance, événement rare, méthodes Monte-Carlo et de « splitting », krigeage, processus gaussien, ordre convexe, encadrement de quantiles, inférence bayésienne, plan d'expériences, stratégies SUR, fonction lipschitzienne.

Abstract

To evaluate the profitability of a production before the launch of the manufacturing process, most industrial companies use numerical simulation. This allows to test virtually several configurations of the parameters of a given product and to decide on its performance (defined by the specifications). In order to measure the impact of industrial process fluctuations on product performance, we are particularly interested in estimating the probability of failure of the product. Since each simulation requires the execution of a complex and expensive calculation code, it is not possible to perform a sufficient number of tests to estimate this probability using, for example, a Monte-Carlo method. Under the constraint of a limited number of code calls, we propose two very different estimation methods.

The first is based on the principles of Bayesian estimation. Our observations are the results of numerical simulation. The probability of failure is seen as a random variable, the construction of which is based on that of a random process to model the expensive calculation code. To correctly define this model, the Kriging method is used. Conditionally to the observations, the posterior distribution of the random variable, which models the probability of failure, is inaccessible. To learn about this distribution, we construct approximations of the following characteristics : expectation, variance, quantiles... We use the theory of stochastic orders to compare random variables and, more specifically, the convex order. An optimal design of experiments is ensured by the implementation of a sequential planning procedure, based on the principle of the SUR (« Stepwise Uncertainty Reduction ») strategies.

The second method is an iterative procedure, particularly adapted to the case where the probability of failure is very small, i.e. the redoubt event is rare. The expensive calculation code is represented by a function that is assumed to be Lipschitz continuous. At each iteration, this hypothesis is used to construct approximations, by default and by excess, of the probability of failure. We show that these approximations converge towards the true value with the number of iterations. In practice, they are estimated using the Monte-Carlo method known as splitting method.

The proposed methods are relatively simple to implement and the results they provide can be easily interpreted. We test them on various examples, as well as on a real case from STMicroelectronics.

Keywords : Probability of failure, rare event, Monte-Carlo methods and splitting, Kriging, Gaussian processes, convex order, bounds on quantiles, Bayesian inference, design of experiments, SUR strategies, Lipschitz function.

Table des matières

Chapitre 1

Introduction

1.1 Contexte et motivations

Une des préoccupations majeures dans l'industrie est de fournir des produits conformes aux attentes des clients. La difficulté est d'y parvenir sans risquer de voir les systèmes, et procédures mis en œuvre pour atteindre cet objectif, devenir contre-productifs en cherchant à atteindre des niveaux de qualité largement supérieurs. Non seulement cela ne répond pas au besoin initial, mais une mauvaise gestion de la qualité donne inévitablement lieu à une augmentation des coûts de production, ce qui pénalise la rentabilité. Afin de garantir leur compétitivité économique, les entreprises industrielles sont donc quotidiennement confrontées à l'évaluation et la maîtrise de la performance de leurs systèmes de production.

Un procédé de fabrication étant composé d'une succession d'opérations complexes, il est nécessairement soumis à de la variabilité (fluctuations des matériaux, conditions de répétabilité et de reproductibilité instables, dégradations...). Cette variabilité peut significativement détériorer certaines caractéristiques du produit manufacturé, engendrant des pièces non-fonctionnelles. Pour s'assurer que les résultats du processus de fabrication satisfont les contraintes imposées par les clients en terme de qualité, c'est-à-dire le cahier des charges, les entreprises ont généralement recours à des outils de simulation numérique (on parle aussi d'expérience simulée). La simulation numérique désigne l'exécution d'un code de calcul élaboré sur la base d'un modèle mathématique, destiné à représenter des phénomènes physiques, réels et complexes. Les industriels peuvent alors tester virtuellement la fonctionnalité d'une pièce selon différents scénarios de conception. En particulier, l'impact de la variabilité intrinsèque au processus de fabrication peut être mesuré en faisant varier les paramètres fluctuants et en prédisant les performances du produit manufacturé. Cette façon de procéder évite la multiplication d'essais réels, généralement coûteux et difficiles à réaliser. Les gains sont alors multiples : optimisation de la main-d'œuvre et de la matière utilisée, réduction de la perte matérielle et de la durée de conception, amélioration de la performance du produit... Dans le cadre de négociations commerciales, il est primordial pour les entreprises de fournir aux clients des mesures d'évaluation de la conformité qui tiennent compte des contraintes de fabrication. C'est pourquoi, les tests par simulation numérique s'accompagnent généralement d'indicateurs qui visent à mesurer la performance du produit et fournissent des critères de décision quant à la rentabilité du projet. Par

exemple, en supposant que toutes les configurations possibles des paramètres impactant la performance d'un produit puissent être explorées, le ratio de produits non-conformes, c'est-à-dire ne respectant pas le cahier des charges, sur le nombre total de produits considérés, est un indicateur de la performance du système de production.

Quand le code de calcul utilisé pour la simulation numérique est coûteux en temps (plusieurs heures ou plusieurs jours pour une seule simulation), ainsi qu'en ressources informatiques, il n'est cependant pas envisageable de procéder ainsi. Seul un nombre limité d'appels au code est réalisable et les résultats des simulations constituent la seule information disponible pour évaluer la qualité du produit. En outre, le code est généralement trop complexe pour disposer d'une expression analytique (boîte-noire). Dans ces conditions, plusieurs questions se posent : quelles sont les configurations des paramètres du produit, à explorer par simulation numérique, qui fournissent le maximum d'information sur la performance du produit manufacturé ? Comment construire un outil fiable d'aide à la décision lorsque les données sont rares ? Comment mesurer l'incertitude sur les indicateurs de la performance du produit ? Cela conduit à formaliser le problème comme ci-après.

1.2 Présentation du problème

1.2.1 Cadre et notations

Dans tout le manuscrit, le cadre dans lequel nous nous plaçons est le suivant. Étant donnée la variabilité inhérente au processus de fabrication, le produit étudié est caractérisé par un nombre $d \in \mathbb{N}^*$ de paramètres fluctuants. Généralement appelés facteurs, ces paramètres varient dans un espace $\mathbb{X} \subseteq \mathbb{R}^d$ supposé connu. Un jeu de facteurs est donc un vecteur $\mathbf{x} \in \mathbb{X}$ représentant des conditions expérimentales, c'est-à-dire une configuration possible des paramètres du produit manufacturé. Une simulation numérique requiert l'évaluation d'un code numérique. Celui-ci est représenté par une fonction g définie sur $\mathbb{X}$ et à valeurs dans $\mathbb{R}$. Elle peut être évaluée en tout point de $\mathbb{X}$, mais elle est coûteuse en temps de calcul. De plus, son expression analytique est inconnue. On parle de fonction boîte-noire et on suppose ici qu'elle est déterministe, c'est-à-dire qu'évaluer g plusieurs fois au même point n'apporte aucune information supplémentaire. Pour une entrée vectorielle $\mathbf{x} \in \mathbb{X}$ donnée, la sortie scalaire $g(\mathbf{x})$ est une valeur numérique, généralement appelée réponse, qui permet de mesurer la performance du produit dans cette configuration. On dit que le produit caractérisé par le jeu de facteurs $\mathbf{x}$ ne satisfait pas les spécifications imposées, c'est-à-dire qu'il est défaillant ou non-fonctionnel, si $g(\mathbf{x}) > T$, où $T \in \mathbb{R}$ est un seuil fixé et connu. Ainsi, on appelle domaine de défaillance l'ensemble $\mathbb{F} \subset \mathbb{X}$ défini par :

$$\mathbb{F} = \{\mathbf{x} \in \mathbb{X} : g(\mathbf{x}) > T\}.$$

1.2.2 Problématique

Pour mettre en place une analyse probabiliste de la défaillance du produit, on considère un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, un espace mesurable $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ et une variable aléatoire $\mathbf{X}$ à valeurs dans $\mathbb{X}$. La distribution de probabilité $\mathbf{P}_{\mathbf{X}}$ de $\mathbf{X}$ modélise la variabilité des facteurs. On suppose que l'on sait simuler suivant cette loi, de sorte qu'une suite $(\mathbf{X}_i)_{1 \leq i \leq N}$,

où $N \in \mathbb{N}^*$, de variables aléatoires i.i.d. suivant $\mathbf{P_X}$ est un échantillon de configurations des paramètres du produit étudié. Pour mesurer l'impact des fluctuations du processus de fabrication sur les performances du produit, on s'intéresse alors à l'estimation de la probabilité p définie par :

$$p = \mathbb{P}(g(\mathbf{X}) > T) = \mathbb{E}[\mathbb{1}_{g(\mathbf{X})>T}] = \int_{\mathbb{X}} \mathbb{1}_{g(\mathbf{x})>T} \mathbf{P_X}(d\mathbf{x}) = \mathbb{P}(\mathbf{X} \in \mathbb{F}). \tag{1.1}$$

Étant donné ce qui précède, on l'appelle naturellement probabilité de défaillance. Avoir accès à une estimation de p en amont du lancement du processus de fabrication est une information primordiale, aussi bien pour les industriels que pour les clients : cela donne une indication sur la qualité de la production, contribue à la gestion et à l'optimisation du système de production, et aide à la décision vis-à-vis de la rentabilité du projet. Néanmoins, sous la contrainte que g est une fonction boîte-noire, la loi de $g(\mathbf{X})$ est inconnue et la probabilité p ne peut pas être calculée directement. En outre, bien que l'on sache simuler suivant $\mathbf{P_X}$, la contrainte supplémentaire que g est une fonction coûteuse en temps de calcul ne permet pas de simuler un grand nombre de réalisations de $g(\mathbf{X})$. La distribution empirique de cette variable aléatoire est donc, elle aussi, hors d'atteinte, et l'estimation de p par le calcul du ratio de produits défaillants sur le nombre total de produits simulés n'est pas envisageable dans cette situation. Plus de détails sur ce point seront donnés dans le chapitre 2, lors de la présentation de la méthode Monte-Carlo naïve.

Pour fixer les idées, un exemple en dimension $d = 1$ est donné figure 1.1. La fonction g, qui est inconnue en pratique, est ici représentée par la courbe noire sur l'intervalle $[0, 1]$. Le seuil T est la droite en pointillés rouges et le domaine de défaillance est l'intervalle $\left[\frac{3}{4}, 1\right]$. Dans cet exemple, la loi $\mathbf{P_X}$ admet une densité $f_{\mathbf{X}}$ (courbe grise) et la probabilité de défaillance $p > 0$ que l'on cherche à estimer est l'aire grise sous la courbe représentative de cette fonction. Les points orange représentent $N = 100$ réalisations des paramètres du produit, tirées suivant leur densité $f_{\mathbf{X}}$. La valeur de g en chacun de ces points est systématiquement en-dessous de T, de sorte que chaque configuration conduit à un produit fonctionnel. L'estimation de p qui en résulte est donc très mauvaise, car égale à 0. On en conclut que le nombre de simulations N n'est pas assez élevé.

1.2.3 Objectifs

L'ensemble des méthodes proposées dans cette thèse a pour objectif de fournir une estimation de la probabilité de défaillance p et ce malgré les difficultés que nous venons d'énoncer. Pour ce faire, on suppose qu'un budget restreint de $n \in \mathbb{N}^*$ appels à la fonction g est disponible. Quelle que soit l'approche statistique considérée, cela implique, d'une part, de choisir une suite de points $(\mathbf{x}_i)_{1 \leq i \leq n}$ de $\mathbb{X}$ où effectuer les évaluations de la fonction g, et d'autre part, de construire un estimateur fiable de la probabilité de défaillance p étant donné ce nombre limité d'observations. En d'autres termes, il s'agit de choisir les configurations des paramètres du produit à explorer en priorité par simulation numérique pour obtenir de l'information pertinente sur ses performances. Enfin, nous nous attacherons à fournir une mesure de l'incertitude sur le résultat obtenu, une donnée essentielle lorsque l'information est rare, et qui n'est pas systématiquement fournie par les méthodes déjà existantes dans la littérature pour répondre à une problématique similaire à la nôtre.

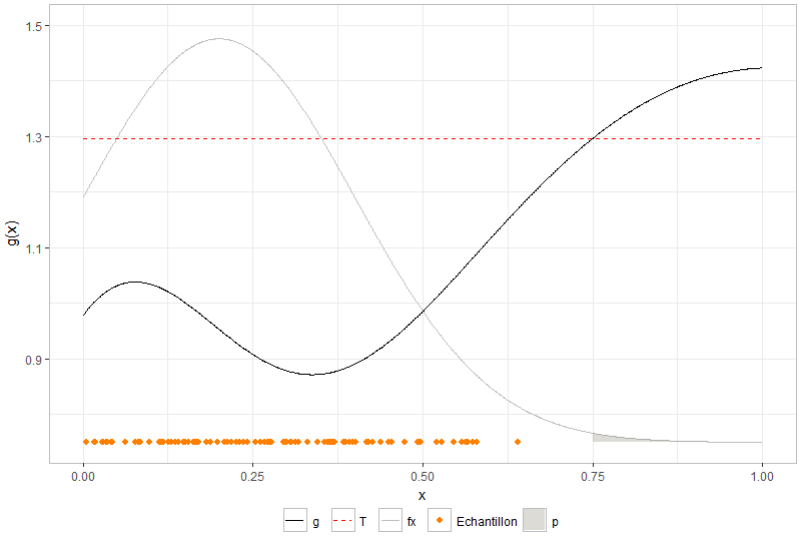

FIGURE 1.1 – Exemple en dimension $d = 1$. Courbe noire : la fonction g. Droite rouge : le seuil T. Courbe grise : la densité $f_{\mathbf{X}}$ de la loi de $\mathbf{X}$. Aire grise : la probabilité de défaillance p que l'on veut estimer. Le domaine de défaillance est l'intervalle $\left[\frac{3}{4}, 1\right]$. Points orange : un échantillon $(\mathbf{X}_i)_{1 \leq i \leq 100}$ i.i.d. suivant $f_{\mathbf{X}}$. L'estimation de p relative à cet échantillon est égale à 0.

1.3 Organisation de la thèse

1.3.1 Résumé du chapitre 2

Si l'objectif principal du chapitre 2 est d'aider le lecteur à la compréhension des méthodes que nous développons dans les chapitres 3 et 4, il a également pour but de présenter quelques méthodes usuelles pour l'estimation d'une probabilité de défaillance. En effet, notre problématique a déjà été étudiée dans divers domaines d'application, et on trouve dans la littérature de nombreuses approches pour y répondre. Cependant, on verra que les contraintes industrielles – à savoir que tout appel au code g est coûteux en temps – restreignent le choix de méthodes qui peuvent véritablement être utilisées au sein d'une entreprise.

On commence par présenter les méthodes Monte-Carlo, qui reposent sur la simulation de variables aléatoires pour l'estimation de p. La plus connue est la méthode Monte-Carlo naïve (ou standard), mais on introduit également des méthodes dites de réduction de variance, qui sont particulièrement adaptées au cas où la probabilité p est proche de 0. En effet, il est raisonnable de supposer que le produit étudié est globalement performant, c'est-à-dire que la probabilité de défaillance est très petite. Dans l'objectif de modéliser la fonction g inconnue à partir de résultats de simulation numérique, la seconde partie de ce chapitre est consacrée à la méthode de modélisation par processus aléatoire gaussien, appelée aussi krigeage. Un modèle de processus gaussien peut être déterminé même pour de petits jeux de données, et il s'agit un modèle très flexible capable de décrire des surfaces de réponses non-continues aussi bien que non-différentiables. Enfin, on énonce en dernière

partie les principes de quelques méthodes usuelles pour le choix des points $(\mathbf{x}_i)_{1\leq i\leq n}$ où évaluer la fonction g. Cet ensemble de points est généralement appelé plan d'expériences. Il est construit dans l'objectif d'identifier à moindre coût les effets des paramètres fluctuants sur la performance du produit.

1.3.2 Résumé du chapitre 3

Dans ce chapitre, on suppose que la fonction g est observée en un nombre $n \in \mathbb{N}^*$ relativement faible de points. Pour mener l'estimation de p, le point de départ est une méthode bayésienne proposée dans (Auffray et al., 2014). Elle consiste à appliquer les principes de la méthode de krigeage, c'est-à-dire à modéliser la fonction g par un processus aléatoire ξ défini sur $\mathbb{X}$ et à valeurs dans $\mathbb{R}$, qui est ensuite conditionné aux évaluations de la fonction g aux points du plan d'expériences. Une prédiction S_n de la probabilité de défaillance, variable aléatoire à valeurs dans $[0,1]$, est alors définie à partir de la loi a posteriori du processus aléatoire. Cependant, on constate rapidement que la loi de S_n est inaccessible et qu'il est difficile de mesurer la dispersion de cette variable aléatoire, c'est-à-dire d'évaluer la pertinence de cette approche bayésienne. Néanmoins, l'espérance de S_n est facile à estimer par une méthode Monte-Carlo naïve et fournit une estimation de la probabilité de défaillance. Il a alors été proposé dans (Oger et al., 2015) une variable aléatoire alternative à S_n, de même valeur moyenne, mais dont la distribution empirique peut plus facilement être estimée. On note R_n cette variable aléatoire. Les travaux menés dans la première partie de ce chapitre consistent à comparer ces deux variables aléatoires afin de statuer quant à leurs avantages et leurs limites respectives. Notre principal résultat concerne l'existence d'une inégalité d'ordre convexe signifiant que, pour toute fonction convexe φ, on a :

$$\mathbb{E}[\varphi(S_n)] \leq \mathbb{E}[\varphi(R_n)].$$

Cette inégalité permet de comparer leur dispersion. Elle implique notamment que la variance de R_n est plus élevée que celle de S_n. Grâce à diverses propriétés relatives à l'ordre convexe que l'on présente en annexe A, on montre que l'on peut apprendre sur la loi de S_n à travers cette inégalité. Par exemple, on en déduit un encadrement des quantiles de S_n. Cela contribue à justifier l'intérêt de l'approche bayésienne pour l'estimation de p telle qu'elle est proposée dans la littérature.

La seconde partie de ce chapitre concerne la construction d'un plan d'expériences $(\mathbf{x}_i)_{1\leq i\leq n}$ où effectuer les évalutions de g. En se basant sur le principe des stratégies SUR proposées dans (Bect et al., 2012), ainsi que sur nos résultats concernant l'ordre convexe, on développe une méthode itérative pour sélectionner les points d'évaluation. Le critère d'échantillonnage que l'on propose repose sur une minimisation de la variance de R_n. Celui-ci est plus facile à calculer que le critère proposé dans (Bect et al., 2012), basé sur une minimisation de la variance de S_n. L'intérêt de cette approche est notamment mis en évidence à travers l'étude d'un cas réel de la société STMicroelectronics.

1.3.3 Résumé du chapitre 4

On souhaite estimer la probabilité de défaillance p dans le cas où sa valeur est très faible, c'est-à-dire dans le cas où l'évènement $\{\mathbf{X} \in \mathbb{F}\}$ est rare. Pour cela, on suppose pour

simplifier que $\mathbb{X} = [0,1]^d$, et on propose un algorithme qui procède par partitionnement dyadique récursif de $\mathbb{X}$. Il s'applique sous l'hypothèse que g est une fonction lipschitzienne de constante $L > 0$ connue. Cette hypothèse, combinée à un découpage dyadique récursif de $[0,1]^d$, est également proposée dans (Cohen et al., 2013) pour mettre en œuvre un algorithme de minimisation de fonction.

À chaque itération, l'ensemble $\mathbb{X}$ est écrit comme une partition de plus en plus fine de cubes dyadiques deux à deux disjoints. L'hypothèse sur g est alors utilisée pour classer les cubes dans trois catégories : ceux qui sont inclus dans le domaine de défaillance $\mathbb{F}$, ceux qui sont disjoints de $\mathbb{F}$, et les autres (i.e. les cubes incertains). Cette répartition est rendue possible grâce à l'évaluation de g aux centres de chacun des cubes. Ainsi, quand un cube est identifié comme inclus dans $\mathbb{F}$ ou disjoint de $\mathbb{F}$, il n'est plus utile d'y évaluer g, et l'algorithme continue à explorer uniquement les cubes au statut incertain.

À l'aide de ces informations, l'algorithme retourne à la fin de chaque itération deux approximations de p : la probabilité que $\mathbf{X}$ appartienne à la réunion des cubes identifiés par l'algorithme comme inclus dans $\mathbb{F}$ (approximation par défaut), et la probabilité que $\mathbf{X}$ appartienne à la réunion des cubes incertains et identifiés par l'algorithme comme inclus dans $\mathbb{F}$ (approximation par excès). Ces approximations sont estimées via une méthode Monte-Carlo dite de splitting, bien adaptée à l'estimation de probabilités d'évènements rares. Nous en présentons le principe au chapitre 2. Cette méthode nécessite la simulation d'échantillons $(\mathbf{X}_i)_{1 \leq i \leq N}$ suivant des restrictions de $\mathbf{P}_{\mathbf{X}}$ à des sous-ensembles de $\mathbb{X}$, que nous obtenons en pratique grâce à l'algorithme de Metropolis-Hastings.

La méthode de splitting peut être appliquée avec une grande valeur de N sans que cela nécessite d'appels supplémentaires à la fonction g. La précision de l'estimation que retourne l'algorithme dépend donc du nombre d'évaluations de la fonction g et de la taille N des échantillons utilisés pour estimer les probabilités.

Chapitre 2

Méthodes usuelles pour l'estimation d'une probabilité de défaillance

2.1 Introduction

Dans ce chapitre, on présente quelques méthodes usuelles pour l'estimation de la probabilité de défaillance p donnée en équation (1.1). La fonction g étant coûteuse à évaluer, si p est proche de 0, on verra que certaines méthodes sont, de prime abord, peu efficaces. C'est notamment le cas de la méthode Monte-Carlo naïve présentée en section 2.2. Puisqu'elle s'appuie sur des simulations de la variable aléatoire $g(\mathbf{X})$ – et que l'erreur d'estimation de cette méthode est d'autant plus faible que le nombre de simulations est grand – alors elle devient inopérante si la fonction g est chère en temps de calcul et la probabilité p petite. Alternativement, on présente deux méthodes classiques dites de « réduction de variance » : l'échantillonnage préférentiel (« Importance Sampling ») et la méthode de splitting (« Multilevel-Splitting »). Cette dernière est utilisée dans le chapitre 4.

Pour contourner les limites des méthodes Monte-Carlo, il est courant de remplacer la fonction g dans la définition de p par un modèle qui peut être évalué à moindre coût (généralement appelé métamodèle ou « surrogate model » en anglais). De cette façon, on peut appliquer une méthode Monte-Carlo avec un nombre suffisant de simulations pour assurer une erreur d'estimation relativement faible. Cette erreur dépend néanmoins de la capacité du modèle à prédire la valeur de g. En effet, il faut, idéalement, que le modèle sélectionné reproduise fidèlement le comportement de g en tout point de $\mathbb{X}$ ou, du moins, au voisinage de la frontière du domaine de défaillance $\mathbb{F}$. Le modèle peut être choisi déterministe (c'est une fonction définie sur $\mathbb{X}$) ou aléatoire (c'est un processus aléatoire indexé par $\mathbb{X}$). Dans les deux cas, il dépend de paramètres qui sont en pratique inconnus et généralement estimés à partir d'un ensemble d'observations de la fonction g. Celle-ci étant déterministe et coûteuse à évaluer, ces observations sont nécessairement réalisées en un nombre limité $n \in \mathbb{N}^*$ de points $(\mathbf{x}_i)_{1 \leq i \leq n} \in \mathbb{X}^n$. L'information étant rare, il est souhaitable que le modèle soit exact en ces points : on parle de modèle d'interpolation. En section 2.3, on présente la méthode de modélisation par un processus gaussien, couramment utilisée en industrie pour la modélisation de codes de calcul coûteux. Cette méthode intervient notamment tout au

long du chapitre 3.

Puisque les paramètres du modèle sont ajustés en fonction des observations, le choix des points $(\mathbf{x}_i)_{1\leq i\leq n}$ en lesquels évaluer la fonction g est primordial. En section 2.4, on présente quelques méthodes pour construire un plan d'expériences. On distingue deux approches : la première consiste à se donner un ensemble de points fixé à l'avance et à construire le modèle à partir des observations de g en ces points. On parle de « model-free design ». La seconde consiste à choisir les points de façon séquentielle, en tenant compte à chaque itération des valeurs de g déjà connues pour déterminer le prochain point d'évaluation. Les paramètres du modèle sont ajustés au fur et à mesure des appels à la fonction g. On parle de « model-based design ». Bien qu'on ne le présente pas comme tel, précisons que l'algorithme que l'on propose au chapitre 4 est une méthode de planification d'expériences séquentielle, adaptée au calcul d'une probabilité de défaillance.

2.2 Méthodes Monte-Carlo

Les méthodes Monte-Carlo sont des méthodes d'estimation qui reposent sur la possibilité de simuler des suites de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) suivant une loi donnée. Elles sont généralement utilisées pour estimer des intégrales s'écrivant sous la forme d'une espérance, comme par exemple :

$$\mathbb{E}\left[\varphi(\mathbf{X})\right] = \int_{\mathbb{X}} \varphi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \tag{2.1}$$

où $\varphi : \mathbb{X} \subseteq \mathbb{R}^d \to \mathbb{R}$ est une fonction donnée et $\mathbf{X}$ est une variable aléatoire de densité $f_{\mathbf{X}}$ suivant laquelle on sait simuler. Pour l'estimation de la probabilité de défaillance (1.1), si l'on suppose que la loi $\mathbf{P}_{\mathbf{X}}$ admet une densité $f_{\mathbf{X}}$ suivant laquelle on sait simuler, alors (2.1) s'écrit : $p = \mathbb{E}[\mathbb{1}_{g(\mathbf{X})>T}] = \int_{\mathbb{X}} \mathbb{1}_{g(\mathbf{x})>T} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$.

2.2.1 Méthode Monte-Carlo naïve

Principe

Soit $N \in \mathbb{N}^*$ et $(\mathbf{X}_i)_{1\leq i\leq N}$ une suite de variables aléatoires i.i.d. suivant $f_{\mathbf{X}}$. L'estimateur de $p = \int_{\mathbb{X}} \mathbb{1}_{g(\mathbf{X})>T} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$ par la méthode Monte-Carlo naïve est la variable aléatoire $\widehat{p}_N$ définie par :

$$\widehat{p}_N = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\mathbf{X}_i\in\mathbb{F}} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{g(\mathbf{X}_i)>T}. \tag{2.2}$$

Il s'agit simplement de la proportion de variables $(\mathbf{X}_i)_{1\leq i\leq N}$ qui sont dans le domaine de défaillance $\mathbb{F}$. Ceci est illustré figure 2.1. Dans cet exemple, la fonction g est définie sur $\mathbb{R}^2$ et à valeurs dans $\mathbb{R}$. La frontière de $\mathbb{F}$ est représentée par la ligne de niveau en pointillés bleus : c'est la ligne de niveau T de g. L'ensemble des points orange et rouges est un échantillon de taille $N = 10^4$ i.i.d. suivant une loi normale bidimensionnelle. L'estimateur Monte-Carlo est la proportion de points rouges, qui vaut ici $\frac{9}{10^4}$.

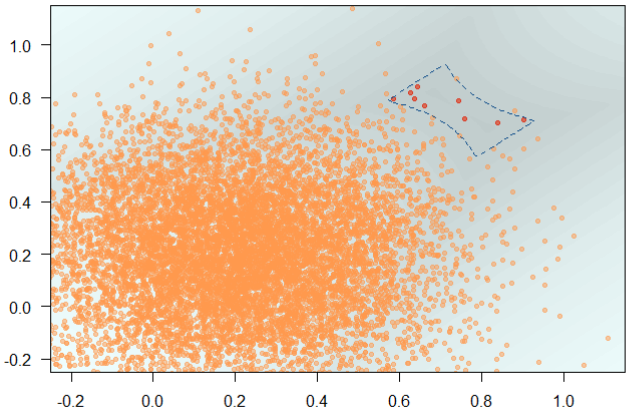

FIGURE 2.1 – Illustration du principe de la méthode Monte-Carlo pour une fonction $g : \mathbb{R}^2 \to \mathbb{R}$. La ligne en pointillés bleus est la ligne de niveau T de g, c'est-à-dire la frontière du domaine de défaillance $\mathbb{F} = \{\mathbf{x} \in \mathbb{R}^2 : g(\mathbf{x}) > T\}$. Les points orange et rouges sont générés suivant la densité $f_{\mathbf{X}}$ de $\mathbf{X}$. La probabilité $p = \mathbb{P}(g(\mathbf{X}) > T)$ est estimée par la proportion de points dans le domaine de défaillance (les points rouges).

Propriétés

Les variables aléatoires $(\mathbb{1}_{g(\mathbf{X}_i)>T})_{1\leq i\leq N}$ étant indépendantes et toutes distribuées suivant une loi de Bernoulli de paramètre p, l'estimateur $\widehat{p}_N$ donné en (2.2) est sans biais et sa variance est égale à $\frac{\sigma^2}{N}$, où $\sigma^2 = p(1-p)$. Par la loi forte des grands nombres, l'estimateur $\widehat{p}_N$ est consistant (fortement convergent) :

$$\widehat{p}_N \xrightarrow[N\to\infty]{p.s.} p. \tag{2.3}$$

En outre, le théorème central-limite s'applique et la vitesse de convergence de $\widehat{p}_N$ est en $\frac{1}{\sqrt{N}}$:

$$\sqrt{N}\big(\widehat{p}_N - p\big) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2). \tag{2.4}$$

Soit $\alpha \in (0,1)$ fixé. D'après (2.4), un intervalle de confiance de niveau asymptotique $1-\alpha$ pour p s'écrit :

$$\left[\widehat{p}_N - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{N}} \quad , \quad \widehat{p}_N + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{N}}\right], \tag{2.5}$$

où $\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$ est le quantile de niveau $1-\frac{\alpha}{2}$ de la loi normale centrée réduite. Puisque p est inconnue, le paramètre σ^2 l'est aussi et on l'estime naturellement par :

$$\widehat{\sigma}^2_N = \widehat{p}_N(1-\widehat{p}_N).$$

Cet estimateur est aussi consistant (d'après le résultat de convergence donné en équation (2.3) associé à la continuité de la fonction $x \in [0,1] \mapsto x(1-x)$). De ce fait, d'après (2.4) et le lemme de Slutsky, on a :

$$\sqrt{N}\,\frac{\widehat{p}_N - p}{\widehat{\sigma}_N} \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}(0,1),$$

de sorte que l'intervalle de confiance de niveau asymptotique $1-\alpha$ que l'on calcule en pratique est :

$$\left[\widehat{p}_N - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\widehat{\sigma}_N}{\sqrt{N}} \quad , \quad \widehat{p}_N + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\widehat{\sigma}_N}{\sqrt{N}}\right].$$

Limites

Lorsque la probabilité p que l'on cherche à estimer est très petite, on mesure généralement l'erreur d'estimation à travers le calcul de l'écart-type relatif, ou erreur relative. Ici, elle s'écrit :

$$\sqrt{\frac{\mathrm{Var}[\widehat{p}_N]}{p^2}} = \sqrt{\frac{\sigma^2}{Np^2}} = \sqrt{\frac{1-p}{Np}} \approx \sqrt{\frac{1}{Np}}.$$

Ainsi, si p est de l'ordre de 10^{-8} par exemple, et que l'on souhaite une erreur relative de l'ordre de 10%, alors il faut générer un échantillon de taille $N = 10^{10}$, puis évaluer g autant de fois pour fournir une estimation de p. Il est évident que sous la contrainte que cette fonction est coûteuse à évaluer, cette approche n'est pas envisageable.

La méthode Monte-Carlo naïve est facile à mettre en œuvre (si l'on sait simuler suivant la densité $f_{\mathbf{X}}$) mais son principal défaut est que le nombre de simulations influe significativement sur la qualité de l'estimation. Dans le cadre des évènements rares, cette méthode risque notamment de renvoyer une estimation égale à 0 si le nombre de simulations est insuffisant. Or, dans un contexte industriel, une sous-estimation de la probabilité de défaillance est une situation dangereuse.

Pour améliorer les performances de l'estimateur $\widehat{p}_N$ donné en (2.2), il existe des méthodes alternatives dites de « réduction de variance » qui ont pour but de réduire sa variance $\frac{\sigma^2}{N}$. Elles sont, par exemple, présentées dans (Calfisch, 1998). Les méthodes d'échantillonnage préférentiel, de stratification, de variables de contrôle ou encore de variables antithétiques agissent sur le facteur σ, tandis que les méthodes quasi Monte-Carlo améliorent la vitesse de convergence. Par exemple, l'utilisation d'une suite à discrépance faible pour la création d'un échantillon $(\mathbf{X}_i)_{1\leq i\leq N}$ conduit à une convergence de la méthode en $\mathcal{O}\big((\log N)^d/N\big)$. Dans le cadre des évènements rares, les deux principales méthodes de réduction de variance sont l'échantillonnage préférentiel et la méthode de splitting. Nous les présentons ci-dessous et la méthode de splitting sera notamment appliquée dans le chapitre 4.

2.2.2 Échantillonnage préférentiel

Lorsque la probabilité que l'on cherche à estimer est celle d'un évènement rare, les observations $(\mathbf{X}_i)_{1\leq i\leq N}$ ont peu de chance de tomber dans le domaine de défaillance $\mathbb{F}$, car il

s'agit d'un sous-domaine de $\mathbb{X}$ où la densité $f_{\mathbf{X}}$ est très faible. L'échantillonnage préférentiel (ou « Importance Sampling ») consiste alors à générer l'échantillon $(\mathbf{X}_i)_{1\leq i\leq N}$ suivant une autre loi que $\mathbf{P}_{\mathbf{X}}$ de façon à forcer l'apparition de l'évènement rare (dans le sens où la densité de cette autre loi prend des valeurs plus élevées que $f_{\mathbf{X}}$ dans le domaine de défaillance). La présentation ci-dessous suit le chapitre 3 de (Robert and Casella, 2004), mais aussi (Bucklew, 2004) et le chapitre 2 de (Rubino and Tuffin, 2009).

Soit h une densité sur $\mathbb{X}$ telle que $h(\mathbf{x}) = 0$ implique $\mathbb{1}_{g(\mathbf{x})>T} f_{\mathbf{X}}(\mathbf{x}) = 0$. Alors, on peut réécrire p de la façon suivante :

$$p = \int_{\mathbb{X}} \mathbb{1}_{g(\mathbf{x})>T} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{X}} \mathbb{1}_{g(\mathbf{x})>T} \frac{f_{\mathbf{X}}(\mathbf{x})}{h(\mathbf{x})} h(\mathbf{x}) d\mathbf{x} = \mathbb{E}\big[\mathbb{1}_{g(\mathbf{Y})>T} L(\mathbf{Y})\big],$$

où on a posé $L(\mathbf{y}) = \frac{f_{\mathbf{X}}(\mathbf{y})}{h(\mathbf{y})}$ et la densité de $\mathbf{Y}$ est h. Celle-ci est généralement appelée « densité instrumentale » ou encore « densité d'importance ». En simulant suivant cette densité – et non suivant la loi naturelle $\mathbf{P}_{\mathbf{X}}$ – on espère, d'une part, forcer l'apparition de l'évènement rare et, d'autre part, réduire la variance de l'estimateur.

Ainsi, si l'on sait simuler un échantillon $(\mathbf{Y}_i)_{1\leq i\leq N}$ i.i.d. suivant h et calculer le rapport $L(\mathbf{y})$ en tout point $\mathbf{y} \in \mathbb{X}$, alors l'estimateur Monte-Carlo par échantillonnage préférentiel prend la forme d'une moyenne pondérée des observations :

$$\widehat{p}_N = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{g(\mathbf{Y}_i)>T} L(\mathbf{Y}_i).$$

On vérifie facilement que cet estimateur est sans biais et que sa variance vaut :

$$\mathrm{Var}[\widehat{p}_N] = \frac{1}{N}\left[\int_{\mathbb{X}} \mathbb{1}_{g(\mathbf{y})>T} L^2(\mathbf{y}) h(\mathbf{y}) d\mathbf{y} - p^2\right] = \frac{1}{N}\left[\int_{\mathbb{X}} \mathbb{1}_{g(\mathbf{x})>T} \frac{f_{\mathbf{X}}^2(\mathbf{x})}{h(\mathbf{x})} d\mathbf{x} - p^2\right]. \quad (2.6)$$

La principale difficulté de la méthode d'échantillonnage préférentiel consiste à trouver une densité instrumentale qui rend cette variance inférieure à celle de l'estimateur de la méthode Monte-Carlo naïve (qui vaut $\frac{\sigma^2}{N} = \frac{p(1-p)}{N}$). Toutefois, on peut montrer (voir, par exemple, le chapitre 5 de (Asmussen and Glynn, 2007)) que la densité h^* qui annule la variance (2.6) est égale à :

$$h^*(\mathbf{y}) = \frac{\mathbb{1}_{g(\mathbf{y})>T} f_{\mathbf{X}}(\mathbf{y})}{\int_{\mathbb{X}} \mathbb{1}_{g(\mathbf{x})>T} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}} = \frac{\mathbb{1}_{g(\mathbf{y})>T} f_{\mathbf{X}}(\mathbf{y})}{p}, \quad \forall \mathbf{y} \in \mathbb{X}. \quad (2.7)$$

En prenant $h = h^*$, on a $\widehat{p}_N = p$, de sorte qu'une seule simulation suffit. Cependant, même si l'on savait simuler suivant h^*, le rapport $L(\mathbf{y}) = \frac{f_{\mathbf{X}}(\mathbf{y})}{h^*(\mathbf{y})}$ est hors de portée puisqu'il fait intervenir la quantité cherchée p.

La réussite de l'échantillonnage préférentiel par rapport à une méthode Monte-Carlo naïve repose donc sur le choix de la densité instrumentale h. Si celle-ci est suffisamment proche de la densité (2.7), alors on peut espérer que la variance de l'estimateur par échantillonnage préférentiel sera réduite. Plusieurs approches pour y parvenir sont proposées dans la littérature et de nombreux exemples d'applications peuvent être trouvés, par exemple,

dans (Asmussen and Glynn, 2007) ou encore (Rubino and Tuffin, 2009). Dans (Rubinstein, 1999), une approche itérative consiste à sélectionner, parmi une famille de fonctions paramétriques, la densité instrumentale minimisant la distance de Kullback-Leibler (appelée aussi « Cross-Entropy ») par rapport à la densité (2.7). Alternativement, des méthodes non-paramétriques d'estimation par noyau de (2.7) sont proposées, par exemple, dans (Ang et al., 1990) et (Au and Beck, 1999). Enfin, dans le cadre de l'estimation de probabilités de défaillance, des procédures itératives sont données dans (Melchers, 1990) et (Bucher, 1988).

2.2.3 Méthode de splitting

2.2.3.1 Principe

La seconde méthode classique de réduction de variance pour la simulation d'évènements rares est la méthode de splitting, aussi appelée « Multilevel Splitting » (voir (Kahn and Harris, 1951)), ou encore « Subset Simulation » dans le domaine de la fiabilité (voir (Au and Beck, 2001)). Cette méthode consiste à écrire l'évènement rare $\{\mathbf{X} \in \mathbb{F}\}$ en une suite de sous-évènements emboîtés dont les probabilités d'occurence sont suffisamment élevées pour être estimées facilement (voir, par exemple, (Kahn and Harris, 1951), (Rosenbluth and Rosenbluth, 1955), le chapitre 3 de (Rubino and Tuffin, 2009) et (Cérou et al., 2012)).

Soit $q \in \mathbb{N}^*$ fixé et $-\infty = T_0 < T_1 < \ldots < T_{q-1} < T_q = T$ une suite croissante de seuils intermédiaires. On pose :

$$\mathbb{F}_j = \{\mathbf{x} \in \mathbb{X} : g(\mathbf{x}) > T_j\} \quad \forall j = 0, \ldots, q,$$

de sorte que $\mathbb{F}_0 = \mathbb{X}$. L'évènement rare $\{\mathbf{X} \in \mathbb{F}\}$ s'écrit comme la suite de sous-évènements emboîtés suivante :

$$\{\mathbf{X} \in \mathbb{F}\} = \{\mathbf{X} \in \mathbb{F}_q\} \subset \{\mathbf{X} \in \mathbb{F}_{q-1}\} \subset \ldots \subset \{\mathbf{X} \in \mathbb{F}_1\} \subset \{\mathbf{X} \in \mathbb{F}_0\}.$$

Par application de la formule de Bayes, on en déduit une écriture de p sous la forme d'un produit de q probabilités conditionnelles :

$$p = \prod_{j=1}^{q} \mathbb{P}(\mathbf{X} \in \mathbb{F}_j \mid \mathbf{X} \in \mathbb{F}_{j-1}) = \prod_{j=1}^{q} p_j.$$

Supposons que $\forall j = 1, \ldots, q$, on dispose d'une suite de variables aléatoires$(\mathbf{X}_i^{j-1})_{1 \leq i \leq N}$ i.i.d. suivant la loi de $\mathbf{X}$ sachant $\mathbf{X} \in \mathbb{F}_{j-1}$. Alors, l'estimateur de p par la méthode de splitting est :

$$\widehat{p}_N = \prod_{j=1}^{q} \widehat{p}_j, \quad \text{où} \quad \widehat{p}_j = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\mathbf{X}_i^{j-1} \in \mathbb{F}_j} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{g(\mathbf{X}_i^{j-1}) > T_j}. \tag{2.8}$$

2.2.3.2 Exemple

Cette approche est illustrée à la figure 2.2. Dans cet exemple, la fonction g est définie sur $\mathbb{R}^2$ et à valeurs dans $\mathbb{R}$. La frontière du domaine de défaillance $\mathbb{F}$ est représentée par la ligne de niveau T de g (ligne en tirets bleus). En haut à gauche, l'échantillon (points orange et rouges) est simulé suivant la loi $\mathbf{P_X}$ (ici, une loi gaussienne multivariée). Puisqu'aucune observation ne tombe dans le domaine de défaillance $\mathbb{F}$, l'estimation de p par la méthode Monte-Carlo naïve est égale à 0. Pour résoudre ce problème, on se donne une suite croissante de seuils $-\infty = T_0 < T_1 < \ldots < T_4 = T$, et on pose :

$$\mathbb{F}_j = \{\mathbf{x} \in \mathbb{R}^2 : g(\mathbf{x}) > T_j\}, \quad \forall j = 0, \ldots 4,$$

avec $\mathbb{F}_0 = \mathbb{R}^2$ et $\mathbb{F}_4 = \mathbb{F}$ (les lignes en pointillés gris représentent les frontières des sous-domaines $\mathbb{F}_1$, $\mathbb{F}_2$ et $\mathbb{F}_3$). On peut alors écrire p de la façon suivante :

$$p = \prod_{j=1}^{4} \mathbb{P}(\mathbf{X} \in \mathbb{F}_j \mid \mathbf{X} \in \mathbb{F}_{j-1}).$$

On estime chaque probabilité conditionnelle par la méthode Monte-Carlo naïve. Comme on le voit sur chaque figure, on simule pour cela un échantillon i.i.d. suivant la restriction de $\mathbf{P_X}$ au sous-ensemble $\mathbb{F}_{j-1}$ (points orange et rouges) et on estime la probabilité conditionnelle $\mathbb{P}(\mathbf{X} \in \mathbb{F}_j \mid \mathbf{X} \in \mathbb{F}_{j-1})$ par la proportion d'observations au-dessus du seuil T_j (points rouges). Les estimations que l'on obtient sont non-nulles, de sorte que l'estimation de p qui en retourne est également non-nulle.

2.2.3.3 Propriétés

Pour étudier les propriétés de l'estimateur $\widehat{p}_N$ donné en (2.8), on fait l'hypothèse que pour tout $j = 1, \ldots, q$, on dispose d'un échantillon de variables aléatoires $(\mathbf{X}_i^{j-1})_{1 \leq i \leq N}$ i.i.d. suivant la loi de $\mathbf{X}$ sachant $\mathbf{X} \in \mathbb{F}_{j-1}$. On fait également l'hypothèse que ces échantillons sont deux à deux indépendants, de sorte que les estimateurs $(\widehat{p}_j)_{1 \leq j \leq q}$ sont indépendants.

Dans ces conditions, les estimateurs $(\widehat{p}_j)_{1 \leq j \leq q}$ vérifient un à un les résultats de la section 2.2.1. Ils sont sans biais et de variance :

$$\mathrm{Var}[\widehat{p}_j] = \frac{p_j(1-p_j)}{N}, \quad \forall j = 1, \ldots, q.$$

En outre, d'après le théorème central-limite, ils vérifient :

$$\sqrt{N}\big(\widehat{p}_j - p_j\big) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}\big(0, p_j(1-p_j)\big). \quad \forall j = 1, \ldots, q. \tag{2.9}$$

Par conséquent, on a le résultat de normalité asymptotique suivant pour $\widehat{p}_N$:

Proposition 2.2.1. *L'estimateur $\widehat{p}_N$ donné en (2.8) est sans biais et satisfait :*

$$\sqrt{N}\big(\widehat{p}_N - p\big) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}\left(0,\ p^2 \sum_{j=1}^{q} \frac{1-p_j}{p_j}\right). \tag{2.10}$$

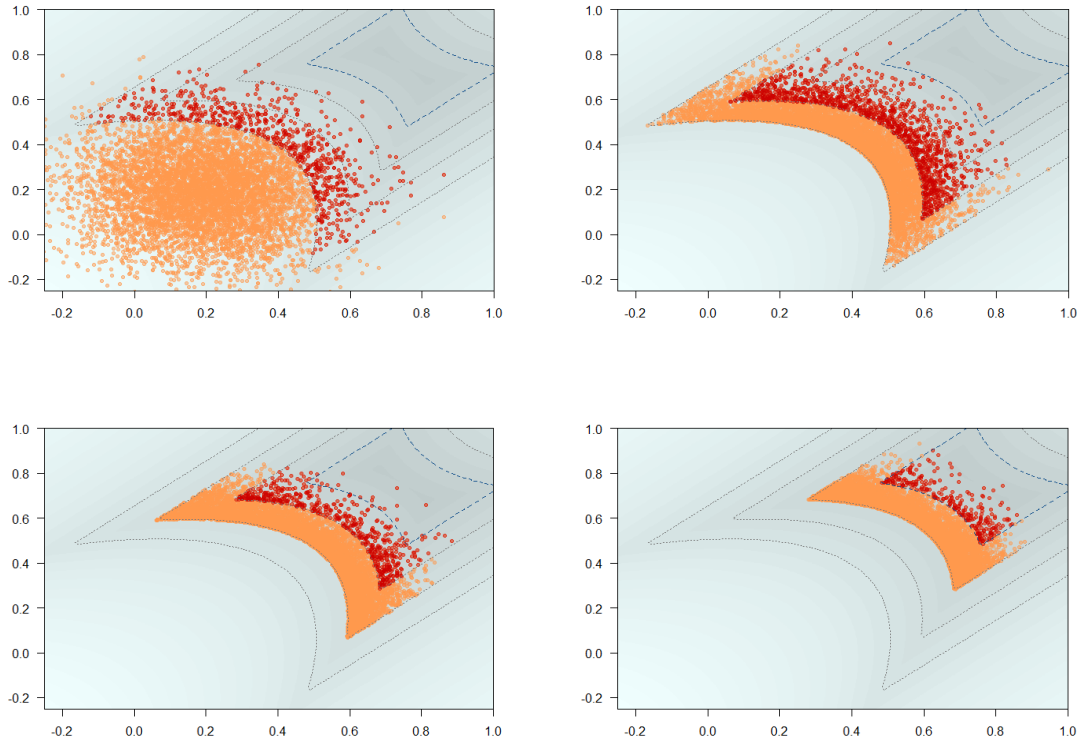

FIGURE 2.2 – Illustration du principe de la méthode de splitting. La frontière du domaine de défaillance $\mathbb{F}$ est la ligne de niveau en tirets bleus. Les frontières des domaines de défaillance intermédiaires sont les lignes de niveaux en pointillés gris. L'estimateur de p est égal au produit des proportions de points rouges. L'estimation de p qui en retourne est donc non-nulle, alors qu'une estimation directe par la méthode Monte-Carlo naïve est égale à 0 (voir figure en haut à gauche, où l'échantillon est simulé suivant la loi initiale et où aucun point ne tombe dans le domaine de défaillance).

Démonstration. On commence par appliquer la méthode delta au résultat de convergence donné en équation (2.9), avec la fonction $x \mapsto \log(x)$. On obtient :

$$\sqrt{N}\big(\log(\widehat{p}_j) - \log(p_j)\big) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}\left(0,\ \frac{1-p_j}{p_j}\right), \quad \forall j = 1, \dots, q. \tag{2.11}$$

Les estimateurs $(\widehat{p}_j)_{1\leq j\leq q}$ étant indépendants, en passant aux fonctions caractéristiques, le théorème de Paul Levy assure alors que :

$$\sqrt{N}\big(\log(\widehat{p}_N) - \log(p)\big) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}\left(0,\ \sum_{j=1}^{q} \frac{1-p_j}{p_j}\right).$$

Pour finir, on applique à nouveau la méthode delta en prenant la fonction $x \to \exp(x)$. On

obtient :

$$\sqrt{N}(\widehat{p}_N - p) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}\left(0,\ p^2 \sum_{j=1}^{q} \frac{1-p_j}{p_j}\right).$$

□

En remplaçant les probabilités p et $(p_j)_{1\leq j\leq q}$ par leurs estimations dans (2.10), on peut calculer des intervalles de confiance asymptotiques pour p. En pratique, il est difficile de satisfaire les hypothèses d'indépendance ci-dessus et la simulation suivant des lois conditionnelles n'est pas triviale. Toutefois, on verra au chapitre 4 comment on peut y parvenir de façon approchée. La procédure proposée est itérative et repose sur des applications successives de l'algorithme de Metropolis-Hastings (voir (Metropolis et al., 1953) et (Hastings, 1970)).

2.2.3.4 Remarques générales

Une des principales difficulté de la méthode de splitting réside dans le choix des seuils intermédiaires. En effet, si les probabilités conditionnelles $(p_j)_{1\leq j\leq q}$ sont trop grandes, alors il va falloir effectuer un nombre q d'itérations relativement élevé. Au contraire, si elles sont trop petites, alors l'estimation par la méthode Monte-Carlo naïve risque d'être inefficace.

Dans (Cérou et al., 2012), les auteurs expliquent que l'estimateur $\widehat{p}_N$ défini en (2.8) est de variance minimale si, pour tout $j = 1, \ldots, q$, la probabilité conditionnelle p_j s'écrit : $p_j = p^{1/q}$. Or, p étant la quantité inconnue que l'on cherche à estimer, il est évidemment impossible de déterminer les seuils qui permettent de vérifier cela. Cependant, ce résultat suggère qu'il faut, autant que possible, choisir les seuils de façon à égaliser les probabilités conditionnelles. C'est pourquoi il existe une version adaptative de la méthode où les seuils sont placés pas à pas durant la procédure d'estimation (voir (Au and Beck, 2001) et (Cérou et al., 2012)). Dans cette approche, le choix des seuils dépend des échantillons simulés. En effet, le principe est de fixer à l'avance une proportion $p_0 \in (0, 1)$, et, à chaque itération, de placer le seuil de façon à avoir Np_0 observations qui le dépassent. L'estimateur qui en découle s'écrit $\widehat{p}_N = \widehat{r}_0 p_0^{\widehat{n}_0}$, où $\widehat{r}_0$ est la proportion d'observations qui dépassent le seuil T à la dernière itération et $\widehat{n}_0$ est le nombre d'itérations, qui est nécessairement aléatoire. Le cas particulier où $p_0 = 1 - \frac{1}{N}$, c'est-à-dire où $N - 1$ observations sont conservées à chaque itération, est étudié dans (Guyader and Hengartner, 2011).

Dans notre contexte, cette méthode d'estimation reste difficile à mettre en œuvre car elle nécessite à chaque itération d'effectuer un grand nombre d'appels à la fonction g. Toutefois, des applications récentes de cette méthode (qui, on le rappelle, porte davantage le nom de « Subset Simulation » dans le domaine de la fiabilité), combinée à la modélisation par processus gaussiens (que l'on présente en section 2.3.3), sont par exemple proposées dans (Xiaoxu et al., 2016), (Bect et al., 2017) et (Walter, 2017). En outre, on présente au chapitre 4 un algorithme pour l'estimation de p qui fait appel à la méthode de splitting pour estimer des probabilités de la forme $\mathbb{P}(\mathbf{X} \in Q)$, où Q est un sous-ensemble de $\mathbb{X}$ de mesure potentiellement très petite pour $\mathbf{P}_{\mathbf{X}}$.

2.3 Méthodes d'approximation

Les méthodes Monte-Carlo que nous venons de présenter sont efficaces si le nombre de simulations de la variable aléatoire $g(\mathbf{X})$ est suffisamment grand. La contrainte que g est une fonction coûteuse à évaluer rend ces méthodes difficiles à appliquer dans notre contexte. On peut alors envisager deux solutions pour contourner ce problème.

La première solution consiste à identifier une approximation $\widehat{\mathbb{F}}$ du domaine de défaillance $\mathbb{F}$ grâce à des évaluations de la fonction g en des points de $\mathbb{X}$ bien choisis. Ensuite, on applique une méthode Monte-Carlo pour estimer la probabilité $\mathbb{P}(\mathbf{X} \in \widehat{\mathbb{F}})$, sans que cela requiert d'appels supplémentaires à la fonction g. C'est le principe des méthodes FORM et SORM, couramment utilisées en fiabilité, que l'on présente brièvement en section 2.3.1. C'est aussi exactement le principe de l'algorithme que l'on développe dans le chapitre 4.

La seconde solution consiste à construire une approximation $\widehat{g}$ de la fonction g (i.e. un modèle, parfois aussi appelé métamodèle), qui est moins chère à évaluer. Pour construire une bonne approximation $\widehat{g}$, on a besoin de connaître des valeurs vraies de g en un certain nombre de points. L'estimation de la probabilité $\mathbb{P}(\widehat{g}(\mathbf{X}) > T)$ par une méthode Monte-Carlo ne nécessite, à nouveau, aucune évaluation supplémentaire de g. Dans le chapitre 3, on utilise la méthode de modélisation par processus gaussiens, appelée aussi krigeage. Celle-ci est présentée en détails en section 2.3.3.

2.3.1 Approximation du domaine de défaillance

On présente ici les méthodes FORM et SORM (« First and Second Order Reliability Methods ») couramment utilisées en fiabilité (voir (Cizeli et al., 1994) et les références qui s'y trouvent). Elles reposent sur le fait que la probabilité p vérifie :

$$p = \int_{\mathbb{X}} \mathbb{1}_{g(\mathbf{x})>T} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) = \int_{\mathbb{X}} \mathbb{1}_{l(\mathbf{x})<0} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}),$$

où $l = T - g$ est appelée fonction d'état limite (« limit state function »). La première étape consiste à transformer la loi de $\mathbf{X}$ en une loi normale centrée réduite, de façon à obtenir un vecteur aléatoire gaussien $\mathbf{Z}$ de même dimension que $\mathbf{X}$, mais dont les composantes sont indépendantes, centrées et réduites. Cela s'effectue à l'aide d'un opérateur de transformation noté ici τ (on parle de transformation isoprobabiliste dans l'espace standard gaussien). Pour plus de détails, voir (Rosenblatt, 1952) et (Nataf, 1962), ainsi que la thèse (Lebrun, 2013). Une fois la transformation effectuée, la probabilité p se réécrit :

$$p = \mathbb{P}(L(\mathbf{Z}) < 0) = \int_{\mathbb{R}^d} \mathbb{1}_{L(\mathbf{z})<0} \times \frac{1}{(2\pi)^{\frac{d}{2}}} \exp\left(-\frac{\|\mathbf{z}\|^2}{2}\right) d\mathbf{z},$$

où $L = l \circ \tau^{-1}$ est la fonction d'état limite transformée et $\|\cdot\|$ est la norme euclidienne sur $\mathbb{R}^d$. On cherche ensuite le point $\mathbf{z}^* \in \mathbb{R}^d$, appelé « point de conception » (ou aussi, « most probable point » en anglais), qui a la plus grande probabilité d'appartenir à l'hypersurface d'équation $L(\mathbf{z}) = 0$ (l'hypersurface d'état limite). Puisque $\mathbf{Z}$ est un vecteur gaussien centré et que la densité de probabilité au point z^* doit être maximale, alors on cherche le point

de l'hypersurface qui est le moins éloigné de l'origine, et supposé ne pas appartenir au domaine $\{\mathbf{z} \in \mathbb{R}^d : L(\mathbf{z}) < 0\}$. Cela revient à résoudre le problème d'optimisation suivant :

$$\mathbf{z}^* = \underset{\mathbf{z}\in\mathbb{R}^d}{\operatorname{argmin}} \|\mathbf{z}\| \quad \text{sous la contrainte} \quad L(\mathbf{z}) = 0. \tag{2.12}$$

Cette étape est cruciale car on doit utiliser un algorithme d'optimisation suffisamment performant pour minimiser le nombre d'appels à la fonction L, dépendant de g. Des propositions d'algorithmes sont, par exemple, données dans (Hasofer and Lind, 1974), (Rackwitz and Flessler, 1978), (Zhang and Der Kiureghian, 1995) et (Ditlevsen and Madsen, 1996). La distance $\beta = \|\mathbf{z}^*\|$ est appelée « indice de fiabilité ». Dans le cas de la méthode FORM, l'hypersurface d'état limite est approchée par un hyperplan tangent passant par $\mathbf{z}^*$. Cela revient à approcher la probabilité de défaillance p par $\mathbb{P}(\mathcal{N}(0,1) > \beta) = 1 - \Phi(\beta)$, où Φ est la fonction de répartition de la loi normale centrée réduite. Concernant la méthode SORM, l'approximation est faite par une hypersurface quadratique tangente, passant également par $\mathbf{z}^*$ et conduisant à une approximation de p plus complexe (voir (Ditlevsen and Madsen, 1996)).

En pratique, il se peut que le domaine de défaillance $\mathbb{F}$ soit la réunion de sous-domaines de $\mathbb{X}$ et qu'il existe par conséquent plusieurs points de conception. L'approche décrite ci-dessus n'est alors pas adaptée car elle suppose qu'il existe une unique solution au problème d'optimisation (2.12). Il s'agit là d'un inconvénient majeur des méthodes FORM et SORM. En outre, puisque la qualité de l'approximation dépend essentiellement du point z^* que retourne l'algorithme d'optimisation, la capacité de ce dernier à identifier le minimum global est déterminante. Enfin, on précise que l'approximation de l'état limite par un hyperplan ou une hypersurface quadratique peut conduire à une sur-estimation ou une sous-estimation de la vraie probabilité, et qu'il n'existe pas de mesure de l'erreur d'approximation pour ces méthodes.

2.3.2 Modèles d'approximation d'une fonction inconnue

Dans le cadre des méthodes FORM et SORM, le choix des points où évaluer g dépend de l'algorithme d'optimisation choisi pour résoudre le problème (2.12). Ici, on se place dans le cas où la fonction g est déjà observée en une suite de points $\mathbf{D}_n = (\mathbf{x}_i)_{1\leq i\leq n}$ fixés. On parle de plan d'expériences pour désigner cet ensemble de points (« design of experiments » en anglais). En section 2.4, on verra les méthodes courantes pour construire un plan d'expériences.

Pour construire un modèle d'approximation de la fonction g, on présente dans ce qui suit la méthode de modélisation par processus gaussiens, car c'est l'unique méthode que nous utilisons dans cette thèse. Cependant, on précise qu'il existe plusieurs méthodes alternatives. Par exemple, la modélisation par des fonctions « splines », qui consiste à découper l'ensemble $\mathbb{X}$ et à considérer un modèle polynomial pour chacun des sous-espaces (pour une présentation générale, voir les livres (Hastie et al., 2001) et (Györfi et al., 2002)). On peut également considérer la modélisation par des réseaux de neurones (voir (Hastie et al., 2001) et (Bishop, 2006)). Dans le chapitre 6 de (Rasmussen and Williams, 2006), les auteurs comparent les modèles de processus gaussiens à ces solutions alternatives.

2.3.3 Krigeage ou modélisation par processus gaussiens

2.3.3.1 Introduction

Introduit par (Krige, 1951), puis formalisé par (Matheron, 1963), le krigeage a initialement été développé en géostatistique pour la prédiction de fonctions par interpolation de données spatiales. Celles-ci sont interprétées comme le résultat d'une réalisation d'un processus aléatoire spatial et utilisées pour construire un modèle d'interpolation (voir (Stein, 1999), (Diggle and Ribeiro, 2007) ou encore (Chiles and Delfiner, 2009)). Par la suite, cette approche a été adaptée à la modélisation de fonctions inconnues dans le cadre des expériences simulées (voir (Sacks et al., 1989a), (Sacks et al., 1989b), ou encore (Santner et al., 2003)) et les domaines d'application se sont diversifiés : machine learning ((Rasmussen and Williams, 2006)), optimisation ((Jones et al., 1998) et (Emmerich et al., 2006)), analyse de sensibilité ((Oakley and O'Hagan, 2004) et (Cannamela et al., 2014)) ou encore fiabilité ((Kaymaz, 2004), (Bichon et al., 2008), (Echard et al., 2011) et (Dubourg et al., 2013)).

La plupart du temps, le modèle choisi est un processus aléatoire gaussien car celui-ci est l'un des rares pour lequel on sache mener des calculs analytiques et qui relève d'une bonne adéquation avec un certain nombre de phénomènes physiques observés. Dans ce qui suit, on parlera alors indifféremment de krigeage ou de modélisation par processus gaussien, bien qu'il s'agisse d'un abus de langage.

2.3.3.2 Principe

Dans tout ce qui suit, on suppose que la fonction g est connue aux points d'un plan d'expériences $\mathbf{D}_n = (\mathbf{x}_i)_{1\leq i\leq n}$ fixé. On note $\mathbf{g}_n = (g(\mathbf{x}_i))_{1\leq i\leq n}$ le vecteur des valeurs observées.

Modélisation

Le principe du krigeage est de modéliser la fonction g par un processus aléatoire indexé par $\mathbb{X}$ et à valeurs dans $\mathbb{R}$. Les observations $\mathbf{g}_n$ sont vues comme de l'information sur une trajectoire de ce processus. Notons ξ ce processus. En le choisissant gaussien, le modèle que l'on considère s'écrit (voir, par exemple, la section 4 de (Sacks et al., 1989a), ainsi que (Rasmussen and Williams, 2006)) :

$$\xi(\mathbf{x}) = \sum_{i=1}^{q} m_i(\mathbf{x})\beta_i + \zeta(\mathbf{x}) = \mathbf{m}(\mathbf{x})^T\boldsymbol{\beta} + \zeta(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{X}, \tag{2.13}$$

où $\mathbf{m}(\mathbf{x}) = (m_i(\mathbf{x}))_{1\leq i\leq q}$ est un vecteur de fonctions fixées, $\boldsymbol{\beta} = (\beta_i)_{1\leq i\leq q}$ est un vecteur de paramètres, et ζ est un processus gaussien centré caractérisé par sa fonction de covariance :

$$\mathrm{Cov}\big(\xi(\mathbf{x}), \xi(\mathbf{x}')\big) = \sigma^2 k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}'), \quad \forall \mathbf{x}, \mathbf{x}' \in \mathbb{X}^2,$$

avec σ^2 un paramètre de variance et $k_{\boldsymbol{\theta}}$ une fonction de corrélation dépendant d'un paramètre $\boldsymbol{\theta} \in \mathbb{R}^r$, $r > 0$. On suppose ici que la variance σ^2 est égale en tout point de $\mathbb{X}$ (i.e. $k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}) = 1$). Ce choix de modélisation peut se justifier de la façon suivante. Dans le

modèle (2.13), le second terme du membre de droite est aléatoire. Il peut être interprété comme un terme d'erreur destiné à apporter de la « souplesse » à la composante déterministe $\mathbf{m}(\mathbf{x})^T\boldsymbol{\beta}$. On considère que cette erreur est nulle en moyenne. En outre, puisqu'il n'y a aucune raison de supposer que cette erreur est plus grande dans certaines régions de $\mathbb{X}$ que dans d'autres, on considère que la variance σ^2 est constante en tout point. Cet argument suggère notamment de choisir la fonction de covariance invariante par translation dans $\mathbb{X}$. Dans ce cas, cela revient à supposer que le processus ζ est stationnaire (au sens faible) et que la covariance s'exprime en fonction de $\mathbf{x} - \mathbf{x}'$ (voir le chapitre 4 de (Rasmussen and Williams, 2006)). Le choix de la fonction de covariance sera discuté en section 2.3.3.5.

La composante déterministe $\mathbf{m}(\cdot)^T\boldsymbol{\beta}$ dans (2.13) caractérise la tendance du processus ξ. Par exemple, on dit que la tendance est linéaire lorsque $\mathbf{m}(\mathbf{x}) = (1, \mathbf{x})$. En pratique, on choisit arbitrairement ces fonctions compte tenu de la connaissance que l'on a, a priori, de la fonction g ou de la complexité souhaitée du modèle. Le vecteur de paramètres $\boldsymbol{\beta}$ donne un poids à chacune de ces fonctions. Il est inconnu en pratique. On peut soit le fixer, soit l'estimer au vu des observations $\mathbf{g}_n$ en utilisant des méthodes qui sont décrites en section 2.3.3.3. La fonction de corrélation paramétrique $k_{\boldsymbol{\theta}}$ caractérise la dépendance au sein du processus ξ. En pratique, elle est choisie arbitrairement parmi une classe de fonctions que nous présenterons en section 2.3.3.5. Dans le modèle (2.13), les paramètres σ et $\boldsymbol{\theta}$ sont inconnus et peuvent être estimés à partir des observations $\mathbf{g}_n$ (voir section 2.3.3.3). Pour fixer les idées, on donne ci-dessous l'exemple de la fonction de corrélation k_θ, avec $\theta \in \mathbb{R}$, appelée exponentielle quadratique :

$$\mathrm{Cov}\big(\xi(\mathbf{x}), \xi(\mathbf{x}')\big) = \sigma^2 k_\theta(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{1}{2}\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{\theta^2}\right), \quad \forall (\mathbf{x}, \mathbf{x}') \in \mathbb{X}^2, \tag{2.14}$$

où $\|\cdot\|$ est la norme euclidienne sur $\mathbb{X}$. Sur la figure 2.3, on considère une fonction linéaire $g : [0,1] \to \mathbb{R}$ et on montre l'influence du choix des paramètres $(\sigma^2, \theta) \in \mathbb{R}^2$ dans la modélisation par un processus gaussien. Ici, la tendance est linéaire et le même vecteur de paramètres $\boldsymbol{\beta} = (1, 2)$ est fixé pour chacun des quatre modèles. La fonction de covariance est celle donnée en (2.14). C'est une fonction croissante de θ. On constate que, selon les valeurs prises par σ^2 et θ, le modèle est plus ou moins représentatif de la vraie fonction. Pour une grande valeur de θ (à gauche), les trajectoires du processus gaussien sont beaucoup plus régulières que pour une petite valeur de θ (à droite). Le paramètre σ^2 caractérisant la variance des variables aléatoires $(\xi(\mathbf{x}))_{\mathbf{x}\in\mathbb{X}}$, on voit que plus sa valeur est grande, plus les trajectoires du processus sont dispersées (en bas).

Remarque 2.3.1. *Lorsque $m(\mathbf{x}) = 0$, on parle de krigeage « simple ». Lorsque $m(\mathbf{x}) = 1$, on parle de krigeage « ordinaire » et, enfin, lorsque $m(\mathbf{x})$ a une forme plus générale, on parle de krigeage « universel ».*

Conditionnement

Notons $\boldsymbol{\xi}_n$ le vecteur aléatoire $(\xi(\mathbf{x}_i))_{1\leq i\leq n}$. Le modèle (2.13) étant donné, celui-ci doit à présent tenir compte du fait qu'une réalisation de $\boldsymbol{\xi}_n$ est observée et vaut $\mathbf{g}_n$. Pour se faire, on va conditionner la loi du processus aléatoire ξ par rapport à la donnée que $\boldsymbol{\xi}_n = \mathbf{g}_n$. Il

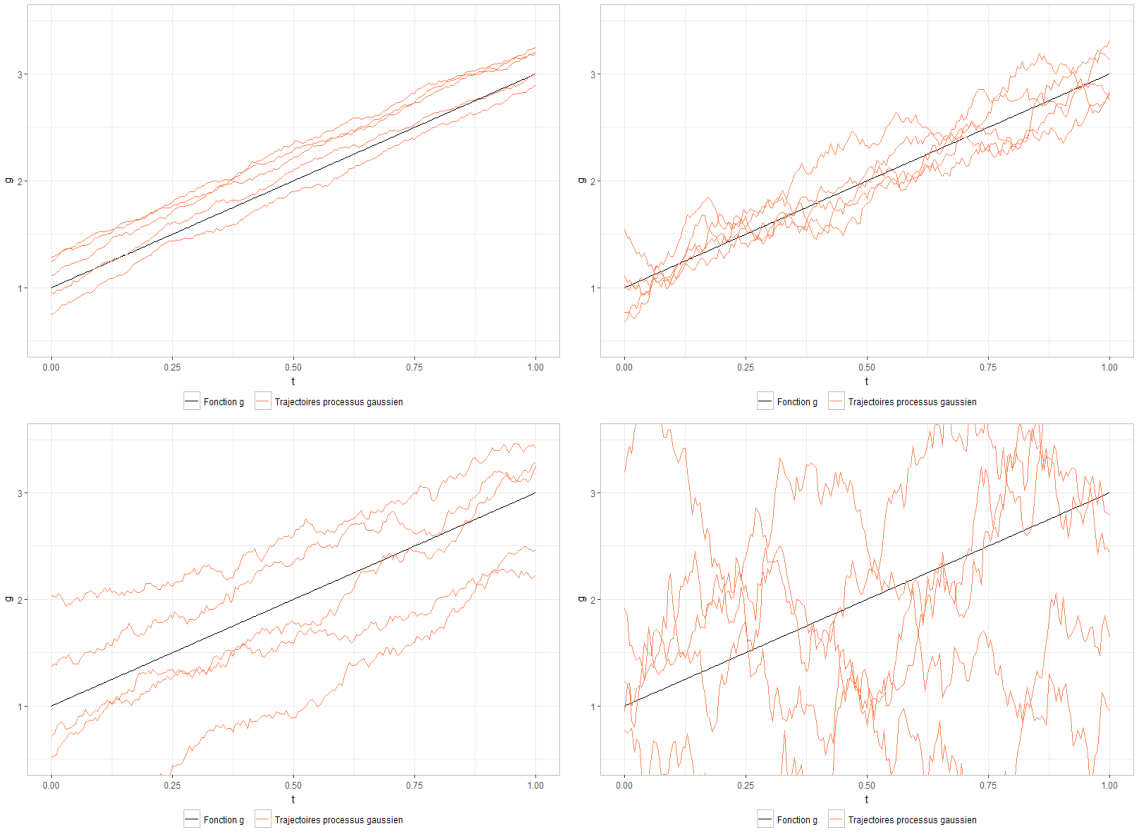

FIGURE 2.3 – Trajectoires d'un processus gaussien indexé par $[0,1]$. La tendance du processus est linéaire, avec $\boldsymbol{\beta} = (1,2)$. La fonction de covariance est donnée en (2.14), avec $(\sigma^2, \theta) = \left(\frac{1}{10}, 10\right)$ en haut à gauche, $(\sigma^2, \theta) = \left(\frac{1}{10}, \frac{1}{2}\right)$ en haut à droite, $(\sigma^2, \theta) = (1, 10)$ en bas à gauche, et $(\sigma^2, \theta) = \left(1, \frac{1}{2}\right)$ en bas à droite.

est facile de vérifier que, d'après le modèle (2.13), on a :

$$\xi(\mathbf{x}) \sim \mathcal{N}(\mathbf{m}(\mathbf{x})^T \boldsymbol{\beta}, \sigma^2), \quad \forall \mathbf{x} \in \mathbb{X}.$$

Dans la proposition suivante, on montre que lorsque l'on conditionne cette loi par rapport à $\boldsymbol{\xi}_n = \mathbf{g}_n$, on obtient toujours une loi gaussienne. Dans tout ce qui suit, nous nous plaçons dans des situations génériques où les observations conduisent à des matrices de covariance inversibles (i.e. définies positives).

Proposition 2.3.1. *Sous les hypothèses du modèle (2.13), avec les paramètres $\boldsymbol{\beta}$, σ et $\boldsymbol{\theta}$ fixés, la loi conditionnelle de $\xi(\mathbf{x})$ sachant $\boldsymbol{\xi}_n = \boldsymbol{g}_n$ est la loi normale :*

$$\mathcal{N}\big(m_n(\mathbf{x}), \sigma_n^2(\mathbf{x})\big), \tag{2.15}$$

où,

$$m_n(\mathbf{x}) = \mathbb{E}\left[\xi(\mathbf{x}) \mid \boldsymbol{\xi}_n = \mathbf{g}_n\right] = \mathbf{m}(\mathbf{x})^T \boldsymbol{\beta} + \boldsymbol{\Sigma}_{n,\mathbf{x},\boldsymbol{\theta}}^T \boldsymbol{\Sigma}_{n,\boldsymbol{\theta}}^{-1} \big(\mathbf{g}_n - \mathbf{M}_n \boldsymbol{\beta}\big), \tag{2.16}$$

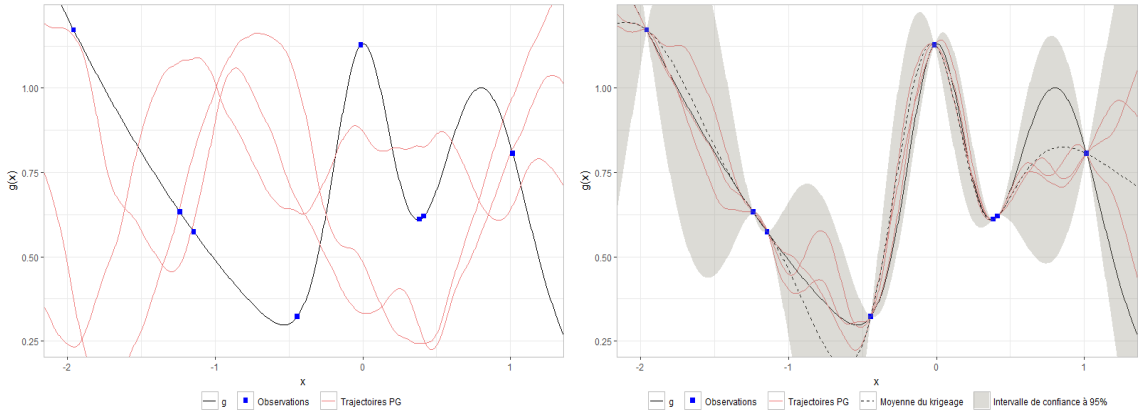

FIGURE 2.4 – Exemple de modélisation par processus gaussien. Courbe noire : fonction g que l'on cherche à modéliser. Points bleus : valeurs observées de g. À gauche : trajectoires du processus gaussien distribué suivant la loi du modèle (2.13). À droite : trajectoires du processus gaussien distribué suivant la loi conditionnée aux observations, donnée en proposition 2.3.1. La courbe noire en pointillés est la moyenne (2.16) et les surfaces grises sont des intervalles de confiance à 95%.

et

$$\sigma_n^2(\mathbf{x}) = \mathrm{Var}\left[\xi(\mathbf{x}) \,|\, \boldsymbol{\xi}_n = \mathbf{g}_n\right] = \sigma^2\big(1 - \boldsymbol{\Sigma}_{n,\mathbf{x},\boldsymbol{\theta}}^T \boldsymbol{\Sigma}_{n,\boldsymbol{\theta}}{}^{-1} \boldsymbol{\Sigma}_{n,\mathbf{x},\boldsymbol{\theta}}\big), \tag{2.17}$$

avec :

$$\mathbf{M}_n = (\mathbf{m}(\mathbf{x}_1), \ldots, \mathbf{m}(\mathbf{x}_n))^T, \quad \boldsymbol{\Sigma}_{n,\mathbf{x},\boldsymbol{\theta}} = \big(k_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x})\big)_{1\leq i\leq n}^T, \quad \text{et} \quad \boldsymbol{\Sigma}_{n,\boldsymbol{\theta}} = \big(k_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j)\big)_{1\leq i,j\leq n}.$$

Une démonstration de cette proposition peut être retrouvée, par exemple, à la section 1.2 de (Oger, 2014). On remarque que pour tout $\mathbf{x} \in \mathbb{X}$, la variance (2.17) vérifie : $\sigma_n^2(\mathbf{x}) \leq \sigma^2$. Cela signifie qu'en conditionnant la loi du processus ξ aux observations, on réduit l'incertitude sur les variables $(\xi(\mathbf{x}))_{\mathbf{x}\in\mathbb{X}}$. Cela se voit immédiatement sur la figure 2.4, où est illustré l'intérêt du conditionnement. À gauche, la loi du processus gaussien ne tient pas compte des observations (les points bleus) et les trajectoires (courbes rouges) ne sont pas vraiment représentatives de la fonction g. À droite, la loi est conditionnée aux observations et la modélisation est bien meilleure. Les aires grises sont des intervalles de confiance à 95%, construits à partir de (2.15). Ils s'écrivent sous la forme :

$$\left[m_n(\mathbf{x}) - 1.96\sigma_n(\mathbf{x}) \quad ; \quad m_n(\mathbf{x}) + 1.96\sigma_n(\mathbf{x})\right], \quad \forall \mathbf{x} \in \mathbb{X},$$

où fonctions m_n et σ_n sont définies en (2.16) et (2.17). Celles-ci vérifient :

$$m_n(\mathbf{x}_i) = g(\mathbf{x}_i) \quad \text{et} \quad \sigma_n^2(\mathbf{x}_i) = 0, \quad \forall i = 1 \ldots, n.$$

Cela signifie que n'importe quel processus distribué suivant la loi (2.15) a ses trajectoires qui interpolent les points d'observations, et peut être vu comme un modèle aléatoire d'interpolation de g (cela se voit à droite de la figure 2.4). Alternativement, la fonction m_n

donnée en (2.16) est souvent utilisée comme modèle déterministe d'approximation de la fonction g. C'est une fonction qui interpole g aux points du plan d'expériences $\mathbf{D}_n$ et qui est relativement peu coûteuse à évaluer. Elle peut donc être utilisée pour prédire ou approcher la valeur de g en tout point de $\mathbb{X}$. Par ailleurs, pour désigner l'espérance conditionnelle $\mathbb{E}\left[\xi(\mathbf{x}) \mid \boldsymbol{\xi}_n\right]$, on rencontre parfois le terme de « meilleur prédicteur linéaire » (BLP pour « Best Linear Predictor »). En effet, comme mentionné par exemple dans (Sacks et al., 1989b) et (Stein, 1999), on a :

Proposition 2.3.2. *Sous les hypothèses du modèle (2.13), avec les paramètres $\boldsymbol{\beta}$, σ et $\boldsymbol{\theta}$ fixés, l'estimateur $\widehat{\xi}_n(\mathbf{x})$ défini par :*

$$\widehat{\xi}_n(\mathbf{x}) = \mathbb{E}\left[\xi(\mathbf{x}) \mid \boldsymbol{\xi}_n\right] = \mathbf{m}(\mathbf{x})^T\boldsymbol{\beta} + \boldsymbol{\Sigma}_{n,\mathbf{x},\boldsymbol{\theta}}^T\boldsymbol{\Sigma}_{n,\boldsymbol{\theta}}^{-1}\left(\boldsymbol{\xi}_n - \mathbf{M}_n\boldsymbol{\beta}\right), \tag{2.18}$$

est le meilleur prédicteur linéaire de $\xi(\mathbf{x})$, c'est-à-dire qu'il vérifie :

$$\forall \mathbf{x} \in \mathbb{X}, \quad \widehat{\xi}_n(\mathbf{x}) = \underset{\tilde{\xi}_n(\mathbf{x})}{\operatorname{argmin}}\, \mathbb{E}\Big[\left(\xi(\mathbf{x}) - \tilde{\xi}_n(\mathbf{x})\right)^2 \mid \boldsymbol{\xi}_n\Big],$$

avec :

$$\tilde{\xi}_n(\mathbf{x}) = \lambda(\mathbf{x}) + \boldsymbol{\lambda}^T\boldsymbol{\xi}_n,$$

où $\lambda(\mathbf{x}) \in \mathbb{R}$ et $\boldsymbol{\lambda} \in \mathbb{R}^n$. Il est sans biais et son erreur en moyenne quadratique est égale à la variance $\sigma_n^2(\mathbf{x})$ donnée en (2.17).

Une démonstration de cette proposition est, par exemple, détaillée au chapitre 2 de la thèse (Bettinger, 2009). Le principe est de montrer que, sous les hypothèses du modèle (2.13), la construction d'un estimateur vérifiant les contraintes de linéarité, de non-biais et d'erreur en moyenne quadratique minimale, conduit à l'espérance conditionnelle (2.18). On parle aussi d'estimateur BLUP pour « Best Linear Unbiased Predictor ». La proposition 2.3.2 mentionne notamment que, pour tout $\mathbf{x} \in \mathbb{X}$, la variance $\sigma_n^2(\mathbf{x})$ donnée en (2.17) est une mesure de l'erreur d'approximation de $\xi(\mathbf{x})$ par (2.18) au sens de l'erreur en moyenne quadratique. C'est pourquoi, comme nous le verrons en section 2.4.2, cette variance peut être utilisée pour la construction de plans d'expériences.

Il n'existe pas à notre connaissance de mesure de l'erreur d'approximation de g par la fonction $m_n = \mathbb{E}\left[\xi(\cdot) \mid \boldsymbol{\xi}_n = \mathbf{g}_n\right]$. Par exemple, pour statuer quant à la pertinence du modèle, il serait souhaitable d'avoir accès à une borne de l'erreur en norme infinie, qui est définie par : $\|m_n - g\|_\infty = \sup_{\mathbf{x}\in\mathbb{X}} |m_n(\mathbf{x}) - g(\mathbf{x})|$, ou de l'erreur L_p, qui est définie par : $\int_{\mathbb{X}} |m_n(\mathbf{x}) - g(\mathbf{x})|^p \mathbf{P}_{\mathbf{X}}(d\mathbf{x})$, avec $p \in \mathbb{N}^*$ (voir (Györfi et al., 2002)). En pratique, on peut néanmoins utiliser une méthode de validation croisée pour se donner une idée de la capacité de prédiction du modèle. Cette méthode est brièvement présentée en section 2.3.3.3.

Approche bayésienne

La procédure de modélisation ci-dessus peut être décrite dans un cadre bayésien, avec les notions de lois a priori et a posteriori (pour une vue d'ensemble sur les méthodes bayésiennes, on pourra consulter le livre (Robert, 2007)). En effet, le modèle (2.13) peut

être vu comme un choix de loi a priori pour ξ, tandis que la loi conditionnelle obtenue à la proposition 2.3.1 peut être vue comme la loi a posteriori sachant les observations. On verra en section 2.3.3.4 que l'approche bayésienne peut aussi être utilisée pour fournir une estimation des paramètres du modèle (2.13).

2.3.3.3 Estimation des paramètres

En pratique, les paramètres $\boldsymbol{\beta}$, σ^2 et $\boldsymbol{\theta}$ du modèle (2.13) sont inconnus. On peut les fixer arbitrairement ou bien les estimer au vu des observations $\mathbf{g}_n$ par des méthodes que nous décrivons ci-dessous. On note $\widehat{\boldsymbol{\beta}}_n$, $\widehat{\sigma}_n^2$ et $\widehat{\boldsymbol{\theta}}_n$ les estimateurs de ces paramètres. Le modèle de processus gaussien dont on se sert en pratique pour modéliser la fonction inconnue g est celui de la proposition 2.3.1, où les paramètres sont remplacés par leurs estimations.

Maximum de vraisemblance

Une méthode couramment utilisée pour l'estimation des paramètres $\boldsymbol{\beta}$, σ^2 et $\boldsymbol{\theta}$ est celle du maximum de vraisemblance. Sous les hypothèses du modèle (2.13), la log-vraisemblance est donnée, à une constante près, par :

$$L\big(\mathbf{g}_n;\boldsymbol{\beta},\sigma^2,\boldsymbol{\theta}\big) = -\frac{1}{2}\left[n\log\sigma^2 + \log\big|\det\boldsymbol{\Sigma}_{\boldsymbol{n},\boldsymbol{\theta}}\big| + \frac{\big(\mathbf{g}_n - \mathbf{M}_n\boldsymbol{\beta}\big)^T\boldsymbol{\Sigma}_{n,\boldsymbol{\theta}}^{-1}\big(\mathbf{g}_n - \mathbf{M}_n\boldsymbol{\beta}\big)}{\sigma^2}\right]. \tag{2.19}$$

Par maximisation cette log-vraisemblance, on en déduit que l'estimateur du maximum de vraisemblance de $\boldsymbol{\beta}$ est :

$$\widehat{\boldsymbol{\beta}}_n = \big(\mathbf{M}_n^T\boldsymbol{\Sigma}_{n,\boldsymbol{\theta}}^{-1}\mathbf{M}_n\big)^{-1}\mathbf{M}_n^T\boldsymbol{\Sigma}_{n,\boldsymbol{\theta}}^{-1}\mathbf{g}_n.$$

Par suite, l'estimateur du maximum de vraisemblance de σ^2 est :

$$\widehat{\sigma}_n^2 = \frac{1}{n}\big(\mathbf{g}_n - \mathbf{M}_n\widehat{\boldsymbol{\beta}}_n\big)^T\boldsymbol{\Sigma}_{n,\boldsymbol{\theta}}^{-1}\big(\mathbf{g}_n - \mathbf{M}_n\widehat{\boldsymbol{\beta}}_n\big).$$

Enfin, l'estimateur du maximum de vraisemblance de $\boldsymbol{\theta}$ est l'estimateur $\widehat{\boldsymbol{\theta}}_n$ qui vérifie :

$$\widehat{\boldsymbol{\theta}}_n = \operatorname*{argmin}_{\boldsymbol{\theta}}\Big[n\log\widehat{\sigma}_n^2 + \log\big|\det\boldsymbol{\Sigma}_{\boldsymbol{n},\boldsymbol{\theta}}\big|\Big].$$

Des informations supplémentaires sur la méthode du maximum de vraisemblance, ainsi que sur sa mise en œuvre pratique, peuvent être retrouvées à la section 6.4 de (Stein, 1999), mais aussi dans les thèses suivantes : (Marrel, 2008), (Bachoc, 2013b), (Le Gratiet, 2013) et (Auffray et al., 2014). Un des inconvénients de la méthode est le coût de calcul algorithmique dépendant de la taille de l'échantillon d'apprentissage $\mathbf{D}_n$. Lorsque le nombre d'observations est faible, les coûts sont moindres mais les extrema de la fonction de vraisemblance (2.19) sont généralement mal identifiés. À l'inverse, lorsque le nombre d'observations de g est trop élevé, l'inversion de la matrice de covariance $\boldsymbol{\Sigma}_{n,\boldsymbol{\theta}}$ (de taille $n \times n$) devient difficile et numériquement coûteux (voir aussi (Rasmussen and Williams, 2006) pour une discussion sur les avantages et les limitations de la méthode du maximum de vraisemblance).

Alternativement, la méthode dite du maximum de vraisemblance restreint peut être appliquée (voir (Patterson and Thompson, 1971), (Bettinger, 2009) et (Le Gratiet, 2013)). Dans (Li and Sudjianto, 2005), les auteurs proposent de pénaliser la vraisemblance pour empêcher d'obtenir, avec cette méthode, des estimateurs de variance trop élevée. Dans (Fang et al., 2006), différents choix de pénalisation de la vraisemblance sont comparés pour des plans d'expériences contenant peu de points.

Validation croisée

La méthode de validation croisée consiste à partitionner les observations $(\mathbf{D}_n, \mathbf{g}_n)$ en deux échantillons : un échantillon d'apprentissage et un échantillon de validation. Lorsque ce dernier ne contient qu'une seule observation, la méthode porte le nom de « leave-one-out ».

Supposons que les paramètres $\boldsymbol{\beta}$ et σ^2 aient été fixés ou estimés, par exemple, par maximum de vraisemblance. Alors, le principe du leave-one-out pour l'estimation de $\boldsymbol{\theta}$ est le suivant. Pour tout $i = 1, \ldots, n$, l'échantillon d'apprentissage est $\{(\mathbf{x}_j, g(\mathbf{x}_j)) : j = 1, \ldots, n, j \neq i\}$ et l'échantillon de validation est nécessairement réduit au couple $(\mathbf{x}_i, g(\mathbf{x}_i))$. On utilise l'échantillon d'apprentissage pour construire le modèle d'approximation de la fonction g donné par la moyenne (2.16), que l'on le note ici $m_n^{-i}(\mathbf{x}|\boldsymbol{\theta})$, $\forall \mathbf{x} \in \mathbb{X}$. On peut alors calculer la différence $g(\mathbf{x}_i) - m_n^{-i}(\mathbf{x}_i|\boldsymbol{\theta})$ et, puisque cela donne de l'information sur la capacité de prédiction du modèle au point $\mathbf{x}_i$, le paramètre $\boldsymbol{\theta}$ que l'on choisit est celui qui réalise le minimum de la quantité suivante :

$$\frac{1}{n}\sum_{i=1}^{n}\left(g(\mathbf{x}_i) - m_n^{-i}(\mathbf{x}|\boldsymbol{\theta})\right)^2.$$

Pour une comparaison entre la méthode de maximum de vraisemblance et la validation croisée, on pourra consulter (Bachoc, 2013a).

2.3.3.4 Estimation bayésienne : paramètres vus comme aléatoires

On peut adopter une approche bayésienne en se donnant des lois a priori sur les paramètres et en utilisant les observations $\mathbf{g}_n$ pour calculer les lois a posteriori. Dans (Santner et al., 2003), on parle de « modèle hiérarchique » car la loi a posteriori du processus ξ est calculée en intégrant les lois a posteriori des paramètres. Par exemple, pour σ^2 et $\boldsymbol{\theta}$ fixés, si on se donne une loi a priori gaussienne pour $\boldsymbol{\beta}$, alors la loi conditionnelle de ξ sachant les observations est toujours gaussienne. Les expressions des paramètres de cette loi peuvent être retrouvées au chapitre 1 de (Le Gratiet, 2013). Divers exemples de choix a priori pour les paramètres sont d'ailleurs étudiés dans cette thèse.

Il existe un certain nombre d'arguments qui peuvent être avancés pour justifier le choix d'une loi a priori. Dans (Kass and Wasserman, 1996) et (Kato, 2009) par exemple, la façon de procéder est de se donner une loi a priori non-informative, c'est-à-dire induisant le minimum d'information sur le paramètre (voir la section 3.5 de (Robert, 2007) pour une présentation générale des lois non-informatives). Un exemple classique de loi non-informative est la loi de Jeffreys qui est construite de façon à favoriser les valeurs du paramètre pour lesquelles l'information de Fisher est grande (voir (Jeffreys, 1961)).

Puisque, dans cette approche, la loi de ξ conditionnée aux observations tient compte de la loi des paramètres, les résultats de la proposition 2.3.1 ne sont plus valables. Autrement dit, la loi conditionnelle n'est pas nécessairement gaussienne. En effet, on peut, par exemple, montrer qu'une loi a priori gaussienne pour $\boldsymbol{\beta}$ sachant σ^2 et une loi a priori de Jeffreys pour σ^2 conduisent à une loi a posteriori de Student pour ξ. Pour plus de détails, on peut consulter (Sacks et al., 1989a), ainsi que les thèses suivantes : (Barbillon, 2010), (Le Gratiet, 2013) et (Oger, 2014).

En pratique, le calcul des lois a posteriori n'est pas trivial. Cependant, des méthodes MCMC peuvent être appliquées pour réaliser ces calculs. Par exemple, l'algorithme de Metropolis-Hastings que l'on présente au chapitre 4 peut être utilisé.

2.3.3.5 Choix de la fonction de covariance

Le choix de la fonction de covariance dans le modèle (2.13) joue un rôle primordial et influe sur les capacités de prédiction. En effet, pour les processus gaussiens, la régularité des trajectoires (par exemple, à travers les notions de continuité höldérienne) est liée à la covariance (voir (Belyaev, 1961), (Fernique, 1971) et (Adler, 1981)). En section 2.3.3.2, on a donné l'exemple de la fonction de corrélation exponentielle quadratique, mais il existe d'autres fonctions de corrélation qui peuvent être utilisées pour calculer la covariance. On en présente quelques-unes ici, et pour une étude plus complète, on pourra consulter (Stein, 1999) et le chapitre 4 de (Rasmussen and Williams, 2006).

Commençons par rappeler qu'une fonction $k_{\boldsymbol{\theta}} : (\mathbf{x}, \mathbf{x}') \in \mathbb{X}^2 \mapsto k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}')$ est une fonction de corrélation si elle est symétrique :

$$k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}') = k_{\boldsymbol{\theta}}(\mathbf{x}', \mathbf{x}), \quad \forall (\mathbf{x}, \mathbf{x}') \in \mathbb{X}^2,$$

et semi-définie positive, c'est-à-dire que $\forall m \in \mathbb{N}^*$, $\forall (\mathbf{x}_i)_{1 \leq i \leq m} \in \mathbb{X}^m$ et $\forall (\lambda_i)_{1 \leq i \leq m} \in \mathbb{R}^m$, on a :

$$\sum_{i,j=1}^{m} \lambda_i \lambda_j k_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j) \geq 0.$$

En outre, une restriction habituelle est de considérer les fonctions de corrélations stationnaires. Cela signifie que pour tout $(\mathbf{x}, \mathbf{x}') \in \mathbb{X}^2$, la fonction $k_{\boldsymbol{\theta}}$ est invariante par translation, c'est-à-dire qu'elle est fonction de $\mathbf{x} - \mathbf{x}'$:

$$k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}') = \rho_{\boldsymbol{\theta}}(\mathbf{x} - \mathbf{x}'),$$

où $\rho_{\boldsymbol{\theta}} : \mathbb{X} \to \mathbb{R}$, avec la convention que $\rho_{\boldsymbol{\theta}}(0) = 1$. Cela implique que le processus ζ dans le modèle (2.13) est stationnaire au sens faible. On dit aussi stationnaire à l'ordre 2. En outre, dans le cas où la fonction moyenne du processus ξ est constante (i.e. $\mathbf{m}(\mathbf{x}) = \mu \in \mathbb{R}$, $\forall \mathbf{x} \in \mathbb{X}$), alors le choix d'une fonction de corrélation stationnaire implique aussi que ξ est faiblement stationnaire (voir le chapitre 4 de (Rasmussen and Williams, 2006)).

Dans ce qui suit, la notation $\|\cdot\|$ désigne la norme euclidienne sur $\mathbb{X}$. On distingue deux types de fonctions de corrélation. Les premières dont nous allons parler sont les fonctions

isotropes, et les secondes sont les fonctions anisotropes. Les fonctions isotropes sont populaires car le nombre de paramètres à estimer est réduit. Par exemple, les fonctions de corrélation qui s'expriment, pour tout $(\mathbf{x}, \mathbf{x}') \in \mathbb{X}^2$, par :

$$k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}') = \rho_{\boldsymbol{\theta}}(\|\mathbf{x} - \mathbf{x}'\|),$$

où $\boldsymbol{\theta} = (\theta, \nu)$, avec $\theta \in \mathbb{R}$ et $\nu \in \mathbb{R}$, sont des fonctions de corrélation isotropes. Le paramètre ν réfère à la régularité du processus, et notamment à la continuité et à la différentiabilité en moyenne quadratique. Dans les thèses (Barbillon, 2010), (Le Gratiet, 2013) et (Bachoc, 2013b) – et les références qui s'y trouvent – ces notions sont explicitement définies et liées aux propriétés de régularité des trajectoires. Des exemples de fonctions de corrélation isotropes bien connues sont les suivants :

— La fonction gaussienne ou exponentielle quadratique :

$$k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2}\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{\theta^2}\right).$$

Des exemples de trajectoires de processus gaussien avec cette fonction de corrélation ont été donnés figure 2.3.

— La fonction de Matérn :

$$k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}') = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}\|\mathbf{x} - \mathbf{x}'\|}{\theta}\right)^{\nu} K_{\nu}\left(\frac{\sqrt{2\nu}\|\mathbf{x} - \mathbf{x}'\|}{\theta}\right), \tag{2.20}$$

où $\theta > 0$ et K_{ν} est une fonction de Bessel modifiée d'ordre $\nu > 0$ (voir (Abramowitz and Stegun, 1965)). Les fonctions de Matérn avec $\nu = \frac{3}{2}$ et $\nu = \frac{5}{2}$ sont les plus couramment utilisées dans le cadre de la modélisation par processus gaussien. Cela est dû au fait que, lorsque $\nu = q + \frac{1}{2}$, où $q \in \mathbb{N}$, la fonction de Matérn s'exprime simplement comme un produit de la fonction exponentielle et de polynômes d'ordre q (voir (Abramowitz and Stegun, 1965)). Par exemple, pour $\nu = \frac{3}{2}$, cela donne :

$$k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}') = \left(1 + \frac{\sqrt{3}\|\mathbf{x} - \mathbf{x}'\|}{\theta}\right)\exp\left(-\frac{\sqrt{3}\|\mathbf{x} - \mathbf{x}'\|}{\theta}\right).$$

On remarque que lorsque ν tend vers l'infini dans (2.20), on retrouve la fonction gaussienne, et lorsque $\nu = \frac{1}{2}$, on retrouve la fonction dite exponentielle, qui s'écrit :

$$k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|}{\theta}\right).$$

Les trajectoires d'un processus gaussien avec une fonction de corrélation exponentielle sont continues mais pas différentiables.

— La fonction ν-exponentielle :

$$k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\left(\frac{\|\mathbf{x} - \mathbf{x}'\|}{\theta}\right)^{\nu}\right),$$

où $0 < \nu \leq 2$. Si $\nu = 1$, on retrouve le noyau exponentiel ci-dessus.

Alternativement, il possible de considérer des variantes anisotropes des fonctions de corrélation isotropes en utilisant, non plus la distance euclidienne, mais une distance quadratique (dite de Mahalanobis) de la forme $(\mathbf{x}-\mathbf{x}')^T M(\mathbf{x}-\mathbf{x}')$, avec une matrice M définie positive fixée (voir (Rasmussen and Williams, 2006)). Pour cette classe de fonctions, le paramètre $\boldsymbol{\theta}$ s'écrit :

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d, \nu)\,,$$

où $\theta_i \in \mathbb{R}$, $\forall i = 1, \ldots, d$, et $\nu \in \mathbb{R}$. Les fonctions anisotropes font donc intervenir un nombre plus élevé de paramètres que les fonctions isotropes. Cependant, elles peuvent être préférables dans le cas de données spatiales, le paramètre θ_i pouvant être vu comme la portée de l'information dans la i-ème direction (voir (Ginsbourger, 2009)). Un exemple simple de fonction de corrélation anisotrope est la fonction gaussienne, encore appelée exponentielle quadratique :

$$k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2}\sum_{i=1}^{d}\frac{(x_i - x_i')^2}{\theta_i^2}\right).$$

2.4 Plans d'expériences

Il existe de nombreuses méthodes pour construire un plan d'expériences qui contient des points suffisamment bien répartis dans $\mathbb{X}$ pour que l'information sur g dont on dispose soit suffisante pour mener une estimation relativement peu biaisée. Chaque appel à g étant coûteux, on privilégie les plans d'expériences qui fournissent le maximum d'information sur cette fonction. On cherche donc des plans qui évitent les réplications (la fonction g étant déterministe, les réplications n'apportent aucune information supplémentaire) et qui échantillonnent dans les zones d'importance (par exemple, on souhaite des points proches de la frontière du domaine de défaillance, ou bien dans les zones où le modèle a de mauvaises capacités de prédiction et doit être amélioré).

On distingue deux types de plans d'expériences. Les premiers que l'on présente sont les plans dits « space-filling ». Ils sont généralement construits avant l'étape de modélisation et ont pour but de couvrir au mieux le domaine $\mathbb{X}$. Les seconds sont les plans dits « adaptatifs » ou « optimaux ». Ils sont construits de façon itérative en tenant compte de toute l'information sur g déjà disponible pour déterminer le prochain point d'évaluation.

2.4.1 Méthodes classiques d'échantillonnage

La plupart des méthodes décrites ci-après peuvent être retrouvées dans (Pronzato and Müller, 2012) et les plans qu'elles génèrent sont généralement appelés « space-filling » en anglais.

Dans la méthode Monte-Carlo naïve, l'échantillon est simulé suivant la loi naturelle $\mathbf{P}_{\mathbf{X}}$ et la vitesse de convergence est en $\mathcal{O}\big(1/\sqrt{N}\big)$, c'est-à-dire indépendante de la dimension d de $\mathbf{X}$, mais relativement lente. Les méthodes quasi Monte-Carlo ont pour objectif d'améliorer cette vitesse. L'erreur de ces méthodes peut aller jusqu'en $\mathcal{O}\big(\log(N)^d/N\big)$, mais elle dépend

cette fois de la dimension d de $\mathbb{X}$. Pour les méthodes quasi Monte-Carlo, $\mathbf{X}$ est typiquement supposé à valeurs dans $[0,1]^d$ et distribué uniformément. Ces méthodes utilisent des suites basées sur des critères de discrépance. Par exemple, la discrépance à l'origine d'une suite en dimension $d = 1$ est la mesure en norme infinie entre sa fonction de répartition empirique et celle de la loi uniforme. Généralement, on utilise des suite à discrépance faible, comme les suites de Halton (Rafajlowicz and Schwabe, 2006) ou les suites de Sobol (Bratley and Fox, 1988). Un exemple d'échantillon de taille $N = 100$, obtenu par tirage d'une suite de Halton, est donné en haut à gauche de la figure 2.5. D'autres formes de discrépance, ainsi que des algorithmes pour obtenir des plans uniformes, sont disponibles dans le livre de (Fang et al., 2006). Pour plus de détails sur les méthodes quasi Monte-Carlo, on pourra consulter, par exemple, (Niederreiter, 1992), (Calfisch, 1998) et (Tuffin, 2010).

Une autre méthode d'échantillonnage quasi aléatoire est la méthode d'échantillonnage par hypercube latin. Le plan d'expériences obtenu par cette méthode est généralement appelé plan LHS pour « Latin Hypercube Sample » (voir (McKay et al., 1979) et (Stein, 1987)). Pour obtenir un plan LHS de taille N, le principe est le suivant : on découpe chaque axe $[0,1]$ en N sous-intervalles $\left[0, \frac{1}{N}\right], \ldots, \left[\frac{N-1}{N}, 1\right]$, ce qui conduit à partitionner $[0,1]^d$ en N^d cellules de même taille. On sélectionne alors N cellules de façon à avoir exactement une cellule sur chaque ligne et sur chaque colonne. Enfin, dans chaque cellule sélectionnée, on tire un point au hasard selon la loi uniforme dans la cellule. Chaque point de l'échantillon ainsi créé est positionné de manière à ne pas avoir de coordonnées communes avec les autres points. Un exemple de plan LHS est proposé en haut à droite de la figure 2.5.

On peut également construire des échantillons en optimisant un critère basé sur la distance entre les points de façon à obtenir une bonne répartition dans $\mathbb{X}$. Deux exemples classiques de plans d'expériences basés sur la distance entre les points sont les plans minimax et maximin (voir (Johnson et al., 1990)). Un plan d'expériences $(\mathbf{x}_i)_{1\leq i\leq n}$ est dit minimax, s'il minimise :

$$\sup_{\mathbf{x}\in\mathbb{X}} \min_{1\leq i\leq n} \|\mathbf{x} - \mathbf{x}_i\|,$$

et maximin, s'il maximise :

$$\min_{1\leq i,j\leq n} \|\mathbf{x}_i - \mathbf{x}_j\|,$$

où $\|\cdot\|$ est la distance euclidienne sur $\mathbb{X}$. Comme on peut le voir sur la figure 2.5, un plan minimax a tendance à placer les plans à l'intérieur de $\mathbb{X}$, tandis qu'un plan maximin a tendance à placer les points aux bornes de $\mathbb{X}$. Dans (Mitchell et al., 1995), les auteurs proposent de chercher un plan d'expérience maximin dans la classe des plans LHS (voir aussi (Joseph and Hung, 2008)).

2.4.2 Planification d'expériences séquentielle

La procédure pour déterminer les points où évaluer la fonction g peut être adaptative : on parle de planification d'expériences séquentielle. Partant d'un plan initial de type « space-filling » (un plan LHS, par exemple), on détermine chaque prochain point d'évaluation

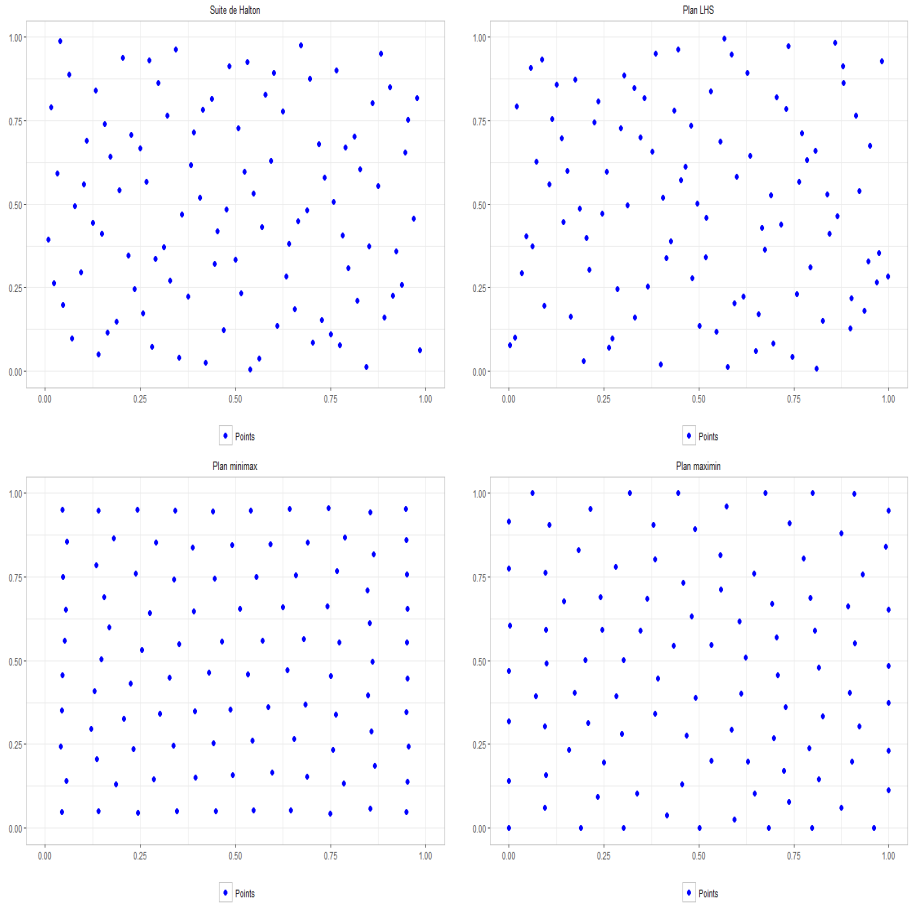

FIGURE 2.5 – Exemples de plans d'expériences de taille $N = 100$. En haut à gauche : une suite de Halton. En haut à droite : un plan LHS. En bas à gauche : un plan minimax. En bas à droite : un plan maximin.

de g en tenant compte des valeurs déjà observées de cette fonction. La procédure repose généralement sur l'optimisation d'un critère, qui dépend du modèle d'approximation de g que l'on s'est fixé, ainsi que de l'objectif visé. Par exemple, on ne considère pas le même critère pour résoudre un problème d'optimisation globale que pour estimer une probabilité de défaillance. À chaque itération, on effectue une ou plusieurs évaluations supplémentaires de g, et les paramètres du modèle peuvent être réajustés au vu de ces nouvelles observations. En anglais, un plan d'expériences construit de cette façon est appelé « model-based design of experiments ». On précise qu'il n'y a aucune raison pour que le plan d'expériences obtenu à la fin de la procédure conserve les mêmes propriétés ou la même structure que le plan initial.

Le cadre dans lequel on se place est le suivant : on suppose que la fonction g est déjà observée aux points d'un plan d'expériences $\mathbf{D}_n = (\mathbf{x}_i)_{1 \leq i \leq n}$ et on cherche le point $\mathbf{x}_{n+1} \in \mathbb{X} \backslash \mathbf{D}_n$ en lequel effectuer la prochaine évaluation de g. Dans ce qui suit, on présente essentiellement des critères d'échantillonnage basés sur la modélisation par processus gaussien, car ils sont couramment utilisés pour répondre à des problématiques industrielles et, par conséquent, relativement faciles à mettre en œuvre. En outre, on s'intéresse dans le chapitre 3 aux stratégies SUR, que l'on présente très brièvement ci-après, et qui ont été développées dans le cadre de la modélisation par processus gaussien. Notons que la version adaptative de la méthode de splitting présentée en section 2.2.3 peut être vue comme une méthode de planification d'expériences séquentielle (voir (Au and Beck, 2001) et (Cérou et al., 2012)). On pourrait bien sûr l'envisager si la fonction g n'était pas coûteuse à évaluer. Par ailleurs, bien qu'on ne le présente pas comme tel, l'algorithme que l'on propose au chapitre 4 rentre également dans ce cadre.

2.4.2.1 Critères d'échantillonnage basés sur la modélisation par processus gaussien

Il existe de nombreux critères développés dans le cadre de la modélisation de g par un processus aléatoire gaussien. Par exemple, dans (Sacks et al., 1989a), les auteurs proposent un critère basé sur la variance σ_n^2 du processus aléatoire gaussien conditionné aux observations (l'expression de cette variance est donnée à l'équation (2.17)). Ce critère porte le nom d'erreur en moyenne quadratique maximale (MMSE pour « Maximum Mean Squared Error »). Il s'agit de choisir le point $\mathbf{x}_{n+1}^*$ qui vérifie :

$$\mathbf{x}_{n+1}^* = \underset{\mathbf{x}_{n+1} \in \mathbb{X}}{\operatorname{argmin}} \max_{\mathbf{x} \in \mathbb{X}} \sigma_{n+1}^2(\mathbf{x}),$$

où $\sigma_{n+1}^2(\mathbf{x})$ est la variance calculée pour le plan d'expériences $\mathbf{D}_n \cup \mathbf{x}_{n+1}$. Même si g n'est pas observée au point $\mathbf{x}_{n+1}$, cette variance peut être calculée car elle ne dépend pas des valeurs observées de g. Dans (Sacks et al., 1989a), les auteurs considèrent également l'erreur en moyenne quadratique intégrée (IMSE pour « Integrated Mean Squared Error ») :

$$\text{IMSE} = \int_{\mathbb{X}} \sigma_n^2(\mathbf{x}) d\mathbf{x}.$$

Elle est reconsidérée dans (Picheny et al., 2010) et porte le nom de « targeted IMSE ». Elle dépend alors d'une fonction de poids w_n, qui tient compte des valeurs de g déjà observées

aux points du plan $\mathbf{D}_n$:

$$\text{tIMSE} = \int_{\mathbb{X}} \sigma_n^2(\mathbf{x}) w_n(\mathbf{x}) d\mathbf{x}.$$

Par exemple, en notant ξ_n le processus aléatoire gaussien conditionné aux observations, les auteurs dans (Picheny et al., 2010) proposent la fonction w_n suivante :

$$w_n(\mathbf{x}) = \mathbb{E}\big[\mathbb{1}_{|\xi_n(\mathbf{x})-T|<\varepsilon}\big], \quad \text{avec} \quad \varepsilon > 0.$$

Dans ces conditions, le critère tIMSE oriente le choix des points de façon à réduire l'incertitude sur le processus ξ_n dans les sous-ensembles de $\mathbb{X}$ où ses trajectoires sont proches du seuil T. Ainsi, le point qui est sélectionné par le critère tIMSE vérifie :

$$\mathbf{x}_{n+1}^* = \underset{\mathbf{x}_{n+1}\in\mathbb{X}}{\text{argmin}}\ \text{tIMSE} = \underset{\mathbf{x}_{n+1}\in\mathbb{X}}{\text{argmin}} \int_{\mathbb{X}} \sigma_{n+1}^2(\mathbf{x}) w_n(\mathbf{x}) d\mathbf{x}.$$

Pour répondre à une problématique d'estimation de probabilité de défaillance similaire à la nôtre, des procédures itératives de sélection de points sont aussi proposées dans (Echard et al., 2011) et (Dubourg et al., 2013). Elles ont en commun le fait de combiner la modélisation par processus gaussien et les méthodes de simulation Monte-Carlo. Supposons que l'on dispose d'un échantillon $(\mathbf{X}_i)_{1\leq i\leq N}$ i.i.d. suivant $\mathbf{P}_{\mathbf{X}}$. Alors, dans (Echard et al., 2011) par exemple, l'estimateur de la probabilité de défaillance est :

$$\widehat{p}_N = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{m_n(\mathbf{X}_i)>T}, \tag{2.21}$$

où m_n est la fonction moyenne du processus gaussien conditionné aux observations donnée en équation (2.16). On rappelle que cette fonction interpole les points où g est observée. Elle est, par conséquent, généralement utilisée comme modèle d'approximation de g. En outre, elle est relativement peu coûteuse à évaluer. Le point choisi pour enrichir le plan d'expériences $\mathbf{D}_n$ est le point $\mathbf{x}_{n+1}^*$ qui réalise le minimum de la fonction U suivante :

$$\mathbf{x}_{n+1}^* = \underset{\mathbf{x}_{n+1}\in\mathbb{X}}{\text{argmin}}\ U(\mathbf{x}_{n+1}) = \underset{\mathbf{x}_{n+1}\in\mathbb{X}}{\text{argmin}}\ \frac{|m_n(\mathbf{x}_{n+1}) - T|}{\sigma_n(\mathbf{x}_{n+1})}.$$

Ce critère réalise un compromis entre les zones où la fonction m_n est proche de T et où le modèle de processus gaussien est incertain, c'est-à-dire mal renseigné. Cela revient à améliorer la précision de l'estimateur (2.21).

2.4.2.2 Stratégies SUR

Toujours dans le cadre de l'estimation d'une probabilité de défaillance, il est proposé dans (Bect et al., 2012) une méthode de planification d'expériences séquentielle portant le nom de stratégie SUR (« Stepwise Uncertainty Reduction ») et basée sur la méthode de krigeage. Dans le chapitre 3, nous utilisons cette méthode pour améliorer la performance de notre procédure d'estimation de la probabilité p et, en particulier, nous l'appliquons à l'étude d'un cas réel de la société STMicroelectronics. Les outils nécessaires à la présentation et

la compréhension de cette approche seront donc essentiellement donnés dans le chapitre 3. Néanmoins, on précise que l'idée sous-adjacente aux stratégies SUR est de sélectionner le point où évaluer la fonction g qui minimise l'incertitude sur une quantité d'intérêt. En quelques mots, le formalisme est le suivant :

La procédure étant itérative, on munit l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ de la filtration $(\mathcal{F}_n)_{n\in\mathbb{N}^*}$ définie comme la tribu engendrée par les variables aléatoires $(\xi(\mathbf{x}_1), \ldots, \xi(\mathbf{x}_n))$:

$$\mathcal{F}_n = \sigma(\xi(\mathbf{x}_1), \ldots, \xi(\mathbf{x}_n)).$$

Pour tout $n \in \mathbb{N}^*$, on pose $\mathbb{E}_n[\cdot] = \mathbb{E}[\cdot \mid \mathcal{F}_n]$ et $\mathrm{Var}_n[\cdot] = \mathrm{Var}[\cdot \mid \mathcal{F}_n]$. Il s'agit alors de déterminer le point $\mathbf{x}^*_{n+1}$ qui vérifie :

$$\mathbf{x}^*_{n+1} = \underset{\mathbf{x}_{n+1}\in\mathbb{X}}{\operatorname{argmin}}\ J_n(\mathbf{x}_{n+1}),$$

où le critère J_n est défini par :

$$J_n(\mathbf{x}_{n+1}) = \mathbb{E}_n\big[H_{n+1} \mid \mathbf{X}_{n+1} = \mathbf{x}_{n+1}\big],$$

avec H_{n+1} une variable aléatoire qui mesure l'incertitude sur une quantité d'intérêt en considérant $n+1$ points, si le $(n+1)$-ème point d'observation est $\mathbf{x}_{n+1}$. En utilisant ce critère, on espère minimiser le nombre d'observations requis pour atteindre une précision voulue sur la quantité d'intérêt. Par exemple, on peut prendre $H_{n+1} = \ell(\alpha, \widehat{\alpha}_{n+1})$, où la fonction $\ell : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ est une fonction de perte et $\widehat{\alpha}_{n+1}$ est un estimateur de la quantité d'intérêt α. Dans le cas où l est la fonction de perte quadratique et $\widehat{\alpha}_{n+1} = \mathbb{E}_{n+1}[\alpha]$, on a :

$$H_{n+1} = (\alpha - \mathbb{E}_{n+1}[\alpha])^2.$$

Puisque $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, on a encore :

$$\mathbb{E}_n[H_{n+1}] = \mathbb{E}_n[(\alpha - \mathbb{E}_{n+1}[\alpha])^2] = \mathbb{E}_n[\mathbb{E}_{n+1}[(\alpha - \mathbb{E}_{n+1}[\alpha])^2]] = \mathbb{E}_n[\mathrm{Var}_{n+1}(\alpha)].$$

La stratégie consiste donc à déterminer le point $\mathbf{x}^*_{n+1}$ qui minimise la variance future espérée de α.

Chapitre 3

Estimation d'une probabilité de défaillance avec l'ordre convexe

3.1 Introduction

Dans ce chapitre, on propose d'estimer la probabilité de défaillance p définie en (1.1) en adoptant la même approche bayésienne que dans (Auffray et al., 2014). En s'appuyant sur les principes de la méthode de krigeage et sur les observations de la fonction g aux points d'un plan d'expériences fixé, on définit une variable aléatoire S_n à valeurs dans $[0,1]$. La probabilité p étant inconnue, on fait l'hypothèse qu'il s'agit d'une réalisation de cette variable aléatoire. D'un point de vue bayésien, cela signifie que l'on mène l'estimation de p par le biais de la distribution de S_n. Cependant, on constate assez rapidement les limites de cette approche : la distribution exacte est hors d'atteinte, la distribution empirique est difficile à calculer et les critères de dispersion usuels (variance, moments, quantile...) ne peuvent être estimés à moindre coût. L'espérance de S_n est néanmoins très facile à estimer par une méthode Monte-Carlo naïve.

Dans (Oger et al., 2015), les auteurs proposent à la place de S_n une variable aléatoire notée R_n. Ils montrent en particulier que S_n et R_n ont la même espérance. Si ces deux variables aléatoires sont comparables en terme de biais, le résultat principal de ce chapitre est plus général. En effet, nous montrons qu'il existe un ordre convexe entre S_n et R_n. Cela signifie que pour toute fonction convexe φ, on a :

$$\mathbb{E}[\varphi(S_n)] \leq \mathbb{E}[\varphi(R_n)].$$

Ce résultat implique que les variables aléatoires sont aussi comparables en terme de variance. Nous l'utilisons notamment pour apprendre sur la loi de S_n et nous montrons notamment que l'on peut borner la fonction quantile de S_n. On en déduit des intervalles de crédibilité pour la loi de cette variable aléatoire. Pour une valeur de $\alpha \in [0,1]$ fixée, un intervalle $[a_\alpha, b_\alpha]$ vérifiant :

$$\mathbb{P}(a_\alpha \leq S_n \leq b_\alpha) \geq 1-\alpha,$$

est un intervalle de crédibilité de niveau $1-\alpha$ pour la loi de S_n. Si la moyenne de S_n apporte au décideur une information qui peut lui permettre de statuer quant à la robustesse du

produit, un intervalle de crédibilité mesure ainsi l'incertitude sur cet indicateur. Parmi les méthodes déjà existantes dans la littérature, qui utilisent aussi la méthode de krigeage pour l'estimation de p, cette information est rarement disponible.

Lorsque l'intervalle de crédibilité de R_n est jugé trop large pour permettre de considérer la valeur moyenne comme une estimation fiable du rendement, on conclut que la modélisation n'est pas suffisamment efficace et que l'on doit fournir de nouvelles observations. En se basant sur le principe des stratégies SUR présentées dans (Bect et al., 2012), on utilise notre résultat sur l'ordre convexe pour mettre en œuvre une méthode de planification d'expériences séquentielle. Grâce à un critère basé sur la variance de R_n, on sélectionne de façon itérative les points à ajouter au plan d'expériences et on ajuste le modèle au vu de ces nouvelles observations. Pour répondre aux besoins de la société STMicroelectronics, on étudie cette méthode sur un cas réel. On dispose de 10^3 résultats de simulations numériques auxquels nous confronter. Nous verrons que, grâce à la planification séquentielle, les estimations que l'on obtient avec un nombre bien inférieur de simulations sont compatibles avec celles obtenues avec 10^3.

En section 3.2, on définit la variable aléatoire S_n, et on rappelle les avantages et les limites de cette approche rencontrés dans les précédents travaux sur le sujet. En section 3.3, on présente la variable aléatoire R_n et, en section 3.4, on montre qu'il existe un ordre convexe entre ces deux variables aléatoires. En section 3.5, on propose plusieurs applications de ce résultat et on s'intéresse notamment à la construction d'intervalles de crédibilité pour la loi de S_n. En section 3.6, on présente les stratégies SUR pour la construction du plan d'expériences. Des exemples d'applications sont donnés tout au long du chapitre, et l'étude sur le cas réel de la société STMicroelectronics est proposée en section 3.7. Pour finir, une conclusion et des perspectives sont données en section 3.8. Enfin, précisons qu'une présentation complémentaire de la théorie des mesures de risque et des ordres stochastiques est proposée en annexe A. On insiste particulièrement sur l'ordre convexe, car des propriétés relatives à cet ordre stochastique sont utilisées tout au long du chapitre.

3.2 Estimateur basé sur le krigeage

En section 2.3.3, on a présenté la méthode krigeage qui consiste à modéliser une fonction déterministe par un processus aléatoire. Lorsque ce processus est choisi gaussien, on parle aussi de modélisation par processus gaussien. L'ensemble des résultats qui seront donnés dans ce chapitre sont valables pour n'importe quel choix de processus aléatoire. Cependant, on privilégiera le choix gaussien afin de faciliter les explications et les calculs. On commence par quelques rappels sur le krigeage et on fixe les notations qui seront utilisées tout au long de ce chapitre.

Soit $\mathbf{X}$ une variable aléatoire à valeurs dans $\mathbb{X} \subseteq \mathbb{R}^d$, de loi $\mathbf{P}_{\mathbf{X}}$ suivant laquelle on sait simuler. On rappelle que l'on souhaite estimer la probabilité de défaillance p définie par :

$$p = \mathbb{P}(g(\mathbf{X}) > T) = \int_{\mathbb{X}} \mathbb{1}_{g(\mathbf{x})>T} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) = \mathbf{P}_{\mathbf{X}}(\mathbb{F}),$$

où $T \in \mathbb{R}$ est un seuil fixé, $g : \mathbb{X} \to \mathbb{R}$ est une fonction boîte-noire coûteuse en temps de calcul, et $\mathbb{F} = \{\mathbf{x} \in \mathbb{X} : g(\mathbf{x}) > T\}$ est le domaine de défaillance. On suppose que la

fonction g est observée aux points d'un plan d'expériences $\mathbf{D}_n = (\mathbf{x}_i)_{1\leq i\leq n} \in \mathbb{X}^n$ fixé, et on note $\mathbf{g}_n = (g(\mathbf{x}_i))_{1\leq i\leq n}$ le vecteur des observations. Le principe du krigeage est de modéliser la fonction g par un processus aléatoire ξ indexé par $\mathbb{X}$. On lui attribue une loi a priori, par exemple gaussienne. Sous l'hypothèse que g est une trajectoire de ce processus aléatoire, le vecteur $\mathbf{g}_n$ est une réalisation observée du vecteur aléatoire gaussien $\boldsymbol{\xi}_n = (\xi(\mathbf{x}_i))_{1\leq i\leq n}$. Pour tenir compte de cette information dans la modélisation, on conditionne la loi a priori de ξ à la donnée que $\boldsymbol{\xi}_n = \mathbf{g}_n$. Le processus a posteriori que l'on obtient est toujours gaussien et toutes ses trajectoires interpolent g aux points du plan $\mathbf{D}_n$ (voir figure 2.4).

Afin de ne pas alourdir la présentation, on adopte tout au long de ce chapitre les notations suivantes. On note ξ_n le processus aléatoire de mêmes lois fini-dimensionnelles que le processus ξ conditionné aux observations, ce qui implique l'égalité en loi suivante :

$$\mathcal{L}(\xi(\mathbf{x}) \mid \boldsymbol{\xi}_n = \mathbf{g}_n) = \mathcal{L}(\xi_n(\mathbf{x})), \quad \forall \mathbf{x} \in \mathbb{X}.$$

Pour tout $\mathbf{x} \in \mathbb{X}$, on pose :

$$m_n(\mathbf{x}) = \mathbb{E}[\xi(\mathbf{x}) \mid \boldsymbol{\xi}_n = \mathbf{g}_n] = \mathbb{E}[\xi_n(\mathbf{x})], \quad \sigma_n^2(\mathbf{x}) = \mathrm{Var}[\xi(\mathbf{x}) \mid \boldsymbol{\xi}_n = \mathbf{g}_n] = \mathrm{Var}[\xi_n(\mathbf{x})],$$

et

$$p_n(\mathbf{x}) = \mathbb{E}\big[\mathbb{1}_{\xi(\mathbf{x})>T} \mid \boldsymbol{\xi}_n = \mathbf{g}_n\big] = \mathbb{P}\big(\xi(\mathbf{x}) > T \mid \boldsymbol{\xi}_n = \mathbf{g}_n\big) = \mathbb{P}\big(\xi_n(\mathbf{x}) > T\big). \tag{3.1}$$

Dans le cas gaussien (avec les paramètres de moyenne, de variance et de covariance fixés), la variable $\xi_n(\mathbf{x})$ est gaussienne, d'espérance $m_n(\mathbf{x})$ donnée en (2.16), de variance $\sigma_n^2(\mathbf{x})$ donnée en (2.17). De plus, on a :

$$p_n(\mathbf{x}) = \Phi\left(\frac{m_n(\mathbf{x}) - T}{\sigma_n(\mathbf{x})}\right),$$

où Φ est la fonction de répartition de la loi normale centrée réduite.

3.2.1 Approche bayésienne

Sous l'hypothèse que la fonction g est une trajectoire du processus aléatoire ξ, la probabilité $p = \mathbb{P}\big(g(\mathbf{X}) > T\big)$ est une réalisation de la variable aléatoire $S \in [0,1]$ définie par :

$$S = \mathbb{P}\big(\xi(\mathbf{X}) > T \mid \xi\big) = \int_{\mathbb{X}} \mathbb{1}_{\xi(\mathbf{x})>T} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}). \tag{3.2}$$

D'un point de vue bayésien, cela signifie que l'on se donne une loi a priori et que l'on va mener l'estimation de p par la donnée de la loi a posteriori. Or, d'après ce qui précède, la loi conditionnelle de S sachant $\boldsymbol{\xi}_n = \mathbf{g}_n$ est exactement celle de la variable aléatoire $S_n \in [0,1]$ définie par :

$$S_n = \mathbb{P}\big(\xi_n(\mathbf{X}) > T \mid \xi_n\big) = \int_{\mathbb{X}} \mathbb{1}_{\xi_n(\mathbf{x})>T} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}). \tag{3.3}$$

Si le processus ξ_n est suffisamment bien renseigné autour du seuil (i.e. ses trajectoires reproduisent fidèlement le comportement de g au voisinage du seuil), alors il est raisonnable de supposer que toute réalisation de S_n est relativement proche de la probabilité p.

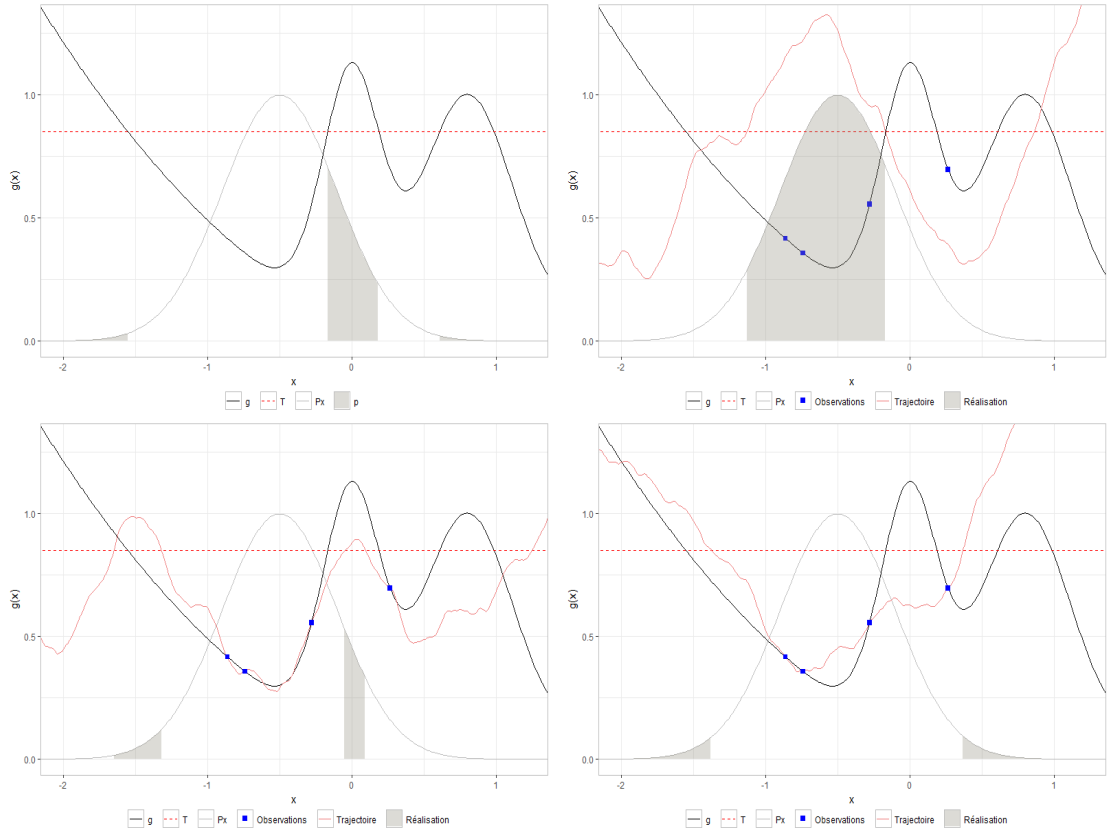

FIGURE 3.1 – En haut à gauche : la probabilité cible p (aire grise). En haut à droite : exemple d'une réalisation de la variable aléatoire S définie en (3.2) (aire grise). La courbe rouge est la trajectoire d'un processus gaussien dont la loi n'est pas conditionnée aux observations (points bleus). En bas : exemple de réalisations de la variable aléatoire S_n définie en (3.3) (aires grises). L'approximation de p est meilleure à gauche qu'à droite. Les courbes rouges sont les trajectoires d'un processus gaussien dont la loi est conditionnée aux observations (points bleus).

Pour fixer les idées, on propose quelques illustrations en figure 3.1. En haut à gauche, on a représenté une fonction $g : \mathbb{R} \to \mathbb{R}$ (courbe noire) et la vraie probabilité de défaillance (aire grise). La fonction est observée en $n = 4$ points. En haut à droite, on a représenté une trajectoire du processus gaussien ξ (courbe rouge) dont la loi n'est pas conditionnée aux observations. La réalisation de S qui en résulte ne fournit pas une bonne approximation de p (aire grise). En bas, les deux trajectoires représentées sont celles du processus gaussien ξ_n dont la loi est conditionnée aux observations. Les aires grises correspondent donc à deux réalisations de S_n. À gauche, l'approximation de p est meilleure car la trajectoire du processus gaussien reproduit mieux le comportement de g au voisinage du seuil.

Puisque l'on a l'égalité en loi :

$$\mathcal{L}(S \mid \boldsymbol{\xi}_n = \mathbf{g}_n) = \mathcal{L}(S_n),$$

la variable d'intérêt est donc ici la variable aléatoire S_n donnée en (3.3). La plupart des

travaux menés dans ce chapitre consistent à obtenir des informations sur la distribution de cette variable aléatoire. Dans (Auffray et al., 2014), où cette approche est également proposée pour estimer une probabilité de défaillance, les auteurs proposent d'approcher les quantiles de S_n par l'intermédiaire des inégalités de Markov et de Bienaymé-Tchebychev. On y reviendra en section 3.5. Avant cela, expliquons comment obtenir une estimation de la probabilité de défaillance.

Commençons par rappeler qu'en statistiques bayésiennes, pour une loi a priori donnée et pour la fonction de perte quadratique, l'estimateur de la quantité d'intérêt que l'on considère est la moyenne de la loi a posteriori, car il s'agit d'un estimateur de Bayes (voir, par exemple, la définition 2.3.3 de (Robert, 2007), ainsi que la proposition 2.5.1 qui montre que l'espérance conditionnelle est un estimateur de Bayes). Dans notre contexte, cela signifie que l'espérance conditionnelle $\mathbb{E}[S \mid \boldsymbol{\xi}_n]$ est un estimateur de Bayes et qu'il est raisonnable de considérer la quantité déterministe $\mathbb{E}[S \mid \boldsymbol{\xi}_n = \mathbf{g}_n]$ comme une bonne approximation de la variable aléatoire S au sens de l'erreur en moyenne quadratique a posteriori. Dans tout ce qui suit, cette quantité est notée μ_n. Elle a une expression très simple et s'écrit :

$$\mu_n = \mathbb{E}[S \mid \boldsymbol{\xi}_n = \mathbf{g}_n] = \int_{\mathbb{X}} \mathbb{E}\big[\mathbb{1}_{\xi(\mathbf{x})>T} \mid \boldsymbol{\xi}_n = \mathbf{g}_n\big] \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) = \int_{\mathbb{X}} p_n(\mathbf{x}) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) = \mathbb{E}[p_n(\mathbf{X})], \tag{3.4}$$

où la fonction p_n donnée en (3.1) est relativement peu coûteuse à évaluer. Pour estimer μ_n, on peut donc appliquer une méthode Monte-Carlo naïve (voir section 2.2.1) et cela ne requiert pas d'appels supplémentaires à la fonction g. On fixe $N \in \mathbb{N}^*$ suffisamment grand et on se donne une suite $(\mathbf{X}_i)_{1 \leq i \leq N}$ de variables aléatoires i.i.d. suivant $\mathbf{P}_{\mathbf{X}}$. On a alors :

$$\mu_n \approx \frac{1}{N} \sum_{i=1}^{N} p_n(\mathbf{X}_i). \tag{3.5}$$

La valeur numérique que retourne (3.5) en pratique peut donc servir comme estimation de la probabilité de défaillance. En outre, puisque l'on a évidemment :

$$\mathbb{E}[S_n] = \mathbb{E}[S \mid \boldsymbol{\xi}_n = \mathbf{g}_n] = \mu_n,$$

la variable aléatoire S_n est un estimateur sans biais de μ_n et l'estimateur (3.5) fournit aussi une estimation de l'espérance de S_n. Notons que l'erreur d'approximation en moyenne quadratique a posteriori de S par μ_n est simplement la variance de S_n.

3.2.2 Limites

Pour déterminer si la loi de S_n est suffisamment concentrée autour de sa moyenne et statuer quant à la précision de l'estimateur (3.5), il est souhaitable de calculer la variance S_n. Comme mentionné dans (Auffray et al., 2014), celle-ci s'écrit :

$$\mathrm{Var}\,[S_n] = \int_{\mathbb{X}^2} \mathrm{Cov}\left(\mathbb{1}_{\xi_n(\mathbf{x}_1)>T}, \mathbb{1}_{\xi_n(\mathbf{x}_2)>T}\right) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}_1) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}_2).$$

De façon équivalente, on écrira plutôt :

$$\mathrm{Var}\left[S_n\right] = \int_{\mathbb{X}^2} \mathbb{P}\big(\xi_n(\mathbf{x}_1) > T, \xi_n(\mathbf{x}_2) > T\big)\mathbf{P}_{\mathbf{X}}(d\mathbf{x}_1)\mathbf{P}_{\mathbf{X}}(d\mathbf{x}_2) - \mu_n^2. \tag{3.6}$$

Cette variance peut être estimée par une méthode Monte-Carlo naïve, un estimateur du terme de droite se déduisant facilement de (3.5). Le terme de gauche (i.e. le moment d'ordre 2) fait intervenir une probabilité jointe et une intégrale double. Par conséquent, son estimation requiert un temps de calcul qui peut être assez élevé. Plus généralement, pour tout $m \in \mathbb{N}^*$, on a :

$$\begin{aligned}\mathbb{E}[S_n^m] &= \mathbb{E}\left[\int_{\mathbb{X}^m} \left(\prod_{i=1}^{m} \mathbb{1}_{\xi_n(\mathbf{x}_i)>T}\right)\mathbf{P}_{\mathbf{X}}(d\mathbf{x}_1)\ldots\mathbf{P}_{\mathbf{X}}(d\mathbf{x}_m)\right] \\ &= \int_{\mathbb{X}^m} \mathbb{P}\left(\bigcap_{i=1}^{m}\left\{\xi_n(\mathbf{x}_i) > T\right\}\right)\mathbf{P}_{\mathbf{X}}(d\mathbf{x}_1)\ldots\mathbf{P}_{\mathbf{X}}(d\mathbf{x}_m). \end{aligned} \tag{3.7}$$

Dès que $m \geq 2$, il devient coûteux d'estimer (3.7) et, par conséquent, difficile d'obtenir des indicateurs de la dispersion de S_n. De plus, la fonction quantile et la fonction de répartition de S_n ne disposent à notre connaissance d'aucune forme explicite et nous n'avons pas accès à la loi de S_n. L'étude de la distribution de S_n par le biais d'une inférence statistique nous semble, de surcroît, déraisonnable. En effet, générer une réalisation de cette variable aléatoire implique d'une part, de simuler une trajectoire du processus aléatoire ξ_n et d'autre part, de réaliser une intégration, par exemple, par l'intermédiaire d'une méthode Monte-Carlo naïve :

$$\forall \omega \in \Omega, \quad S_n(\omega) = \int_{\mathbb{X}} \mathbb{1}_{\xi_n(\omega,\mathbf{x})>T}\mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \approx \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\xi_n(\omega,\mathbf{X}_i)>T}, \tag{3.8}$$

où $(\mathbf{X}_i)_{1\leq i\leq N}$ est une suite de variables aléatoires i.i.d. suivant $\mathbf{P}_{\mathbf{X}}$. Une méthode basée sur la discrétisation de l'ensemble $\mathbb{X}$ pour la simulation de trajectoires conditionnées aux observations est proposée dans (Auffray et al., 2014). Cependant, les auteurs reconnaissent eux-mêmes que cette approche est fastidieuse, car elle nécessite l'inversion d'une matrice de covariance dont la taille dépend du nombre de points retenus pour la discrétisation. Or, ce dernier croît exponentiellement avec la dimension de $\mathbb{X}$.

D'un point de vue bayésien, il est naturel de considérer la variable aléatoire S_n. Cependant, en dehors de son espérance, la distribution de cette variable aléatoire est hors d'atteinte. Les méthodes développées dans le reste de ce chapitre ont donc pour objectif d'apprendre sur la loi de S_n et, par conséquent, de justifier l'intérêt de l'approche bayésienne pour l'estimation de la probabilité de défaillance. Pour ce faire, on s'appuie sur une méthode d'estimation alternative, proposée dans (Oger et al., 2015) et que l'on présente ci-après.

3.3 Estimateur alternatif

Soit U une variable aléatoire de loi uniforme sur $[0,1]$. Considérons la variable aléatoire $R_n \in [0,1]$ définie par :

$$R_n = \mathbb{P}\big(p_n(\mathbf{X}) > U \mid U\big) = \int_{\mathbb{X}} \mathbb{1}_{p_n(\mathbf{x})>U}\mathbf{P}_{\mathbf{X}}(d\mathbf{x}). \tag{3.9}$$

Il est facile de vérifier que R_n est, au même titre que S_n, un estimateur sans biais de l'espérance μ_n donnée en (3.4) :

$$\mathbb{E}[R_n] = \int_{\mathbb{X}} \mathbb{E}[\mathbb{1}_{p_n(\mathbf{x})>U}]\mathbf{P}_{\mathbf{X}}(d\mathbf{x}) = \int_{\mathbb{X}} p_n(\mathbf{x})\mathbf{P}_{\mathbf{X}}(d\mathbf{x}) = \mu_n.$$

Notons que, dans (Oger et al., 2015), les auteurs montrent que la loi uniforme sur $[0,1]$ est la seule loi pour U telle que, pour toute loi $\mathbf{P}_{\mathbf{X}}$ et pour tout choix de distribution pour le processus ξ, on a :

$$\mathbb{E}[R_n] = \mathbb{E}[S_n] = \mu_n.$$

À l'exception de ce résultat d'égalité en moyenne, la relation entre S_n et R_n n'est pas davantage étudiée dans (Oger et al., 2015), les auteurs s'étant attachés à justifier intrinsèquement le choix de R_n pour l'estimation de la probabilité p. À ce stade, les avantages de R_n sur S_n sont précisément les suivants : la loi de R_n ne dépend pas directement de la loi de ξ_n, puisque l'information fournie par la donnée que $\boldsymbol{\xi}_n = \mathbf{g}_n$ est prise en compte à travers la fonction p_n. Cela garantit que seules les lois marginales du processus ξ_n sont impliquées dans la modélisation. En outre, la fonction p_n présente l'intérêt d'être relativement peu coûteuse à évaluer et la variable aléatoire U est unidimensionnelle, quelle que soit la dimension d de $\mathbb{X}$. De plus, si on sait simuler suivant la loi de U, on peut approximativement simuler suivant la loi de R_n à moindre coût. En effet, si on se donne d'une part, un échantillon $(U_j)_{1\leq j\leq M}$ i.i.d. suivant la loi de U et d'autre part, un échantillon $(\mathbf{X}_i)_{1\leq i\leq N}$ i.i.d. suivant $\mathbf{P}_{\mathbf{X}}$, alors on obtient une suite de M variables aléatoires approximativement i.i.d. distribuées suivant la loi de R_n. En effet, on peut considérer que :

$$\int_{\mathbb{X}} \mathbb{1}_{p_n(\mathbf{x})>U_j}\mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \approx \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{p_n(\mathbf{X}_i)>U_j}, \quad \forall j = 1,\ldots,M.$$

3.4 Inégalité d'ordre convexe entre les estimateurs

Tous les résultats qui vont suivre sont originaux et on insiste sur le fait qu'ils sont indépendants du choix de la loi a priori pour le processus ξ.

3.4.1 Comparaison de la dispersion

Commençons par comparer la dispersion des variables aléatoires S_n et R_n, qui ont la même espérance μ_n. Le moment d'ordre $m \in \mathbb{N}^*$ de R_n existe et s'écrit :

$$\begin{aligned}
\mathbb{E}[R_n^m] &= \int_0^1 \left(\int_{\mathbb{X}} \mathbb{1}_{p_n(\mathbf{x})>u}\mathbf{P}_{\mathbf{X}}(d\mathbf{x})\right)^m du \\
&= \int_{\mathbb{X}^m} \left(\int_0^1 \mathbb{1}_{p_n(\mathbf{x}_1)>u}\ldots\mathbb{1}_{p_n(\mathbf{x}_m)>u} du\right)\mathbf{P}_{\mathbf{X}}(d\mathbf{x}_1)\ldots\mathbf{P}_{\mathbf{X}}(d\mathbf{x}_m) \\
&= \int_{\mathbb{X}^m} \min\big(p_n(\mathbf{x}_1),\ldots,p_n(\mathbf{x}_m)\big)\mathbf{P}_{\mathbf{X}}(d\mathbf{x}_1)\ldots\mathbf{P}_{\mathbf{X}}(d\mathbf{x}_m). \qquad (3.10)
\end{aligned}$$

La variance de R_n est donc définie par :

$$\mathrm{Var}[R_n] = \int_{\mathbb{X}^2} \min\big(p_n(\mathbf{x}_1), p_n(\mathbf{x}_2)\big)\mathbf{P}_{\mathbf{X}}(d\mathbf{x}_1)\mathbf{P}_{\mathbf{X}}(d\mathbf{x}_2) - \mu_n^2. \tag{3.11}$$

Les expressions des moments et de la variance de S_n ont déjà été données en section 3.2.2. Puisque $\mathbb{P}(A \cap B) \leq \min\big(\mathbb{P}(A), \mathbb{P}(B)\big)$, on en déduit que :

$$\mathbb{E}[S_n^m] \leq \mathbb{E}[R_n^m], \ \forall m \in \mathbb{N}^*, \quad \text{et} \quad \mathrm{Var}[S_n] \leq \mathrm{Var}[R_n]. \tag{3.12}$$

Ces deux inégalités montrent que la distribution de R_n est plus dispersée que celle de S_n et, par conséquent, que R_n est un estimateur moins performant que S_n en terme de précision. Néanmoins, nous avons constaté les limites de S_n en section 3.2.2 et nous privilégions malgré tout R_n, car il est plus facile de travailler avec cette variable aléatoire. De plus, de notre point de vue, ces inégalités sont intéressantes car elles nous assurent que l'on dispose de majorants des moments et de la variance de S_n. En effet, contrairement à ce que suggère l'équation (3.10), il est très facile d'estimer les moments de R_n et le coût de calcul n'augmente pas avec l'ordre m. En section 3.3, on a expliqué comment simuler une suite de variables aléatoires approximativement i.i.d. suivant la loi de R_n. Il suffit alors de calculer les moments empiriques. Des détails sur la façon de procéder pour l'estimation de la variance de R_n sont donnés en annexe B.

Remarque 3.4.1. *Dans certaines situations, on peut constater que la variance de S_n conduit à des conclusions erronées. En effet, considérons le cas particulier où les variables aléatoires $(\xi_n(\mathbf{x}))_{\mathbf{x}\in\mathbb{X}}$ sont indépendantes et où on a $p_n(\mathbf{x}) = \frac{1}{2}$, $\forall \mathbf{x} \in \mathbb{X}$. Alors, on peut montrer que la loi de S_n est une loi de Dirac en $\frac{1}{2}$ et celle de R_n est une loi de Bernoulli de paramètre $\frac{1}{2}$. En effet, sous ces hypothèses, le moment d'ordre 2 de S_n donné en (3.7) satisfait :*

$$\begin{aligned}
\mathbb{E}[S_n^2] &= \int_{\mathbb{X}^2} \mathbb{P}\big(\xi_n(\mathbf{x}_1) > T, \xi_n(\mathbf{x}_2) > T\big)\mathbf{P}_{\mathbf{X}}(d\mathbf{x}_1)\mathbf{P}_{\mathbf{X}}(d\mathbf{x}_2) \\
&= \left(\int_{\mathbb{X}} \mathbb{P}\big(\xi_n(\mathbf{x}) > T\big)\mathbf{P}_{\mathbf{X}}(d\mathbf{x})\right)^2 \\
&= \mathbb{E}[S_n]^2,
\end{aligned}$$

ce qui implique que $\mathrm{Var}[S_n] = 0$. De plus, $\mathbb{E}[S_n] = \int_{\mathbb{X}} \frac{1}{2}\mathbf{P}_{\mathbf{X}}(d\mathbf{x}) = \frac{1}{2}$, donc la loi de S_n est une loi de Dirac en $\frac{1}{2}$. Concernant la variable R_n, celle-ci s'écrit ici :

$$R_n = \int_{\mathbb{X}} \mathbb{1}_{\frac{1}{2}>U}\mathbf{P}_{\mathbf{X}}(d\mathbf{x}) = \mathbb{1}_{\frac{1}{2}>U}.$$

Puisque U est uniforme sur $[0,1]$, il s'agit bien d'une loi de Bernoulli de paramètre $\frac{1}{2}$. Ainsi, lorsque la modélisation à l'aide du processus aléatoire ξ_n est insuffisante, à savoir, $p_n(\mathbf{x}) = \frac{1}{2}$, $\forall \mathbf{x} \in \mathbb{X}$, la variance de S_n est nulle alors que celle de R_n est maximale. L'incertitude sur le résultat est correctement propagée lorsque l'on a recours à R_n.

	$n = 5$	$n = 20$
Moyenne empirique de S_n	0.010	0.169
Moyenne empirique de R_n	0.010	0.169
Variance empirique de S_n	$1.4 \cdot 10^{-3}$	$9.93 \cdot 10^{-5}$
Variance empirique de R_n	$3.2 \cdot 10^{-3}$	$1.69 \cdot 10^{-4}$

TABLE 3.1 – Estimations de la moyenne et de la variance des variables aléatoires S_n et R_n, en fonction du nombre d'évaluations de la fonction g. La vraie probabilité est $p = 1.66 \times 10^{-1}$.

3.4.2 Exemple

Avant d'aller plus loin, nous proposons de comparer les capacités de prédiction de R_n et S_n sur un exemple en dimension $d = 1$, issu de (Bect et al., 2012), et sur lequel nous nous appuyons tout au long de ce chapitre pour illustrer les différents résultats. On considère la fonction $g : \mathbb{R} \to \mathbb{R}$ définie par :

$$g(\mathbf{x}) = (0.4\mathbf{x} - 0.3)^2 + e^{-11.534|\mathbf{x}|^{1.95}} + e^{-5(\mathbf{x}-0.8)^2}.$$

La probabilité de défaillance est $p = \mathbb{P}(g(\mathbf{X}) > T)$, avec $T = 0.85$ et $\mathbf{P_X} = \mathcal{N}(-\frac{1}{2}, \frac{2}{5})$. Puisqu'ici la fonction g est connue, on sait par intégration numérique que la vraie valeur de p est 1.66×10^{-1}. Le modèle de krigeage que l'on se donne est un processus gaussien avec une fonction moyenne constante et une fonction de corrélation Matérn $\frac{3}{2}$. Tous les paramètres du modèle sont estimés par la méthode du maximum de vraisemblance et on a utilisé pour cela la fonction `"km"` du package R `"DiceKriging"`. À gauche de la figure 3.2, on a représenté la modélisation pour $n = 5$ observations, et à gauche de la figure 3.3, pour $n = 20$ observations. Dans le deuxième cas, les trajectoires du processus aléatoire gaussien conditionné aux observations reproduisent correctement le comportement de la fonction g au voisinage du seuil T. Comme il s'agit d'un exemple en dimension $d = 1$, ces trajectoires peuvent être facilement simulées à l'aide de la fonction `"simulate"` de ce même package R.

Pour l'estimation de p, nous avons effectué 10^4 simulations de trajectoires et généré autant de réalisations de S_n par l'intermédiaire d'une méthode Monte-Carlo naïve, comme décrit à l'équation (3.8). L'échantillon $(\mathbf{X}_i)_{1 \leq i \leq N}$ utilisé pour cela est distribué suivant $\mathbf{P_X}$ et de taille $N = 10^4$. À l'aide de la procédure décrite en section 3.3, on a également simulé 10^4 réalisations de R_n. On a utilisé pour cela le même échantillon $(\mathbf{X}_i)_{1 \leq i \leq N}$ que pour S_n. Le temps de calcul pour obtenir les réalisations de R_n est négligeable devant celui nécessaire pour obtenir les réalisations de S_n. Comme on peut le constater en comparant les distributions empiriques représentées dans les figures 3.2 et 3.3, la variable R_n est plus dispersée que S_n, et la variabilité diminue avec le nombre d'observations. Cela se voit également en comparant les valeurs numériques du tableau 3.1. Comme attendu, les moyennes de deux variables aléatoires sont quasiment égales et fournissent une bonne estimation de la probabilité de défaillance pour $n = 20$.

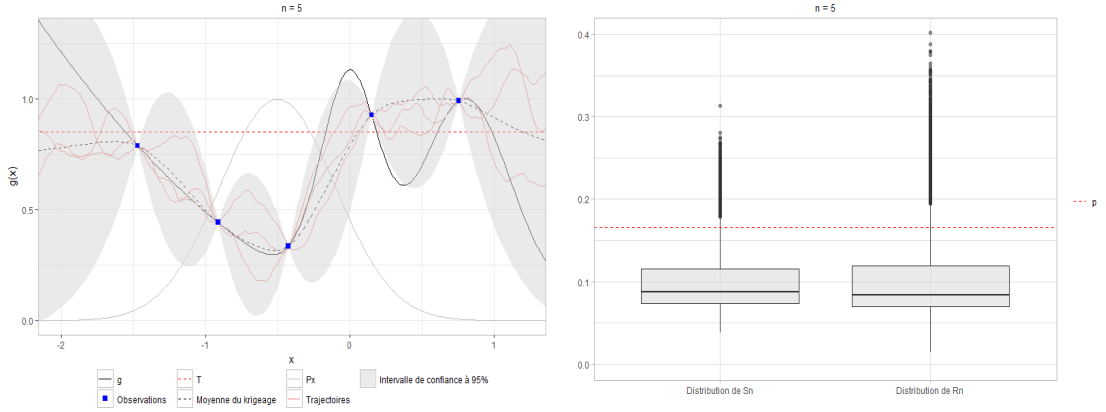

FIGURE 3.2 – À gauche : modèle de processus gaussien pour $n = 5$ évaluations de la fonction g (représentée par la courbe noire). À droite : distributions empiriques de S_n et R_n.

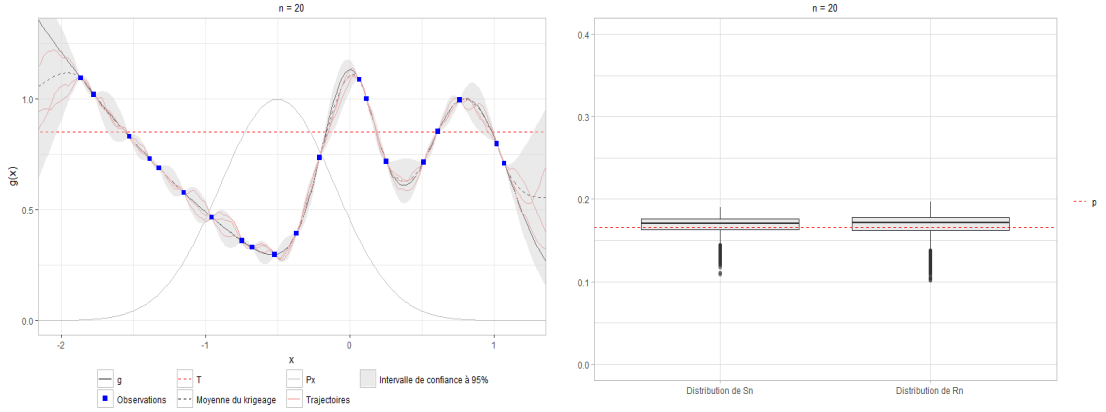

FIGURE 3.3 – À gauche : modèle de processus gaussien pour $n = 20$ évaluations de la fonction g (représentée par la courbe noire). À droite : distributions empiriques de S_n et R_n.

3.4.3 Inégalité d'ordre convexe

En annexe A, nous présentons les ordres stochastiques. Il s'agit d'outils permettant de comparer des variables aléatoires pour aider à la prise de décision. Rappelons la définition d'un ordre stochastique convexe :

Définition 3.4.1. *On dit que X est plus petite que Y au sens de l'ordre convexe, et on note $X \leq_{cx} Y$, si pour toute fonction convexe $\varphi : \mathbb{R} \to \mathbb{R}$ telle que les espérances existent, on a :*

$$\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(Y)]. \tag{3.13}$$

On remarque immédiatement que :

$$X \leq_{cx} Y \quad \Rightarrow \quad \mathbb{E}[X] = \mathbb{E}[Y] \quad \text{et} \quad \text{Var}[X] \leq \text{Var}[Y].$$

Dans la proposition suivante, on montre qu'il existe un ordre convexe entre S_n et R_n.

Proposition 3.4.1. *Soient S_n et R_n les variables aléatoires définies en (3.3) et (3.9), alors :*

$$S_n \leq_{cx} R_n. \tag{3.14}$$

Démonstration. Voir section 3.9. □

Comme nous allons le voir dans les prochaines sections, la relation d'ordre convexe (3.14) peut être exploitée pour apprendre sur la loi de S_n. Il existe un très grand nombre de propriétés relatives à l'ordre convexe qui sont alors vérifiées par S_n et R_n (voir le chapitre 3 de (Shaked and Shanthikumar, 2007) pour une vue d'ensemble de ces propriétés). En annexe A.2.3, on énonce quelques-unes de ces propriétés. En particulier, celles de la proposition A.2.3 sont satisfaites et nous verrons notamment à la section 3.5.1 comment exploiter les propriétés *(iii)* et *(iv)* pour encadrer les quantiles de S_n. Dans un cadre bayésien, avoir accès aux quantiles de la loi a posteriori est une information essentielle pour statuer quant à la précision de l'estimation.

3.5 Applications

Dans cette section, on commence par montrer que l'on a accès aux expressions de la fonction quantile et de la fonction de survie de R_n. Puis on tire profit de la relation d'ordre convexe établie en proposition 3.4.1 pour borner la fonction quantile de S_n. On en déduit des intervalles de crédibilité pour la loi de S_n, permettant de statuer quant à la pertinence de l'espérance de S_n comme estimateur de la probabilité de défaillance.

3.5.1 Encadrement et estimation de quantiles

3.5.1.1 Fonction quantile et fonction de survie

La proposition suivante montre que la fonction quantile de R_n dispose d'une expression explicite.

Proposition 3.5.1. *Soit R_n la variable aléatoire définie à l'équation (3.9) Alors, la fonction quantile $F_{R_n}^{-1}$ de R_n vérifie :*

$$F_{R_n}^{-1}(\alpha) = \int_{\mathbb{X}} \mathbb{1}_{p_n(\mathbf{x})>1-\alpha} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}), \quad \forall \alpha \in [0,1].$$

Démonstration. Voir section 3.9. □

Pour calculer la fonction de survie G_{R_n} de R_n, on pourra utiliser le fait que, pour tout $t \in [0,1]$, celle-ci vérifie :

$$G_{R_n}(t) = \mathbb{P}(R_n > t) = \inf\{\alpha \in [0,1] : F_{R_n}^{-1}(1-\alpha) \leq t\}. \tag{3.15}$$

En annexe B, on explique comment estimer point par point ces deux fonctions. La complexité de calcul de ces méthodes n'augmente pas avec la dimension d de $\mathbb{X}$ car l'estimation est faite par une méthode Monte-Carlo naïve.

3.5.1.2 Nouvelles bornes

Soit $\alpha \in (0,1)$ et $F_{S_n}^{-1}(\alpha)$ le quantile de S_n d'ordre α. On cherche à obtenir une approximation de $F_{S_n}^{-1}(\alpha)$. Dans (Auffray et al., 2014), les auteurs proposent d'utiliser pour cela les inégalités de Markov et de Bienaymé-Tchebychev, que l'on rappelle en annexe A.3. Les majorations obtenues s'écrivent :

$$F_{S_n}^{-1}(\alpha) \leq \frac{\mu_n}{1-\alpha} \quad \text{et} \quad F_{S_n}^{-1}(\alpha) \leq \mu_n + \sqrt{\frac{\mathrm{Var}[S_n]}{1-\alpha}}. \tag{3.16}$$

En section 3.2.1, on a expliqué comment estimer à moindre coût la moyenne μ_n par une méthode Monte-Carlo naïve, de sorte que la borne de gauche dans (3.16) est très facile à estimer. Par contre – et cela est souligné dans (Auffray et al., 2014) – la méthode Monte-Carlo pour l'estimation de la variance de S_n nécessite des temps de calculs relativement élevés (on a déjà fait cette remarque en section 3.2.2). Une façon simple de contourner le problème est de majorer la variance de S_n par la variance de R_n dans (3.16) :

$$F_{S_n}^{-1}(\alpha) \leq \mu_n + \sqrt{\frac{\mathrm{Var}[R_n]}{1-\alpha}}. \tag{3.17}$$

La variance de R_n est moins coûteuse à estimer que celle de S_n (on rappelle qu'une méthode facile à mettre en œuvre pour y parvenir est proposée en annexe B). Néanmoins, on doit reconnaître que la borne (3.17) basée sur la variance de R_n offre un moins bon contrôle du quantile que la borne (3.16) basée sur la variance de S_n. C'est pourquoi, afin de réaliser un compromis entre précision et complexité de calcul, on propose l'encadrement suivant.

Proposition 3.5.2. *Pour tout $\alpha \in (0,1)$, le quantile $F_{S_n}^{-1}(\alpha)$ satisfait :*

$$\delta_n^-(\alpha) \leq \gamma_n^-(\alpha) \leq F_{S_n}^{-1}(\alpha) \leq \gamma_n^+(\alpha) \leq \delta_n^+(\alpha), \tag{3.18}$$

où :

$$\delta_n^-(\alpha) = \frac{\mu_n + \alpha - 1}{\alpha} \quad \textit{et} \quad \delta_n^+(\alpha) = \frac{\mu_n}{1-\alpha},$$

et

$$\gamma_n^-(\alpha) = \frac{1}{\alpha}\int_0^\alpha F_{R_n}^{-1}(t)dt \quad et \quad \gamma_n^+(\alpha) = \frac{1}{1-\alpha}\int_\alpha^1 F_{R_n}^{-1}(t)dt.$$

Démonstration. Voir section 3.9. □

Les bornes $\gamma_n^-(\alpha)$ et $\gamma_n^+(\alpha)$ offrent un meilleur contrôle de $F_{S_n}^{-1}(\alpha)$. Dans la démonstration, on prouve qu'elles sont obtenues grâce à l'inégalité d'ordre convexe (3.14) entre S_n et R_n. Nous les appellerons donc « bornes issues de l'ordre convexe ». Les bornes $\delta_n^-(\alpha)$ et $\delta_n^+(\alpha)$ découlent des bornes issues de l'ordre convexe, par minoration pour la première et par majoration pour la seconde. Néanmoins, la borne supérieure $\delta_n^+(\alpha)$ faisant référence à l'inégalité de Markov (3.16), nous les appellerons « bornes de Markov ». En outre, il est facile de vérifier que la borne inférieure $\delta_n^-(\alpha)$ est une borne améliorée de celle que l'on obtient avec l'inégalité de Markov (voir équation (A.12) en annexe A.3).

En pratique, on doit prendre $\max(0, \delta_n^-(\alpha))$ et $\min(1, \delta_n^+(\alpha))$, car les bornes de Markov ne sont pas systématiquement informatives. En effet, si $\mu_n \leq 1-\alpha$ (resp. $\mu_n \geq 1-\alpha$), alors on a $\delta_n^-(\alpha) \leq 0 \leq F_{S_n}^{-1}(\alpha)$ (resp. $F_{S_n}^{-1}(\alpha) \leq 1 \leq \delta_n^+(\alpha)$). Par contre, il est facile de vérifier que ce n'est pas le cas pour les bornes issues de l'ordre convexe, car elles sont nécessairement à valeurs dans $(0,1)$. De plus, on montre dans la proposition suivante que ces bornes peuvent s'écrire comme une intégrale par rapport à la loi $\mathbf{P_X}$.

Proposition 3.5.3. *Pour tout $\alpha \in (0,1)$, considérons les bornes $\gamma_n^-(\alpha)$ et $\gamma_n^+(\alpha)$ données à la proposition 3.5.2. On a alors :*

(i) $\gamma_n^-(\alpha) = 1 - \displaystyle\int_{\mathbb{X}} \min\left(1, \frac{1-p_n(\mathbf{x})}{\alpha}\right) \mathbf{P_X}(d\mathbf{x})$.

(ii) $\gamma_n^+(\alpha) = \displaystyle\int_{\mathbb{X}} \min\left(1, \frac{p_n(\mathbf{x})}{1-\alpha}\right) \mathbf{P_X}(d\mathbf{x})$.

Démonstration. Voir section 3.9. □

En pratique, il n'y a donc aucune difficulté majeure à estimer les bornes issues de l'ordre convexe et cela nécessite la même complexité de calcul que celle nécessaire à l'estimation des bornes de Markov. Il suffit d'appliquer une méthode Monte-Carlo naïve, de la même façon que pour l'estimation de $\mu_n = \int_{\mathbb{X}} p_n(\mathbf{x})\mathbf{P_X}(d\mathbf{x})$, comme décrit en section 3.2.1. Notons que, pour comparer les performances de ces différentes bornes, un exemple sera proposé en section 3.5.3.

Remarque 3.5.1. *On peut remarquer que les bornes issues de l'ordre convexe réfèrent à la mesure de risque appelée « Tail-Value-at-Risk », dont la définition et les propriétés sont données en annexe A.1.3 :*

$$\gamma_n^-(\alpha) = \mathrm{TVaR}_\alpha^-[R_n], \quad \text{et} \quad \gamma_n^+(\alpha) = \mathrm{TVaR}_\alpha^+[R_n]. \tag{3.19}$$

Dans (Brazauskas et al., 2008), les auteurs expliquent que la façon usuelle d'estimer la Tail-Value-at-Risk de niveau $\alpha \in [0,1)$ d'une variable aléatoire X est de procéder par

approche non-paramétrique. Typiquement, il s'agit de simuler un échantillon $X_1, \dots, X_N$ i.i.d. suivant la loi de X et de considérer l'estimateur :

$$\mathrm{TVaR}_\alpha^+[X] \approx \frac{1}{1-\alpha}\int_\alpha^1 F_{X,N}^{-1}(t)dt,$$

où $F_{X,N}^{-1}(t)$ est le quantile de la distribution empirique. Ici, on ne sait pas exactement simuler suivant la loi de R_n. Cependant, grâce aux expressions données en proposition 3.5.3, il est très facile d'estimer les bornes données en équations (3.19). Cela ne nécessite pas de simuler suivant la loi de R_n car elles s'expriment comme une intégrale par rapport à la loi $\mathbf{P_X}$.

Remarque 3.5.2. *Dans (Meilijson and Nadas, 1979), le chapitre 4 de (Hürlimann, 1999) ou encore (Hürlimann, 2003), il est dit que la transformée de Hardy-Littlewood d'une variable aléatoire X est la variable aléatoire notée X^H, dont la fonction quantile $F_{X^H}^{-1}$ est définie pour tout $\alpha \in [0,1]$ par :*

$$F_{X^H}^{-1}(\alpha) = \begin{cases} \dfrac{1}{1-\alpha}\displaystyle\int_\alpha^1 F_X^{-1}(t)dt, & si\ \alpha < 1, \\ \\ F_X^{-1}(1), & si\ \alpha = 1. \end{cases}$$

De plus, si $S_X = \{Y : Y \leq_{icx} X\}$ est l'ensemble de toutes les variables aléatoires plus petites que X au sens de l'ordre convexe croissant, alors le plus petit majorant Z vérifiant :

$$Y \leq_{st} Z, \quad \forall\ Y \in S_X$$

est la transformée de Hardy-Littlewood X^H de X.

Dans notre contexte, puisque $S_n \leq_{cx} R_n \Rightarrow S_n \leq_{icx} R_n$, cela signifie qu'il existe une variable aléatoire notée R_n^H telle que :

$$S_n \leq_{st} R_n^H.$$

La variable R_n^H vérifie toutes les propriétés de dominance stochastique à l'ordre 1 présentées annexe A.2.1. En particulier, sa fonction quantile, qui est la Tail-Value-at-Risk de R_n, majore la fonction quantile de S_n en tout point. Ceci a effectivement été vérifié en proposition 3.5.2. En outre, d'après le théorème 2.3 de (Hürlimann, 1999), on a :

$$R_n^H = R_n + m_{R_n}(R_n),$$

où $m_{R_n}(x) = \mathbb{E}[R_n - x | R_n > x] = \dfrac{1}{G_{R_n}(x)}\displaystyle\int_x^1 G_{R_n}(t)dt$, *et G_{R_n} est la fonction de survie de R_n donnée en équation (3.15).*

3.5.2 Construction et estimation d'intervalles de crédibilité

3.5.2.1 Définition

La notion d'intervalle de crédibilité (ou encore de région de crédibilité) intervient dans le cadre de l'estimation bayésienne et fait référence aux valeurs probables d'une variable aléatoire au vu de sa loi a posteriori. Une définition formelle peut, par exemple, être retrouvée dans (Robert, 2007). Dans notre contexte, celle-ci s'écrit :

Définition 3.5.1. *Soit $\alpha \in [0,1]$ fixé et S la variable aléatoire donnée en (3.2). On appelle intervalle de crédibilité de niveau $1-\alpha$ pour la loi a posteriori de S tout intervalle $[a_\alpha, b_\alpha] \subseteq [0,1]$ vérifiant :*

$$\mathbb{P}(a_\alpha \leq S \leq b_\alpha \mid \boldsymbol{\xi}_n = \mathbf{g}_n) \geq 1 - \alpha.$$

Puisque $\mathcal{L}(S \mid \boldsymbol{\xi}_n = \mathbf{g}_n) = \mathcal{L}(S_n)$, notre objectif est de déterminer, pour une valeur de $\alpha \in [0,1]$ fixé, un intervalle $[a_\alpha, b_\alpha] \subseteq [0,1]$ vérifiant :

$$\mathbb{P}(a_\alpha \leq S_n \leq b_\alpha) \geq 1 - \alpha. \tag{3.20}$$

On souhaite évidement que cet intervalle puisse être facilement estimé, mais aussi qu'il ne soit pas trop large afin d'apporter des informations pertinentes sur les valeurs probables de S_n. En effet, pour une valeur de $\alpha \in [0,1]$, il existe une infinité d'intervalles $[a_\alpha, b_\alpha]$ vérifiant (3.20).

3.5.2.2 Construction et estimation

Soit $\alpha \in (0,1)$ fixé et $F_{S_n}^{-1}(\alpha)$ le quantile de S_n de niveau α. Considérons les bornes issues de l'ordre convexe $\gamma_n^-(\alpha)$ et $\gamma_n^+(\alpha)$ introduites à la section précédente. D'après la proposition 3.5.2, elles vérifient :

$$\gamma_n^-(\alpha) \leq F_{S_n}^{-1}(\alpha) \leq \gamma_n^+(\alpha).$$

Puisque, par définition, $F_{S_n}^{-1}(\alpha) = \inf\{t \in [0,1] : \mathbb{P}(S_n \leq t) \geq \alpha\}$, on peut écrire :

$$\mathbb{P}\big(S_n \leq \gamma_n^-(\alpha\beta)\big) \leq \alpha\beta, \quad \forall \beta \in (0,1),$$

et

$$\mathbb{P}\big(S_n \leq \gamma_n^+(1-\alpha(1-\beta))\big) \geq 1-\alpha(1-\beta), \quad \forall \beta \in (0,1).$$

Autrement dit,

$$\mathbb{P}\big(\gamma_n^-(\alpha\beta) \leq S_n \leq \gamma_n^+(1-\alpha(1-\beta))\big) \geq 1-\alpha, \quad \forall \beta \in (0,1).$$

La proposition suivante récapitule ce qui vient d'être dit et utilise les expressions des bornes $\gamma_n^-(\alpha)$ et $\gamma_n^+(\alpha)$ données à la proposition 3.5.3.

Proposition 3.5.4. *Soit $\alpha \in (0,1)$ fixé. Pour tout $\beta \in (0,1)$, l'intervalle $I_n^{cx}(\alpha,\beta)$ défini par :*

$$I_n^{cx}(\alpha,\beta) \quad = \left[1 - \int_{\mathbb{X}} \min\left(1, \frac{1-p_n(\mathbf{x})}{\alpha\beta}\right) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \quad , \quad \int_{\mathbb{X}} \min\left(1, \frac{p_n(\mathbf{x})}{\alpha(1-\beta)}\right) \mathbf{P}_{\mathbf{X}}(d\mathbf{x})\right], \tag{3.21}$$

est un intervalle de crédibilité de niveau $1-\alpha$ pour la loi de S_n, c'est-à-dire qu'il vérifie :

$$\mathbb{P}\big(S_n \in I_n^{cx}(\alpha,\beta)\big) \geq 1-\alpha, \quad \forall \beta \in (0,1).$$

$n = 10$	$\mathbb{E}[S_n] = 0.148$	$\mathbb{P}\big(S_n \in [0.039; 0.383]\big) \geq 0.95$
$n = 20$	$\mathbb{E}[S_n] = 0.169$	$\mathbb{P}\big(S_n \in [0.128; 0.188]\big) \geq 0.95$
$n = 30$	$\mathbb{E}[S_n] = 0.166$	$\mathbb{P}\big(S_n \in [0.160; 0.175]\big) \geq 0.95$
$n = 40$	$\mathbb{E}[S_n] = 0.166$	$\mathbb{P}\big(S_n \in [0.163; 0.169]\big) \geq 0.95$

TABLE 3.2 – Intervalles de crédibilité à 95% en fonction du nombre n d'appels à la fonction g.

Pour une valeur de α fixée dans $(0, 1)$, la longueur de l'intervalle $I_n(\alpha, \beta)$ dépend de la valeur du paramètre $\beta \in (0, 1)$. C'est pourquoi on suggère d'utiliser en pratique un algorithme classique d'optimisation qui détermine la valeur de β pour laquelle la longueur de $I_n^{cx}(\alpha, \beta)$ est minimale. Sinon, on peut simplement prendre, par exemple, $\beta = \frac{1}{2}$. En outre, on peut vérifier de la même façon que l'intervalle $I_n^{Markov}(\alpha, \beta)$ défini pour tout $\beta \in (0, 1)$ par :

$$I_n^{Markov}(\alpha, \beta) = \left[\frac{\mu_n + \alpha\beta - 1}{\alpha\beta} \quad , \quad \frac{\mu_n}{\alpha(1-\beta)} \right], \tag{3.22}$$

est aussi un intervalle de crédibilité de niveau $1-\alpha$ pour la loi de S_n. D'après la proposition 3.5.2, il est néanmoins plus large que $I_n^{cx}(\alpha, \beta)$.

3.5.3 Exemple

Reprenons l'exemple traité en section 3.4.2 – avec $n = 20$ évaluations de la fonction g – et comparons les performances des bornes issues des inégalités de Markov, de Bienaymé-Tchebychev et de l'ordre convexe pour l'approximation de la fonction quantile de S_n.

Puisque nous avons précédemment calculé la distribution empirique de S_n, on peut estimer sa fonction quantile (par exemple, en remplaçant la fonction de répartition par la fonction de répartition empirique dans la définition de $F_{S_n}^{-1}$). Cette fonction (estimée) est représentée en rouge dans la figure 3.4, et on utilise les bornes de la proposition 3.5.2 pour l'approcher. L'aire grise la plus foncée (la plus resserrée) correspond à l'information fournie par les bornes issues de l'ordre convexe, tandis que l'aire grise la moins foncée (la plus large) correspond à l'information fournie par les bornes de Markov. La courbe en pointillés noirs (en haut) correspond à l'inégalité de Bienaymé-Tchebychev appliquée à la variable R_n (voir (3.17)) et la courbe en tirets noirs (en bas) correspond à l'inégalité de Bienaymé-Tchebychev appliquée la variable S_n (voir (3.16)). On précise que toutes les bornes sont estimées par une méthode Monte-Carlo naïve, avec le même échantillon i.i.d. suivant $\mathbf{P_X}$ de taille $N = 10^4$. Au vu des résultats de la figure 3.4, les bornes de Markov sont très peu informatives, tandis que celles issues de l'ordre convexe offrent un encadrement relativement précis de la fonction $F_{S_n}^{-1}$.

À la figure 3.5.3, on donne des estimations de l'intervalle de crédibilité $I_n^{cx}(\alpha, \beta)$ donné en (3.21) pour $\alpha = 0.05$ et $\beta = \frac{1}{2}$. On constate qu'en augmentant progressivement le nombre n d'évaluations de la fonction g, l'intervalle de crédibilité est de plus en plus resserré. En outre, il contient systématiquement la vraie valeur de la probabilité de défaillance (ici $p = 1.66 \times 10^{-1}$).

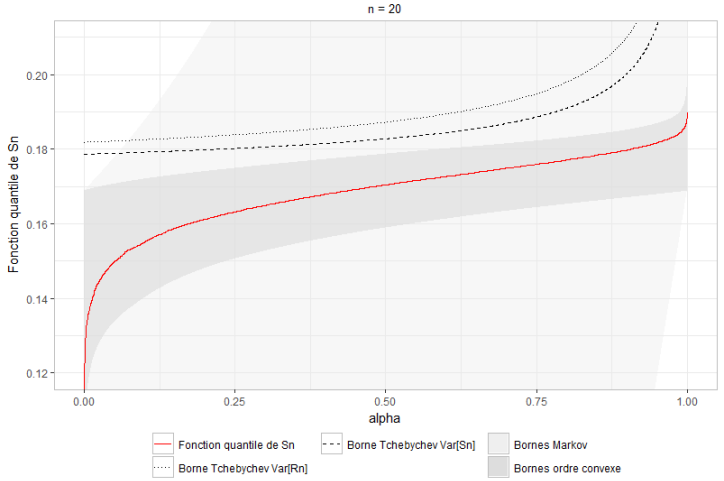

FIGURE 3.4 – La vraie probabilité de défaillance est $p = 1.66 \times 10^{-1}$. On a évalué $n = 20$ fois la fonction g (voir figure 3.3). La courbe rouge est une estimation de la fonction quantile de S_n et on compare les bornes issues des inégalités de Markov (aire en gris clair), de Bienaymé-Tchebychev (courbe en pointillés noirs) et de l'ordre convexe (aire en gris foncé) pour l'approximation de cette fonction.

3.5.4 Extension aux mesures de risque spectrales

Comme mentionné en annexe A.1.4, un grand nombre de mesures de risque peuvent s'écrire comme une moyenne pondérée de la fonction quantile (c'est le cas, par exemple, de l'espérance ou de la Tail-Value-at-Risk). Elles sont appelées « mesures de risque spectrales » et vérifient la propriété de cohérence (voir (Artzner et al., 1997) et (Artzner et al., 1999)). Dans notre contexte, une mesure de risque spectrale pour S_n s'écrit :

$$\rho_\Psi[S_n] = \int_0^1 F_{S_n}^{-1}(t)\Psi(t)dt, \tag{3.23}$$

où $\Psi : [0,1] \to \mathbb{R}^+$ est une fonction croissante et d'intégrale 1. Comme mentionné dans (Acerbi, 2002), l'approche (non-paramétrique) usuelle pour estimer (3.23) est de simuler un grand nombre de réalisations de la variable aléatoire S_n et de remplacer $F_{S_n}^{-1}(t)$ par le quantile empirique de niveau t. Ici, cette méthode d'estimation n'est évidemment pas applicable compte tenu du coût engendré par la simulation de réalisations de S_n (voir les remarques faites en section 3.2.2). À nouveau, on propose de contourner le problème en construisant des majorants de (3.23). Les résultats que nous avons obtenus sont donnés dans la proposition ci-dessous.

Proposition 3.5.5. *Soit $\Psi : [0,1] \to \mathbb{R}_+$ une fonction spectrale avec :*

$$M = \sup_{t\in[0,1]} \Psi(t) = \Psi(1).$$

La mesure spectrale ρ_Ψ de S_n vérifie :

$$\rho_\Psi(S_n) \leq \rho_\Psi(R_n) \leq \min\left(1, M\mu_n\right), \tag{3.24}$$

où $\rho_\Psi(R_n) = \displaystyle\int_0^1 F_{R_n}^{-1}(t)\Psi(t)dt.$

Démonstration. Voir section 3.9. □

Étant donné ce qui précède, il n'y a aucune difficulté pratique à estimer les majorants dans (3.24). En outre, on a le corollaire suivant.

Corollaire 3.5.1. *Soit* $\Psi : [0,1] \to \mathbb{R}_+$ *une fonction spectrale. La mesure spectrale* ρ_Ψ *de* S_n *vérifie :*

$$\rho_\Psi(S_n) \leq \int_{\mathbb{X}} \left[\int_0^{p_n(\mathbf{x})} \Psi(1-t)dt \right] \mathbf{P}_{\mathbf{X}}(d\mathbf{x}).$$

Démonstration. Voir section 3.9. □

Le résultat du corollaire 3.5.1 signifie que l'on dispose d'un majorant de $\rho_\Psi(S_n)$ qui s'exprime comme une intégrale par rapport à la loi de $\mathbf{X}$ et qui, par conséquent, peut être facilement estimé par une méthode Monte-Carlo naïve.

Remarque 3.5.3. *On précise que nous n'avons pas réussi à établir une minoration de la mesure spectrale* $\rho_\Psi(S_n)$ *qui soit valable pour toute fonction spectrale* Ψ *telle que définie ci-dessus. Le problème de la détermination d'une borne inférieure pour* $\rho_\Psi(S_n)$ *reste donc ouvert.*

3.6 Planification d'expériences séquentielle et stratégies SUR

Dans le cas où l'intervalle de crédibilité de la loi de S_n (voir proposition 3.5.4) est jugé trop grand pour permettre de considérer la valeur moyenne comme une estimation fiable de la probabilité de défaillance p, on conclut que le modèle de krigeage n'est pas suffisamment renseigné. Pour améliorer la prédiction, il est nécessaire de rajouter de l'information au modèle, c'est-à-dire fournir des nouvelles données. Dans ce but, on propose de mettre en œuvre une méthode de planification d'expériences séquentielle qui repose sur le principe des stratégies SUR (« Stepwise Uncertainty Reduction »), que nous avons très brièvement introduites en section 2.4.2.2. Dans (Bect et al., 2012), les auteurs ont formalisé, dans le cadre de la modélisation par processus gaussien, le principe des stratégies SUR pour l'estimation d'une probabilité de défaillance. Pour faciliter la compréhension, nous simplifions leur formalisme et l'adaptons au cadre de notre étude. Pour une présentation plus générale des stratégies SUR, on pourra évidemment consulter (Bect et al., 2012), mais aussi la thèse (Chevalier, 2013), ainsi que les références qui s'y trouvent. En outre, on précise que des résultats théoriques justifiant la performance de ces méthodes ont récemment été proposés dans (Bect et al., 2019).

Pour plus de clarté, le processus ξ est toujours choisi gaussien, bien que les résultats ci-dessous restent valides pour n'importe quelle loi.

3.6.1 Critères d'échantillonnage

3.6.1.1 Minimisation d'un critère

La situation dans laquelle on se place est celle où g est observée aux points du plan d'expériences $\mathbf{D}_n = (\mathbf{x}_i)_{1\leq i\leq n}$, et où l'on cherche un point supplémentaire $\mathbf{x}_{n+1} \in \mathbb{X}\backslash\mathbf{D}_n$ en lequel évaluer la fonction afin de se constituer un ensemble de $n+1$ observations. La procédure de sélection de points étant itérative, l'information dont on dispose sur la fonction g évolue au fur et à mesure que l'on ajoute des points au plans d'expériences. C'est pourquoi on munit ici l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ de la filtration $(\mathcal{F}_n)_{n\in\mathbb{N}^*}$, définie pour tout $n \in \mathbb{N}^*$ comme la tribu engendrée par les variables aléatoires $\xi(\mathbf{x}_1), \ldots, \xi(\mathbf{x}_n)$:

$$\mathcal{F}_n = \sigma(\xi(\mathbf{x}_1), \ldots, \xi(\mathbf{x}_n)).$$

Dans ces conditions, ξ_n est le processus aléatoire de mêmes lois fini-dimensionnelles que le processus ξ sachant $\mathcal{F}_n$, la variable aléatoire S_n est la variable distribuée suivant la loi de S sachant $\mathcal{F}_n$, et R_n est la variable aléatoire définie par :

$$R_n = \int_{\mathbb{X}} \mathbb{1}_{p_n(\mathbf{x})>U} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}),$$

où $p_n(\mathbf{x}) = \mathbb{P}(\xi(\mathbf{x}) > T \mid \mathcal{F}_n) = \mathbb{P}(\xi_n(\mathbf{x}) > T)$, $\forall \mathbf{x} \in \mathbb{X}$. Par ailleurs, on adopte la notation suivante :

$$\mathbb{E}_n[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_n], \quad \forall n \in \mathbb{N}^*.$$

Étant donné ce qui précède, une stratégie idéale serait de déterminer le point $\mathbf{x}^*_{n+1}$ qui minimise la variance a posteriori de la variable aléatoire S sachant $n+1$ observations, c'est-à-dire $\mathrm{Var}[S_{n+1}]$. Cependant, en tout point $\mathbf{x}_{n+1} \in \mathbb{X} \setminus \mathbf{D}_n$, la variable aléatoire $\xi(\mathbf{x}_{n+1})$ n'est pas encore observée et, par conséquent, la variance $\mathrm{Var}[S_{n+1}]$ est à ce stade une variable aléatoire qui dépend de $\xi(\mathbf{x}_{n+1})$. La seule information à disposition sur cette dernière est qu'il s'agit d'une variable aléatoire vérifiant :

$$\mathcal{L}(\xi(\mathbf{x}_{n+1}) \mid \mathcal{F}_n) = \mathcal{N}(m_n(\mathbf{x}_{n+1}), \sigma^2_n(\mathbf{x}_{n+1})),$$

où l'espérance $m_n(\mathbf{x})$ est donnée en (2.18) et la variance $\sigma^2_n(\mathbf{x})$ est donnée en (2.17). Par conséquent, il paraît légitime de sélectionner le point $\mathbf{x}^*_{n+1}$ qui vérifie :

$$\mathbf{x}^*_{n+1} = \underset{\mathbf{x}_{n+1}\in\mathbb{X}}{\operatorname{argmin}} \mathbb{E}_n\big[\mathrm{Var}[S_{n+1}] \mid \mathbf{X}_{n+1} = \mathbf{x}_{n+1}\big].$$

Ceci conduit à définir le critère d'échantillonnage suivant, que l'on note J_{S_n}, et dont la minimisation permet d'identifier où effectuer la prochaine évaluation de g :

$$J_{S_n}(\mathbf{x}_{n+1}) = \mathbb{E}_n\big[\mathrm{Var}[S_{n+1}] \mid \mathbf{X}_{n+1} = \mathbf{x}_{n+1}\big].$$

Remarque 3.6.1. *La stratégie décrite ci-dessus vise à enrichir le plan d'expériences en ajoutant les points un par un. En anglais, on parle de « 1-step lookahead strategy ». Dans (Chevalier et al., 2014), les auteurs montrent que l'on peut développer des stratégies, appelées « q-step lookahead strategies », qui permettent d'ajouter $q \in \mathbb{N}^*$ points en même temps. Malgré son intérêt, cette problématique n'est pas abordée dans cette thèse.*

3.6.1.2 Critères alternatifs

En pratique, le critère J_{S_n} est difficile à calculer sans que cela ne requiert des temps de calculs élevés (voir les remarques sur la variance de S_n faites en section 3.2.2). Dans (Bect et al., 2012), les auteurs reconnaissent eux-mêmes les limites de cette approche et proposent des critères alternatifs. Pour tout $\mathbf{x} \in \mathbb{X}$, posons :

$$\tau_n(\mathbf{x}) = \min(p_n(\mathbf{x}), 1 - p_n(\mathbf{x})), \quad \text{et} \quad \nu_n(\mathbf{x}) = p_n(\mathbf{x})(1 - p_n(\mathbf{x})).$$

Les critères alternatifs proposés dans (Bect et al., 2012) sont les suivants :

$$J_{1,n}(\mathbf{x}_{n+1}) = \mathbb{E}_n\left[\left(\int_{\mathbb{X}} \tau_{n+1}(\mathbf{x})^{\frac{1}{2}} \mathbf{P}_{\mathbf{X}}(d\mathbf{x})\right)^2 \,\middle|\, \mathbf{X}_{n+1} = \mathbf{x}_{n+1}\right], \tag{3.25}$$

$$J_{2,n}(\mathbf{x}_{n+1}) = \mathbb{E}_n\left[\left(\int_{\mathbb{X}} \nu_{n+1}(\mathbf{x})^{\frac{1}{2}} \mathbf{P}_{\mathbf{X}}(d\mathbf{x})\right)^2 \,\middle|\, \mathbf{X}_{n+1} = \mathbf{x}_{n+1}\right], \tag{3.26}$$

$$J_{3,n}(\mathbf{x}_{n+1}) = \mathbb{E}_n\left[\int_{\mathbb{X}} \tau_{n+1}(\mathbf{x}) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \,\middle|\, \mathbf{X}_{n+1} = \mathbf{x}_{n+1}\right], \tag{3.27}$$

$$J_{4,n}(\mathbf{x}_{n+1}) = \mathbb{E}_n\left[\int_{\mathbb{X}} \nu_{n+1}(\mathbf{x}) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \,\middle|\, \mathbf{X}_{n+1} = \mathbf{x}_{n+1}\right]. \tag{3.28}$$

Ces critères sont plus faciles à calculer que J_{S_n} car ils s'expriment en fonction d'une seule intégrale par rapport à la loi $\mathbf{P}_{\mathbf{X}}$. Pour les obtenir, la démonstration proposée dans (Bect et al., 2012) repose sur une majoration de la fonction J_{S_n}, à travers des applications successives de l'inégalité de Minkowski généralisée (voir (Vestrup, 2003)) et de l'inégalité de Jensen. Nous verrons en section 3.6.2.3 comment les estimer facilement en pratique.

Remarque 3.6.2. *Rappelons que, dans le cadre de la modélisation par processus gaussien, la fonction moyenne m_n du processus ξ_n est parfois utilisée comme modèle déterministe d'interpolation de la fonction g (voir section 2.3.3). C'est pourquoi, une approximation de la probabilité de défaillance p est aussi parfois donnée par la quantité :*

$$\int_{\mathbb{X}} \mathbb{1}_{m_n(\mathbf{x})>T} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}), \tag{3.29}$$

qui peut facilement être estimée par une méthode Monte-Carlo naïve (voir (Picheny et al., 2010), (Echard et al., 2011) et (Dubourg et al., 2013) pour des applications avec cet estimateur). Comme démontré dans (Bect et al., 2012), les critères $J_{1,n}$ et $J_{3,n}$ réfèrent à la quantité (3.29) pour l'estimation de p, tandis que les critères $J_{2,n}$ et $J_{4,n}$ réfèrent à l'espérance μ_n de S_n donnée en (3.4) et dont l'intérêt à été justifié en section 3.2.

3.6.2 Stratégie SUR basée sur l'ordre convexe

3.6.2.1 Un nouveau critère

On rappelle que les variables aléatoires S_n et R_n vérifient la relation d'ordre convexe (3.14) et que cela implique que :

$$\mathrm{Var}[S_n] \leq \mathrm{Var}[R_n], \quad \forall n \in \mathbb{N}^*.$$

Étant donnée cette inégalité, la stratégie alternative que l'on propose consiste à trouver le point $\mathbf{x}_{n+1}^*$ tel que :

$$\mathbf{x}_{n+1}^* = \underset{\mathbf{x}_{n+1}\in\mathbb{X}}{\operatorname{argmin}}\, J_{R_n}(\mathbf{x}_{n+1}),$$

où

$$J_{R_n}(\mathbf{x}_{n+1}) = \mathbb{E}_n\big[\operatorname{Var}[R_{n+1}] \mid \mathbf{X}_{n+1} = \mathbf{x}_{n+1}\big]. \tag{3.30}$$

En section 3.6.2.3, nous verrons comment estimer facilement ce critère, les calculs étant équivalents, en terme de complexité algorithmique, à ceux nécessaires pour l'estimation des critères $(J_{k,n})_{1\leq k\leq 4}$. Concernant les applications, on pourra directement consulter les exemples proposés en section 3.7. Avant cela, on propose de discuter de la pertinence du critère J_{R_n}.

Remarque 3.6.3. *Comme mentionné dans (Bect et al., 2012), la quantité $Var[S_{n+1}]$ réfère, dans le cadre bayésien usuel, au risque de Bayes pour la fonction de perte quadratique. Il s'en suit immédiatement, d'après l'inégalité d'ordre convexe (3.14), que l'on peut fournir un majorant de cette mesure d'incertitude pour toute fonction de perte convexe. Ainsi, en substituant S_n par R_n, le champ d'application des stratégies SUR peut significativement être étendu.*

3.6.2.2 Comparaison de critères

Commençons par souligner ce fait évident : la fonction J_{R_n} est toujours au-dessus de la fonction J_{S_n}. En outre, on peut montrer qu'elle est localement plus proche que les fonctions $(J_{k,n})_{1\leq k\leq 4}$.

Proposition 3.6.1. *Pour tout $k = 1, \ldots, 4$, les critères J_{S_n}, J_{R_n} et $J_{k,n}$ satisfont :*

$$J_{S_n}(\mathbf{x}) \leq J_{R_n}(\mathbf{x}) \leq J_{k,n}(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{X}. \tag{3.31}$$

Démonstration. Voir section 3.9. □

La relation d'ordre (3.31) signifie simplement que si l'on souhaite une approximation ponctuelle de J_{S_n}, alors la fonction J_{R_n} est à privilégier. Néanmoins, on reconnaît que cela ne garantit pas que l'approximation du minimum de J_{S_n} est meilleure en considérant J_{R_n} plutôt que $(J_{k,n})_{1\leq k\leq 4}$. Par ailleurs, il est facile de vérifier que les fonctions τ_n et ν_n sont maximales quand $p_n = \frac{1}{2}$. Cela correspond à une situation où le modèle de krigeage est non-informatif. Par ailleurs, ces fonctions sont nulles lorsque $p_n = 0$ ou $p_n = 1$, et il s'agit d'une situation saine où le modèle de krigeage est suffisamment renseigné. Ainsi, on en conclut que les critères $(J_{k,n})_{1\leq k\leq 4}$ font intervenir des fonctions qui donnent de l'information sur les sous-domaines de $\mathbb{X}$ où le modèle de krigeage doit être amélioré. Nous allons montrer que l'on peut interpréter le critère J_{R_n} de la même façon. Dans tout ce qui suit, on fait l'hypothèse que la loi de la variable aléatoire $p_n(\mathbf{X})$ est continue et admet une densité.

Proposition 3.6.2. *Soit $n \in \mathbb{N}^*$. La variance de R_n est donnée par :*

$$\mathrm{Var}[R_n] = \int_{\mathbb{X}} \eta_n(\mathbf{x}) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}),$$

où $\eta_n : [0,1] \to \mathbb{R}$ est une fonction de p_n définie pour tout $\mathbf{x} \in \mathbb{X}$ par :

$$\eta_n(\mathbf{x}) = \big(1 - p_n(\mathbf{x})\big) \int_{\mathbb{X}} p_n(\mathbf{y}) \mathbb{1}_{p_n(\mathbf{y}) \leq p_n(\mathbf{x})} \mathbf{P}_{\mathbf{X}}(d\mathbf{y}) + p_n(\mathbf{x}) \int_{\mathbb{X}} \big(1 - p_n(\mathbf{y})\big) \mathbb{1}_{p_n(\mathbf{y}) > p_n(\mathbf{x})} \mathbf{P}_{\mathbf{X}}(d\mathbf{y}).$$

Démonstration. Voir section 3.9. □

Nous remarquons que la fonction η_n est nulle en $p_n = 0$ et $p_n = 1$, tout comme les fonctions τ_n et ν_n. On montre dans la proposition suivante que cette fonction admet également un maximum.

Proposition 3.6.3. *Soit η_n la fonction définie à la proposition 3.6.2 et $q_n^* \in [0,1]$ tel que :*

$$\int_{\mathbb{X}} \mathbb{1}_{p_n(\mathbf{x}) > q_n^*} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) = \mu_n \Leftrightarrow q_n^* = \mathbb{P}(R_n > \mu_n). \tag{3.32}$$

Alors, η_n admet un maximum global sur $[0,1]$ au point q_n^*.

Démonstration. Voir section 3.9. □

Suite à la proposition 3.5.1, on a donné l'expression de la fonction de survie de R_n. Par conséquent, il est possible en pratique d'estimer la valeur de q_n^*. Bien évidemment, la valeur obtenue dépend de la loi $\mathbf{P}_{\mathbf{X}}$. On a ainsi montré que le critère J_{R_n} peut se réécrire :

$$J_{R_n}(\mathbf{x}_{n+1}) = \mathbb{E}_n\left[\int_{\mathbb{X}} \eta_{n+1}(\mathbf{x}) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \mid \mathbf{X}_{n+1} = \mathbf{x}_{n+1}\right],$$

où, au même titre que les fonctions τ_n et ν_n, la fonction η_n peut être interprétée comme une probabilité de classification erronée. En effet, comme on peut le constater en figure 3.5, les points où la fonction η_n est maximale correspondent à des sous-domaines de $\mathbb{X}$ où les trajectoires du processus ξ_n prennent des valeurs proches du seuil T. En outre, on voit immédiatement que, si le modèle de krigeage est davantage renseigné dans ces sous-domaines, alors l'estimation de p peut considérablement être améliorée.

Remarque 3.6.4. *Il est possible d'étudier les propriétés du critère J_{R_n} dans le cadre de la théorie des ensembles aléatoires (pour une présentation complète de cette théorie, on pourra consulter le livre (Molchanov, 2005)). En effet, on peut voir l'ensemble Γ_n défini par :*

$$\Gamma_n = \{\mathbf{x} \in \mathbb{X} : \xi_n(\mathbf{x}) > T\},$$

comme un ensemble aléatoire. En outre, il vérifie : $\mathbb{E}[\mathbf{P}_{\mathbf{X}}(\Gamma_n)] = \mu_n$. Considérons l'ensemble déterministe :

$$Q_n(q_n^*) = \{\mathbf{x} \in \mathbb{X} : p_n(\mathbf{x}) > q_n^*\},$$

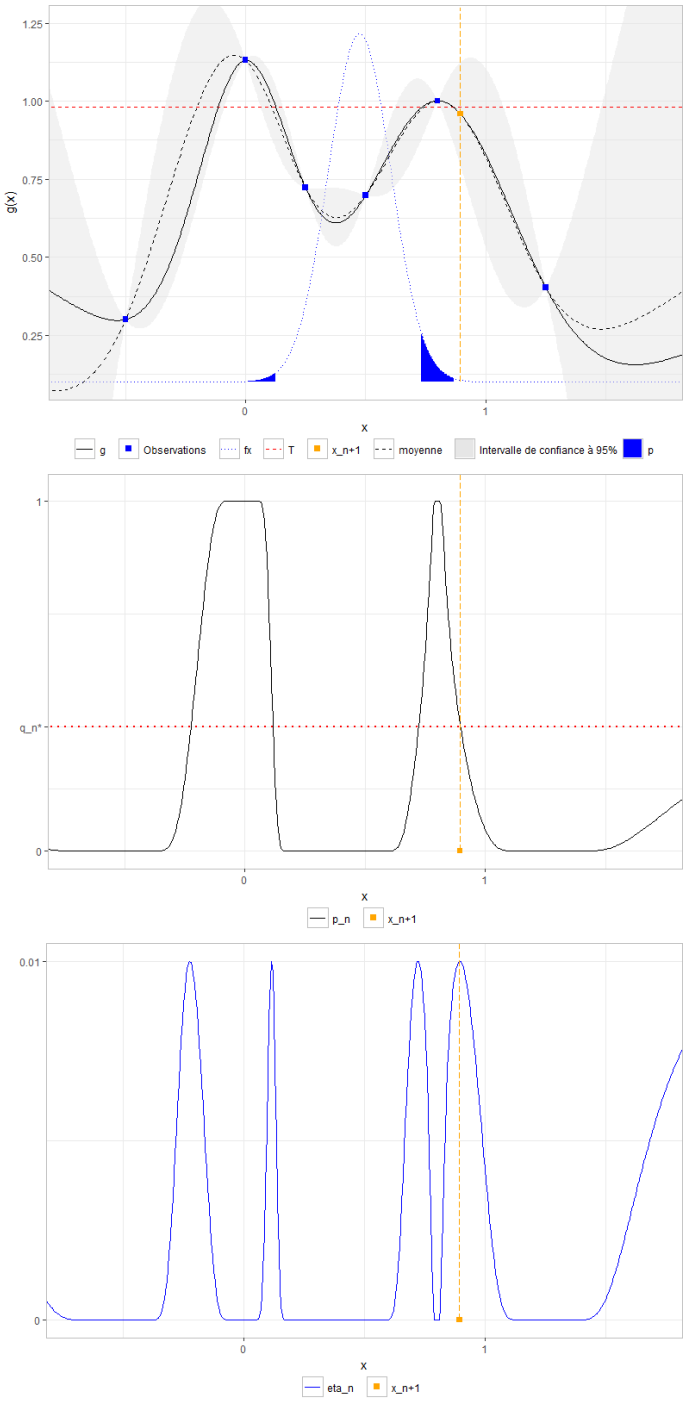

FIGURE 3.5 – Exemple simple pour aider à l'interprétation du critère J_{R_n}. En haut : la fonction g (courbe noire) est observée en $n = 6$ points et l'aire bleue sous la densité $f_{\mathbf{X}}$ est la probabilité p. Au milieu : représentation de la fonction $\mathbf{x} \in \mathbb{X} \mapsto p_n(\mathbf{x})$ et de la valeur de q_n^* donnée en proposition 3.6.3. En bas : représentation de la fonction $\mathbf{x} \in \mathbb{X} \mapsto \eta_n(\mathbf{x})$ définie dans la proposition 3.6.2. Les points dans $\mathbb{X}$ où η_n est maximale (par exemple, le point orange) conduisent à $p_n = q_n^*$. Il s'agit de points où le modèle de processus gaussien a ses trajectoires relativement proches du seuil T et doit être renseigné pour améliorer l'estimation de p.

où q_n^* *est donné en proposition 3.6.3, de sorte que* $\mathbf{P}_{\mathbf{X}}(Q_n(q_n^*)) = \mu_n$. *Notons* $A\Delta B$ *la différence symétrique entre deux ensembles* A *et* B : $A\Delta B = (A \cap B^c) \cup (A^c \cap B)$. *Alors, dans la théorie des ensembles aléatoires, la quantité :*

$$\begin{aligned} D_{Dev,n} = \mathbb{E}\big[\mathbf{P}_{\mathbf{X}}(\Gamma_n \Delta Q_n(q_n^*))\big] &= \mathbb{E}\left[\int_{\mathbb{X}} \mathbb{1}_{\xi_n(\mathbf{y})>T}\mathbb{1}_{p_n(\mathbf{y})\leq q_n^*}\mathbf{P}_{\mathbf{X}}(d\mathbf{y}) + \int_{\mathbb{X}} \mathbb{1}_{\xi_n(\mathbf{y})\leq T}\mathbb{1}_{p_n(\mathbf{y})>q_n^*}\mathbf{P}_{\mathbf{X}}(d\mathbf{y})\right] \\ &= \int_{\mathbb{X}} p_n(\mathbf{y})\mathbb{1}_{p_n(\mathbf{y})\leq q_n^*}\mathbf{P}_{\mathbf{X}}(d\mathbf{y}) + \int_{\mathbb{X}} (1-p_n(\mathbf{y}))\mathbb{1}_{p_n(\mathbf{y})>q_n^*}\mathbf{P}_{\mathbf{X}}(d\mathbf{y}), \end{aligned}$$

s'appelle « déviation de Vorob'ev ». On peut voir $Q_n(q_n^*)$ *comme une approximation (déterministe) de* Γ_n *et* $D_{Dev,n}$ *comme une mesure de l'erreur d'approximation. Dans (Chevalier, 2013), alternativement à* J_{S_n}, *il est proposé un critère basé sur cette quantité. Noté* $J_{Dev,n}$, *ce critère est défini pour tout* $\mathbf{x}_{n+1} \in \mathbb{X}$ *par :*

$$J_{Dev,n}(\mathbf{x}_{n+1}) = \mathbb{E}_n\left[D_{Dev,n+1} \mid \mathbf{X}_{n+1} = \mathbf{x}_{n+1}\right].$$

Des applications récentes de ce critère peuvent être retrouvées dans le chapitre 5 de (Azzimonti, 2016), ainsi que dans (El Amri et al., 2018). Les auteurs utilisent notamment ce critère pour construire des stratégies permettant de sélectionner plusieurs points simultanément (voir remarque 3.6.1). Pour finir, notons η_n^* *la valeur maximale de la fonction* η_n, *i.e. quand* $p_n = q_n^*$. *Alors, il est facile de vérifier que :*

$$\eta_n(\mathbf{x}) \leq \eta_n^* = \frac{D_{Dev,n}}{2}, \quad \forall \mathbf{x} \in \mathbb{X}.$$

Par conséquent, le critère J_{R_n} *vérifie :*

$$J_{R_n}(\mathbf{x}) \leq \frac{J_{Dev,n}(\mathbf{x})}{2}, \quad \forall \mathbf{x} \in \mathbb{X}.$$

Étant donnée cette relation d'ordre, serait intéressant de comparer la performance de ces deux critères sur un cas concret. En effet, cela signifie que J_{R_n} *est, une fois de plus, une meilleure approximation du critère* J_{S_n}. *Nous n'avons pas étudié plus en détails cette approche dans cette thèse, mais il s'agit d'une perspective très intéressante comme tenu des travaux récents sur ce sujet.*

3.6.2.3 Implémentation

La procédure d'implémentation pour le critère J_{R_n} est fondamentalement la même que pour les critères $(J_{n,k})_{1\leq k\leq 4}$, qui est décrite dans (Bect et al., 2012). Nous avons, dans le cas d'un processus aléatoire gaussien :

$$\begin{aligned} J_{R_n}(\mathbf{x}_{n+1}) &= \mathbb{E}_n\big[\mathrm{Var}[R_{n+1}] \mid \mathbf{X}_{n+1} = \mathbf{x}_{n+1}\big] \\ &= \mathbb{E}_n\big[\mathrm{Var}[R_{n+1}(\mathbf{x}_{n+1}, \xi(\mathbf{x}_{n+1}))] \mid \mathbf{X}_{n+1} = \mathbf{x}_{n+1}\big] \\ &= \int_{\mathbb{R}} \mathrm{Var}[R_{n+1}(\mathbf{x}_{n+1}, z)] \frac{1}{\sigma_n(\mathbf{x}_{n+1})\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{z-m_n(\mathbf{x}_{n+1})}{\sigma_n(\mathbf{x}_{n+1})}\right)^2} dz. \end{aligned}$$

On utilise une méthode de quadrature de Gauss-Hermite pour approcher l'intégrale :

$$J_{R_n}(\mathbf{x}_{n+1}) \approx \frac{1}{\sqrt{\pi}} \sum_{q=1}^{Q} w_q \mathrm{Var}[R_{n+1}(\mathbf{x}_{n+1}, m_n(\mathbf{x}_{n+1}) + \sqrt{2} u_q \sigma_n(\mathbf{x}_{n+1}))],$$

où $(w_q)_{1 \leq q \leq Q}$ et $(u_q)_{1 \leq q \leq Q}$ représentent respectivement les poids et les nœuds de la quadrature. De plus, pour approcher la variance $\mathrm{Var}[R_{n+1}(\mathbf{x}_{n+1}, m_n(\mathbf{x}_{n+1}) + \sqrt{2} u_q \sigma(\mathbf{x}_{n+1}))]$, il suffit de suivre la procédure donnée dans l'annexe B, en considérant que la fonction g est évaluée aux points du plan d'expériences $\mathbf{D}_n \cup \{\mathbf{x}_{n+1}\}$ et que le vecteur des observations est $(\mathbf{g}_n, m_n(\mathbf{x}_{n+1}) + \sqrt{2} u_q \sigma_n(\mathbf{x}_{n+1}))$.

3.7 Exemples

Nous proposons deux exemples pour mettre en avant l'intérêt des méthodes développées dans ce chapitre. Le premier est proposé dans de nombreux articles pour mettre en œuvre des analyses de fiabilité. Le second est un cas réel de la société STMicroelectronics, qui a déjà été traité dans (Oger, 2014).

3.7.1 Exemple en dimension 6

On teste ici nos résultats sur la version modifiée d'un exemple en dimension $d = 6$ proposé, par exemple, dans (Gayton et al., 2003), (Echard et al., 2011) et (Echard et al., 2013). La fonction $g : \mathbb{R}^6 \to \mathbb{R}$ est connue et définie par :

$$g(\mathbf{x}) = 3x_4 - \left| \frac{2x_5}{x_2 + x_3} \sin\left(\frac{\omega_0 x_6}{2}\right) \right|, \quad \text{où } \omega_0 = \sqrt{\frac{x_2 + x_3}{x_1}}.$$

On veut estimer la probabilité $p = \mathbb{P}\big(g(\mathbf{X}) > T\big)$, où $T = 0.7$ et $\mathbf{X} = (X_1, \ldots, X_6)$ est un vecteur aléatoire gaussien. Les paramètres des variables gaussiennes sont donnés dans le tableau 3.3.

Pour un échantillon de taille $N = 10^7$, la méthode Monte-Carlo naïve donne $p \approx 1.77 \times 10^{-2}$ et l'intervalle de confiance à 95% qui en résulte est $[1.76 \times 10^{-2},\ 1.78 \times 10^{-2}]$. On se donne un budget total de 150 évaluations de g pour estimer p. On part d'un plan d'expériences initial, un LHS de taille $n = 25$, et on détermine 125 points supplémentaires où évaluer g grâce à une stratégie SUR. Pour cela, les critères que l'on utilise sont J_{R_n}, donné en (3.30), et $J_{4,n}$, donné en (3.28). Nous avons souhaité comparer J_{R_n} à $J_{4,n}$ car ce dernier est le plus couramment utilisé en pratique (voir (Chevalier et al., 2014)). Pour choisir les points candidats $\mathbf{x}_{n+1} \in \mathbb{X} \setminus \mathbf{D}_n$, nous procédons comme dans (Bect et al., 2012) : on simule un échantillon i.i.d. de points suivant $\mathbf{P}_\mathbf{X}$ et le critère sera évalué en chaque point.

Le modèle est un processus gaussien de tendance polynomiale d'ordre 2 et de noyau Matérn $\frac{3}{2}$, dont les paramètres sont estimés par maximum de vraisemblance à chaque nouvel ajout de points. Ces choix ont été validés par la méthode de « leave-one-out », que nous

avons présentée en section 2.3.3.3. La figure 3.6 illustre les capacités de prédiction du modèle, pour les $n = 25$ observations initiales ainsi qu'à l'issue des 150 évaluations de g, via cette méthode de validation croisée. Quel que soit soit le nombre d'observations ou le critère considéré, on constate que le modèle prédit correctement les valeurs de g déjà observées. En effet, les points rouges sont relativement bien alignés par rapport à la droite $y = x$.

À chaque itération, on calcule une estimation de la probabilité de défaillance p grâce à l'espérance de R_n. On utilise pour cela l'estimateur donné en (3.5). On estime également un intervalle de crédibilité à 95% pour S_n donné en (3.21), via une méthode Monte-Carlo naïve, avec $\beta = \frac{1}{2}$.

Les résultats obtenus après avoir épuisé le budget d'appels à la fonction g sont donnés dans le tableau 3.4 et les distributions empiriques de R_n sont représentées en figure 3.7. Les estimations fournies par J_{R_n} et $J_{4,n}$ sont compatibles avec la valeur estimée par la méthode Monte-Carlo naïve. De plus, comme on peut le constater sur la figure 3.8, quel que soit le critère considéré, l'intervalle de crédibilité tend à diminuer au fur et à mesure que le nombre de données croît. Néanmoins, on observe encore un léger biais, bien que les estimations semblent se stabiliser à partir de 100 évaluations. Il est probable qu'en augmentant le nombre d'appels à la fonction g, les estimations de p ne soient pas meilleures. Plusieurs solutions sont alors envisageables : choisir une autre fonction de covariance (par exemple, par l'intermédiaire d'une méthode de validation croisée), augmenter la taille des échantillons Monte-Carlo pour l'estimation des diverses intégrales suivant $\mathbf{P}_{\mathbf{X}}$, ou encore, procéder à l'optimisation des fonctions J_{R_n} et $J_{4,n}$ autrement que part l'intermédiaire d'un échantillon de point candidats distribués suivant $\mathbf{P}_{\mathbf{X}}$. En effet, il serait certainement préférable d'utiliser un algorithme d'optimisation. Cela a été traité dans le cadre d'un stage sur le sujet (plus d'informations sont données ci-après lors de la présentation du cas industriel) et des algorithmes génétiques pour l'optimisation des critères ont été intégrés à la procédure. Des résultats de simulations peuvent être retrouvés dans (Smith, 2018).

3.7.2 Étude d'un cas réel de la société STMicroelectronics

Les méthodes d'estimation décrites dans ce chapitre ont été testées sur un cas réel de la société STMicroelectronics. Celui-ci a déjà été présenté et traité dans une précédente thèse Cifre (voir (Oger, 2014)), afin de répondre à une problématique d'estimation de probabilité de défaillance identique à la nôtre. Cependant, on propose ici d'enrichir l'approche proposée dans (Oger, 2014) en intégrant une procédure de planification d'expériences séquentielle basée sur le principe des stratégies SUR présentées en section 3.6. Pour analyser la pertinence de nos résultats, nous disposons d'un ensemble de 10^3 simulations à partir desquelles des études statistiques ont déjà été menées au sein de STMicroelectronics. Nous verrons que, grâce à l'ensemble des méthodes présentées dans ce chapitre, la valeur estimée de p que l'on obtient avec seulement 10^2 simulations coïncide avec celle obtenue avec 10^3. Les résultats présentés ci-après sont issus d'une collaboration avec Raphaël Smith, stagiaire en Master 2 MApI3 de l'université Toulouse III Paul Sabatier. Dans le cadre de son stage, Raphaël a notamment adapté la méthode de modélisation par processus gaussien pour prendre en compte les valeurs du gradient de g aux points du plan d'expériences. En effet, le simulateur utilisé pour réaliser les simulations numériques est susceptible de fournir cette

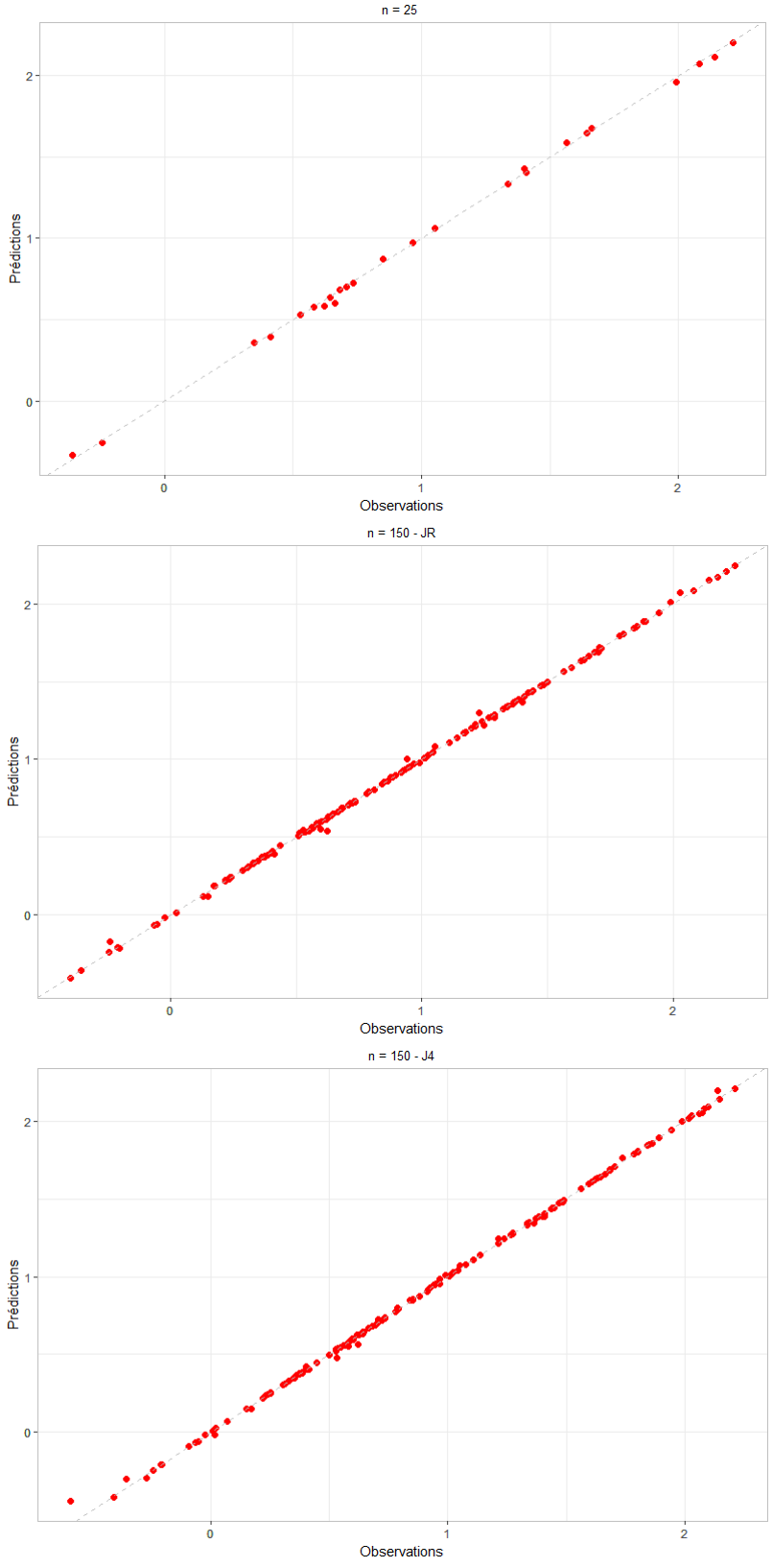

FIGURE 3.6 – Validation du modèle de krigeage par « Leave-One-Out ».

Variable	X_1	X_2	X_3	X_4	X_5	X_6
Moyenne	1	1	0.1	0.5	0.45	1
Variance	0.05	0.1	0.01	0.05	0.075	0.2

TABLE 3.3 – Distributions des marginales de $\mathbf{X}$.

	Espérance de R_n	Variance de R_n	Intervalle de crédibilité à 95%
J_{R_n} ; $n = 150$	1.796×10^{-2}	5.3×10^{-7}	$[1.629 \times 10^{-2},\ 1.975 \times 10^{-2}]$
$J_{4,n}$; $n = 150$	1.802×10^{-2}	6.9×10^{-7}	$[1.612 \times 10^{-2},\ 2.007 \times 10^{-2}]$

TABLE 3.4 – Estimations de l'espérance de R_n donnée en (3.4), de la variance de R_n donnée en (3.11), et de l'intervalle de crédibilité à 95% donné en (3.21). Les estimations sont obtenues après 150 appels à la fonction g. La méthode Monte-Carlo naïve avec un échantillon de taille $N = 10^7$ conduit à $p \approx 1.77 \times 10^{-2}$.

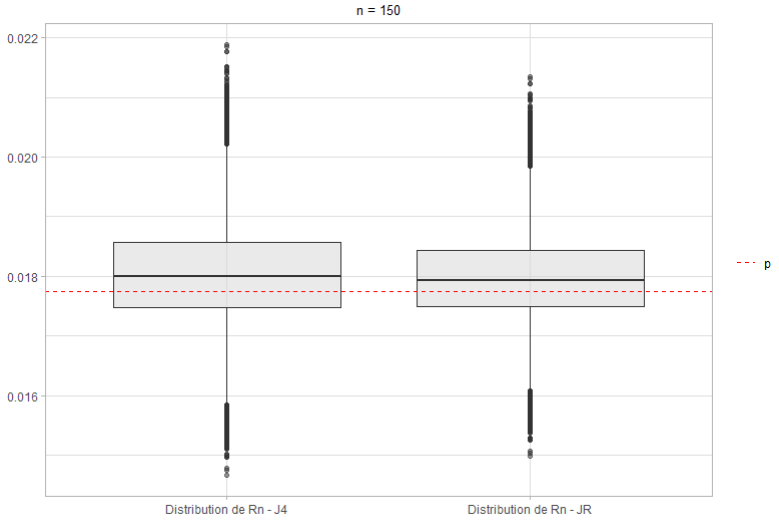

FIGURE 3.7 – Distribution de la variable aléatoire R_n après 150 appels à la fonction g.

information supplémentaire, pour un coût de calcul faible (voir (Vardapetyan et al., 2008)). On verra par la suite que les résultats sont bien meilleurs une fois cette information prise en compte. L'ensemble des résultats obtenus lors de ce stage est contenu dans (Smith, 2018).

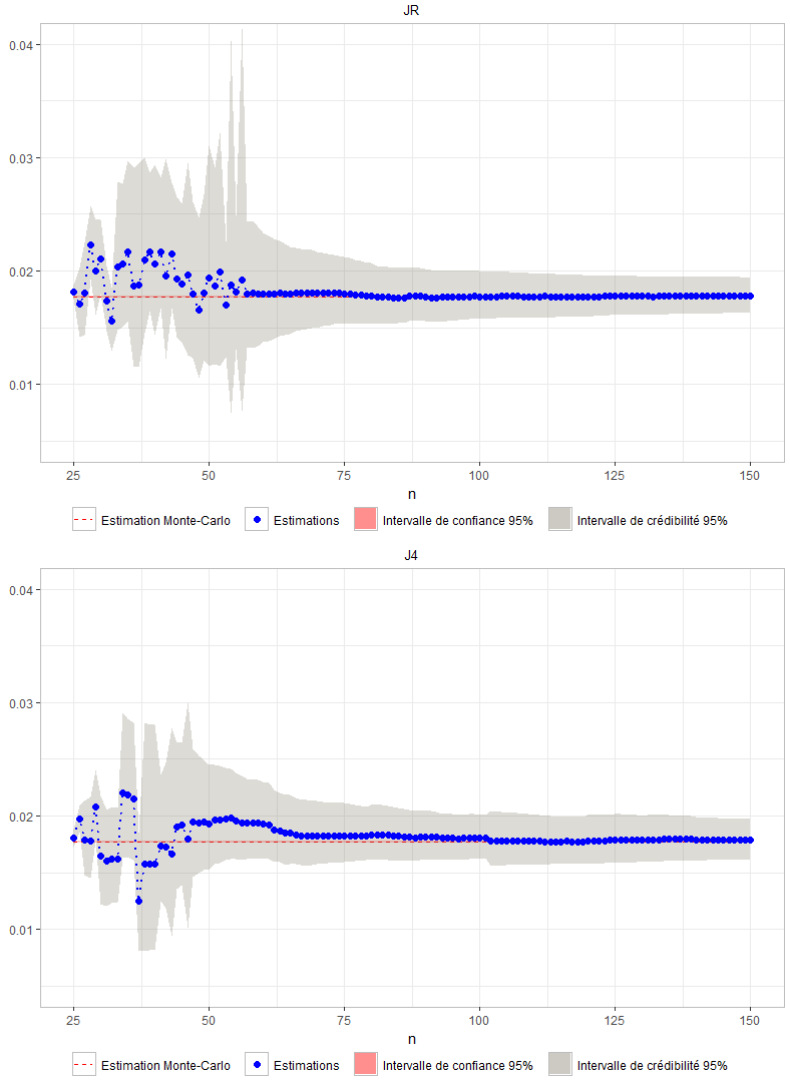

FIGURE 3.8 – Estimations successives de la probabilité de défaillance avec les critères J_{R_n} (en haut) et $J_{4,n}$ (en bas). Ligne en pointillés rouges : l'estimation de la probabilité de défaillance obtenue par application de la méthode Monte-Carlo naïve avec $N = 10^7$. Aire rouge : intervalle de confiance à 95% pour p obtenu également par la méthode Monte-Carlo naïve. Puisqu'il est égal à $[1.76 \times 10^{-2}, 1.78 \times 10^{2}]$, cet intervalle est à peine visible. Points bleus : estimations successives de p obtenues grâce à l'estimateur (3.5). Aire grise : estimations par la méthode Monte-Carlo naïve de l'intervalle de crédibilité à 95% donné en (3.21).

3.7.2.1 Description du cas réel

Le cas industriel concerne l'étude d'un composant électronique appelé duplexeur. Il s'agit d'un dispositif utilisé pour filtrer un signal sur différentes bandes de fréquence en vue de le séparer. La société STMicroelectronics couvrant le marché de la téléphonie mobile, le duplexeur étudié fonctionne aux radiofréquences ($f \sim 1$ GHz). La voie principale de transmission du signal est divisée en deux voies distinctes (voir figure 3.9). Chaque voie comporte un filtre passe-bande permettant de sélectionner certaines composantes fréquentielles du signal.

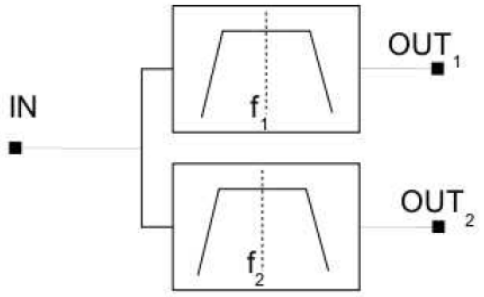

FIGURE 3.9 – Schéma d'un duplexeur. IN : entrée du signal, OUT : sorties pour chacun des signaux séparés.

Le processus de fabrication d'un filtre passe-bande est une succession d'opérations complexes, difficiles à maintenir constantes au cours du temps : dépôt de couches métalliques par électrolyse, gravure par plasma, photolithographie... La technologie qui en résulte (nommée RLC06A) peut alors être vue comme un empilement de couches isolantes et conductrices. Les paramètres sujets à des variations sont typiquement les épaisseurs de ces dépôts. Pour le duplexeur étudié, exactement 4 épaisseurs de dépôts sont influantes, au sens où leurs variations impactent significativement les performances du filtre. Les couches concernées sont représentées à la figure 3.10. Il est possible de modéliser la variabilité naturelle de l'outil industriel en associant à ces grandeurs une distribution, de sorte qu'elles deviennent de fait des variables aléatoires. On les note $X_1, \ldots, X_4$ et leurs distributions sont extraites par la mesure en ligne (c'est-à-dire directement sur les lignes de production) de motifs de test. On peut correctement approcher celles-ci par des distributions normales. Les paramètres de chacune d'entre elles sont donnés dans la table 3.5.

Dans le cadre de l'électronique Haute-Fréquence, on montre que le duplexeur peut être complètement caractérisé par sa matrice de dispersion. Il s'agit ici d'une matrice 3×3, qui donne pour chaque entrée–sortie du dispositif la proportion du signal transmise ou réfléchie. Chaque paramètre dépend de la fréquence. Les spécifications du client sont définies sur ces caractéristiques fréquentielles : il donne un gabarit que la réponse fréquentielle du duplexeur doit respecter. Ce gabarit ne porte pas sur toute la plage de fréquences mais uniquement sur certaines bandes. Ici, cela se traduit par 12 réponses, chacune d'entre elles étant une fonction inconnue des 4 variables d'entrée données dans la table 3.5. Chaque réponse est associée à une plage de fréquences et possède sa propre contrainte, exprimée

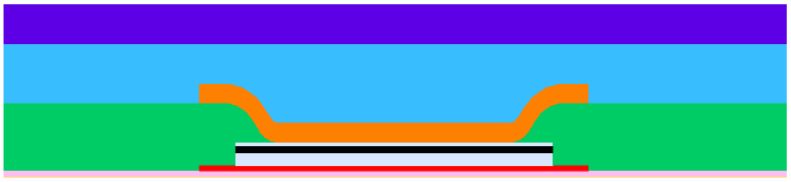

FIGURE 3.10 – Vue en coupe de la technologie RLC06A dont les paramètres variables sont : Capa2 (noir), Meta2 (orange), BCB1 (vert) et Meta1 (rouge).

en décibels, comme indiqué dans le tableau 3.6. On verra ci-après comment adapter les méthodes développées dans ce chapitre au fait que la sortie est vectorielle.

Les simulations numériques sont effectuées grâce au logiciel commercial HFSS (« High Frequency Structural Simulator »), de la société ANSYS, qui permet la résolution d'un système d'équations intégro-différentielles : ce type d'outil est extrêmement complexe et intègre l'état de l'art des méthodes d'analyse numérique. En outre, il est basé sur la méthode des éléments finis, très populaire dans le monde du calcul scientifique. Le logiciel HFSS dispose d'un module de représentation 3D qui permet la génération d'un modèle géométrique fidèle (voir figure 3.11). Ce modèle est paramétré, les variables étant les épaisseurs des matériaux. Il est donc possible d'accéder à différentes configurations (virtuelles) du produit et de déduire pour chacune d'entre elles, les caractéristiques électriques. La simulation physique de ce modèle nécessite 25Go de mémoire, 5 heures d'exécution avec 8 processeurs (« threads ») soit environ 12 heures de temps CPU.

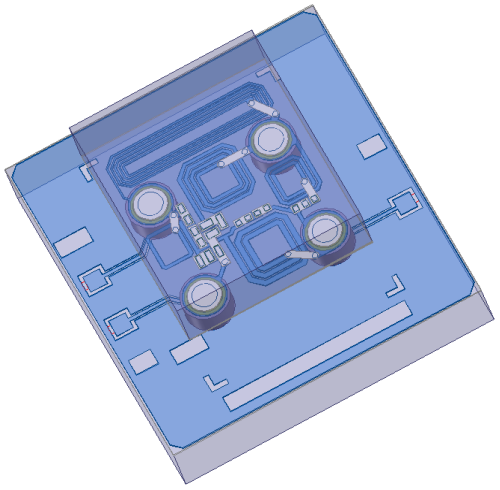

FIGURE 3.11 – Modélisation 3D du diplexeur.

Facteur	Nom	Unité	Valeur minimale	Valeur maximale	Distribution	Paramètre de position	Paramètre d'échelle
X_1	BCB1	µm	2.7504	5.1011	Normale tronquée	3.3477	0.19108
X_2	Capa2	µm	160.41	188.31	Normale tronquée	174.31	1.6831
X_3	Meta1b	µm	0.62241	0.90364	Normale tronquée	0.73389	0.031933
X_4	Meta2	µm	4.9964	7.4946	Normale tronquée	6.1457	0.26781

TABLE 3.5 – Distribution des 4 épaisseurs des dépôts qui impactent la réponse du duplexeur.

Réponse	Plage de fréquences	Unité	Fonction agrégation	Valeur minimale	Valeur maximale	Contrainte
S_{11}	2.4 - 2.485GHz	dB	Min	0	$+\infty$	≥ 17
S_{21}	2.4 - 2.485GHz	dB	Max	0	$+\infty$	≤ 0.7
S_{22}	2.4 - 2.485GHz	dB	Min	0	$+\infty$	≥ 17
S_{31}	2.4 - 2.485GHz	dB	Min	0	$+\infty$	≥ 18
S_{21}	4.9 - 5.85GHz	dB	Min	0	$+\infty$	≥ 20
S_{31}	4.9 - 5.85GHz	dB	Max	0	$+\infty$	≤ 0.7
S_{11}	4.9 - 5.85GHz	dB	Min	0	$+\infty$	≥ 17
S_{33}	4.9 - 5.85GHz	dB	Min	0	$+\infty$	≥ 17
S_{21}	5.85 - 7GHz	dB	Min	0	$+\infty$	≥ 15
S_{21}	7 - 9.5GHz	dB	Min	0	$+\infty$	≥ 7
S_{21}	9.8 - 10.5GHz	dB	Min	0	$+\infty$	≥ 16
S_{31}	9.8 - 11.65GHz	dB	Min	0	$+\infty$	≥ 11

TABLE 3.6 – Réponses qui mesurent la performance du duplexeur et les contraintes associées.

3.7.2.2 Formalisme et étude préliminaire

Pour une meilleure compréhension de ce qui va suivre, commençons par formaliser le problème. Ici, une simulation numérique via le logiciel HFSS requiert l'exécution d'un code de calcul, représenté par la fonction g définie par :

$$\begin{array}{rccl} g & : & \mathbb{X} & \to & \mathbb{R}^{12} \\ & & \mathbf{x} & \mapsto & g(\mathbf{x}) = (g_1(\mathbf{x}), \dots, g_{12}(\mathbf{x})), \end{array}$$

où l'ensemble $\mathbb{X} \subseteq \mathbb{R}^4$ est défini dans le tableau 3.5 (colonnes « Valeur mininale » et « Valeur maximale »). Les spécifications imposées au produit industriel portent sur chaque réponse individuellement (voir tableau 3.6). Pour un jeu de facteurs $\mathbf{x} \in \mathbb{X}$, le produit est dit défaillant si :

$$g_1(\mathbf{x}) \geq T_1 \quad \text{ou} \quad \dots \quad \text{ou} \quad g_{12}(\mathbf{x}) \geq T_{12}.$$

où les valeurs des seuils $T_1, \dots, T_{12}$ sont données dans la colonne « Contrainte » du tableau 3.6. Cela signifie qu'une pièce est défaillante si au moins une des réponses l'est. Ainsi, la probabilité de défaillance p que l'on cherche à estimer s'écrit :

$$p = \mathbb{P}\left(\bigcup_{j=1}^{12} g_j(\mathbf{X}) \geq T_j \right), \tag{3.33}$$

où $\mathbf{X} = (X_1, \dots, X_4)$ contient les 4 épaisseurs fluctuantes du produit dont les distributions sont données dans le tableau 3.5. D'un point de vue formel, les méthodes développées dans les sections précédentes peuvent continuer à s'appliquer, avec la construction d'un processus aléatoire multi-dimensionnel. En effet, il existe des modèles de krigeage adaptés au cas où la sortie est vectorielle, appelés modèles de co-krigeage (voir (Chilès and Delfiner, 1999), (Wackernagel, 2003) ainsi que la thèse (Le Gratiet, 2013)). Malgré leur intérêt, nous ne les avons pas étudiés dans le cadre de cette thèse. En outre, cela introduit une complexité considérable pour la mise en œuvre pratique : ce type de modèle s'avère difficile à renseigner (détermination des paramètres du modèle) et à exploiter numériquement. Dans la section suivante, on explique comment nous avons procédé pour traiter le cas multi-dimensionnel.

Avant cela, précisons qu'une première étude de ce cas industriel a été réalisée d'après $N = 10^3$ simulations via le logiciel HFSS, l'idée étant de mener une estimation de la probabilité de défaillance par une méthode Monte-Carlo naïve. Étant donné un échantillon $(\mathbf{X}_i)_{1 \leq i \leq N}$ i.i.d. distribué suivant les lois des facteurs d'entrée, l'estimateur $\widehat{p}$ de la probabilité p donnée en équation (3.33) est :

$$\widehat{p} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\bigcup_{j=1}^{12} g_j(\mathbf{X}_i) \geq T_j}.$$

Nous en avons déduit que la probabilité de défaillance vaut approximativement 6.7×10^{-2}. Avec le théorème central limite, on peut obtenir un intervalle de confiance à 95% sur la valeur de p. Ici, cet intervalle vaut $[4.5 \times 10^{-2},\ 8.8 \times 10^{-2}]$. On utilisera ces résultats de simulation comme référence afin de valider ceux obtenus.

3.7.2.3 Stratégie pour une sortie vectorielle

La performance du produit est ici déterminée par plusieurs réponses. Dans le cadre des stratégies SUR présentées en section 3.6, les critères d'échantillonnage que l'on propose ne sont définis que pour une seule réponse. On décrit ici la procédure que nous avons adoptée pour traiter les 12 réponses.

Supposons que la fonction g soit déjà connue aux points d'un plan d'expériences $(\mathbf{x}_i)_{1\leq i\leq n}$ de taille $n \in \mathbb{N}^*$. Pour chacune des réponses, on utilise ces observations pour construire un modèle de processus gaussien. Pour tout $j = 1, \ldots, 12$, on note $\xi_{n,j}$ le modèle de processus gaussien conditionné aux observations pour la j-ème réponse. On définit ainsi :

$$p_n(\mathbf{x}) = \mathbb{P}\left(\bigcup_{j=1}^{12} \xi_{n,j}(\mathbf{x}) \geq T_j\right), \quad \forall \mathbf{x} \in \mathbb{X}. \tag{3.34}$$

De même que dans le cas unidimensionnel (voir section 3.2.1), l'estimateur de la probabilité p donnée en (3.33) est la quantité $\int_{\mathbb{X}} p_n(\mathbf{x})\mathbf{P}_{\mathbf{X}}(d\mathbf{x})$. Néanmoins, la quantité (3.34) étant inacessible, sauf à supposer un modèle de dépendance entre les processus aléatoires (ce qui revient à construire un champ multi-dimensionnel), on procède par majoration. Pour tout $\mathbf{x} \in \mathbb{X}$ et pour tout $j = 1, \ldots, 12$, on pose $p_{n,j}(\mathbf{x}) = \mathbb{P}\left(\xi_{n,j}(\mathbf{x}) \geq T_j\right)$ et on définit :

$$p_n^+(\mathbf{x}) = \min\left(1, \sum_{j=1}^{12} p_{n,j}(\mathbf{x})\right).$$

La borne de l'union implique que :

$$p_n(\mathbf{x}) \leq p_n^+(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{X}. \tag{3.35}$$

Si tous les modèles sont bien renseignés, les probabilités $(p_{n,j}(\mathbf{x}))_{1\leq j\leq 12}$ sont proches de 0 ou 1, et cela n'induit pas d'erreur importante. Ainsi, l'estimateur de la probabilité p donnée en (3.33) que l'on utilise en pratique est la quantité :

$$\int_{\mathbb{X}} p_n^+(\mathbf{x})\mathbf{P}_{\mathbf{X}}(d\mathbf{x}). \tag{3.36}$$

L'intervalle de crédibilité à 95% que l'on considère pour évaluer la précision de cet estimateur est issu de la proposition 3.5.4 (avec $\beta = \frac{1}{2}$) et défini par :

$$\left[1 - \int_{\mathbb{X}} \min\left(1, \frac{1 - p_n^+(\mathbf{x})}{0.025}\right)\mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \quad , \quad \int_{\mathbb{X}} \min\left(1, \frac{p_n^+(\mathbf{x})}{0.025}\right)\mathbf{P}_{\mathbf{X}}(d\mathbf{x})\right]. \tag{3.37}$$

Pour estimer (3.36) et (3.37), on utilise une méthode Monte-Carlo naïve en simulant suivant $\mathbf{P}_{\mathbf{X}}$. Notons que, avec cette approche, on induit nécessairement un biais et l'estimation de p que l'on obtient excède la vraie valeur de la probabilité de défaillance.

Concernant la planification séquentielle via une stratégie SUR basée sur la variance de R_n (voir section 3.6.2), nous procédons de la façon suivante : pour chacune des réponses, un

modèle de processus gaussien étant construit, on estime la variance de la variable aléatoire R_n (à nouveau, voir l'annexe B pour plus de détails sur la façon de procéder). Pour déterminer le point $\mathbf{x}_{n+1}$ où effectuer la prochaine évaluation de g, on optimise le critère d'échantillonnage J_{R_n}, donné en (3.30), qui correspond à la réponse pour laquelle la variance de R_n est la plus élevée. Cette procédure consiste simplement à retenir la réponse pour laquelle l'incertitude sur le résultat est la plus grande. On ajoute le point sélectionné au plan d'expériences déjà existant. Enfin, on calcule la valeur des 12 réponses en ce nouveau point et on procède à l'estimation de p en appliquant l'approche ci-dessus.

3.7.2.4 Résultats

Pour chaque réponse, le modèle de krigeage que nous avons retenu comporte une tendance constante et une fonction de corrélation Matérn $\frac{3}{2}$ isotrope. Ces choix ont été validés par une méthode de validation croisée (voir (Smith, 2018)). Notre plan d'expériences initial est un plan LHS de taille $n = 50$ auquel nous avons ajouté 150 points en appliquant la stratégie SUR basée sur le critère J_{R_n}. Cependant, au lieu d'optimiser le critère sur tout l'ensemble $\mathbb{X}$, on détermine le point à ajouter au plan d'expériences parmi l'ensemble de points $(\mathbf{X}_i)_{1\leq i\leq N}$ où les valeurs de g sont déjà disponibles. Cela évite de faire des simulations supplémentaires. On rappelle que $N = 10^3$ et que ces observations ont conduit à $p \approx 6.7 \times 10^{-2}$.

Les résultats que nous avons obtenus sont donnés dans le tableau 3.8, et les estimations successives sont représentées sur la figure 3.12. La valeur estimée de p par la méthode Monte-Carlo est représentée par la droite en pointillés rouges : c'est notre mesure de référence, bien qu'il ne s'agisse pas de la vraie valeur de p. En effet, la taille de l'échantillon utilisé pour cette estimation est relativement faible. On constate que, très rapidement, les estimations successives (points bleus) sont dans l'intervalle de confiance fourni par la méthode Monte-Carlo naïve avec 10^3 observations (aire rouge). À partir de 150 évaluations, les estimations semblent se stabiliser autour la mesure de référence. Cela signifie qu'avec seulement 150 évaluations, l'estimation de la probabilité de défaillance que l'on obtient par application de la stratégie SUR basée sur J_{R_n} est équivalente à celle obtenue pour 1000 évaluations. De ce point de vue, les résultats sont satisfaisants. Néanmoins, on reconnaît que les intervalles de crédibilité que l'on obtient ne sont pas informatifs. En effet, ils sont très larges et il serait souhaitable qu'ils se réduisent au fur et à mesure que le plan d'expériences est enrichi, dans le sens où il s'agit de mesures de l'incertitude sur les estimations successives de la probabilité de défaillance. Ainsi, l'ajout de données semble, ici, ne pas se traduire par une sélection significative des trajectoires du processus compatibles avec celles-ci. Il est possible que le fait de sélectionner les points à ajouter au plan d'expériences parmi l'ensemble $(\mathbf{X}_i)_{1\leq i\leq N}$ de points candidats distribués suivant la loi de $\mathbf{P}_{\mathbf{X}}$ diminue l'efficacité de la stratégie. On peut également conclure que la fonction de corrélation (et donc la base d'interpolation) retenue n'est finalement pas adaptée, ou encore que le choix des paramètres du modèle n'est pas approprié, ce qui signifie que la méthode de vraisemblance est ici mise en difficulté.

	Estimation par méthode Monte-Carlo naïve	Intervalle de confiance à 95%
p_N^{MC} ; $N = 10^3$	6.7×10^{-2}	$\left[4.5\times10^{-2}\,;\,8.8\times10^{-2}\right]$

TABLE 3.7 – Estimation de la probabilité de défaillance avec la méthode Monte-Carlo naïve.

	Espérance de R_n	Variance de R_n	Intervalle de crédibilité à 95%
$J_n^{R_n}$; $n = 200$	0.069	1×10^{-3}	$\left[1.3\times10^{-3}\,;\,1.9\times10^{-1}\right]$

TABLE 3.8 – Estimation de la probabilité de défaillance après l'ajout de 150 points à un plan d'expériences initial LHS de taille $n = 50$, via une stratégie SUR basée sur le critère J_{R_n}.

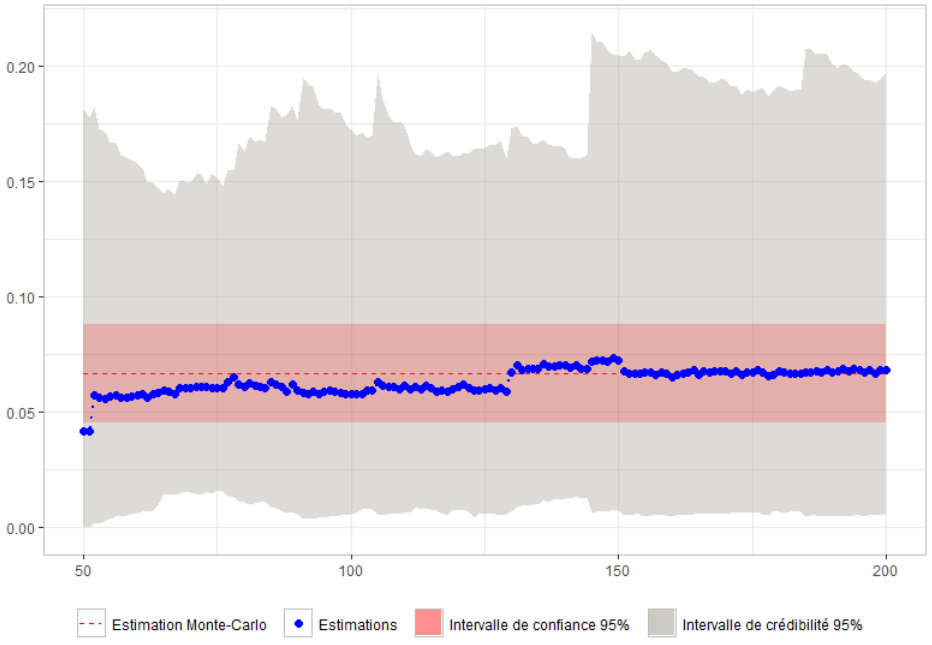

FIGURE 3.12 – Estimations successives de p pour le cas industriel.

3.7.2.5 Étude avec prise en compte de la dérivée

Lors de la réalisation de simulations numériques, il est possible de connaître les valeurs des dérivées aux points mesurés. En effet, le simulateur HFSS permet non seulement d'accéder aux valeurs de g en tout de $\mathbb{X}$, mais également aux différentes dérivées partielles selon chaque dimension, i.e. son gradient. Pour chaque réponse, nous avons donc construit un modèle de krigeage qui tient compte du gradient aux points d'évaluation. Pour plus d'informations sur la prise en compte de la valeur de la dérivée dans la modélisation par processus gaussien, on pourra consulter (Solak et al., 2003) et (Wu et al., 2018).

L'utilisation des dérivées dans la modélisation par processus gaussien a plusieurs conséquences sur celle-ci. D'une part, cela constitue de l'information supplémentaire sur le processus et il faut la prendre en compte lors du conditionnement. D'autre part, cela a des répercussions sur le choix de la fonction de covariance. Par exemple, la fonction exponentielle est exclue car, avec cette fonction de corrélation, les trajectoires d'un processus gaussien sont continues mais pas différentiables (voir section 2.3.3.5 pour une présentation des fonctions de corrélation usuelles pour la modélisation par processus gaussiens et les propriétés de régularité des trajectoires). À nouveau, nous avons choisi une tendance constante et une fonction de corrélation Matérn $\frac{3}{2}$ isotrope (voir (Smith, 2018) pour une discussion sur ce choix de modélisation). Notons que, lorsque l'on prend en compte les dérivées, il devient fondamental de mettre en place une méthode robuste d'inversion de la matrice de covariance. En effet, la taille et le conditionnement numérique des matrices de covariance augmentent considérablement. La qualité numérique du problème est bien souvent très mauvaise, ce qui pose des problèmes pour identifier les paramètres du modèle, voire même lors de l'étape de conditionnement. La solution utilisée dans (Smith, 2018) est d'ajouter une constante $\varepsilon > 0$ sur la diagonale de la matrice. On appelle cela « l'effet nugget » en géostatistique. Cela revient à considérer que les observations sont très légèrement bruitées et que l'on autorise le modèle à ne plus interpoler les points où la fonction est observée. Cette astuce permet d'améliorer considérablement le conditionnement de la matrice à inverser. La valeur de ε est choisie de façon à réaliser un compromis entre faciliter l'inversion de la matrice de covariance (ε suffisamment grand) et minimiser l'imprécision dans la modélisation (ε suffisamment petit).

Les résultats que nous avons obtenus en intégrant les valeurs des dérivées observées sont donnés en figure 3.13. Ils doivent être comparés avec ceux donnés en figure 3.14, où le modèle de krigeage est construit sans tenir compte des valeurs des dérivées, mais pour lequel nous avons également utilisé l'effet nugget ($\varepsilon = 10^{-6}$). Nous avons traité ces deux cas avec le même plan initial de type LHS constitué de 50 points. Notons que, cette fois, les points candidats n'ont pas été choisis parmi l'échantillon Monte-Carlo $(\mathbf{X}_i)_{1 \leq i \leq N}$ pour construire le plan d'expériences. En effet, les simulations ont été réalisées au fur et à mesure de l'identification des points par le critère. On remarque immédiatement que les résultats sont bien meilleurs lorsque les valeurs des dérivées sont connues. En effet, en figure 3.13, on voit que dès 50 points, l'intervalle de crédibilité est inclus dans l'intervalle de confiance estimé par la méthode Monte-Carlo naïve. De plus, on vérifie qu'il diminue légèrement avec l'ajout des points aux plans d'expériences. Ces résultats montrent que l'ajout des dérivées au modèle permet une meilleure modélisation et une estimation plus fiable de la probabilité de défaillance.

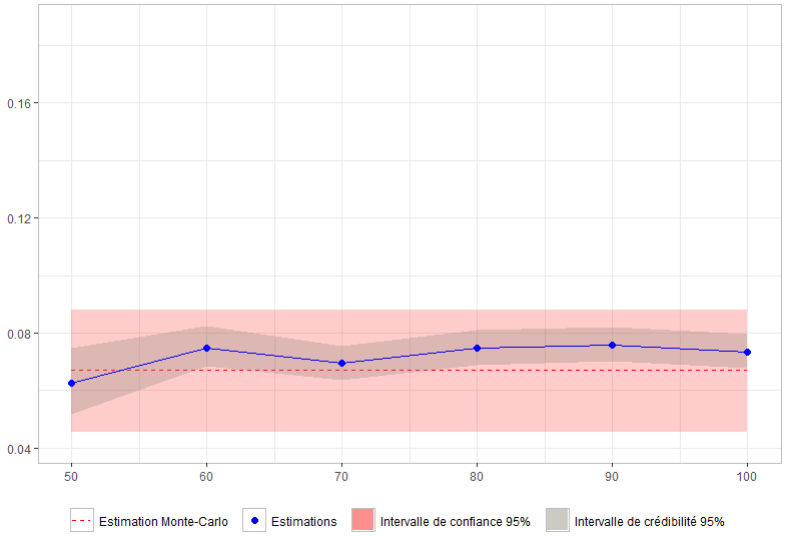

FIGURE 3.13 – Estimations successives de p pour le cas industriel avec prise en compte de la dérivée.

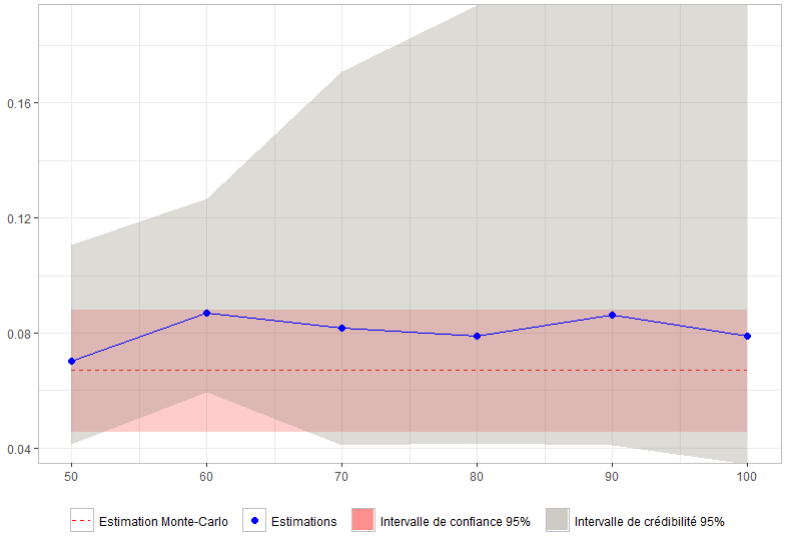

FIGURE 3.14 – Estimations successives de p pour le cas industriel sans prise en compte de la dérivée.

En outre, on remarque en figure 3.14 que l'incertitude représentée par les intervalles de crédibilité croît fortement et que l'estimation de p est moins bonne. Nous avons observé ce phénomène à plusieurs reprises lors des différentes simulations menées pour l'étude de ce cas industriel. Cela nous a conduits à l'hypothèse suivante : au moins une des fonctions que l'on cherche à modéliser présente des caractéristiques bien différentes suivant la dimension considérée. Il est donc possible qu'un choix de fonction de corrélation anisotrope serait plus approprié. Cependant, la méthode proposée dans (Smith, 2018), pour intégrer l'information sur les dérivées dans la modélisation, nécessite des calculs analytiques relativement complexes et a été développée uniquement pour la fonction de corrélation isotrope Matérn,

avec $\nu = q + \frac{1}{2}$, $q \in \mathbb{N}^*$. Cela pourrait être une perspective de développement intéressante.

La conclusion de cette analyse est qu'il est possible de faire des prédictions quantitatives de risque sur des cas réels, avec un nombre de simulations tout à fait raisonnable, dans une configuration informatique relativement légère pour une entreprise.

3.8 Conclusion et perspectives

Lorsque peu de données sont disponibles, il est courant d'adopter une approche bayésienne pour mener une estimation de la quantité d'intérêt. Chaque observation étant ici coûteuse à obtenir, il semble raisonnable de considérer la variable aléatoire S_n définie en équation (3.3) et de considérer que sa distribution aide à mesurer la sensibilité du produit aux variations du processus de fabrication. Néanmoins, celle-ci étant inaccessible, il était nécessaire de proposer des solutions nouvelles pour valider cette approche. Notre objectif était donc d'améliorer la procédure d'estimation par approche bayésienne et, par conséquent, de justifier son intérêt. Grâce à notre résultat principal sur l'ordre convexe entre S_n et la variable aléatoire R_n définie en équation (3.9), nous répondons à cet objectif. Par exemple, on déduit de la relation d'ordre convexe des approximations, par défaut et par excès, des quantiles de S_n. On montre également que l'on peut majorer les moments et la variance de S_n. Ces quantités sont très faciles à estimer par une méthode Monte-Carlo naïve.

Les perspectives concernant nos travaux sont nombreuses. Tout d'abord, il existe un nombre important de propriétés relatives à l'ordre convexe (voir (Shaked and Shanthikumar, 2007)). Il est donc possible, qu'en dehors de celles utilisées dans cette thèse, d'autres puissent être exploitées afin d'améliorer la connaissance que l'on a de la distribution de S_n. D'un point de vue théorique, il serait également intéressant d'étudier plus en profondeur la relation entre S_n et R_n. Sur ce sujet, plusieurs pistes de réflexion et références à l'appui ont été données dans ce chapitre. D'un point de vue pratique, on pourrait envisager de construire des approximations des quantiles de S_n plus performantes que celles données en proposition 3.5.2. En effet, nous avons pu constater dans différents exemples d'application que ces bornes ne convergent pas systématiquement vers la vraie valeur de p. Bien que cela semble difficile, il serait également pertinent de vérifier que la vraie probabilité p est contenue dans tout intervalle de crédibilité pour la loi de S_n.

Concernant les stratégies SUR, nous avons proposé un critère basé sur la variance de R_n. Afin de confirmer l'intérêt de ce critère, une justification théorique complète serait intéressante, par exemple, en se basant sur les résultats donnés dans (Bect et al., 2019) ou en se plaçant dans le cadre de la théorie des ensembles aléatoires (voir, par exemple, (Chevalier et al., 2013) et (Azzimonti et al., 2018)). De plus, il serait envisageable de mettre en œuvre une stratégie qui sélectionne plusieurs points à la fois, de façon à réduire le temps de calcul. Les travaux menés dans (Chevalier et al., 2014) sur ce sujet précis doivent pouvoir être étendus. En outre, les stratégies SUR n'ayant pas été initialement développées pour une sortie vectorielle, il serait intéressant de réfléchir à une approche dédiée, le cas multi-réponse étant une situation courante dans bon nombre de problématiques industrielles.

3.9 Démonstrations

Démonstration de la proposition 3.4.1

En préambule, on présente une définition et deux propositions qui sont utilisées dans la démonstration. La définition est la suivante :

Définition 3.9.1. *Soit $N \in \mathbb{N}^*$. Un vecteur aléatoire $(X_i)_{1\leq i\leq N} \in \mathbb{R}^N$ est dit comonotone si pour tout N-uplet $(x_i)_{1\leq i\leq N} \in \mathbb{R}^N$, on a :*

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_N \leq x_N) = \min_{1\leq i\leq N} \big(\mathbb{P}(X_i \leq x_i)\big).$$

Le concept de comonotonicité fait référence à la dépendance entre les marginales d'un vecteur aléatoire et s'applique, par exemple, à la gestion des risques financiers. Pour plus d'informations, on pourra consulter les sections 4 et 5 de (Dhaene et al., 2002), ou encore (Kaas et al., 2002). Les deux propositions suivantes font référence à l'ordre convexe présenté en annexe A, section A.2.3. Ces propositions peuvent respectivement être retrouvées en section 5 de (Kaas et al., 2002) et dans le chapitre 3 de (Shaked and Shanthikumar, 2007).

Proposition 3.9.1. *Soit $N \in \mathbb{N}^*$. Si le vecteur aléatoire $(X_i)_{1\leq i\leq N}$ est comonotone et que ses lois marginales sont les mêmes que celles du vecteur aléatoire $(Y_i)_{1\leq i\leq N}$, alors :*

$$\sum_{i=1}^{N} Y_i \leq_{cx} \sum_{i=1}^{N} X_i.$$

Proposition 3.9.2. *Soient X et Y deux variables aléatoires, et $(X_N)_{N\in\mathbb{N}^*}$ et $(Y_N)_{N\in\mathbb{N}^*}$ deux suites de variables aléatoires telles que $X_N \xrightarrow[N\to\infty]{\mathcal{L}} X$ et $Y_N \xrightarrow[N\to\infty]{\mathcal{L}} Y$. Si les deux propriétés suivantes sont vérifiées :*

(i) $\lim_{N\to\infty} \mathbb{E}|X_N| = \mathbb{E}|X|$ *et* $\lim_{N\to\infty} \mathbb{E}|Y_N| = \mathbb{E}|Y|$,

(ii) $X_N \leq_{cx} Y_N$, $\forall N \in \mathbb{N}^*$.

alors $X \leq_{cx} Y$.

Dans cette démonstration, on veut montrer qu'il existe une inégalité d'ordre convexe entre S_n et R_n. Pour cela, considérons une variable aléatoire U de loi uniforme sur $[0,1]$ et, pour tout $\mathbf{x} \in \mathbb{X}$, posons :

$$B_{\mathbf{x}} = \mathbb{1}_{p_n(\mathbf{x})>U} \in \{0,1\}.$$

Cette variable suit une loi de Bernoulli de paramètre $p_n(\mathbf{x})$ et sa fonction de répartition s'écrit :

$$\mathbb{P}(B_{\mathbf{x}} \leq b) = \begin{cases} 1 - p_n(\mathbf{x}) & \text{si } b \in [0,1), \\ 1 & \text{si } b \geq 1. \end{cases}$$

▷ Application de la définition 3.9.1. Commençons par montrer que, pour tout N-uplet $(\mathbf{x}_i)_{1\leq i\leq N} \in \mathbb{X}^N$, le vecteur aléatoire $(B_{\mathbf{x}_i})_{1\leq i\leq N}$ est comonotone, c'est-à-dire que :

$$\mathbb{P}(B_{\mathbf{x}_1} \leq b_1, \ldots, B_{\mathbf{x}_N} \leq b_N) = \min_{1\leq i\leq N} \big(\mathbb{P}(B_{\mathbf{x}_i} \leq b_i)\big), \tag{3.38}$$

où $b_i \in \mathbb{R}$, $i \in \{1, \dots, N\}$. Les variables aléatoires $(B_{\mathbf{x}_i})_{1 \leq i \leq N}$ ne prenant que les valeurs 0 et 1, il suffit de le montrer pour $b_i \in \{0, 1\}$. Pour cela, posons $E = \{b_1, \dots, b_N\}$ et commençons par considérer le cas particulier où tous les éléments de E sont égaux à 1. On a alors :

$$\mathbb{P}(B_{\mathbf{x}_i} \leq b_i) = 1, \ \forall i = 1, \dots, N,$$

et cela implique que :

$$\mathbb{P}(B_{\mathbf{x}_1} \leq b_1, \dots, B_{\mathbf{x}_N} \leq b_N) = 1 = \min_{1 \leq i \leq N} \big(\mathbb{P}(B_{\mathbf{x}_i} \leq b_i)\big).$$

L'égalité (3.38) est donc satisfaite. À présent, considérons le cas où E contient exactement $j \in \{1, \dots, N\}$ éléments égaux à 0. En notant $\mathfrak{S}(E)$ le groupe symétrique sur E, il existe une permutation $\sigma \in \mathfrak{S}(E)$ telle que :

$$\begin{cases} b_{\sigma(i)} = 0, & \forall i = 1, \dots, j, \\ b_{\sigma(i)} = 1, & \forall i = j+1, \dots, N. \end{cases}$$

Il en découle que :

$$\begin{aligned} &\mathbb{P}(B_{\mathbf{x}_1} \leq b_1, \dots, B_{\mathbf{x}_j} \leq b_j, B_{\mathbf{x}_{j+1}} \leq b_{j+1}, \dots, B_{\mathbf{x}_N} \leq b_N) \\ &= \mathbb{P}(B_{\mathbf{x}_{\sigma(1)}} \leq 0, \dots, B_{\mathbf{x}_{\sigma(j)}} \leq 0, B_{\mathbf{x}_{\sigma(j+1)}} \leq 1, \dots, B_{\mathbf{x}_{\sigma(N)}} \leq 1) \\ &= \mathbb{P}(B_{\mathbf{x}_{\sigma(1)}} \leq 0, \dots, B_{\mathbf{x}_{\sigma(j)}} \leq 0) \\ &= \mathbb{P}(p_n(\mathbf{x}_{\sigma(1)}) \leq U, \dots, p_n(\mathbf{x}_{\sigma(j)}) \leq U) \\ &= 1 - \max_{1 \leq i \leq j} \big(p_n(\mathbf{x}_{\sigma(i)})\big), \end{aligned}$$

car U est supposée uniforme sur $[0, 1]$. Par conséquent :

$$\begin{aligned} \mathbb{P}(B_{\mathbf{x}_1} \leq b_1, \dots, B_{\mathbf{x}_N} \leq b_N) &= \min_{1 \leq i \leq j} \big(1 - p_n(\mathbf{x}_{\sigma(i)})\big) \\ &= \min_{1 \leq i \leq j} \big(\mathbb{P}(B_{\mathbf{x}_{\sigma(i)}} \leq 0)\big) \\ &= \min_{1 \leq i \leq N} \big(\mathbb{P}(B_{\mathbf{x}_{\sigma(i)}} \leq b_{\sigma(i)})\big), \end{aligned}$$

car $\mathbb{P}(B_{\mathbf{x}_{\sigma(i)}} \leq b_{\sigma(i)}) = 1$, $\forall i = j+1, \dots, N$. Finalement,

$$\mathbb{P}(B_{\mathbf{x}_1} \leq b_1, \dots, B_{\mathbf{x}_N} \leq b_N) = \min_{1 \leq i \leq N} \big(\mathbb{P}(B_{\mathbf{x}_{\sigma(i)}} \leq b_{\sigma(i)})\big) = \min_{1 \leq i \leq N} \big(\mathbb{P}(B_{\mathbf{x}_i} \leq b_i)\big),$$

et l'égalité (3.38) est de nouveau vérifiée. On a donc démontré que, pour tout $(\mathbf{x}_i)_{1 \leq i \leq N} \in \mathbb{X}^N$, le vecteur $(B_{\mathbf{x}_i})_{1 \leq i \leq N}$ est comonotone.

▷ Application de la proposition 3.9.1. Pour tout N-uplet $(\mathbf{x}_i)_{1 \leq i \leq N} \in \mathbb{X}^N$, le vecteur $(B_{\mathbf{x}_i})_{1 \leq i \leq N} = (\mathbb{1}_{p_n(\mathbf{x}_i) > U})_{1 \leq i \leq N}$ étant comonotone, et ses distributions marginales étant les mêmes que celles du vecteur $(\mathbb{1}_{\xi_n(\mathbf{x}_i) > T})_{1 \leq i \leq N}$, la proposition 3.9.1 s'applique. Par conséquent, on a :

$$\sum_{i=1}^{N} \mathbb{1}_{\xi_n(\mathbf{x}_i) > T} \leq_{cx} \sum_{i=1}^{N} \mathbb{1}_{p_n(\mathbf{x}_i) > U},$$

ou, de façon équivalente,

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{\xi_n(\mathbf{x}_i)>T} \leq_{cx} \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{p_n(\mathbf{x}_i)>U}. \tag{3.39}$$

▷ Application de la proposition 3.9.2. Considérons une suite de variables aléatoires $(\mathbf{X}_i)_{1\leq i\leq N}$ i.i.d. suivant $\mathbf{P}_{\mathbf{X}}$. Pour tout $N \in \mathbb{N}^*$, on pose :

$$V_N = \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{\xi_n(\mathbf{X}_i)>T} \quad \text{et} \quad W_N = \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{p_n(\mathbf{X}_i)>U}.$$

D'après la loi forte des grands nombres, on a :

$$V_N \xrightarrow[N\to\infty]{p.s.} \mathbb{E}[\mathbb{1}_{\xi_n(\mathbf{X})>T} \mid \xi_n] = S_n,$$

et

$$W_N \xrightarrow[N\to\infty]{p.s.} \mathbb{E}[\mathbb{1}_{p_n(\mathbf{X})>U} \mid U] = R_n.$$

Soit h une fonction continue, alors par le théorème de continuité, sachant ξ_n, on a :

$$h(V_N) \xrightarrow[N\to\infty]{p.s.} h(S_n),$$

et, sachant U, on a :

$$h(W_N) \xrightarrow[N\to\infty]{p.s.} h(R_n).$$

Toutes les variables ci-dessus étant à valeurs dans $[0,1]$ et la fonction h continue, c'est-à-dire que les espérances existent, alors par le théorème de convergence dominée de Lebesgue (conditionnel), on a :

$$\mathbb{E}[h(V_N) \mid \xi_n] \xrightarrow[N\to\infty]{p.s.} \mathbb{E}[h(S_n) \mid \xi_n],$$

et

$$\mathbb{E}[h(W_N) \mid U] \xrightarrow[N\to\infty]{p.s.} \mathbb{E}[h(R_n) \mid U].$$

Cela implique que :

$$\lim_{N\to\infty} \mathbb{E}[h(V_N)] = \mathbb{E}[h(S_n)], \tag{3.40}$$

et

$$\lim_{N\to\infty} \mathbb{E}[h(W_N)] = \mathbb{E}[h(R_n)]. \tag{3.41}$$

Par conséquent, on a :

$$V_N \xrightarrow[N\to\infty]{\mathcal{L}} S_n \quad \text{et} \quad W_N \xrightarrow[N\to\infty]{\mathcal{L}} R_n.$$

En particulier, en prenant $h(x) = |x|$ dans (3.40) et (3.41), l'hypothèse *(i)* de la proposition 3.9.2 est vérifiée. De plus, on remarque que, en tenant compte simultanément de la définition A.2.3 d'un ordre convexe et de l'inégalité (3.39), pour toute fonction convexe φ et pour tout $N \in \mathbb{N}^*$, on a :

$$\mathbb{E}\left[\varphi\left(V_N\right) \mid \mathbf{X}_1, \ldots, \mathbf{X}_N\right] \leq \mathbb{E}\left[\varphi\left(W_N\right) \mid \mathbf{X}_1, \ldots, \mathbf{X}_N\right].$$

Cela implique que :

$$\mathbb{E}\left[\varphi(V_N)\right] \leq \mathbb{E}\left[\varphi(W_N)\right],$$

ou, de façon équivalente, que :

$$V_N \leq_{cx} W_N.$$

Ainsi, l'hypothèse *(ii)* de la proposition 3.9.2 est aussi vérifiée et cette proposition s'applique. Il en découle que $S_n \leq_{cx} R_n$.

Démonstration de la proposition 3.5.1

Soit U une variable aléatoire de loi uniforme sur $[0,1]$ et R_n la variable aléatoire à valeurs dans $[0,1]$ définie par : $R_n = \int_{\mathbb{X}} \mathbb{1}_{p_n(\mathbf{x})>U} \mathbf{P}_{\mathbf{X}}(d\mathbf{x})$. Dans cette démonstration, on veut montrer que, pour tout $\alpha \in [0,1]$, le quantile $F_{R_n}^{-1}(\alpha)$ d'ordre α de R_n satisfait :

$$F_{R_n}^{-1}(\alpha) = \int_{\mathbb{X}} \mathbb{1}_{p_n(\mathbf{x})>1-\alpha} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}).$$

Soient F_{R_n} la fonction de répartition de R_n et $G_{R_n} = 1 - F_{R_n}$ sa fonction de survie. On rappelle que $F_{R_n}^{-1}(\alpha)$ est défini par :

$$F_{R_n}^{-1}(\alpha) = \inf\{t \in [0,1] : F_{R_n}(t) \geq \alpha\} = \inf\{t \in [0,1] : G_{R_n}(t) \leq 1-\alpha\}.$$

Soit $G_{p_n(\mathbf{X})}$ la fonction de survie de la variable aléatoire $p_n(\mathbf{X})$ définie par :

$$G_{p_n(\mathbf{X})}(u) = \mathbb{P}(p_n(\mathbf{X}) > u) = \int_{\mathbb{X}} \mathbb{1}_{p_n(\mathbf{x})>u} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}), \quad \forall u \in [0,1].$$

Il vient immédiatement que : $G_{p_n(\mathbf{X})}(U) = \mathbb{P}(p_n(\mathbf{X}) > U \mid U) = R_n$. Considérons à présent la fonction $G_{p_n(\mathbf{X})}^{-1}$ définie pour tout $t \in [0,1]$ par :

$$G_{p_n(\mathbf{X})}^{-1}(t) = \inf\{u \in [0,1] : G_{p_n(\mathbf{X})}(u) \leq t\}.$$

Il est facile de vérifier que les fonctions $G_{p_n(\mathbf{X})}$ et $G_{p_n(\mathbf{X})}^{-1}$ vérifient :

$$G_{p_n(\mathbf{X})}(u) \leq t \Leftrightarrow u \geq G_{p_n(\mathbf{X})}^{-1}(t), \quad \forall u \in [0,1] \text{ et } \forall t \in [0,1].$$

Pour s'en convaincre, il suffit d'adapter la démonstration de la propriété d'équivalence (A.3) qui est proposée, par exemple, dans (Shorack, 2000). Pour tout $t \in [0,1]$, la fonction de survie G_{R_n} vérifie :

$$G_{R_n}(t) = \mathbb{P}(R_n > t) = \mathbb{P}(G_{p_n(\mathbf{X})}(U) > t) = \mathbb{P}(U < G_{p_n(\mathbf{X})}^{-1}(t)) = G_{p_n(\mathbf{X})}^{-1}(t).$$

Ainsi, pour tout $\alpha \in [0,1]$, on a :

$$\begin{aligned} F_{R_n}^{-1}(\alpha) = \inf\{t \in [0,1] : G_{R_n}(t) \leq 1-\alpha\} &= \inf\{t \in [0,1] : G^{-1}_{p_n(\mathbf{X})}(t) \leq 1-\alpha\} \\ &= \inf\{t \in [0,1] : G_{p_n(\mathbf{X})}(1-\alpha) \leq t\} \\ &= G_{p_n(\mathbf{X})}(1-\alpha) \\ &= \int_{\mathbb{X}} \mathbb{1}_{p_n(\mathbf{x})>1-\alpha} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}). \end{aligned}$$

Démonstration de la proposition 3.5.2

Dans cette démonstration, on veut montrer que pour tout $\alpha \in (0,1)$, on a :

$$\frac{\mu_n + \alpha - 1}{\alpha} \leq \frac{1}{\alpha}\int_0^\alpha F_{R_n}^{-1}(t)dt \leq F_{S_n}^{-1}(\alpha) \leq \frac{1}{1-\alpha}\int_\alpha^1 F_{R_n}^{-1}(t)dt \leq \frac{\mu_n}{1-\alpha}, \tag{3.42}$$

où :

$$\mu_n = \mathbb{E}[R_n] = \int_0^1 F_{R_n}^{-1}(t)dt.$$

D'après la proposition 3.4.1, on a : $S_n \leq_{cx} R_n$. Par conséquent, les propriétés *(iii)* et *(iv)* de la proposition A.2.3 s'appliquent. On peut alors affirmer que, pour tout $\alpha \in (0,1)$:

$$\int_0^\alpha F_{S_n}^{-1}(t)\,dt \geq \int_0^\alpha F_{R_n}^{-1}(t)\,dt, \quad \text{et} \quad \int_\alpha^1 F_{S_n}^{-1}(t)\,dt \leq \int_\alpha^1 F_{R_n}^{-1}(t)\,dt.$$

La fonction $F_{S_n}^{-1}$ étant croissante, on en déduit que :

$$F_{S_n}^{-1}(\alpha) = \frac{1}{\alpha}\int_0^\alpha F_{S_n}^{-1}(\alpha)dt \geq \frac{1}{\alpha}\int_0^\alpha F_{S_n}^{-1}(t)dt \geq \frac{1}{\alpha}\int_0^\alpha F_{R_n}^{-1}(t)dt.$$

Or,

$$\mu_n = \int_0^1 F_{R_n}^{-1}(t)dt \leq \int_0^\alpha F_{R_n}^{-1}(t)dt + 1 - \alpha,$$

car $0 \leq F_{R_n}^{-1}(t) \leq 1$, $\forall t \in [0,1]$. Par conséquent,

$$\mu_n + \alpha - 1 \leq \int_0^\alpha F_{R_n}^{-1}(t)dt.$$

On a ainsi :

$$\frac{\mu_n + \alpha - 1}{\alpha} \leq \frac{1}{\alpha}\int_0^\alpha F_{R_n}^{-1}(t)dt \leq F_{S_n}^{-1}(\alpha),$$

et les inégalités de gauche dans (3.42) sont démontrées. De la même façon, on a :

$$F_{S_n}^{-1}(\alpha) = \frac{1}{1-\alpha}\int_\alpha^1 F_{S_n}^{-1}(\alpha)dt \leq \frac{1}{1-\alpha}\int_\alpha^1 F_{S_n}^{-1}(t)dt \leq \frac{1}{1-\alpha}\int_\alpha^1 F_{R_n}^{-1}(t)dt,$$

et

$$\frac{1}{1-\alpha}\int_\alpha^1 F_{R_n}^{-1}(t)dt \leq \frac{1}{1-\alpha}\int_0^1 F_{R_n}^{-1}(t)dt = \frac{\mu_n}{1-\alpha},$$

ce qui démontre les inégalités de droite dans (3.42).

Démonstration de la proposition 3.5.3

D'après la proposition 3.5.1, la fonction quantile $F_{R_n}^{-1}$ de R_n vérifie :

$$F_{R_n}^{-1}(\alpha) = \int_{\mathbb{X}} \mathbb{1}_{\alpha > 1-p_n(\mathbf{x})} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}), \quad \forall \alpha \in [0,1].$$

Par conséquent, pour tout $\alpha \in (0,1)$, on a :

$$\begin{aligned}
\frac{1}{\alpha}\int_0^\alpha F_{R_n}^{-1}(t)dt &= \frac{1}{\alpha}\int_{\mathbb{X}} \left(\int_0^\alpha \mathbb{1}_{t>1-p_n(\mathbf{x})} dt \right) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \\
&= \frac{1}{\alpha}\int_{\mathbb{X}} \max\big(0, \alpha - 1 + p_n(\mathbf{x})\big) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \\
&= 1 - \int_{\mathbb{X}} \min\left(1, \frac{1-p_n(\mathbf{x})}{\alpha}\right) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}),
\end{aligned}$$

et

$$\begin{aligned}
\frac{1}{1-\alpha}\int_\alpha^1 F_{R_n}^{-1}(t)dt &= \frac{1}{1-\alpha}\int_{\mathbb{X}} \left(\int_\alpha^1 \mathbb{1}_{t>1-p_n(\mathbf{x})} dt \right) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \\
&= \frac{1}{1-\alpha}\int_{\mathbb{X}} \min\big(1-\alpha, p_n(\mathbf{x})\big) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \\
&= \int_{\mathbb{X}} \min\left(1, \frac{p_n(\mathbf{x})}{1-\alpha}\right) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}).
\end{aligned}$$

Démonstration de la proposition 3.5.5

Considérons les variables aléatoires S_n et R_n à valeurs dans $[0,1]$. Dans la preuve, nous avons besoin de montrer qu'il existe un ordre stochastique de dilatation entre ces deux variables aléatoires. Pour plus de clarté, on donne la définition de cet ordre stochastique, noté $\leq_{dil}$, et on évoque deux théorèmes qui seront utiles par la suite (voir le théorème 3.A.8 de (Shaked and Shanthikumar, 2007)) :

Définition 3.9.2. *Soit X et Y deux variables aléatoires de moyenne finie. Alors,*

$$X \leq_{dil} Y \Leftrightarrow [X - \mathbb{E}[X]] \leq_{cx} [Y - \mathbb{E}[Y]].$$

Théorème 3.9.1. *Soit X et Y deux variables aléatoires de moyenne finie. Alors,*

$$X \leq_{dil} Y \Leftrightarrow \frac{1}{1-\alpha}\int_\alpha^1 \left(F_X^{-1}(t) - F_Y^{-1}(t)\right) dt \leq \int_0^1 \left(F_X^{-1}(t) - F_Y^{-1}(t)\right) dt, \quad \forall \alpha \in (0,1)$$

Théorème 3.9.2. *Soit X et Y deux variables aléatoires de moyenne finie. Alors,*

$$X \leq_{dil} Y \Leftrightarrow \int_0^1 \left(F_X^{-1}(t) - \mathbb{E}[X]\right) \Psi(t) dt \leq \int_0^1 \left(F_Y^{-1}(t) - \mathbb{E}[Y]\right) \Psi(t) dt.$$

pour toute fonction croissante $\Psi : [0,1] \to \mathbb{R}^+$ telle que les intégrales existent.

Les deux variables aléatoires S_n et R_n ont la même espérance μ_n. Cela implique que :

$$0 = \mu_n - \mu_n = \int_0^1 \left(F_{S_n}^{-1}(t) - F_{R_n}^{-1}(t)\right) dt.$$

En outre, d'après la propriété *(iv)* de la proposition A.2.3 (voir annexe A), nous avons :

$$\frac{1}{1-\alpha}\int_\alpha^1 F_{S_n}^{-1}(t)dt \leq \frac{1}{1-\alpha}\int_\alpha^1 F_{R_n}^{-1}(t)dt, \quad \forall \alpha \in (0,1),$$

ou encore,

$$\frac{1}{1-\alpha}\int_\alpha^1 \left(F_{S_n}^{-1}(t) - F_{R_n}^{-1}(t)\right) dt \leq 0, \quad \forall \alpha \in (0,1).$$

Autrement dit,

$$\frac{1}{1-\alpha}\int_\alpha^1 \left(F_{S_n}^{-1}(t) - F_{R_n}^{-1}(t)\right) dt \leq \int_0^1 \left(F_{S_n}^{-1}(t) - F_{R_n}^{-1}(t)\right) dt, \quad \forall \alpha \in (0,1).$$

D'après le théorème 3.9.1, il existe donc un ordre de dilatation entre S_n et R_n.

Soit $\Psi : [0,1] \to \mathbb{R}^+$ une fonction croissante avec $M = \sup_{t\in[0,1]} \Psi(t) = \Psi(1)$ et vérifiant $\int_0^1 \Psi(t)dt = 1$. D'après le théorème 3.9.2, on a :

$$\int_0^1 \left(F_{S_n}^{-1}(t) - \mu_n\right) \Psi(t)dt \leq \int_0^1 \left(F_{R_n}^{-1}(t) - \mu_n\right) \Psi(t)dt,$$

ou encore,

$$\int_0^1 F_{S_n}^{-1}(t)\Psi(t)dt \leq \int_0^1 F_{R_n}^{-1}(t)\Psi(t)dt. \tag{3.43}$$

De plus, la fonction $F_{R_n}^{-1}$ étant à valeurs dans $[0,1]$, Ψ d'intégrale 1 et majorée par M, cela implique que :

$$\int_0^1 F_{R_n}^{-1}(t)\Psi(t)dt \leq \min\left(1, M\mu_n\right).$$

Démonstration du corollaire 3.5.1

Pour démontrer le corollaire 3.5.1, il suffit de remarquer que, en combinant les résultats de la proposition 3.5.1 avec ceux de la proposition 3.5.5, on a :

$$\begin{aligned}
\rho_\Psi(S_n) &\leq \int_0^1 F_{R_n}^{-1}(t)\Psi(t)dt \\
&= \int_0^1 \left[\int_{\mathbb{X}} \mathbb{1}_{t>1-p_n(\mathbf{x})} \mathbf{P}_{\mathbf{X}}(d\mathbf{x})\right] \Psi(t)dt \\
&= \int_{\mathbb{X}} \left[\int_0^{p_n(\mathbf{x})} \Psi(1-t)dt\right] \mathbf{P}_{\mathbf{X}}(d\mathbf{x}).
\end{aligned}$$

Démonstration de la proposition 3.6.1

Soit $n \in \mathbb{N}^*$. La démonstration qui suit s'inpire de celle de la proposition 3 dans (Bect et al., 2012), qui montre que $J_{S_n} \leq J_{n,k}$, $\forall k = 1, \ldots, 4$. Nous allons préciser cette relation d'ordre et montrer notamment que le critère J_{R_n} offre une meilleure approximation locale de J_{S_n} que les critères $(J_{n,k})_{k=1,\ldots,4}$.

D'après la relation d'ordre convexe établie à la proposition 3.4.1, on a $\mathrm{Var}[S_n] \leq \mathrm{Var}[R_n]$ et, par conséquent, $J_{S_n}(\mathbf{x}) \leq J_{R_n}(\mathbf{x})$, $\forall \mathbf{x} \in \mathbb{X}$. Par définition de R_n, on a :

$$R_n - \mathbb{E}[R_n] = \int_{\mathbb{X}} \left(\mathbb{1}_{p_n(\mathbf{x})>U} - p_n(\mathbf{x}) \right) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}).$$

Notons $\|X\| = \mathbb{E}[X^2]^{\frac{1}{2}}$ la norme euclidienne sur l'espace $L^2(\Omega, \mathcal{F}, \mathbb{P})$ des variables aléatoires de carré intégrable. D'après l'inégalité de Minkowski généralisée (voir (Vestrup, 2003)) et, compte tenu que :

$$\int_0^1 \mathbb{1}_{p_n(\mathbf{x})>u} du = p_n(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{X},$$

on peut écrire que :

$$\begin{aligned}
\|R_n - \mathbb{E}[R_n]\| &= \left(\int_0^1 \left(\int_{\mathbb{X}} \mathbb{1}_{p_n(\mathbf{x})>u} - p_n(\mathbf{x}) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \right)^2 du \right)^{\frac{1}{2}} \\
&\leq \int_{\mathbb{X}} \left(\int_0^1 \left(\mathbb{1}_{p_n(\mathbf{x})>u} - p_n(\mathbf{x}) \right)^2 du \right)^{\frac{1}{2}} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \\
&= \int_{\mathbb{X}} \left(p_n(\mathbf{x})(1 - p_n(\mathbf{x})) \right)^{\frac{1}{2}} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}).
\end{aligned}$$

Ainsi,

$$\begin{aligned}
\mathrm{Var}[R_n] = \|R_n - \mathbb{E}[R_n]\|^2 &\leq \left(\int_{\mathbb{X}} \left(p_n(\mathbf{x})(1 - p_n(\mathbf{x})) \right)^{\frac{1}{2}} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \right)^2 \\
&\leq \int_{\mathbb{X}} p_n(\mathbf{x})(1 - p_n(\mathbf{x})) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}),
\end{aligned}$$

d'après l'inégalité de Jensen. En outre, puisque pour tout $x \in [0, 1]$, on a $x(1 - x) \leq \min(x, 1 - x)$, il s'ensuit que :

$$\mathrm{Var}[R_n] \leq \int_{\mathbb{X}} \min(p_n(\mathbf{x}), 1 - p_n(\mathbf{x})) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}).$$

On a donc montré que :

$$J_{R_n}(\mathbf{x}) \leq J_{n,2}(\mathbf{x}) \leq J_{n,4}(\mathbf{x}) \leq J_{n,3}(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{X}.$$

D'après ce qui précède, on a de plus :

$$\mathrm{Var}[R_n] \leq \left(\int_{\mathbb{X}} \left(p_n(\mathbf{x})(1 - p_n(\mathbf{x})) \right)^{\frac{1}{2}} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \right)^2 \leq \left(\int_{\mathbb{X}} \min \left(p_n(\mathbf{x}), 1 - p_n(\mathbf{x}) \right)^{\frac{1}{2}} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \right)^2,$$

c'est-à-dire :

$$J_{R_n}(\mathbf{x}) \leq J_{n,2}(\mathbf{x}) \leq J_{n,1}(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{X}.$$

On a donc montré la relation d'ordre suivante :

$$J_{S_n} \leq J_{R_n} \leq J_{n,k}, \quad \forall k = 1, \ldots, 4.$$

Démonstration de la proposition 3.6.2

Soit U une variable aléatoire uniforme sur $[0,1]$ et R_n la variable aléatoire définie en équation (3.9). L'espérance de R_n s'écrit :

$$\mathbb{E}[R_n] = \int_{\mathbb{X}} p_n(\mathbf{x}) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}).$$

Par conséquent, la variance de R_n est égale à :

$$\begin{aligned}
\mathrm{Var}[R_n] &= \mathbb{E}\left[\left(\int_{\mathbb{X}} \mathbb{1}_{p_n(\mathbf{x})>U} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) - \int_{\mathbb{X}} p_n(\mathbf{x}) \mathbf{P}_{\mathbf{X}}(d\mathbf{x})\right)^2\right] \\
&= \int_{\mathbb{X}^2} \mathbb{E}\left[\mathbb{1}_{p_n(\mathbf{x})>U} \mathbb{1}_{p_n(\mathbf{y})>U}\right] \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \mathbf{P}_{\mathbf{X}}(d\mathbf{y}) - \int_{\mathbb{X}^2} p_n(\mathbf{x}) p_n(\mathbf{y}) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \mathbf{P}_{\mathbf{X}}(d\mathbf{y}) \\
&= \int_{\mathbb{X}^2} \big(\min(p_n(\mathbf{x}), p_n(\mathbf{y})) - p_n(\mathbf{x}) p_n(\mathbf{y})\big) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \mathbf{P}_{\mathbf{X}}(d\mathbf{y}), \\
&= \int_{\mathbb{X}^2} \left(p_n(\mathbf{x})(1 - p_n(\mathbf{y})) \mathbb{1}_{p_n(\mathbf{x})<p_n(\mathbf{y})} + p_n(\mathbf{y})(1 - p_n(\mathbf{x})) \mathbb{1}_{p_n(\mathbf{x})\geq p_n(\mathbf{y})}\right) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) \mathbf{P}_{\mathbf{X}}(d\mathbf{y}), \\
&= \int_{\mathbb{X}} \Bigg(p_n(\mathbf{x}) \int_{\mathbb{X}} (1 - p_n(\mathbf{y})) \mathbb{1}_{p_n(\mathbf{x})<p_n(\mathbf{y})} \mathbf{P}_{\mathbf{X}}(d\mathbf{y}) \\
&\quad + (1 - p_n(\mathbf{x})) \int_{\mathbb{X}} p_n(\mathbf{y}) \mathbb{1}_{p_n(\mathbf{x})\geq p_n(\mathbf{y})} \mathbf{P}_{\mathbf{X}}(d\mathbf{y})\Bigg) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}).
\end{aligned}$$

On pose $\eta_n(\mathbf{x}) = p_n(\mathbf{x}) \int_{\mathbb{X}} (1 - p_n(\mathbf{y})) \mathbb{1}_{p_n(\mathbf{x})<p_n(\mathbf{y})} \mathbf{P}_{\mathbf{X}}(d\mathbf{y}) + (1 - p_n(\mathbf{x})) \int_{\mathbb{X}} p_n(\mathbf{y}) \mathbb{1}_{p_n(\mathbf{x})\geq p_n(\mathbf{y})} \mathbf{P}_{\mathbf{X}}(d\mathbf{y})$.
Il vient que :

$$\mathrm{Var}[R_n] = \int_{\mathbb{X}} \eta_n(\mathbf{x}) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}).$$

Démonstration de la proposition 3.6.3

On suppose dans cette démonstration que la loi de la variable aléatoire $p_n(\mathbf{X})$ admet une densité f_{p_n} par rapport à la mesure de Lebesgue. On note G_{p_n} sa fonction de survie et F_{p_n} sa fonction de répartition :

$$G_{p_n}(u) = \mathbb{P}(p_n(\mathbf{X}) > u) = \int_{\mathbb{X}} \mathbb{1}_{p_n(\mathbf{x})>u} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) = 1 - F_{p_n}(u), \quad \forall u \in [0,1].$$

L'espérance μ_n de S_n s'écrit alors :

$$\mu_n = \mathbb{E}[S_n] = \mathbb{E}[p_n(\mathbf{X})] = \int_0^1 u f_{p_n}(u)du = \int_0^1 \mathbb{P}(p_n(\mathbf{X}) > u)du = \int_0^1 G_{p_n}(u)du.$$

Considérons la fonction η_n donnée à la proposition 3.6.2 et définie pour tout $\mathbf{x} \in \mathbb{X}$ par :

$$\eta_n(\mathbf{x}) = p_n(\mathbf{x}) \int_{\mathbb{X}} \big(1 - p_n(\mathbf{y})\big) \mathbb{1}_{p_n(\mathbf{x}) < p_n(\mathbf{y})} \mathbf{P}_{\mathbf{X}}(d\mathbf{y}) + \big(1 - p_n(\mathbf{x})\big) \int_{\mathbb{X}} p_n(\mathbf{y}) \mathbb{1}_{p_n(\mathbf{x}) \geq p_n(\mathbf{y})} \mathbf{P}_{\mathbf{X}}(d\mathbf{y}). \tag{3.44}$$

On va montrer que cette fonction admet un maximum global sur $[0,1]$ au point :

$$q_n^* \;=\; \mathbb{P}(R_n > \mu_n).$$

Pour cela, commençons par remarquer que, d'après ce qui précède, la fonction η_n se réécrit :

$$\begin{aligned}
\eta_n(\mathbf{x}) &= p_n(\mathbf{x}) \mathbb{E}\big[(1 - p_n(\mathbf{X})) \mathbb{1}_{p_n(\mathbf{x}) \leq p_n(\mathbf{X})}\big] + \big(1 - p_n(\mathbf{x})\big) \mathbb{E}\big[p_n(\mathbf{X}) \mathbb{1}_{p_n(\mathbf{x}) \geq p_n(\mathbf{X})}\big] \\
&= p_n(\mathbf{x}) \int_{p_n(\mathbf{x})}^1 (1-u) f_{p_n}(u)du + \big(1 - p_n(\mathbf{x})\big) \int_0^{p_n(\mathbf{x})} u f_{p_n}(u)du \\
&= p_n(\mathbf{x}) \int_{p_n(\mathbf{x})}^1 f_{p_n}(u)du - \mu_n p_n(\mathbf{x}) + \int_0^{p_n(\mathbf{x})} u f_{p_n}(u)du
\end{aligned}$$

À l'aide d'une intégration par parties appliquée au dernier membre de droite, on obtient :

$$\begin{aligned}
\eta_n(\mathbf{x}) &= p_n(\mathbf{x})\big(1 - F_{p_n}(p_n(\mathbf{x}))\big) - \mu_n p_n(\mathbf{x}) + \Big[p_n(\mathbf{x}) F_{p_n}(p_n(\mathbf{x})) - \int_0^{p_n(\mathbf{x})} F_{p_n}(u)du\Big] \\
&= \int_0^{p_n(\mathbf{x})} G_{p_n}(u)du - \mu_n p_n(\mathbf{x}).
\end{aligned}$$

À présent, posons $\varphi(q) = \displaystyle\int_0^q G_{p_n}(u)du - \mu_n q$, $\forall q \in [0,1]$. Alors,

$$\varphi'(q) = 0 \Leftrightarrow G_{p_n}(q) = \mu_n \Leftrightarrow \int_{\mathbb{X}} \mathbb{1}_{p_n(\mathbf{x}) > q} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) = \mu_n. \tag{3.45}$$

Soit $G_{p_n}^{-1}$ la fonction définie pour tout $t \in [0,1]$ par :

$$G_{p_n}^{-1}(t) = \inf\{u \in [0,1] : G_{p_n}(u) \leq t\}.$$

Comme mentionné dans la démonstration 3.9 de la proposition 3.5.1, on a :

$$G_{p_n}^{-1}(t) = \mathbb{P}(R_n > t), \quad \text{et} \quad G_{p_n}(u) \leq t \Leftrightarrow u \geq G_{p_n}^{-1}(t), \quad \forall u \in [0,1],\ \forall t \in [0,1].$$

Ainsi,

$$G_{p_n}(q) = \mu_n \Leftrightarrow q = \mathbb{P}(R_n > \mu_n),$$

et la fonction η_n admet donc un unique extremum au point $q_n^* = \mathbb{P}(R_n > \mu_n)$ qui, d'après (3.45), vérifie : $\int_{\mathbb{X}} \mathbb{1}_{p_n(\mathbf{x}) > q_n^*} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}) = \mu_n$. En étudiant le signe de φ', on vérifie aisément qu'il s'agit d'un maximum.

Chapitre 4

Estimation d'une probabilité de défaillance par découpage dyadique récursif

4.1 Introduction

Soit $\mathbb{X} = [0,1]^d$ et $g : \mathbb{X} \to \mathbb{R}$ une fonction boîte-noire qui peut être évaluée en tout point $\mathbf{x} \in \mathbb{X}$, mais qui est coûteuse en temps de calcul. Considérons un vecteur aléatoire noté $\mathbf{X} = (X_1, \ldots, X_d)$, à valeurs dans $\mathbb{X} = [0,1]^d$, dont la loi $\mathbf{P}_{\mathbf{X}}$ admet une densité $f_{\mathbf{X}}$ par rapport à la mesure de Lebesgue. On suppose que $f_{\mathbf{X}}$ est connue, éventuellement à une constante de normalisation près. Cela signifie que l'on peut appliquer un algorithme de Metropolis-Hastings pour simuler suivant la loi $\mathbf{P}_{\mathbf{X}}$ ou suivant une restriction de cette loi à un sous-ensemble de $\mathbb{X}$. Étant donné un seuil $T \in \mathbb{R}$, on s'intéresse à l'estimation d'une probabilité de défaillance notée p et définie par :

$$p = \mathbb{P}(g(\mathbf{X}) > T) = \mathbf{P}_{\mathbf{X}}(\mathbb{F}) > 0, \tag{4.1}$$

où $\mathbb{F} = \{\mathbf{x} \in \mathbb{X} : g(\mathbf{x}) > T\} \subset \mathbb{X}$ est le domaine de défaillance. En pratique, la probabilité p est très petite (mais non nulle) et on dit que l'évènement $\{\mathbf{X} \in \mathbb{F}\}$ est rare. Typiquement, la méthode Monte-Carlo naïve ne peut être envisagée dans ce contexte (voir section 2.2.1). Une solution pour simuler de façon à forcer l'apparition de l'évènement rare est d'utiliser une méthode de réduction de variance. Une des plus connues est la méthode de splitting (« Multilevel Splitting » ou encore « Subset Simulation ») que nous avons présentée en section 2.2.3. Pour rappel, elle consiste à écrire l'évènement rare en une suite emboîtée de sous-évènements dont la probabilité d'occurrence est suffisamment élevée pour être estimée, par exemple, par une méthode Monte-Carlo naïve. Plus précisément, considérons les ensembles $\mathbb{F} = \mathbb{F}_q \subset \ldots \subset \mathbb{F}_1 \subset \mathbb{F}_0 = \mathbb{X}$, avec $q \in \mathbb{N}^*$. Alors, par la formule de Bayes, on peut écrire p comme le produit de probabilités conditionnelles suivant :

$$p = \mathbb{P}(\mathbf{X} \in \mathbb{F}) = \prod_{j=1}^{q} \mathbb{P}(\mathbf{X} \in \mathbb{F}_j \mid \mathbf{X} \in \mathbb{F}_{j-1}) = \prod_{j=1}^{q} p_j.$$

Le principe est de déterminer les $\mathbb{F}_j$ de sorte que les probabilités $(p_j)_{1\leq j\leq q}$ ne soient pas trop petites, donc plus faciles à estimer que p (pour plus de détails, voir le chapitre 3 de (Rubino and Tuffin, 2009), (Au and Beck, 2001) et (Cérou et al., 2012)).

Dans le chapitre 3, nous avons développé des méthodes d'estimation de la probabilité de défaillance p qui nécessitent de faire l'hypothèse que la fonction g est une trajectoire d'un processus aléatoire. Cette modélisation entraîne un a priori sur la régularité de la fonction g à travers le choix de la fonction de corrélation du processus (voir (Adler, 1981) et (Rasmussen and Williams, 2006)). Dans ce chapitre, nous développons une méthode d'estimation de probabilité de défaillance basée sur des hypothèses moins fortes. Une propriété requise pour la fonction g est notamment d'être lipschitzienne. Cette hypothèse est également utilisée dans (Cohen et al., 2013) – et les références qui s'y trouvent – pour mettre en œuvre un algorithme itératif de minimisation de fonction. Celui-ci s'appuie sur une partition dyadique de l'ensemble $\mathbb{X}$ pour déterminer les points en lesquels évaluer la fonction g et en déduire une valeur approchée du minimum de la fonction.

L'algorithme que l'on propose repose sur le même principe que celui donné dans (Cohen et al., 2013). À chaque itération, on procède à un découpage dyadique récursif de $\mathbb{X}$ pour déterminer les points en lesquels évaluer la fonction g, et on en déduit une approximation du domaine de défaillance $\mathbb{F}$. Une méthode de splitting est utilisée en parallèle pour fournir une estimation de la probabilité de défaillance. L'estimation de p que retourne notre algorithme est donc entachée de deux types d'erreur : une erreur d'approximation et une erreur d'estimation. On montre que l'erreur d'approximation décroît à vitesse géométrique avec le nombre d'appels à la fonction g et que l'erreur d'estimation est en $\mathcal{O}\big(\frac{1}{\sqrt{N}}\big)$, où N est la taille de l'échantillon aléatoire pour la méthode de splitting. En pratique, N peut être choisi grand car la méthode de splitting ne requiert pas d'appels supplémentaires à la fonction g. À titre de comparaison, les méthodes FORM et SORM, que nous avons présentées en section 2.3.1, s'appuient également sur une approximation du domaine de défaillance $\mathbb{F}$ pour estimer p. Néanmoins, elles requièrent des hypothèses de régularité et de différentiabilité sur la fonction g, dans le but d'approcher le domaine de défaillance $\mathbb{F}$ par des hypersurfaces linéaires et quadratiques via des développements de Taylor à l'ordre 1 et 2 (voir le chapitre 4 de (Lebrun, 2013) et les références qui s'y trouvent). L'erreur d'approximation commise est difficile à mesurer pour ces méthodes.

Le plan de ce chapitre est le suivant. En section 4.2, on commence par énoncer les hypothèses que doit vérifier la fonction g pour que l'algorithme puisse fonctionner. Celui-ci est entièrement décrit et analysé en section 4.3, où il est notamment expliqué que l'on peut lui associer une structure d'arbre 2^d-régulier. Un exemple simple en dimension $d = 1$ est aussi proposé pour aider à la compréhension. En section 4.4, on présente la méthode de splitting appliquée à notre problème. Enfin, on donne quelques remarques et perspectives en section 4.5.

4.2 Hypothèses

L'algorithme que l'on propose nécessite de faire deux hypothèses sur la fonction g. La première sert à la mise en œuvre pratique (voir ci-dessous et la section 4.3.3) et la seconde

à l'analyse théorique (voir section 4.3.5). Enfin, nous ferons une hypothèse supplémentaire sur la densité $f_{\mathbf{X}}$. On verra que ces hypothèses assurent, d'une part, que le nombre d'appels à la fonction g qu'effectue l'algorithme croît linéairement (et non exponentiellement) en fonction du nombre d'itérations et, d'autre part, que l'erreur d'approximation converge vers 0 à vitesse géométrique.

4.2.1 Fonction lipschitzienne

L'algorithme fonctionne sous l'hypothèse que g est une fonction lipschitzienne, c'est-à-dire qu'il existe une constante $L > 0$ telle que pour tout $(\mathbf{x}, \mathbf{x}') \in \mathbb{X}^2$:

$$|g(\mathbf{x}) - g(\mathbf{x}')| \leq L \left\|\mathbf{x} - \mathbf{x}'\right\|, \tag{$\mathcal{H}_1$}$$

où $\|\mathbf{x}\| = \max_{1 \leq i \leq d} |\mathbf{x}_i|$ est la norme infinie sur $\mathbb{X}$.

Par exemple, la fonction représentée à gauche de la figure 4.1 est valide (il s'agit de la fonction que l'on étudie en section 4.3.4), tandis que celle représentée à droite ne l'est pas (il s'agit de la fonction définie par $g(x) = \sin\left(\frac{1}{x}\right)$ si $x \in (0, 1]$ et $g(0) = 0$).

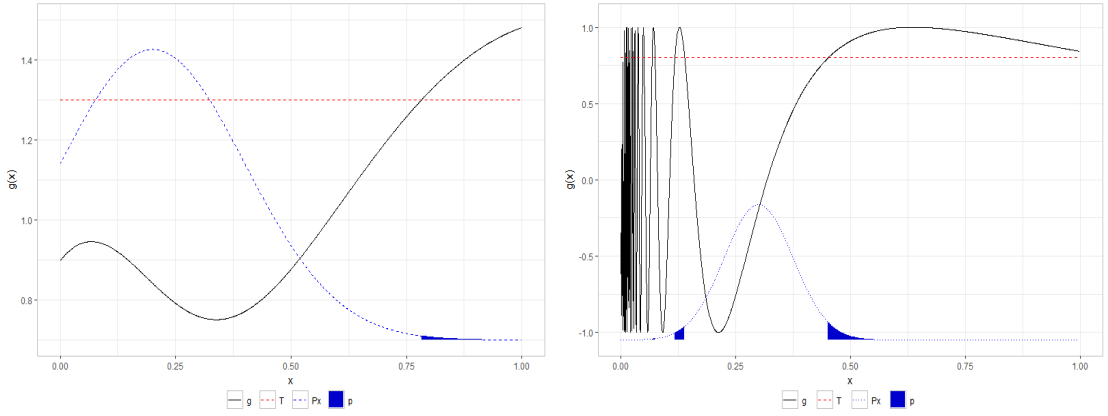

FIGURE 4.1 – À gauche : la fonction vérifie ($\mathcal{H}_1$). À droite : la fonction ne vérifie pas ($\mathcal{H}_1$).

Soit $k \in \mathbb{N}^*$. Pour comprendre l'intérêt de l'hypothèse ($\mathcal{H}_1$), considérons un cube Q de côté $\frac{1}{2^{k-1}}$ et notons $\mathbf{c}_Q$ son centre. Par définition de Q, on a :

$$\|\mathbf{x} - \mathbf{c}_Q\| \leq \frac{1}{2^k}, \quad \forall \mathbf{x} \in Q.$$

L'hypothèse ($\mathcal{H}_1$) implique donc que :

$$g(\mathbf{c}_Q) - \frac{L}{2^k} \leq g(\mathbf{x}) \leq g(\mathbf{c}_Q) + \frac{L}{2^k}, \quad \forall \mathbf{x} \in Q.$$

Ainsi, si $T < g(\mathbf{c}_Q) - \frac{L}{2^k}$ (resp. $g(\mathbf{c}_Q) + \frac{L}{2^k} \leq T$), alors on a nécessairement $Q \subset \mathbb{F}$ (resp. $Q \cap \mathbb{F} = \emptyset$). Tandis que si $|g(\mathbf{c}_Q) - T| \leq \frac{L}{2^k}$, alors on ne peut pas statuer sur $Q \cap \mathbb{F}$.

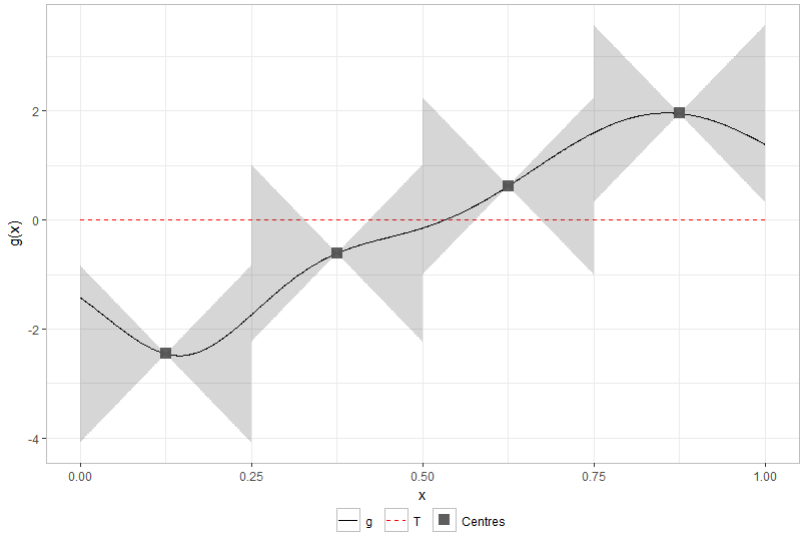

FIGURE 4.2 – Exemple d'application de $(\mathcal{H}_1)$ en dimension 1. La fonction g représentée par la courbe noire est inconnue en pratique. Seules ses valeurs aux points $\left\{\frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}\right\}$ sont observées. Les aires grises sont les valeurs susceptibles d'être prises par g dans les intervalles $\left\{\left[0, \frac{1}{4}\right], \left[\frac{1}{4}, \frac{1}{2}\right], \left[\frac{1}{2}, \frac{3}{4}\right], \left[\frac{3}{4}, 1\right]\right\}$ sous l'hypothèse $(\mathcal{H}_1)$. Lorsqu'on compare une aire grise au seuil T, trois conclusions sont possibles : soit l'intervalle correspondant à cette aire est inclus dans le domaine de défaillance (c'est le cas de $\left[\frac{3}{4}, 1\right]$), soit l'intersection est vide (c'est le cas de $\left[0, \frac{1}{4}\right]$), soit on ne peut pas conclure (c'est le cas de $\left[\frac{1}{4}, \frac{1}{2}\right], \left[\frac{1}{2}, \frac{3}{4}\right]$).

Sur la figure 4.2, où la fonction $g : [0,1] \to \mathbb{R}$ est L-lipschitzienne, on fait apparaître chacune de ces trois situations. Pour tout intervalle $Q \in \left\{\left[0, \frac{1}{4}\right], \left[\frac{1}{4}, \frac{1}{2}\right], \left[\frac{1}{2}, \frac{3}{4}\right], \left[\frac{3}{4}, 1\right]\right\}$, de centre $\mathbf{c}_Q \in \left\{\frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}\right\}$, on a représenté la surface délimitée par les courbes représentatives des fonctions $\mathbf{x} \in Q \mapsto g(\mathbf{c}_Q) + L|\mathbf{x} - \mathbf{c}_Q|$ et $\mathbf{x} \in Q \mapsto g(\mathbf{c}_Q) - L|\mathbf{x} - \mathbf{c}_Q|$. La plus petite valeur que prend g dans un intervalle Q est $g(\mathbf{c}_Q) - \frac{L}{8}$, et la plus grande est $g(\mathbf{c}_Q) + \frac{L}{8}$. Ainsi, puisque $g\left(\frac{1}{8}\right) + \frac{L}{8} < T$, alors $\left[0, \frac{1}{4}\right] \cap \mathbb{F} = \emptyset$. Au contraire, puisque $g\left(\frac{7}{8}\right) - \frac{L}{8} > T$, alors $\left[\frac{3}{4}, 1\right] \subset \mathbb{F}$. Pour les intervalles $\left[\frac{1}{4}, \frac{1}{2}\right]$ et $\left[\frac{1}{2}, \frac{3}{4}\right]$, le domaine dans lequel peut évoluer g sous $(\mathcal{H}_1)$ est traversé par le seuil T, donc on ne peut pas se prononcer.

4.2.2 Ensemble borné

Dans ce qui suit, on note $\lambda(E)$ la mesure de Lebesgue d'un ensemble mesurable $E \subseteq \mathbb{X}$.

Considérons un point $\mathbf{x}_T \in \mathbb{X}$ tel que $g(\mathbf{x}_T) = T$. L'existence de ce point est assurée car p est strictement positive et g est continue d'après $(\mathcal{H}_1)$. La seconde hypothèse que l'on fait sur g garantit pour tout $\delta \in \mathbb{R}^*$ un contrôle de la mesure de Lebesgue de l'ensemble :

$$\{\mathbf{x} \in \mathbb{X} : |g(\mathbf{x}) - g(\mathbf{x}_T)| \leq \delta\}.$$

Pour fixer les idées, comparons au préalable les deux fonctions représentées figure 4.3, vérifiant $(\mathcal{H}_1)$ avec la même constante $L \approx 0.4$, et traversant le seuil T au point $\mathbf{x}_T \approx 0.6$.

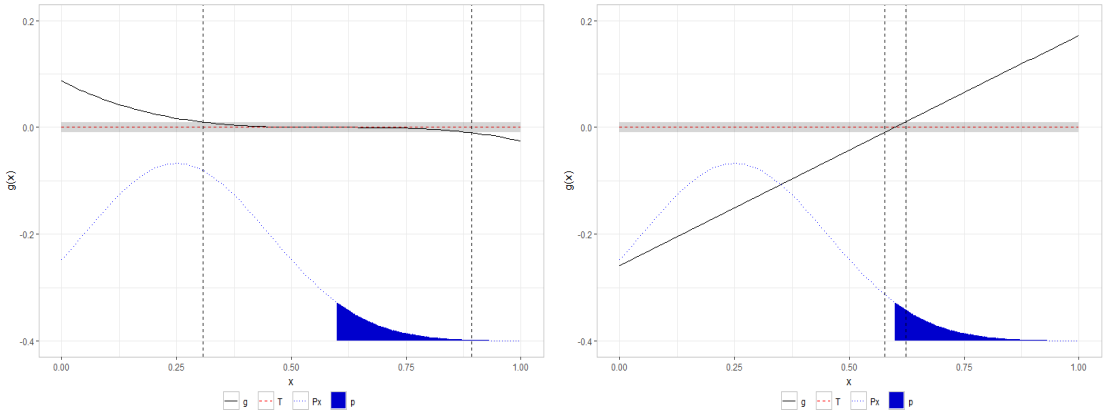

FIGURE 4.3 – Exemples de fonctions qui vérifient $(\mathcal{H}_1)$ avec la même constante $L \approx 0.4$. À gauche : la fonction vérifie $(\mathcal{H}_2)$ avec $C \gg 1$. À droite : la fonction vérifie $(\mathcal{H}_2)$ avec $C = 1$.

Pour chacune d'elle, on a tracé l'intervalle $\{\mathbf{x} \in [0,1] : |g(\mathbf{x}) - g(\mathbf{x}_T)| \leq 0.01\}$. À gauche, la longueur de cet intervalle est grande. C'est un exemple typique de situation que l'on cherche à éviter car la fonction traverse lentement le seuil et, par conséquent, les bornes du domaine de défaillance sont difficiles à localiser. À droite, la longueur de l'intervalle est au contraire très petite, et la fonction traverse favorablement le seuil T dans le sens où la position du point $\mathbf{x}_T$ est facile à repérer.

Notons que la fonction représentée à droite est affine, c'est-à-dire lipschitzienne de constante la pente de la droite. On peut alors vérifier que pour tout $\delta > 0$, la longueur exacte de l'intervalle $\{\mathbf{x} \in [0,1] : |g(\mathbf{x}) - g(\mathbf{x}_T)| \leq \delta\}$ est $\frac{2\delta}{L}$. Dans la classe des fonctions $g : [0,1] \to \mathbb{R}$ qui vérifient $(\mathcal{H}_1)$ avec la même constante L, cette longueur est la plus petite possible. En effet, pour toute fonction L-lipschitzienne $g : \mathbb{X} \to \mathbb{R}$, on a :

$$\|\mathbf{x} - \mathbf{x}_T\| \leq \frac{\delta}{L} \Rightarrow |g(\mathbf{x}) - g(\mathbf{x}_T)| \leq \delta, \quad \forall \delta > 0.$$

Par conséquent,

$$\lambda\big(\{\mathbf{x} \in \mathbb{X} : |g(\mathbf{x}) - g(\mathbf{x}_T)| \leq \delta\}\big) \geq \lambda\left(\left\{\mathbf{x} \in \mathbb{X} : \|\mathbf{x} - \mathbf{x}_T\| \leq \frac{\delta}{L}\right\}\right) = \left(\frac{2\delta}{L}\right)^d, \quad \forall \delta > 0. \tag{4.2}$$

Posons :

$$C = \sup_{\delta > 0} \frac{\lambda\big(\{\mathbf{x} \in \mathbb{X} : |g(\mathbf{x}) - g(\mathbf{x}_T)| \leq \delta\}\big)}{\left(\frac{2\delta}{L}\right)^d} \in [0, +\infty].$$

On a nécessairement $C \geq 1$ d'après l'inégalité (4.2) et la seconde hypothèse que l'on fait sur g est alors :

$$C < +\infty. \tag{$\mathcal{H}_2$}$$

Autrement dit, on suppose que la fonction g satisfait :

$$\lambda\big(\{\mathbf{x} \in \mathbb{X} : |g(\mathbf{x}) - g(\mathbf{x}_T)| \leq \delta\}\big) \leq C \left(\frac{2\delta}{L}\right)^d, \ \forall \delta > 0,$$

où $1 \leq C < +\infty$.

Les fonctions représentées figure 4.3 vérifient $(\mathcal{H}_2)$, avec une grande valeur de C à gauche, et $C = 1$ à droite. Contrairement à $(\mathcal{H}_1)$, l'hypothèse $(\mathcal{H}_2)$ n'a pas besoin d'être vérifiée pour appliquer l'algorithme, dans le sens où il n'est pas nécessaire de connaître la valeur de C pour la mise en œuvre pratique. Cependant, on verra en section 4.3.5 qu'elle est utile d'un point de vue théorique pour analyser les performances de ce dernier.

Dans tout ce qui suit, on suppose que la fonction g vérifie les hypothèses $(\mathcal{H}_1)$ et $(\mathcal{H}_2)$, et notamment qu'une constante L vérifiant $(\mathcal{H}_1)$ est connue.

4.2.3 Densité bornée

La troisième hypothèse que nous ferons concerne la densité $f_{\mathbf{X}}$. On pose :

$$\|f_{\mathbf{X}}\| = \sup_{\mathbf{x} \in \mathbb{X}} f_{\mathbf{X}}(\mathbf{x}),$$

et on suppose que $f_{\mathbf{X}}$ est majorée et strictement positive :

$$\forall \mathbf{x} \in \mathbb{X}, \quad 0 < f_{\mathbf{X}}(\mathbf{x}) < \|f_{\mathbf{X}}\| < \infty. \tag{$\mathcal{H}_3$}$$

Nous verrons en section 4.3.5 que cette hypothèse assure que l'erreur d'approximation de l'algorithme n'explose pas.

4.3 Algorithme avec probabilités connues

Commençons par rappeler qu'un cube dyadique $Q \subset \mathbb{X}$ de côté $\frac{1}{2^m}$, où $m \in \mathbb{N}^*$, est un pavé qui s'écrit sous la forme :

$$Q = \prod_{i=1}^{d} \left[\frac{k_i}{2^m}, \frac{k_i + 1}{2^m}\right[,$$

où $m \in \mathbb{N}^*$ et k_i un entier tel que : $0 \leq k_i \leq 2^m - 1, \forall 1 \leq i \leq d$.

On peut écrire $\mathbb{X}$ comme la réunion de trois sous-ensembles de cubes dyadiques disjoints : ceux qui sont inclus dans le domaine de défaillance $\mathbb{F}$, ceux qui sont disjoints de $\mathbb{F}$, et les autres. Respectivement notés $\mathcal{I}$ pour « in », $\mathcal{O}$ pour « out » et $\mathcal{A}$ pour « alive », on les définit de la façon suivante :

$$\mathcal{I} = \{Q \in \mathbb{X} : Q \subseteq \mathbb{F}\}, \ \mathcal{O} = \{Q \in \mathbb{X} : Q \cap \mathbb{F} = \emptyset\}, \text{ et } \mathcal{A} = \mathbb{X} \setminus (\mathcal{I} \cup \mathcal{O}). \tag{4.3}$$

Le domaine de défaillance $\mathbb{F}$ vérifie ainsi : $\mathcal{I} \subseteq \mathbb{F} \subseteq \mathcal{I} \cup \mathcal{A}$.

4.3.1 Partitionnement dyadique et structure d'arbre étiqueté

L'algorithme que l'on propose procède par découpage dyadique récursif de $\mathbb{X}$. Par conséquent, on peut lui associer une suite d'arbres étiquetés 2^d-réguliers.

4.3.1.1 Formalisme

Un arbre 2^d-régulier est un arbre dont chaque nœud a 2^d enfants. Ici, chaque nœud est associé à un unique cube dyadique dans $\mathbb{X}$. L'ensemble des cubes identifiés par les nœuds terminaux forment une partition de $\mathbb{X}$. La racine est le seul nœud à ne pas avoir de parent.

Dans la suite, on utilise le formalisme de (Neveu, 1986). La racine s'écrit $\emptyset$ et elle correspond à $\mathbb{X}$ tout entier. On dit que c'est la génération 0 de l'arbre. Un nœud à distance $k \in \mathbb{N}^*$ de la racine appartient à la génération k, et s'écrit comme un vecteur à k coordonnées $u = (u_1, \dots, u_k)$, où $u_i \in \{1, 2, \dots, 2^d\}$, $\forall i = 1, \dots, k$. On dit aussi qu'il est de profondeur k et on note $|u| = k$, avec la convention que la racine est de profondeur nulle. La hauteur de l'arbre est la profondeur du nœud le plus éloigné de la racine.

Un nœud u de profondeur $k \in \mathbb{N}$ réfère à un cube dyadique, de côté $\frac{1}{2^k}$, que l'on note $Q(u)$ (avec $Q(\emptyset) = \mathbb{X}$) et qui vérifie donc : $\lambda\big(Q(u)\big) = \frac{1}{2^{dk}}$. Pour écrire u comme un vecteur de coordonnées, il convient de choisir une façon de numéroter les sous-cubes et de conserver l'ordre de numérotation.

Par exemple, si l'on compare l'arbre de la figure 4.4 à la partition dyadique de $[0,1]^2$ représentée figure 4.5, on voit immédiatement la correspondance entre nœuds, cubes et ordre de numérotation : l'ensemble $[0,1]^2$ est identifié par la racine $\emptyset$, le nœud **(2)** réfère au cube $\left[\frac{1}{2}, 1\right]^2$ (en rouge), le nœud **(4, 2)** au cube $\left[\frac{1}{4}, \frac{1}{2}\right]^2$ (en orange) et le nœud **(1, 3, 4)** au cube $\left[\frac{1}{4}, \frac{3}{8}\right] \times \left[\frac{1}{2}, \frac{5}{8}\right]$ (en gris). Dans cet exemple, l'arbre est de hauteur 3.

4.3.1.2 Arbre étiqueté

En référence aux ensembles $\mathcal{I}$, $\mathcal{O}$ et $\mathcal{A}$ définis en (4.3), on attribue à chaque nœud de l'arbre une étiquette choisie dans $\{\mathcal{I}, \mathcal{O}, \mathcal{A}\}$. En notant $\mathbf{t}$ un arbre étiqueté construit par l'algorithme, un nœud $u \in \mathbf{t}$ étiqueté $\mathcal{I}$ (resp. $\mathcal{O}$) réfère alors à un cube $Q(u)$, qui vérifie $Q(u) \subseteq \mathbb{F}$ (resp. $Q(u) \cap \mathbb{F} = \emptyset$). L'ensemble des feuilles (c'est-à-dire les nœuds terminaux) est noté $\mathcal{L}(\mathbf{t})$, avec $\mathcal{L}$ pour « leaves ». On l'écrit comme la réunion de trois sous-ensembles :

$$\mathcal{L}(\mathbf{t}) = \mathcal{I}(\mathbf{t}) \cup \mathcal{O}(\mathbf{t}) \cup \mathcal{A}(\mathbf{t}),$$

où $\mathcal{I}(\mathbf{t})$ est l'ensemble des feuilles étiquetées $\mathcal{I}$, $\mathcal{O}(\mathbf{t})$ celles étiquetées $\mathcal{O}$, et $\mathcal{A}(\mathbf{t})$ celles étiquetées $\mathcal{A}$.

4.3.1.3 Notations

Pour tout $u \in \mathbf{t} \setminus \{\emptyset\}$, il existe une unique branche de $\mathbf{t}$ qui relie la racine à u (voir, par exemple, la branche rouge représentée figure 4.6). L'ensemble des nœuds qui appartiennent

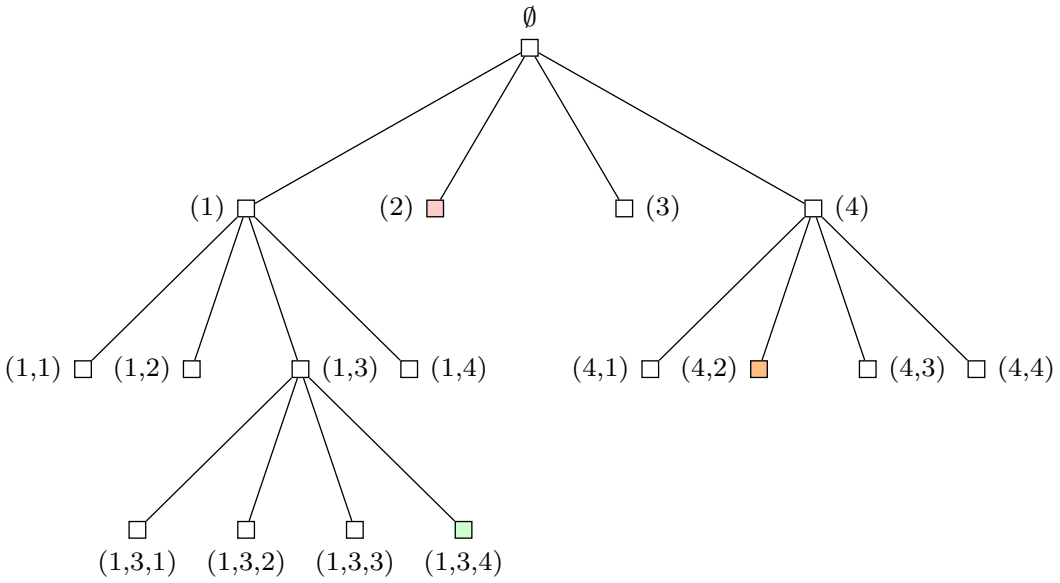

FIGURE 4.4 – Arbre de hauteur 3 correspondant à la partition dyadique de $[0,1]^2$ représentée figure 4.5.

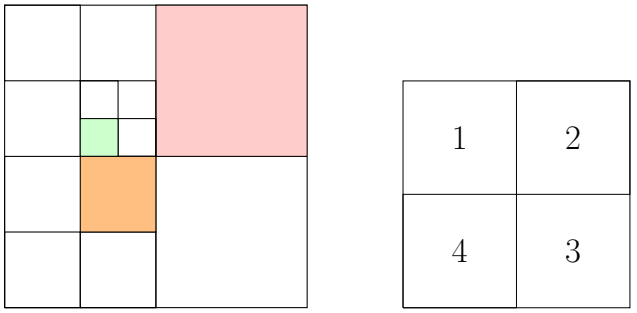

FIGURE 4.5 – À gauche : partitionnement dyadique de $[0,1]^2$ correspondant à l'arbre de la figure 4.4. À droite : ordre de numérotation imposé.

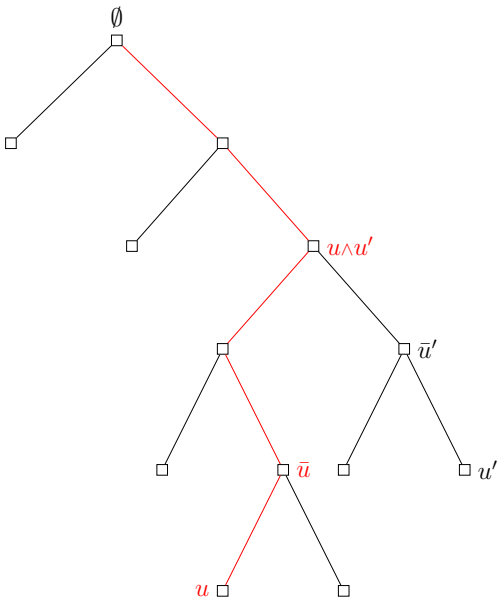

FIGURE 4.6 – Exemple d'arbre 2−régulier. Le nœud $\bar{u}$ est le parent de u. Tous les nœuds sur le chemin rouge sont des ancêtres de u. Les nœuds u et u' sont des feuilles de l'arbre. Leur ancêtre commun le plus récent est le nœud $u \wedge u'$.

à cette branche – et qui ne sont pas u – sont appelés les ancêtres de u. Si $v \in \mathbf{t}$ est un ancêtre de u, alors on note $v \prec u$. L'ancêtre le plus récent de u est son parent et on le note $\bar{u}$. Autrement dit, $\forall u \in \mathbf{t} \setminus \{\emptyset\}$, on a :

$$\emptyset \prec \ldots \prec \bar{u} \prec u,$$

ou, de façon équivalente,

$$\mathbb{X} \supset \ldots \supset Q(\bar{u}) \supset Q(u).$$

On note $\mathbf{t}_u$ la branche contenant u et tous ses ancêtres, sauf la racine :

$$\mathbf{t}_u = \{u\} \cup \{v \in \mathbf{t} \setminus \{\emptyset\} : v \prec u\}.$$

Enfin, pour $u, u' \in \mathbf{t}$, avec $u \neq u'$, on note $u \wedge u'$ leur ancêtre commun le plus récent. C'est le nœud de plus grande profondeur dans $\mathbf{t}_u \cap \mathbf{t}_{u'}$ (voir, par exemple, figure 4.6).

4.3.2 Principe général

L'algorithme est entièrement décrit en section 4.3.3. Avant cela, on présente les principales idées sur lesquelles il repose. On verra qu'il procède en parallèle à une identification du domaine de défaillance $\mathbb{F}$ et à une approximation de la probabilité de défaillance p.

4.3.2.1 Identification du domaine de défaillance

Le principe de notre algorithme est d'identifier le domaine de défaillance $\mathbb{F}$ à travers une répartition de cubes dyadiques de $\mathbb{X}$ dans les sous-ensembles $\mathcal{I}$, $\mathcal{O}$ et $\mathcal{A}$ définis en (4.3). Il procède pour cela par découpe dyadique récursive de $\mathbb{X}$. À chaque itération, l'ensemble $\mathbb{X}$ est écrit comme une partition de plus en plus fine de cubes dyadiques deux à deux disjoints.

Pour déterminer à quel sous-ensemble parmi $\mathcal{I}$, $\mathcal{O}$ et $\mathcal{A}$ appartient un cube, on procède comme expliqué en section 4.2.1 : on évalue g en son centre et on utilise l'hypothèse $(\mathcal{H}_1)$ pour la classification.

Lorsqu'un cube est dans $\mathcal{I}$ (resp. $\mathcal{O}$), il n'est plus nécessaire d'y évaluer g par la suite car on sait qu'il est inclus dans $\mathbb{F}$ (resp. disjoint de $\mathbb{F}$). Seuls les cubes incertains (i.e. les cubes de $\mathcal{A}$) requièrent de nouvelles évaluations de cette fonction. Autrement dit, les nœuds dans l'arbre qui sont étiquetés $\mathcal{I}$ ou $\mathcal{O}$ sont nécessairement des feuilles, c'est-à-dire qu'ils n'ont pas d'enfants. Seuls les nœuds étiquetés $\mathcal{A}$ ont des enfants puisqu'ils réfèrent à des cubes dont l'intersection avec $\mathbb{F}$ est incertaine.

4.3.2.2 Approximation de la probabilité de défaillance

La relation d'ordre $\mathcal{I} \subseteq \mathbb{F} \subseteq \mathcal{I} \cup \mathcal{A}$ implique l'encadrement de p suivant :

$$\mathbb{P}\big(\mathbf{X} \in \mathcal{I}\big) \leq p \leq \mathbb{P}\big(\mathbf{X} \in \mathcal{I} \cup \mathcal{A}\big).$$

Pour chaque arbre étiqueté $\mathbf{t}$ construit par l'algorithme, on dispose alors de deux approximations de la probabilité de défaillance. La première est une approximation par défaut. On la note $p^-(\mathbf{t})$ et elle est définie par :

$$p^-(\mathbf{t}) = \sum_{u \in \mathcal{I}(\mathbf{t})} \mathbb{P}\big(\mathbf{X} \in Q(u)\big). \tag{4.4}$$

C'est la probabilité que $\mathbf{X}$ appartienne à l'ensemble des cubes identifiés par l'algorithme comme inclus dans le domaine de défaillance $\mathbb{F}$. La seconde est une approximation par excès. On la note $p^+(\mathbf{t})$ et elle est définie par :

$$p^+(\mathbf{t}) = \sum_{u \in \mathcal{I}(\mathbf{t}) \cup \mathcal{A}(\mathbf{t})} \mathbb{P}\big(\mathbf{X} \in Q(u)\big) = p^-(\mathbf{t}) + \sum_{u \in \mathcal{A}(\mathbf{t})} \mathbb{P}\big(\mathbf{X} \in Q(u)\big). \tag{4.5}$$

C'est la probabilité que $\mathbf{X}$ appartienne à l'ensemble des cubes identifiés par l'algorithme comme inclus dans $\mathbb{F}$ ou comme incertains.

En pratique, une méthode Monte-Carlo spécifique est appliquée pour estimer les quantités $p^-(\mathbf{t})$ et $p^+(\mathbf{t})$. Cependant, pour ne pas alourdir les explications, celle-ci est détaillée à part – en section 4.4 – et on suppose dans ce qui suit que l'on sait exactement calculer ces bornes, c'est-à-dire qu'elles constituent l'encadrement de p que l'algorithme retourne.

4.3.3 Implémentation

Dans ce qui suit, la notation $|E|$ désigne le cardinal d'un ensemble fini $E \subseteq \mathbb{X}$.

On explique ici comment se déroulent les itérations successives de l'algorithme. À chaque étape, l'arbre $\mathbf{t}$ est augmenté d'une génération et l'encadrement de p est mis à jour à travers le calcul des quantités $p^-(\mathbf{t})$ et $p^+(\mathbf{t})$ définies en (4.4) et (4.5). On se donne un nombre maximal $n \in \mathbb{N}^*$ d'appels à la fonction g et l'algorithme s'arrête lorsque ce budget est atteint.

L'étape d'initialisation consiste à poser $\mathbf{t} = \emptyset$ et à affecter l'étiquette $\mathcal{A}$ à la racine. La hauteur de l'arbre initial est donc nulle et on a : $\mathcal{I}(\mathbf{t}) = \mathcal{O}(\mathbf{t}) = \emptyset$ et $\mathcal{A}(\mathbf{t}) = \{\emptyset\}$.

Au début de l'étape $k \in \mathbb{N}^*$, l'arbre $\mathbf{t}$ est de hauteur $k-1$, de sorte que les feuilles appartenant à $\mathcal{I}(\mathbf{t}) \cup \mathcal{O}(\mathbf{t})$ sont de profondeur au plus $k-2$ et celles appartenant à $\mathcal{A}(\mathbf{t})$ sont de profondeur $k-1$. Ces dernières constituent la dernière génération de $\mathbf{t}$ et sont assimilées à des cubes dyadiques de longueur $\frac{1}{2^{k-1}}$. On note n_{k-1} le nombre d'appels à la fonction g déjà effectués, avec $n_0 = 0$.

L'algorithme construit la génération k de $\mathbf{t}$ et retourne une approximation de p comme décrit ci-dessous :

— **Classification des feuilles de la dernière génération**

Soit $\tilde{\mathcal{A}}(\mathbf{t})$ un ensemble de feuilles que l'on initialise à $\emptyset$ au début de chaque étape.

Pour tout $u \in \mathcal{A}(\mathbf{t})$ de profondeur $k-1$,

— Si $n_{k-1} < n$, alors on évalue g au centre $\mathbf{c}_Q$ du cube dyadique $Q(u)$ et on actualise le nombre d'appels à la fonction : $n_{k-1} \leftarrow n_{k-1} + 1$.

On applique l'hypothèse $(\mathcal{H}_1)$:

— Si $g(\mathbf{c}_Q) > T + \frac{L}{2^k}$, alors u est etiqueté $\mathcal{I}$:

$$\mathcal{I}(\mathbf{t}) \leftarrow \mathcal{I}(\mathbf{t}) \cup \{u\}.$$

— Si $g(\mathbf{c}_Q) \leq T - \frac{L}{2^k}$, alors u est étiqueté $\mathcal{O}$:

$$\mathcal{O}(\mathbf{t}) \leftarrow \mathcal{O}(\mathbf{t}) \cup \{u\}.$$

— Si $|g(\mathbf{c}_Q) - T| \leq \frac{L}{2^k}$, alors u est étiqueté $\mathcal{A}$:

$$\tilde{\mathcal{A}}(\mathbf{t}) \leftarrow \tilde{\mathcal{A}}(\mathbf{t}) \cup \{u\}.$$

— Si $n_{k-1} = n$, alors aucune évaluation de g n'est effectuée au centre de $Q(u)$ et u est, par défaut, étiqueté $\mathcal{A}$:

$$\tilde{\mathcal{A}}(\mathbf{t}) \leftarrow \tilde{\mathcal{A}}(\mathbf{t}) \cup \{u\}.$$

— **Mise à jour de l'arbre**

On construit la génération k de $\mathbf{t}$: ce sont des feuilles de profondeur k qui sont les enfants de $\tilde{\mathcal{A}}(\mathbf{t})$ et qui sont étiquetées $\mathcal{A}$. Le nouvel ensemble $\mathcal{A}(\mathbf{t})$ vérifie alors :

$$|\mathcal{A}(\mathbf{t})| = 2^d |\tilde{\mathcal{A}}(\mathbf{t})|.$$

— **Mise à jour des probabilités**

On met à jour les approximations de la probabilité de défaillance p en calculant :

$$p^-(\mathbf{t}) = \sum_{u \in \mathcal{I}(\mathbf{t})} \mathbb{P}\big(\mathbf{X} \in Q(u)\big) \quad \text{et} \quad p^+(\mathbf{t}) = p^-(\mathbf{t}) + \sum_{u \in \mathcal{A}(\mathbf{t})} \mathbb{P}\big(\mathbf{X} \in Q(u)\big).$$

— On actualise le nombre d'appels à la fonction : $n_k = n_{k-1}$. L'algorithme s'arrête si $n_k = n$.

Remarque 4.3.1. *En pratique, d'autres critères d'arrêt que le nombre d'appels à la fonction g peuvent être pris en compte. Par exemple, quand bien même le budget d'évaluations de g ne serait pas épuisé, l'algorithme doit s'arrêter à la fin d'une étape si, pour une valeur $\epsilon > 0$ fixée, les approximations $p^-(\mathbf{t})$ et $p^+(\mathbf{t})$ vérifient : $p^+(\mathbf{t}) - p^-(\mathbf{t}) < \epsilon$.*

4.3.4 Exemple

Afin d'illustrer le principe de fonctionnement de l'algorithme, on propose ici un exemple simple en dimension $d = 1$ pour lequel la valeur de la probabilité de défaillance p est connue. En haut à gauche de la figure 4.7, on a représenté la fonction g (courbe noire), le seuil T (droite en pointillés rouges), la densité $f_{\mathbf{X}}$ de la loi $\mathbf{P}_{\mathbf{X}}$ (courbe bleue) et la probabilité p (aire bleue) que l'on cherche à estimer. On a précisement ici :

$$g(\mathbf{x}) = (0.8\mathbf{x} - 0.3) + \exp\big(-11.534\mathbf{x}^{1.95}\big) + \exp\big(-2(\mathbf{x} - 0.9)^2\big).$$

On rappelle que si $g : I \subset \mathbb{R} \to \mathbb{R}$ une fonction dérivable en tout point et $\sup_{x \in I} |g'(x)| \leq L$, alors g est L-lipschitzienne. Ici, $L = \sup_{\mathbf{x} \in [0,1]} |g'(\mathbf{x})| = 1.61$. En prenant $\mathbf{X} \sim \mathcal{N}\big(\frac{1}{5}, \frac{1}{5}^2\big)$ et $T = 1.3$, la probabilité p est égale à 2.08×10^{-3}.

À l'étape $k = 1$, une première évaluation de g est effectuée au point $\frac{1}{2}$. On en déduit la plus petite (resp. grande) valeur que peut prendre g dans l'intervalle $[0, 1]$. D'après $(\mathcal{H}_1)$, elle vaut $g\left(\frac{1}{2}\right) - \frac{L}{2}$ (resp. $g\left(\frac{1}{2}\right) + \frac{L}{2}$). On compare ces quantités à la valeur du seuil T et on a bien entendu $\left|g\left(\frac{1}{2}\right) - T\right| \leq \frac{L}{2}$. Comme représenté en haut à droite de la figure 4.7, cela signifie que la ligne de niveau T traverse les courbes représentatives des fonctions $\mathbf{x} \mapsto g\left(\frac{1}{2}\right) + L|\mathbf{x} - \frac{1}{2}|$ et $\mathbf{x} \mapsto g\left(\frac{1}{2}\right) - L|\mathbf{x} - \frac{1}{2}|$. Dans l'arbre étiqueté $\mathbf{t}$ en bas à droite, on conclut :

$$\mathcal{I}(\mathbf{t}) = \emptyset, \quad \mathcal{O}(\mathbf{t}) = \emptyset \quad \text{et} \quad \mathcal{A}(\mathbf{t}) = \{(1), (2)\}.$$

À ce stade, les approximations $p^-(\mathbf{t})$ et $p^+(\mathbf{t})$ vérifient donc :

$$p^-(\mathbf{t}) = 0 \quad \text{et} \quad p^+(\mathbf{t}) = p^-(\mathbf{t}) + \mathbb{P}\big(\mathbf{X} \in \big[0, \tfrac{1}{2}\big] \cup \big[\tfrac{1}{2}, 1\big]\big) = 1.$$

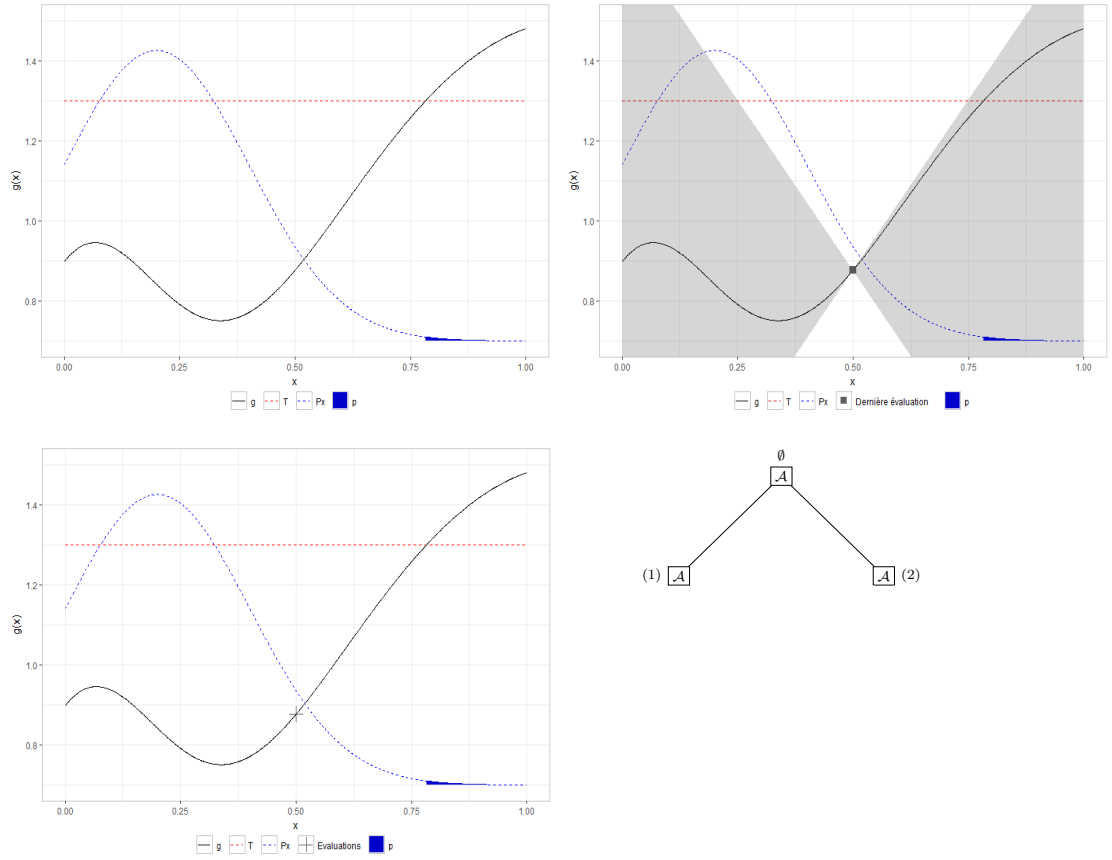

FIGURE 4.7 – La probabilité p que l'on cherche à estimer correspond à l'aire bleue sous la courbe de la densité de la loi $\mathbf{P}_{\mathbf{X}}$. Étape 1 : l'algorithme évalue g au point $\frac{1}{2}$. On a $\mathcal{I}(\mathbf{t}) = \emptyset$, $\mathcal{O}(\mathbf{t}) = \emptyset$ et $\mathcal{A}(\mathbf{t}) = \{(1), (2)\}$. Par conséquent, on a $p^-(\mathbf{t}) = 0$ et $p^+(\mathbf{t}) = p^-(\mathbf{t}) + \mathbb{P}\big(\mathbf{X} \in \big[0, \frac{1}{2}\big] \cup \big[\frac{1}{2}, 1\big]\big) = 1$.

Sur la figure 4.8, on montre comment procède l'algorithme à l'étape $k = 2$: une première évaluation est effectuée au point $\frac{1}{4}$ et on observe que : $g\left(\frac{1}{4}\right) + \frac{M}{4} \leq T$, c'est-à-dire que $g(\mathbf{x}) \leq T$, pour tout $\mathbf{x} \in \big[0, \frac{1}{2}\big]$. Plus aucune évaluation de g ne sera effectuée dans le sous-domaine $\big[0, \frac{1}{2}\big]$ et on affecte l'étiquette $\mathcal{O}$ au nœud (1). Après une évaluation supplémentaire de g au point $\frac{3}{4}$, on a :

$$\mathcal{I}(\mathbf{t}) = \emptyset, \quad \mathcal{O}(\mathbf{t}) = \{(1)\} \quad \text{et} \quad \mathcal{A}(\mathbf{t}) = \{(2,1), (2,2)\}.$$

On met à jour les approximations de p :

$$p^-(\mathbf{t}) = 0, \quad \text{et} \quad p^+(\mathbf{t}) = p^-(\mathbf{t}) + \mathbb{P}\big(\mathbf{X} \in \big[\tfrac{1}{2}, \tfrac{3}{4}\big] \cup \big[\tfrac{3}{4}, 1\big]\big).$$

Sur les figures 4.9 à 4.11, on a représenté de la même façon les étapes 3 à 5. À partir de l'étape 4, on remarque que l'algorithme commence à identifier des sous-domaines de $\mathbb{X}$ inclus dans le domaine de défaillance et dans lesquels plus aucune évaluation de g n'est effectuée par la suite. Sur la figure 4.12, on donne à gauche le résultat de l'algorithme après 35 évaluations de g. On constate que celles-ci sont concentrées autour de l'unique point $\mathbf{x}_T$

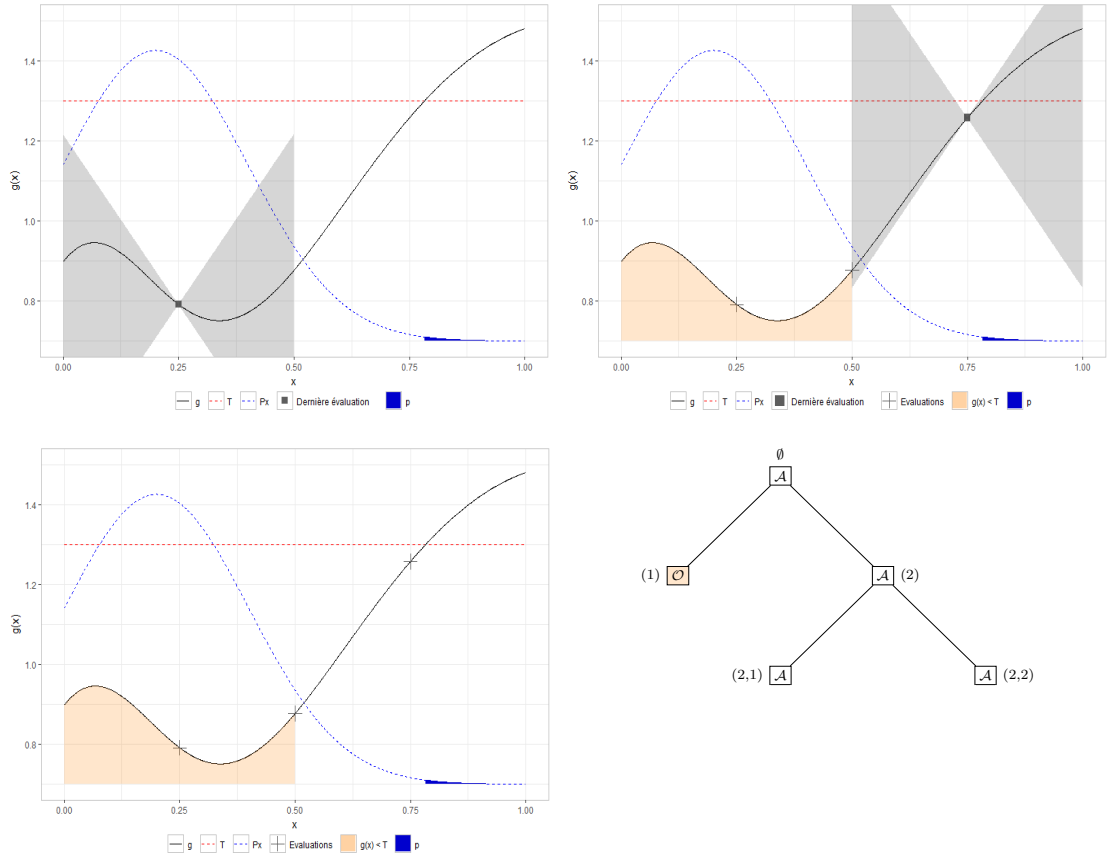

FIGURE 4.8 – Étape 2 : l'algorithme évalue g aux points $\frac{1}{4}$ et $\frac{3}{4}$. On a $\mathcal{I}(\mathbf{t}) = \emptyset$, $\mathcal{O}(\mathbf{t}) = \{(1)\}$ et $\mathcal{A}(\mathbf{t}) = \{(2,1),(2,2)\}$. Plus aucune évaluation de g ne sera effectuée dans l'intervalle $\left[0, \frac{1}{2}\right]$. On a par conséquent $p^-(\mathbf{t}) = 0$ et $p^+(\mathbf{t}) = p^-(\mathbf{t}) + \mathbb{P}\left(\mathbf{X} \in \left[\frac{1}{2}, \frac{3}{4}\right] \cup \left[\frac{3}{4}, 1\right]\right)$.

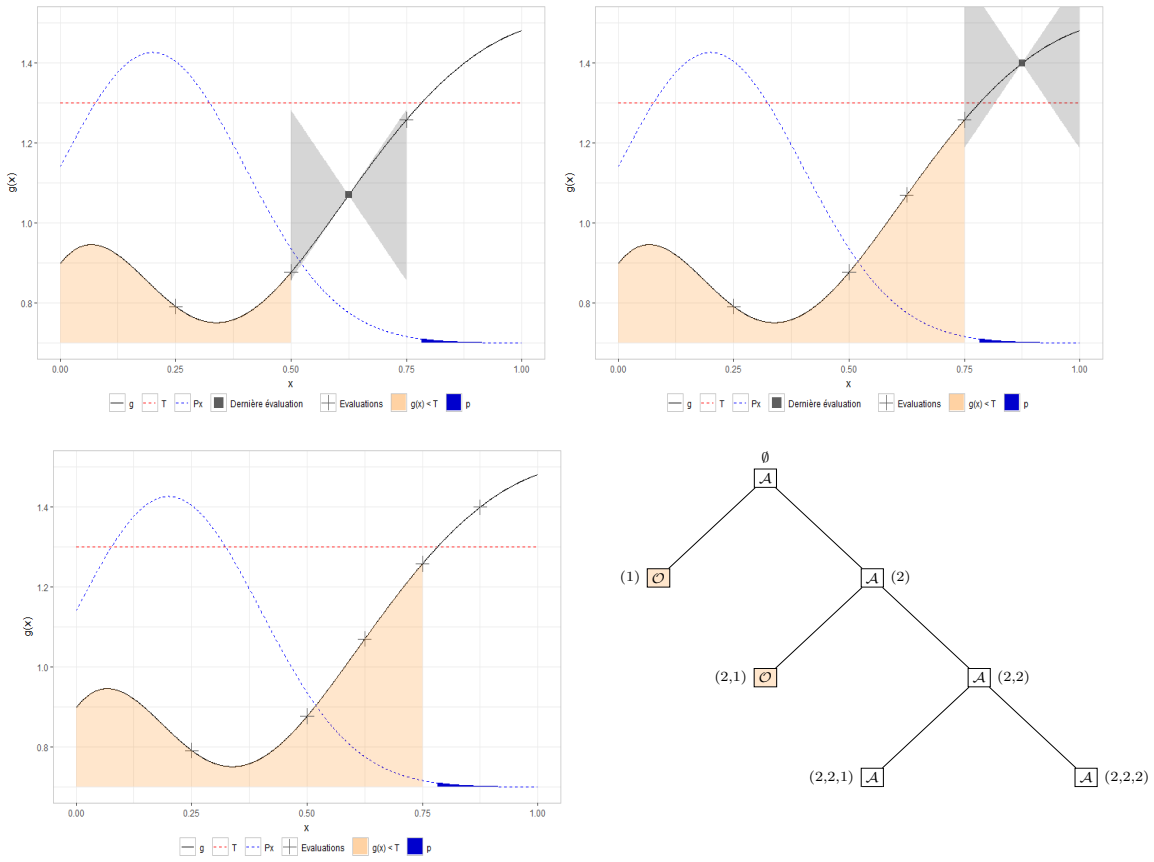

FIGURE 4.9 – Étape 3 : l'algorithme évalue g aux points $\frac{5}{8}$ et $\frac{7}{8}$. On a $\mathcal{I}(\mathbf{t}) = \emptyset$, $\mathcal{O}(\mathbf{t}) = \{(1), (2,1)\}$ et $\mathcal{A}(\mathbf{t}) = \{(2,2,1), (2,2,2)\}$. Par conséquent, $p^-(\mathbf{t}) = 0$ et $p^+(\mathbf{t}) = p^-(\mathbf{t}) + \mathbb{P}\big(\mathbf{X} \in \big[\frac{3}{4}, \frac{7}{8}\big] \cup \big[\frac{7}{8}, 1\big]\big)$.

tel que $g(\mathbf{x}_T) = T$, et que les intervalles en lesquels elles sont effectuées sont de longueur très petite. À droite, on a représenté les approximations successives de p, avec $p^+(\mathbf{t})$ en gris et $p^-(\mathbf{t})$ en vert.

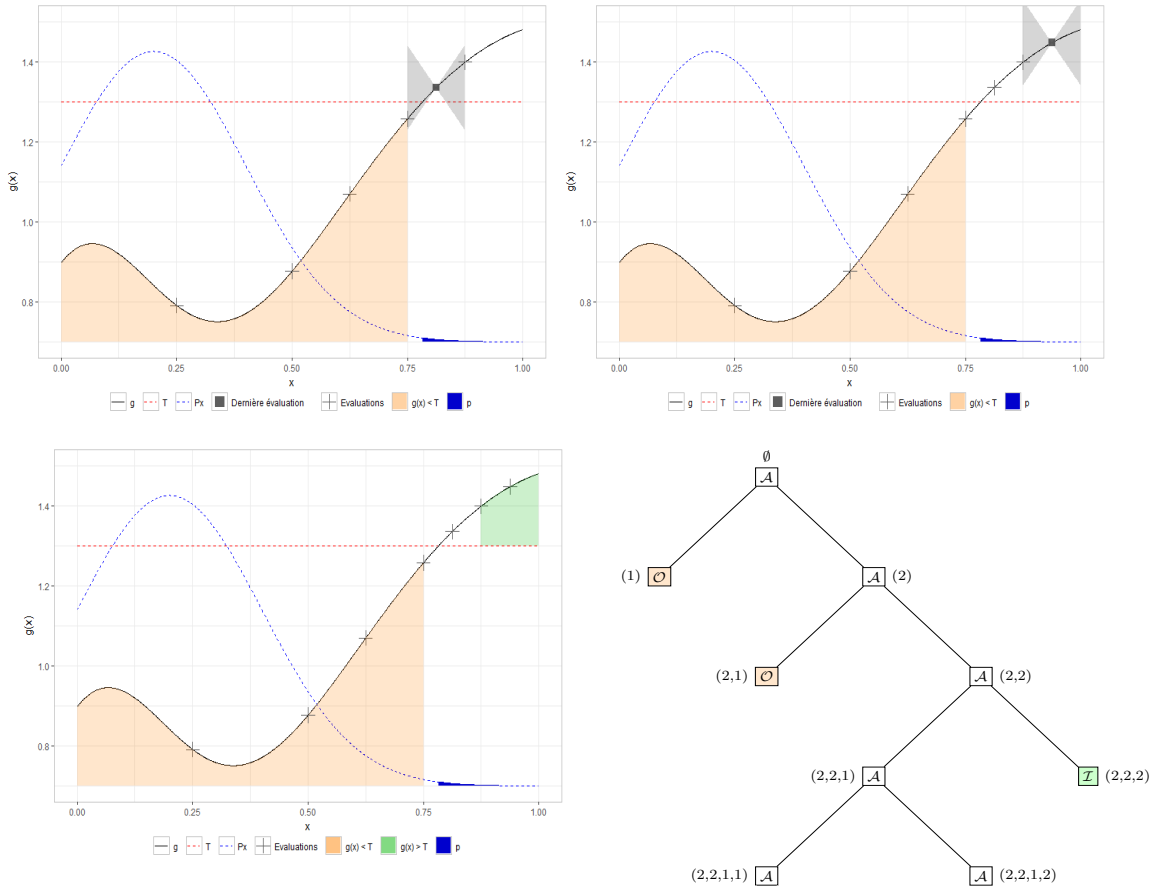

FIGURE 4.10 – Étape 4 : l'algorithme évalue g aux points $\frac{13}{16}$ et $\frac{15}{16}$. On a $\mathcal{I}(\mathbf{t}) = \{(2,2,2)\}$, $\mathcal{O}(\mathbf{t}) = \{(1),(2,1)\}$ et $\mathcal{A}(\mathbf{t}) = \{(2,2,1,1),(2,2,1,2)\}$. Plus aucune évaluation de g ne sera effectuée dans l'intervalle $\left[0,\frac{3}{4}\right] \cup \left[\frac{7}{8},1\right]$ (orange et vert). On a alors $p^-(\mathbf{t}) = \mathbb{P}\big(\mathbf{X} \in \left[\frac{7}{8},1\right]\big)$ et $p^+(\mathbf{t}) = p^-(\mathbf{t}) + \mathbb{P}\big(\mathbf{X} \in \left[\frac{3}{4},\frac{13}{16}\right] \cup \left[\frac{13}{16},\frac{7}{8}\right]\big)$.

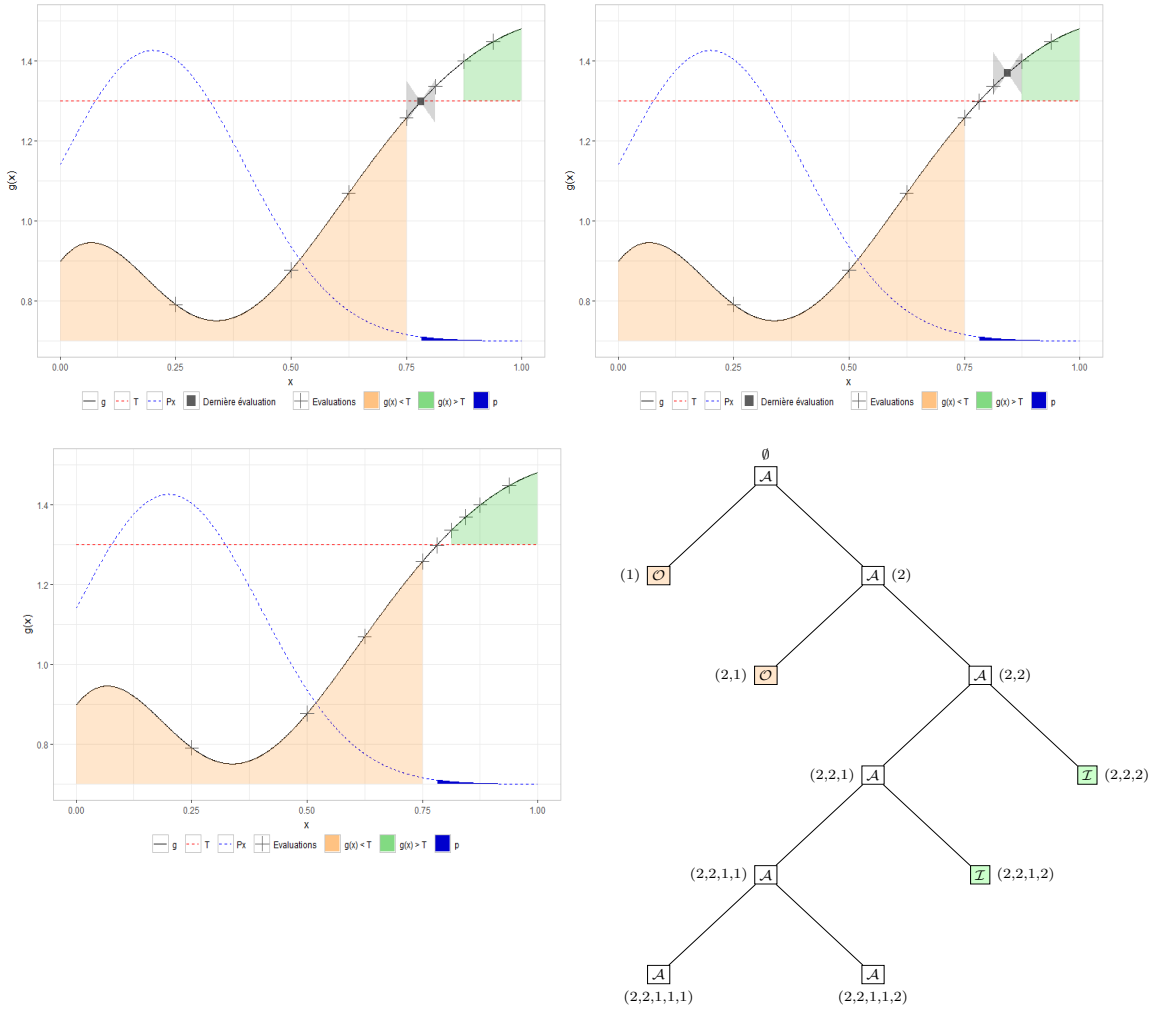

FIGURE 4.11 – Étape 5 : l'algorithme évalue g aux points $\frac{25}{32}$ et $\frac{27}{32}$. On a $\mathcal{I}(\mathbf{t}) = \{(2,2,2),(2,2,1,2)\}$, $\mathcal{O}(\mathbf{t}) = \{(1),(2,1)\}$ et $\mathcal{A}(\mathbf{t}) = \{(2,2,1,1,1),(2,2,1,1,2)\}$. Plus aucune évaluation de g ne sera effectuée dans l'intervalle $\left[0,\frac{3}{4}\right] \cup \left[\frac{13}{16},1\right]$ (orange et vert). On a $p^-(\mathbf{t}) = \mathbb{P}\left(\mathbf{X} \in \left[\frac{13}{16},\frac{7}{8}\right] \cup \left[\frac{7}{8},1\right]\right)$ et $p^+(\mathbf{t}) = p^-(\mathbf{t}) + \mathbb{P}\left(\mathbf{X} \in \left[\frac{3}{4},\frac{25}{32}\right] \cup \left[\frac{25}{32},\frac{13}{16}\right]\right)$.

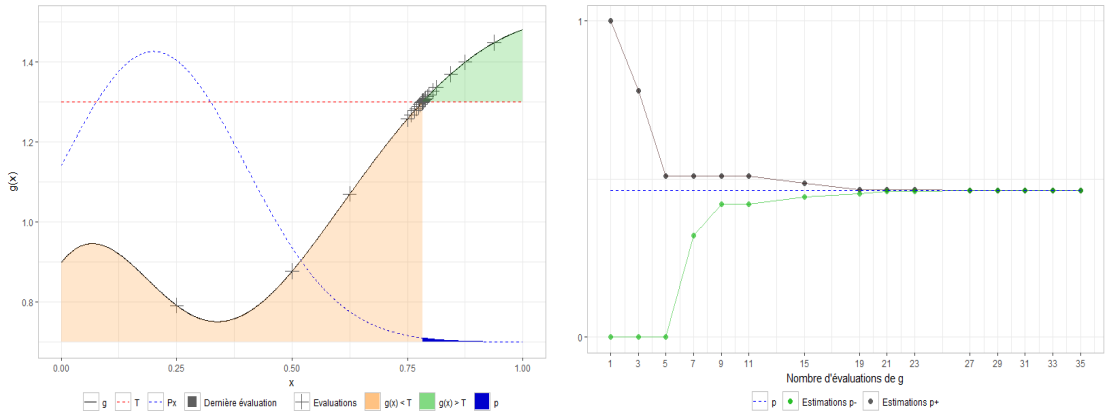

FIGURE 4.12 – À gauche : résultat de l'algorithme une fois le budget de $n = 35$ évaluations atteint. À droite : approximations successives en fonction du nombre d'appels à la fonction g. En bleu : la vraie probabilité de défaillance p. En gris : l'approximation par excès $p^+(\mathbf{t})$. En vert : l'approximation par défaut $p^-(\mathbf{t})$.

Remarque 4.3.2. *À chaque itération, l'algorithme effectue un certain nombre d'appels à la fonction g aux centres de cubes dyadiques. On peut envisager deux façons d'ordonner les évaluations en fonction des cubes. Soit on respecte l'ordre de numérotation imposé par l'arbre (c'est le cas dans l'exemple ci-dessus), soit on ordonne les cubes de façon à explorer en priorité ceux de plus forte probabilité. Ainsi, si le budget est atteint au cours d'une itération, on aura tranché sur les cubes de forte probabilité et cela peut impacter considérablement la valeur estimée de p.*

4.3.5 Propriétés et analyse de l'algorithme

Dans l'exemple précédent (figure 4.12 à droite), on constate que les approximations convergent vers la vraie valeur de p au fur et à mesure que l'algorithme progresse. Dans ce qui suit, on étudie la précision de l'algorithme en fonction du nombre d'appels à la fonction g.

Pour tout $k \in \mathbb{N}^*$, on note n_k le nombre d'évaluations de g à la fin de l'étape k et $\mathbf{t}$ l'arbre étiqueté de hauteur k construit. On commence par montrer dans la propriété ci-dessous que n_k croît linéairement avec k, et non exponentiellement comme pourrait le suggérer le partitionnement de $\mathbb{X}$ en cubes dyadiques. Ceci est rendu possible grâce aux hypothèses de la section 4.2.

Propriété 4.3.1. *Pour tout $k \in \mathbb{N}^*$, le nombre total n_k d'évaluations de la fonction g à la fin de l'étape k vérifie :*

$$n_k \leq 1 + 4^d C(k-1).$$

Démonstration. Voir section 4.6. □

Remarque 4.3.3. *Dans la preuve de la propriété 4.3.1, on montre qu'à chaque étape de l'algorithme, le cardinal $|\mathcal{A}(\mathbf{t})|$ de $\mathcal{A}(\mathbf{t})$ est nécessairement majoré par $4^d C$. Cela signifie qu'à la fin de chaque étape, le nombre de cubes dyadiques de $\mathbb{X}$ où il faut encore évaluer g est borné indépendamment de k.*

Propriété 4.3.2. *Pour tout $k \in \mathbb{N}^*$, on note $\mathbf{t}$ l'arbre de hauteur k fourni par l'algorithme. Les approximations $p^-(\mathbf{t})$ et $p^+(\mathbf{t})$ de p satisfont les encadrements suivants :*

$$p^-(\mathbf{t}) \leq p \leq p^-(\mathbf{t}) + \|f_{\mathbf{X}}\| 4^d C 2^{-dk},$$

et

$$p \leq p^+(\mathbf{t}) \leq p + \|f_{\mathbf{X}}\| 4^d C 2^{-dk}.$$

Démonstration. Voir section 4.6. □

Le calcul des quantités $p^-(\mathbf{t})$ et $p^+(\mathbf{t})$ à la fin de l'étape k a nécessité n_k appels à la fonction g. C'est pourquoi la précision de l'algorithme à la fin d'une étape k est mesurée à travers les erreurs d'approximation $e^-_{n_k}(\mathbf{t})$ et $e^+_{n_k}(\mathbf{t})$ définies par :

$$e^-_{n_k}(\mathbf{t}) = p - p^-(\mathbf{t}) \quad \text{et} \quad e^+_{n_k}(\mathbf{t}) = p^+(\mathbf{t}) - p.$$

Ces erreurs peuvent être majorées comme suit :

Propriété 4.3.3. *Pour tout $k \in \mathbb{N}^*$, les erreurs d'approximation $e^-_{n_k}(\mathbf{t})$ et $e^+_{n_k}(\mathbf{t})$ vérifient :*

$$0 \leq \max\left(e^-_{n_k}(\mathbf{t}), e^+_{n_k}(\mathbf{t})\right) \leq \|f_{\mathbf{X}}\| 2^d C 2^{\frac{-d(n_k-1)}{4^d C}}. \tag{4.6}$$

Démonstration. Voir section 4.6. □

Enfin, supposons que le budget d'appels à la fonction g est $n \in \mathbb{N}^*$ et qu'il est atteint au cours d'une étape $k \in \mathbb{N}^*$. En notant $\mathbf{t}$ l'arbre de hauteur k fourni par l'algorithme, l'erreur d'approximation finale est :

$$e_n(\mathbf{t}) = \max\left(e^-_n(\mathbf{t}), e^+_n(\mathbf{t})\right),$$

où $e^-_n(\mathbf{t}) = p - p^-(\mathbf{t})$ et $e^+_n(\mathbf{t}) = p^+(\mathbf{t}) - p$. D'après la propriété 4.3.1, on a :

$$n \leq 1 + 4^d C(k-1).$$

Cela implique que $k \geq \left\lceil \frac{n-1}{4^d C} \right\rceil + 1$, ou encore que :

$$2^{-dk} \leq 2^{-d} 2^{-d\left\lceil \frac{n-1}{4^d C} \right\rceil} \leq 2^{-d} 2^{-\frac{d(n-1)}{4^d C}}.$$

Au vu des résultats donnés à la propriété 4.3.2, on en déduit que l'erreur d'approximation finale $e_n(\mathbf{t})$ peut être majorée comme suit :

Propriété 4.3.4. *Pour tout $n \in \mathbb{N}^*$, l'erreur d'approximation $e_n(\mathbf{t}) = \max\left(e^-_n(\mathbf{t}), e^+_n(\mathbf{t})\right)$ vérifie :*

$$0 \leq e_n(\mathbf{t}) \leq \|f_{\mathbf{X}}\| 2^d C 2^{\frac{-d(n-1)}{4^d C}}.$$

Ce résultat signifie que l'erreur d'approximation de notre algorithme converge vers 0 à vitesse géométrique et qu'il est d'autant plus performant que C est proche de 1. Cette dernière remarque est cohérente avec les exemples de fonctions vérifiant $(\mathcal{H}_2)$ donnés en figure 4.3.

4.4 Estimation des probabilités par méthode de splitting

Soit $\mathbf{t}\setminus\{\emptyset\}$ l'arbre étiqueté 2^d-régulier construit par l'algorithme. Pour tout $u \in \mathbf{t}$, on pose :

$$p(u) = \mathbb{P}\big(\mathbf{X} \in Q(u)\big),$$

et pour tous $u, u' \in \mathbf{t}$ tels que u' est un ancêtre de u ($u' \prec u$ et $Q(u) \subset Q(u')$), on pose :

$$p(u', u) = \mathbb{P}\big(\mathbf{X} \in Q(u) \mid \mathbf{X} \in Q(u')\big).$$

Pour simplifier, on regroupe les approximations $p^-(\mathbf{t})$ et $p^+(\mathbf{t})$ définies en (4.4) et (4.5) sous une même notation :

$$p(\mathbf{t}) = \sum_{u \in \mathcal{U}(\mathbf{t})} p(u) = \begin{cases} p^-(\mathbf{t}) & \text{si} \quad \mathcal{U}(\mathbf{t}) = \mathcal{I}(\mathbf{t}), \\ p^+(\mathbf{t}) & \text{si} \quad \mathcal{U}(\mathbf{t}) = \mathcal{I}(\mathbf{t}) \cup \mathcal{A}(\mathbf{t}). \end{cases} \tag{4.7}$$

Jusqu'à présent, on a supposé que l'on sait exactement calculer la probabilité $p(\mathbf{t})$. En pratique, il n'y a aucune raison pour que cette hypothèse soit vérifiée, d'autant que la densité $f_{\mathbf{X}}$ est supposée connue seulement à une constante de normalisation près. C'est pourquoi il est nécessaire d'ajouter à l'algorithme une étape d'estimation de probabilités.

Pour toute feuille $u \in \mathcal{U}(\mathbf{t})$, on note $\widehat{p}(u)$ un estimateur de la probabilité $p(u)$. On note également $\widehat{p}(\mathbf{t})$ un estimateur de la probabilité $p(\mathbf{t})$. D'après (4.7), ce dernier est défini par :

$$\widehat{p}(\mathbf{t}) = \sum_{u \in \mathcal{U}(\mathbf{t})} \widehat{p}(u).$$

L'estimateur $\widehat{p}(\mathbf{t})$ doit être adapté au fait que, au fur et à mesure que l'algorithme progresse, la taille des cubes dyadiques de $\mathbb{X}$ se réduit et les probabilités $(p(u))_{u \in \mathcal{U}(\mathbf{t})}$ deviennent très petites. Typiquement, la méthode Monte-Carlo naïve est exclue, car elle risque de renvoyer une estimation systématiquement égale à 0 dès que $p(u)$ est la probabilité que $\mathbf{X}$ appartienne à un cube $Q(u)$ de petite taille dans lequel la densité $f_{\mathbf{X}}$ est faible.

Pour contourner ce problème, on propose d'estimer les probabilités $(p(u))_{u \in \mathcal{U}(\mathbf{t})}$ par la méthode de splitting, qui a déjà été présentée en section 2.2.3. On rappelle que cette méthode consiste à écrire chaque probabilité en un produit de probabilités conditionnelles suffisamment grandes pour être estimées par une méthode Monte-Carlo naïve. Compte tenu de la structure arborescente de notre algorithme, cette décomposition est immédiate. On décrit cela ci-après.

4.4.1 Estimation de la probabilité d'une feuille

Soit $u \in \mathcal{U}(\mathbf{t})$ une feuille de $\mathbf{t}$. On va spécifier l'estimateur $\widehat{p}(u)$ de la probabilité $p(u)$ en se basant sur le fait qu'elle peut s'écrire comme un produit de probabilités conditionnelles.

Rappelons que $\mathbf{t}_u$ est la branche de $\mathbf{t}$ contenant la feuille u et tous ses ancêtres, sauf la racine. Pour tout $v \in \mathbf{t}_u$ de parent $\bar{v}$, puisque l'on a $Q(v) \subset Q(\bar{v})$, la formule de Bayes implique que :

$$p(v) = p(\bar{v}, v)p(\bar{v}).$$

La décomposition de $p(u)$ en produit de probabilités conditionnelles est donc immédiate et s'écrit :

$$p(u) = p(\bar{u}, u)p(\bar{u}) = \prod_{v \in \mathbf{t}_u} p(\bar{v}, v).$$

Cela revient à voir $\mathbf{t}$ comme un arbre pondéré. La pondération de la branche allant de $\bar{v}$ vers v est la probabilité conditionnelle $p(\bar{v}, v)$ et la probabilité d'un chemin est le produit des probabilités des branches qui le composent.

Supposons que pour tout nœud $v \in \mathbf{t} \setminus \{\emptyset\}$, on dispose d'une suite de variables aléatoires $\mathbf{X}^v = (\mathbf{X}_i^v)_{1 \leq i \leq N}$ i.i.d. suivant la loi conditionnelle de $\mathbf{X}$ sachant $\mathbf{X} \in Q(v)$. Dans ces conditions, on peut définir un estimateur $\widehat{p}(u)$ de la probabilité $p(u)$ de la façon suivante :

Définition 4.4.1. *Pour toute feuille $u \in \mathcal{U}(\mathbf{t})$, un estimateur noté $\widehat{p}(u)$ de la probabilité $p(u) = \prod_{v \in \mathbf{t}_u} p(\bar{v}, v)$ est la variable aléatoire définie par :*

$$\widehat{p}(u) = \prod_{v \in \mathbf{t}_u} \widehat{p}(\bar{v}, v), \tag{4.8}$$

où les estimateurs $(\widehat{p}(\bar{v}, v))_{v \in \mathbf{t}_u}$ sont définis par :

$$\widehat{p}(\bar{v}, v) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\mathbf{X}_i^{\bar{v}} \in Q(v)}, \quad \forall v \in \mathbf{t}_u, \tag{4.9}$$

avec $\mathbf{X}^{\bar{v}} = (\mathbf{X}_i^{\bar{v}})_{1 \leq i \leq N}$ un échantillon i.i.d. suivant la loi de $\mathbf{X}$ sachant $\mathbf{X} \in Q(\bar{v})$.

La méthode d'estimation que l'on propose repose sur la possibilité de simuler des suites de variables aléatoires i.i.d. suivant la restriction de $\mathbf{P}_{\mathbf{X}}$ à des cubes dyadiques de $\mathbb{X}$ (i.e. suivant des lois conditionnelles). Pour garantir les bonnes propriétés de nos estimateurs, ces suites doivent également être indépendantes deux à deux. Par exemple, cela assure que les estimateurs $(\widehat{p}(\bar{v}, v))_{v \in \mathbf{t}_u}$ définis en équation (4.9) sont indépendants, ce qui implique que l'estimateur $\widehat{p}(u)$ donné en (4.8) a de bonnes propriétés (voir la remarque 4.4.1 ci-dessous). Pour parvenir, de façon approchée, à construire de telles suites, on applique l'algorithme de Metropolis-Hastings (voir (Tierney, 1994) pour une présentation dans notre contexte). Nous présentons cet algorithme et l'appliquons à notre problème en section 4.4.3.

Remarque 4.4.1. *Pour tout $u \in \mathbf{t} \setminus \{\emptyset\}$, l'hypothèse d'indépendance des estimateurs $(\widehat{p}(\bar{v}, v))_{v \in \mathbf{t}_u}$ garantit de bonnes propriétés de l'estimateur $\widehat{p}(u)$ donné en (4.8). En effet, remarquons que, pour tout $v \in \mathbf{t} \setminus \{\emptyset\}$, l'estimateur $\widehat{p}(\bar{v}, v)$ donné en (4.9) est sans biais, fortement consistant, et de variance :*

$$\mathrm{Var}\big[\widehat{p}(\bar{v}, v)\big] = \frac{p(\bar{v}, v)(1 - p(\bar{v}, v))}{N}.$$

D'après le théorème central limite, il satisfait :

$$\sqrt{N}\big(\widehat{p}(\bar{v}, v) - p(\bar{v}, v)\big) \xrightarrow[N \to \infty]{\mathcal{L}} \mathcal{N}\big(0,\ p(\bar{v}, v)(1 - p(\bar{v}, v))\big). \tag{4.10}$$

Par la proposition de 2.2.1, on peut alors vérifier que sous l'hypothèse d'indépendance des estimateurs $(\widehat{p}(\bar{v}, v))_{v \in \mathbf{t}_u}$, l'estimateur $\widehat{p}(u)$ est lui aussi sans biais, fortement consistant et asymptotiquement normal :

$$\sqrt{N}\big(\widehat{p}(u) - p(u)\big) \xrightarrow[N \to \infty]{\mathcal{L}} \mathcal{N}\left(0,\ p(u)^2 \sum_{v \in \mathbf{t}_u} \frac{1 - p(\bar{v}, v)}{p(\bar{v}, v)}\right). \tag{4.11}$$

Grâce à (4.11), on pourrait construire des intervalles de confiance asymptotiques pour $p(u)$. Il suffit pour cela de remplacer les probabilités qui interviennent dans la variance asymptotique par leurs estimations. On rappelle que la densité $f_{\mathbf{X}}$ est supposée strictement positive, ce qui assure que $p(\bar{v}, v) > 0$, quel que soit $v \in \mathbf{t}$.

4.4.2 Estimateur : définition et propriétés

On rappelle que l'on cherche à estimer la probabilité $p(\mathbf{t})$ donnée en (4.7). D'après ce qui précède, elle est définie par :

$$p(\mathbf{t}) = \sum_{u \in \mathcal{U}(\mathbf{t})} p(u) = \sum_{u \in \mathcal{U}(\mathbf{t})} \prod_{v \in \mathbf{t}_u} p(\bar{v}, v).$$

Notons $\mathcal{V}(\mathbf{t})$ l'ensemble des nœuds internes de $\mathbf{t}$ (c'est-à-dire les nœuds de $\mathbf{t}$ qui ne sont pas des feuilles). Pour tout nœud $v \in \mathcal{V}(\mathbf{t})$, on se donne une suite de variables aléatoires $\mathbf{X}^v = (\mathbf{X}_i^v)_{1 \le i \le N}$ i.i.d. suivant la loi de $\mathbf{X}$ sachant $\mathbf{X} \in Q(v)$. On suppose de plus que les suites $(\mathbf{X}^v)_{v \in \mathcal{V}(\mathbf{t})}$ sont indépendantes. Sous ces hypothèses, l'estimateur $\widehat{p}(\mathbf{t})$ de la probabilité $p(\mathbf{t})$ est défini comme ci-dessous.

Définition 4.4.2. *L'estimateur $\widehat{p}(\mathbf{t})$ de la probabilité $p(\mathbf{t}) = \sum_{u \in \mathcal{U}(\mathbf{t})} \prod_{v \in \mathbf{t}_u} p(\bar{v}, v)$ est la variable aléatoire définie par :*

$$\widehat{p}(\mathbf{t}) = \sum_{u \in \mathcal{U}(\mathbf{t})} \prod_{v \in \mathbf{t}_u} \widehat{p}(\bar{v}, v), \tag{4.12}$$

où, pour tout $u \in \mathcal{U}(\mathbf{t})$, les estimateurs $(\widehat{p}(\bar{v}, v))_{v \in \mathbf{t}_u}$ sont définis par :

$$\widehat{p}(\bar{v}, v) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\mathbf{X}_i^{\bar{v}} \in Q(v)}, \quad \forall v \in \mathbf{t}_u, \tag{4.13}$$

avec $\mathbf{X}^{\bar{v}} = (\mathbf{X}_i^{\bar{v}})_{1 \le i \le N}$ un échantillon i.i.d. suivant la loi de $\mathbf{X}$ sachant $\mathbf{X} \in Q(\bar{v})$.

Dans ce qui suit, on s'intéresse aux propriétés de l'estimateur (4.12). Auparavant, on donne un résultat qui fournit la consistance et la normalité asymptotique des estimateurs $\big(\widehat{p}(\bar{v},v)\big)_{v\in\mathbf{t}}$ (qui ne sont pas indépendants).

Lemme 4.4.1. *Pour tout $v\in\mathbf{t}$, l'estimateur $\widehat{p}(\bar{v},v)$ donné en (4.13) est sans biais. En outre, il vérifie :*

$$\widehat{p}(\bar{v},v)\xrightarrow[N\to\infty]{p.s.}p(\bar{v},v). \tag{4.14}$$

Notons $\widehat{\mathbf{q}}$ le vecteur $(\widehat{p}(\bar{v},v))_{v\in\mathbf{t}}$ et $\mathbf{q}$ le vecteur $(p(\bar{v},v))_{v\in\mathbf{t}}$. On a :

$$\sqrt{N}(\widehat{\mathbf{q}}-\mathbf{q})\xrightarrow[N\to\infty]{\mathcal{L}}\mathcal{N}(0,\Gamma),$$

où la matrice de covariance Γ est donnée par :

$$\Gamma(v,v')=\begin{cases}p(\bar{v},v)\big(1-p(\bar{v},v)\big) & si\ v=v',\\ -p(\bar{v},v)p(\bar{v},v') & si\ v\neq v'\ et\ \bar{v}=\bar{v}',\\ 0 & si\ v\neq v'\ et\ \bar{v}\neq\bar{v}'.\end{cases} \tag{4.15}$$

Démonstration. Voir section 4.6. □

On rappelle que $u\wedge u'$ désigne l'ancêtre commun le plus récent des nœuds $u,u'\in\mathbf{t}$. On rappelle également que la densité $f_{\mathbf{X}}$ est supposée strictement positive, ce qui assure que $p(\bar{v},v)>0$, quel que soit $v\in\mathbf{t}$. Le résultat suivant fournit la normalité asymptotique de l'estimateur $\widehat{p}(\mathbf{t})$ défini en (4.12).

Théorème 4.4.1. *Pour tout ensemble $\mathcal{U}(\mathbf{t})$ de feuilles de $\mathbf{t}$, l'estimateur $\widehat{p}(\mathbf{t})$ est sans biais. En outre, il est consistant :*

$$\widehat{p}(\mathbf{t})\xrightarrow[N\to\infty]{p.s.}p(\mathbf{t}),$$

et satisfait :

$$\sqrt{N}\big(\widehat{p}(\mathbf{t})-p(\mathbf{t})\big)\xrightarrow[N\to\infty]{\mathcal{L}}\mathcal{N}(0,\sigma(\mathbf{t})^2), \tag{4.16}$$

où

$$\sigma(\mathbf{t})^2=\sum_{u\in\mathcal{U}(\mathbf{t})}p(u)^2\sum_{v\in\mathbf{t}_u}\frac{1-p(\bar{v},v)}{p(\bar{v},v)}+\sum_{\substack{u,u'\in\mathcal{U}(\mathbf{t})\\ u\neq u'}}p(u)p(u')\left[\sum_{v\in\mathbf{t}_{u\wedge u'}}\frac{1-p(\bar{v},v)}{p(\bar{v},v)}-1\right].$$

Démonstration. Voir section 4.6. □

En pratique, la variance asymptotique $\sigma(\mathbf{t})^2$ est inconnue, mais peut être estimée par :

$$\widehat{\sigma}(\mathbf{t})^2=\sum_{u\in\mathcal{U}(\mathbf{t})}\widehat{p}(u)^2\sum_{v\in\mathbf{t}_u}\frac{1-\widehat{p}(\bar{v},v)}{\widehat{p}(\bar{v},v)}+\sum_{\substack{u,u'\in\mathcal{U}(\mathbf{t})\\ u\neq u'}}\widehat{p}(u)\widehat{p}(u')\left[\sum_{v\in\mathbf{t}_{u\wedge u'}}\frac{1-\widehat{p}(\bar{v},v)}{\widehat{p}(\bar{v},v)}-1\right].$$

Tous les estimateurs étant consistants, celui-ci en hérite :

$$\widehat{\sigma}(\mathbf{t})^2 \xrightarrow[N\to\infty]{p.s.} \sigma(\mathbf{t})^2.$$

Ainsi, en appliquant le lemme de Slutsky au résultat de convergence en loi (4.16), on a :

$$\sqrt{N}\,\frac{\widehat{p}(\mathbf{t}) - p(\mathbf{t})}{\widehat{\sigma}(\mathbf{t})} \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Soit $\alpha \in (0,1)$ fixé. Un intervalle de confiance de niveau asymptotique $1-\alpha$ pour $p(\mathbf{t})$ est alors donné par :

$$\left[\widehat{p}(\mathbf{t}) - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\widehat{\sigma}(\mathbf{t})}{\sqrt{N}};\ \widehat{p}(\mathbf{t}) + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\widehat{\sigma}(\mathbf{t})}{\sqrt{N}}\right], \tag{4.17}$$

où $\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$ est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale centrée réduite. En section 4.4.4, on donne un exemple pour illustrer la performance de ces résultats.

4.4.3 Simulation d'échantillons Monte-Carlo

Précédemment, on a supposé que pour tout $v \in \mathcal{V}(\mathbf{t})$, on dispose de suites $\mathbf{X}^v = (\mathbf{X}_i^v)_{1\le i\le N}$ i.i.d. suivant la loi de $\mathbf{X}$ sachant $\mathbf{X} \in Q(v)$. On a également supposé que les suites $(\mathbf{X}^v)_{v\in\mathcal{V}(\mathbf{t})}$ sont indépendantes. On montre ici comment y parvenir de façon approchée avec l'algorithme de Metropolis-Hastings. On rappelle que pour tout $\mathbf{x} \in \mathbb{X}$, on fait l'hypothèse que :

$$0 < f_{\mathbf{X}}(\mathbf{x}) < \|f_{\mathbf{X}}\| < \infty.$$

Soit $v \in \mathcal{V}(\mathbf{t})$. Notons μ_v la loi de $\mathbf{X}$ sachant $\mathbf{X} \in Q(v)$:

$$\mu_v(d\mathbf{x}) = \frac{1}{p(v)}\mathbb{1}_{\mathbf{x}\in Q(v)}\mathbf{P}_{\mathbf{X}}(d\mathbf{x}). \tag{4.18}$$

La densité associée à la loi μ_v est la fonction f_v définie par :

$$f_v(\mathbf{x}) = \frac{1}{p(v)}\mathbb{1}_{\mathbf{x}\in Q(v)}f_{\mathbf{X}}(\mathbf{x}),\ \forall \mathbf{x} \in \mathbb{X}. \tag{4.19}$$

La probabilité $p(v)$ étant inconnue et $f_{\mathbf{X}}$ pouvant n'être connue qu'à une constante de normalisation près, cette densité est connue à une constante de normalisation près. On est donc typiquement dans le cadre d'application de l'algorithme de Metropolis-Hastings.

4.4.3.1 Algorithme de Metropolis-Hastings

La première version de l'algorithme a été introduite par (Metropolis et al., 1953), puis généralisée par (Hastings, 1970). Le principe de cet algorithme est de construire une chaîne de Markov dont la loi stationnaire est la loi suivant laquelle on souhaite simuler (ici μ_v). Ici, on présente l'algorithme pour des chaînes de Markov à espace d'états continus. On suit pour cela (Tierney, 1994), (Roberts and Rosenthal, 2004) et (Jourdain, 2019). Pour une

présentation avec des chaînes de Markov à espace d'états discret, on pourra consulter, par exemple, le chapitre 2 de (Delmas and Jourdain, 2006).

Soit $h_v : \mathbb{X} \times \mathbb{X} \to \mathbb{R}_*^+$ la densité d'un noyau de transition suivant laquelle on sait simuler et que l'on suppose strictement positive pour tout $(\mathbf{x}, \mathbf{y}) \in \mathbb{X}^2$. On pose :

$$r_v(\mathbf{x}, \mathbf{y}) = \frac{f_v(\mathbf{y}) h_v(\mathbf{y}, \mathbf{x})}{f_v(\mathbf{x}) h_v(\mathbf{x}, \mathbf{y})}, \quad \text{et} \quad \alpha_v(\mathbf{x}, \mathbf{y}) = \min\big(1, r_v(\mathbf{x}, \mathbf{y})\big).$$

L'algorithme de Metropolis-Hastings construit une suite de variables aléatoires $(\mathbf{X}_m)_{m \in \mathbb{N}}$ à valeurs dans $Q(v)$ de la façon suivante. Soit $\mathbf{X}_m \in Q(v)$ la dernière valeur retenue.

1. On génère la proposition $\mathbf{Y}$ suivant $h_v(\mathbf{X}_m, \cdot)$.
2. On calcule $\alpha_v(\mathbf{X}_m, \mathbf{Y})$.
3. On tire indépendamment une variable aléatoire U suivant une loi uniforme sur $[0, 1]$ et on pose :

$$\mathbf{X}_{m+1} = \mathbf{Y} \mathbb{1}_{U \leq \alpha_v(\mathbf{X}_m, \mathbf{Y})} + \mathbf{X}_m \mathbb{1}_{U > \alpha_v(\mathbf{X}_m, \mathbf{Y})},$$

   c'est-à-dire que l'on accepte la proposition $\mathbf{Y}$ avec la probabilité $\alpha_v(\mathbf{X}_m, \mathbf{Y})$ et que l'on conserve $\mathbf{X}_m$ avec la probabilité $1 - \alpha_v(\mathbf{X}_m, \mathbf{Y})$.

Lors de la mise en œuvre pratique, il n'est pas utile d'exécuter les étapes 2 et 3 si la proposition $\mathbf{Y}$ n'est pas à valeurs dans $Q(v)$, car elle sera nécessairement rejetée. En effet, on a $f_v(\mathbf{Y}) = 0 \Rightarrow \alpha_v(\mathbf{X}_m, \mathbf{Y}) = 0$. Dans notre cadre d'application, on prendra $\mathbf{X}_0$ dans $Q(v)$ de sorte que pour tout $m \in \mathbb{N}^*$, $\mathbf{X}_m$ appartient aussi à $Q(v)$.

La suite $(\mathbf{X}_m)_{m \in \mathbb{N}}$ peut contenir des répétitions. Cependant, sous des hypothèses que nous allons détailler, si on déroule la procédure un nombre suffisant de fois, on peut donc considérer que le dernier état généré est quasiment indépendant de la position initiale $\mathbf{X}_0$ et distribué suivant la loi cible μ_v.

On verra ci-après deux exemples de noyaux h_v qui peuvent être choisis pour la mise en œuvre pratique. Par ailleurs, on peut remarquer que le rapport $r_v(\mathbf{x}, \mathbf{y})$ ne fait pas intervenir les constantes de normalisation des densités $f_v(\mathbf{x})$ et $f_v(\mathbf{y})$, car elles s'annulent entre elles. Précisons que, dans la version originale de Metropolis (Metropolis et al., 1953), le noyau h_v est supposé symétrique :

$$h_v(\mathbf{x}, \mathbf{y}) = h_v(\mathbf{y}, \mathbf{x}) \quad \text{et} \quad r_\alpha(\mathbf{x}, \mathbf{y}) = \frac{f_v(\mathbf{y})}{f_v(\mathbf{x})}.$$

4.4.3.2 Propriétés de l'algorithme de Metropolis-Hastings

La suite $(\mathbf{X}_m)_{m \in \mathbb{N}}$ décrite par l'algorithme de Metropolis-Hastings est une chaîne de Markov à valeurs dans $\mathbb{X}$ (et même dans $Q(v)$). Dans la proposition suivante, on explicite le noyau de transition de cette chaîne. Il s'agit de l'application $M_v : \mathbb{X} \times \mathcal{B}(\mathbb{X}) \to [0, 1]$ telle que :

(i) $\forall \mathbf{x} \in \mathbb{X}$, $B \in \mathcal{B}(\mathbb{X}) \mapsto M_v(\mathbf{x}, B)$ est une mesure de probabilité sur $\mathbb{X}$, avec en particulier $M_v(\mathbf{x}, \mathbb{X}) = 1$.

(ii) $\forall B \in \mathcal{B}(\mathbb{X})$, $\mathbf{x} \in \mathbb{X} \mapsto M_v(\mathbf{x}, B)$ est mesurable.

Proposition 4.4.1. *La suite $(\mathbf{X}_m)_{m\in\mathbb{N}}$ est une chaîne de Markov à valeurs dans $\mathbb{X}$ et son noyau de transition $M_v : \mathbb{X} \times \mathcal{B}(\mathbb{X}) \to [0,1]$ est défini par :*

$$M_v(\mathbf{x}, d\mathbf{y}) = \mathbb{1}_{\mathbf{x}\notin Q(v)}\delta_{\mathbf{x}}(d\mathbf{y}) + \mathbb{1}_{\mathbf{x}\in Q(v)}K_v(\mathbf{x}, d\mathbf{y}), \tag{4.20}$$

où $K_v(\mathbf{x}, d\mathbf{y})$ est le noyau de transition défini par :

$$K_v(\mathbf{x}, d\mathbf{y}) = \alpha_v(\mathbf{x}, \mathbf{y})h_v(\mathbf{x}, \mathbf{y})d\mathbf{y} + \delta_{\mathbf{x}}(d\mathbf{y})\left(1 - \int_{\mathbb{X}} \alpha_v(\mathbf{x}, \mathbf{z})h_v(\mathbf{x}, \mathbf{z})d\mathbf{z}\right).$$

Bien que ce résultat soit connu, on rappelle la démonstration en section 4.6 pour aider à la compréhension. On peut remarquer que le noyau $K_v(\mathbf{x}, \cdot)$ est la somme de deux termes. Celui de gauche signifie que, partant de la réalisation $\mathbf{x}$, la proposition $\mathbf{y}$ générée par la densité $h_v(\mathbf{x}, \cdot)$ est acceptée avec une probabilité $\alpha_v(\mathbf{x}, \mathbf{y})$. Celui de droite traduit qu'en cas de rejet, la nouvelle réalisation vaut toujours $\mathbf{x}$. Puisque $\int_{\mathbb{X}} \delta_{\mathbf{x}}(d\mathbf{y}) = 1$, on est certain d'avoir l'égalité $\int_{\mathbb{X}} K_v(\mathbf{x}, d\mathbf{y}) = 1$, $\forall \mathbf{x} \in \mathbb{X}$. Le terme de gauche dans (4.20) assure que le noyau M_v est bien défini sur tout l'ensemble $\mathbb{X}$ dans le sens où, pour tout $\mathbf{x} \in \mathbb{X}$, on a bien l'égalité $\int_{\mathbb{X}} M_v(\mathbf{x}, d\mathbf{y}) = 1$.

Notre objectif est de simuler suivant la loi μ_v définie à l'équation (4.18). Aussi, dans la proposition suivante, on montre que le noyau M_v est réversible pour μ_v et, en particulier, μ_v est stationnaire pour la chaîne de Markov construite par l'algorithme de Metropolis-Hastings.

Proposition 4.4.2. *La loi μ_v définie à l'équation (4.18) vérifie :*

$$\forall (\mathbf{x}, \mathbf{y}) \in \mathbb{X} \times \mathbb{X}, \quad \mu_v(d\mathbf{x})M_v(\mathbf{x}, d\mathbf{y}) = \mu_v(d\mathbf{y})M_v(\mathbf{y}, d\mathbf{x}),$$

ce qui implique :

$$\int_{\mathbb{X}} \mu_v(d\mathbf{x})M_v(\mathbf{x}, d\mathbf{y}) = \mu_v(d\mathbf{y}),$$

où $M_v(\mathbf{x}, \cdot)$ est le noyau de transition donné en (4.20).

À nouveau, ce résultat est classique mais une démonstration adaptée à notre contexte est proposée en section 4.6. Ce résultat assure que, sous certaines conditions sur le noyau $M_v(\mathbf{x}, \cdot)$ et lorsque $m \to +\infty$, la variable aléatoire $\mathbf{X}_m$ est approximativement distribuée suivant la loi cible μ_v. En effet, notons $M_v^m(\mathbf{x}, d\mathbf{y}) = \mathbb{P}(\mathbf{X}_m \in d\mathbf{y} \mid \mathbf{X}_0 = \mathbf{x})$, et rappelons que la distance en variation totale entre deux mesures de probabilié μ_1 et μ_2 est donnée par :

$$\|\mu_1 - \mu_2\|_{TV} = \sup_{B\in\mathcal{B}(\mathbb{X})} |\mu_1(B) - \mu_2(B)|.$$

On dit que la chaîne est uniformément ergodique sur $Q(v)$ s'il existe deux constantes $A > 0$ et $0 < r < 1$ telles que :

$$\sup_{\mathbf{x}\in Q(v)} \|M_v^m(\mathbf{x}, \cdot) - \mu_v\|_{TV} \leq Ar^m.$$

Le corollaire 3 de (Tierney, 1994) montre que s'il existe des constantes c, c_h et C_h telles que $\inf_{\mathbf{x}\in Q(v)} f_{\mathbf{X}}(\mathbf{x}) \geq c > 0$ et

$$\forall (\mathbf{x},\mathbf{y}) \in Q(v)\times Q(v), \quad 0 < c_h \leq h_v(\mathbf{x},\mathbf{y}) \leq C_h,$$

alors la chaîne est uniformément ergodique. Pour notre exemple, puisque $\mathbb{X}$ est compact, ceci est assuré à chaque étape de l'algorithme dès lors que l'on suppose $f_{\mathbf{X}}$ (respectivement h_v) continue et strictement positive sur $\mathbb{X}$ (respectivement $Q(v)$). C'est en particulier vrai pour le noyau gaussien présenté plus loin. Par ailleurs, on peut aussi considérer un noyau dit de Metropolis indépendant, i.e. de la forme $h_v(\mathbf{x},\mathbf{y}) = g_v(\mathbf{y})$ avec g_v une densité strictement positive sur $Q(v)$. Dans ce cas, le corollaire 4 de (Tierney, 1994) montre que si $\beta^{-1} = \sup_{\mathbf{y}\in Q(v)} f_v(\mathbf{y})/g_v(\mathbf{y})$ est fini, alors la chaîne est uniformément ergodique, de taux de convergence $r \leq 1-\beta$. Dans notre cadre, ceci est par exemple assuré si $f_{\mathbf{X}}$ est continue et strictement positive sur $\mathbb{X}$ et si, à chaque étape de l'algorithme, g_v correspond à la densité de la loi uniforme sur $Q(v)$.

4.4.3.3 Paramètres de l'algorithme de Metropolis-Hastings

La mise en œuvre pratique de l'algorithme de Metropolis-Hastings, et notamment le choix du noyau h_v, dépend du problème à traiter. Ce choix est discuté dans (Tierney, 1994) ou encore dans (Chib and Greenberg, 1995). Pour la méthode de splitting, des exemples de densités sont proposés dans (Guyader and Hengartner, 2011) et (Cérou et al., 2012).

Par exemple, une façon simple de procéder est de prendre la densité gaussienne symétrique :

$$h_v(\mathbf{x},\mathbf{y}) = \frac{1}{\sigma_v\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_v^2}\|\mathbf{y}-\mathbf{x}\|_2^2\right), \tag{4.21}$$

où $\|\cdot\|_2$ est la norme euclidienne. Partant de $\mathbf{X}_m = \mathbf{x}$, cela revient à proposer $\mathbf{Y} = \mathbf{x}+\sigma_v\mathbf{W}$, avec $\mathbf{W} \sim \mathcal{N}(0,\mathbf{I}_d)$ et $\sigma_v > 0$ fixé. Puisqu'alors $h_v(\mathbf{x},\mathbf{y}) = h_v(\mathbf{y},\mathbf{x})$, on est dans le cadre de Metropolis. La valeur du paramètre $\sigma_v > 0$ doit être réglée en fonction de la taille du cube dans lequel on cherche à simuler. Il faut s'assurer que les transitions ne soient pas systématiquement rejetées (σ_v choisi trop grand) ou acceptées (σ_v choisi trop petit et la chaîne ne s'éloigne pas assez vite de sa position initiale). Comme mentionné dans (Guyader and Hengartner, 2011), il est raisonnable de compter la proportion de transitions acceptées à chaque étape de l'algorithme de Metropolis-Hastings et, si cette proportion est inférieure à un certain taux, par exemple 20%, alors on réduit σ_v de 10% par exemple.

4.4.3.4 Création de l'échantillon Monte-Carlo

Grâce à l'algorithme de Metropolis-Hastings, on sait à partir de n'importe quel état initial $\mathbf{X}_0 \in Q(v)$ simuler une chaîne de Markov de loi stationnaire μ_v. On explique ici comment obtenir un échantillon Monte-Carlo $\mathbf{X}^v = (\mathbf{X}_i^v)_{1\leq i\leq N}$ approximativement i.i.d. suivant cette loi.

Soit $\bar{v} \in \mathcal{V}(\mathbf{t})$ et $v \in \mathbf{t}$ de parent $\bar{v}$, avec $Q(v) \subset Q(\bar{v})$. Supposons que l'on dispose d'un échantillon $\mathbf{X}^{\bar{v}} = (\mathbf{X}_i^{\bar{v}})_{1\leq i\leq N}$ i.i.d. suivant la loi $\mu_{\bar{v}}$, c'est-à-dire à valeurs dans $Q(\bar{v})$. Partant de $\mathbf{X}^{\bar{v}}$, on veut obtenir un échantillon i.i.d. $\mathbf{X}^v$ à valeurs dans $Q(v)$.

On commence par diviser $\mathbf{X}^{\bar{v}}$ en deux sous-groupes de points : ceux qui appartiennent déjà à $Q(v)$ et les autres. Le premier groupe va être conservé et le second va être éliminé, puis redistribué sur les points conservés. Notons :

$$G_{\bar{v}} = \{1 \leq i \leq N : \mathbf{X}_i^{\bar{v}} \in Q(v)\} \quad \text{et} \quad G'_{\bar{v}} = \{1 \leq i \leq N : \mathbf{X}_i^{\bar{v}} \notin Q(v)\} \quad .$$

La suite $(\mathbf{X}_i^{\bar{v}})_{i \in G_{\bar{v}}}$ étant déjà distribuée suivant μ_v, on conserve ce groupe de points. En revanche, les points de $(\mathbf{X}_i)_{i \in \bar{G}'_{\bar{v}}}$ sont redistribués indépendamment et de façon uniforme sur les points $(\mathbf{X}_i)_{i \in G_{\bar{v}}}$. À ce stade, nous disposons donc d'un ensemble de N points chacun distribué suivant la loi de $\mathbf{X}$ sachant $\mathbf{X} \in Q(v)$. Cependant, ils ne sont clairement pas indépendants. L'idée est d'appliquer l'algorithme de Metropolis-Hastings à chacun de ces points indépendamment un nombre suffisant de fois. On obtient ainsi un échantillon quasiment i.i.d. de la loi $\mathbf{P}_{\mathbf{X}}$ restreinte à $Q(v)$.

4.4.4 Exemple

Nous reprenons ici l'exemple en dimension $d = 1$ proposé en section 4.3.4. On rappelle que la vraie valeur de p est 2.08×10^{-3}. Nous mettons en place la procédure d'estimation par méthode de splitting décrite ci-dessus pour estimer les approximations $p^-(\mathbf{t})$ et $p^+(\mathbf{t})$. La taille de l'échantillon $(\mathbf{X}_i)_{1 \leq i \leq N}$ utilisé pour la méthode de splitting est $N = 10^5$. Pour mettre en œuvre l'algorithme de Metropolis-Hastings, on choisit la densité gaussienne symétrique (4.21) et la valeur du paramètre σ_v est égal à un quart de la longueur du cube dyadique $Q(v)$ dans lequel on souhaite simuler.

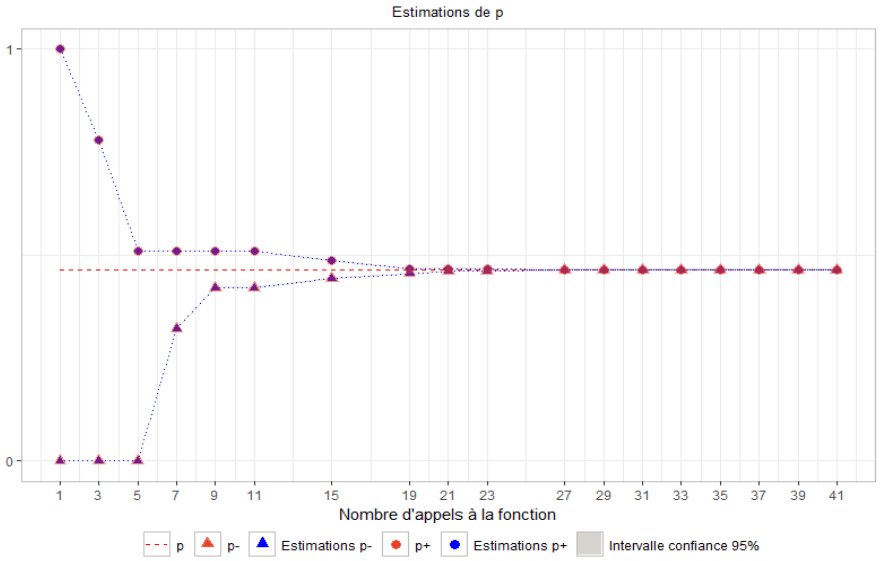

FIGURE 4.13 – Ligne pointillés rouge : la vraie valeur de p. Triangles rouges : les approximations successives $p^-(\mathbf{t})$. Points rouges : les approximations successives $p^+(\mathbf{t})$. Triangles bleus : les estimations successives $\widehat{p}^-(\mathbf{t})$ de $p^-(\mathbf{t})$. Points bleus : les estimations successives $\widehat{p}^+(\mathbf{t})$ de $p^+(\mathbf{t})$. Aires grises : les intervalles de confiance à 95% pour les approximations $p^-(\mathbf{t})$ et $p^+(\mathbf{t})$. Ces intervalles étant à peine visibles, on donne en figure 4.14 une version agrandie de ce graphique.

Dans les figures 4.13 et 4.14, on compare les vraies valeurs de $p^-(\mathbf{t})$ et $p^+(\mathbf{t})$ (en rouge) et leurs estimations (en bleu), en fonction du nombre d'appels à g. Les deux aires grises (qui sont visibles uniquement sur la figure agrandie 4.14) sont les intervalles de confiance à 95% définis en équation (4.17). Ils fournissent approximativement la même information quand la taille des cubes se réduit considérablement. Les résultats donnés dans le tableau 4.1 sont les valeurs numériques que retourne l'algorithme une fois le budget épuisé.

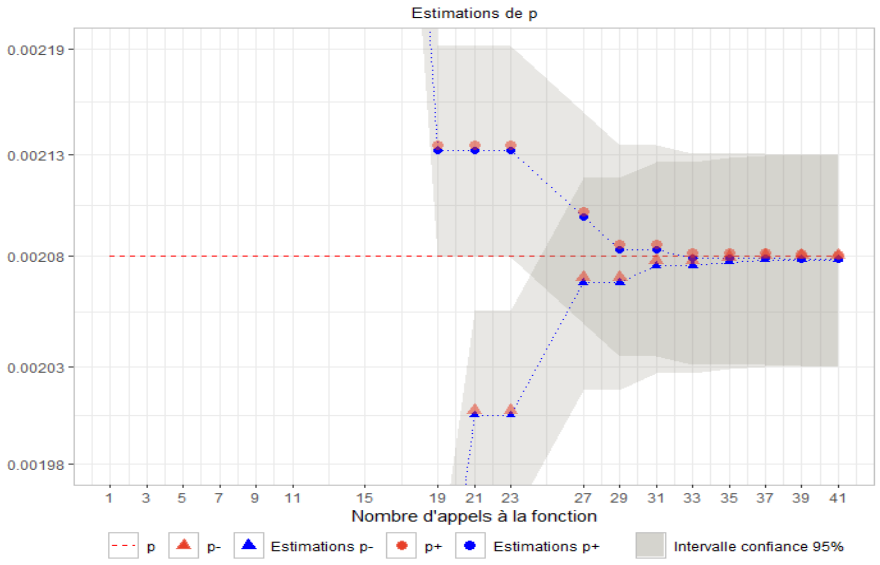

FIGURE 4.14 – Agrandissement de la figure 4.13.

	Approximation	Estimation	Intervalle de confiance à 95%
$p^-(\mathbf{t})$	2.0807×10^{-3}	2.0782×10^{-3}	$[2.0253 \times 10^{-3}\ ,\ 2.1311 \times 10^{-3}]$
$p^+(\mathbf{t})$	2.0807×10^{-3}	2.0785×10^{-3}	$[2.0256 \times 10^{-3}\ ,\ 2.1314 \times 10^{-3}]$

TABLE 4.1 – Résultats numériques associés aux figures 4.13 et 4.14 pour $n = 41$ appels à la fonction g. La taille de l'échantillon Monte-Carlo utilisé pour l'estimation est $N = 10^5$. La vraie valeur de p est 2.08×10^{-3}.

Dans la figure 4.15, on montre comment évolue l'échantillon Monte-Carlo (points orange) pour effectuer les estimations successives. En haut à gauche, l'échantillon est distribué suivant $\mathbf{P_X}$ afin d'estimer la probabilité $p^+(\mathbf{t}) = \mathbb{P}(\mathbf{X} \in [0, \frac{1}{2}]) + \mathbb{P}(\mathbf{X} \in [\frac{1}{2}, 1])$. En haut à droite, il est distribué suivant la restriction de $\mathbf{P_X}$ à $[\frac{1}{2}, 1]$ et permet d'estimer les probabilités $p^+(\mathbf{t}) = \mathbb{P}(\mathbf{X} \in [\frac{1}{2}, \frac{3}{4}]) + \mathbb{P}(\mathbf{X} \in [\frac{3}{4}, 1])$. En bas à gauche, il est distribué suivant la restriction de $\mathbf{P_X}$ à $[\frac{3}{4}, 1]$ et est utilisé pour estimer la probabilité

$p^+(\mathbf{t}) = \mathbb{P}\big(\mathbf{X} \in \left[\frac{3}{4}, \frac{7}{8}\right]\big) + \mathbb{P}\big(\mathbf{X} \in \left[\frac{7}{8}, 1\right]\big)$. Enfin, en bas à droite, l'échantillon est distribué suivant la restriction de $\mathbf{P_X}$ à $\left[\frac{3}{4}, \frac{7}{8}\right]$ et permet d'estimer la probabilité $p^+(\mathbf{t}) = p^-(\mathbf{t}) + \mathbb{P}\big(\mathbf{X} \in \left[\frac{3}{4}, \frac{13}{16}\right]\big) + \mathbb{P}\big(\mathbf{X} \in \left[\frac{13}{16}, \frac{7}{8}\right]\big)$, où $p^-(\mathbf{t}) = \mathbb{P}\big(\mathbf{X} \in \left[\frac{7}{8}, 1\right]\big)$ a déjà été estimée à l'étape précédente.

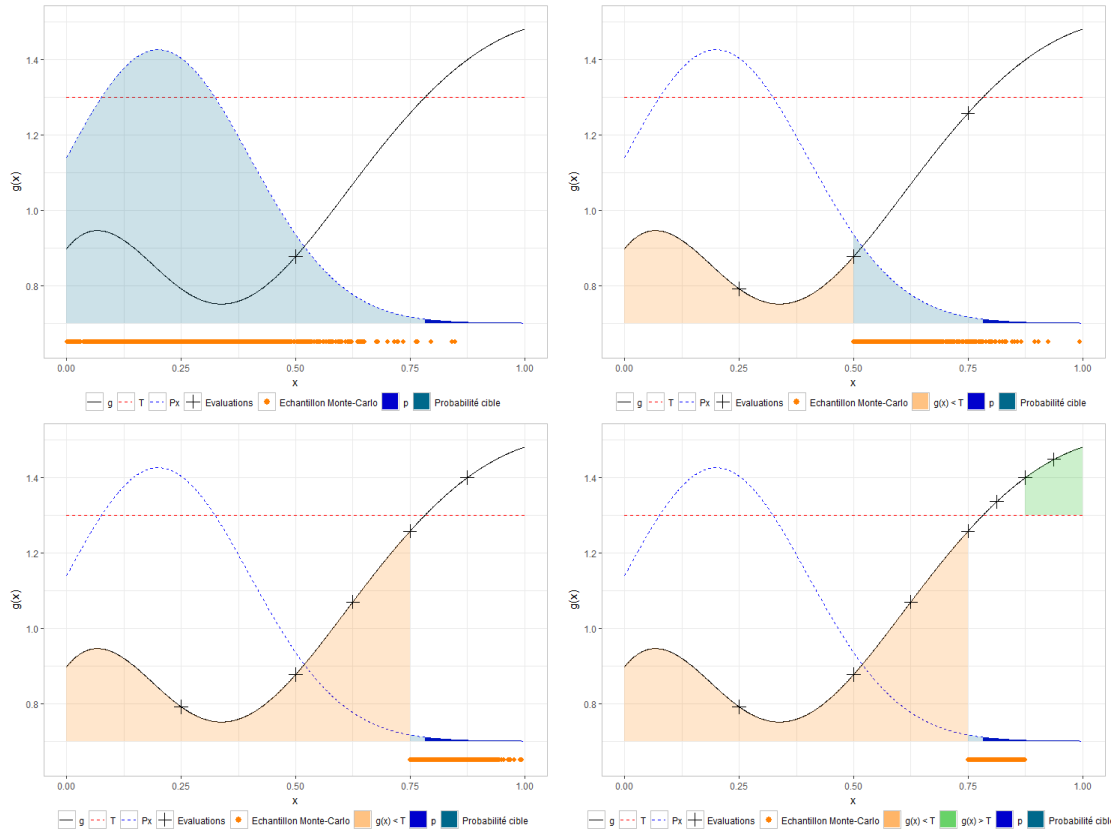

FIGURE 4.15 – Représentation graphique des échantillons Monte-Carlo pour la méthode de splitting.

4.5 Conclusion et perspectives

Pour répondre à une problématique d'estimation de probabilité d'évènement rare, notre algorithme présente plusieurs avantages. Tout d'abord, la procédure basée sur un découpage dyadique récursif de $\mathbb{X}$ permet de lui associer une représentation sous forme d'arbre 2^d-régulier. Cela induit un formalisme simplifié, permettant de mener une étude théorique des performances de l'algorithme. De plus, celui-ci fournit une approximation par défaut et une approximation par excès de la probabilité p. On montre que les erreurs d'approximation diminuent au fur et à mesure que l'algorithme progresse, et cela se vérifie effectivement dans les résultats de simulations numériques. Grâce à la méthode de splitting, on met en place une procédure efficace pour estimer ces approximations, qui, à partir d'un certain

nombre d'itérations, prennent nécessairement des valeurs très faibles. La précision des estimations que retourne l'algorithme peut être mesurée, et dépend du nombre d'évaluations de la fonction g et de la taille de l'échantillon Monte-Carlo utilisé pour estimer les probabilités. On dispose notamment d'intervalles de confiance, permettant de statuer quant à l'incertitude sur les résultats obtenus. Parmi les méthodes qui existent dans la littérature pour répondre à une problématique similaire à la nôtre, cette précision des estimateurs n'est pas systématique.

Rappelons que l'algorithme repose sur l'hypothèse que la fonction g est lipschitzienne et qu'une constante L vérifiant $(\mathcal{H}_1)$ est connue. En pratique, il n'est pas évident de vérifier cela et, en particulier, d'avoir accès à une valeur de L lorsque la fonction est inconnue. Dans (Cohen et al., 2013), où cette hypothèse est aussi utilisée pour mettre en œuvre un algorithme d'optimisation, les auteurs proposent une solution pour cela : ils développent une autre version de l'algorithme et, à chaque itération, utilisent l'information fournie sur les valeurs déjà observées de g pour approcher une constante L vérifiant $(\mathcal{H}_1)$. Malgré l'intérêt de cette procédure, nous ne l'avons pas traitée dans cette thèse, mais il s'agit d'une perspective qu'il serait intéressant d'explorer. Notons que des méthodes pour estimer une constante de Lipschitz peuvent aussi être trouvées, par exemple, dans (De Haan, 1981), (Jones et al., 1993) et (Wood and Zhang, 1996). De plus, nous l'avons vu au chapitre précédent, les simulateurs numériques en industrie disposent parfois des valeurs du gradient de g aux points d'évaluation. Pour estimer une constante L vérifiant $(\mathcal{H}_1)$, on pourrait donc envisager de mettre en œuvre une méthode itérative, en parallèle de l'algorithme pour l'estimation de p, qui utilise l'information disponible sur la dérivée.

4.6 Démonstrations

On rappelle que la notation $|E|$ désigne le cardinal d'un ensemble fini $E \subseteq \mathbb{X}$ et que $\lambda(E)$ désigne la mesure de Lebesgue d'un ensemble mesurable. On rappelle également que les hypothèses $(\mathcal{H}_1)$, $(\mathcal{H}_2)$ et $(\mathcal{H}_3)$, énoncées en section 4.2, sont supposées vérifiées.

Démonstration de la propriété 4.3.1

Considérons l'algorithme décrit en section 4.3.3. On suppose un nombre $k \in \mathbb{N}^*$ d'itérations déjà effectuées et on note $\mathbf{t}$ l'arbre de hauteur k qui en résulte. On rappelle que tout cube dyadique identifié par un nœud u dans $\mathbf{t}$ est noté $Q(u)$ et que $\mathbf{c}_Q$ désigne son centre.

Au cours de l'itération $j = 1, \ldots, k$, l'algorithme génère un ensemble $\tilde{\mathcal{A}}(\mathbf{t})$ composé de nœuds $u \in \mathbf{t}$ qui vérifient, d'une part :

$$\begin{cases} |u| = j - 1, \\ \lambda\left(Q(u)\right) = \frac{1}{2^{d(j-1)}}, \\ \|\mathbf{x} - \mathbf{c}_Q\| \leq \frac{1}{2^j}, \ \forall\ \mathbf{x} \in Q(u), \end{cases}$$

et, d'autre part :

$$|g(\mathbf{c}_Q) - T| \leq \frac{L}{2^j}.$$

Cela implique que, pour tout $\mathbf{x} \in Q(u)$, on a :

$$\begin{aligned} g(\mathbf{x}) &= g(\mathbf{c}_Q) + g(\mathbf{x}) - g(\mathbf{c}_Q) \\ g(\mathbf{x}) &\leq T + \frac{L}{2^j} + L \left\| \mathbf{x} - \mathbf{c}_Q \right\|, \text{ car } g \text{ est } L\text{-lipschitzienne d'après } (\mathcal{H}_1), \\ g(\mathbf{x}) &\leq T + \frac{L}{2^{j-1}}, \end{aligned}$$

et, de la même façon, $g(\mathbf{x}) \geq T - \frac{L}{2^{j-1}}$. Autrement dit, pour tout $\mathbf{x} \in Q(u)$, on a :

$$|g(\mathbf{x}) - T| \leq \frac{L}{2^{j-1}}.$$

Par conséquent, le cardinal $|\tilde{\mathcal{A}}(\mathbf{t})|$ de $\tilde{\mathcal{A}}(\mathbf{t})$ satisfait :

$$\frac{|\tilde{\mathcal{A}}(\mathbf{t})|}{2^{d(j-1)}} \leq \lambda \left(\left\{ \mathbf{x} \in \mathbb{X} : |g(\mathbf{x}) - T| \leq \frac{L}{2^{j-1}} \right\} \right).$$

Or, d'après $(\mathcal{H}_2)$, cela implique que :

$$\frac{|\tilde{\mathcal{A}}(\mathbf{t})|}{2^{d(j-1)}} \leq C \left(\frac{2}{L} \times \frac{L}{2^{(j-1)}} \right)^d,$$

c'est-à-dire :

$$|\tilde{\mathcal{A}}(\mathbf{t})| \leq 2^d C. \tag{4.22}$$

Pour tout $j = 1, \ldots, k$, notons μ_j le nombre de nœuds étiquetés $\mathcal{A}$ à la fin de la j-ième itération (c'est-à-dire après l'étape de mise à jour de l'arbre) :

$$\mu_j = |\mathcal{A}(\mathbf{t})| = 2^d |\tilde{\mathcal{A}}(\mathbf{t})|.$$

D'après ce qui précède, on peut dire que :

$$\mu_j \leq 4^d C, \ \forall j = 1, \ldots, k. \tag{4.23}$$

Ainsi, puisque le nombre n_k d'évaluations de g à la fin de l'étape $k \geq 2$ vérifie :

$$n_k = 1 + \sum_{j=1}^{k-1} \mu_j.$$

On a :

$$n_k \leq 1 + \sum_{j=1}^{k-1} 4^d C = 1 + 4^d C(k-1).$$

Démonstration de la propriété 4.3.2

Soit $k \in \mathbb{N}^*$. Supposons k itérations déjà effectuées et notons $\mathbf{t}$ l'arbre de profondeur k construit par l'algorithme. Pour toute feuille $u \in \mathcal{A}(\mathbf{t})$, on a :

$$\lambda\left(Q(u)\right) = \frac{1}{2^{dk}},$$

où $Q(u)$ est le cube dyadique identifié par u dans l'arbre $\mathbf{t}$. D'après $(\mathcal{H}_3)$, la densité $f_{\mathbf{X}}$ vérifie : $0 < \|f_{\mathbf{X}}\| = \sup_{\mathbf{x}\in\mathbb{X}} f_{\mathbf{X}}(\mathbf{x}) < \infty$. On a alors :

$$\mathbb{P}\big(\mathbf{X} \in Q(u)\big) \leq \frac{\|f_{\mathbf{X}}\|}{2^{dk}}.$$

Dans la démonstration de la propriété 4.3.1, on a montré à l'inégalité (4.23) que le cardinal $|\mathcal{A}(\mathbf{t})|$ de $\mathcal{A}(\mathbf{t})$ est majoré à chaque itération par $4^d C$, où $C \geq 1$ est la constante de l'hypothèse $(\mathcal{H}_2)$. Ainsi,

$$\sum_{u\in\mathcal{A}(\mathbf{t})} \mathbb{P}\big(\mathbf{X} \in Q(u)\big) \leq \sum_{u\in\mathcal{A}(\mathbf{t})} \frac{\|f_{\mathbf{X}}\|}{2^{dk}} = \frac{\|f_{\mathbf{X}}\|}{2^{dk}} |\mathcal{A}(\mathbf{t})| \leq \|f_{\mathbf{X}}\| 4^d C 2^{-dk}.$$

Par conséquent, puisque les approximations $p^-(\mathbf{t})$ et $p^+(\mathbf{t})$ vérifient l'inégalité :

$$p^-(\mathbf{t}) \leq p \leq p^+(\mathbf{t}) = p^-(\mathbf{t}) + \sum_{u\in\mathcal{A}(\mathbf{t})} \mathbb{P}\big(\mathbf{X} \in Q(u)\big).$$

Elles satisfont également :

$$p^-(\mathbf{t}) \leq p \leq p^-(\mathbf{t}) + \|f_{\mathbf{X}}\| 4^d C 2^{-dk}, \tag{4.24}$$

et

$$p \leq p^+(\mathbf{t}) \leq p + \|f_{\mathbf{X}}\| 4^d C 2^{-dk}. \tag{4.25}$$

Démonstration de la propriété 4.3.3

Pour tout $k \in \mathbb{N}^*$, notons n_k le nombre total d'évaluations de la fonction g à la fin de l'étape k de l'algorithme. D'après la propriété 4.3.1, on a :

$$n_k \leq 1 + 4^d C(k-1).$$

où $C \geq 1$. On a donc :

$$\frac{n_k - 1}{4^d C} + 1 \leq k,$$

ou encore :

$$2^{-dk} \leq 2^{-d} 2^{-\frac{d(n_k-1)}{4^d C}}.$$

Par la propriété 4.3.2, on a alors :

$$p - p^-(\mathbf{t}) \leq \|f_{\mathbf{X}}\| 2^d C 2^{-\frac{d(n_k-1)}{4^d C}},$$

et

$$p^+(\mathbf{t}) - p \leq \|f_{\mathbf{X}}\| 2^d C 2^{-\frac{d(n_k-1)}{4^d C}}.$$

Par conséquent, les erreurs d'approximations $e^-_{n_k}(\mathbf{t}) = p - p^-(\mathbf{t})$ et $e^+_{n_k}(\mathbf{t}) = p^+(\mathbf{t}) - p$ vérifient :

$$\max\left(e^-_{n_k}(\mathbf{t}), e^+_{n_k}(\mathbf{t})\right) \leq \|f_{\mathbf{X}}\| 2^d C 2^{-\frac{d(n_k-1)}{4^d C}}.$$

Démonstration du lemme 4.4.1.

Soit $\mathbf{t}$ un sous-arbre fini de l'arbre 2^d-régulier. On note $\mathcal{V}(\mathbf{t})$ l'ensemble des nœuds internes de $\mathbf{t}$ (c'est-à-dire les nœuds de $\mathbf{t}$ qui ne sont pas des feuilles). On note $\mathcal{C}(v)$ l'ensemble des 2^d enfants d'un nœud $v \in \mathcal{V}(\mathbf{t})$.

Hypothèses

Pour tout $v \in \mathcal{V}(\mathbf{t})$, on dispose d'un échantillon $\mathbf{X}^v = (\mathbf{X}^v_i)_{1\leq i\leq N}$ i.i.d. suivant la loi de $\mathbf{X}$ sachant $\mathbf{X} \in Q(v)$. On suppose de plus que les échantillons $(\mathbf{X}^v)_{v\in\mathcal{V}(\mathbf{t})}$ sont indépendants.

Tirages

Soit $\bar{v} \in \mathcal{V}(\mathbf{t})$ et $Q(\bar{v}) = \bigcup_{v\in\mathcal{C}(\bar{v})} Q(v)$. On pose :

$$C_N^{\bar{v},v} = \sum_{i=1}^N \mathbb{1}_{\mathbf{X}_i^{\bar{v}}\in Q(v)}, \quad \forall\, v \in \mathcal{C}(\bar{v}).$$

Pour tout $v \in \mathcal{C}(\bar{v})$, la variable aléatoire $C_N^{\bar{v},v}$ est le nombre de variables aléatoires $(\mathbf{X}_i^{\bar{v}})_{1\leq i\leq N}$ qui sont dans le cube $Q(v) \subset Q(\bar{v})$. D'après les hypothèses ci-dessus, elle suit une loi binomiale de paramètres N et $p(\bar{v}, v)$. Sa variance est donc égale à :

$$\mathrm{Var}\big[C_N^{\bar{v},v}\big] = Np(\bar{v},v)(1-p(\bar{v},v)).$$

De plus, on a :

$$\sum_{v\in\mathcal{C}(\bar{v})} C_N^{\bar{v},v} = N, \quad \text{et} \quad \sum_{v\in\mathcal{C}(\bar{v})} p(\bar{v},v) = 1,$$

et le vecteur $\big(C_N^{\bar{v},v}\big)_{v\in\mathcal{C}(\bar{v})}$ suit une loi multinomiale de paramètres N et $\big(p(\bar{v},v)\big)_{v\in\mathcal{C}(\bar{v})}$. Par conséquent, pour tous $v, v' \in \mathcal{C}(\bar{v})$, $v \neq v'$, on a :

$$\mathrm{Cov}\big(C_N^{\bar{v},v}, C_N^{\bar{v},v'}\big) = -Np(\bar{v},v)p(\bar{v},v').$$

Soient $\bar{v}, \bar{v}' \in \mathcal{V}(\mathbf{t})$, $\bar{v} \neq \bar{v}'$. Les échantillons $(\mathbf{X}^v)_{v\in\mathcal{V}(\mathbf{t})}$ étant indépendantes, les vecteurs $(C_N^{\bar{v},v})_{v\in\mathcal{C}(\bar{v})}$ et $(C_N^{\bar{v}',v'})_{v'\in\mathcal{C}(\bar{v}')}$ sont nécessairement indépendants.

Dépendance entre les estimateurs

Pour tout $\bar{v} \in \mathcal{V}(\mathbf{t})$ et pour tout $v \in \mathcal{C}(\bar{v})$, on a : $p(\bar{v}, v) = \mathbb{P}\big(\mathbf{X} \in Q(v) \mid \mathbf{X} \in Q(\bar{v})\big)$. L'estimateur $\widehat{p}(\bar{v}, v)$ de cette probabilité par la méthode Monte-Carlo naïve est défini par :

$$\widehat{p}(\bar{v}, v) = \frac{C_N^{\bar{v},v}}{N}.$$

Cet estimateur est sans biais et sa variance est :

$$\mathrm{Var}\big[\widehat{p}(\bar{v}, v)\big] = \frac{p(\bar{v}, v)(1 - p(\bar{v}, v))}{N}.$$

Pour tout $v, v' \in \mathcal{C}(\bar{v})$, $v \neq v'$, les estimateurs $\widehat{p}(\bar{v}, v)$ et $\widehat{p}(\bar{v}, v')$ sont corrélées négativement :

$$\mathrm{Cov}\big(\widehat{p}(\bar{v}, v), \widehat{p}(\bar{v}, v')\big) = -\frac{p(\bar{v}, v)p(\bar{v}, v')}{N}.$$

Pour tout $\bar{v}, \bar{v}' \in \mathcal{V}(\mathbf{t})$, $\bar{v} \neq \bar{v}'$, les vecteurs $(\widehat{p}(\bar{v}, v))_{v \in \mathcal{C}(\bar{v})}$ et $(\widehat{p}(\bar{v}', v'))_{v' \in \mathcal{C}(\bar{v}')}$ sont indépendants donc décorrélés.

À propos du lemme

Pour tout $v \in \mathbf{t}$, on estime la probabilité conditionnelle $p(\bar{v}, v)$ par :

$$\widehat{p}(\bar{v}, v) = \frac{C_N^{\bar{v},v}}{N}.$$

D'après ce qui précède, et d'après la loi forte des grands nombres, on a :

$$\widehat{p}(\bar{v}, v) \xrightarrow[N \to \infty]{p.s.} p(\bar{v}, v). \tag{4.26}$$

En notant $\Gamma = (\Gamma(v, v'))_{v \in \mathbf{t}, v' \in \mathbf{t}}$ la matrice de covariance suivante :

$$\Gamma(v, v') = \begin{cases} p(\bar{v}, v)\big(1 - p(\bar{v}, v)\big) & \text{si } v = v', \\ -p(\bar{v}, v)p(\bar{v}, v') & \text{si } v \neq v' \text{ et } \bar{v} = \bar{v}', \\ 0 & \text{si } v \neq v' \text{ et } \bar{v} \neq \bar{v}', \end{cases}$$

et en posant $\widehat{\mathbf{q}} = (\widehat{p}(\bar{v}, v))_{v \in \mathbf{t}}$ et $\mathbf{q} = (p(\bar{v}, v))_{v \in \mathbf{t}}$, le théorème central-limite multidimensionnel s'applique, et on a :

$$\sqrt{N}(\widehat{\mathbf{q}} - \mathbf{q}) \xrightarrow[N \to \infty]{\mathcal{L}} \mathcal{N}(0, \Gamma).$$

Démonstration du théorème 4.4.1.

Soit $\mathbf{t} \setminus \{\emptyset\}$ un arbre étiqueté 2^d-régulier construit par l'algorithme et $\mathcal{U}(\mathbf{t})$ un ensemble de feuilles de $\mathbf{t} \setminus \{\emptyset\}$. Dans cette démonstration, on s'intéresse à la consistance et à la normalité de l'estimateur suivant :

$$\widehat{p}(\mathbf{t}) = \sum_{u \in \mathcal{U}(\mathbf{t})} \prod_{v \in \mathbf{t}_u} \widehat{p}(\bar{v}, v),$$

où $\mathbf{t}_u$ est la branche de $\mathbf{t}$ de feuille $u \in \mathcal{U}(\mathbf{t})$. Afin d'alléger les notations, on remplace dans les calculs ci-dessous $p(\bar{v}, v)$ par $q(v)$.

D'après le lemme 4.4.1, la consistance de $\widehat{p}(\mathbf{t})$ est évidente. La normalité asymptotique est une application de la méthode delta. La quantité $p(\mathbf{t})$ s'écrit comme une fonction F du vecteur $\mathbf{q} = \big(q(v)\big)_{v\in\mathbf{t}}$:

$$p(\mathbf{t}) = F(\mathbf{q}) = \sum_{u\in\mathcal{U}(\mathbf{t})} \prod_{v\in\mathbf{t}_u} q(v).$$

Puisque F est à valeurs réelles, on assimile ∂F à un vecteur ligne : $\partial F = (\partial_v F)_{v\in\mathbf{t}}$, où $\partial_v F$ est la dérivée partielle de F par rapport à $q(v)$. Elle est donnée par :

$$\partial_v F(\mathbf{q}) = \sum_{\substack{u\in\mathcal{U}(\mathbf{t})\\ v\prec u}} \prod_{\substack{w\in\mathbf{t}_u\\ w\neq v}} q(w) = \sum_{\substack{u\in\mathcal{U}(\mathbf{t})\\ v\prec u}} \frac{p(u)}{q(v)},$$

où $p(u) = \prod_{v\in\mathbf{t}_u} q(v)$. D'après le résultat de convergence en loi du lemme 4.4.1, en notant le vecteur des estimateurs $\hat{\mathbf{q}} = (\widehat{q}(v))_{v\in\mathbf{t}} = (\widehat{p}(\bar{v}, v))_{v\in\mathbf{t}}$, on a :

$$\sqrt{N}(\hat{\mathbf{q}} - \mathbf{q}) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}(0, \Gamma),$$

où Γ est donnée à l'équation (4.15). En appliquant la méthode delta, cela implique :

$$\sqrt{N}\big(F(\hat{\mathbf{q}}) - F(\mathbf{q})\big) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}(0, \sigma(\mathbf{t})^2),$$

où :

$$\begin{aligned}
\sigma(\mathbf{t})^2 = (\partial F)\Gamma(\partial F)^T &= \sum_{v,w\in\mathbf{t}} (\partial_v F)\Gamma(v,w)(\partial_w F)\\
&= \sum_{v\in\mathbf{t}} \Gamma(v,v)(\partial_v F)^2 + \sum_{\substack{v\neq w\in\mathbf{t}\\ \bar{v}=\bar{w}}} \Gamma(v,w)(\partial_v F)(\partial_w F)\\
&= \sum_{v\in\mathbf{t}} q(v)\big(1-q(v)\big)\left[\sum_{\substack{u\in\mathcal{U}(\mathbf{t})\\ v\prec u}} \frac{p(u)}{q(v)}\right]^2 - \sum_{\substack{v\neq w\in\mathbf{t}\\ \bar{v}=\bar{w}}} q(v)q(w)\left[\sum_{\substack{u\in\mathcal{U}(\mathbf{t})\\ v\prec u}} \frac{p(u)}{q(v)}\right]\left[\sum_{\substack{u'\in\mathcal{U}(\mathbf{t})\\ w\prec u'}} \frac{p(u')}{q(w)}\right]\\
&= \sum_{v\in\mathbf{t}} \frac{1-q(v)}{q(v)}\left[\sum_{\substack{u\in\mathcal{U}(\mathbf{t})\\ v\prec u}} p(u)\right]^2 - \sum_{\substack{v\neq w\in\mathbf{t}\\ \bar{v}=\bar{w}}} \left[\sum_{\substack{u\in\mathcal{U}(\mathbf{t})\\ u\prec v}} p(u)\right]\left[\sum_{\substack{u'\in\mathcal{U}(\mathbf{t})\\ w\prec u'}} p(u')\right].
\end{aligned}$$

Posons :

$$A = \sum_{v\in\mathbf{t}} \frac{1-q(v)}{q(v)}\left[\sum_{\substack{u\in\mathcal{U}(\mathbf{t})\\ v\prec u}} p(u)\right]^2 \quad \text{et} \quad B = \sum_{\substack{v\neq w\in\mathbf{t}\\ \bar{v}=\bar{w}}} \left[\sum_{\substack{u\in\mathcal{U}(\mathbf{t})\\ v\prec u}} p(u)\right]\left[\sum_{\substack{u'\in\mathcal{U}(\mathbf{t})\\ w\prec u'}} p(u')\right].$$

On a :

$$A = \sum_{v\in\mathbf{t}} \sum_{\substack{u\in\mathcal{U}(\mathbf{t})\\ v\prec u}} p(u)^2 \frac{1-q(v)}{q(v)} + \sum_{v\in\mathbf{t}} \sum_{\substack{u,u'\in\mathcal{U}(\mathbf{t})\\ v\prec u,\ v\prec u'}} p(u)p(u') \frac{1-q(v)}{q(v)}$$
$$= \sum_{u\in\mathcal{U}(\mathbf{t})} \sum_{v\in\mathbf{t}_u} p(u)^2 \frac{1-q(v)}{q(v)} + \sum_{u\neq u'\in\mathcal{U}(\mathbf{t})} \sum_{v\in\mathbf{t}_u\cap\mathbf{t}_{u'}} p(u)p(u') \frac{1-q(v)}{q(v)}.$$

Pour permuter les sommes ci-dessus, on a utilisé l'équivalence entre $v \prec u$ et $v \in \mathbf{t}_u$. On peut enfin remarquer que $\mathbf{t}_u \cap \mathbf{t}_{u'} = \mathbf{t}_{u\wedge u'}$:

$$A = \sum_{u\in\mathcal{U}(\mathbf{t})} \sum_{v\in\mathbf{t}_u} p(u)^2 \frac{1-q(v)}{q(v)} + \sum_{u\neq u'\in\mathcal{U}(\mathbf{t})} \sum_{v\in\mathbf{t}_{u\wedge u'}} p(u)p(u') \frac{1-q(v)}{q(v)}.$$

On a aussi,

$$B = \sum_{\substack{v\neq w\in\mathbf{t}\\ \bar{v}=\bar{w}}} \sum_{\substack{u,u'\in\mathcal{U}(\mathbf{t})\\ v\prec u,\ w\prec u'}} p(u)p(u') = \sum_{u\neq u'\in\mathcal{U}(\mathbf{t})} p(u)p(u').$$

Ainsi,

$$\sigma(\mathbf{t})^2 = A - B$$
$$= \sum_{u\in\mathcal{U}(\mathbf{t})} p(u)^2 \sum_{v\in\mathbf{t}_u} \frac{1-q(v)}{q(v)} + \sum_{\substack{u,u'\in\mathcal{U}(\mathbf{t})\\ u\neq u'}} p(u)p(u') \left[\sum_{v\in\mathbf{t}_{u\wedge u'}} \frac{1-q(v)}{q(v)} - 1\right].$$

On retrouve bien la variance $\sigma(\mathbf{t})^2$ annoncée.

Démonstration de la proposition 4.4.1.

Considérons l'algorithme de Metropolis-Hastings décrit en section 4.4.3.1. Notons $M_v(\mathbf{x}, d\mathbf{y})$ le noyau de transition de la chaîne de la chaîne de Markov $(\mathbf{X}_m)_{m\in\mathbb{N}}$ associée. Par construction, $\mathbf{X}_0 \in Q(v)$, donc $\mathbf{X}_m \in Q(v)$ pour tout $m \in \mathbb{N}$.

Soit $\varphi : \mathbb{X} \to \mathbb{R}$ une fonction mesurable bornée. Partant de $\mathbf{X}_m \in Q(v)$, la proposition $\mathbf{Y}$ est générée suivant $h_v(\mathbf{X}_m, \cdot)$ et on a alors :

$$\mathbb{E}\big[\varphi(\mathbf{X}_{m+1}) \mid \mathbf{X}_m\big] = \mathbb{E}\big[\mathbb{E}\big[\varphi(\mathbf{X}_{m+1}) \mid \mathbf{X}_m, \mathbf{Y}\big] \mid \mathbf{X}_m\big]$$
$$= \mathbb{E}\big[\mathbb{E}\big[\varphi(\mathbf{Y})\mathbb{1}_{U\leq\alpha_v(\mathbf{X}_m,\mathbf{Y})} + \varphi(\mathbf{X}_m)\mathbb{1}_{U>\alpha_v(\mathbf{X}_m,\mathbf{Y})} \mid \mathbf{X}_m, \mathbf{Y}\big] \mid \mathbf{X}_m\big]$$
$$= \mathbb{E}\big[\varphi(\mathbf{Y})\alpha_v(\mathbf{X}_m, \mathbf{Y}) + \varphi(\mathbf{X}_m)\big(1 - \alpha_v(\mathbf{X}_m, \mathbf{Y})\big) \mid \mathbf{X}_m\big]$$
$$= \int_{\mathbb{X}} \varphi(\mathbf{y})\alpha_v(\mathbf{X}_m, \mathbf{y})h_v(\mathbf{X}_m, \mathbf{y})d\mathbf{y} + \varphi(\mathbf{X}_m) \int_{\mathbb{X}} \big(1 - \alpha_v(\mathbf{X}_m, \mathbf{y})\big)h_v(\mathbf{X}_m, \mathbf{y})d\mathbf{y}.$$

Soit $\mathbf{x} \in Q(v)$. On pose $r_v(\mathbf{x}) = 1 - \int_{\mathbb{X}} \alpha_v(\mathbf{x}, \mathbf{z}) h_v(\mathbf{x}, \mathbf{z}) d\mathbf{z}$. Puisque $\int_{\mathbb{X}} h_v(\mathbf{x}, \mathbf{z}) d\mathbf{z} = 1$, le terme de droite s'écrit :

$$\begin{aligned}\varphi(\mathbf{X}_m) \int_{\mathbb{X}} \big(1 - \alpha_v(\mathbf{X}_m, \mathbf{y})\big) h_v(\mathbf{X}_m, \mathbf{y}) d\mathbf{y} &= \varphi(\mathbf{X}_m) \left(1 - \int_{\mathbb{X}} \alpha_v(\mathbf{X}_m, \mathbf{y}) h_v(\mathbf{X}_m, \mathbf{y}) d\mathbf{y}\right) \\ &= \left[\int_{\mathbb{X}} \varphi(\mathbf{y}) \delta_{\mathbf{X}_m}(d\mathbf{y})\right] r_v(\mathbf{X}_m).\end{aligned}$$

Par conséquent,

$$\begin{aligned}\mathbb{E}\big[\varphi(\mathbf{X}_{m+1}) \mid \mathbf{X}_m\big] &= \int_{\mathbb{X}} \varphi(\mathbf{y}) \Big[\alpha_v(\mathbf{X}_m, \mathbf{y}) h_v(\mathbf{X}_m, \mathbf{y}) d\mathbf{y} + \delta_{\mathbf{X}_m}(d\mathbf{y}) r_v(\mathbf{X}_m)\Big] d\mathbf{y} \\ &= \int_{\mathbb{X}} \varphi(\mathbf{y}) K_v(\mathbf{X}_m, d\mathbf{y}),\end{aligned}$$

où $K_v(\mathbf{x}, d\mathbf{y}) = \alpha_v(\mathbf{x}, \mathbf{y}) h_v(\mathbf{x}, \mathbf{y}) d\mathbf{y} + \delta_{\mathbf{x}}(d\mathbf{y}) r_v(\mathbf{x})$. Puisque par convention $M_v(\mathbf{x}, d\mathbf{y}) = \delta_{\mathbf{x}}(d\mathbf{y})$ si $\mathbf{x} \notin Q(v)$, on a ainsi montré que le noyau de transition $M_v(\mathbf{x}, d\mathbf{y})$ s'écrit :

$$M_v(\mathbf{x}, d\mathbf{y}) = \begin{cases} \delta_{\mathbf{x}}(d\mathbf{y}) & \text{si } \mathbf{x} \notin Q(v), \\ K_v(\mathbf{x}, d\mathbf{y}) & \text{si } \mathbf{x} \in Q(v), \end{cases}$$

et $\forall \mathbf{x} \in \mathbb{X}$, on a bien $\int_{\mathbb{X}} M_v(\mathbf{x}, d\mathbf{y}) = 1$.

Démonstration de la proposition 4.4.2.

Dans cette démonstration, on veut montrer que loi μ_v est stationnaire pour la chaîne de Markov de noyau transition $M_v(\mathbf{x}, \cdot)$ décrite à la proposition 4.4.2. Pour cela, on commence par montrer que M_v est réversible pour μ_v, c'est-à-dire que

$$\mu_v(d\mathbf{x}) M_v(\mathbf{x}, d\mathbf{y}) = \mu_v(d\mathbf{y}) M_v(\mathbf{y}, d\mathbf{x}), \ \forall (\mathbf{x}, \mathbf{y}) \in \mathbb{X}^2, \tag{4.27}$$

ce qui assurera en particulier que μ_v est stationnaire.

∘ Si $\mathbf{x} \notin Q(v)$ et $\mathbf{y} \notin Q(v)$, alors on a $f_v(\mathbf{x}) = f_v(\mathbf{y}) = 0$, et l'égalité (4.27) est vérifiée.

∘ Si $\mathbf{x} \notin Q(v)$ et $\mathbf{y} \in Q(v)$, alors on a $f_v(\mathbf{x}) = 0$ et $M_v(\mathbf{y}, d\mathbf{x}) = K_v(\mathbf{y}, d\mathbf{x}) = \alpha_v(\mathbf{y}, \mathbf{x}) h_v(\mathbf{y}, \mathbf{x}) d\mathbf{x}$ car nécessairement $\mathbf{x} \neq \mathbf{y}$. Puisque $f_v(\mathbf{x}) = 0 \Rightarrow \alpha_v(\mathbf{y}, \mathbf{x}) = 0$, l'égalité (4.27) est vérifiée.

∘ Si $\mathbf{x} \in Q(v)$ et $\mathbf{y} \in Q(v)$, avec $\mathbf{x} = \mathbf{y}$, l'égalité (4.27) est évidemment vérifiée.

∘ Si $\mathbf{x} \in Q(v)$ et $\mathbf{y} \in Q(v)$, avec $\mathbf{x} \neq \mathbf{y}$, alors on a :

$$\begin{aligned}f_v(\mathbf{x}) h(\mathbf{x}, \mathbf{y}) \alpha_v(\mathbf{x}, \mathbf{y}) &= f_v(\mathbf{x}) h_v(\mathbf{x}, \mathbf{y}) \min\left(1, \frac{f_v(\mathbf{y}) h_v(\mathbf{y}, \mathbf{x})}{f_v(\mathbf{x}) h_v(\mathbf{x}, \mathbf{y})}\right) \\ &= \min\big(f_v(\mathbf{x}) h_v(\mathbf{x}, \mathbf{y}), f_v(\mathbf{y}) h_v(\mathbf{y}, \mathbf{x})\big).\end{aligned}$$

On en déduit que :

$$f_v(\mathbf{x}) h_v(\mathbf{x}, \mathbf{y}) \alpha_v(\mathbf{x}, \mathbf{y}) = f_v(\mathbf{y}) h_v(\mathbf{y}, \mathbf{x}) \alpha_v(\mathbf{y}, \mathbf{x}).$$

Cela entraîne que :

$$\mu_v(d\mathbf{x})\alpha_v(\mathbf{x},\mathbf{y})h(\mathbf{x},\mathbf{y})d\mathbf{y} = \mu_v(d\mathbf{y})\alpha_v(\mathbf{y},\mathbf{x})h_v(\mathbf{y},\mathbf{x})d\mathbf{x},$$

et, par extension, que :

$$\mu_v(d\mathbf{x})M_v(\mathbf{x},d\mathbf{y}) = \mu_v(d\mathbf{y})M_v(\mathbf{y},d\mathbf{x}).$$

L'égalité (4.27) étant vérifiée pour tout $(\mathbf{x},\mathbf{y}) \in \mathbb{X}^2$, on a alors :

$$\begin{aligned}\big(\mu_v M_v\big)(d\mathbf{y}) &= \int_{\mathbb{X}} \mu_v(d\mathbf{x})M_v(\mathbf{x},d\mathbf{y}) \\ &= \int_{\mathbb{X}} \mu_v(d\mathbf{y})M_v(\mathbf{y},d\mathbf{x}) \\ &= \mu_v(\mathbf{y}), \quad \text{car } \int_{\mathbb{X}} M_v(\mathbf{y},d\mathbf{x}) = 1, \quad \forall \mathbf{y} \in \mathbb{X}.\end{aligned}$$

Cela signifie que μ_v est une loi stationnaire pour la chaîne de Markov de noyau M_v.

Annexe

Annexe A

Introduction aux mesures de risque et aux ordres stochastiques

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. On note $\mathcal{B}(\mathbb{R})$ la tribu borélienne de $\mathbb{R}$. On rappelle que la fonction de répartition d'une variable aléatoire $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est la fonction F_X qui caractérise la loi de X et qui est définie pour tout $t \in \mathbb{R}$ par :

$$F_X(t) = \mathbb{P}(X \leq t). \tag{A.1}$$

On rappelle également que la fonction quantile de X est la fonction F_X^{-1} définie $\forall \alpha \in [0,1]$ par :

$$F_X^{-1}(\alpha) = \inf\{t \in \mathbb{R} : F_X(t) \geq \alpha\}. \tag{A.2}$$

Le nombre $F_X^{-1}(\alpha)$ est appelé quantile de X de niveau α. Les fonctions F_X et F_X^{-1} vérifient la propriété d'équivalence suivante (pour une démonstration, voir par exemple (Shorack, 2000)) :

$$F_X(t) \geq \alpha \Leftrightarrow t \geq F_X^{-1}(\alpha), \quad \forall t \in \mathbb{R} \text{ et } \forall \alpha \in [0,1]. \tag{A.3}$$

A.1 Mesures de risque

Une mesure de risque est généralement utilisée pour la prise de décision et le terme « risque » fait référence à une variable aléatoire modélisant l'incertitude que l'on a sur une quantité d'intérêt.

A.1.1 Définition et propriétés

La définition d'une mesure de risque, telle qu'elle est donnée dans (Tasche, 2002), est la suivante :

Définition A.1.1. *Soit V un ensemble non vide de variables aléatoires $\mathcal{F}$-mesurables et à valeurs dans $\mathbb{R}$. On appelle mesure de risque toute application $\rho : V \to \mathbb{R}$.*

Les propriétés basiques des mesures de risque sont données ci-dessous :

Définition A.1.2. *Soit $(X, Y) \in V^2$. Une mesure de risque ρ est dite :*

(i) Invariante en loi si $X \overset{\mathcal{L}}{=} Y \Rightarrow \rho[X] = \rho[Y]$,

(ii) Monotone si $\forall \omega \in \Omega$, $X(\omega) \leq Y(\omega) \Rightarrow \rho[X] \leq \rho[Y]$,

(iii) Invariante par translation si tout pour $\lambda \in \mathbb{R}$, alors $\rho[X + \lambda] = \rho[X] + \lambda$,

(iv) Homogène si pour tout $\lambda > 0$, alors $\rho[\lambda X] = \lambda \rho[X]$,

(v) Sous-additive si $\rho[X + Y] \leq \rho[X] + \rho[Y]$,

(vi) Convexe si pour tout $\lambda \in [0, 1]$, alors $\rho[\lambda X + (1 - \lambda)Y] \leq \lambda \rho[X] + (1 - \lambda)\rho[Y]$.

Une mesure de risque est dite cohérente si elle vérifie les propriétés *(ii)*, *(iii)*, *(iv)* et *(v)*. La notion de cohérence a été introduite et justifiée à travers divers exemples dans (Artzner et al., 1997) et (Artzner et al., 1999). Les auteurs montrent qu'une mesure de risque cohérente quantifie correctement, par exemple, les risques de type marché (risque d'inflation), crédit (risque pour un emprunteur de ne satisfaire ses engagements) ou encore opérationnel (risque de défaillance d'un produit).

A.1.2 Value-at-Risk

Un exemple typique de mesure de risque couramment utilisée dans la gestion des risques financiers est la « Value-at-Risk » (voir (Choudhry, 2013) ou encore le chapitre 7 de (Ruschendorf, 2013)).

Définition A.1.3. *On appelle Value-at-Risk de X au niveau $\alpha \in [0, 1]$ la mesure de risque notée $\mathrm{VaR}_\alpha[X]$ et définie par :*

$$\mathrm{VaR}_\alpha[X] = F_X^{-1}(\alpha),$$

où $F_X^{-1}(\alpha)$ est le quantile de niveau α de X défini en (A.2).

Proposition A.1.1. *La mesure de risque Value-at-Risk est invariante par translation, homogène, monotone et invariante en loi.*

Démonstration. Voir, par exemple, la preuve de la proposition 2.3 de (Tasche, 2002) ou encore celle de la proposition 7.4 de (Ruschendorf, 2013). □

Proposition A.1.2. *Pour tout $\alpha \in [0, 1]$, si φ est une fonction strictement croissante et continue à gauche, alors :*

$$\mathrm{VaR}_\alpha[\varphi(X)] = \varphi\big(\mathrm{VaR}_\alpha[X]\big).$$

Démonstration. Voir, par exemple, la preuve du théorème 1 de (Dhaene et al., 2002). □

Puisque la Value-at-Risk ne vérifie pas la propriété de sous-additivité, il ne s'agit pas d'une mesure de risque cohérente (pour une vue d'ensemble des critiques faites à la Value-at-Risk en tant que mesure de risque, on pourra consulter (Embrechts, 2000) ou encore l'exemple 2.4 de (Tasche, 2002)). Une mesure de risque qui lui est alors généralement préférée est la « Tail-Value-at-Risk ».

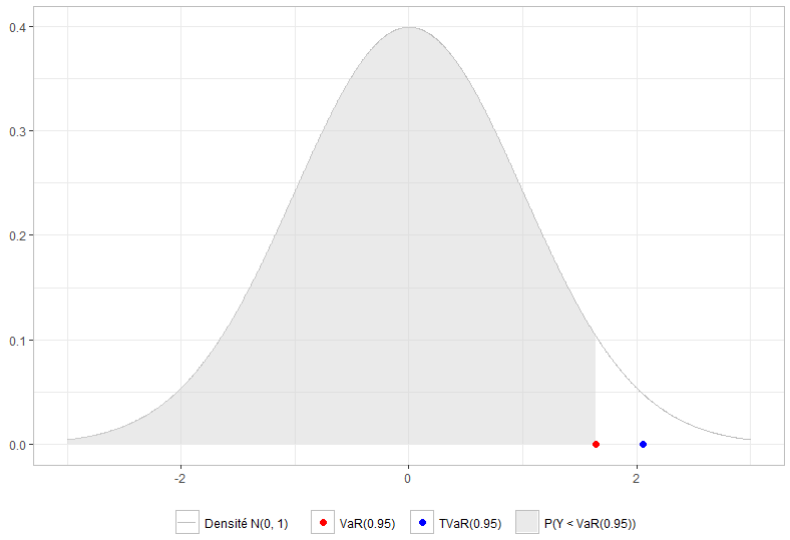

FIGURE A.1 – Valeurs de $\mathrm{VaR}_{0.95}[X]$ (point rouge) et $\mathrm{TVaR}_{0.95}[X]$ (point bleu) pour X distribuée suivant une loi normale centrée et réduite.

A.1.3 Tail-Value-at-Risk

Définition A.1.4. *On appelle Tail-Value-at-Risk de X de niveau $\alpha \in [0,1)$ la mesure de risque notée $\mathrm{TVaR}_\alpha[X]$ et définie par :*

$$\mathrm{TVaR}_\alpha[X] = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_t[X]dt = \frac{1}{1-\alpha}\int_\alpha^1 F_X^{-1}(t)dt. \tag{A.4}$$

On peut remarquer que, si $\mathbb{E}[X]$ désigne l'espérance de X, alors on a $\mathrm{TVaR}_0[X] = \mathbb{E}[X]$. En outre, puisque la fonction $t \in [0,1] \longmapsto \mathrm{VaR}_t[X]$ est croissante, alors pour tout $\alpha \in [0,1)$ fixé, $\mathrm{TVaR}_\alpha[X]$ est un majorant de $\mathrm{VaR}_\alpha[X]$:

$$\mathrm{VaR}_\alpha[X] = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_\alpha[X]dt \leq \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_t[X]dt = \mathrm{TVaR}_\alpha[X].$$

Nous illustrons cette inégalité dans la figure A.1. Les quantités $\mathrm{VaR}_{0.95}[X]$ et $\mathrm{TVaR}_{0.95}[X]$ sont représentées pour X distribuée suivant une loi normale centrée et réduite.

Les propriétés de la Tail-Value-at-Risk sont discutées en détails dans (Rockafellar and Uryasev, 2001), (Acerbi and Tasche, 2002) et (Tasche, 2002). Cette mesure de risque vérifie les propriétés d'homogénéité, de monotonie, d'invariance par translation et d'invariance en loi. De plus, elle est cohérente (voir l'annexe A de (Acerbi and Tasche, 2002) pour une démonstration de la propriété de sous-additivité, ainsi que la section 2.4 de (Denuit et al., 2005) ou encore le chapitre 7 de (Ruschendorf, 2013)). Plus précisement, il s'agit de la plus petite mesure de risque cohérente et invariante en loi qui majore la Value-at-Risk (voir le théorème 9 de (Kusuoka, 2001)).

Par ailleurs, comme mentionné dans (Brazauskas et al., 2008), lorsque la fonction de ré-

partition de X est continue, on a :

$$\mathrm{TVaR}_\alpha[X] = \mathbb{E}\big[X|X > \mathrm{VaR}_\alpha[X]\big], \tag{A.5}$$

où $\mathbb{E}\big[X|X > \mathrm{VaR}_\alpha[X]\big]$ est une mesure de risque de niveau $\alpha \in [0,1)$ appelée « Conditional-Tail-Expectation », ou « Conditional-Value-at-Risk », ou encore « Average-Value-at-Risk ». Pour démontrer l'égalité (A.5), rappelons que si U est une variable aléatoire uniforme sur $[0,1]$, alors les variables aléatoires X et $F_X^{-1}(U)$ ont la même loi (cela se vérifie facilement par application de la propriété d'équivalence (A.3)). De plus, si F_X est continue, alors on a $F_X \circ F_X^{-1} = Id$ et, dans ce cas, la variable aléatoire $F_X(X)$ est distribuée uniformément sur $[0,1]$. Ainsi,

$$\begin{aligned}
\mathbb{E}\big[X|X > \mathrm{VaR}_\alpha[X]\big] &= \mathbb{E}[X|X > F_X^{-1}(\alpha)] \\
&= \mathbb{E}\big[X\mathbb{1}_{X>F_X^{-1}(\alpha)}\big] \times \frac{1}{\mathbb{P}\big(X > F_X^{-1}(\alpha)\big)} \\
&= \mathbb{E}\big[X\mathbb{1}_{F_X(X)>\alpha}\big] \times \frac{1}{1-\mathbb{P}\big(X \leq F_X^{-1}(\alpha)\big)} \\
&= \mathbb{E}\big[F_X^{-1}(U)\mathbb{1}_{U>\alpha}\big] \times \frac{1}{1-\alpha} \\
&= \frac{1}{1-\alpha}\int_\alpha^1 F_X^{-1}(u)du.
\end{aligned}$$

Considérons à présent la mesure de risque suivante :

$$\alpha \in (0,1] \mapsto \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_t[X]dt. \tag{A.6}$$

Pour $\alpha = 1$, il s'agit de l'espérance de X, et, pour $\alpha \in (0,1)$, il s'agit d'un minorant de $\mathrm{VaR}_\alpha[X]$. En effet :

$$\mathrm{VaR}_\alpha[X] = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_\alpha[X]dt \geq \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_t[X]dt.$$

À nouveau, si la distribution de X est continue, on peut montrer que :

$$\frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_t[X]dt = \mathbb{E}\big[X|X \leq \mathrm{VaR}_\alpha[X]\big].$$

Pour distinguer ces deux mesures de risque, on pourra adopter les notations suivantes :

$$\mathrm{TVaR}_\alpha^+[X] = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_\alpha[X]dt, \quad \text{et} \quad \mathrm{TVaR}_\alpha^-[X] = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_\alpha[X]dt, \quad \forall \alpha \in (0,1).$$

A.1.4 Mesure de risque spectrale

Définition A.1.5. *Une fonction spectrale est une fonction $\Psi : [0,1] \to \mathbb{R}_+$ croissante, continue à droite et qui vérifie :*

$$\int_0^1 \Psi(t)dt = 1.$$

On appelle mesure de risque spectrale de X induite par Ψ la mesure de risque notée $\rho_\Psi[X]$ et définie par :

$$\rho_\Psi[X] = \int_0^1 \mathrm{VaR}_t[X]\Psi(t)dt = \int_0^1 F_X^{-1}(t)\Psi(t)dt.$$

Une mesure de risque spectrale est donc une mesure de risque basée sur une moyenne pondérée de la Value-at-Risk. La fonction Ψ est connue sous le nom de « risk adverse function » et traduit, par exemple, l'attitude subjective d'un agent face au risque. L'espérance (avec $\Psi(t) = 1$) et la Tail-Value-at-Risk de niveau α (avec $\Psi(t) = (1-\alpha)^{-1}\mathbb{1}_{t>\alpha}$) sont des cas particuliers de mesures de risque spectrale. Une mesure de risque spectrale est toujours cohérente (voir (Acerbi, 2002)) et peut toujours s'exprimer comme une moyenne pondérée de la Tail-Value-at-Risk (voir annexe C de (Adam et al., 2008)).

A.1.5 Remarques générales

Bien qu'il ne soit pas utile de les présenter ici, on insiste sur le fait qu'il existe d'autres mesures de risque couramment utilisées dans les domaines de l'assurance, de la gestion du risque ou encore de la finance (voir (Denuit et al., 2005) pour une liste exhaustive des mesures de risque usuelles et quelques exemples d'application).

Les mesures de risque sont des quantités qui ne sont pas toujours faciles à calculer. Pour une vue d'ensemble des méthodes d'estimation de la Value-at-Risk et de la Tail-Value-at-Risk, on pourra par exemple consulter (Choudhry, 2013), (Davis, 2013) et (Pitera and Schmidt, 2018), ainsi que les références qui s'y trouvent. Généralement, l'approche proposée est non-paramétrique : on simule un échantillon i.i.d. suivant la loi de X et on considère un estimateur empirique de la mesure de risque (voir (Brazauskas et al., 2008) pour l'estimation de la Conditional-Value-at-Risk et le chapitre 7 de (J. Shervish, 2010) pour l'estimation de la Value-at-Risk).

A.2 Ordres stochastiques

Dans de nombreuses situations, la comparaison de deux variables aléatoires X et Y à travers l'analyse de leur espérance, de leur médiane ou encore de leur écart-type s'avère peu efficace. L'objectif des ordres stochastiques est d'apporter des outils de comparaison alternatifs et informatifs pour aider à la prise de décision. Par exemple, l'ordre stochastique usuel (présenté en section A.2.1) s'appuie sur la Value-at-Risk pour comparer les valeurs prises par les variables aléatoires X et Y, tandis que l'ordre stochastique convexe (présenté en section A.2.3) se base sur la Tail-Value-at-Risk.

Les ordres stochastiques sont couramment utilisés en analyse de survie, en fiabilité (par exemple, pour confronter l'efficacité de plusieurs médicaments à travers une analyse de la durée de vie des patients) ou encore en finance et en gestion des risques (comparaison des pertes et/ou des profits pour le choix de stratégies d'investissement). Pour une vue d'ensemble sur les ordres stochastiques, on pourra consulter les livres (Shaked and Shanthikumar, 2007) et (Belzunce et al., 2015), qui sont fréquement cités dans ce chapitre. Des

exemples d'application divers sont donnés dans (Müller and Stoyan, 2002), mais aussi dans (Chan et al., 1990) et (Ahmed et al., 1996) pour la fiabilité, et dans (Dhaene et al., 2001) et (Dhaene et al., 2002) pour l'actuariat.

Puisque ce sont les seuls que nous utilisons dans cette thèse, on présente uniquement l'ordre stochastique usuel (appelé aussi dominance stochastique à l'ordre 1), l'ordre convexe croissant (appelé aussi dominance stochastique à l'ordre 2) et l'ordre convexe.

A.2.1 Ordre stochastique usuel

On rappelle que la fonction de survie d'une variable aléatoire X est la fonction G_X définie pour tout $t \in \mathbb{R}$ par :

$$G_X(t) = \mathbb{P}(X > t) = 1 - F_X(t),$$

où F_X est la fonction de répartition de X définie à l'équation (A.1).

Définition A.2.1. *On dit que X est plus petite que Y au sens de l'ordre stochastique usuel (ou encore, que Y domine stochastiquement X à l'ordre 1), et on note $X \leq_{st} Y$, si et seulement si pour tout $t \in \mathbb{R}$,*

$$G_X(t) \leq G_Y(t).$$

Une remarque immédiate est que : $X \leq_{st} Y \Rightarrow \mathbb{E}[X] \leq \mathbb{E}[Y]$. Comme en témoigne le résultat suivant, il existe par ailleurs un bon nombre d'inégalités intéressantes équivalentes à $X \leq_{st} Y$:

Proposition A.2.1. *Les propriétés ci-dessous sont équivalentes :*

(i) $X \leq_{st} Y$,

(ii) $G_X(t) \leq G_Y(t), \quad \forall t \in \mathbb{R}$,

(iii) $F_X(t) \geq F_Y(t), \quad \forall t \in \mathbb{R}$,

(iv) $F_X^{-1}(\alpha) \leq F_Y^{-1}(\alpha), \quad \forall \alpha \in (0,1)$,

(v) $\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(Y)]$ *pour toute fonction croissante* $\varphi : \mathbb{R} \to \mathbb{R}$ *telle que les espérances existent.*

Lorsque la distribution de X est inconnue[1], utiliser à la place une variable aléatoire Y telle que $X \leq_{st} Y$ permet d'apprendre sur la loi de X. Cependant, dans certaines situations, l'ordre stochastique usuel n'est pas assez informatif. Pour illustrer cela, nous comparons en figure A.2 les fonctions de survie et les fonctions quantiles de deux variables aléatoires X et Y telles que $X \backsim \text{Beta}\left(\frac{1}{2}, 2\right)$ et $Y \backsim \text{Beta}(10, 3)$ en haut (respectivement $Y \backsim \text{Beta}\left(\frac{4}{5}, 2\right)$ en bas). Dans les deux cas, on a $X \leq_{st} Y$ et les propriétés *(ii)*, *(iv)* et *(v)* (avec $\varphi = Id$) de la proposition A.2.1 sont vérifiées. Dans la première situation, on a une assez mauvaise approximation des différents aspects (moyenne, quantiles, ...) de la loi de X par ceux de Y. Dans la seconde, en revanche, on est plutôt bien informé.

1. En pratique, il se peut, en effet, que la distribution de la variable aléatoire d'intérêt soit inconnue. Par exemple, dans un cadre bayésien, le calcul de la distribution a posteriori peut être hors de portée.

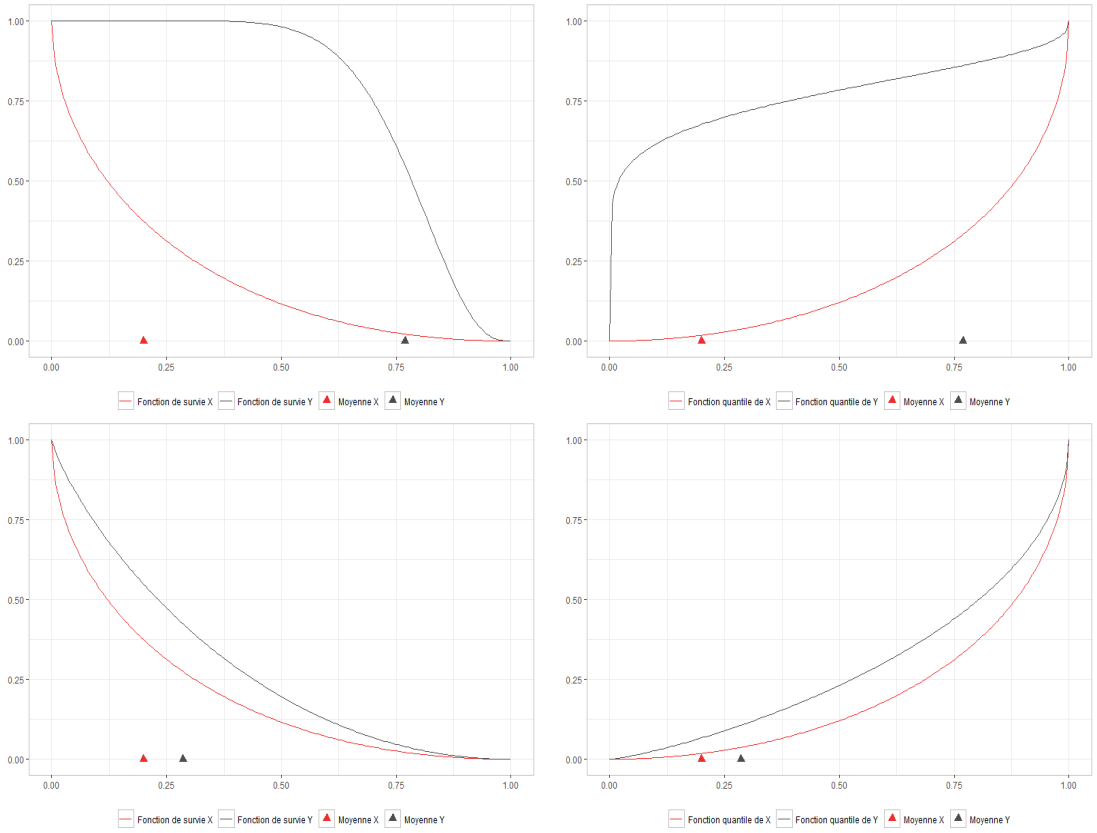

FIGURE A.2 – Illustrations des inégalités *(ii)*, *(iv)* et *(v)* (avec $\varphi = Id$) de la proposition A.2.1, avec $X \sim \text{Beta}\left(\frac{1}{2}, 2\right)$ et $Y \sim \text{Beta}(10, 3)$ en haut (resp. $Y \sim \text{Beta}\left(\frac{4}{5}, 2\right)$ en bas). À gauche : les fonctions de survie de X et Y. À droite : les fonctions quantile de X et Y.

En économie et actuariat, on rencontre plutôt la notation $\leq_{VaR}$, en référence à la Value-at-Risk qui intervient dans la propriété *(iv)* de la proposition A.2.1. En effet, cette propriété s'écrit encore : $\text{VaR}_\alpha[X] \leq \text{VaR}_\alpha[Y]$, $\forall \alpha \in (0,1)$. Dans le théorème suivant, on montre que cette inégalité se généralise à une certaine classe de mesures de risque (pour une démonstration, voir le chapitre 3 de (Denuit et al., 2005) ou la section 4 de (Bäuerle and Müller, 2006)).

Théorème A.2.1. *Soit X et Y deux variables aléatoires de lois continues et ρ une mesure de risque monotone et invariante en loi. Alors,*

$$X \leq_{st} Y \Rightarrow \rho[X] \leq \rho[Y].$$

A.2.2 Ordre stochastique convexe croissant

Définition A.2.2. *On dit que X est plus petite que Y au sens de l'ordre convexe croissant (ou encore que Y domine stochastiquement X à l'ordre 2), et on note $X \leq_{icx} Y$, si et*

seulement si pour toute fonction convexe et croissante $\varphi : \mathbb{R} \to \mathbb{R}$,

$$E[\varphi(X)] \leq \mathbb{E}[\varphi(Y)], \tag{A.7}$$

telle que les espérances existent.

On a $X \leq_{st} Y \Rightarrow X \leq_{icx} Y$ et $X \leq_{icx} Y \Rightarrow \mathbb{E}[X] \leq \mathbb{E}[Y]$. La notation $\leq_{icx}$ fait référence au nom anglais : « increasing convex order ». En science actuarielle, on parle plutôt de « stop-loss order » et on utilise la notation $\leq_{sl}$, car l'inégalité (A.7) est équivalente à :

$$\mathbb{E}[(X-t)_+] \leq \mathbb{E}[(Y-t)_+], \quad \forall t \in \mathbb{R}, \tag{A.8}$$

où la fonction :

$$t \in \mathbb{R} \mapsto \mathbb{E}[(X-t)_+] = \mathbb{E}[\max(0, X-t)] = \int_t^{+\infty} G_X(u)du,$$

s'appelle « stop-loss function » (voir (Kaas et al., 1994), (Müller, 1996), (Dhaene et al., 2002), la section 3.4 de (Denuit et al., 2005) et (Bäuerle and Müller, 2006)).

Proposition A.2.2. *Les propriétés ci-dessous sont équivalentes :*

(i) $X \leq_{icx} Y$,

(ii) $\mathbb{E}[(X-t)_+] \leq \mathbb{E}[(Y-t)_+], \quad \forall t \in \mathbb{R}$,

(iii) $\displaystyle\int_\alpha^1 F_X^{-1}(t)dt \leq \int_\alpha^1 F_Y^{-1}(t)dt, \quad \forall \alpha \in (0,1)$.

La propriété *(ii)* s'écrit encore :

$$\int_t^{+\infty} G_X(u)du \leq \int_t^{+\infty} G_Y(u)du,$$

et la propriété *(iii)* est équivalente à $\mathrm{TVaR}_\alpha[X] \leq \mathrm{TVaR}_\alpha[Y]$.

A.2.3 Ordre convexe

Définition A.2.3. *On dit que* X *est plus petite que* Y *au sens de l'ordre convexe, et on note* $X \leq_{cx} Y$*, si et seulement si pour toute fonction convexe* $\varphi : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(Y)], \tag{A.9}$$

telle que les espérances existent.

Puisque $X \leq_{cx} Y \Rightarrow \mathbb{E}[X] = \mathbb{E}[Y]$ (avec $\varphi(t) = t$ et $\varphi(t) = -t$), l'ordre convexe compare uniquement les variables aléatoires de même valeur moyenne. De plus, puisque l'on a $X \leq_{cx} Y \Rightarrow \mathrm{Var}[X] \leq \mathrm{Var}[Y]$ (avec $\varphi(t) = t^2$), cet ordre stochastique est souvent utilisé pour comparer les variables aléatoires à travers leur dispersion. Dans le chapitre 3 de (Shaked and Shanthikumar, 2007), il est donné les conditions d'équivalence suivantes :

Proposition A.2.3. *Soient X et Y deux variables aléatoires telles que $\mathbb{E}[X] = \mathbb{E}[Y]$. Alors, $X \leq_{cx} Y$ si et seulement si :*

$$(i)\ \mathbb{E}[(X-t)_+] \leq \mathbb{E}[(Y-t)_+],\ ou\ \int_t^{+\infty} G_X(u)du \leq \int_t^{+\infty} G_Y(u)du, \quad \forall t \in \mathbb{R},$$

$$(ii)\ \mathbb{E}[(t-X)_+] \leq \mathbb{E}[(t-Y)_+],\ ou\ \int_{-\infty}^{t} F_X(u)du \leq \int_{-\infty}^{t} F_Y(u)du, \quad \forall t \in \mathbb{R},$$

$$(iii)\ \int_0^{\alpha} F_X^{-1}(t)dt \geq \int_0^{\alpha} F_Y^{-1}(t)dt,\ ou\ \mathrm{TVaR}_\alpha^-[X] \geq \mathrm{TVaR}_\alpha^-[Y], \quad \forall \alpha \in (0,1),$$

$$(iv)\ \int_\alpha^1 F_X^{-1}(t)dt \leq \int_\alpha^1 F_Y^{-1}(t)dt,\ ou\ \mathrm{TVaR}_\alpha^+[X] \leq \mathrm{TVaR}_\alpha^+[Y], \quad \forall \alpha \in (0,1).$$

La propriété (i) signifie que si $\mathbb{E}[X] = \mathbb{E}[Y]$, alors $X \leq_{icx} Y \Leftrightarrow X \leq_{cx} Y$. En référence à l'inégalité (A.8) caractérisant l'ordre convexe croissant et signifiant que les queues de distribution de Y sont uniformément plus grandes que celles de X, on dit alors que Y est un « étalement de X à moyenne constante » (ou encore « mean preserving spread » en anglais). Par ailleurs, en lien avec le « stop-loss order » présenté en section A.2.2, on rencontre aussi parfois la notation $\leq_{sl,=}$ pour désigner l'ordre convexe. Alternativement, il existe un ordre stochastique concave, noté $\leq_{cv}$, et défini par :

$$X \leq_{cv} Y \Leftrightarrow Y \leq_{cx} X.$$

À la figure A.3, on illustre les propriétés de la proposition A.2.3 pour $X \backsim \mathrm{Beta}(10,5)$ et $Y \backsim \mathrm{Beta}(4,2)$. Il existe en effet un ordre convexe entre ces deux variables aléatoires ($X \leq_{cx} Y$) car, d'après le théorème 4.6 de (Malrieu and Zitt, 2017), si $X \backsim \mathrm{Beta}(\alpha,\beta)$ et $X' \backsim \mathrm{Beta}(\alpha',\beta')$, et si $\alpha < \alpha'$, $\beta < \beta'$ et $\alpha/(\alpha+\beta) = \alpha'/(\alpha'+\beta')$, alors $X' \leq_{cx} X$. De fait, les valeurs moyennes de X et Y sont égales. En haut, les inégalités (i) et (ii) sont vérifiées, pour $t = 0.75$ (à gauche) et $t = 0.5$ (à droite). En bas à gauche, on compare les densités de X et de Y, et il apparaît que celle de Y est plus étalée. On vérifie aussi les inégalités (iii) et (iv) à travers la représentation des quantités $\mathrm{TVaR}_\alpha^-[X]$, $\mathrm{TVaR}_\alpha^-[Y]$, $\mathrm{TVaR}_\alpha^+[X]$ et $\mathrm{TVaR}_\alpha^+[Y]$ pour $\alpha = 0.75$. Enfin, en bas à droite, l'inégalité (iv) est à nouveau vérifiée pour cette même valeur de α.

On peut remarquer que si φ_1 une fonction convexe et φ_2 une fonction convexe croissante, alors la fonction composée $\varphi_2 \circ \varphi_1$ est encore convexe et, on a :

$$X \leq_{cx} Y \Leftrightarrow \mathbb{E}\big[\varphi_2 \circ \varphi_1(X)\big] \leq \mathbb{E}\big[\varphi_2 \circ \varphi_1(Y)\big] \Leftrightarrow \varphi_1(X) \leq_{icx} \varphi_1(Y).$$

Par ailleurs, dans (Boutsikas and Vaggelatou, 2002), les auteurs étudient les distances entre les distributions de variables aléatoires vérifiant des relations de dominance stochastique. Un des résultats qu'ils ont obtenus est le suivant : si F_X et F_Y sont les fonctions de répartition de X et Y, alors la distance de Wasserstein entre F_X et F_Y vérifie :

$$\begin{aligned}\int_0^1 \big|F_X(t) - F_Y(t)\big|dt &= \sup_{t\in[0,1]} \big|\mathbb{E}[(X-t)_+] - \mathbb{E}[(Y-t)_+]\big| \\ &\leq \sqrt{\frac{1}{2}\big(\mathbb{E}[Y^2] - \mathbb{E}[X]^2\big)} = \sqrt{\frac{1}{2}\big(\mathrm{Var}[X] - \mathrm{Var}[Y]\big)}.\end{aligned}$$

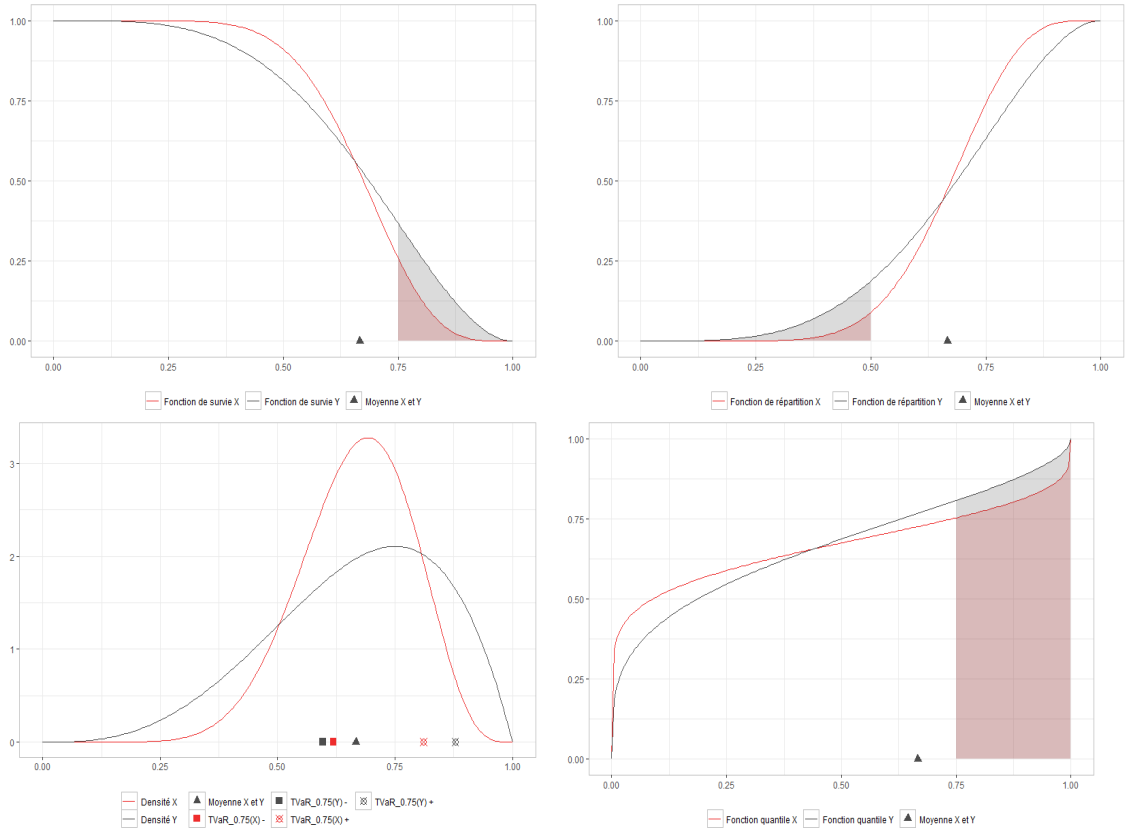

FIGURE A.3 – Illustrations de la proposition A.2.3 pour $X \sim \text{Beta}(10, 5)$ et $Y \sim \text{Beta}(4, 2)$. En haut à gauche : fonctions de survie (en référence à l'inégalité (i)). En haut à droite : fonctions de répartition (inégalité (ii)). En bas à gauche : densités (inégalités (iii) et (iv)). En bas à droite : fonctions quantiles (inégalité (iv)).

Pour plus de détails sur la distance de Wasserstein, on pourra par exemple consulter le chapitre 17 de (Chafaï and Malrieu, 2016).

Enfin, on précise que dans (Bäuerle and Müller, 2006) et (Ruschendorf, 2013), on peut trouver des résultats généraux sur les ordres stochastiques et les mesures de risque. Typiquement, il s'agit de déterminer, pour une relation d'ordre stochastique donnée entre Y et X, les conditions nécessaires que doit vérifier une mesure de risque ρ pour satisfaire $\rho[X] \leq \rho[Y]$.

A.3 Inégalités de Markov et Bienaymé-Tchebychev

Dans le théorème suivant, on rappelle comment s'écrivent les inégalités de Markov et Bienaymé-Tchebychev pour une variable aléatoire à valeurs dans $[0,1]$. Pour une généralisation, on pourra, par exemple, consulter la section 3 de (Dufour, 2003), et les références qui s'y trouvent. On rappelle que l'inégalité de Bienaymé-Tchebychev est un cas particulier de l'inégalité de Markov.

Théorème A.3.1. *Soit X une variable aléatoire à valeurs dans $[0,1]$. Alors, pour tout $t > 0$, on a :*

(i) (Inégalité de Markov)

$$\mathbb{E}[X^m] - t^m \leq \mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X^m]}{t^m}, \quad \text{où } m \in \mathbb{N}^*. \tag{A.10}$$

(ii) (Inégalité de Bienaymé-Tchebychev)

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}[X]}{t^2}. \tag{A.11}$$

Soit $\alpha \in (0,1)$ et $F_X^{-1}(\alpha)$ le quantile de X de niveau α. D'après la définition de la fonction quantile (A.2), il découle respectivement de (A.10) et de (A.11) que $F_X^{-1}(\alpha)$ vérifie :

$$\left(\mathbb{E}[X^m] + \alpha - 1\right)^{\frac{1}{m}} \leq F_X^{-1}(\alpha) \leq \left(\frac{\mathbb{E}[X^m]}{1-\alpha}\right)^{\frac{1}{m}}, \tag{A.12}$$

et

$$F_X^{-1}(\alpha) \leq \mathbb{E}[X] + \sqrt{\frac{\mathrm{Var}[X]}{1-\alpha}}. \tag{A.13}$$

Pour $m \in \{1,2,3\}$, on compare dans la figure A.4 (à gauche) la qualité d'approximation des bornes (A.10) pour l'encadrement de la fonction de répartition d'une loi normale $\mathcal{N}\left(\frac{1}{2}, \frac{1}{20}\right)$. À droite, il s'agit des bornes (A.12) pour l'encadrement de la fonction quantile de cette même loi normale.

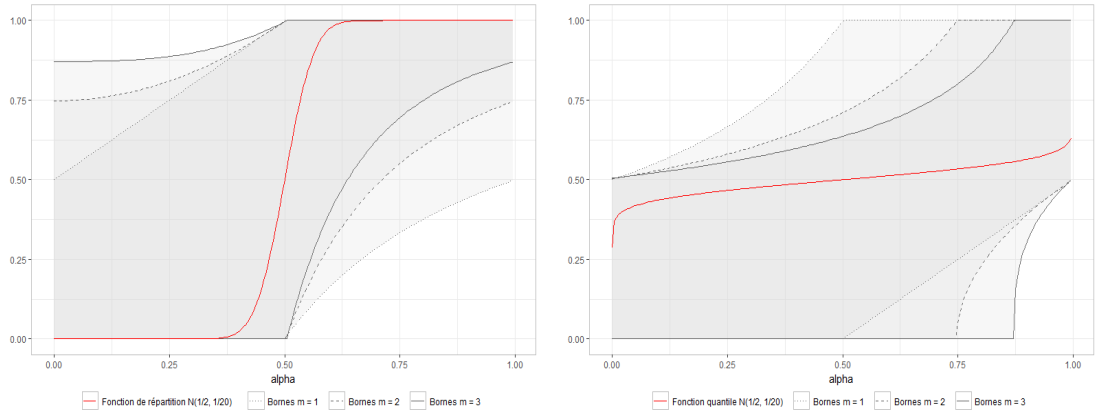

FIGURE A.4 – À gauche : encadrement de la fonction de répartition (courbe rouge) d'une loi normale $\mathcal{N}(\frac{1}{2}, \frac{1}{20})$, par application de (A.10), pour $m \in \{1, 2, 3\}$. À droite : encadrement de la fonction quantile (courbe rouge) d'une loi normale $\mathcal{N}\left(\frac{1}{2}, \frac{1}{20}\right)$, par application de (A.12), pour $m \in \{1, 2, 3\}$.

Annexe B

Mise en œuvre pratique

Dans le chapitre 3, on s'intéresse à l'estimation de la probabilité de défaillance :

$$p = \int_{\mathbb{X}} \mathbb{1}_{g(\mathbf{x})>T} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}),$$

où $g : \mathbb{X} \subseteq \mathbb{R}^d \to \mathbb{R}$ est une fonction boîte-noire coûteuse à évaluer, $T \in \mathbb{R}$ est un seuil et $\mathbf{X}$ est une variable aléatoire à valeurs dans $\mathbb{X}$. La loi de $\mathbf{X}$ est notée $\mathbf{P}_{\mathbf{X}}$ et on suppose savoir simuler suivant cette loi. Par conséquent, on dispose ici d'une suite de variables aléatoires $(\mathbf{X}_i)_{1\leq i\leq N}$ i.i.d. suivant $\mathbf{P}_{\mathbf{X}}$.

On se place dans le cadre suivant : la fonction g est observée aux points d'un plan d'expériences $(\mathbf{x}_i)_{1\leq i\leq n} \in \mathbb{X}^n$ et, au vu de ces observations, on modélise g par un processus aléatoire $(\xi_n(\mathbf{x}))_{\mathbf{x}\in\mathbb{X}}$. On pose :

$$m_n(\mathbf{x}) = \mathbb{E}[\xi_n(\mathbf{x})], \quad \text{et} \quad \sigma_n(\mathbf{x})^2 = \mathrm{Var}[\xi_n(\mathbf{x})], \quad \forall \mathbf{x} \in \mathbb{X},$$

et

$$p_n(\mathbf{x}) = \mathbb{P}(\xi_n(\mathbf{x}) > T).$$

Dans le cas où le processus ξ_n est gaussien, on a par exemple :

$$p_n(\mathbf{x}) = \Phi\left(\frac{m_n(\mathbf{x}) - T}{\sigma(\mathbf{x})}\right),$$

où Φ est la fonction de répartition de la loi normale centrée réduite. Cette modélisation est possible par application de la méthode de krigeage présentée en section 2.3.3, pour un modèle de processus gaussien où la matrice de covariance des observations est bien inversible. En adoptant un point de vue bayésien, on peut alors faire l'hypothèse que la fonction g est une trajectoire du processus ξ_n et donc que la probabilité p est une réalisation de la variable aléatoire $S_n \in [0, 1]$ définie par :

$$S_n = \int_{\mathbb{X}} \mathbb{1}_{\xi_n(\mathbf{x})>T} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}).$$

Compte tenu du fait qu'il est difficile de déterminer explicitement la loi de S_n – et qu'il est numériquement coûteux de générer des réalisations de cette variable aléatoire – on propose de considérer la variable aléatoire alternative R_n définie par :

$$R_n = \int_{\mathbb{X}} \mathbb{1}_{p_n(\mathbf{x})>U} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}),$$

où U est une variable aléatoire uniforme sur $[0,1]$. Ce choix de loi pour U garantit que l'on a l'égalité des moyennes, i.e. $\mathbb{E}[S_n] = \mathbb{E}[R_n]$ (voir (Oger et al., 2015)). Ces espérances fournissent une estimation de la probabilité de défaillance p.

Dans la suite, on propose deux méthodes pour calculer numériquement les quantités suivantes :

▷ L'espérance de R_n qui vérifie :

$$\mathbb{E}[R_n] = \int_{\mathbb{X}} p_n(\mathbf{x}) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}).$$

▷ La variance de R_n qui vérifie :

$$\mathrm{Var}[R_n] = \mathbb{E}[R_n^2] - \mathbb{E}[R_n]^2 = \int_{\mathbb{X}^2} \min\big(p_n(\mathbf{x}_1), p_n(\mathbf{x}_2)\big) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}_1) \mathbf{P}_{\mathbf{X}}(d\mathbf{x}_2) - \mathbb{E}[R_n]^2.$$

▷ La fonction quantile de R_n définie pour tout $\alpha \in [0,1]$ par :

$$F_{R_n}^{-1}(\alpha) = \inf\{t \in [0,1] : \mathbb{P}(R_n \leq t) \geq \alpha\}.$$

D'après la proposition 3.5.1, on a :

$$F_{R_n}^{-1}(\alpha) = \int_{\mathbb{X}} \mathbb{1}_{\alpha>1-p_n(\mathbf{x})} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}).$$

▷ La fonction de survie de R_n définie pour tout $t \in [0,1]$ par :

$$G_{R_n}(t) = \mathbb{P}(R_n > t).$$

On peut vérifier que l'on a :

$$G_{R_n}(t) = \inf\{u \in [0,1] : F_{R_n}^{-1}(1-u) \leq t\}.$$

Méthode 1

Soit $(\mathbf{X}_i)_{1\leq i\leq N}$ un échantillon i.i.d. suivant la loi $\mathbf{P}_{\mathbf{X}}$. Supposons que la fonction p_n soit évaluée en chacun de ces points et que, parmi les N probabilités $p_n(\mathbf{X}_1), \ldots, p_n(\mathbf{X}_N)$, il y en ait N' distinctes. On peut alors les numéroter de 1 à N', puis les classer par ordre croissant :

$$0 \leq p_n^{(1)} \leq \ldots \leq p_n^{(N')} \leq 1,$$

où $p_n^{(1)} = \min\big(p_n(\mathbf{X}_1), \ldots, p_n(\mathbf{X}_N)\big), \ldots, p_n^{(N')} = \max\big(p_n(\mathbf{X}_1), \ldots, p_n(\mathbf{X}_N)\big)$.

Pour $1 \leq i \leq N'$, on note l_i le nombre d'occurrences de la probabilité $p_n^{(i)}$, de sorte que l'égalité suivante soit vérifiée : $\sum_{i=1}^{N'} l_i = N$. On introduit également la notation suivante : $n_i = N - \sum_{j=1}^{i} l_j$, avec $n_{N'} = 0$ et $n_{N'-1} = l_{N'}$.

Calcul de la moyenne

Pour estimer l'espérance de R_n, on fait l'approximation suivante :

$$\mathbb{E}[R_n] \approx \frac{1}{N} \sum_{i=1}^{N'} l_i p_n^{(i)}.$$

On peut remarquer que pour $N' = N$ et $l_i = 1$, avec $1 \leq i \leq N$, on retrouve l'estimateur de la méthode Monte-Carlo naïve.

Calcul de la variance

Pour estimer le moment d'ordre 2 de R_n, on peut faire l'approximation suivante :

$$\begin{aligned}
\mathbb{E}[R_n^2] &\approx \frac{1}{N^2} \sum_{i=1}^{N'} \sum_{j=1}^{N'} l_i l_j \min\left(p_n^{(i)}, p_n^{(j)}\right) \\
&= \frac{1}{N^2} \left(l_1 p_n^{(1)} \left(l_1 + 2 \sum_{i=2}^{N'} l_i \right) + l_2 p_n^{(2)} \left(l_2 + 2 \sum_{i=3}^{N'} l_i \right) + \ldots + \right. \\
&\quad \left. l_{N'-1} p_n^{(N'-1)} \left(l_{N'-1} + 2 l_{N'} \right) + l_{N'} p_n^{(N')} (l_{N'}) \right) \\
&= \frac{1}{N^2} \sum_{i=1}^{N'} l_i p_n^{(i)} (l_i + 2 n_i),
\end{aligned}$$

car $\sum_{j=i+1}^{N'} l_j = n_i$ pour $1 \leq i \leq N' - 1$ et $n_{N'} = 0$. Par conséquent, pour estimer la variance de R_n, on fait l'approximation suivante :

$$\mathrm{Var}[R_n] \approx \frac{1}{N^2} \sum_{i=1}^{N'} l_i p_n^{(i)} (l_i + 2 n_i) - \left(\frac{1}{N} \sum_{i=1}^{N'} l_i p_n^{(i)} \right)^2.$$

À nouveau, on remarque que si $N' = N$ et $l_i = 1$, avec $1 \leq i \leq N$, alors on retrouve l'estimateur de la Monte-Carlo naïve.

Calcul de la fonction quantile

Pour une valeur de $\alpha \in [0, 1]$ fixée, on propose d'estimer le quantile $F_{R_n}^{-1}(\alpha)$ de la façon suivante :

$$F_{R_n}^{-1}(\alpha) \approx \frac{1}{N} \sum_{i=1}^{N'} l_i \mathbb{1}_{\alpha > 1 - p_n^{(i)}}.$$

Calcul de la fonction de survie

Puisque pour tout $\alpha \in [0,1]$, on a :

$$F_{R_n}^{-1}(1-\alpha) \approx \frac{1}{N}\sum_{i=1}^{N'} l_i \mathbb{1}_{p_n^{(i)}>\alpha},$$

alors, pour une valeur de $t \in [0,1]$ fixée, on peut estimer $G_{R_n}(t)$ par :

$$G_{R_n}(t) \approx p_n^{(1)} \mathbb{1}_{\left[\frac{n_1}{N},1\right[}(t) + \sum_{i=2}^{N'} p_n^{(i)} \mathbf{1}_{\left[\frac{n_i}{N},\frac{n_{i-1}}{N}\right[}(t).$$

Méthode 2

La loi de R_n étant inconnue, on peut simuler des réalisations de cette variable aléatoire, et estimer empiriquement les quantités ci-dessus (moyenne, variance, fonction quantile et fonction de répartition), ainsi que les moments de tout ordre.

Soient $(U_j)_{1\leq j\leq M}$ un échantillon i.i.d. suivant une loi uniforme sur $[0,1]$, et $(\mathbf{X}_i)_{1\leq i\leq N}$ un échantillon i.i.d. suivant $\mathbf{P}_{\mathbf{X}}$. Pour obtenir une suite $(R_n^j)_{1\leq j\leq M}$ de variables aléatoires distribuées approximativement suivant la loi de R_n, il suffit de poser :

$$R_n^j = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{p_n(\mathbf{X}_i)>U_j}, \quad \forall j = 1,\dots,M.$$

Exemple

L'efficacité des deux méthodes est complètement indépendante de la dimension d de $\mathbb{X}$. Les temps d'exécution sont équivalents et très rapides. En outre, les résultats qu'elles fournissent sont équivalents. Cela se voit dans l'exemple ci-après, issu de (Oger et al., 2015).

Soit $T > 0$, $a_1,\dots,a_d > T^2$, $\mathbb{X} = [-1,1]^d$ et $\mathbf{X}$ distribué suivant une loi uniforme sur $\mathbb{X}$. Pour tout $\mathbf{x} \in \mathbb{X}$, on pose :

$$g(\mathbf{x}) = \left(\sum_{i=1}^{d} a_i \mathbf{x}_i^2\right)^{\frac{1}{2}}.$$

On peut alors montrer que :

$$p = \mathbb{P}(g(\mathbf{X}) > T) = 1 - \frac{1}{2^d} \times \frac{\pi^{\frac{d}{2}} T^d}{\Gamma\left(\frac{d}{2}+1\right)} \prod_{i=1}^{d} a_i^{-\frac{1}{2}}.$$

On s'intéresse à l'estimation de p dans le cas où $d = 5$, et $a_1 = 4$, $a_2 = 4.2$, $a_3 = 4.4$, $a_4 = 4.6$ et $a_5 = 4.8$. Avec $T = 1.9$, on a $p = 8.99 \times 10^{-1}$. Les résultats que l'on obtient sont donnés dans la figure B.1. On rappelle que la valeur estimée de $\mathbb{E}[R_n]$ fournie une estimation de p.

FIGURE B.1 – Distribution empirique de R_n par application de la méthode 2.

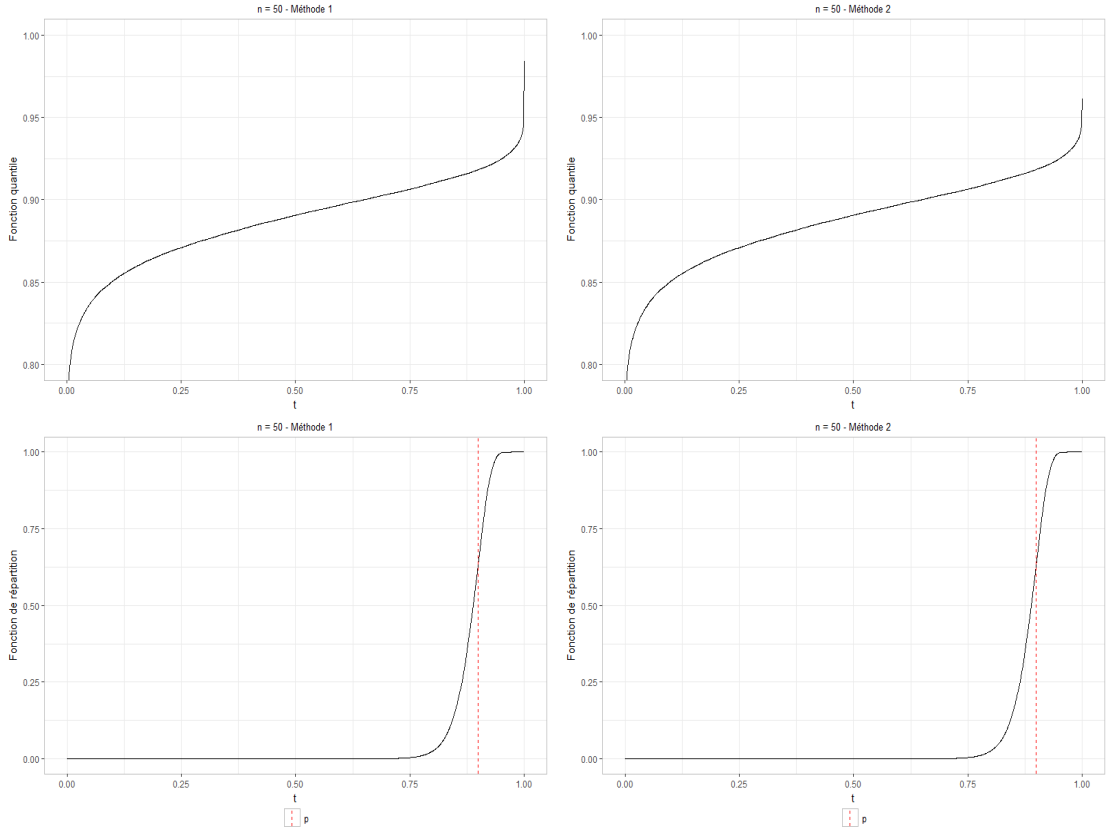

FIGURE B.2 – Droite en pointillés rouge : la vraie probabilité de défaillance p. À gauche : la fonction quantile (en haut) et la fonction de répartition (en bas) de R_n, par application de la méthode 1. À droite : la fonction quantile (en haut) et la fonction de répartition (en bas) R_n par application de la méthode 2.

Bibliographie

Abramowitz, M. and Stegun, I. A. (1965). *Handbook of Mathematical Functions*. Dover, New York.

Acerbi, C. (2002). Spectral measure of risk : a coherent representation of subjective risk aversion. *Journal of Banking and Finance*, 26(7) :1505–1518.

Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance*, 26 :1487–1503.

Adam, A., Houkari, M., and Laurent, J.-P. (2008). Spectral risk measures and portfolio selection. *Journal of Banking and Finance*, 9(32) :1870–1882.

Adler, R. J. (1981). *The Geometry of random fields*. John Wiley & Sons Inc.

Ahmed, A., Soliman, A., and Khider, S. (1996). On some partial ordering of interest in reliability. *Microelectronics Reliability*, (36) :1337–1346.

Ang, G., Ang, A.-S., and Tang, W. (1990). Kernel method in importance sampling density estimation. *Structural Safety and Reliability*, pages 1193–1200.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1997). Thinking coherently. *Risk*, 10(11) :68–71.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3) :203–228.

Asmussen, S. and Glynn, P. W. (2007). *Stochastic simulation : algorithms and analysis*. Springer.

Au, S. and Beck, J. (1999). A new adaptive importance sampling scheme for reliability calculations. *Structural Safety*, 21 :135–158.

Au, S. and Beck, J. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Journal of Probabilistic Engineering Mechanics*, 16(4) :263–277.

Auffray, Y., Barbillon, P., and Marin, J.-M. (2014). Bounding rare event probabilities in computer experiments. *Computational Statistics and Data Analysis*, 80 :153–166.

Azzimonti, D. (2016). *Contributions to Bayesian set estimation relying on random field priors*. PhD thesis, University of Bern.

Azzimonti, D., Ginsbourger, D., Chevalier, C., Bect, J., and Richet, Y. (2018). Adaptive Design of Experiments for Conservative Estimation of Excursion Sets. working paper or preprint.

Bachoc, F. (2013a). Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model misspecification. *Computational Statistics & Data Analysis*, 66 :55–69.

Bachoc, F. (2013b). *Estimation paramétrique de la fonction de covariance dans le modèle de krigeage par processus gaussiens : application à la quantification des incertitudes en simulation numérique.* PhD thesis, Université Paris-Diderot Paris VII.

Barbillon, P. (2010). *Méthodes d'interpolation à noyaux pour l'approximation de fonctions type boîte noire coûteuses.* PhD thesis, Université Paris-sud 11.

Bäuerle, N. and Müller, A. (2006). Stochastic orders and risk measures : consistency and bounds. *Insurance Mathematics and Economics*, 38(1) :132 :148.

Bect, J., Bachoc, F., and Ginsbourger, D. (2019). A supermartingale approach to Gaussian process based sequential design of experiments. *Bernoulli*.

Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3) :773–793.

Bect, J., Li, L., and Vazquez, E. (2017). Bayesian subset simulation. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1) :762–786.

Belyaev, Y. (1961). Continuity and Hölder's conditions for sample functions of stationary Gaussian processes. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, 2 :627–632.

Belzunce, F., Martinez-Riquelme, C., and Mulero, J. (2015). *An introduction to stochastic orders.* Academic Press.

Bettinger, R. (2009). *Inversion d'un système par krigeage : application à la synthèse des catalyseurs à haut débit.* PhD thesis, Université Nice-Sophia Antipolis.

Bichon, B. J., Eldred, M. S., Swiler, L. P., Mahadevan, S., and McFarland, J. M. (2008). Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA Journal*, 46 :2459–2468.

Bishop, C. (2006). *Pattern Recognition and Machine Learning.* Springer, New York.

Boutsikas, M. and Vaggelatou, E. (2002). On the distance between convex-ordered random variables, with applications. *Advances in Applied Probability*, 34 :349–374.

Bratley, P. and Fox, B. (1988). Algorithm 659 : Implementing Sobol's quasirandom sequence generator. *ACM Transactions on Methematical Software (TOMS)*, 14(1) :88–100.

Brazauskas, V., Jones, B. L., Puri, M. L., and Zitikis, R. (2008). Estimating conditional tail expectation with acturial applications in view. *Journal of Statistical Planning and Inference*, 138 :3590–3604.

Bucher, C. (1988). Adaptive sampling : an iterative fast monte carlo procedure. *Structural Safety*, 5 :119–126.

Bucklew, J. (2004). *Introduction to rare event simulation.* Springer.

Calfisch, R. (1998). Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*, 7(4) :1–49.

Cannamela, C., Iooss, B., and Le Gratiet, L. (2014). A Bayesian approach for global sensitivity analysis of (multi-fidelity) computer codes. *SIAM/ASA Journal of Uncertainty Quantification*, pages 336–363.

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.

Chafaï, D. and Malrieu, F. (2016). *Recueil de modèles aléatoires*, volume 78 of *Mathématiques & Applications*. Springer.

Chan, W., Proschan, F., and Sethuraman, J. (1990). Convex-ordering among functions, with applications to reliability and mathematical statistics. *Topics in Statistical Dependance*.

Chevalier, C. (2013). *Fast uncertainty reduction strategies relying on Gaussian process models*. PhD thesis, University of Bern.

Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2014). Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56(4) :455–465.

Chevalier, C., Ginsbourger, B., Bect, J., and Molchanov, I. (2013). Estimating and quantifying uncertainties on level sets using the Vorob'ev expectation and deviation with Gaussian process models. *mODa 10 Advances in Model-Oriented Design and Analysis*.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4) :327–335.

Chilès, J. and Delfiner, P. (1999). *Geostatistis : modeling spatial uncertainty*, volume 2. Wiley series in probability and statistics.

Chiles, J.-P. and Delfiner, P. (2009). *Geostatistics : modeling spatial uncertainty*. Wiley Series in Probability and Statistics. John Wiley & Sons.

Choudhry, M. (2013). *An introduction to Value at Risk*. Wiley.

Cizeli, L., Mavko, B., and Riesch-Oppermann, H. (1994). Application of first and second order reliabilty methods in the safety assessment of cracked steam generator tubing. *Nuclear Engineering and Design*, 147 :359–368.

Cohen, A., Devore, R., Petrova, G., and Wojtaszczyk, P. (2013). Finding the minimum of a function. *Methods and Applications of Analysis*, 20(4) :365–381.

Davis, M. (2013). Consistency of risk measure estimates.

De Haan, L. (1981). Estimation of the minimum of a function using order statistics. *Journal of the American Statistical Association*, 76(374) :467–469.

Delmas, J.-F. and Jourdain, B. (2006). *Modèles aléatoires : applications aux sciences de l'ingénieur et du vivant*. Springer Science & Business Media.

Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2005). *Actuarial theory for dependant risks : measures, orders and models*. Wiley.

Dhaene, J., Denuit, M., Goovaerts, M., Kaas, R., and Vyncke, D. (2002). The concept of comonotonicity in actuarial science and finance : theory. *Insurance : Mathematics & Economics*, 31 :3–33.

Dhaene, J., Kaas, R., Goovaerts, M., and Denuit, M. (2001). *Modern actuarial risk theory*. Kluwer Academic Publishers.

Diggle, P. and Ribeiro, P.-J. (2007). *Model-based geostatistics*. Springer Series in Statistics.

Ditlevsen, O. and Madsen, H. (1996). *Structural reliability methods*, volume 178. Wiley New York.

Dubourg, V., Deheeger, F., and Sudret, B. (2013). Metamodel-based importance sampling for structural reliability analysis. *Probabilistic Engineering Mechanics*, 33 :47–57.

Dufour, J.-M. (2003). Properties of moments of random variables. *Cours de l'Université de Montréal*.

Echard, B., Gayton, N., and Lemaire, M. (2011). AK-MCS : an active learning reliability method combining Kriging and Monte Carlo Simulation. *Structural Safety*, 33 :145–154.

Echard, B., Gayton, N., Lemaire, M., and Relun, N. (2013). A combined importance sampling and Kriging reliability method for small failure probabilities with time-demanding numerical models. *Reliability Engineering & System Safety*, 111 :232–240.

El Amri, M. R., Helbert, C., Lepreux, O., Munoz Zuniga, M., Prieur, C., and Sinoquet, D. (2018). Data-driven stochastic inversion under functional uncertainties. working paper or preprint.

Embrechts, P. (2000). Extreme value theory : potential and limitations as an integrated risk management tool. *Derivatives Use, Trading & Regulation*, pages 449–456.

Emmerich, M., Giannakoglou, K., and Naujoks, B. (2006). Single and multiobjective evolutionary optimization assisted by Gaussian random field metamodels. *IEEE Transactions on Evolutionary Computation*, 10 :421–439.

Fang, K.-T., LI, R., and Sudjianto, A. (2006). *Design and modeling for computer experiments*. Computer science and data analysis, Chapman and Hall/CRC. Computer science and data analysis.

Fernique, X. (1971). Régularité de processus gaussiens. *Inventiones mathematicae*, 12(4).

Gayton, N., Bourinet, J., and Lemaire, M. (2003). CQ2RS : a new statisticial approach to the response surface method for reliability analysis. *Structural Safety*, 25(1) :99–121.

Ginsbourger, D. (2009). *Multiples métamodèles pour l'approximation et l'optimisation de fonctions numériques multivariables*. PhD thesis, École nationale supérieure des Mines de Saint-Étienne.

Guyader, A. and Hengartner, N. (2011). Simulation and estimation of extreme quantiles and extreme probabilities. *Applied Mathematics & Optimization*, 64 :171–196.

Györfi, L., Kohler, M., Krzyżak, A., and Walk, H. (2002). *A distribution-free theroy of nonparametric regression*. Springer.

Hasofer, A. and Lind, N. (1974). Exact and invariant second moment code format. *Journal of Engineering Mechanics*, 100 :111–121.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, New York.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1) :97–109.

Hürlimann, W. (1999). Extermal moments methods and stochastic orders. Monograph.

Hürlimann, W. (2003). Conditional Value-at-Risk bounds for compound Poisson risks and a normal approximation. *Applied Mathematics*, pages 141–153.

J. Shervish, M. (2010). *Theory of statistics*. Springer.

Jeffreys, H. (1961). *Theory of probability*. Oxford University Press, London.

Johnson, M., Moore, J., and Ylvisaker, D. (1990). Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 26(2) :455–492.

Jones, D. R., Perttunen, C. D., and Stuckman, B. E. (1993). Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Applications*, 79(1) :157–181.

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. Workshop on Global Optimization (Trier, 1997).

Joseph, V. and Hung, Y. (2008). Orthogonal-maximim latin hypercube designs. *Statistica Sinica*, (18) :171–186.

Jourdain, B. (2019). Méthodes de Monte Carlo par chaînes de Markov et méthodes particulaires. *Format électronique*.

Kaas, R., Dhaene, J., Vyncke, D., Goovaerts, M., and Denuit, M. (2002). A simple geometric proof that comonotonic risks have the convex-largest sum. *Astin Bulletin*, 32 :71–80.

Kaas, R., Van Heerwarden, A., and Goovaerts, M. (1994). *Ordering of actuarial risks*. Institute for Actuarial Science and Econometrics. University of Amsterdam.

Kahn, H. and Harris, T. (1951). Estimation of particle transmission by random sampling. *National Bureau of Standards applied mathematics series*, 12 :27–30.

Kass, E. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91 :1343–1370.

Kato, K. (2009). Improved prediction for a multivariate normal distribution with unknown mean and variance. *Annals of the Institute of Statistical Mathematics*, 61 :531–542.

Kaymaz, I. (2004). Application of kriging method to structural reliability problems. *Structural Safety*.

Krige, D. (1951). A statistical approach to some mine valuations and allied problems at the witwatersrand. Mémoire de D.E.A., University of Witwatersrand.

Kusuoka, S. (2001). *On law invariant coherent risk measures*. Advances in mathematical economics, Vol. 3. Springer, Tokyo. Dependence, risk bounds, optimal allocations and portfolios.

Le Gratiet, L. (2013). *Multi-fidelity Gaussian process regression for computer experiments*. PhD thesis, Université Paris-Diderot Paris VII.

Lebrun, R. (2013). *Contribution à la modélisation de la dépendance stochastique*. PhD thesis, Université Paris-Diderot Paris VII.

Li, R. and Sudjianto, A. (2005). Analysis of computer experiments using penalized likelihood in Gaussian Kriging models. *Technometrics*, 47 :111–120.

Malrieu, F. and Zitt, P.-A. (2017). On the persistence regime for Lotka-Volterra in randomly fluctuating environments. *ALEA*, 14(2) :733–749.

Marrel, A. (2008). *Mise en œuvre et utilisation du métamodèle processus gaussien pour l'analyse de sensibilité de modèles numériques.* PhD thesis, Institut national des sciences appliquées de Toulouse.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology.*

McKay, M., Beckman, R., and Conover, W. (1979). A comparaison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2) :239–245.

Meilijson, I. and Nadas, A. (1979). Convex majorization with an application to the length of critical paths. *Journal of Applied Probability*, 16 :671–677.

Melchers, R. (1990). Search-based importance sampling. *Structural Safety*, 9 :117–128.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equation of state calculation by fast computing machines. *The journal of Chemical Physics*, 21(6) :1087–1091.

Mitchell, T., Morris, M., and Ylvisaker, D. (1995). Existence of smoothed stationary processes on an interval. *Stochastic Processes and Their Application*, 35 :109–119.

Molchanov, I. (2005). *Theory of random sets.* Springer.

Müller, A. (1996). Orderings of risks : a comparative study via stop-loss transforms. *Insurance : Mathematics and Economics*, 17 :215–222.

Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks.* Wiley Series in Probability and Statistics. Hohn Wiley & Sons Ltd, Chichester.

Nataf, A. (1962). Détermination des distributions dont les marges sont données. *Comptes Rendus de l'Académie des Sciences*, 225 :42–43.

Neveu, J. (1986). Arbres et processus de Galton-Watson. *Annales de l'I.H.P*, 22(2) :199–207.

Niederreiter, H. (1992). *Random number generation and quasi-Monte Carlo methods.* Society for Industrial and Applied Mathematics (SIAM).

Oakley, J. E. and O'Hagan, A. (2004). Probabilistic sensitivity analysis of complex models : a Bayesian approach. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 66(3) :751–769.

Oger, J. (2014). *Méthodes probabilistes pour l'évaluation de risques en production industrielle.* PhD thesis, Université de Tours.

Oger, J., Leduc, P., and Lesigne, E. (2015). A random field model and decision support in industrial production. *J. SFdS*, 156(3) :1–26.

Patterson, H. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58 :545–554.

Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R., and Kim, N. (2010). Adaptive designs of experiments for accurate approximation of a target region. *Journal of Mechanical Design*, 132.

Pitera, M. and Schmidt, T. (2018). Unbiased estimation of risk. *Journal of Banking and Finance*, 91.

Pronzato, L. and Müller, W. (2012). Design of computer experiments : space filling and beyond. *Statistics and Computing*, 22(3) :381–701.

Rackwitz, R. and Flessler, B. (1978). Structural reliability under combined random load sequences. *Computer & Structures*, 9(5) :489–494.

Rafajlowicz, E. and Schwabe, R. (2006). Halton and Hammersley sequences in multivariate nonparametric regression. *Statistics & Probability Letters*, 76(8) :803–812.

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA.

Robert, C. (2007). *The Bayesian Choice*. Springer.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer.

Roberts, G. and Rosenthal, J. (2004). General state space Markov Chains and MCMC algorithm. *Probability Surveys*, 1 :20–71.

Rockafellar, R. and Uryasev, S. (2001). Conditional Value-at-Risk for general loss distribution. *Journal of Banking and Finance*, 26(7) :1443–1471.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23 :470–472.

Rosenbluth, M. and Rosenbluth, A. W. (1955). Monte Carlo calculation of the average extension of molecular chains. *Journal of Chemical Physics*, 23(2) :356–359.

Rubino, G. and Tuffin, B. (2009). *Rare event simulation using Monte Carlo methods*. Wiley.

Rubinstein, R. (1999). The cross-entropy method for combinatorial and continuous optimization. *Methodology and computing in applied probability*, 1(2) :217–190.

Ruschendorf, L. (2013). *Mathematical risk analysis*. Springer, Heidelberg.

Sacks, J., Mitchell, T. J., Welch, W. J., and Wynn, H. P. (1989a). Design and analysis of computer experiments. *Statistical Science*, 4(4) :409–435.

Sacks, J., Schiller, B. S., and Welch, W. J. (1989b). Designs for computer experiments. *Technometrics*, 31(1).

Santner, T. J., Williams, B. J., and Notz, W. I. (2003). *The design and analysis of computer experiments*. Springer Science & Business Media.

Shaked, M. and Shanthikumar, J. (2007). *Stochastic orders*. Springer Series in Statistics.

Shorack, G. (2000). *Probability for Statisticians*. Springer Texts in Statistics.

Smith, R. (2018). Méthodes probabilistes pour l'évaluation de risques en production industrielle. Master's thesis, master 2 MApI3 de l'université Toulouse III Paul Sabatier.

Solak, E., Murray-Smith, R., Leithead, W. E., Leith, D. J., and Rasmussen, C. (2003). Derivative observations in gaussian process models of dynamic systems. *Advances in Neural Information Processing Systems 15*, 15 :1057–1064.

Stein, M. (1987). Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, 29 :143–151.

Stein, M. L. (1999). *Interpolation of spatial data*. Springer Series in Statistics.

Tasche, D. (2002). Expected shortall and beyond. *Journal of Banking and Finance*, 26 :1519–1533.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Stat.*, 22(4) :1701–1762.

Tuffin, B. (2010). *La simulation de Monte Carlo*. Hermes Science Publications.

Vardapetyan, L., Manges, J., and Cendes, Z. (2008). Sensitivity analysis of s-parameters including port variations using the transfinite element method. pages 527 – 530.

Vestrup, E. (2003). *The Theory of Measures and Integration*. Wiley.

Wackernagel, H. (2003). *Multivariate Geostatistics*. Springer-Verlag, Berlin.

Walter, C. (2017). *Using Poisson processes for rare event simulation*. PhD thesis, Université Paris-Diderot Paris VII.

Wood, G. R. and Zhang, B. P. (1996). Estimation of the Lipschitz constant of a function. *Journal of Global Optimization*, 8(1) :91–103.

Wu, A., Aoi, M., and Pillow, J. (2018). Exploiting gradients and Hessians in Bayesian optimization and Bayesian quadrature.

Xiaoxu, H., Jianqiao, C., and Hongping, Z. (2016). Assessing small failure probabilities by AK-SS : An active learning method cimbining Kriging and Subset Simulation. *Structural Safety*, 59 :86–95.

Zhang, Y. and Der Kiureghian, A. (1995). *Two improved algorithms for reliability analysis*, volume 178 of *Reliability and optimization of structural systems*. Springer.

Résumé :

Pour évaluer la rentabilité d'une production en amont du lancement de son processus de fabrication, la plupart des entreprises industrielles ont recours à la simulation numérique. Cela permet de tester virtuellement plusieurs configurations des paramètres d'un produit donné et de statuer quant à ses performances (i.e. les spécifications imposées par le cahier des charges). Afin de mesurer l'impact des fluctuations des procédés industriels sur les performances du produit, nous nous intéressons en particulier à l'estimation de sa probabilité de défaillance. Chaque simulation exigeant l'exécution d'un code de calcul complexe et coûteux, il n'est pas possible d'effectuer un nombre de tests suffisant pour estimer cette probabilité via, par exemple, une méthode Monte-Carlo. Sous la contrainte d'un nombre limité d'appels au code, nous proposons deux méthodes d'estimation très différentes.

La première s'appuie sur les principes de l'estimation bayésienne. Nos observations sont les résultats de simulation numérique. La probabilité de défaillance est vue comme une variable aléatoire, dont la construction repose sur celle d'un processus aléatoire destiné à modéliser le code de calcul coûteux. Pour définir correctement ce modèle, on utilise la méthode de krigeage. Conditionnellement aux observations, la loi a posteriori de la variable aléatoire, qui modélise la probabilité de défaillance, est inaccessible. Pour apprendre sur cette loi, nous construisons des approximations des caractéristiques suivantes : espérance, variance, quantiles... On utilise pour cela la théorie des ordres stochastiques pour la comparaison de variables aléatoires et, plus particulièrement, l'ordre convexe. La construction d'un plan d'expériences optimal est assurée par la mise en place d'une procédure de planification d'expériences séquentielle, basée sur le principe des stratégies SUR (« Stepwise Uncertainty Reduction »).

La seconde méthode est une procédure itérative, particulièrement adaptée au cas où la probabilité de défaillance est très petite, i.e. l'événement redouté est rare. Le code de calcul coûteux est représenté par une fonction que l'on suppose lipschitzienne. À chaque itération, cette hypothèse est utilisée pour construire des approximations, par défaut et par excès, de la probabilité de défaillance. Nous montrons que ces approximations convergent vers la valeur vraie avec le nombre d'itérations. En pratique, on les estime grâce à la méthode Monte-Carlo dite de « splitting ».

Les méthodes que l'on propose sont relativement simples à mettre en œuvre et les résultats qu'elles fournissent peuvent être interprétés sans difficulté. Nous les testons sur divers exemples, ainsi que sur un cas réel provenant de la société STMicroelectronics.

Mots clés :

Probabilité de défaillance, événement rare, méthodes Monte-Carlo et de » splitting« , krigeage, processus gaussien, ordre convexe, encadrement de quantiles, inférence bayésienne, plan d'expériences, stratégies SUR, fonction lipschitzienne.

Abstract :

To evaluate the profitability of a production before the launch of the manufacturing process, most industrial companies use numerical simulation. This allows to test virtually several configurations of the parameters of a given product and to decide on its performance (defined by the specifications). In order to measure the impact of industrial process fluctuations on product performance, we are particularly interested in estimating the probability of failure of the product. Since each simulation requires the execution of a complex and expensive calculation code, it is not possible to perform a sufficient number of tests to estimate this probability using, for example, a Monte-Carlo method. Under the constraint of a limited number of code calls, we propose two very different estimation methods.

The first is based on the principles of Bayesian estimation. Our observations are the results of numerical simulation. The probability of failure is seen as a random variable, the construction of which is based on that of a random process to model the expensive calculation code. To correctly define this model, the Kriging method is used. Conditionally to the observations, the posterior distribution of the random variable, which models the probability of failure, is inaccessible. To learn about this distribution, we construct approximations of the following characteristics : expectation, variance, quantiles... We use the theory of stochastic orders to compare random variables and, more specifically, the convex order. An optimal design of experiments is ensured by the implementation of a sequential planning procedure, based on the principle of the SUR (« Stepwise Uncertainty Reduction ») strategies.

The second method is an iterative procedure, particularly adapted to the case where the probability of failure is very small, i.e. the redoubt event is rare. The expensive calculation code is represented by a function that is assumed to be Lipschitz continuous. At each iteration, this hypothesis is used to construct approximations, by default and by excess, of the probability of failure. We show that these approximations converge towards the true value with the number of iterations. In practice, they are estimated using the Monte-Carlo method known as splitting method.

The proposed methods are relatively simple to implement and the results they provide can be easily interpreted. We test them on various examples, as well as on a real case from STMicroelectronics.

Keywords :

Probability of failure, rare event, Monte-Carlo methods and splitting, Kriging, Gaussian processes, convex order, bounds on quantiles, Bayesian inference, design of experiments, SUR strategies, Lipschitz function.