HAL
open science

# Méthodes probabilistes pour l'évaluation de risques en production industrielle.

Julie Oger

## To cite this version:

Julie Oger. Méthodes probabilistes pour l'évaluation de risques en production industrielle.. Modélisation et simulation. Université François Rabelais - Tours, 2014. Français. NNT : . tel-00982740

**HAL Id: tel-00982740**

**https://theses.hal.science/tel-00982740**

Submitted on 24 Apr 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# UNIVERSITÉ FRANÇOIS RABELAIS DE TOURS

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

**THÈSE** présentée par :

**Julie OGER**

soutenue le : 16 avril 2014

pour obtenir le grade de : Docteur de l'université François - Rabelais de Tours

Discipline/ Spécialité : Mathématiques

## Méthodes probabilistes pour l'évaluation de risques en production industrielle.

THÈSE DIRIGÉE PAR :

| | |
|---|---|
| LESIGNE Emmanuel | Professeur, Université François - Rabelais de Tours |

RAPPORTEURS :

| | |
|---|---|
| BIERMÉ Hermine | Professeur, Université de Poitiers |
| PRIEUR Clémentine | Professeur, Université Joseph Fourier de Grenoble |

JURY :

| | |
|---|---|
| BIERMÉ Hermine | Professeur, Université de Poitiers |
| CHAUVEAU Didier | Professeur, Université d'Orléans |
| GALLARDO Léonard | Professeur, Université François - Rabelais de Tours |
| LEDUC Philippe | Docteur Ingénieur, STMicroelectronics Tours |
| LESIGNE Emmanuel | Professeur, Université François - Rabelais de Tours |
| PRIEUR Clémentine | Professeur, Université Joseph Fourier de Grenoble |
| ROCHETTE Michel | Directeur de recherche, ANSYS Villeurbanne |

# Remerciements

Je souhaite exprimer toute ma gratitude à Emmanuel Lesigne et Philippe Leduc pour leur investissement, leur attention, leur patience et leur enthousiasme tout au long de ces trois années de collaboration.

Je remercie Hermine Biermé et Clémentine Prieur pour avoir accepté de juger et de rapporter ce travail. Merci également à Didier Chauveau, Léonard Gallardo et Michel Rochette pour avoir accepté d'être membres du jury.

Ce travail de thèse s'est déroulé dans le cadre d'une convention CIFRE entre l'université François-Rabelais et la société STMicroelectronics. Je souhaite ainsi remercier les membres des différentes institutions académiques de Tours, le laboratoire de mathématiques et physique théorique, la fédération Denis Poisson, l'école doctorale MIPTIS et l'université François-Rabelais, pour m'avoir fourni un cadre de travail agréable durant ces trois années. Je remercie également mes collègues de la société STMicroelectronics de Tours pour m'avoir aidé à m'épanouir dans ce travail. Merci à l'ANRT pour avoir permis cette collaboration entre l'univers académique et l'univers industriel.

Plus personnellement, je souhaite exprimer une pensée sincère à Pierre Damphousse, sans qui mon aventure tourangelle n'aurait pu avoir lieu.

Ce travail de thèse a pu également être mené à son terme grâce de très bons amis nantais et tourangeaux qui ont su être présents aux bons moments. Merci à vous.

Je termine ces remerciements par ma famille dont mes essentiels : Philippe, Michelle, Justine et Thomas. Je ne peux exprimer toute ma gratitude et tous mes sentiments pour votre soutien, votre reconnaissance et votre amour. Merci du fond du cœur.

# Résumé

Dans un contexte industriel compétitif, une prévision fiable du rendement est une information primordiale pour déterminer avec précision les coûts de production et donc assurer la rentabilité d'un projet. La quantification des risques en amont du démarrage d'un processus de fabrication permet des prises de décision efficaces. Durant la phase de conception d'un produit, les efforts de développement peuvent être alors identifiés et ordonnés par priorité. Afin de mesurer l'impact des fluctuations des procédés industriels sur les performances d'un produit donné, la construction de *la probabilité du risque défaillance* est développée dans cette thèse. La relation complexe entre le processus de fabrication et le produit conçu (non linéaire, caractéristiques multi-modales...) est assurée par une méthode de régression bayésienne. Un champ aléatoire représente ainsi, pour chaque configuration du produit, l'information disponible concernant la probabilité de défaillance. Après une présentation du modèle gaussien, nous décrivons un raisonnement bayésien évitant le choix a priori des paramètres de position et d'échelle. Dans notre modèle, le mélange gaussien a priori, conditionné par des données mesurées (ou calculées), conduit à un posterior caractérisé par une distribution de Student multivariée.

La nature probabiliste du modèle est alors exploitée pour construire une probabilité de risque de défaillance, définie comme une variable aléatoire. Pour ce faire, notre approche consiste à considérer comme aléatoire toutes les données inconnues, inaccessibles ou fluctuantes. Afin de propager les incertitudes, une approche basée sur les ensembles flous fournit un cadre approprié pour la mise en œuvre d'un modèle bayésien imitant le raisonnement d'expert. L'idée sous-jacente est d'ajouter un minimum d'information a priori dans le modèle du risque de défaillance.

Notre méthodologie a été mise en œuvre dans un logiciel nommé GoNoGo. La pertinence de cette approche est illustrée par des exemples théoriques ainsi que sur un exemple réel provenant de la société STMicroelectronics.

**Mots clés :** Krigeage, inférence bayésienne, mélange de processus gaussiens, distribution de Student multivariée, analyse d'incertitude, évaluation de rendement industriel

# Abstract

In competitive industries, a reliable yield forecasting is a prime factor to accurately determine the production costs and therefore ensure profitability. Indeed, quantifying the risks long before the effective manufacturing process enables fact-based decision-making. From the development stage, improvement efforts can be early identified and prioritized. In order to measure the impact of industrial process fluctuations on the product performances, the construction of a failure risk probability estimator is developed in this thesis. The complex relationship between the process technology and the product design (non linearities, multi-modal features...) is handled via random process regression. A random field encodes, for each product configuration, the available information regarding the risk of non-compliance. After a presentation of the Gaussian model approach, we describe a Bayesian reasoning avoiding a priori choices of location and scale parameters. The Gaussian mixture prior, conditioned by measured (or calculated) data, yields a posterior characterized by a multivariate Student distribution.

The probabilistic nature of the model is then operated to derive a failure risk probability, defined as a random variable. To do this, our approach is to consider as random all unknown, inaccessible or fluctuating data. In order to propagate uncertainties, a fuzzy set approach provides an appropriate framework for the implementation of a Bayesian model mimicking expert elicitation. The underlying leitmotiv is to insert minimal a priori information in the failure risk model.

Our reasoning has been implemented in a software called GoNoGo. The relevancy of this concept is illustrated with theoretical examples and on real-data example coming from the company STMicroelectronics.

**Keywords :** Kriging, Bayesian inference, Gaussian processes mixture prior, multivariate t-distribution, uncertainty analysis, manufacturing yield evaluation

# Table des matières

**Introduction** **13**

0.1 Cadre et motivations . . . . . 13

0.2 Construction d'un modèle . . . . . 15

0.3 Exploitation du modèle . . . . . 17

0.4 Application . . . . . 18

**1 Modèles aléatoires** **21**

1.1 Principe général de la construction d'un modèle aléatoire d'une fonction imparfaitement connue . . . . . 21

1.1.1 Conditionnement et champ aléatoire . . . . . 21

1.1.2 Choix d'un prior . . . . . 23

1.2 Champ Gaussien . . . . . 24

1.3 Mélange de champs gaussiens avec moyenne aléatoire . . . . . 28

1.3.1 Distribution conditionnelle du champ aléatoire . . . . . 28

1.3.2 Distribution conditionnelle de la moyenne aléatoire . . . . . 31

1.4 Mélange de champs gaussiens avec moyenne et variance aléatoires . . . . . 33

1.4.1 Distribution de Student multivariée . . . . . 33

1.4.2 Distribution conditionnelle du champ aléatoire . . . . . 33

1.4.3 Distribution conditionnelle de la moyenne aléatoire . . . . . 39

1.4.4 Distribution conditionnelle de la variance aléatoire . . . . . 41

1.5 Modèle mixte aléatoire . . . . . 44

1.5.1 Distribution conditionnelle du champ aléatoire . . . . . 44

1.5.2 Distribution conditionnelle de la moyenne aléatoire . . . . . 48

1.5.3 Distribution conditionnelle de la variance aléatoire . . . . . 51

1.6 Paramétrisation des modèles . . . . . 53

1.6.1 Fonctions de corrélation . . . . . 53

1.6.2 Maximum de vraisemblance . . . . . 56

1.6.3 Validation croisée . . . . . 58

1.6.4 Paramètres vus comme aléatoires . . . . . 58

1.7 Exemples et résultats . . . . . 59
1.7.1 Comparaison du champ de Student et du champ gaussien . . . . . 59
1.7.2 Étude de l'erreur de prédiction du champ de Student . . . . . 60
1.7.3 Performance de l'estimateur du maximum de vraisemblance . . . . . 61
1.7.4 Comparaison du champ de Student et du champ mixte . . . . . 63
1.8 Compléments . . . . . 65
1.8.1 Gestion du bruit . . . . . 65
1.8.2 Transformation de la réponse . . . . . 68
1.8.3 Réponse multidimensionnelle . . . . . 69
Formulaire . . . . . 71

**2 Exploitation d'un modèle aléatoire 71**
2.1 Probabilité du risque de défaillance . . . . . 72
2.1.1 La probabilité de défaillance vue comme une variable aléatoire . . . 72
2.1.2 Choix d'une distribution uniforme pour le seuil . . . . . 74
2.1.3 Monte-Carlo préférentiel . . . . . 76
2.1.4 Conclusion sur le modèle d'évaluation du risque de défaillance . . . . 77
2.1.5 Mise en œuvre numérique du calcul du risque de défaillance . . . . . 77
2.2 Probabilité du risque d'insuffisance du modèle . . . . . 80
2.2.1 Description de l'approche . . . . . 80
2.2.2 Choisir la distribution des seuils . . . . . 81
2.2.3 Monte-Carlo préférentiel . . . . . 83
2.3 Calcul du risque de défaillance : synoptique . . . . . 84
2.4 Calcul de rendement sur des exemples théoriques . . . . . 87
2.4.1 Exemple de la forme quadratique. . . . . 88
2.4.2 Exemple du sinus . . . . . 88
2.4.3 Exemple des gaussiennes . . . . . 90
2.5 Influence des facteurs . . . . . 93
2.5.1 Méthodologie . . . . . 93
2.5.2 Critère de comparaison . . . . . 94
2.5.3 Exemple théorique . . . . . 95

**3 Applications pratiques 97**
3.1 Description du logiciel GoNoGo . . . . . 97
3.1.1 Cœur de calcul . . . . . 98
3.1.2 Interface utilisateur . . . . . 98
3.2 Étude d'un cas réel de la société STMicroelectronics . . . . . 101
3.2.1 Diplexeur . . . . . 101

3.2.2 HFSS . . . . . 103
3.2.3 CornerStone et GoNoGo . . . . . 104
3.2.4 Étude “marketing/design” : choix des spécifications . . . . . 105
3.2.5 Sélection de transformations pour les réponses . . . . . 108
3.2.6 Étude ingénierie de test : influence des réponses . . . . . 109
3.2.7 Étude “manufacturing” : dérive de la distribution des facteurs . . . . 109

**Conclusion** **113**

**Annexes** **117**

**A Démonstration du lemme 4** **117**

# Introduction

## 0.1 Cadre et motivations

Durant la mise en œuvre d'une production industrielle, le processus de fabrication ne peut être entièrement et parfaitement contrôlé. Il existe une variabilité intrinsèque à ce processus qui peut considérablement détériorer les caractéristiques des pièces manufacturées, induisant des pièces non-fonctionnelles. De telles variations peuvent être critiques lorsqu'il s'agit de systèmes techniques complexes, comme par exemple des circuits intégrés développés par l'industrie de la microélectronique. Ainsi, bien souvent, lors de l'élaboration d'un composant, d'une fonction ou d'un circuit électronique, les ingénieurs peuvent être contraints à la sur-qualité : des précautions, des marges sont prises afin d'assurer que le produit satisfasse le cahier des charges, c'est-à-dire une liste de spécifications. Ces mesures s'accompagnent en général de coûts supplémentaires qui détériorent la profitabilité. C'est pourquoi, il est nécessaire de prendre en compte lors du développement cette variabilité du processus de fabrication, et d'en quantifier les conséquences.

Pour évaluer l'influence des paramètres fluctuants, il est souvent effectué une analyse aux limites ("corner process") en appliquant par exemple la méthodologie des plans d'expériences. Cette technique consiste à caractériser des réalisations du produit correspondant à certaines configurations limites de ses paramètres. Cependant, les conclusions de ces campagnes de mesures sont, en général, qualitatives et par conséquent incomplètes. De plus, le temps et le coût requis pour la réalisation de prototypes physiques sont souvent prohibitifs. Comme alternative, les modèles numériques (déterministes) implémentés dans des logiciels de simulation d'ingénierie offrent un moyen pour calculer les caractéristiques pertinentes (thermiques, mécaniques, électriques...) d'un produit. Ainsi, de nos jours, les ingénieurs peuvent virtuellement explorer diverses configurations de conception et obtenir un aperçu de l'ensemble des performances du produit final. Dans ce contexte, des études statistiques peuvent être conduites pour évaluer l'effet de la variabilité du processus sur un rendement de production. La plus populaire des méthodes statistiques est la méthode Monte-Carlo, qui consiste à comptabiliser les évènements "échecs" parmi des configurations des paramètres du produit, tirées suivant leur distribution de probabilités. Une estimation de la distribution des paramètres étudiés est effectivement accessible étant donné que le processus de fabrication est surveillé en continu. La méthode Monte-Carlo est facile à mettre en œuvre mais son efficacité dépend essentiellement de la complexité du modèle déterministe. Quand le simulateur numérique considéré est consommateur en temps nous avons seulement accès à une information incomplète. Par exemple, la durée de simulations basées sur

la méthode des éléments finis, de quelques heures à plusieurs jours, n'est pas compatible avec une approche Monte-Carlo seule.

Ainsi, lorsque les données sont rares, il est nécessaire de propager l'information disponible dans l'espace des facteurs (i.e. les paramètres du produit). Pour cela, nous choisissons d'utiliser un modèle aléatoire dont nous proposons plusieurs constructions dans le premier chapitre de cette thèse. La méthode d'échantillonnage Monte-Carlo est ensuite appliquée à cette représentation du code informatique (du simulateur), comme le font Pfingsten et al. (2006). Compte tenu des distributions de probabilités des paramètres du produit, nous allons au-delà de l'approche de Pfingsten, en exploitant le modèle aléatoire pour définir la probabilité du risque de défaillance et la probabilité du risque d'insuffisance du modèle dont les constructions constitueront le deuxième chapitre de cet écrit. La mise en œuvre pratique de cette approche et l'illustration sur un exemple réel forment le troisième chapitre de cette thèse.

Plus généralement, l'approche présentée dans cette thèse peut s'appliquer à toute situation où un champ déterministe, indexé par un espace $X$ muni d'une probabilité, n'est connu qu'en un nombre limité de points de cet espace. Elle permet de construire un modèle aléatoire de ce champ, puis d'approcher le sous-espace de $X$ pour lequel le champ déterministe atteint des valeurs critiques. Cette approximation du "sous-espace critique" conduit à la construction d'une probabilité d'être dans le domaine critique, et d'une probabilité que le modèle soit insuffisamment informé pour que la réponse précédente soit significative.

Pour décrire notre approche avec plus de précision, nous revenons à notre cadre : des pièces manufacturées sont caractérisées par un jeu de facteurs sujets à des variations numériques, ainsi que par une grandeur mesurant la performance du produit pour une application donnée, appelée réponse. Celle-ci est soumise à une spécification (performance minimum), qui détermine si une pièce est fonctionnelle ou non.

Introduisons quelques données et notations. Nous considérons que le produit étudié (c'est-à-dire chaque pièce individuelle fabriquée) est caractérisé par un nombre $D$ de facteurs numériques. Chaque facteur peut varier dans un intervalle donné et permet la définition de l'espace des facteurs $X \subseteq \mathbb{R}^D$. Chaque jeu de facteurs $x$ ($\in X$) caractérise une valeur numérique $y(x)$, et c'est sur cette valeur de la *réponse* $y$ que sont imposées les spécifications du produit (i.e le cahier des charges). Ces spécifications permettent de définir *l'espace hors-spécifications* $A \subseteq \mathbb{R}$ : le produit caractérisé par le jeu de facteur $x$ ne satisfait pas les spécifications si et seulement si $y(x) \in A$. L'espace des facteurs $X$ et l'espace hors-spécifications $A$ sont supposés connus, mais, de la fonction déterministe $x \mapsto y(x)$, nous n'avons qu'une information partielle.

De plus, la distribution de probabilité $P$ est donnée sur l'espace des facteurs $X$. Cette distribution reflète la variabilité des facteurs et est également considérée connue.

Comme nous le disions précédemment, la connaissance de la quantité $y(x)$ est rarement disponible pour tout $x$. Nous avons seulement accès à un nombre restreint de données. Supposons que nous connaissions $n$ valeurs de la réponse déterministe $(y_i = y(x_i))_{1 \leq i \leq n}$ respectivement pour les valeurs de jeu facteurs $(x_i)_{1 \leq i \leq n}$. Avec ces données, notre but est de définir une variable aléatoire nommée *probabilité du risque de défaillance*. Sa distribution devrait aider à mesurer la sensibilité du produit aux variations du processus de fabrication et donc à valider la conception proposée. On propose également la construction de la

*probabilité du risque d'insuffisance* du modèle qui permet d'évaluer la pertinence de cette distribution pour cette validation.

Commençons avec la construction d'un champ aléatoire $(Y_x)_{x\in X}$ modélisant la réponse inconnue $(y(x))_{x\in X}$.

## 0.2 Construction d'un modèle

Plusieurs méthodes sont décrites dans la littérature pour inférer un modèle $(Y_x)_{x\in X}$ d'après un nombre limité de données déterministes disponibles

$$Y_{x_1} = y_1, Y_{x_2} = y_2, \ldots, Y_{x_n} = y_n.$$

L'analyse de régression linéaire est la méthode de choix dans la communauté scientifique pour estimer les relations entre les variables prédictives. Elle consiste à approcher la fonction inconnue $x \mapsto y(x)$ par un modèle qui est une combinaison linéaire de fonctions connues. Un modèle de régression linéaire est alors défini par le système suivant

$$y_i = \boldsymbol{f}(\boldsymbol{x_i})\boldsymbol{\beta}^T + \epsilon_i \quad \text{pour} \quad 1 \leq i \leq n,$$

où les vecteurs $(\boldsymbol{x_i})$, les valeurs $(y_i)$ et le vecteur $\boldsymbol{f}$ de fonctions sont connus, et où le vecteur de paramètres $\boldsymbol{\beta}$ et les erreurs (ou résidus) $\epsilon_i$ sont des inconnues à déterminer (voir Atkinson et Donev (1992)). Usuellement, on recherche le vecteur $\boldsymbol{\beta}$ tel que les valeurs $\epsilon_i$ soient les plus petites possibles (méthode des moindres carrés).

Quand le phénomène étudié est complexe, l'ordre du modèle nécessaire pour correctement représenter les données, implique de nombreuses inconnues. Cet inconvénient est d'autant plus grand que l'espace des facteurs est de grande dimension. Ainsi, pour un modèle polynomial de degré $d$ en $D$ variables, c'est-à-dire quand $\boldsymbol{f}\left(\boldsymbol{x^{(1)}}, \boldsymbol{x^{(2)}}, \ldots, \boldsymbol{x^{(D)}}\right)$ est le vecteur des monômes de degré total $\leq d$ en les variables $x^{(1)}, x^{(2)}, \ldots, x^{(D)}$, il faut déterminer $\binom{D+d}{d}$ coefficients. Si le nombre de données disponibles est inférieur à ce nombre, le modèle est singulier. De plus, dans le cas où les données proviennent de mesures, les résidus $\epsilon_i$ sont souvent interprétés comme du bruit et ainsi considérés comme aléatoires. Ils sont alors fréquemment supposés centrés, de même variance et non corrélés entre eux. Cependant, ces hypothèses semblent difficiles à justifier dans le cas où les données proviennent de calculs numériques, reproductibles par définition.

Ces difficultés sont surmontées par les modèles basés sur le processus gaussien (GP) dont la méthode originale, "kriging", a été formalisée par Matheron (1989) dans le domaine des géo-sciences. Cette méthode consiste, dans un premier temps, à sélectionner a priori un champ aléatoire indexé par $X$, nommé usuellement *prior*. On le conditionne ensuite par les données disponibles, pour obtenir, a posteriori, un modèle appelé *posterior* (voir section 1.2). Cette approche a été largement utilisée dans différentes applications telles que les études géo-statistiques (Berger et al. (2000)), l'optimisation (Emmerich et al. (2006)), l'aérologie (Cornford et al. (2002)), ou l'analyse de sensibilité (Oakley et O'Hagan (2004)), justifiée par de meilleurs performances prédictives que la plupart des autres méthodes de régression. Le livre de (Rasmussen et Williams (2006)) offre un panorama complet de cette approche désormais classique.

La modélisation par processus gaussien a plusieurs propriétés intéressantes lorsque l'on traite des données issues de simulations numériques. Nous développerons ces aspects plus en détail dans la section 1.1.1.

Un problème majeur et souvent négligé de cette méthode est le choix a priori des paramètres (moyenne, variance et fonction de corrélation) du modèle, laissé à la responsabilité de l'utilisateur. Afin de réduire la nature arbitraire de ce choix, nous avons, dans un premier temps, proposé de remplacer le champ gaussien par un mélange de champs gaussiens, en évitant le choix a priori de la moyenne. Pour cela, dans le but d'introduire le minimum d'information dans le modèle, la moyenne est considérée comme une variable aléatoire avec une distribution de probabilités uniforme. Le modèle a priori est alors du type :

$$Y_x := U + W_x,$$

où $U$ est une variable aléatoire réelle suivant une distribution uniforme sur un grand intervalle $[-m, m]$ et où $(W_x)_{x\in X}$ est un champ gaussien centré de variance constante. (Ces données aléatoires sont supposées mutuellement indépendantes.) On obtient ainsi un posterior gaussien (section 1.3).

Nous avons ensuite poussé plus avant ce raisonnement et proposé, a priori, le mélange de champs gaussiens suivant :

$$Y_x := U + VW_x,$$

où $U$ est une variable aléatoire réelle suivant une distribution uniforme sur un grand intervalle $[-m, m]$, où $V$ est une variable aléatoire réelle suivant une distribution uniforme sur un grand intervalle $[\epsilon, 1/\epsilon]$ et où $(W_x)_{x\in X}$ est un champ gaussien centré réduit. (Ces données aléatoires sont supposées mutuellement indépendantes.) Ce faisant, les vecteurs aléatoires a posteriori suivent une distribution de Student multivariée (section 1.4).

Nous avons remarqué, à la suite de nos travaux, que des résultats similaires à ces deux précédents champs aléatoires sont obtenus par une approche par modèle hiérarchique bayésien décrits, par exemple, par Sacks et al. (1989) et Santner et al. (2003). En fait, les deux approches s'avèrent équivalentes. Nous reviendrons plus en détail, au prochain chapitre, sur les similitudes et les différences méthodologiques entre ces deux modèles.

En s'inspirant des travaux de Santner et al. (2003), la méthode du conditionnement a été ensuite mixée avec un modèle linéaire, considérant a priori le champ :

$$Y_x := \sum_{i=1}^{\ell} U_i f_i(x) + VW_x$$

où $(U_i)_{1\leq i\leq \ell}$ est un vecteur aléatoire réel suivant une distribution uniforme sur un grand espace $[-m, m]^\ell$, où $V$ est une variable aléatoire réelle suivant une distribution uniforme sur un grand intervalle $[\epsilon, 1/\epsilon]$, où $(W_x)_{x\in X}$ est un champ gaussien centré réduit et où les fonctions $f_i, 1 \leq i \leq \ell$ sont déterministes. (Là aussi, ces données aléatoires sont supposées mutuellement indépendantes.) Le posterior obtenu suit également une distribution de Student multivariée (voir section 1.5).

Une fois le modèle aléatoire disponible, exploitons-le pour calculer un risque de défaillance et un risque d'insuffisance.

## 0.3 Exploitation du modèle

Le champ aléatoire $(Y_x)_{x \in X}$ représente l'information aussi bien que l'incertitude au regard des valeurs observées de la fonction réponse $x \mapsto y(x)$. Une fois ce champ construit, nous pouvons calculer pour chaque point $x$ de l'espace des facteurs $X$, la *probabilité de défaillance*, $\mathbb{P}(Y_x \in A)$.

Nous notons ici, par $\mathbb{P}$, la mesure de probabilité associée au champ aléatoire (signifiant que le champ aléatoire est défini sur l'espace abstrait de probabilité $(\Omega, \mathbb{P})$). De plus, nous rappelons que nous notons $P$ la mesure de probabilité assignée à l'espace des facteurs $X$.

Nous décidons qu'il y a risque de défaillance quand le probabilité de défaillance dépasse un certain seuil fixé $\alpha \in [0, 1]$. Ainsi, *la probabilité du risque de défaillance* au seuil $\alpha$ est définie par

$$R(\alpha) = P(\{x \in X \mid \mathbb{P}(Y_x \in A) > \alpha\}).$$

Afin d'améliorer la lisibilité de ce mémoire, nous utiliserons également pour nommer cette quantité le terme *risque de défaillance* et nous convenons qu'il y a là un (léger) abus de langage. Pour $\alpha$ fixé, la distribution du risque de défaillance est une masse de Dirac au point $R(\alpha)$. Cette masse de Dirac est notée par $\delta_{R(\alpha)}$.

Donner une valeur correcte à $\alpha$ est un problème délicat. D'une part, la décision va influer significativement sur l'évaluation du risque du produit. D'autre part, ce choix est éminemment subjectif puisqu'il dépend fortement de l'attitude face au risque de l'individu. Pour contourner cette difficulté, nous proposons de considérer $\alpha$ comme un variable aléatoire. En conséquence, $R(\alpha)$ est également une variable aléatoire et a pour distribution

$$\mathcal{R} := \int_0^1 \delta_{R(\alpha)} \, \eta(\mathrm{d}\alpha)$$

où $\eta$ est la distribution de probabilité du seuil $\alpha$. Nous verrons par la suite que la distribution uniforme est un choix pour $\eta$ qui donne des propriétés intéressantes comme la préservation de la moyenne ou l'invariance de la distribution uniforme.

En pratique, la quantité $R(\alpha)$ ne peut être calculée analytiquement, on l'approche par une méthode de simulation Monte-Carlo. Suivant les méthodes bayésiennes standards pour les sondages, nous considérons que la quantité $R(\alpha)$ suit une distribution beta, notée $\beta_{n(\alpha)}$. Finalement, la distribution du risque de défaillance est définie, dans le cas où $\eta$ est une distribution uniforme, par

$$\mathcal{R} := \int_0^1 \beta_{n(\alpha)} \, \mathrm{d}\alpha.$$

Ce raisonnement est décrit plus en détail dans la section 2.1.

En complément du risque de défaillance, un raisonnement similaire nous permet d'aboutir au risque d'insuffisance du modèle (voir section 2.2).

Remarquons que les calculs des risques de défaillance et d'insuffisance du modèle sont effectifs pour tout modèle, aléatoire ou non, dès lors que les probabilités de défaillance sont accessibles. **C'est pourquoi, la lecture du deuxième chapitre peut être indépendante de celle du premier.**

Une fois les distributions de ces risques calculées, nous pouvons en extraire des quantités statistiques classiques telles que la moyenne, la variance et un intervalle de confiance.

La moyenne du risque de défaillance apporte au décideur une information qui peut lui permettre de statuer quant à la robustesse du produit, relative au procédé industriel de fabrication. Son intervalle de confiance mesure l'incertitude sur cet indicateur, la source d'incertitude étant en premier lieu la connaissance partielle de la fonction $y$. Le risque d'insuffisance du modèle informe quant à la capacité prédictive de celui-ci.

Dans le cas où l'intervalle de confiance du risque de défaillance est jugé trop grand pour permettre de considérer la valeur moyenne comme une prédiction fiable du risque ou lorsque le risque d'insuffisance est jugé élevé, on conclut que le modèle n'est pas suffisamment renseigné. Pour améliorer la prédiction du risque de défaillance, il est nécessaire de rajouter de l'information au modèle, c'est-à-dire de fournir de nouvelles données. Dans ce but, nous proposons l'utilisation de méthodes d'optimisation, voir section 2.3.

Pour valider notre approche pour le calcul du risque de défaillance, nous avons traité trois exemples théoriques présentant différents niveaux de complexité (section 2.4). Nous avons ensuite comparé notre méthode de calcul du risque avec une méthode statistique classique sur chacun de ces exemples et avec un nombre croissant de données.

Le premier exemple, peu complexe, nous permet de valider notre méthode. Nous remarquons que celle-ci est, dans ce cas, plus performante que la méthode classique ; le risque réel est approché de manière satisfaisante avec beaucoup moins de données.

Le deuxième exemple est plus complexe mais de risque réel peu élevé. Cette complexité le rend difficile à modéliser, même pour notre modèle aléatoire. Et le risque peu élevé favorise une approche de calcul de risque par la méthode statistique classique. Ainsi, on observe, dans ce cas, que cette méthode donne des résultats comparables à notre approche probabiliste.

Le troisième exemple met en évidence une limite caractéristique de la méthode statistique classique. En effet, dans cet exemple de complexité modérée, le risque réel est faible. On observe ainsi que la méthode statistique tend à sous-évaluer le risque réel de défaillance alors que notre approche probabiliste donne une prédiction de ce risque plus robuste.

## 0.4 Application

Afin de mettre en application notre approche, l'ensemble des méthodes précédemment décrites a été implémenté dans un logiciel nommé GoNoGo (section 3.1) où une interface graphique permet, entre autres, à l'utilisateur de saisir toutes les données de l'étude à réaliser, de choisir un modèle aléatoire approprié, de lancer un calcul de rendement et de déterminer un plan d'expériences. Le cœur de calcul a été construit à l'aide de l'environnement de calcul scientifique Matlab (version R2012.a).

Nous illustrons son utilisation sur un cas concret, un produit fourni et conçu par STMicroelectronics : un diplexeur. Dans la section 3.2, nous explicitons trois différentes études qu'il est possible de mener, selon les points de vues d'un commercial, d'un concepteur et d'un gestionnaire de la chaîne de production. Le premier peut s'appuyer sur des prédictions du risque de défaillance pour négocier de meilleures spécifications auprès du client.

Le second peut bien-sûr évaluer si le produit tel qu'il l'a conçu est satisfaisant ou non. Et le dernier peut ordonner les tests de validations des lots de produits et étudier l'impact d'une dérive de la distribution des facteurs sur le produit.

Une première étude permet ainsi de proposer un nouveau jeu de spécifications qui donne un rendement compatible avec les standards de l'industrie micro-électronique ; trois d'entre elles peuvent être plus restrictives sans que le rendement ne soit dégradé alors que l'une d'entre elle doit être allégée pour avoir un rendement non nul.

Dans une deuxième étude, il nous est possible de classer les 12 réponses étudiées suivant l'ordre de leur contribution au risque de défaillance. Cela peut permettre d'optimiser les tests de validation des lots en sortie d'usine.

Enfin, une troisième étude donne la possibilité de statuer sur un lot jugé non-conforme, dû à une variabilité anormale sur un ou plusieurs facteurs. Habituellement la décision relative à l'avenir de ce lot intervient suite à la réunion des différents acteurs de la réalisation et de la fabrication du produit. Il est en général décidé la mise en œuvre d'un ensemble de procédures de tests relativement lourdes. GoNoGo peut permettre un gain de temps considérable en proposant des abaques d'aide à la décision.

# Chapitre 1

# Modèles aléatoires

Dans le cadre de la modélisation d'une fonction à partir d'un ensemble restreint de données, nous avons retenu l'utilisation de processus gaussiens. La construction de ces modèles aléatoires repose sur le principe d'inférence bayésienne que nous rappelons dans la section 1.1. Nous y justifions également le choix de cette approche par le fait que les données traitées sont déterministes, pour lesquelles les modèles GP présentent quelques propriétés intéressantes.

La réalisation d'un champ aléatoire par inférence bayésienne s'effectuant à partir d'un prior, nous exposons quatre modèles aléatoires issus de quatre différents priors dans les sections 1.2, 1.3, 1.4 et 1.5.

Les relations entre les données sont prises en compte via l'utilisation d'une fonction de corrélation. La section 1.6 liste quelques fonctions de corrélation possibles ainsi que des méthodes pour les paramétrer.

Quelques exemples sont traités dans la section 1.7 pour illustrer les différences entre les modèles présentés.

Pour conclure ce chapitre, la section 1.8 abordera trois compléments possibles aux modèles aléatoires : introduction d'un bruit, transformation de la réponse et traitement d'une réponse multidimensionnelle.

## 1.1 Principe général de la construction d'un modèle aléatoire d'une fonction imparfaitement connue

### 1.1.1 Conditionnement et champ aléatoire

Le contexte général dans lequel nous nous plaçons pour la construction d'un modèle est celui de l'inférence bayésienne d'un champ aléatoire. Cela signifie que, après observation des données à notre disposition, nous procédons en deux étapes :

- choix d'un champ aléatoire a priori, donné par sa loi
- conditionnement de ce champ par l'information connue, de façon à obtenir le champ aléatoire a posteriori.

Rappelons qu'étant donnés un espace $X$ (qui sera pour nous une partie de l'espace $\mathbb{R}^D$) et un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$, un *champ aléatoire* réel défini sur $\Omega$ et indexé par $X$ est une application de $X$ dans l'ensemble des variables réelles définies sur $\Omega$ :

$$x \mapsto Y_x.$$

Cette application est notée $(Y_x)_{x \in X}$.

La *loi du champ aléatoire* est la donnée de la distribution de chacun des vecteurs aléatoires $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_r})$ pour tout choix d'une famille finie $(t_1, t_2, \ldots, t_r)$ de points de $X$. Le champ est dit gaussien si toutes ces lois fini-dimensionnelles sont des lois de Gauss. Nous rencontrerons également dans la suite de cet écrit le champ de Student pour lequel notre référence sera le livre de Kotz et Nadarajah (2004).

Rappelons la *formule fondamentale de calcul des densités de probabilités conditionnelles* ou *densités conditionnelles*. Soient $Z_1$ et $Z_2$ deux vecteurs aléatoires de dimensions respectives $r$ et $n$, définis sur le même espace probabilisé, et tels que le vecteur $(Z_1, Z_2)$ suit une loi de densité $f$ par rapport à la mesure de Lebesgue sur $\mathbb{R}^{r+n}$. La loi conditionnelle de $Z_1$ sachant $Z_2$ possède une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^r$ ; cette densité, appelée densité conditionnelle de $Z_1$ sachant $Z_2$, est donnée par la formule

$$z_1 \mapsto \frac{f(z_1, Z_2)}{\int f(z_1', Z_2)\ \mathrm{d}z_1'}.$$

Voici comment nous utiliserons ce formalisme : la loi de $(Z_1, Z_2)$ est donnée a priori par sa densité $f$, et $z_2$ est un vecteur fixé de $\mathbb{R}^n$ ; quelle est la loi du vecteur aléatoire $Z_1$ sachant que $Z_2 = z_2$ ? C'est la loi de densité

$$z_1 \mapsto \frac{f(z_1, z_2)}{\int f(z_1', z_2)\ \mathrm{d}z_1'}.$$

Cette méthode de conditionnement d'un champ aléatoire par inférence bayésienne, dans l'intention de modéliser une fonction inconnue, porte le nom de "kriging". Elle a été développée dans le domaine des géosciences par Matheron (1989) et étendue à un certain nombre de domaines dont quelques exemples ont été listés dans l'introduction, section 0.2.

La modélisation par processus gaussien a plusieurs propriétés intéressantes. L'interpolation respecte les valeurs exactes aux points d'observation, c'est-à-dire qu'il n'y a pas d'erreurs de prédiction (ou résidus) en ces points. C'est donc un outil approprié pour l'analyse d'expériences numériques (Sacks et al. (1989)), pour lesquelles il est difficile de donner un sens aux résidus observés. Un modèle GP peut être déterminé (en théorie) même pour de petits jeux de données, une propriété clé quand l'information est manquante, ce qui est une situation habituelle quand le nombre de facteurs ($D$) est grand. De plus, c'est un modèle très flexible, capable de décrire des surfaces de réponses non-continues aussi bien que non-différentiables. Cette caractéristique est particulièrement utile pour gérer les discontinuités des réponses qui peuvent survenir en raison du schéma de résolution numérique des codes informatiques, en particulier lorsque des algorithmes de maillage sont impliqués. Finalement, la nature probabiliste des prédictions peut être interprétée comme un modèle pour l'incertitude, un intervalle de confiance représentant un degré de confiance dans la prédiction réalisée.

Il existe plusieurs méthodes alternatives aux processus gaussiens comme par exemple la modélisation par des fonction "splines" (DiMatteo et al. (2001)) qui consiste à découper l'espace des facteurs $X$ et à considérer un modèle polynomial pour chacun des sous-espaces sous des conditions de continuités et de différentiabilités aux limites des sous-espaces. Il y a également la modélisation par des réseaux de neurones (Neal (1996)) qui formellement consiste à créer un automate, c'est-à-dire une liste d'états ainsi que des règles pour passer d'un état à un autre, capable d'apprendre, c'est-à-dire capable de faire évoluer ses états et ses règles en fonction des données qui le renseignent. Rasmussen et Williams (2006) consacrent un chapitre entier à la comparaison des modèles GP à des solutions alternatives.

### 1.1.2 Choix d'un prior

Nous considérerons dans un premier temps le prior le plus utilisé, le champ gaussien, dû à sa simplicité de manipulation d'un point de vue algébrique et à une bonne adéquation avec un certain nombre de phénomènes physiques observés. Le théorème central limite illustre la place prépondérante jouée par la loi de Gauss dans les phénomènes aléatoires complexes. La construction du champ gaussien, a posteriori, obtenu par conditionnement bayésien, est décrite dans la section 1.2.

La difficulté majeure et souvent négligée de cette méthode est le choix a priori des paramètres (moyenne, variance et fonction de corrélation dans le cas d'un prior gaussien), laissé à la responsabilité de l'utilisateur. Afin de réduire la nature arbitraire de ce choix, nous proposons de remplacer le champ gaussien par un mélange de champs gaussiens.

Pour cela, dans le but d'introduire le minimum d'information a priori lors de la construction du modèle aléatoire, nous supposons la moyenne aléatoire suivant une distribution de probabilité uniforme. Il vient, a posteriori, une distribution gaussienne décrite par le théorème 2. Par la suite, nous avons supposé la variance aléatoire suivant également une distribution de probabilité uniforme. Ce faisant, les vecteurs aléatoires a posteriori suivent une distribution proche d'une distribution de Student multivariée décrite par le théorème 4.

Comme nous l'avons annoncé dans l'introduction de cette thèse, nous avons constaté, à la suite de nos travaux, que ces modèles donnent des résultats équivalents à ceux obtenus avec une approche par modèle hiérarchique bayésien décrite, par exemple, par Sacks et al. (1989) et Santner et al. (2003). Cette approche consiste à considérer une hiérarchie d'aléas provenant de différents priors, par exemple, l'aléa d'un prior gaussien, puis celui d'un prior sur sa moyenne, et ensuite celui d'un autre prior sur les paramètres du prior de la moyenne.

Sacks et al. (1989) et Santner et al. (2003) ont choisi d'utiliser des priors dits impropres, c'est-à-dire tels que le modèle hiérarchique considéré, avant conditionnement, est décrit par une distribution dont sa densité voit son intégrale diverger. Ils appliquent, cependant, le conditionnement bayésien, en effectuant une manipulation formelle des écritures mathématiques. Le posterior obtenu est lui parfaitement décrit, c'est-à-dire qu'il suit une distribution de probabilité.

C'est là une différence clé vis-à-vis de notre méthode qui est correctement définie à tout moment de la construction du champ aléatoire, ce qui est plus satisfaisant d'un point de vue probabiliste. De plus, nous ne supposons pas de relation de hiérarchie entre les aléas intervenant dans les priors considérés dans ce manuscrit. Néanmoins, les modèles

a posteriori sont équivalents, ce qui illustre l'interprétation des priors impropres comme des limites de priors propres faite, par exemple, par DeGroot (1970), même si d'autres interprétations sont étudiées (voir Akaike (1980)).

En s'inspirant des travaux de Sacks et al. (1989) et Santner et al. (2003), nous avons développé un modèle mixant la régression linéaire et le conditionnement bayésien (théorème 7).

Dans les modèles aléatoires décrits par les théorèmes 4 et 7, des priors uniformes sur les paramètres d'échelle ont été choisis. D'autres priors auraient pu être pris. Santner et al. (2003) utilise notamment les priors de Jeffreys. Ils existent un certain nombre d'arguments qui peuvent être avancés pour justifier le choix de tel ou tel prior comme ceux présentés par Kass et Wasserman (1996) et Kato (2009). Nous justifions l'utilisation d'un prior uniforme par le fait que le posterior en dépend fortement dans le cas où le modèle est peu informé. Le prior entraîne ainsi nécessairement un biais sur l'estimation de la variance a posteriori. Un prior uniforme conduit à une sur-estimation de cette variance, voir Gelman (2006), ce qui est en adéquation avec notre prise en compte des incertitudes.

Notons que tout le long de ce chapitre, nous nous plaçons dans des situations génériques où les données ne conduisent pas à des distributions dégénérées, i.e. les matrices de covariances sont définies positives.

## 1.2 Champ Gaussien

Dans cette section, est rappelée la définition d'un champ gaussien conditionné dont nous exprimons la moyenne et la variance.

Un champ gaussien $(Y_x)_{x\in X}$ se caractérise par le fait que toutes les distributions de ses marginales de dimensions finies sont gaussiennes et par :
- les valeurs moyennes : $\mu(x) := \mathbb{E}(Y_x)$,
- les covariances : $\rho(x, x') := \operatorname{cov}(Y_x, Y_{x'})$.

**Théorème 1.** *Soient* $(\boldsymbol{x_i}) := (x_1, x_2, \ldots, x_n) \in X^n$ *et* $\boldsymbol{y} := (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$.
*La distribution conditionnelle d'un champ gaussien* $(Y_x)_{x\in X}$ *sachant* $(Y_{x_i} = y_i)_{1\leq i\leq n}$ *est également gaussienne avec :*
- *pour valeurs moyennes :*

$$\mathbb{E}\left(Y_x \mid (Y_{x_i} = y_i)_{1\leq i\leq n}\right) = \mu(x) + \boldsymbol{\rho}(\boldsymbol{x}, (\boldsymbol{x_i}))\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}((\boldsymbol{x_i})))^T$$

- *pour covariances :*

$$\operatorname{cov}\left(Y_x, Y_{x'} \mid (Y_{x_i} = y_i)_{1\leq i\leq n}\right) = \rho(x, x') - \boldsymbol{\rho}(\boldsymbol{x}, (\boldsymbol{x_i}))\boldsymbol{\Sigma}^{-1}\boldsymbol{\rho}(\boldsymbol{x'}, (\boldsymbol{x_i}))^T$$

*où* $\boldsymbol{\Sigma} := (\rho(x_i, x_j))_{1\leq i,j\leq n}$ *est la matrice de covariance aux points observés,* $\boldsymbol{\mu}((\boldsymbol{x_i})) := (\mu(x_1), \mu(x_2), \ldots, \mu(x_n))$ *est le vecteur des moyennes et* $\boldsymbol{\rho}(\boldsymbol{x}, (\boldsymbol{x_i})) := (\rho(x, x_1), \rho(x, x_2), \ldots, \rho(x, x_n))$ *est le vecteur de covariance.*

Ce théorème (qui est un résultat classique de la théorie des probabilités) se déduit de la proposition suivante.

**Proposition 1.** *Si les vecteurs aléatoires* $\boldsymbol{X_1}$ *et* $\boldsymbol{X_2}$ *ont pour loi conjointe la loi normale :*

$$\mathcal{N}\left(\left(\boldsymbol{\mu_1}, \boldsymbol{\mu_2}\right), \boldsymbol{\Psi}\right) \quad \textit{avec} \quad \boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Sigma_{11}} & \boldsymbol{\Sigma_{12}} \\ \boldsymbol{\Sigma_{21}} & \boldsymbol{\Sigma_{22}} \end{bmatrix}$$

*et si les matrices* $\boldsymbol{\Psi}$, $\boldsymbol{\Sigma_{11}}$ *et* $\boldsymbol{\Sigma_{22}}$ *sont non dégénérées, alors la loi conditionnelle de* $\boldsymbol{X_1}$ *sachant* $\boldsymbol{X_2}$ *est la loi normale :*

$$\mathcal{N}\left(\boldsymbol{\mu_1} + \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}}^{-1}\left(\boldsymbol{X_2} - \boldsymbol{\mu_2}\right)^T, \boldsymbol{\Sigma_{11}} - \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}}^{-1}\boldsymbol{\Sigma_{21}}\right).$$

*Démonstration.* On note $d_i$ la dimension du vecteur $\boldsymbol{X_i}$, $i = 1, 2$. La loi de $\boldsymbol{X_2}$ a pour densité, sur $\mathbb{R}^{d_2}$ :

$$f_2 : \boldsymbol{x_2} \mapsto \frac{1}{(2\pi)^{\frac{d_2}{2}} \left|\boldsymbol{\Sigma_{22}}\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\left(\boldsymbol{x_2} - \boldsymbol{\mu_2}\right)\boldsymbol{\Sigma_{22}}^{-1}\left(\boldsymbol{x_2} - \boldsymbol{\mu_2}\right)^T\right).$$

La loi conjointe des vecteurs aléatoires $\boldsymbol{X_1}$ et $\boldsymbol{X_2}$ a pour densité, sur $\mathbb{R}^d$, avec $d = d_1 + d_2$ :

$$f : \left(\boldsymbol{x_1}, \boldsymbol{x_2}\right) \mapsto \frac{1}{(2\pi)^{\frac{d}{2}} \left|\boldsymbol{\Psi}\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\left(\boldsymbol{x_1} - \boldsymbol{\mu_1}, \boldsymbol{x_2} - \boldsymbol{\mu_2}\right)\boldsymbol{\Psi}^{-1}\left(\boldsymbol{x_1} - \boldsymbol{\mu_1}, \boldsymbol{x_2} - \boldsymbol{\mu_2}\right)^T\right).$$

On a, par le lemme 1 (voir ci-après) :

$$\boldsymbol{\Psi}^{-1} = \begin{bmatrix} \boldsymbol{V}^{-1} & -\boldsymbol{V}^{-1}\boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}}^{-1} \\ -\boldsymbol{\Sigma_{22}}^{-1}\boldsymbol{\Sigma_{21}}\boldsymbol{V}^{-1} & \boldsymbol{\Sigma_{22}}^{-1} + \boldsymbol{\Sigma_{22}}^{-1}\boldsymbol{\Sigma_{21}}\boldsymbol{V}^{-1}\boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}}^{-1} \end{bmatrix}$$

avec :

$$\boldsymbol{V} = \boldsymbol{\Sigma_{11}} - \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}}^{-1}\boldsymbol{\Sigma_{21}}$$

d'où :

$$\begin{aligned} &\left(\boldsymbol{x_1} - \boldsymbol{\mu_1}, \boldsymbol{x_2} - \boldsymbol{\mu_2}\right)\boldsymbol{\Psi}^{-1}\left(\boldsymbol{x_1} - \boldsymbol{\mu_1}, \boldsymbol{x_2} - \boldsymbol{\mu_2}\right)^T \\ &\qquad = \left(\boldsymbol{x_2} - \boldsymbol{\mu_2}\right)\boldsymbol{\Sigma_{22}}^{-1}\left(\boldsymbol{x_2} - \boldsymbol{\mu_2}\right)^T + \left(\boldsymbol{x_1} - \boldsymbol{E}\right)\boldsymbol{V}^{-1}\left(\boldsymbol{x_1} - \boldsymbol{E}\right)^T \end{aligned}$$

avec :

$$\boldsymbol{E} = \boldsymbol{\mu_1} + \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}}^{-1}\left(\boldsymbol{x_2} - \boldsymbol{\mu_2}\right)^T.$$

De plus (lemme 1) :

$$\left|\boldsymbol{\Psi}\right| = \left|\boldsymbol{\Sigma_{22}}\right|\left|\boldsymbol{V}\right|.$$

En utilisant la formule fondamentale indiquant que la loi conditionnelle de $\boldsymbol{X_1}$ sachant $\boldsymbol{X_2}$ a pour densité

$$\boldsymbol{x_1} \mapsto \frac{f\left(\boldsymbol{x_1}, \boldsymbol{X_2}\right)}{f_2\left(\boldsymbol{X_2}\right)},$$

on en déduit la loi conditionnelle de $\boldsymbol{X_1}$ sachant $\boldsymbol{X_2}$ :

$$\frac{\mathcal{N}\left(\left(\boldsymbol{\mu_1}, \boldsymbol{\mu_2}\right), \begin{bmatrix} \boldsymbol{\Sigma_{11}} & \boldsymbol{\Sigma_{12}} \\ \boldsymbol{\Sigma_{21}} & \boldsymbol{\Sigma_{22}} \end{bmatrix}\right)}{\mathcal{N}\left(\boldsymbol{\mu_2}, \boldsymbol{\Sigma_{22}}\right)} = \mathcal{N}\left(\boldsymbol{\mu_1} + \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}}^{-1}(\boldsymbol{X_2} - \boldsymbol{\mu_2})^T, \boldsymbol{\Sigma_{11}} - \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}}^{-1}\boldsymbol{\Sigma_{21}}\right)$$

□

**Lemme 1.** *Soient $p$ et $q$ deux nombres entiers strictement positifs et $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ et $\boldsymbol{D}$ des matrices de types respectifs $(p,p)$, $(p,q)$, $(q,p)$ et $(q,q)$. On suppose que les matrices carrées $\boldsymbol{A}$ et $\boldsymbol{D}$ sont inversibles. Il y a alors équivalence entre les affirmations suivantes :*

*(i) La matrice $\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}$ est inversible,*

*(ii) La matrice $\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}$ est inversible,*

*(iii) La matrice $\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}$ est inversible,*

*(iv)*

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^{-1} = \begin{bmatrix} \left(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \\ -\boldsymbol{D}^{-1}\boldsymbol{C}\left(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)^{-1} & \left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \end{bmatrix}.$$

*De plus, quand ces conditions sont satisfaites, on a*

$$\boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} = \left(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)^{-1}\boldsymbol{B}\boldsymbol{D}^{-1}, \tag{1.1}$$

*et*

$$\left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} = \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{C}\left(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)^{-1}\boldsymbol{B}\boldsymbol{D}^{-1}. \tag{1.2}$$

*Enfin, on a toujours la formule suivante sur les déterminants :*

$$\begin{vmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{vmatrix} = |\boldsymbol{D}| \times \left|\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right|.$$

*Démonstration.* Par définition, la matrice $\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}$ est inversible si et seulement si il existe une matrice $\begin{bmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H} \end{bmatrix}$ telle que :

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}\begin{bmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H} \end{bmatrix} = \mathbf{Id_{p+q}}$$

Ce qui est équivalent au système suivant :

$$\begin{cases} \boldsymbol{A}\boldsymbol{E} + \boldsymbol{B}\boldsymbol{G} = \mathbf{Id_p} \\ \boldsymbol{C}\boldsymbol{E} + \boldsymbol{D}\boldsymbol{G} = \mathbf{0} \\ \boldsymbol{A}\boldsymbol{F} + \boldsymbol{B}\boldsymbol{H} = \mathbf{0} \\ \boldsymbol{C}\boldsymbol{F} + \boldsymbol{D}\boldsymbol{H} = \mathbf{Id_q} \end{cases} \equiv \begin{cases} \boldsymbol{A}\boldsymbol{E} + \boldsymbol{B}\boldsymbol{G} = \mathbf{Id_p} \\ \boldsymbol{G} = -\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{E} \\ \boldsymbol{F} = -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{H} \\ \boldsymbol{C}\boldsymbol{F} + \boldsymbol{D}\boldsymbol{H} = \mathbf{Id_q} \end{cases} \equiv \begin{cases} \left(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)\boldsymbol{E} = \mathbf{Id_p} \\ \boldsymbol{G} = -\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{E} \\ \boldsymbol{F} = -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{H} \\ \left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)\boldsymbol{H} = \mathbf{Id_q} \end{cases}$$

Cela revient à dire que $\boldsymbol{E}$ et $\boldsymbol{H}$ sont inversibles et que :

$$\begin{cases} \boldsymbol{E} = \left(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)^{-1} \\ \boldsymbol{G} = -\boldsymbol{D}^{-1}\boldsymbol{C}\left(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)^{-1} \\ \boldsymbol{F} = -\boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \\ \boldsymbol{H} = \left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \end{cases}$$

Démontrant ainsi les implications $(i) \Rightarrow (iv) \Rightarrow (i)$ et $(i) \Rightarrow (ii)$ et $(i) \Rightarrow (iii)$.

De plus, en développant :

$$\begin{bmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H} \end{bmatrix} \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \mathbf{Id_{p+q}}$$

on obtient le système suivant :

$$\begin{cases} \boldsymbol{EA} + \boldsymbol{FC} = \mathbf{Id_p} \\ \boldsymbol{GA} + \boldsymbol{HC} = \mathbf{0} \\ \boldsymbol{EB} + \boldsymbol{FD} = \mathbf{0} \\ \boldsymbol{GB} + \boldsymbol{HD} = \mathbf{Id_q} \end{cases}$$

qui permet de prouver les formules (1.1) et (1.2).

Pour compléter la démonstration, notons qu'il est possible de factoriser la matrice $\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}$ en un produit de deux matrices triangulaires. Soit la matrice $\begin{bmatrix} \boldsymbol{R} & \boldsymbol{S} \\ \boldsymbol{T} & \boldsymbol{U} \end{bmatrix}$ telle que :

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \mathbf{Id_p} & \boldsymbol{B} \\ \mathbf{0} & \boldsymbol{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{R} & \boldsymbol{S} \\ \boldsymbol{T} & \boldsymbol{U} \end{bmatrix}$$

En développant, on trouve le système :

$$\begin{cases} \boldsymbol{R} = \boldsymbol{A} - \boldsymbol{BD}^{-1}\boldsymbol{C} \\ \boldsymbol{S} = \mathbf{0} \\ \boldsymbol{T} = \boldsymbol{D}^{-1}\boldsymbol{C} \\ \boldsymbol{U} = \mathbf{Id_q} \end{cases}$$

On obtient :

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \mathbf{Id_p} & \boldsymbol{B} \\ \mathbf{0} & \boldsymbol{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{A} - \boldsymbol{BD}^{-1}\boldsymbol{C} & \mathbf{0} \\ \boldsymbol{D}^{-1}\boldsymbol{C} & \mathbf{Id_q} \end{bmatrix}.$$

Ce qui démontre l'implication $(ii) \Rightarrow (i)$ et donne la formule

$$\begin{vmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{vmatrix} = |\boldsymbol{D}| \times \left|\boldsymbol{A} - \boldsymbol{BD}^{-1}\boldsymbol{C}\right|.$$

L'implication $(iii) \Rightarrow (i)$ s'obtient de façon similaire. □

Une approche standard pour résoudre notre problème initial est d'utiliser le théorème 1 avec un champ gaussien donné a priori où la valeur moyenne $\mu := \mu(x)$ ne dépend pas de $x$ et la variance $\sigma^2 := \rho(x, x)$ non plus. Dans ce cas, nous noterons $k(x, x') = \rho(x, x')/\sigma^2$, le coefficient de corrélation de $Y_x$ et $Y_{x'}$. Dans ce cadre, nous obtenons le corollaire suivant.

**Corollaire 1.** *Soient* $(x_1, x_2, \ldots, x_n) \in X^n$ *et* $\boldsymbol{y} := (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$.
*La distribution conditionnelle du champ gaussien* $(Y_x)_{x \in X}$ *sachant* $(Y_{x_i} = y_i)_{1 \leq i \leq n}$ *est gaussienne avec :*

- *pour valeurs moyennes :*

$$\mathbb{E}\left(Y_x \mid (Y_{x_i} = y_i)_{1 \leq i \leq n}\right) = \mu + \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \mu\boldsymbol{1})^T$$

- *pour covariances :*

$$\operatorname{cov}\left(Y_x, Y_{x'} \mid (Y_{x_i} = y_i)_{1 \leq i \leq n}\right) = \sigma^2\left(k(x, x') - \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{x'})^T\right)$$

*où* $\boldsymbol{\Sigma} := (k(x_i, x_j))_{1 \leq i,j \leq n}$ *est la matrice de corrélation,* $\boldsymbol{k}(\boldsymbol{x}) := (k(x, x_1), k(x, x_2), \ldots, k(x, x_n))$ *est le vecteur de corrélation et* $\boldsymbol{1} := (1, 1, \ldots, 1) \in \mathbb{R}^n$.

## 1.3 Mélange de champs gaussiens avec moyenne aléatoire

Dans cette section, nous décrivons un modèle aléatoire construit à partir d'une distribution a priori qui est un mélange gaussien, avec pour but d'éviter de choisir a priori une valeur pour la moyenne. Pour une moyenne a priori aléatoire uniforme sur un grand intervalle, le champ aléatoire conditionné tend à être gaussien. Nous donnons des formules pour sa moyenne et sa variance. Ce faisant, nous reconnaissons les formules données par Sacks et al. (1989) et Santner et al. (2003).

Nous proposons donc de considérer a priori le champ aléatoire $Y_x = U + W_x$ où $U$ est une variable aléatoire réelle suivant une distribution uniforme sur un intervalle $[-m, m]$ et où $(W_x)_{x\in X}$ est un champ gaussien centré de variance constante. De plus, nous supposons que $U$ et $(W_x)_{x\in X}$ sont indépendants.

Le champ gaussien $(W_x)_{x\in X}$ est caractérisé par la variance $\sigma^2$ et la fonction de corrélation $k$ (rappelons que la moyenne est nulle). Pour tous $x, x' \in X$, $\operatorname{cov}(W_x, W_{x'}) = \sigma^2 k(x, x')$ et, pour tout $x \in X$, $k(x, x) = 1$.

Soient $(x_1, x_2, \ldots, x_n) \in X^n$ et $\boldsymbol{y} := (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$. Notons $\boldsymbol{\Sigma} := (k(x_i, x_j))_{1\le i,j\le n}$ la matrice des corrélations et $\boldsymbol{k}(\boldsymbol{x}) := (k(x, x_j))_{1\le j\le n}$ le vecteur de corrélation.

Sous ces hypothèses et notations, nous déterminons la distribution, conditionnellement à l'information $(Y_{x_i} = y_i)_{1\le i\le n}$, du champ aléatoire $(Y_x)_{x\in X}$ (théorème 2) et de la variable aléatoire $U$ (théorème 3)

Rappelons que nous nous plaçons dans la situation générique où les matrices de covariance sont non dégénérées.

### 1.3.1 Distribution conditionnelle du champ aléatoire

**Théorème 2.** *La distribution conditionnelle du champ aléatoire $(Y_x)_{x\in X}$ sachant que $(Y_{x_i} = y_i)_{1\le i\le n}$ est donnée par des formules explicites des densités des marginales de dimensions finies.*
*Quand le paramètre $m$ tend vers l'infini, ces distributions conditionnelles deviennent gaussiennes. En particulier, quand $m \to \infty$, la distribution conditionnelle de la variable aléatoire $Y_x$ devient gaussienne avec comme moyenne*

$$\mu + \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \mu\boldsymbol{1})^T \quad \textit{avec} \quad \mu := \frac{\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T} \tag{1.3}$$

*et variance*

$$\sigma^2\left(1 - \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{x})^T + \frac{\left(1 - \boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{x})^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right) \tag{1.4}$$

*où $\boldsymbol{1} = (1, 1, \ldots, 1) \in \mathbb{R}^n$.*

*Remarque.* La moyenne $\mu + \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \mu\boldsymbol{1})^T$ est une fonction linéaire du vecteur $\boldsymbol{y}$. En effet, on peut également l'exprimer de la manière suivante :

$$\left(\left(1 - \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)\frac{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T} + \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{y}^T.$$

L'intérêt de l'expression (1.3) est de pouvoir la comparer plus directement à la moyenne du champ gaussien conditionné (section 1.2). De plus, la formule donnant $\mu$ pourra être comparée à la formule donnant la moyenne du champ aléatoire a priori gaussien obtenue par la méthode du maximum de vraisemblance (1.6.2).

*Démonstration.*

**Partie 1 : Distribution conditionnelle multivariée**

Soit $r$ un entier positif et $(t_1, \ldots, t_r) \in X^r$. Nous avons :

$$(Y_{t_1}, \ldots, Y_{t_r}, Y_{x_1}, \ldots, Y_{x_n}) = (U, \ldots, U) + (W_{t_1}, \ldots, W_{t_r}, W_{x_1}, \ldots, W_{x_n})$$

et $(W_{t_1}, \ldots, W_{t_r}, W_{x_1}, \ldots, W_{x_n})$ suit la distribution $\mathcal{N}(0, \boldsymbol{\Delta})$ avec

$$\boldsymbol{\Delta} := \sigma^2 \begin{pmatrix} \boldsymbol{\Sigma_2} & \boldsymbol{k}(\boldsymbol{t}) \\ \boldsymbol{k}(\boldsymbol{t})^T & \boldsymbol{\Sigma} \end{pmatrix}, \text{ une matrice définie positive}$$

où $\boldsymbol{k}(\boldsymbol{t}) := (k(t_i, x_j))_{1 \le i \le r, 1 \le j \le n}$ et $\boldsymbol{\Sigma_2} := (k(t_i, t_j))_{1 \le i \le r, 1 \le j \le r}$.
Notons $f$, la densité du vecteur aléatoire $(Y_{t_1}, \ldots, Y_{t_r}, Y_{x_1}, \ldots, Y_{x_n})$ :

$$f(\boldsymbol{\zeta}) = \frac{1}{2m} \int_{-m}^{m} \frac{1}{\left(\sqrt{2\pi}\right)^{n+r} \sqrt{|\boldsymbol{\Delta}|}} \exp\left(-\frac{1}{2} (\boldsymbol{\zeta} - \boldsymbol{u}) \boldsymbol{\Delta}^{-1} (\boldsymbol{\zeta} - \boldsymbol{u})^T\right) \mathrm{d}u$$

où $\boldsymbol{\zeta} := (y_{t_1}, y_{t_2}, \ldots, y_{t_r}, y_{x_1}, y_{x_2}, \ldots, y_{x_n})$ et $\boldsymbol{u} := (u, u, \ldots, u) \in \mathbb{R}^{n+r}$.

La densité conditionnelle de $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_r})$ sachant $(Y_{x_i} = y_i)_{1 \le i \le n}$ est

$$\boldsymbol{z} := (y_{t_1}, y_{t_2}, \ldots, y_{t_r}) \mapsto \frac{f(\boldsymbol{z}, \boldsymbol{y})}{\int_{\mathbb{R}^r} f(\boldsymbol{z'}, \boldsymbol{y}) \ \mathrm{d}\boldsymbol{z'}} =: g(\boldsymbol{z}).$$

Après simplification, nous obtenons :

$$g(\boldsymbol{z}) = \frac{\int_{-m}^{m} \exp\left(-\frac{1}{2} \left(\begin{pmatrix}\boldsymbol{z} & \boldsymbol{y}\end{pmatrix} - \boldsymbol{u}\right) \boldsymbol{\Delta}^{-1} \left(\begin{pmatrix}\boldsymbol{z} & \boldsymbol{y}\end{pmatrix} - \boldsymbol{u}\right)^T\right) \mathrm{d}u}{\int_{\mathbb{R}^r} \int_{-m}^{m} \exp\left(-\frac{1}{2} \left(\begin{pmatrix}\boldsymbol{z'} & \boldsymbol{y}\end{pmatrix} - \boldsymbol{u}\right) \boldsymbol{\Delta}^{-1} \left(\begin{pmatrix}\boldsymbol{z'} & \boldsymbol{y}\end{pmatrix} - \boldsymbol{u}\right)^T\right) \mathrm{d}\boldsymbol{z'} \ \mathrm{d}u}.$$

Notons

$$\mathrm{NUM}(\boldsymbol{z}) := \int_{-m}^{m} \exp\left(-\frac{1}{2} \left(\begin{pmatrix}\boldsymbol{z} & \boldsymbol{y}\end{pmatrix} - \boldsymbol{u}\right) \boldsymbol{\Delta}^{-1} \left(\begin{pmatrix}\boldsymbol{z} & \boldsymbol{y}\end{pmatrix} - \boldsymbol{u}\right)^T\right) \mathrm{d}u.$$

Considérons le polynôme en la variable $u$ suivant :

$$\left(\begin{pmatrix}\boldsymbol{z} & \boldsymbol{y}\end{pmatrix} - \boldsymbol{u}\right) \boldsymbol{\Delta}^{-1} \left(\begin{pmatrix}\boldsymbol{z} & \boldsymbol{y}\end{pmatrix} - \boldsymbol{u}\right)^T = au^2 + 2bu + c.$$

Notons que le polynôme $au^2 + 2bu + c$ est toujours strictement positif puisque nous avons exclu les situations dégénérées. Donc son discriminant est strictement négatif, c'est-à-dire $\left(c - b^2/a\right) > 0$.
Identifions $a$, $b$ et $c$ :

$$\begin{aligned} a &= \boldsymbol{1} \boldsymbol{\Delta}^{-1} \boldsymbol{1}^T \\ b &= -\boldsymbol{1} \boldsymbol{\Delta}^{-1} \begin{pmatrix}\boldsymbol{z} & \boldsymbol{y}\end{pmatrix}^T \\ c &= \begin{pmatrix}\boldsymbol{z} & \boldsymbol{y}\end{pmatrix} \boldsymbol{\Delta}^{-1} \begin{pmatrix}\boldsymbol{z} & \boldsymbol{y}\end{pmatrix}^T \end{aligned}$$

où $\boldsymbol{1} = (1, 1, \ldots, 1)$ est ici de dimension $n + r$.
Rappelons que :

$$\int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(at^2 + 2bt + c\right)\right) \, \mathrm{d}t = \frac{\sqrt{2\pi}}{\sqrt{a}} \exp\left(-\frac{1}{2}\left(c - \frac{b^2}{a}\right)\right).$$

Nous obtenons :

$$\mathrm{NUM}\left(\boldsymbol{z}\right) \underset{m \to +\infty}{\longrightarrow} \frac{\sqrt{2\pi}}{\sqrt{a}} \exp\left(-\frac{1}{2}\left(c - \frac{b^2}{a}\right)\right).$$

C'est-à-dire :

$$\mathrm{NUM}\left(\boldsymbol{z}\right) \underset{m \to +\infty}{\longrightarrow} \frac{\sqrt{2\pi}}{\sqrt{\boldsymbol{1}\boldsymbol{\Delta}^{-1}\boldsymbol{1}^T}} \exp\left(-\frac{1}{2}\left(\begin{pmatrix}\boldsymbol{z} & \boldsymbol{y}\end{pmatrix}\boldsymbol{\Delta}^{-1}\begin{pmatrix}\boldsymbol{z} & \boldsymbol{y}\end{pmatrix}^T - \frac{\left(\boldsymbol{1}\boldsymbol{\Delta}^{-1}\begin{pmatrix}\boldsymbol{z} & \boldsymbol{y}\end{pmatrix}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Delta}^{-1}\boldsymbol{1}^T}\right)\right).$$

Par convergence monotone, l'intégrale de $\mathrm{NUM}\left(\boldsymbol{z}\right)$ tend vers l'intégrale de sa limite quand $m$ tend vers l'infini. On obtient donc une expression de la limite de la densité $g$ :

$$g\left(\boldsymbol{z}\right) \underset{m \to +\infty}{\longrightarrow} \frac{\exp\left(-\frac{1}{2}\left(\begin{pmatrix}\boldsymbol{z} & \boldsymbol{y}\end{pmatrix}\boldsymbol{\Delta}^{-1}\begin{pmatrix}\boldsymbol{z} & \boldsymbol{y}\end{pmatrix}^T - \frac{\left(\boldsymbol{1}\boldsymbol{\Delta}^{-1}\begin{pmatrix}\boldsymbol{z} & \boldsymbol{y}\end{pmatrix}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Delta}^{-1}\boldsymbol{1}^T}\right)\right)}{\int_{\mathbb{R}^r} \exp\left(-\frac{1}{2}\left(\begin{pmatrix}\boldsymbol{z}' & \boldsymbol{y}\end{pmatrix}\boldsymbol{\Delta}^{-1}\begin{pmatrix}\boldsymbol{z}' & \boldsymbol{y}\end{pmatrix}^T - \frac{\left(\boldsymbol{1}\boldsymbol{\Delta}^{-1}\begin{pmatrix}\boldsymbol{z}' & \boldsymbol{y}\end{pmatrix}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Delta}^{-1}\boldsymbol{1}^T}\right)\right) \mathrm{d}\boldsymbol{z}'}.$$

(Il peut être établi que cette convergence est uniforme en la variable $\boldsymbol{z}$.)
Dans l'exponentielle du numérateur de cette expression, nous observons un polynôme du second degré non-négatif en la variable $\boldsymbol{z}$. Nous reconnaissons une distribution gaussienne (de dimension $r$).
Dans la situation limite $m \to \infty$, la distribution conditionnelle du champ aléatoire est alors bien définie et est gaussienne.

**Partie 2 : Distribution conditionnelle univariée**

Plaçons-nous maintenant dans le cas univarié. La variable $z$ est alors de dimension $r = 1$ et il existe des nombres $v$, $l$ et $p$ tels que :

$$\begin{pmatrix}z & \boldsymbol{y}\end{pmatrix}\boldsymbol{\Delta}^{-1}\begin{pmatrix}z & \boldsymbol{y}\end{pmatrix}^T - \frac{\left(\boldsymbol{1}\boldsymbol{\Delta}^{-1}\begin{pmatrix}z & \boldsymbol{y}\end{pmatrix}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Delta}^{-1}\boldsymbol{1}^T} = \frac{1}{v}\left(z - l\right)^2 + p.$$

Ce qui donne :

$$g\left(\boldsymbol{z}\right) \underset{m \to +\infty}{\longrightarrow} \frac{\exp\left(-\frac{1}{2}p\right)\exp\left(-\frac{1}{2v}\left(z - l\right)^2\right)}{\int_{\mathbb{R}} \exp\left(-\frac{1}{2}p\right)\exp\left(-\frac{1}{2v}\left(z' - l\right)^2\right) \mathrm{d}z'} = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{1}{2v}\left(z - l\right)^2\right).$$

D'après le lemme 1 de la section 1.2 :

$$\sigma^2\boldsymbol{\Delta}^{-1} = \lambda^{-1}\begin{pmatrix}1 & -\boldsymbol{k}\left(\boldsymbol{t}\right)\boldsymbol{\Sigma}^{-1} \\ -\boldsymbol{\Sigma}^{-1}\boldsymbol{k}\left(\boldsymbol{t}\right)^T & \lambda\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\boldsymbol{k}\left(\boldsymbol{t}\right)^T\boldsymbol{k}\left(\boldsymbol{t}\right)\boldsymbol{\Sigma}^{-1}\end{pmatrix}$$

avec $\lambda := 1 - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{t})^T$. Ainsi, en identifiant les quantités $l$ et $v$, on montre qu'au point $x$ de l'espace $X$, la marginale unidimensionnelle de la distribution gaussienne du champ a pour moyenne :

$$\mu + \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \mu\boldsymbol{1})^T \quad \text{avec} \quad \mu := \frac{\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}$$

et variance :

$$\sigma^2\left(1 - \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{x})^T + \frac{\left(1 - \boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{x})^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right).$$

□

### 1.3.2 Distribution conditionnelle de la moyenne aléatoire

Bien que nous n'en ayons pas l'usage dans les prochains chapitres, nous nous sommes intéressés à la description de la loi conditionnelle de $U$. En effet, ayant défini un prior pour la moyenne du champ aléatoire, il nous est possible d'effectuer de l'inférence bayésienne sur ce paramètre. Nous remarquerons que le mode de la distribution conditionnelle de $U$ est égal à l'estimateur du maximum de vraisemblance (voir section 1.6.2 et le tableau récapitulatif 1.5).

**Théorème 3.** *Pour $m$ fixé, la distribution conditionnelle de la variable aléatoire $U$ sachant que $(Y_{x_i} = y_i)_{1\leq i\leq n}$ est celle d'une gaussienne restreinte à $[-m, m]$.*
*Quand $m$ tend vers l'infini, la distribution conditionnelle de la variable aléatoire $U$ est gaussienne de moyenne :*

$$\frac{\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}$$

*et de variance :*

$$\frac{\sigma^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}$$

*où $\boldsymbol{1} = (1, 1, \ldots, 1) \in \mathbb{R}^n$.*

*Démonstration.* Nous avons :

$$(Y_{x_1}, \ldots, Y_{x_n}) = (U, \ldots, U) + (W_{x_1}, \ldots, W_{x_n})$$

où $(W_{x_1}, \ldots, W_{x_n})$ suit la distribution $\mathcal{N}(0, \boldsymbol{\Delta})$ avec $\boldsymbol{\Delta} := \sigma^2\boldsymbol{\Sigma}$.
Notons $f$ la densité du vecteur aléatoire $(U, Y_{x_1}, \ldots, Y_{x_n})$ :

$$f(u, \boldsymbol{\zeta}) = \frac{1}{2m}\frac{1}{\left(\sqrt{2\pi}\right)^n\sigma^n\sqrt{|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\zeta} - \boldsymbol{u})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\zeta} - \boldsymbol{u})^T\right)\mathbb{1}_{[-m,m]}(u)$$

avec $\boldsymbol{\zeta} := (y_{x_1}, y_{x_2}, \ldots, y_{x_n})$ et $\boldsymbol{u} := (u, u, \ldots, u) \in \mathbb{R}^n$.

La densité conditionnelle de $U$ sachant que $(Y_{x_i} = y_i)_{1\leq i\leq n}$ est :

$$u \mapsto \frac{f(u, \boldsymbol{y})}{\int_{-m}^{m} f(u', \boldsymbol{y})\ \mathrm{d}u'} =: g(u).$$

Après simplification, nous obtenons :

$$g(u) = \frac{\exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{u})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{u})^T\right)}{\int_{-m}^{m}\exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{u'})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{u'})^T\right)\,\mathrm{d}u'}\,\mathbb{1}_{[-m,m]}(u).$$

Nous observons que pour $m$ fixé, la distribution conditionnelle de la variable aléatoire $U$ est une gaussienne restreinte à $[-m, m]$.
Notons :

$$\mathrm{NUM}(u) := \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{u})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{u})^T\right).$$

Remarquons que le terme dans l'exponentielle peut être développé comme suit :

$$(\boldsymbol{y}-\boldsymbol{u})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{u})^T = au^2 + 2bu + c = a\left(u+\frac{b}{a}\right)^2 + c - \frac{b^2}{a}.$$

avec

$$\begin{aligned} a &= \boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T \\ b &= -\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T \\ c &= \boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T \end{aligned}$$

On obtient :

$$\mathrm{NUM}(u) := \exp\left(-\frac{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{2\sigma^2}\left(u-\frac{\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)^2\right)\exp\left(-\frac{1}{2\sigma^2}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T - \frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)\right).$$

Notons :

$$\mathrm{DEN} := \int_{-m}^{m}\exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{u})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{u})^T\right)\,\mathrm{d}u$$

et rappelons que :

$$\int_{-m}^{m}\exp\left(-\frac{1}{2\sigma^2}\left(au^2+2bu+c\right)\right)\,\mathrm{d}u \underset{m\to+\infty}{\longrightarrow} \frac{\sigma\sqrt{2\pi}}{\sqrt{a}}\exp\left(-\frac{1}{2\sigma^2}\left(c-\frac{b^2}{a}\right)\right).$$

Nous pouvons en déduire en utilisant le résultat ci-dessus :

$$\mathrm{DEN} \underset{m\to\infty}{\longrightarrow} \frac{\sigma\sqrt{2\pi}}{\sqrt{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}}\exp\left(-\frac{1}{2\sigma^2}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T - \frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)\right).$$

Nous avons ainsi :

$$g(u) \underset{m\to\infty}{\longrightarrow} \frac{\sqrt{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}}{\sigma\sqrt{2\pi}}\exp\left(-\frac{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{2\sigma^2}\left(u-\frac{\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)^2\right).$$

Nous reconnaissons que la distribution conditionnelle limite de la variable aléatoire $U$ est une distribution gaussienne de moyenne :

$$\frac{\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}$$

et de variance :

$$\frac{\sigma^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}.$$

□

## 1.4 Mélange de champs gaussiens avec moyenne et variance aléatoires

Dans cette section, nous présentons une distribution a priori, composée d'un mélange gaussien avec pour but d'éviter le choix a priori de la moyenne et la variance. Nous observons qu'avec des priors uniformes sur des grands intervalles, la distribution du champ aléatoire conditionné tend vers une distribution de Student multivariée dont nous donnons les formules pour ses paramètres de position et d'échelle.

Considérons le champ aléatoire a priori $Y_x = U + VW_x$, où $U$ est une variable aléatoire réelle suivant une distribution uniforme sur un intervalle $[-m, m]$, où $V$ est une variable aléatoire réelle positive suivant une distribution uniforme sur un intervalle $[\epsilon, 1/\epsilon]$ et où $(W_x)_{x\in X}$ est un champ gaussien centré réduit. De plus, nous supposons que $U$, $V$ et $(W_x)_{x\in X}$ sont indépendants.

Le paramètre caractérisant le champ gaussien $(W_x)_{x\in X}$ est la fonction de corrélation $k$ (rappelons que la moyenne est nulle et la variance vaut 1). Nous supposons que $n \geq 3$. Soient $(x_1, x_2, \ldots, x_n) \in X^n$ et $\boldsymbol{y} := (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$. Notons $\boldsymbol{\Sigma} := (k(x_i, x_j))_{1\leq i,j\leq n}$ la matrice de corrélation définie positive et $\boldsymbol{k}(\boldsymbol{x}) := (k(x, x_j))_{1\leq j\leq n}$ le vecteur de corrélation.

Sous ces hypothèses et notations, nous déterminons la distribution, conditionnellement à l'information $(Y_{x_i} = y_i)_{1\leq i\leq n}$, du champ aléatoire $(Y_x)_{x\in X}$ (théorème 4), de la variable aléatoire $U$ (théorème 5) et de la variable aléatoire $V$ (théorème 6).

Dans un premier temps, rappelons la définition de la distribution de Student multivariée.

### 1.4.1 Distribution de Student multivariée

Nous nous référons au livre de Kotz et Nadarajah (2004). Une distribution de Student $p$-variée de degré $\nu$ (ou "multivariate t-distribution" en anglais) a une densité de la forme :

$$\boldsymbol{t} \mapsto \frac{1}{(\sqrt{\pi\nu})^p \sqrt{|\boldsymbol{\Sigma}|}} \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{1}{\nu}(\boldsymbol{t}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{t}-\boldsymbol{\mu})^T\right)^{-\frac{\nu+p}{2}} \qquad (\boldsymbol{t} \in \mathbb{R}^p)$$

où $\boldsymbol{\Sigma}$ est la matrice $p \times p$ définie positive des paramètres d'échelles et $\boldsymbol{\mu}$ est le vecteur $1 \times p$ des paramètres de position.

Pour $\nu > 1$, le vecteur moyenne de la distribution de Student est correctement défini et est égal à $\boldsymbol{\mu}$. Pour $\nu > 2$, la matrice de covariance de la distribution de Student est correctement définie et est égale à $\frac{\nu}{\nu-2}\boldsymbol{\Sigma}$.

C'est une généralisation de la distribution de Student. Quand $\nu = 1$, c'est une distribution multivariée de Cauchy. Quand $\nu$ tend vers l'infini, la distribution tend vers une distribution gaussienne multivariée.

### 1.4.2 Distribution conditionnelle du champ aléatoire

**Théorème 4.** *La distribution conditionnelle du champ aléatoire $(Y_x)_{x\in X}$ sachant que $(Y_{x_i} = y_i)_{1\leq i\leq n}$ est donnée par des formules explicites des marginales de dimensions fi-*

*nies.*
*Quand le paramètre $m$ tend vers l'infini et $\epsilon$ tend vers zéro, pour $n > 2$, ces distributions conditionnelles deviennent des distributions de Student multivariées.*
*En particulier, quand $m \to \infty$ et $\epsilon \to 0$, pour $n > 2$, la distribution conditionnelle de la variable aléatoire $Y_x$ devient une distribution de Student de degré $n-2$, ayant pour paramètre de position :*

$$\mu + \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \mu\boldsymbol{1})^T \quad \textit{avec } \mu := \frac{\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}$$

*et pour paramètre d'échelle :*

$$\sqrt{\frac{1}{n-2}\left((\boldsymbol{y} - \mu\boldsymbol{1})\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T\right)\left(1 - \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{x})^T + \frac{\left(1 - \boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{x})^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)}$$

*où $\boldsymbol{1} = (1, 1, \ldots, 1) \in \mathbb{R}^n$.*

*Remarque.* L'expression $(\boldsymbol{y} - \mu\boldsymbol{1})\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T$ peut également s'écrire $(\boldsymbol{y} - \mu\boldsymbol{1})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \mu\boldsymbol{1})^T$. Cette formule pourra être comparée à la formule donnant la variance du champ aléatoire a priori gaussien, obtenue par la méthode du maximum de vraisemblance (1.6.2).

*Démonstration.*

**Partie 1 : Distribution conditionnelle multivariée**

Soit $r$ un entier positif et soit $(t_1, \ldots, t_r) \in X^r$.
Nous avons

$$(Y_{t_1}, \ldots, Y_{t_r}, Y_{x_1}, \ldots, Y_{x_n}) = (U, \ldots, U) + V(W_{t_1}, \ldots, W_{t_r}, W_{x_1}, \ldots, W_{x_n})$$

où $(W_{t_1}, \ldots, W_{t_r}, W_{x_1}, \ldots, W_{x_n})$ suit la distribution $\mathcal{N}(0, \boldsymbol{\Delta})$ avec :

$$\boldsymbol{\Delta} := \begin{pmatrix} \boldsymbol{\Sigma_2} & \boldsymbol{k}(\boldsymbol{t}) \\ \boldsymbol{k}(\boldsymbol{t})^T & \boldsymbol{\Sigma} \end{pmatrix}, \text{ une matrice définie positive}$$

où $\boldsymbol{k}(\boldsymbol{t}) := (k(t_i, x_j))_{1 \le i \le r, 1 \le j \le n}$ et $\boldsymbol{\Sigma_2} := (k(t_i, t_j))_{1 \le i \le r, 1 \le j \le r}$. (Rappelons que nous décrivons une situation générique.)
Notons $f$ la densité du vecteur aléatoire $(Y_{t_1}, \ldots, Y_{t_r}, Y_{x_1}, \ldots, Y_{x_n})$ :

$$f(\boldsymbol{\zeta}) = \frac{1}{2m}\frac{1}{\frac{1}{\epsilon} - \epsilon}\int_{-m}^{m}\int_{\epsilon}^{\frac{1}{\epsilon}} \frac{1}{\left(\sqrt{2\pi}\right)^{n+r} v^{n+r}\sqrt{|\boldsymbol{\Delta}|}} \exp\left(-\frac{1}{2v^2}(\boldsymbol{\zeta} - \boldsymbol{u})\boldsymbol{\Delta}^{-1}(\boldsymbol{\zeta} - \boldsymbol{u})^T\right) \mathrm{d}u\, \mathrm{d}v$$

avec $\boldsymbol{\zeta} := (y_{t_1}, y_{t_2}, \ldots, y_{t_r}, y_{x_1}, y_{x_2}, \ldots, y_{x_n})$ et $\boldsymbol{u} := (u, u, \ldots, u) \in \mathbb{R}^{n+r}$.

La distribution conditionnelle de $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_r})$ sachant $(Y_{x_i} = y_i)_{1 \le i \le n}$ est :

$$\boldsymbol{z} := (y_{t_1}, y_{t_2}, \ldots, y_{t_r}) \mapsto \frac{f(\boldsymbol{z}, \boldsymbol{y})}{\int_{\mathbb{R}^r} f(\boldsymbol{z'}, \boldsymbol{y})\, \mathrm{d}\boldsymbol{z'}} =: g(\boldsymbol{z}).$$

Après simplification, nous obtenons :

$$g(\boldsymbol{z}) = \frac{\int_{-m}^{m}\int_{\epsilon}^{\frac{1}{\epsilon}} \frac{1}{v^{n+r}} \exp\left(-\frac{1}{2v^2}\left(\left(\boldsymbol{z} \quad \boldsymbol{y}\right) - \boldsymbol{u}\right)\boldsymbol{\Delta}^{-1}\left(\left(\boldsymbol{z} \quad \boldsymbol{y}\right) - \boldsymbol{u}\right)^T\right) \mathrm{d}u\ \mathrm{d}v}{\int_{\mathbb{R}^r}\int_{-m}^{m}\int_{\epsilon}^{\frac{1}{\epsilon}} \frac{1}{v^{n+r}} \exp\left(-\frac{1}{2v^2}\left(\left(\boldsymbol{z'} \quad \boldsymbol{y}\right) - \boldsymbol{u}\right)\boldsymbol{\Delta}^{-1}\left(\left(\boldsymbol{z'} \quad \boldsymbol{y}\right) - \boldsymbol{u}\right)^T\right) \mathrm{d}\boldsymbol{z'}\ \mathrm{d}u\ \mathrm{d}v}.$$

Notons :

$$\mathrm{NUM}(\boldsymbol{z}) := \int_{-m}^{m}\int_{\epsilon}^{\frac{1}{\epsilon}} \frac{1}{v^{n+r}} \exp\left(-\frac{1}{2v^2}\left(\left(\boldsymbol{z} \quad \boldsymbol{y}\right) - \boldsymbol{u}\right)\boldsymbol{\Delta}^{-1}\left(\left(\boldsymbol{z} \quad \boldsymbol{y}\right) - \boldsymbol{u}\right)^T\right) \mathrm{d}u\ \mathrm{d}v.$$

Considérons le polynôme en la variable $u$ suivant :

$$\left(\left(\boldsymbol{z} \quad \boldsymbol{y}\right) - \boldsymbol{u}\right)\boldsymbol{\Delta}^{-1}\left(\left(\boldsymbol{z} \quad \boldsymbol{y}\right) - \boldsymbol{u}\right)^T = au^2 + 2bu + c.$$

Rappelons que :

$$\int_{-m}^{m} \exp\left(-\frac{1}{2v^2}\left(au^2 + 2bu + c\right)\right) \mathrm{d}u \underset{m\to+\infty}{\longrightarrow} \frac{v\sqrt{2\pi}}{\sqrt{a}} \exp\left(-\frac{1}{2v^2}\left(c - \frac{b^2}{a}\right)\right).$$

Ce qui donne, par le théorème de convergence monotone :

$$\mathrm{NUM}(\boldsymbol{z}) \underset{m\to+\infty}{\longrightarrow} \int_{\epsilon}^{\frac{1}{\epsilon}} \frac{\sqrt{2\pi}}{v^{n+r-1}\sqrt{a}} \exp\left(-\frac{1}{2v^2}\left(c - \frac{b^2}{a}\right)\right) \mathrm{d}v.$$

Avec le changement de variable $q = \frac{1}{2v^2}\left(c - \frac{b^2}{a}\right)$ et la définition de la fonction d'Euler $\Gamma$, nous obtenons :

$$\mathrm{NUM}(\boldsymbol{z}) \underset{m\to+\infty,\epsilon\to 0}{\longrightarrow} \frac{\sqrt{2\pi}}{\sqrt{a}} 2^{\frac{n+r-4}{2}} \left(c - \frac{b^2}{a}\right)^{-\frac{n+r-2}{2}} \Gamma\left(\frac{n+r-2}{2}\right).$$

Notons que le polynôme $au^2 + 2bu + c$ est toujours strictement positif puisque nous avons exclu les situations dégénérées. Donc son discriminant est strictement négatif, c'est-à-dire $\left(c - b^2/a\right) > 0$. De plus, l'expression $\left(c - b^2/a\right)$ est un polynôme du second degré en la variable $\boldsymbol{z}$, et nous écrivons

$$c - \frac{b^2}{a} = Q(\boldsymbol{z}) + L(\boldsymbol{z}) + M$$

où $Q$ est une forme quadratique, $L$ est une forme linéaire et $M$ est constant.
Par convergence monotone, l'intégrale de $\mathrm{NUM}(\boldsymbol{z})$ tend vers l'intégrale de sa limite quand $m$ tend vers l'infini. On obtient donc une expression de la limite de la densité $g$ :

$$g(\boldsymbol{z}) \underset{m\to+\infty,\epsilon\to 0}{\longrightarrow} \frac{\frac{\sqrt{2\pi}}{\sqrt{a}} 2^{\frac{n+r-4}{2}} \left(Q(\boldsymbol{z}) + L(\boldsymbol{z}) + M\right)^{-\frac{n+r-2}{2}} \Gamma\left(\frac{n+r-2}{2}\right)}{\int_{\mathbb{R}^r} \frac{\sqrt{2\pi}}{\sqrt{a}} 2^{\frac{n+r-4}{2}} \left(Q(\boldsymbol{z'}) + L(\boldsymbol{z'}) + M\right)^{-\frac{n+r-2}{2}} \Gamma\left(\frac{n+r-2}{2}\right) \mathrm{d}\boldsymbol{z'}},$$

c'est-à-dire :

$$g(\boldsymbol{z}) \underset{m\to+\infty,\epsilon\to 0}{\longrightarrow} \frac{\left(Q(\boldsymbol{z}) + L(\boldsymbol{z}) + M\right)^{-\frac{n+r-2}{2}}}{\int_{\mathbb{R}^r} \left(Q(\boldsymbol{z'}) + L(\boldsymbol{z'}) + M\right)^{-\frac{n+r-2}{2}} \mathrm{d}\boldsymbol{z'}}.$$

Comme $Q(\boldsymbol{z}) + L(\boldsymbol{z}) + M$ est un polynôme positif, avec le lemme 2 où $\ell = 1$, nous avons, pour $n > 2$ :

$$\int_{\mathbb{R}^r} \left(Q(\boldsymbol{z}') + L(\boldsymbol{z}') + M\right)^{-\frac{n+r-2}{2}} \, \mathrm{d}\boldsymbol{z}' < +\infty.$$

Puis, à l'aide du lemme 3, avec $\ell = 1$, nous concluons que $g(\boldsymbol{z})$ est une distribution de Student multivariée de degré $n-2$, quand $m$ tend vers l'infini et $\epsilon$ tend vers zéro.

**Partie 2 : Distribution conditionnelle univariée**

Dans le cas univarié ($r = 1$), nous avons :

$$g(z) \underset{m\to+\infty,\epsilon\to 0}{\longrightarrow} \frac{\left(Q(z) + L(z) + M\right)^{-\frac{n-1}{2}}}{\int_{\mathbb{R}} \left(Q(z') + L(z') + M\right)^{-\frac{n-1}{2}} \, \mathrm{d}z'}.$$

Revenons sur les définitions de $a$, $b$ et $c$. Rappelons que, d'après le lemme 1 :

$$\boldsymbol{\Delta}^{-1} = \lambda^{-1} \begin{pmatrix} 1 & -\boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1} \\ -\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{t})^T & \lambda\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{t})^T\boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1} \end{pmatrix}$$

avec $\lambda := 1 - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{t})^T$. Alors :

$$\begin{aligned} a &= \boldsymbol{1}\boldsymbol{\Delta}^{-1}\boldsymbol{1}^T \\ b &= \lambda^{-1}\left(z - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T\right)\left(1 - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right) + \boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T \\ c &= \lambda^{-1}\left(z - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T\right)^2 + \boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T \end{aligned}$$

avec $\boldsymbol{1} = (1, 1, \ldots, 1) \in \mathbb{R}^n$. Soit $\tilde{z} := z - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T$ qui donne

$$c - \frac{b^2}{a} = q\tilde{z}^2 - 2l\tilde{z} + M$$

avec :

$$\begin{aligned} q &= \lambda^{-1}\left(1 - \lambda^{-1}\left(1 - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)a^{-1}\left(1 - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^T\right) \\ l &= \lambda^{-1}\left(1 - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)a^{-1}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^T \\ M &= \boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T - \left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)a^{-1}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^T. \end{aligned}$$

En la variable $\tilde{z}$, nous obtenons une distribution de Student de degré $n-2$, avec comme paramètre de position :

$$\frac{l}{q} = \mu\left(1 - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right) \text{ avec } \mu := \frac{\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}$$

et paramètre d'échelle :

$$\begin{aligned} &\sqrt{\frac{1}{n-2}\left(\frac{M}{q} - \left(\frac{l}{q}\right)^2\right)} \\ &\quad = \sqrt{\frac{1}{n-2}\left((\boldsymbol{y} - \mu\boldsymbol{1})\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T\right)\left(1 - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{t})^T + \frac{\left(1 - \boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{t})^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)}. \end{aligned}$$

Ainsi la distribution conditionnelle de la variable aléatoire $Y_t$ est une distribution de Student de degré $n-2$, ayant comme paramètre de position :

$$\mu + \boldsymbol{k}\left(\boldsymbol{t}\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-\mu\boldsymbol{1}\right)^T \text{ avec } \mu := \frac{\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}$$

et paramètre d'échelle :

$$\sqrt{\frac{1}{n-2}\left(\left(\boldsymbol{y}-\mu\boldsymbol{1}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T\right)\left(1-\boldsymbol{k}\left(\boldsymbol{t}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{k}\left(\boldsymbol{t}\right)^T+\frac{\left(1-\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{k}\left(\boldsymbol{t}\right)^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)}$$

où $\boldsymbol{1} = (1, 1, \ldots, 1) \in \mathbb{R}^n$. □

**Lemme 2.** *Soit $p(\boldsymbol{z})$, un polynôme du second degré en $r$ variables. Supposons que*
- *pour tout $\boldsymbol{z} \in \mathbb{R}^r, p(\boldsymbol{z}) > 0$,*
- *la partie quadratique de $p$ est définie positive.*

*Soit $\ell > 0$. Si $n > \ell + 1$ alors*

$$\int_{\mathbb{R}^r} p\left(\boldsymbol{z}\right)^{-\frac{n+r-\ell-1}{2}} \, d\boldsymbol{z} < \infty.$$

*Démonstration.* Avec un changement de base orthogonale, nous pouvons écrire :

$$p\left(\boldsymbol{z}\right) = \sum_{i=1}^{r} d_i e_i^2 + \sum_{i=1}^{r} s_i e_i + M$$

où $(e_1, e_2, \ldots, e_r) := \boldsymbol{e}$ sont les coordonnées de $\boldsymbol{z}$ dans la nouvelle base, et où, pour tout $1 \leq i \leq r$, les nombres $d_i$ sont strictement positifs.
Considérons une boule $B$ dans $\mathbb{R}^r$, contenant zéro et telle que pour tout $z \ni B$, nous avons :

$$p\left(\boldsymbol{z}\right) \geq \frac{1}{2}\sum_{i=1}^{r} d_i e_i^2 \geq \frac{C}{2}\sum_{i=1}^{r} e_i^2$$

où $C = \min\left(d_1, d_2, \ldots, d_r\right)$. Notons $B^c$ le complémentaire de $B$ dans l'espace $\mathbb{R}^r$. Comme $p > 0$, l'intégrale de la fonction $p\left(\boldsymbol{z}\right)^{-\frac{n+r-\ell-1}{2}}$ sur la boule $B$ est finie.

De plus, nous avons :

$$\int_{B^c} p\left(\boldsymbol{z}\right)^{-\frac{n+r-\ell-1}{2}} \, \mathrm{d}\boldsymbol{z} \leq C' \int_{B^c} \left(\sum_{i=1}^{r} e_i^2\right)^{-\frac{n+r-\ell-1}{2}} \mathrm{d}e_1 \, \mathrm{d}e_2 \ldots \, \mathrm{d}e_r$$

où $C' := \left(\frac{C}{2}\right)^{-\frac{n+r-\ell-1}{2}}$.
De par la définition de la norme euclidienne :

$$\int_{B^c} p\left(\boldsymbol{z}\right)^{-\frac{n+r-\ell-1}{2}} \, \mathrm{d}\boldsymbol{z} \leq C' \int_{B^c} \|\boldsymbol{e}\|^{-(n+r-\ell-1)} \, \mathrm{d}e_1 \, \mathrm{d}e_2 \ldots \, \mathrm{d}e_r.$$

Mais,

$$\int_{B^c} \|\boldsymbol{e}\|^{-(n+r-\ell-1)} \, \mathrm{d}e_1 \, \mathrm{d}e_2 \ldots \, \mathrm{d}e_r < \infty \quad \text{si et seulement si} \quad n > \ell + 1$$

ainsi, nous avons :

$$\int_{\mathbb{R}^r} p(\boldsymbol{z})^{-\frac{n+r-\ell-1}{2}} \, \mathrm{d}\boldsymbol{z} < \infty \quad \text{si} \quad n > \ell + 1.$$

□

**Lemme 3.** *Soient $\ell$, $n$ et $r$ trois entiers positifs avec $n > \ell + 1$. Soient $Q$ une forme quadratique définie positive en $r$ variables, $L$ une forme linéaire en $r$ variables et $M$ une constante positive. Supposons que pour tout $\boldsymbol{z} \in \mathbb{R}^r$, $Q(\boldsymbol{z}) + L(\boldsymbol{z}) + M > 0$.*
*Définissons $C := \left(\int_{\mathbb{R}^r} (Q(\boldsymbol{z}) + L(\boldsymbol{z}) + M)^{-\frac{n+r-\ell-1}{2}} \, d\boldsymbol{z}\right)^{-1}$. La densité de probabilité $f$ définie sur $\mathbb{R}^r$ par :*

$$f : \boldsymbol{z} \mapsto C\,(Q(\boldsymbol{z}) + L(\boldsymbol{z}) + M)^{-\frac{n+r-\ell-1}{2}}$$

*est une densité de Student multivariée avec $n - \ell - 1$ degrés de liberté.*

*Démonstration.* Notons $\boldsymbol{z} := (z_1, z_2, \ldots, z_r)$, $Q(\boldsymbol{z}) := \sum_{i=1}^r q_{ii} z_i^2 + 2\sum_{1 \le i < j \le r} q_{ij} z_i z_j$ et $L(\boldsymbol{z}) := \sum_{i=1}^r \ell_i z_i$. Notons également $\boldsymbol{L} = (\ell_1, \ell_2, \ldots, \ell_r)$ et $\boldsymbol{Q} = (q_{ij})_{1 \le i,j \le r}$.
Alors :

$$L(\boldsymbol{z}) = \boldsymbol{L}\boldsymbol{z}^T \quad \text{et} \quad Q(\boldsymbol{z}) = \boldsymbol{z}\boldsymbol{Q}\boldsymbol{z}^T.$$

Diagonalisons la matrice définie positive $\boldsymbol{Q}$ ; il existe une matrice orthogonale $\boldsymbol{P}$, de type $(r, r)$ telle que :

$$\boldsymbol{Q} = \boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1} = \boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^T$$

où $\boldsymbol{A}$ est une matrice diagonale avec les coefficients diagonaux $a_1, a_2, \ldots, a_r > 0$.
Notons $\boldsymbol{s} = \boldsymbol{z}\boldsymbol{P}$ et $\boldsymbol{s} = (s_1, s_2, \ldots, s_r)$. Nous avons :

$$f(\boldsymbol{z}) = C\left(M + \boldsymbol{L}\boldsymbol{P}\boldsymbol{s}^T + \boldsymbol{s}\boldsymbol{A}\boldsymbol{s}^T\right)^{-\frac{n+r-\ell-1}{2}}.$$

En notant $\boldsymbol{L}\boldsymbol{P} = (b_1, b_2, \ldots, b_r)$, nous déduisons :

$$\boldsymbol{L}\boldsymbol{P}\boldsymbol{s}^T + \boldsymbol{s}\boldsymbol{A}\boldsymbol{s}^T = \sum_{i=1}^r b_i s_i + a_i s_i^2 = \sum_{i=1}^r a_i \left(s_i + \frac{b_i}{2a_i}\right)^2 - \frac{b_i^2}{4a_i}.$$

Notons $\boldsymbol{B}$ le vecteur $\left(\frac{b_i}{2a_i}\right)_{1 \le i \le r}$ et $\boldsymbol{t} := \boldsymbol{s} + \boldsymbol{B}$. Notons également $d = M - \sum_{i=1}^r \frac{b_i^2}{4a_i}$.
Nous avons

$$f(\boldsymbol{z}) = C\left(d + \boldsymbol{t}\boldsymbol{A}\boldsymbol{t}^T\right)^{-\frac{n+r-\ell-1}{2}}.$$

Remarquons que $d > 0$ puisque $d + \boldsymbol{t}\boldsymbol{A}\boldsymbol{t}^T = Q(\boldsymbol{z}) + L(\boldsymbol{z}) + M$ qui est strictement positif pour tout $z$. Il est alors possible de mettre $d$ en facteur :

$$\begin{aligned} f(\boldsymbol{z}) &= Cd^{-\frac{n+r-\ell-1}{2}} \left(1 + \frac{1}{d}(\boldsymbol{z}\boldsymbol{P} + \boldsymbol{B})\,\boldsymbol{A}\,(\boldsymbol{z}\boldsymbol{P} + \boldsymbol{B})^T\right)^{-\frac{n+r-\ell-1}{2}} \\ &= Cd^{-\frac{n+r-\ell-1}{2}} \left(1 + \left(\boldsymbol{z} + \boldsymbol{B}\boldsymbol{P}^T\right)\left(\frac{1}{d}\boldsymbol{Q}\right)\left(\boldsymbol{z} + \boldsymbol{B}\boldsymbol{P}^T\right)^T\right)^{-\frac{n+r-\ell-1}{2}}. \end{aligned}$$

avec $C' := Cd^{-\frac{n+r-\ell-1}{2}}$, $\boldsymbol{\mu} := -\boldsymbol{B}\boldsymbol{P}^T$ et $\boldsymbol{\Pi} := \frac{d}{n-\ell-1}\boldsymbol{Q}^{-1}$, nous avons :

$$f(\boldsymbol{z}) = C'\left(1 + \frac{1}{n-\ell-1}(\boldsymbol{z}-\boldsymbol{\mu})\boldsymbol{\Pi}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})^T\right)^{-\frac{n+r-\ell-1}{2}}.$$

Suivant Kotz et Nadarajah (2004), nous reconnaissons une distribution de Student multivariée en dimension $r$ de degré de liberté $n-\ell-1$, ayant pour paramètre d'échelle $\boldsymbol{\Pi}$ et pour paramètre de position $\boldsymbol{\mu}$. □

### 1.4.3 Distribution conditionnelle de la moyenne aléatoire

Nous avons mis en évidence dans la section 1.1.2 le lien fort existant entre le champ aléatoire décrit dans le théorème 4 et l'approche dite "hiérarchique", la différence tenant plus finalement à la manière de formaliser le problème. Ainsi, on peut également effectuer de l'inférence bayésienne sur ce qui est considéré comme des paramètres dans l'approche hiérarchique, à savoir $U$ et $V$. Nous débutons cette étude avec la moyenne du champ aléatoire et déterminons, dans cette section, la loi conditionnelle de $U$. Nous remarquons, à nouveau, une certaine proximité avec les résultats obtenus par la méthode du maximum de vraisemblance, puisque le mode de la distribution conditionnelle de $U$ est égal à l'estimateur du maximum de vraisemblance (voir section 1.6.2 et le tableau récapitulatif 1.5).

**Théorème 5.** *La distribution conditionnelle de la variable aléatoire $U$ sachant que $(Y_{x_i} = y_i)_{1\leq i\leq n}$ est donnée par l'expression explicite de sa densité.*
*Pour $m$ fixé et quand $\epsilon$ tend vers zéro, pour $n > 2$, la distribution conditionnelle de la variable aléatoire $U$ est une distribution de Student restreinte à l'intervalle $[-m, m]$.*
*Quand $m$ tend vers l'infini et $\epsilon$ tend vers zéro, pour $n > 2$, la distribution conditionnelle de la variable aléatoire $U$ devient une distribution de Student de degré $n-2$, ayant pour paramètre de position :*

$$\mu := \frac{\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}$$

*et pour paramètre d'échelle :*

$$\sqrt{\frac{1}{n-2}\frac{1}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T - \frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)} = \sqrt{\frac{1}{n-2}\frac{1}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\left((\boldsymbol{y}-\mu\boldsymbol{1})\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T\right)}$$

*où $\boldsymbol{1} = (1, 1, \ldots, 1) \in \mathbb{R}^n$.*

*Démonstration.* Nous avons :

$$(Y_{x_1}, \ldots, Y_{x_n}) = (U, \ldots, U) + V(W_{x_1}, \ldots, W_{x_n})$$

où $(W_{x_1}, \ldots, W_{x_n})$ suit la distribution $\mathcal{N}(0, \boldsymbol{\Sigma})$.
Notons $f$ la densité du vecteur aléatoire $(U, Y_{x_1}, \ldots, Y_{x_n})$ :

$$f(u, \boldsymbol{\zeta}) = \frac{1}{2m}\frac{1}{\frac{1}{\epsilon}-\epsilon}\int_{\epsilon}^{\frac{1}{\epsilon}} \frac{1}{\left(\sqrt{2\pi}\right)^n v^n\sqrt{|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2v^2}(\boldsymbol{\zeta}-\boldsymbol{u})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\zeta}-\boldsymbol{u})^T\right)\,\mathrm{d}v\;\mathbb{1}_{[-m,m]}(u)$$

avec $\boldsymbol{\zeta} := (y_{x_1}, y_{x_2}, \ldots, y_{x_n})$ et $\boldsymbol{u} := (u, u, \ldots, u) \in \mathbb{R}^n$.

La distribution conditionnelle de $U$ sachant que $(Y_{x_i} = y_i)_{1 \leq i \leq n}$ est :

$$u \mapsto \frac{f(u, \boldsymbol{y})}{\int_{-m}^{m} f(u', \boldsymbol{y}) \ \mathrm{d}u'} =: g(u).$$

Après simplification, nous obtenons :

$$g(u) = \frac{\int_{\epsilon}^{\frac{1}{\epsilon}} \frac{1}{v^n} \exp\left(-\frac{1}{2v^2} (\boldsymbol{y} - \boldsymbol{u}) \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{u})^T\right) \ \mathrm{d}v}{\int_{-m}^{m} \int_{\epsilon}^{\frac{1}{\epsilon}} \frac{1}{v^n} \exp\left(-\frac{1}{2v^2} (\boldsymbol{y} - \boldsymbol{u'}) \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{u'})^T\right) \ \mathrm{d}u' \ \mathrm{d}v} \mathbb{1}_{[-m,m]}(u).$$

Notons :

$$\mathrm{NUM}(u) := \int_{\epsilon}^{\frac{1}{\epsilon}} \frac{1}{v^n} \exp\left(-\frac{1}{2v^2} (\boldsymbol{y} - \boldsymbol{u}) \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{u})^T\right) \ \mathrm{d}v \ \mathbb{1}_{[-m,m]}(u).$$

Avec le changement de variable $q = \frac{1}{2v^2} (\boldsymbol{y} - \boldsymbol{u}) \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{u})^T$ et la définition de la fonction d'Euler $\Gamma$, nous obtenons :

$$\mathrm{NUM}(u) \underset{\epsilon \to 0}{\longrightarrow} 2^{\frac{n-3}{2}} \left((\boldsymbol{y} - \boldsymbol{u}) \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{u})^T\right)^{-\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right) \mathbb{1}_{[-m,m]}(u).$$

A l'aide de quelques calculs supplémentaires, on montre que, pour $m$ fixé, la distribution conditionnelle de la variable aléatoire $U$ tend vers une distribution de Student restreinte à l'intervalle $[-m, m]$ lorsque $\epsilon$ tend vers zéro.
Notons :

$$\mathrm{DEN} := \int_{-m}^{m} \int_{\epsilon}^{\frac{1}{\epsilon}} \frac{1}{v^n} \exp\left(-\frac{1}{2v^2} (\boldsymbol{y} - \boldsymbol{u}) \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{u})^T\right) \ \mathrm{d}u \ \mathrm{d}v.$$

Remarquons que :

$$(\boldsymbol{y} - \boldsymbol{u}) \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{u})^T = au^2 + 2bu + c$$

avec

$$\begin{aligned} a &= \boldsymbol{1} \boldsymbol{\Sigma}^{-1} \boldsymbol{1}^T \\ b &= -\boldsymbol{y} \boldsymbol{\Sigma}^{-1} \boldsymbol{1}^T \\ c &= \boldsymbol{y} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}^T \end{aligned}$$

et rappelons que :

$$\int_{-m}^{m} \exp\left(-\frac{1}{2v^2} \left(au^2 + 2bu + c\right)\right) \ \mathrm{d}u \underset{m \to +\infty}{\longrightarrow} \frac{v\sqrt{2\pi}}{\sqrt{a}} \exp\left(-\frac{1}{2v^2} \left(c - \frac{b^2}{a}\right)\right)$$

Nous obtenons, par convergence monotone :

$$\mathrm{DEN} \underset{m \to \infty}{\longrightarrow} \int_{\epsilon}^{\frac{1}{\epsilon}} \frac{1}{v^{n-1}} \frac{\sqrt{2\pi}}{\sqrt{\boldsymbol{1} \boldsymbol{\Sigma}^{-1} \boldsymbol{1}^T}} \exp\left(-\frac{1}{2v^2} \left(\boldsymbol{y} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}^T - \frac{\left(\boldsymbol{y} \boldsymbol{\Sigma}^{-1} \boldsymbol{1}^T\right)^2}{\boldsymbol{1} \boldsymbol{\Sigma}^{-1} \boldsymbol{1}^T}\right)\right) \ \mathrm{d}v$$

Avec le changement de variable $q = \frac{1}{2v^2}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T - \frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)$ et la définition de la fonction d'Euler $\Gamma$, nous obtenons :

$$\text{DEN} \underset{m\to\infty,\epsilon\to 0}{\longrightarrow} \frac{\sqrt{2\pi}}{\sqrt{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}} 2^{\frac{n-4}{2}} \left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T - \frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)^{-\frac{n-2}{2}} \Gamma\left(\frac{n-2}{2}\right).$$

Ainsi, on a

$$g(u) \underset{m\to\infty,\epsilon\to 0}{\longrightarrow} \frac{2^{\frac{n-3}{2}}\left((\boldsymbol{y}-\boldsymbol{u})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{u})^T\right)^{-\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}{\frac{\sqrt{2\pi}}{\sqrt{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}} 2^{\frac{n-4}{2}}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T - \frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)^{-\frac{n-2}{2}}\Gamma\left(\frac{n-2}{2}\right)}.$$

C'est-à-dire, après quelques calculs :

$$g(u) \underset{m\to\infty,\epsilon\to 0}{\longrightarrow} \frac{1}{\sqrt{\pi s^2}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)}\left(1+\frac{(u-l)^2}{s^2}\right)^{-\frac{n-1}{2}}$$

avec

$$l := \frac{\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}$$

et

$$s^2 := \frac{1}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T - \frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right).$$

Suivant la définition de Kotz et Nadarajah (2004), nous reconnaissons que la distribution conditionnelle de la variable aléatoire $U$ est une distribution de Student de degré $n-2$, de paramètre de position $l$ et de paramètre d'échelle $\frac{1}{\sqrt{n-2}}s$. □

### 1.4.4 Distribution conditionnelle de la variance aléatoire

Il est également possible d'effectuer de l'inférence bayésienne sur la variance du champ aléatoire. Nous observerons que son mode est identique à l'estimateur du maximum de vraisemblance (voir section 1.6.2 et le tableau récapitulatif 1.5).

Dans un premier temps, rappelons la définition de la distribution inverse du $\chi^2$ réduite. Nous nous référons pour cela au livre de Gelman et al. (2004).

La distribution inverse du $\chi^2$ réduite de degré $\nu$ et de paramètre d'échelle $\tau$ est la loi de densité, pour $x > 0$ :

$$x \mapsto \left(\frac{\tau^2\nu}{2}\right)^{\frac{\nu}{2}} \frac{1}{\Gamma\left(\frac{\nu}{2}\right)}\frac{1}{x^{1+\frac{\nu}{2}}}\exp\left(-\frac{\nu\tau^2}{2x}\right).$$

Pour $\nu > 2$, la moyenne de la distribution inverse du $\chi^2$ réduite est définie et est égale à $\frac{\nu\tau^2}{\nu-2}$. Pour $\nu > 4$, sa variance est également définie et vaut $\frac{2\nu^2\tau^4}{(\nu-2)^2(\nu-4)}$.

**Théorème 6.** *La distribution conditionnelle de la variable aléatoire $V$ sachant que $(Y_{x_i} = y_i)_{1\leq i\leq n}$ est donnée par l'expression explicite de sa densité.*
*Pour $\epsilon$ fixé, quand $m$ tend vers l'infini, pour $n > 2$, la distribution conditionnelle de la variable aléatoire $V^2$ devient une distribution inverse du $\chi^2$ réduite, restreinte à l'intervalle $\left[\epsilon^2, 1/\epsilon^2\right]$.*
*Quand $m$ tend vers l'infini et $\epsilon$ tend vers zéro, pour $n > 2$, la distribution conditionnelle de la variable aléatoire $V^2$ devient une distribution inverse du $\chi^2$ réduite de degré $n-2$ et de paramètre d'échelle :*

$$\sqrt{\frac{1}{n-2}\left(\boldsymbol{y} - \mu\boldsymbol{1}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T} \quad \textit{avec} \quad \mu := \frac{\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}$$

*où $\boldsymbol{1} = (1, 1, \ldots, 1) \in \mathbb{R}^n$.*

*Démonstration.* Nous avons :

$$(Y_{x_1}, \ldots, Y_{x_n}) = (U, \ldots, U) + V\left(W_{x_1}, \ldots, W_{x_n}\right)$$

où $(W_{x_1}, \ldots, W_{x_n})$ suit la distribution $\mathcal{N}(0, \boldsymbol{\Sigma})$.
Notons $f$ la densité du vecteur aléatoire $(V, Y_{x_1}, \ldots, Y_{x_n})$ :

$$f(v, \boldsymbol{\zeta}) = \frac{1}{2m}\frac{1}{\frac{1}{\epsilon} - \epsilon}\int_{-m}^{m}\frac{1}{\left(\sqrt{2\pi}\right)^n v^n\sqrt{|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2v^2}\left(\boldsymbol{\zeta} - \boldsymbol{u}\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\zeta} - \boldsymbol{u}\right)^T\right)\,\mathrm{d}u\;\mathbb{1}_{\left[\epsilon, \frac{1}{\epsilon}\right]}(v)$$

avec $\boldsymbol{\zeta} := (y_{x_1}, y_{x_2}, \ldots, y_{x_n})$ et $\boldsymbol{u} := (u, u, \ldots, u) \in \mathbb{R}^n$.

La distribution conditionnelle de $V$ sachant $(Y_{x_i} = y_i)_{1\leq i\leq n}$ est :

$$v \mapsto \frac{f(v, \boldsymbol{y})}{\int_{\epsilon}^{1/\epsilon} f(v', \boldsymbol{y})\;\mathrm{d}v'} =: g(v).$$

Après simplification, nous obtenons :

$$g(v) = \frac{\int_{-m}^{m}\frac{1}{v^n}\exp\left(-\frac{1}{2v^2}\left(\boldsymbol{y} - \boldsymbol{u}\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y} - \boldsymbol{u}\right)^T\right)\,\mathrm{d}u}{\int_{-m}^{m}\int_{\epsilon}^{\frac{1}{\epsilon}}\frac{1}{v'^n}\exp\left(-\frac{1}{2v'^2}\left(\boldsymbol{y} - \boldsymbol{u}\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y} - \boldsymbol{u}\right)^T\right)\,\mathrm{d}u\;\mathrm{d}v'}\;\mathbb{1}_{\left[\epsilon, \frac{1}{\epsilon}\right]}(v).$$

Notons :

$$\mathrm{NUM}(v) := \int_{-m}^{m}\frac{1}{v^n}\exp\left(-\frac{1}{2v^2}\left(\boldsymbol{y} - \boldsymbol{u}\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y} - \boldsymbol{u}\right)^T\right)\,\mathrm{d}u\;\mathbb{1}_{\left[\epsilon, \frac{1}{\epsilon}\right]}(v).$$

Remarquons que

$$\left(\boldsymbol{y} - \boldsymbol{u}\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y} - \boldsymbol{u}\right)^T = au^2 + 2bu + c$$

avec

$$\begin{aligned} a &= \boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T \\ b &= -\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T \\ c &= \boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T \end{aligned}$$

et rappelons que :

$$\int_{-m}^{m} \exp\left(-\frac{1}{2v^2}\left(au^2+2bu+c\right)\right)\,\mathrm{d}u \underset{m\to+\infty}{\longrightarrow} \frac{v\sqrt{2\pi}}{\sqrt{a}}\exp\left(-\frac{1}{2v^2}\left(c-\frac{b^2}{a}\right)\right).$$

Nous obtenons :

$$\mathrm{NUM}\left(v\right) \underset{m\to\infty}{\longrightarrow} \frac{1}{v^{n-1}}\frac{\sqrt{2\pi}}{\sqrt{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}}\exp\left(-\frac{1}{2v^2}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T-\frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)\right)\mathbb{1}_{\left[\epsilon,\frac{1}{\epsilon}\right]}\left(v\right).$$

Pour $\epsilon$ fixé, l'expression ci-dessus, permet, après quelques calculs de démontrer que la distribution conditionnelle de la variable aléatoire $V^2$ tend vers une distribution inverse du $\chi^2$ réduite, restreinte à l'intervalle $\left[\epsilon^2, 1/\epsilon^2\right]$ lorsque $m$ tend vers l'infini.
Notons :

$$\mathrm{DEN} := \int_{-m}^{m}\int_{\epsilon}^{\frac{1}{\epsilon}} \frac{1}{v^n}\exp\left(-\frac{1}{2v^2}\left(\boldsymbol{y}-\boldsymbol{u}\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-\boldsymbol{u}\right)^T\right)\,\mathrm{d}u\,\mathrm{d}v.$$

Par convergence monotone, l'intégrale de $\mathrm{NUM}\left(v\right)$ tend vers l'intégrale de sa limite quand $m$ tend vers l'infini. Par conséquent, nous avons :

$$\mathrm{DEN} \underset{m\to\infty}{\longrightarrow} \int_{\epsilon}^{\frac{1}{\epsilon}} \frac{1}{v^{n-1}}\frac{\sqrt{2\pi}}{\sqrt{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}}\exp\left(-\frac{1}{2v^2}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T-\frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)\right)\,\mathrm{d}v.$$

Avec le changement de variable $q = \frac{1}{2v^2}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T-\frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)$ et la définition de la fonction d'Euler $\Gamma$, nous obtenons :

$$\mathrm{DEN} \underset{m\to\infty,\epsilon\to 0}{\longrightarrow} \frac{\sqrt{2\pi}}{\sqrt{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}}2^{\frac{n-4}{2}}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T-\frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)^{-\frac{n-2}{2}}\Gamma\left(\frac{n-2}{2}\right).$$

On obtient donc une expression analytique de la limite de la densité $g$ :

$$g\left(v\right) \underset{m\to\infty,\epsilon\to 0}{\longrightarrow} \frac{\frac{1}{v^{n-1}}\frac{\sqrt{2\pi}}{\sqrt{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}}\exp\left(-\frac{1}{2v^2}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T-\frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)\right)}{\frac{\sqrt{2\pi}}{\sqrt{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}}2^{\frac{n-4}{2}}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T-\frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)^{-\frac{n-2}{2}}\Gamma\left(\frac{n-2}{2}\right)}\mathbb{1}_{[v>0]}.$$

C'est-à-dire :

$$g\left(v\right) \underset{m\to\infty,\epsilon\to 0}{\longrightarrow} \frac{\frac{1}{v^{n-1}}\exp\left(-\frac{1}{2v^2}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T-\frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)\right)}{2^{\frac{n-4}{2}}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T-\frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)^{-\frac{n-2}{2}}\Gamma\left(\frac{n-2}{2}\right)}\mathbb{1}_{[v>0]}.$$

Notons que la variable aléatoire $W := V^2$ admet pour densité $h\left(w\right) := \frac{1}{2\sqrt{w}}g\left(\sqrt{w}\right)$. Nous avons :

$$h\left(w\right) \underset{m\to\infty,\epsilon\to 0}{\longrightarrow} \frac{\frac{1}{w^{n/2}}\exp\left(-\frac{1}{2w}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T-\frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)\right)}{2^{\frac{n-2}{2}}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T-\frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)^{-\frac{n-2}{2}}\Gamma\left(\frac{n-2}{2}\right)}.$$

Nous reconnaissons que la variable aléatoire $V^2$ suit la distribution inverse du $\chi^2$ réduite de degré $n-2$ et de paramètre d'échelle :

$$\sqrt{\frac{1}{n-2}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T - \frac{\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)} = \sqrt{\frac{1}{n-2}\left(\boldsymbol{y}-\mu\boldsymbol{1}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T} \quad \text{avec} \quad \mu := \frac{\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}$$

où $\boldsymbol{1} = (1, 1, \ldots, 1) \in \mathbb{R}^n$. □

## 1.5 Modèle mixte aléatoire

Le modèle aléatoire défini dans la section précédente peut être généralisé, enrichi d'un ensemble de fonctions de régression $(f_1, f_2, \ldots, f_\ell)$. En s'inspirant de l'article de Sacks et al. (1989) et du livre de Santner et al. (2003), nous proposons de considérer le champ aléatoire a priori suivant :

$$Y_x = \sum_{i=1}^{\ell} U_i f_i(x) + V W_x$$

où $(U_i)_{1\leq i\leq \ell}$ est un vecteur aléatoire réel suivant une distribution uniforme sur un hypercube $[-m, m]^\ell$, où $V$ est une variable aléatoire réelle positive suivant une distribution uniforme sur un intervalle $[\epsilon, 1/\epsilon]$, où $(W_x)_{x\in X}$ est un champ gaussien centré réduit et où les fonctions $f_i, 1 \leq i \leq \ell$ sont déterministes. De plus, nous supposons que les données aléatoires $(U_i)_{1\leq i\leq \ell}$, $V$ et $(W_x)_{x\in X}$ sont indépendantes.

Le champ gaussien $(W_x)_{x\in X}$ est caractérisé par la fonction de corrélation $k$ (rappelons que la moyenne est nulle et que la variance vaut 1). Nous supposons que $n \geq \ell + 2$. Soient $\boldsymbol{x} := (x_1, x_2, \ldots, x_n) \in X^n$, $\boldsymbol{y} := (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$ et $\boldsymbol{U} = (U_1, U_2, \ldots, U_\ell)$. Pour tous les éléments $x$ de l'espace des facteurs $X$, nous notons $\boldsymbol{f}(\boldsymbol{x})$ le vecteur colonne $(\ell, 1)$, $\boldsymbol{f}(\boldsymbol{x}) := (f_1(x), f_2(x), \ldots, f_\ell(x))^T$ et $\boldsymbol{F}(\boldsymbol{x}) := (f_i(x_j))_{1\leq i\leq \ell, 1\leq j\leq n}$ la matrice $(\ell, n)$. Notons $\boldsymbol{k}(\boldsymbol{x}) := (k(x, x_j))_{1\leq j\leq n}$ le vecteur de corrélation et $\boldsymbol{\Sigma} := (k(x_i, x_j))_{1\leq i,j\leq n}$, la matrice de corrélation.

Nous exprimons, dans cette section, la densité conditionnellement à $(Y_{x_i} = y_i)_{1\leq i\leq n}$ du champ aléatoire $(Y_x)_{x\in X}$ (théorème 7), du vecteur aléatoire $\boldsymbol{U}$ (théorème 8) et de la variable aléatoire $V$ (théorème 9).

Répétons que nous considérons que nous sommes toujours dans une situation générique qui nous permet de supposer que la matrice de corrélation est définie positive.

### 1.5.1 Distribution conditionnelle du champ aléatoire

Soit $\boldsymbol{t} = (t_1, t_2, \ldots, t_r) \in X^r$. Avec les notations introduites précédemment on peut considérer la matrice $\boldsymbol{k}(\boldsymbol{t}) := (k(t_i, x_j))_{1\leq i\leq r, 1\leq j\leq n}$ et la matrice $\boldsymbol{F}(\boldsymbol{t}, \boldsymbol{x}) := \boldsymbol{F}(t_1, \ldots, t_r, x_1, \ldots, x_n)$ de type $(\ell, r+n)$. On pose $\boldsymbol{\Sigma_2} := (k(t_i, t_j))_{1\leq i\leq r, 1\leq j\leq r}$ et

$$\boldsymbol{\Delta} := \begin{pmatrix} \boldsymbol{\Sigma_2} & \boldsymbol{k}(\boldsymbol{t}) \\ \boldsymbol{k}(\boldsymbol{t})^T & \boldsymbol{\Sigma} \end{pmatrix}.$$

On suppose que les matrices symétriques positives $\boldsymbol{\Delta}$ et $\boldsymbol{a} := \boldsymbol{F}(\boldsymbol{t},\boldsymbol{x})\boldsymbol{\Delta}^{-1}\boldsymbol{F}(\boldsymbol{t},\boldsymbol{x})^T$ sont non-dégénérées.

**Théorème 7.** *La distribution conditionnelle du champ aléatoire $(Y_x)_{x\in X}$ sachant que $(Y_{x_i} = y_i)_{1\leq i\leq n}$ est donnée par les formules explicites des densités des marginales de dimensions finies. Supposons que $n-\ell > 1$.*
*Quand le paramètre $m$ tend vers l'infini et $\epsilon$ tend vers zéro, ces distributions conditionnelles deviennent des distributions de Student multivariées.*
*En particulier, quand $m \to \infty$ et $\epsilon \to 0$, la distribution conditionnelle de la variable aléatoire $Y_x$ devient une distribution de Student de degré $n-\ell-1$, pour paramètre de position :*

$$\boldsymbol{\mu}^T\boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y} - \boldsymbol{\mu}^T\boldsymbol{F}(\boldsymbol{x})\right)^T = \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T + \left(\boldsymbol{f}(\boldsymbol{x})^T - \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)\boldsymbol{\mu}$$

*où*

$$\boldsymbol{\mu} := \left(\boldsymbol{F}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)^{-1}\boldsymbol{F}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T$$

*et pour paramètre d'échelle :*

$$\sqrt{\frac{1}{n-\ell-1}\left(\left(\boldsymbol{y} - \boldsymbol{\mu}^T\boldsymbol{F}(\boldsymbol{x})\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T\right)\left(1 - \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{x})^T + \boldsymbol{B}\left(\boldsymbol{F}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)^{-1}\boldsymbol{B}^T\right)}$$

*où*

$$\boldsymbol{B} := \left(\boldsymbol{f}(\boldsymbol{x})^T - \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right).$$

*Remarque.* Nous vérifions aisément que si $\ell = 1$ et si la fonction $f_1$ est constante, alors nous obtenons le modèle aléatoire décrit dans la section 1.4.

*Démonstration.*

**Partie 1 : Distribution conditionnelle multivariée**

Soit $r$ un entier positif et soit $(t_1,\ldots,t_r) \in X^r$. Nous avons

$$(Y_{t_1},\ldots,Y_{t_r},Y_{x_1},\ldots,Y_{x_n}) = \boldsymbol{U}\boldsymbol{F}(\boldsymbol{t},\boldsymbol{x}) + V(W_{t_1},\ldots,W_{t_r},W_{x_1},\ldots,W_{x_n})$$

où $(W_{t_1},\ldots,W_{t_r},W_{x_1},\ldots,W_{x_n})$ suit la distribution $\mathcal{N}(0,\boldsymbol{\Delta})$.
Soit $f$ la densité du vecteur aléatoire $(Y_{t_1},\ldots,Y_{t_r},Y_{x_1},\ldots,Y_{x_n})$ :

$$f(\boldsymbol{\zeta}) = \frac{1}{(2m)^\ell}\frac{1}{\frac{1}{\epsilon}-\epsilon}\int_{[-m,m]^\ell}\int_\epsilon^{\frac{1}{\epsilon}}\frac{1}{\left(\sqrt{2\pi}\right)^{n+r}v^{n+r}\sqrt{|\boldsymbol{\Delta}|}}$$
$$\exp\left(-\frac{1}{2v^2}\left(\boldsymbol{\zeta} - \boldsymbol{u}\boldsymbol{F}(\boldsymbol{t},\boldsymbol{x})\right)\boldsymbol{\Delta}^{-1}\left(\boldsymbol{\zeta} - \boldsymbol{u}\boldsymbol{F}(\boldsymbol{t},\boldsymbol{x})\right)^T\right)\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}v$$

avec $\boldsymbol{\zeta} := (y_{t_1},\ldots,y_{t_r},y_{x_1},\ldots,y_{x_n})$.

La densité conditionnelle de $(Y_{t_1},Y_{t_2},\ldots,Y_{t_r})$ sachant $(Y_{x_i} = y_i)_{1\leq i\leq n}$ est :

$$\boldsymbol{z} := (y_{t_1},y_{t_2},\ldots,y_{t_r}) \mapsto \frac{f(\boldsymbol{z},\boldsymbol{y})}{\int_{\mathbb{R}^r} f(\boldsymbol{z'},\boldsymbol{y})\,\mathrm{d}\boldsymbol{z'}} =: g(\boldsymbol{z}).$$

Après simplification, nous avons :

$$g(\boldsymbol{z}) = \frac{\int_{[-m,m]^\ell} \int_\epsilon^{\frac{1}{\epsilon}} \frac{1}{v^{n+r}} \exp\left(-\frac{1}{2v^2}\left(\left(\boldsymbol{z} \quad \boldsymbol{y}\right) - \boldsymbol{uF}(\boldsymbol{t},\boldsymbol{x})\right)\boldsymbol{\Delta}^{-1}\left(\left(\boldsymbol{z} \quad \boldsymbol{y}\right) - \boldsymbol{uF}(\boldsymbol{t},\boldsymbol{x})\right)^T\right) \mathrm{d}\boldsymbol{u}\,\mathrm{d}v}{\int_{\mathbb{R}^r} \int_{[-m,m]^\ell} \int_\epsilon^{\frac{1}{\epsilon}} \frac{1}{v^{n+r}} \exp\left(-\frac{1}{2v^2}\left(\left(\boldsymbol{z}' \quad \boldsymbol{y}\right) - \boldsymbol{uF}(\boldsymbol{t},\boldsymbol{x})\right)\boldsymbol{\Delta}^{-1}\left(\left(\boldsymbol{z}' \quad \boldsymbol{y}\right) - \boldsymbol{uF}(\boldsymbol{t},\boldsymbol{x})\right)^T\right) \mathrm{d}\boldsymbol{z}'\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}v}.$$

Notons

$$\mathrm{NUM}(\boldsymbol{z}) := \int_{[-m,m]^\ell} \int_\epsilon^{\frac{1}{\epsilon}} \frac{1}{v^{n+r}} \exp\left(-\frac{1}{2v^2}\left(\left(\boldsymbol{z} \quad \boldsymbol{y}\right) - \boldsymbol{uF}(\boldsymbol{t},\boldsymbol{x})\right)\boldsymbol{\Delta}^{-1}\left(\left(\boldsymbol{z} \quad \boldsymbol{y}\right) - \boldsymbol{uF}(\boldsymbol{t},\boldsymbol{x})\right)^T\right) \mathrm{d}\boldsymbol{u}\,\mathrm{d}v.$$

Nous pouvons écrire

$$\left(\left(\boldsymbol{z} \quad \boldsymbol{y}\right) - \boldsymbol{uF}(\boldsymbol{t},\boldsymbol{x})\right)\boldsymbol{\Delta}^{-1}\left(\left(\boldsymbol{z} \quad \boldsymbol{y}\right) - \boldsymbol{uF}(\boldsymbol{t},\boldsymbol{x})\right)^T = \boldsymbol{uau}^T + 2\boldsymbol{bu}^T + c$$

avec

$\boldsymbol{a} = \boldsymbol{F}(\boldsymbol{t},\boldsymbol{x})\boldsymbol{\Delta}^{-1}\boldsymbol{F}(\boldsymbol{t},\boldsymbol{x})^T$, matrice de type $(\ell,\ell)$,
$\boldsymbol{b} = -\left(\boldsymbol{z} \quad \boldsymbol{y}\right)\boldsymbol{\Delta}^{-1}\boldsymbol{F}(\boldsymbol{t},\boldsymbol{x})^T$, vecteur ligne de dimension $\ell$,
$c = \left(\boldsymbol{z} \quad \boldsymbol{y}\right)\boldsymbol{\Delta}^{-1}\left(\boldsymbol{z} \quad \boldsymbol{y}\right)^T$, scalaire.

Rappelons une formule classique d'intégration gaussienne. Si $\boldsymbol{a}$ est une matrice définie positive $\ell \times \ell$, $\boldsymbol{b}$ un vecteur $1 \times \ell$ et $c$ un nombre, alors

$$\int_{\mathbb{R}^\ell} \exp\left(-\frac{1}{2}\left(\boldsymbol{xax}^T + 2\boldsymbol{bx}^T + c\right)\right) \mathrm{d}\boldsymbol{x} = \frac{(2\pi)^{\ell/2}}{|\boldsymbol{a}|^{1/2}} \exp\left(-\frac{1}{2}\left(c - \boldsymbol{ba}^{-1}\boldsymbol{b}^T\right)\right).$$

Nous obtenons, par convergence monotone :

$$\mathrm{NUM}(\boldsymbol{z}) \underset{m\to+\infty}{\longrightarrow} \int_\epsilon^{\frac{1}{\epsilon}} \frac{(2\pi)^{\ell/2}}{v^{n+r-\ell}\,|\boldsymbol{a}|^{1/2}} \exp\left(-\frac{1}{2v^2}\left(c - \boldsymbol{ba}^{-1}\boldsymbol{b}^T\right)\right) \mathrm{d}v.$$

Avec le changement de variable $v \mapsto q = \frac{1}{2v^2}\left(c - \boldsymbol{ba}^{-1}\boldsymbol{b}^T\right)$ et la définition de la fonction d'Euler $\Gamma$, nous obtenons :

$$\mathrm{NUM}(\boldsymbol{z}) \underset{m\to+\infty,\epsilon\to 0}{\longrightarrow} \frac{(2\pi)^{\ell/2}}{|\boldsymbol{a}|^{1/2}} 2^{\frac{n+r-\ell-3}{2}} \left(c - \boldsymbol{ba}^{-1}\boldsymbol{b}^T\right)^{-\frac{n+r-\ell-1}{2}} \Gamma\left(\frac{n+r-\ell-1}{2}\right).$$

Remarquons que le polynôme $\boldsymbol{uau}^T + 2\boldsymbol{bu}^T + c$ est toujours strictement positif puisque nous excluons les situations dégénérées. En posant $\boldsymbol{u} = -\boldsymbol{ba}^{-1}$ et en se rappelant que $\boldsymbol{a}^T = \boldsymbol{a}$, on a $\left(c - \boldsymbol{ba}^{-1}\boldsymbol{b}^T\right) > 0$. L'expression $c - \boldsymbol{ba}^{-1}\boldsymbol{b}^T$ est un polynôme du second degré en la variable $\boldsymbol{z}$ et il est positif d'après ce qui précède. On écrit

$$c - \boldsymbol{ba}^{-1}\boldsymbol{b}^T = Q(\boldsymbol{z}) + L(\boldsymbol{z}) + M$$

où $Q$ est une forme quadratique, $L$ est forme linéaire et $M$ est constant.
(Nous ne considérons pas le cas exceptionnel, mais possible, où $Q = L = 0$.)
Donc

$$\mathrm{NUM}(\boldsymbol{z}) \underset{m\to+\infty,\epsilon\to 0}{\longrightarrow} \frac{(2\pi)^{\ell/2}}{|\boldsymbol{a}|^{1/2}} 2^{\frac{n+r-\ell-3}{2}} \left(Q(\boldsymbol{z}) + L(\boldsymbol{z}) + M\right)^{-\frac{n+r-\ell-1}{2}} \Gamma\left(\frac{n+r-\ell-1}{2}\right).$$

Par convergence monotone, l'intégrale de NUM $(\boldsymbol{z})$ tend vers l'intégrale de sa limite quand $m$ tend vers l'infini. On obtient donc une expression de la limite de la densité $g$ :

$$g(\boldsymbol{z}) \underset{m\to+\infty,\epsilon\to 0}{\longrightarrow} \frac{\frac{(2\pi)^{\ell/2}}{|\boldsymbol{a}|^{1/2}} 2^{\frac{n+r-\ell-3}{2}} \left(Q(\boldsymbol{z})+L(\boldsymbol{z})+M\right)^{-\frac{n+r-\ell-1}{2}} \Gamma\left(\frac{n+r-\ell-1}{2}\right)}{\int_{\mathbb{R}^r} \frac{(2\pi)^{\ell/2}}{|\boldsymbol{a}|^{1/2}} 2^{\frac{n+r-\ell-3}{2}} \left(Q(\boldsymbol{z}')+L(\boldsymbol{z}')+M\right)^{-\frac{n+r-\ell-1}{2}} \Gamma\left(\frac{n+r-\ell-1}{2}\right) \mathrm{d}\boldsymbol{z}'}.$$

C'est-à-dire :

$$g(\boldsymbol{z}) \underset{m\to+\infty,\epsilon\to 0}{\longrightarrow} \frac{\left(Q(\boldsymbol{z})+L(\boldsymbol{z})+M\right)^{-\frac{n+r-\ell-1}{2}}}{\int_{\mathbb{R}^r} \left(Q(\boldsymbol{z}')+L(\boldsymbol{z}')+M\right)^{-\frac{n+r-\ell-1}{2}} \mathrm{d}\boldsymbol{z}'}.$$

Comme $Q(\boldsymbol{z}')+L(\boldsymbol{z}')+M$ est un polynôme positif, avec le lemme 2, nous obtenons, pour $n > \ell+1$ :

$$\int_{\mathbb{R}^r} \left(Q(\boldsymbol{z}')+L(\boldsymbol{z}')+M\right)^{-\frac{n+r-\ell-1}{2}} \mathrm{d}\boldsymbol{z}' < \infty$$

Avec le lemme 3, nous concluons que $g(\boldsymbol{z})$ est une distribution de Student multivariée de degré $n-\ell-1$, quand $m$ tend vers l'infini et $\epsilon$ tend vers zéro.

**Partie 2 : Distribution conditionnelle univariée**

Dans le cas particulier univarié, c'est-à-dire si $r=1$, nous avons

$$g(z) \underset{m\to+\infty,\epsilon\to 0}{\longrightarrow} \frac{\left(Q(z)+L(z)+M\right)^{-\frac{n-\ell}{2}}}{\int_{\mathbb{R}} \left(Q(z')+L(z')+M\right)^{-\frac{n-\ell}{2}} \mathrm{d}z'}.$$

Revenons aux définitions de $\boldsymbol{a}$, $\boldsymbol{b}$ et $c$. Rappelons que, d'après le lemme 1 :

$$\boldsymbol{\Delta}^{-1} = \lambda^{-1} \begin{pmatrix} 1 & -\boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1} \\ -\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{t})^T & \lambda\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{t})^T\boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1} \end{pmatrix}$$

avec $\lambda := 1 - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{t})^T$. Alors

$$\begin{aligned}
\boldsymbol{a} &= \boldsymbol{F}(\boldsymbol{t},\boldsymbol{x})\boldsymbol{\Delta}^{-1}\boldsymbol{F}(\boldsymbol{t},\boldsymbol{x})^T \\
\boldsymbol{b} &= \lambda^{-1}\left(z - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T\right)\left(\boldsymbol{f}(\boldsymbol{t})^T - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right) + \boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T \\
c &= \lambda^{-1}\left(z - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T\right)^2 + \boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T
\end{aligned}$$

Posons $\tilde{z} := z - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T$ qui donne

$$c - \boldsymbol{b}\boldsymbol{a}^{-1}\boldsymbol{b}^T = q\tilde{z}^2 - 2l\tilde{z} + M$$

avec

$$\begin{aligned}
q &= \lambda^{-1}\left(1 - \lambda^{-1}\left(\boldsymbol{f}(\boldsymbol{t})^T - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)\boldsymbol{a}^{-1}\left(\boldsymbol{f}(\boldsymbol{t})^T - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)^T\right) \\
l &= \lambda^{-1}\left(\boldsymbol{f}(\boldsymbol{t})^T - \boldsymbol{k}(\boldsymbol{t})\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)\boldsymbol{a}^{-1}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)^T \\
M &= \boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T - \left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)\boldsymbol{a}^{-1}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)^T
\end{aligned}$$

En la variable $\tilde{z}$, nous reconnaissons une distribution de Student de degré $n-\ell-1$, avec pour paramètre de position $l/q$ et paramètre d'échelle $M/q-(l/q)^2$. Nous pouvons appliquer le lemme 4 avec $\boldsymbol{s} := \boldsymbol{f}(\boldsymbol{t})^T$, $\boldsymbol{C}$ tel que $\boldsymbol{C}^T\boldsymbol{C} := \boldsymbol{\Sigma}^{-1}$, $\boldsymbol{S} := \boldsymbol{C}\boldsymbol{F}(\boldsymbol{x})^T$, $\boldsymbol{v} := \boldsymbol{k}(\boldsymbol{t})\boldsymbol{C}^T$ et $\boldsymbol{w} := \boldsymbol{y}\boldsymbol{C}^T$ et nous obtenons finalement, après quelques calculs, les résultats du théorème. □

**Lemme 4.** *On se donne un vecteur ligne $\boldsymbol{s}$ de dimension $\ell$, une matrice $\boldsymbol{S}$ de type $(n,\ell)$ et deux vecteurs ligne $\boldsymbol{v}$ et $\boldsymbol{w}$ de dimension $n$ tels que $1-\boldsymbol{v}\boldsymbol{v}^T \neq 0$ et $\boldsymbol{s} \neq \boldsymbol{v}\boldsymbol{S}$. Notons $\boldsymbol{a}$ la matrice de type $(\ell,\ell)$ suivante :*

$$\begin{aligned}\boldsymbol{a} := \frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}^T\boldsymbol{s}-\boldsymbol{S}^T\boldsymbol{v}^T\boldsymbol{s}-\boldsymbol{s}^T\boldsymbol{v}\boldsymbol{S}+\boldsymbol{S}^T\boldsymbol{v}^T\boldsymbol{v}\boldsymbol{S}\right)+\boldsymbol{S}^T\boldsymbol{S}\\ = \frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)+\boldsymbol{S}^T\boldsymbol{S}.\end{aligned}$$

*Supposons, de plus, que les matrices $\boldsymbol{a}$ et $\boldsymbol{S}^T\boldsymbol{S}$ sont inversibles.*
*Nous avons alors une identité de vecteurs de dimension $\ell$ :*

$$\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1} = \frac{\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}}{1-\frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T}.$$

*et une identité de scalaires :*

$$\begin{aligned}&\frac{\boldsymbol{w}\boldsymbol{w}^T-\left(\boldsymbol{w}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{w}\boldsymbol{S}\right)^T}{\frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(1-\frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\right)}-\left(\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{w}\boldsymbol{S}\right)^T\right)^2\\ &\quad=\left(\boldsymbol{w}\boldsymbol{w}^T-\left(\boldsymbol{w}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{w}\boldsymbol{S}\right)^T\right)\left(1-\boldsymbol{v}\boldsymbol{v}^T+\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\right).\end{aligned}$$

La démonstration du lemme 4 est décrite dans l'annexe A.

### 1.5.2 Distribution conditionnelle de la moyenne aléatoire

Dans la même lignée que ce nous avons proposé dans la section 1.4.3, effectuons de l'inférence bayésienne sur le vecteur moyenne de ce champ. Sa distribution conditionnée par les données $(Y_{x_i}=y_i)_{1\leq i\leq n}$ est décrite comme suit.

**Théorème 8.** *La distribution conditionnelle du vecteur aléatoire $\boldsymbol{U}$ sachant que $(Y_{x_i}=y_i)_{1\leq i\leq n}$ est donnée par l'expression explicite de sa densité. Supposons que $n-\ell>1$.*
*Lorsque $\epsilon$ tend vers zéro, pour $m$ fixé, la distribution conditionnelle du vecteur aléatoire $\boldsymbol{U}$ tend vers une distribution de Student multivariée restreinte à l'hypercube $[-m,m]^\ell$.*
*Quand le paramètre $m$ tend vers l'infini et $\epsilon$ tend vers zéro, la distribution conditionnelle du vecteur aléatoire $\boldsymbol{U}$ devient une distribution de Student multivariée de degré $n-\ell-1$, avec pour vecteur de paramètres de position :*

$$\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)\left(\boldsymbol{F}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)^{-1}$$

*et comme matrice des paramètres d'échelle :*

$$\frac{1}{n-\ell-1}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T-\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}\left(\boldsymbol{x}\right)^T\right)\left(\boldsymbol{F}\left(\boldsymbol{x}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}\left(\boldsymbol{x}\right)^T\right)^{-1}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}\left(\boldsymbol{x}\right)^T\right)^T\right)\left(\boldsymbol{F}\left(\boldsymbol{x}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}\left(\boldsymbol{x}\right)^T\right)^{-1}.$$

*Démonstration.* Nous avons :

$$\left(Y_{x_1},\ldots,Y_{x_n}\right)=\boldsymbol{U}\boldsymbol{F}\left(x_1,\ldots,x_n\right)+V\left(W_{x_1},\ldots,W_{x_n}\right)$$

où $\left(W_{x_1},\ldots,W_{x_n}\right)$ suit la distribution $\mathcal{N}\left(0,\boldsymbol{\Sigma}\right)$ et où $\boldsymbol{U}:=\left(U_1,U_2,\ldots,U_\ell\right)$.
Soit $f$ la densité du vecteur aléatoire $\left(\boldsymbol{U},Y_{x_1},\ldots,Y_{x_n}\right)$ :

$$f\left(\boldsymbol{u},\boldsymbol{\zeta}\right)=\frac{1}{(2m)^\ell}\frac{1}{\frac{1}{\epsilon}-\epsilon}\int_\epsilon^{\frac{1}{\epsilon}}\frac{1}{\left(\sqrt{2\pi}\right)^n v^n\sqrt{\left|\boldsymbol{\Sigma}\right|}}\exp\left(-\frac{1}{2v^2}\left(\boldsymbol{\zeta}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\zeta}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)^T\right)\,\mathrm{d}v\;\mathbb{1}_{[-m,m]^\ell}\left(\boldsymbol{u}\right)$$

avec $\boldsymbol{\zeta}:=\left(y_{x_1},y_{x_2},\ldots,y_{x_n}\right)$ et $\boldsymbol{u}:=\left(u_1,u_2,\ldots,u_\ell\right)$.

La densité conditionnelle du vecteur aléatoire $\boldsymbol{U}$ sachant $\left(Y_{x_i}=y_i\right)_{1\leq i\leq n}$ est :

$$\boldsymbol{u}\mapsto\frac{f\left(\boldsymbol{u},\boldsymbol{y}\right)}{\int_{[-m,m]^\ell}f\left(\boldsymbol{u}',\boldsymbol{y}\right)\,\mathrm{d}\boldsymbol{u}'}=:g\left(\boldsymbol{u}\right).$$

Après simplification, nous avons :

$$g\left(\boldsymbol{u}\right)=\frac{\int_\epsilon^{\frac{1}{\epsilon}}\frac{1}{v^n}\exp\left(-\frac{1}{2v^2}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)^T\right)\,\mathrm{d}v}{\int_{[-m,m]^\ell}\int_\epsilon^{\frac{1}{\epsilon}}\frac{1}{v^n}\exp\left(-\frac{1}{2v^2}\left(\boldsymbol{y}-\boldsymbol{u}'\boldsymbol{F}\left(\boldsymbol{x}\right)\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-\boldsymbol{u}'\boldsymbol{F}\left(\boldsymbol{x}\right)\right)^T\right)\,\mathrm{d}\boldsymbol{u}'\,\mathrm{d}v}\;\mathbb{1}_{[-m,m]^\ell}\left(\boldsymbol{u}\right).$$

Notons

$$\mathrm{NUM}\left(\boldsymbol{u}\right):=\int_\epsilon^{\frac{1}{\epsilon}}\frac{1}{v^n}\exp\left(-\frac{1}{2v^2}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)^T\right)\,\mathrm{d}v\;\mathbb{1}_{[-m,m]^\ell}\left(\boldsymbol{u}\right).$$

Avec le changement de variable $v\mapsto q=\frac{1}{2v^2}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)^T$ et la définition de la fonction d'Euler $\Gamma$, nous obtenons :

$$\mathrm{NUM}\left(\boldsymbol{u}\right)\underset{\epsilon\to0}{\longrightarrow}2^{\frac{n-3}{2}}\left(\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)^T\right)^{-\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)\;\mathbb{1}_{[-m,m]^\ell}\left(\boldsymbol{u}\right).$$

Pour $m$ fixé, l'expression précédente permet, après quelques calculs, de démontrer que la distribution conditionnelle du vecteur aléatoire $\boldsymbol{U}$ tend vers une distribution de Student multivariée restreinte à l'hypercube $[-m,m]^\ell$, lorsque $\epsilon$ tend vers zéro.
Notons

$$\mathrm{DEN}:=\int_{[-m,m]^\ell}\int_\epsilon^{\frac{1}{\epsilon}}\frac{1}{v^n}\exp\left(-\frac{1}{2v^2}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)^T\right)\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}v.$$

Considérons le polynôme en $\boldsymbol{u}$ suivant :

$$(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}(\boldsymbol{x}))\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}(\boldsymbol{x}))^T=\boldsymbol{u}\boldsymbol{a}\boldsymbol{u}^T+2\boldsymbol{b}\boldsymbol{u}^T+c.$$

Identifions $\boldsymbol{a}$, $b$ et $c$ :

$$\begin{aligned}
&\boldsymbol{a}=\boldsymbol{F}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T,\text{ matrice de type }(\ell,\ell),\\
&\boldsymbol{b}=-\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T,\text{ vecteur ligne de dimension }\ell,\\
&c=\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T,\text{ scalaire.}
\end{aligned}$$

Nous avons donc, par convergence monotone :

$$\mathrm{DEN}\underset{m\to+\infty}{\longrightarrow}\int_{\epsilon}^{\frac{1}{\epsilon}}\frac{(2\pi)^{\ell/2}}{v^{n-\ell}\,|\boldsymbol{a}|^{1/2}}\exp\left(-\frac{1}{2v^2}\left(c-\boldsymbol{b}\boldsymbol{a}^{-1}\boldsymbol{b}^T\right)\right)\,\mathrm{d}v.$$

Avec le changement de variable $v\mapsto q=\frac{1}{2v^2}\left(c-\boldsymbol{b}\boldsymbol{a}^{-1}\boldsymbol{b}^T\right)$ et la définition de la fonction d'Euler $\Gamma$, nous obtenons :

$$\mathrm{DEN}\underset{m\to+\infty,\epsilon\to0}{\longrightarrow}\frac{(2\pi)^{\ell/2}}{|\boldsymbol{a}|^{1/2}}2^{\frac{n-\ell-3}{2}}\left(c-\boldsymbol{b}\boldsymbol{a}^{-1}\boldsymbol{b}^T\right)^{-\frac{n-\ell-1}{2}}\Gamma\left(\frac{n-\ell-1}{2}\right).$$

Ainsi, nous déduisons la limite de la densité $g$ :

$$g(\boldsymbol{u})\underset{m\to+\infty,\epsilon\to0}{\longrightarrow}\frac{2^{\frac{n-3}{2}}\left((\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}(\boldsymbol{x}))\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}(\boldsymbol{x}))^T\right)^{-\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}{\frac{(2\pi)^{\ell/2}}{|\boldsymbol{a}|^{1/2}}2^{\frac{n-\ell-3}{2}}\left(c-\boldsymbol{b}\boldsymbol{a}^{-1}\boldsymbol{b}^T\right)^{-\frac{n-\ell-1}{2}}\Gamma\left(\frac{n-\ell-1}{2}\right)}.$$

Comme nous avons

$$(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}(\boldsymbol{x}))\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}(\boldsymbol{x}))^T=(\boldsymbol{u}-\boldsymbol{l})\,\boldsymbol{Q}^{-1}\,(\boldsymbol{u}-\boldsymbol{l})^T+M$$

avec

$$\begin{aligned}
&\boldsymbol{Q}=\left(\boldsymbol{F}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)^{-1}\\
&\boldsymbol{l}=\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)\left(\boldsymbol{F}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)^{-1}\\
&M=c-\boldsymbol{b}\boldsymbol{a}^{-1}\boldsymbol{b}^T
\end{aligned}$$

On a

$$g(\boldsymbol{u})\underset{m\to+\infty,\epsilon\to0}{\longrightarrow}\frac{2^{\frac{n-3}{2}}\left(M+(\boldsymbol{u}-\boldsymbol{l})\,\boldsymbol{Q}^{-1}\,(\boldsymbol{u}-\boldsymbol{l})^T\right)^{-\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}{(2\pi)^{\ell/2}\,|\boldsymbol{Q}|^{1/2}\,2^{\frac{n-\ell-3}{2}}M^{-\frac{n-\ell-1}{2}}\Gamma\left(\frac{n-\ell-1}{2}\right)}.$$

C'est-à-dire

$$g(\boldsymbol{u})\underset{m\to+\infty,\epsilon\to0}{\longrightarrow}\frac{1}{\pi^{\ell/2}M^{\ell/2}\,|\boldsymbol{Q}|^{1/2}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-\ell-1}{2}\right)}\left(1+\frac{1}{M}\,(\boldsymbol{u}-\boldsymbol{l})\,\boldsymbol{Q}^{-1}\,(\boldsymbol{u}-\boldsymbol{l})^T\right)^{-\frac{n-1}{2}}.$$

Nous reconnaissons une distribution de Student multivariée de dimension $\ell$, de degré $n-\ell-1$, de paramètres de position, le vecteur :

$$\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)\left(\boldsymbol{F}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)^{-1}$$

et de paramètres d'échelle, la matrice :

$$\frac{1}{n-\ell-1}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T-\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}\left(\boldsymbol{x}\right)^T\right)\left(\boldsymbol{F}\left(\boldsymbol{x}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}\left(\boldsymbol{x}\right)^T\right)^{-1}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}\left(\boldsymbol{x}\right)^T\right)^T\right)\left(\boldsymbol{F}\left(\boldsymbol{x}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}\left(\boldsymbol{x}\right)^T\right)^{-1}.$$

□

### 1.5.3 Distribution conditionnelle de la variance aléatoire

Il nous est également possible d'expliciter la distribution de la variable aléatoire $V^2$ conditionnellement aux données $(Y_{x_i}=y_i)_{1\leq i\leq n}$.

**Théorème 9.** *La distribution conditionnelle de la variable aléatoire $V$ sachant que $(Y_{x_i}=y_i)_{1\leq i\leq n}$ est donnée par l'expression explicite de sa densité.*
*Lorsque $m$ tend vers l'infini, pour $\epsilon$ fixé, la distribution conditionnelle de la variable aléatoire $V^2$ tend vers une distribution inverse du $\chi^2$ réduite restreinte à l'espace $\left[\epsilon,\frac{1}{\epsilon}\right]$.*
*Quand le paramètre $m$ tend vers l'infini et $\epsilon$ tend vers zéro, la distribution conditionnelle de la variable aléatoire $V^2$ devient une distribution inverse du $\chi^2$ réduite de degré $n-\ell-1$ et de paramètre d'échelle :*

$$\sqrt{\frac{1}{n-\ell-1}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T-\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}\left(\boldsymbol{x}\right)^T\right)\left(\boldsymbol{F}\left(\boldsymbol{x}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}\left(\boldsymbol{x}\right)^T\right)^{-1}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}\left(\boldsymbol{x}\right)^T\right)^T\right)}.$$

*Démonstration.* Nous avons

$$\left(Y_{x_1},\ldots,Y_{x_n}\right)=\boldsymbol{U}\boldsymbol{F}\left(x_1,\ldots,x_n\right)+V\left(W_{x_1},\ldots,W_{x_n}\right)$$

où $\left(W_{x_1},\ldots,W_{x_n}\right)$ suit la distribution $\mathcal{N}\left(0,\boldsymbol{\Sigma}\right)$.
Soit $f$ la densité du vecteur aléatoire $\left(V,Y_{x_1},\ldots,Y_{x_n}\right)$ :

$$f\left(v,\boldsymbol{\zeta}\right)=\frac{1}{\left(2m\right)^\ell}\frac{1}{\frac{1}{\epsilon}-\epsilon}\int_{[-m,m]^\ell}\frac{1}{\left(\sqrt{2\pi}\right)^n v^n\sqrt{\left|\boldsymbol{\Sigma}\right|}}\exp\left(-\frac{1}{2v^2}\left(\boldsymbol{\zeta}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\zeta}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)^T\right)\,\mathrm{d}\boldsymbol{u}\;\mathbb{1}_{\left[\epsilon,\frac{1}{\epsilon}\right]}\left(v\right)$$

avec $\boldsymbol{\zeta}:=\left(y_{x_1},y_{x_2},\ldots,y_{x_n}\right)$.

La densité conditionnelle $V$ sachant $(Y_{x_i}=y_i)_{1\leq i\leq n}$ est :

$$v\mapsto\frac{f\left(v,\boldsymbol{y}\right)}{\int_\epsilon^{\frac{1}{\epsilon}}f\left(v',\boldsymbol{y}\right)\,\mathrm{d}v'}=:g\left(v\right).$$

Après simplification, nous avons :

$$g\left(v\right)=\frac{\int_{[-m,m]^\ell}\frac{1}{v^n}\exp\left(-\frac{1}{2v^2}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)^T\right)\,\mathrm{d}\boldsymbol{u}}{\int_\epsilon^{\frac{1}{\epsilon}}\int_{[-m,m]^\ell}\frac{1}{v'^n}\exp\left(-\frac{1}{2v'^2}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)^T\right)\,\mathrm{d}v'\,\mathrm{d}\boldsymbol{u}}\;\mathbb{1}_{\left[\epsilon,\frac{1}{\epsilon}\right]}\left(v\right).$$

Notons

$$\mathrm{NUM}\left(v\right) := \int_{[-m,m]^{\ell}} \frac{1}{v^n} \exp\left(-\frac{1}{2v^2}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)^T\right)\,\mathrm{d}\boldsymbol{u}.$$

Nous avons

$$\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)^T = \boldsymbol{u}\boldsymbol{a}\boldsymbol{u}^T + 2\boldsymbol{b}\boldsymbol{u}^T + c$$

avec

$\boldsymbol{a} = \boldsymbol{F}\left(\boldsymbol{x}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}\left(\boldsymbol{x}\right)^T$, matrice de type $(\ell,\ell)$,
$\boldsymbol{b} = -\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}\left(\boldsymbol{x}\right)^T$, vecteur ligne de dimension $\ell$,
$c = \boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T$, scalaire.

En appliquant le même raisonnement que dans la section 1.5.2, on montre que

$$\mathrm{NUM}\left(v\right) \underset{m\to+\infty}{\longrightarrow} \frac{\left(2\pi\right)^{\ell/2}}{v^{n-\ell}\left|\boldsymbol{a}\right|^{1/2}} \exp\left(-\frac{1}{2v^2}\left(c-\boldsymbol{b}\boldsymbol{a}^{-1}\boldsymbol{b}^T\right)\right).$$

Pour $\epsilon$ fixé, cette précédente expression, permet, après quelques calculs de démontrer que la distribution conditionnelle de la variable aléatoire $V^2$ tend vers une distribution inverse du $\chi^2$ réduite restreinte à l'espace $\left[\epsilon, \frac{1}{\epsilon}\right]$ lorsque $m$ tend vers l'infini.
Notons

$$\mathrm{DEN} := \int_{\epsilon}^{\frac{1}{\epsilon}} \int_{[-m,m]^{\ell}} \frac{1}{v^n} \exp\left(-\frac{1}{2v^2}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-\boldsymbol{u}\boldsymbol{F}\left(\boldsymbol{x}\right)\right)^T\right)\,\mathrm{d}v\,\mathrm{d}\boldsymbol{u}.$$

Par convergence monotone, l'intégrale de $\mathrm{NUM}\left(v\right)$ tend vers l'intégrale de sa limite quand $m$ tend vers l'infini. Par conséquent, nous avons

$$\mathrm{DEN} := \int_{\epsilon}^{\frac{1}{\epsilon}} \frac{\left(2\pi\right)^{\ell/2}}{v^{n-\ell}\left|\boldsymbol{a}\right|^{1/2}} \exp\left(-\frac{1}{2v^2}\left(c-\boldsymbol{b}\boldsymbol{a}^{-1}\boldsymbol{b}^T\right)\right)\,\mathrm{d}v.$$

Avec le changement de variable $v \mapsto q = \frac{1}{2v^2}\left(c-\boldsymbol{b}\boldsymbol{a}^{-1}\boldsymbol{b}^T\right)$ et la définition de la fonction d'Euler $\Gamma$, nous obtenons :

$$\mathrm{DEN} \underset{m\to+\infty,\epsilon\to 0}{\longrightarrow} \frac{\left(2\pi\right)^{\ell/2}}{\left|\boldsymbol{a}\right|^{1/2}} 2^{\frac{n-\ell-3}{2}} \left(c-\boldsymbol{b}\boldsymbol{a}^{-1}\boldsymbol{b}^T\right)^{-\frac{n-\ell-1}{2}} \Gamma\left(\frac{n-\ell-1}{2}\right).$$

Ainsi, on obtient la limite de la densité $g$ :

$$g\left(v\right) \underset{m\to+\infty,\epsilon\to 0}{\longrightarrow} \frac{\frac{\left(2\pi\right)^{\ell/2}}{v^{n-\ell}\left|\boldsymbol{a}\right|^{1/2}} \exp\left(-\frac{1}{2v^2}\left(c-\boldsymbol{b}\boldsymbol{a}^{-1}\boldsymbol{b}^T\right)\right)}{\frac{\left(2\pi\right)^{\ell/2}}{\left|\boldsymbol{a}\right|^{1/2}} 2^{\frac{n-\ell-3}{2}} \left(c-\boldsymbol{b}\boldsymbol{a}^{-1}\boldsymbol{b}^T\right)^{-\frac{n-\ell-1}{2}} \Gamma\left(\frac{n-\ell-1}{2}\right)} \mathbb{1}_{[v>0]}.$$

C'est-à-dire :

$$g\left(v\right) \underset{m\to+\infty,\epsilon\to 0}{\longrightarrow} \frac{\frac{1}{v^{n-\ell}} \exp\left(-\frac{1}{2v^2}\left(c-\boldsymbol{b}\boldsymbol{a}^{-1}\boldsymbol{b}^T\right)\right)}{2^{\frac{n-\ell-3}{2}} \left(c-\boldsymbol{b}\boldsymbol{a}^{-1}\boldsymbol{b}^T\right)^{-\frac{n-\ell-1}{2}} \Gamma\left(\frac{n-\ell-1}{2}\right)} \mathbb{1}_{[v>0]}.$$

Notons que la variable aléatoire $W := V^2$ admet pour densité $h(w) := \frac{1}{2\sqrt{w}} g(\sqrt{w})$. Nous avons :

$$h\left(v^2\right) \underset{m\to+\infty,\epsilon\to 0}{\longrightarrow} \frac{\frac{1}{v^{n-\ell-1}} \exp\left(-\frac{1}{2v^2}\left(c - \boldsymbol{b}\boldsymbol{a}^{-1}\boldsymbol{b}^T\right)\right)}{2^{\frac{n-\ell-1}{2}}\left(c - \boldsymbol{b}\boldsymbol{a}^{-1}\boldsymbol{b}^T\right)^{-\frac{n-\ell-1}{2}}\Gamma\left(\frac{n-\ell-1}{2}\right)}.$$

Nous reconnaissons que la variable aléatoire $V^2$ suit la distribution inverse du $\chi^2$ réduite de degré $n-\ell-1$ et de paramètre d'échelle :

$$\sqrt{\frac{1}{n-\ell-1}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T - \left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)\left(\boldsymbol{F}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)^{-1}\left(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(\boldsymbol{x})^T\right)^T\right)}.$$

□

## 1.6 Paramétrisation des modèles

Les modèles aléatoires présentés dans les sections précédentes dépendent d'une fonction de corrélation. Celle-ci permet de lier les données observées de manière à représenter leurs influences dans l'espace des facteurs. Le choix de la fonction de corrélation s'effectuant a priori, impacte d'autant plus significativement le posterior que le nombre de données utilisées pour le conditionnement est faible. C'est pourquoi, ce choix est une des difficultés majeures dans la modélisation à l'aide de champs aléatoires et ce sujet est hélas souvent évité. Cependant, dans certains cas particuliers, il peut être justifié par les connaissances de l'expert en charge de la modélisation (Cornford et al. (2002)).

Dans la section 1.6.1, nous présentons quelques exemples de fonctions de corrélation classées par familles.

La détermination des paramètres de la fonction de corrélation est le sujet des méthodes d'étalonnage. La plus populaire d'entre elles, est la méthode du maximum de vraisemblance qui sera détaillée dans la section 1.6.2. Puis dans la section 1.6.3, est décrite une méthode qui permet d'évaluer le modèle par validation croisée. Nous présenterons également, dans la section 1.6.4, l'approche qui consiste à considérer la fonction de corrélation comme aléatoire, donnée par une loi a priori sur ses paramètres.

Il n'existe pas, à notre connaissance, d'étude complète et rigoureuse sur la problématique du choix de la "meilleure" fonction de corrélation pour un problème donné. Cependant, Abrahamsen (1997) liste quelques propriétés (de continuité et de dérivabilité) des fonctions de corrélations qui peuvent guider le choix de l'expert.

### 1.6.1 Fonctions de corrélation

Les fonctions de corrélations sont généralement classées par familles (Abrahamsen (1997)). Une étude complète des principales caractéristiques des fonctions usuelles est présentée dans le livre de Rasmussen et Williams (2006)). La condition générale d'admissibilité d'une fonction $(x, x') \mapsto \rho(x, x')$ comme fonction de corrélation est le fait d'être symétrique, $\rho(x, x') = \rho(x', x)$ et de type positif, c'est-à-dire que pour tous $x_1, x_2, \ldots, x_n \in X$, la matrice $(\rho(x_i, x_j))_{1\leq i,j\leq n}$ définit une forme quadratique positive. Voici une sélection des plus

connues, présentées d'après les critères de sélection de modèles couramment retenus dans la littérature. (On se place dans le cas où $X$ est une partie d'un espace euclidien.)

#### 1.6.1.1 Stationnaire

Une fonction de corrélation $\rho(x, x')$ est *stationnaire* si elle est invariante par translation, c'est-à-dire si elle est fonction de $x - x'$.

**Isotrope** Une fonction de corrélation $\rho(x, x')$ est *isotrope* si elle est invariante par les isométries affines, c'est-à-dire si elle est en fonction de $|x - x'|$ où $|x|$ est la norme euclidienne de $x$.

C'est la plus populaire des familles de fonctions de corrélation. Les fonctions de corrélations isotropes requièrent, en général, l'ajustement d'un faible nombre de paramètres. À l'inverse, les fonctions anisotropes font intervenir un nombre de paramètre dépendant de la dimension $D$ de l'espace des facteurs $X$.

**Exemple.** $\gamma$-exponentielle :

$$\rho(x, x') = \exp\left(-\left(\frac{|x - x'|}{l}\right)^{\gamma}\right).$$

L'exposant $\gamma \in (0, 2]$ contrôle la régularité du processus aléatoire (qui est infiniment dérivable seulement quand $\gamma = 2$) et le réel positif $l$ est la longueur caractéristique définissant un domaine autour de chaque point $x$ où sa corrélation avec un autre point est non négligeable.

**Exemple.** Matérn :

$$\rho(x, x') = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}\,|x - x'|}{l}\right)^{\nu} K_{\nu}\left(\frac{\sqrt{2\nu}\,|x - x'|}{l}\right)$$

où $l > 0$, $\nu > 0$ et $K_{\nu}$ est la fonction modifiée de Bessel. Le processus aléatoire est de classe $C^k$ si et seulement si $\nu > k$. Quand $\nu$ tend vers l'infini, cela conduit à la fonction de corrélation $\gamma$-exponentielle avec $\gamma = 2$.

**Exemple.** Quadratique rationnelle :

$$\rho(x, x') = \left(1 + \frac{|x - x'|^2}{2\alpha l^2}\right)^{-\alpha}$$

où $l > 0$ et $\alpha > 0$.

**Anisotrope** L'isotropie impose un choix cohérent des échelles (ou des unités) sur les différentes coordonnées pour qu'elles soient bien prise en compte par la norme euclidienne. Lorsque ce choix s'avère difficile, une fonction anisotrope peut être plus adéquate.

Currin et al. (1988) propose l'utilisation d'une fonction *stationnaire définie par un produit de fonctions de coordonnées* :

$$\rho(x, x') = \prod_{i=1}^{D} f_i(x^{(i)} - x'^{(i)})$$

où on note $x = \left(x^{(1)}, x^{(1)}, \ldots, x^{(D)}\right)$.

**Exemple.** Mockus et al. (1997) :

$$\rho(x, x') = \prod_{i=1}^{D} \left(1 - \frac{\left|x^{(i)} - x'^{(i)}\right|}{2}\right)$$

où $x, x' \in [-1, 1]^D$.

Il est également possible de considérer des variantes anisotropes des fonctions de corrélation isotropes en utilisant, non-plus la distance euclidienne, mais une distance quadratique de la forme $(x, x')\, M\, (x, x')^T$ avec une matrice $M$ définie-positive fixée (voir Rasmussen et Williams (2006)). Ci-dessous, un exemple où $M$ est une matrice diagonale.

**Exemple.** $\gamma$-exponentielle :

$$\rho\left(x, x'\right) := \exp\left(-\left(\sum_{i=1}^{D}\left(\frac{x^{(i)} - x'^{(i)}}{l_i}\right)^2\right)^{\frac{\gamma}{2}}\right) \quad \text{pour} \quad x, x' \in X.$$

Comme pour le cas isotrope, l'exposant $\gamma \in (0, 2]$ contrôle la régularité du processus aléatoire (infiniment dérivable seulement quand $\gamma = 2$). Les réels positifs $l_1, l_2, \ldots, l_D$ sont les longueurs caractéristiques de chacune des dimensions qui définissent l'espace elliptique autour de chaque point $x$ caractérisant, en quelque sorte, la zone d'influence de celui-ci.

#### 1.6.1.2 Non-Stationnaire

**Produit scalaire** Une fonction de corrélation $\rho(x, x')$ est dite *produit scalaire* si elle est en fonction de $x \cdot x'$ (voir Rasmussen et Williams (2006)).

**Exemple.** Linéaire :

$$\rho(x, x') = \sigma_0 + x \cdot x'$$

si $\sigma_0$ est nul, la fonction de corrélation est dite homogène.

**Exemple.** Polynomiale :

$$\rho(x, x') = \left(\sigma_0 + x \cdot x'\right)^p$$

où $p > 0$.

### 1.6.2 Maximum de vraisemblance

La méthode classique d'estimation par maximum de vraisemblance (EMV) définit un estimateur d'un vecteur de paramètres inconnu $\boldsymbol{\theta}$ d'une distribution de probabilité $f_{\boldsymbol{\theta}}$. Cet estimateur est la valeur $\boldsymbol{\theta}_{\mathbf{max}}$ qui maximise la densité de la distribution d'un échantillon de loi $f_{\boldsymbol{\theta}}$ calculée aux valeurs observées de cet échantillon.

Sous la même dénomination, cette méthode a été adaptée aux problèmes d'identification des paramètres d'un champ aléatoire $(Y_x)_{x\in X}$ sous la donnée d'une famille de valeurs $Y_{x_1}, Y_{x_2}, \ldots, Y_{x_n}$.

Dans le contexte de la recherche d'un modèle aléatoire pour approcher un phénomène déterministe, la méthode EMV permet l'étalonnage des paramètres pour un champ aléatoire a priori $(Y_x)_{x\in X}$ sachant que $Y_{x_i} = y_i$, $1 \leq i \leq n$ (de façon sûre).

Bien que nous n'ayons pas trouvé de raisonnement théorique justifiant l'application de la méthode EMV à cette problématique, cette méthode est devenue, en pratique, une approche classique dans la théorie des champs aléatoires, après le choix de fonctions de corrélation dépendant de paramètres à ajuster, voir Rasmussen et Williams (2006), Stein (1999) et Robert (2007). Ces derniers décrivent cette approche et discutent sur ses capacités et limitations telles que les problématiques numériques qu'entraînent des échantillons de données trop grands, les optima multiples pour les petits échantillons et les problèmes de sur-apprentissage ("overfitting") pour lesquels un échantillon n'est pas représentatif du phénomène étudié étant donné la complexité du modèle retenu.

Dans les deux sous-sections suivantes, nous calculons exactement la vraisemblance du modèle gaussien et nous proposons une approximation dans le cas du modèle de Student. Pour cela, dans un premier temps, nous considérons une famille paramétrique de fonctions de corrélation $k_{\boldsymbol{\theta}}$ dépendant du vecteur de paramètres $\boldsymbol{\theta}$, qui définit ainsi la matrice de corrélation paramétrique $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = (k_{\boldsymbol{\theta}}(x_i, x_j))_{1\leq i,j\leq n}$.

#### 1.6.2.1 Application au modèle gaussien

La méthode EMV, appliquée au champ aléatoire homogène décrit dans la section 1.2, corollaire 1, consiste à choisir $\boldsymbol{\theta}$, $\mu$ et $\sigma$ tels qu'ils maximisent

$$f(\boldsymbol{\theta}, \mu, \sigma) = \frac{1}{(2\pi\sigma^2)^{n/2} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{1/2}} \exp\left(-\frac{1}{2\sigma^2} (\boldsymbol{y} - \mu \boldsymbol{1}) \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{y} - \mu \boldsymbol{1})^T\right)$$

avec $\boldsymbol{y} := (y_1, y_2, \ldots, y_n)$ et $\boldsymbol{1} := (1, 1, \ldots, 1) \in \mathbb{R}^n$.

Cela revient à minimiser la quantité

$$-\ln(f(\boldsymbol{\theta}, \mu, \sigma)) = \frac{n}{2}\ln(2\pi) + \frac{n}{2}\ln(\sigma^2) + \frac{1}{2}\ln(|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|) + \frac{1}{2\sigma^2} (\boldsymbol{y} - \mu \boldsymbol{1}) \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{y} - \mu \boldsymbol{1})^T.$$

Pour estimer $\mu$ et $\sigma$, la méthode EMV conduit à résoudre :

$$\frac{\partial f(\boldsymbol{\theta}, \mu, \sigma)}{\partial \sigma} = 0 \quad \text{et} \quad \frac{\partial f(\boldsymbol{\theta}, \mu, \sigma)}{\partial \mu} = 0.$$

On retrouve le résultat classique suivant (voir Currin et al. (1988) ou Stein (1999)) :

$$\sigma_{\boldsymbol{\theta}}^2 = \frac{s_{\boldsymbol{\theta}}}{n} \quad \text{avec} \quad s_{\boldsymbol{\theta}} := (\boldsymbol{y} - \mu_{\boldsymbol{\theta}} \boldsymbol{1}) \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{y} - \mu_{\boldsymbol{\theta}} \boldsymbol{1})^T$$

et

$$\mu_{\boldsymbol{\theta}} = \frac{\boldsymbol{1}\boldsymbol{\Sigma_\theta}^{-1}\boldsymbol{y}^T}{\boldsymbol{1}\boldsymbol{\Sigma_\theta}^{-1}\boldsymbol{1}^T}.$$

Fait remarquable : avec cette valeur de $\mu_{\boldsymbol{\theta}}$, la moyenne du champ aléatoire conditionné dans le corollaire 1 coïncide avec celles des théorèmes 2 et 4.
Avec cette valeur de $\sigma^2_{\boldsymbol{\theta}}$, les variances des champs aléatoires conditionnés dans le corollaire 1 et dans le théorème 4 présentent des similitudes (cf. tableau 1.5).

Pour la recherche de $\boldsymbol{\theta}$, il nous reste à minimiser :

$$\frac{n}{2}\ln(2\pi) + \frac{n}{2}\ln\left(\sigma^2_{\boldsymbol{\theta}}\right) + \frac{1}{2}\ln\left(|\boldsymbol{\Sigma_\theta}|\right) + \frac{n}{2},$$

c'est-à-dire :

$$n\ln(s_{\boldsymbol{\theta}}) + \ln\left(|\boldsymbol{\Sigma_\theta}|\right).$$

#### 1.6.2.2 Application au modèle de Student

La méthode EMV, appliquée au champ aléatoire décrit dans la section 1.4, théorème 4, consiste à choisir $\boldsymbol{\theta}$ maximisant la densité du vecteur aléatoire $(Y_{x_1}, \ldots, Y_{x_n})$ :

$$f_{m,\epsilon}(\boldsymbol{\theta}) = \frac{1}{2m}\frac{1}{\frac{1}{\epsilon}-\epsilon}\int_{-m}^{m}\int_{\epsilon}^{\frac{1}{\epsilon}} \frac{1}{\left(\sqrt{2\pi}\right)^n v^n\sqrt{|\boldsymbol{\Sigma_\theta}|}}\exp\left(-\frac{1}{2v^2}(\boldsymbol{y}-\boldsymbol{u})\boldsymbol{\Sigma_\theta}^{-1}(\boldsymbol{y}-\boldsymbol{u})^T\right)\,\mathrm{d}u\,\mathrm{d}v$$

avec $\boldsymbol{y} := (y_1, y_2, \ldots, y_n)$ et $\boldsymbol{u} := (u, u, \ldots, u) \in \mathbb{R}^n$.

L'estimateur du maximum de vraisemblance $\boldsymbol{\theta_{\max}}(\boldsymbol{m}, \boldsymbol{\epsilon})$ maximise $f_{m,\epsilon}(\boldsymbol{\theta})$. Il maximise également :

$$\tilde{f}_{m,\epsilon}(\boldsymbol{\theta}) = \int_{-m}^{m}\int_{\epsilon}^{\frac{1}{\epsilon}} \frac{1}{v^n\sqrt{|\boldsymbol{\Sigma_\theta}|}}\exp\left(-\frac{1}{2v^2}(\boldsymbol{y}-\boldsymbol{u})\boldsymbol{\Sigma_\theta}^{-1}(\boldsymbol{y}-\boldsymbol{u})^T\right)\,\mathrm{d}u\,\mathrm{d}v.$$

Nous ne connaissons pas d'expression analytique de $\tilde{f}_{m,\epsilon}(\boldsymbol{\theta})$ et a fortiori de $\boldsymbol{\theta_{\max}}(\boldsymbol{m}, \boldsymbol{\epsilon})$ mais puisque nous nous intéressons aux larges valeurs de $m$ et aux petites valeurs de $\epsilon$, nous considérons la valeur limite $\tilde{f}_{\infty,0}(\boldsymbol{\theta})$. Ainsi, nous proposons d'étudier :

$$\tilde{f}_{\infty,0}(\boldsymbol{\theta}) = \int_{-\infty}^{+\infty}\int_{0}^{+\infty} \frac{1}{v^n\sqrt{|\boldsymbol{\Sigma_\theta}|}}\exp\left(-\frac{1}{2v^2}(\boldsymbol{y}-\boldsymbol{u})\boldsymbol{\Sigma_\theta}^{-1}(\boldsymbol{y}-\boldsymbol{u})^T\right)\,\mathrm{d}u\,\mathrm{d}v.$$

Un petit calcul donne :

$$\tilde{f}_{\infty,0}(\boldsymbol{\theta}) = \frac{\sqrt{\pi}2^{\frac{n-3}{2}}\left((\boldsymbol{y}-\mu_{\boldsymbol{\theta}}\boldsymbol{1})\boldsymbol{\Sigma_\theta}^{-1}(\boldsymbol{y}-\mu_{\boldsymbol{\theta}}\boldsymbol{1})^T\right)^{\frac{2-n}{2}}\Gamma\left(\frac{n-2}{2}\right)}{\sqrt{|\boldsymbol{\Sigma_\theta}|\left(\boldsymbol{1}\boldsymbol{\Sigma_\theta}^{-1}\boldsymbol{1}^T\right)}},$$

c'est-à-dire avec les notations $s_{\boldsymbol{\theta}}$ et $\mu_{\boldsymbol{\theta}}$ introduites dans la section précédente :

$$\tilde{f}_{\infty,0}(\boldsymbol{\theta}) = \frac{\sqrt{\pi}2^{\frac{n-3}{2}}(s_{\boldsymbol{\theta}})^{\frac{2-n}{2}}\Gamma\left(\frac{n-2}{2}\right)}{\sqrt{|\boldsymbol{\Sigma_\theta}|\left(\boldsymbol{1}\boldsymbol{\Sigma_\theta}^{-1}\boldsymbol{1}^T\right)}}$$

où

$$s_{\boldsymbol{\theta}} := (\boldsymbol{y} - \mu_{\boldsymbol{\theta}} \boldsymbol{1})\, \boldsymbol{\Sigma_\theta}^{-1} (\boldsymbol{y} - \mu_{\boldsymbol{\theta}} \boldsymbol{1})^T \quad \text{et} \quad \mu_{\boldsymbol{\theta}} := \frac{\boldsymbol{1}\boldsymbol{\Sigma_\theta}^{-1}\boldsymbol{y}^T}{\boldsymbol{1}\boldsymbol{\Sigma_\theta}^{-1}\boldsymbol{1}^T}.$$

Maximiser $\tilde{f}_{\infty,0}(\boldsymbol{\theta})$ est équivalent à minimiser $-\ln\left(\tilde{f}_{\infty,0}(\boldsymbol{\theta})\right)$ et après simplification, nous obtenons :

$$\boldsymbol{\theta_{\max}}\,(\boldsymbol{\infty}, \mathbf{0}) \quad \text{minimise} \quad n \ln(s_{\boldsymbol{\theta}}) + \ln(|\boldsymbol{\Sigma_\theta}|) + \ln\left(\frac{\boldsymbol{1}\boldsymbol{\Sigma_\theta}^{-1}\boldsymbol{1}^T}{s_{\boldsymbol{\theta}}^2}\right).$$

Cette quantité à optimiser, est à comparer avec celle du cas gaussien.

*Remarque.* Un raisonnement similaire à celui-ci, appliqué au champ aléatoire du théorème 2, amène des résultats proches, résumés dans le tableau 1.5.

### 1.6.3 Validation croisée

Une seconde approche pour aider au choix de la fonction de corrélation, est la méthode de validation croisée.

Elle consiste, dans un premier temps, à partitionner les données en deux ensembles : un jeu d'apprentissage et un jeu de validation. Puis, pour chacune des fonctions de corrélation candidates, un modèle aléatoire conditionné avec le jeu d'apprentissage est généré, puis testé sur le jeu de validation. La fonction de corrélation induisant le modèle a posteriori le plus performant est sélectionnée. Pour cela, il existe différents critères mesurant la capacité prédictive d'un modèle, comme, par exemple, l'erreur relative, quadratique (Currin et al. (1988)).

Cependant, comme nous n'avons à disposition qu'un nombre limité $n$ de données, en mettre une partie de côté à titre de validation, peut détériorer les capacités prédictives des modèles testés si le jeu d'apprentissage n'est plus représentatif de la fonction $y$ initiale et donc biaiser l'évaluation de ces modèles.

Pour limiter cela, pour chacune des fonctions de corrélation candidates, nous effectuons $n$ partitions où le jeu de validation est, pour chacune, constitué d'un seul point. Nous sélectionnons ensuite la fonction de corrélation induisant le modèle a posteriori le plus performant en moyenne sur ces $n$ partitions. Cette méthode nommée "Leave-one-out cross validation" se révèle être, pour nous, la méthode de validation croisée la plus pertinente ; les modèles construits sont le mieux renseignés possible.

Une fois une fonction de corrélation sélectionnée, on conditionne le champ aléatoire avec les $n$ données à notre disposition.

### 1.6.4 Paramètres vus comme aléatoires

Comme alternative à ces méthodes classiques d'identification des paramètres d'un modèle, Rasmussen et Williams (2006) se placent dans un cadre bayésien qui consiste à choisir une famille paramétrique de fonctions de corrélation et une distribution a priori sur ses paramètres, pour construire un modèle hiérarchique (décrit notamment par Santner et al. (2003)). Cette méthode nécessite l'approximation numérique d'intégrales. Les méthodes

de Monte-Carlo par chaîne de Markov (MCMC) comptent parmi les solutions les plus répandues pour réaliser ces calculs, voir Robert (2007) pour une description détaillée de ces méthodes. Une limitation de cette approche est son coût de calcul.

## 1.7 Exemples et résultats

### 1.7.1 Comparaison du champ de Student et du champ gaussien

Revenons un instant sur le champ gaussien décrit par le corollaire 1. Supposons que sa moyenne $\mu$ et sa variance $\sigma^2$ sont estimés par la méthode du maximum de vraisemblance, décrite dans la section précédente.

Il est intéressant de comparer sa moyenne et sa variance a posteriori avec celles du champ aléatoire défini dans le théorème 4 (voir aussi le tableau récapitulatif 1.5). Remarquons, que pour la même fonction de corrélation, la moyenne reste la même et la variance croît. Nous illustrons cette différence sur le simple exemple de dimension 1 suivant.

Soit $X = [-5, 5]$. Soit un jeu de 5 points $(x_1 = -4, x_2 = -3, x_3 = -1, x_4 = 0, x_5 = 2)$ avec les réponses associées $(y_1 = -2, y_2 = 0, y_3 = 1, y_4 = 2, y_5 = -1)$. Nous considérons, une fonction de corrélation arbitrairement fixée, $\rho(x, x') := \exp\left(-100\,|x - x'|^2\right)$, un champ gaussien et un champ de Student conditionnés par les données précédentes.

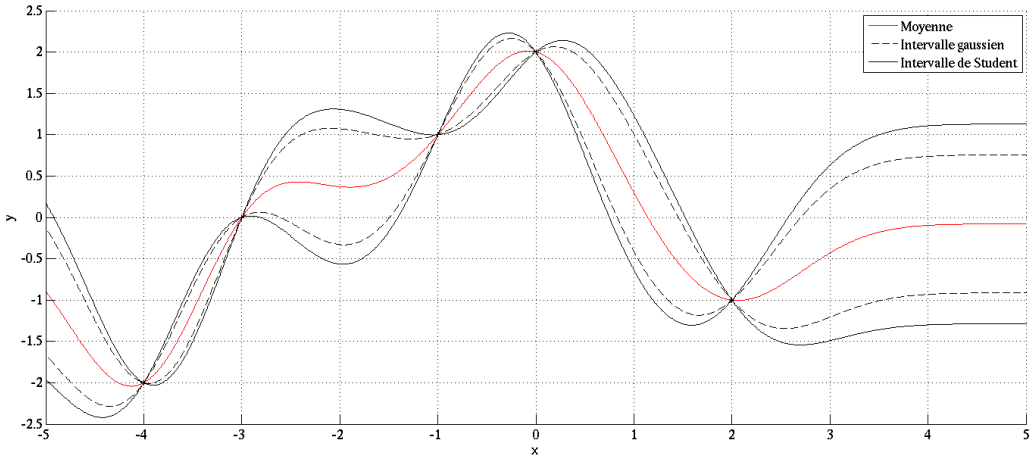



FIGURE 1.1 – Illustration d'un cas univarié. Les champs gaussien et de Student sont obtenus a posteriori à partir de 5 évaluations de la fonction inconnue $y$.

La figure 1.1 montre, pour chaque champ conditionné, la moyenne et un intervalle de confiance à 90%. Comme attendu, les intervalles de confiance sont de largeurs nulles aux 5 points observés, puisqu'en ces points les distributions dégénèrent en des masses de Dirac. L'intervalle de confiance donné par la distribution de Student est plus grand que celui obtenu par la distribution gaussienne. Le champ de Student conduit ainsi à faire des prédictions avec moins d'assurance qu'avec le champ gaussien. Il est cependant plus fiable,

puisque nous avons pris en compte toutes les valeurs possibles pour la moyenne $\mu$ et la variance $\sigma^2$, qui ont été arbitrairement fixées dans le cas du champ gaussien.

### 1.7.2 Étude de l'erreur de prédiction du champ de Student

Afin de mettre en évidence les capacités prédictives du champ de Student, nous utilisons comme fonction théorique $y$, l'exemple de la forme quadratique de dimension 5, avec $(a_1 = 4, a_2 = 4.2, a_3 = 4.4, a_4 = 4.6, a_5 = 4.8)$, défini précisément dans la section 2.4.1.1, pour étudier l'évolution de l'erreur de prédiction du modèle au fur et à mesure que le nombre de données croît.

Nous étudions le champ de Student $(Y_x)_{x \in X}$, conditionné par $n$ données, décrit par le théorème 4. Nous retenons la fonction de corrélation anisotrope $\gamma$-exponentielle (section 1.6.1.1) dont les paramètres sont ajustés par la méthode du maximum de vraisemblance (section 1.6.2).

Nous sélectionnons comme critère de la qualité de prédiction les quantités suivantes. Pour chaque $t \in X$, définissons l'erreur relative comme suit

$$\epsilon_r(t) := \left| \frac{\mathbb{E}(Y_t) - y(t)}{y(t)} \right|.$$

Nous appelons l'erreur relative moyenne, l'intégrale sur $X$ de $\epsilon_r$ et considérons la fonction de répartition complémentaire (ou fonction de queue) de $\epsilon_r$ :

$$\alpha \mapsto P(\{t \in X \mid \epsilon_r(t) > \alpha\}).$$

Ces statistiques sont calculées pour un nombre $n$ de données croissant et le résultat est représenté dans la figure 1.2. Comme attendu, l'erreur relative chute quand $n$ augmente. Considérant un jeu de 10 données, $\epsilon_r$ est inférieur à 42% avec une probabilité de 0.9 (modèle pauvre). Cette valeur descend à 1.9% avec 100 données et à 0.21% avec 1000 données.

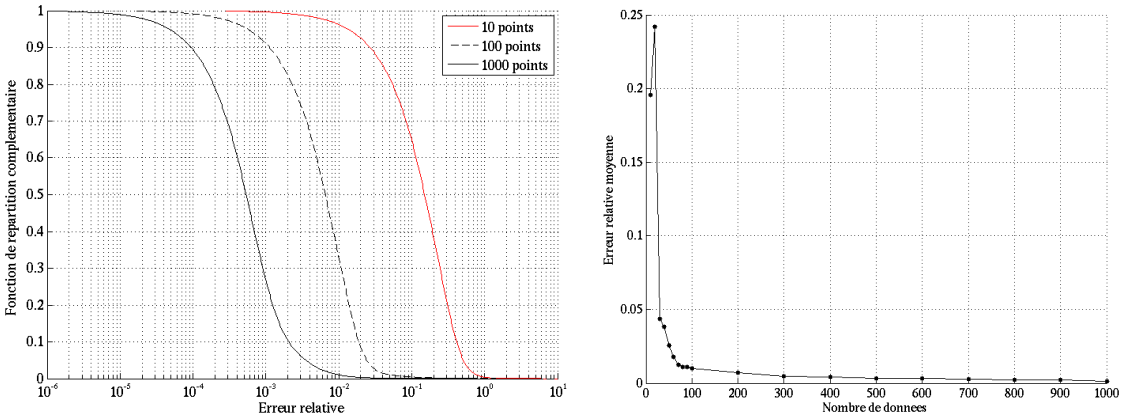



FIGURE 1.2 – Erreur commise par le modèle de Student, cas de l'exemple de la forme quadratique.

### 1.7.3 Performance de l'estimateur du maximum de vraisemblance

Pour cette courte discussion sur la méthode EMV, nous anticipons sur des définitions précises qui seront données dans le deuxième chapitre.

Plaçons-nous sous les hypothèses décrites dans le théorème 4 et considérons la fonction de corrélation $\gamma$-exponentielle isotrope décrite dans la section 1.6.1.1.

Dans cet exemple, nous cherchons à illustrer les performances de l'estimateur du maximum de vraisemblance des paramètres de la fonction de corrélation $l$ et $\gamma$ en le comparant à un certain nombre de jeux de paramètres arbitrairement fixés.

Nous étudions deux exemples théoriques de complexités différentes et retenons la différence relative entre le risque réel et celui prédit ainsi que la taille de l'intervalle de confiance du risque de défaillance (dont la construction est décrite dans la section 2.1) comme deux quantités pertinentes pour cette étude.

Dans les tableaux qui suivent, la notation "PB" indique que la matrice de corrélation obtenue est numériquement singulière, ce qui interdit tout calcul qui conduirait à des résultats absurdes.

#### 1.7.3.1 Exemple de la forme quadratique

Prenons l'exemple de la forme quadratique décrite plus en détail dans la section 2.4.1.1, de dimension 5, avec $(a_1 = 4, a_2 = 4.2,\ \ a_3 = 4.4, a_4 = 4.6, a_5 = 4.8)$, un seuil de 1.9 et un échantillon de 100 points tirés à l'aide de la séquence de Sobol. Le risque de défaillance réel vaut ici 10.08%.

La méthode du maximum de vraisemblance nous donne, comme estimateur, le couple $(l, \gamma) = (0.954, 1.998)$, autrement dit cette méthode propose un champ dont les variables sont fortement corrélées. L'erreur relative entre le risque réel et le risque prédit est alors de 6.40%, valeur relativement faible qui valide le modèle. Notons, de plus, que l'intervalle de confiance du risque de défaillance, de probabilité 0.9, est de taille 0.021, contient la valeur réelle.

Comparons ces résultats avec ceux issus de modèles aléatoires conditionnés construits avec d'autres valeurs de $l$ et $\gamma$.

En fonction d'un certain nombre de valeurs du couple $(l, \gamma)$, les erreurs relatives moyennes entre le risque réel et les risques prédits sont listées dans le tableau 1.1 et la taille des intervalles de confiance dans le tableau 1.2. On note que tous les intervalles de confiance contiennent, dans cet exemple, le risque réel.

On observe qu'il existe d'autres modèles, comme celui où $l$ vaut 1 et $\gamma$ vaut 1.5, qui conduisent à une très bonne estimation du risque réel (0.73% d'erreur relative). Cependant, l'intervalle de confiance associé à la prédiction est presque six fois plus grand (plus exactement d'une taille de 0.12). L'estimation du risque est incertaine et donc plus difficile à exploiter. Ainsi, selon le critère que l'on retient, la méthode EMV est plus ou moins efficace. Elle conduit néanmoins à un modèle a posteriori pertinent.

| $l\backslash\gamma$ | 0.1 | 0.5 | 1 | 1.5 | 2 |
|---|---|---|---|---|---|
| 0.001 | 2.39% | 6.00% | 6.00% | 6.00% | 6.00% |
| 0.01 | 4.36% | 5.12% | 6.00% | 6.00% | 6.00% |
| 0.1 | 6.18% | 15.8% | 1.11% | 6.02% | 6.32% |
| 1 | 7.34% | 18.3% | 8.59% | 0.73% | 6.15% |
| 10 | 8.38% | 16.8% | 8.18% | 3.59% | 10.8% |
| 100 | 98.5% | 16.0% | 8.00% | 3.32% | PB |
| 1000 | 9.64% | 15.8% | 7.74% | 3.22% | PB |

TABLE 1.1 – Exemple de la forme quadratique : différence relative entre le risque réel et celui prédit.

| $l\backslash\gamma$ | 0.1 | 0.5 | 1 | 1.5 | 2 |
|---|---|---|---|---|---|
| 0.001 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| 0.01 | 0.989 | 1.000 | 1.000 | 1.000 | 1.000 |
| 0.1 | 0.956 | 0.525 | 0.982 | 0.996 | 0.997 |
| 1 | 0.908 | 0.725 | 0.183 | 0.120 | 0.020 |
| 10 | 0.864 | 0.274 | 0.172 | 0.129 | 0.027 |
| 100 | 0.824 | 0.250 | 0.173 | 0.130 | PB |
| 1000 | 0.794 | 0.243 | 0.171 | 0.133 | PB |

TABLE 1.2 – Exemple de la forme quadratique : taille de l'intervalle de confiance du risque de défaillance.

#### 1.7.3.2 Exemple sinus

Considérons maintenant l'exemple du sinus, décrit en détail dans la section 2.4.2.1, de dimension 3, avec $(a_1 = 1, a_2 = -1, a_3 = 2)$, un seuil de 0.9995 et un échantillon de 100 points une fois de plus obtenus à l'aide de la séquence de Sobol. Le risque réel de défaillance vaut 1.01%.

La méthode du maximum de vraisemblance nous donne ici le couple $(l, \gamma) = \big(9.19 \times 10^6, 7.20 \times 10^{-1}\big)$. L'erreur relative entre le risque réel et le risque prédit est très élevé, de l'ordre 832% et la taille de l'intervalle de confiance du risque de défaillance prédit est de 0.540, indiquant que le modèle n'est pas suffisamment renseigné pour estimer correctement le risque de défaillance.

En fonction des valeurs du couple $(l, \gamma)$, les erreurs relatives entre le risque réel et les risques prédits sont indiquées dans le tableau 1.3, les tailles des intervalles de confiance dans le tableau 1.4.

Parmi les couples de paramètres testés, aucune combinaison ne donne un modèle franchement meilleur. Par exemple, le couple $(1, 2)$ (en rouge) conduit à une prédiction totalement aberrante. En outre, l'intervalle de confiance associé ne contient pas le risque réel.

Dans cet exemple, plus complexe (la fonction $y$ est oscillante), la méthode du maximum

de vraisemblance ne semble, au premier abord, pas adéquate. Il n'existe, cependant, pas de configuration de paramètres satisfaisante ; quelque soit la paramétrisation de la fonction de corrélation, le champ conditionné est insuffisant pour représenter la fonction inconnue $y$. Il est ici nécessaire d'augmenter la taille de l'échantillon pour améliorer le modèle aléatoire conditionné ou modifier celui-ci en prenant une fonction de corrélation plus adaptée (périodique par exemple).

| $l\backslash\gamma$ | 0.1 | 0.5 | 1 | 1.5 | 2 |
|---|---|---|---|---|---|
| 0.001 | 708% | 699% | 699% | 699% | 699% |
| 0.01 | 711% | 694% | 699% | 699% | 699% |
| 0.1 | 713% | 734% | 709% | 671% | 676% |
| 1 | 714% | 775% | 870% | 1010% | 2963% |
| 10 | 718% | 788% | 884% | 968% | PB |
| 100 | 719% | 791% | 876% | 970% | PB |
| 1000 | 720% | 765% | 883% | 961% | PB |

TABLE 1.3 – Exemple du sinus : différence relative entre le risque réel et celui prédit.

| $l\backslash\gamma$ | 0.1 | 0.5 | 1 | 1.5 | 2 |
|---|---|---|---|---|---|
| 0.001 | 0.947 | 1.000 | 1.000 | 1.000 | 1.000 |
| 0.01 | 0.913 | 0.980 | 0.999 | 0.999 | 0.999 |
| 0.1 | 0.888 | 0.666 | 0.617 | 0.626 | 0.676 |
| 1 | 0.870 | 0.598 | 0.511 | 0.480 | 0.339 |
| 10 | 0.856 | 0.584 | 0.516 | 0.480 | PB |
| 100 | 0.845 | 0.580 | 0.508 | 0.480 | PB |
| 1000 | 0.835 | 0.580 | 0.511 | 0.480 | PB |

TABLE 1.4 – Exemple du sinus : taille de l'intervalle de confiance du risque de défaillance.

### 1.7.4 Comparaison du champ de Student et du champ mixte

Nous allons comparer le modèle de Student donné dans le théorème 4 et le modèle mixte décrit par le théorème 7, construit avec des monômes, sur un exemple théorique polynomial défini comme suit.

Supposons que la fonction $y$ est un polynôme de degré $d$ en $D$ variables qui possède alors $\binom{D+d}{d}$ coefficients. Les $n$ points d'observations $\boldsymbol{x} := (x_1, x_2, \ldots, x_n)$ sont les éléments d'un plan d'expériences $D$-optimal. De cette manière, les coefficients du modèle mixte sont déterminés plus précisément (variance des posteriors plus faible). Pour plus de détails, voir Atkinson et Donev (1992). Les conditions sont ainsi réunies pour que le modèle mixte donne de bons résultats : $y$ est polynomial et les $n$ données sont collectées à l'aide d'un plan $D$-optimal.

Les deux modèles aléatoires sont ensuite conditionnés par les $n$ données observées. Pour

chaque modèle, nous effectuons un tirage Monte-Carlo sur l'espace des facteurs $X$ : pour chaque tirage $x_m$, nous effectuons une prédiction par le modèle de $y_m$ et calculons l'erreur commise sur cette prédiction. On obtient ainsi une erreur moyenne de prédiction pour le modèle de Student et pour le modèle mixte, ce qui permet de les comparer.

Plus précisément, nous avons étudié 30 polynômes de degré $d = 4$ en $D = 4$ variables (soit $70 = \binom{8}{4}$ coefficients) et retenu le nombre de fois où le modèle de Student est meilleur que le modèle mixte, en fonction du nombre $n$ de données. Nous avons représenté le taux de réussite du modèle de Student, c'est-à-dire la proportion des essais pour lesquels le modèle de Student est plus performant, en fonction du nombre $n$ de données dans la figure 1.3.

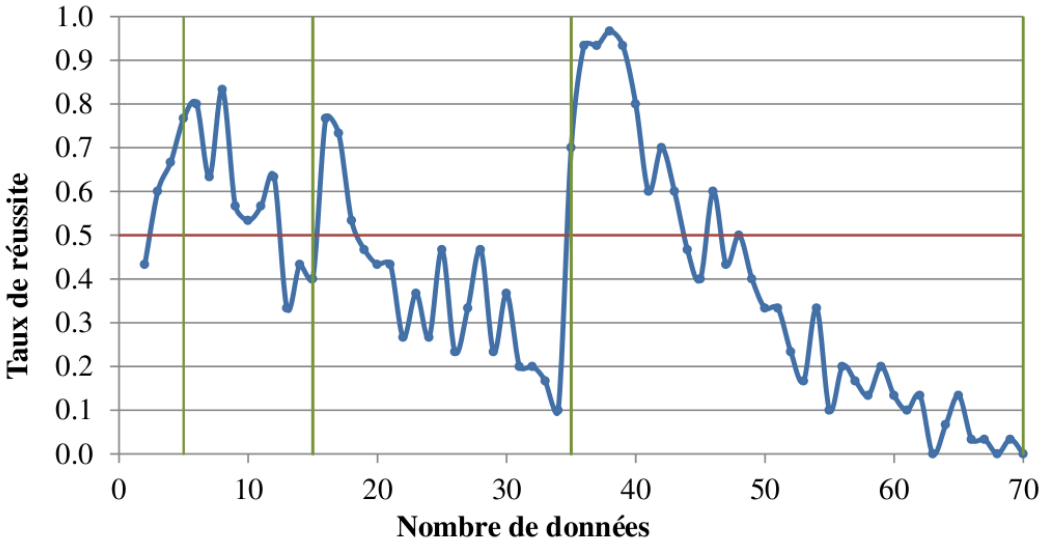



FIGURE 1.3 – Performance du modèle de Student vis-à-vis du modèle mixte en fonction du nombre de données.

Sans surprise, nous observons, pour 70 données, que le modèle mixte permet une prédiction parfaite (sans aucune erreur) ; le modèle a posteriori est effectivement égal à la fonction $y$.

Nous remarquons ensuite des "sauts" du taux de réussite, à chaque fois que le degré du modèle mixte augmente, ainsi que de petites oscillations. Les lignes verticales vertes indiquent le nombre de données au-delà duquel l'ordre supérieur est accessible : 5, 15, 35 et 70 données permettent respectivement l'utilisation d'un modèle polynomial de degré 1, 2, 3 et 4. Lorsque le nombre de données est légèrement supérieur à ces seuils, le modèle de Student est meilleur que le modèle mixte. Ainsi, bien que la complexité du modèle mixte augmente, les performances observées se dégradent, ce qui est déroutant au premier abord. Notons que la structure même du modèle n'est pas en cause puisque lorsque le nombre de données augmente (à structure constante), le modèle mixte s'améliore pour systématiquement finir par donner de meilleurs résultats que le modèle de Student. De plus, fait remarquable, il semble que la modélisation des résidus par le mélange de champs gaussiens joue un rôle mineur.

Nous sommes ici dans une situation d'"overfitting", c'est-à-dire que le modèle mixte est trop complexe pour le nombre de données disponibles. Un modèle plus simple, comme celui de Student ne tombe pas dans cet écueil (comme le notent Jones et al. (1998)). Ce phénomène d'"overfitting" s'estompe quand le nombre $n$ de données croît et réapparait à chaque augmentation de la complexité du modèle mixte (augmentation de degré). Ainsi, cet exemple, construit initialement pour mettre en valeur le modèle mixte, pointe du doigt un défaut de ce modèle : le risque d'"overfitting". Ce risque est plus faible avec un modèle de Student, car ce dernier est plus simple.

## 1.8 Compléments

Nous rassemblons ici trois contributions indépendantes et partielles prolongeant l'étude qui précède dans le chapitre. Ces contributions portent sur la gestion de bruit sur la réponse pour le traitement de données expérimentales ; sur la prise en compte d'une transformation sur la réponse dans le but d'améliorer le modèle ; et enfin sur le traitement d'une réponse multidimensionnelle.

### 1.8.1 Gestion du bruit

Lorsque les moyens de simulations n'existent pas, nous avons à traiter des données expérimentales qui par essence peuvent être bruitées. Il faut donc prendre en compte ce bruit dans notre démarche.

Nous appelons bruit une incertitude aléatoire portant sur les données $y_1, y_2, \ldots, y_n$. Dans ce cas, le vecteur $\boldsymbol{y} := (y_1, y_2, \ldots, y_n)$ est une variable aléatoire à valeur dans $\mathbb{R}^n$. Il est possible alors d'expliciter la probabilité d'accident de deux manières.

Dans une première approche, nous construisons le champ aléatoire $(Y_x)$ à partir du champ aléatoire $(Y_x \mid \boldsymbol{y})$. Nous allons observer que si le bruit suit une distribution gaussienne et que le champ aléatoire $(Y_x \mid \boldsymbol{y})$ est également gaussien, alors le champ aléatoire $(Y_x)$ est explicite et gaussien également.

Une deuxième approche consiste à calculer les probabilités d'accident à partir des probabilités d'accident conditionnées aux bruits. Cette démarche peut être suffisante pour nous, car, dans la suite de notre travail, nous n'utiliserons que ces probabilités d'accident associées à chaque point $x$ de l'espace des facteurs $X$.

Il existe d'autres façons d'insérer le bruit dans un modèle aléatoire, où il est souvent supposé gaussien, voir Rasmussen et Williams (2006) et Goldberg et al. (1997).

#### 1.8.1.1 Sur le champ aléatoire

Dans un premier temps, rappelons la définition de la variance conditionnelle.

**Définition** (Variance conditionnelle)**.** Soient $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, $\mathcal{B}$ une sous-tribu de $\mathcal{T}$, et $X$ une variable aléatoire de carré intégrable sur cet espace.
La *variance conditionnelle* de $X$ sachant $\mathcal{B}$ est définie comme la variable aléatoire

$$\operatorname{var}(X \mid \mathcal{B}) = \mathbb{E}\left[X^2 \mid \mathcal{B}\right] - \mathbb{E}\left[\left(\mathbb{E}\left(X \mid \mathcal{B}\right)\right)^2\right].$$

La *formule de la variance conditionnelle* est

$$\operatorname{var}(X) = \mathbb{E}\left[\operatorname{var}(X \mid \mathcal{B})\right] + \operatorname{var}\left(\mathbb{E}[X \mid \mathcal{B}]\right).$$

Ensuite, rappelons quelques formules de calcul pour les transformés linéaires et quadratiques de vecteurs aléatoires.

**Proposition 2** (Variance de l'image d'un vecteur par une forme linéaire)**.** *Soient* $(\Omega, \mathcal{T}, \mathbb{P})$ *un espace probabilisé et* $X = (X_1, X_2, \ldots, X_n)$ *un vecteur aléatoire de carré intégrable sur cet espace. Soit* $L$ *un vecteur (déterministe) de dimension* $n$*. On a :*

$$\operatorname{var}\left(LX^T\right) = L \operatorname{cov}(X) L^T,$$

*où* $\operatorname{cov}(X) = \mathbb{E}\left[X^T X\right] - \mathbb{E}[X]^T \mathbb{E}[X]$ *est la matrice des covariances du vecteur* $X$*.*

**Proposition 3** (Moyenne de l'image d'un vecteur par une forme quadratique)**.** *Soient* $(\Omega, \mathcal{T}, \mathbb{P})$ *un espace probabilisé et* $X = (X_1, X_2, \ldots, X_n)$ *un vecteur aléatoire de carré intégrable sur cet espace. Soit* $S$ *une matrice carrée de dimension* $n$*. On a :*

$$\mathbb{E}\left(XSX^T\right) = \operatorname{tr}\left(S\operatorname{cov}(X)^T\right) + \mathbb{E}(X) S \mathbb{E}(X)^T$$

*où* $\operatorname{cov}(X) = \mathbb{E}\left[X^T X\right] - \mathbb{E}[X]^T \mathbb{E}[X]$ *est la matrice des covariances du vecteur* $X$*.*

**Application au cas gaussien** On se place sous les hypothèses du théorème 1. Notons $\boldsymbol{E}(\boldsymbol{y})$ et $\mathbf{cov}(\boldsymbol{y})$, respectivement la moyenne et la covariance de $\boldsymbol{y}$. Dans le cas où les moyennes et les covariances sont fixées, a priori, indépendamment de $\boldsymbol{y}$, la distribution de probabilité conditionnelle sachant que $Y_{x_i} = y_i, 1 \leq i \leq n$, a pour espérance conditionnelle :

$$\mathbb{E}\left(Y_x \mid \boldsymbol{y}\right) = \mu(x) + \boldsymbol{\rho}(\boldsymbol{x}, (\boldsymbol{x_i}))\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y} - \boldsymbol{\mu}((\boldsymbol{x_i}))\right)^T$$

et pour variance conditionnelle, indépendante de $y$,

$$\operatorname{var}\left(Y_x \mid \boldsymbol{y}\right) = \rho(x, x) - \boldsymbol{\rho}(\boldsymbol{x}, (\boldsymbol{x_i}))\boldsymbol{\Sigma}^{-1}\boldsymbol{\rho}(\boldsymbol{x}, (\boldsymbol{x_i}))^T.$$

Calculons dans un premier temps $\mathbb{E}\left(Y_x\right)$. Nous avons

$$\mathbb{E}\left(Y_x\right) = \mathbb{E}\left(\mathbb{E}\left(Y_x \mid \boldsymbol{y}\right)\right) = \mu(x) + \boldsymbol{\rho}(\boldsymbol{x}, (\boldsymbol{x_i}))\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{E}(\boldsymbol{y}) - \boldsymbol{\mu}((\boldsymbol{x_i}))\right)^T.$$

Calculons maintenant la quantité $\operatorname{var}\left(Y_x\right)$. Pour appliquer la formule de la variance conditionnelle (en considérant la sous-tribu $\mathcal{B}$ comme la tribu engendrée par l'aléa sur la réponse $\boldsymbol{y}$, c'est-à-dire le bruit), il nous faut calculer la variance de $\mathbb{E}\left(Y_x \mid \boldsymbol{y}\right)$. D'après la formule précédente on a

$$\operatorname{var}\left(\mathbb{E}\left(Y_x \mid \boldsymbol{y}\right)\right) = \operatorname{var}\left(\boldsymbol{\rho}(\boldsymbol{x}, (\boldsymbol{x_i}))\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T\right).$$

D'après la proposition 2, on a alors

$$\operatorname{var}\left(\mathbb{E}\left(Y_x \mid \boldsymbol{y}\right)\right) = \boldsymbol{\rho}(\boldsymbol{x}, (\boldsymbol{x_i}))\boldsymbol{\Sigma}^{-1}\mathbf{cov}(\boldsymbol{y})\boldsymbol{\Sigma}^{-1}\boldsymbol{\rho}(\boldsymbol{x}, (\boldsymbol{x_i}))^T.$$

On a également, ici,

$$\mathbb{E}\left(\operatorname{var}\left(Y_x \mid \boldsymbol{y}\right)\right) = \operatorname{var}\left(Y_x \mid \boldsymbol{y}\right).$$

En appliquant la formule de la variance conditionnelle, on obtient finalement

$$\operatorname{var}\left(Y_x\right) = \operatorname{var}(Y_x \mid \boldsymbol{y}) + \boldsymbol{\rho}(\boldsymbol{x}, (\boldsymbol{x_i}))\boldsymbol{\Sigma}^{-1}\mathbf{cov}(\boldsymbol{y})\boldsymbol{\Sigma}^{-1}\boldsymbol{\rho}(\boldsymbol{x}, (\boldsymbol{x_i}))^T.$$

**Proposition 4.** *Si les données* $\boldsymbol{y} := (y_1, y_2, \ldots, y_n)$ *sont aléatoires et suivent une loi de Gauss,* $\mathcal{N}(\boldsymbol{E}(\boldsymbol{y}), \mathbf{cov}(\boldsymbol{y}))$ *et si on choisit a priori un champ aléatoire gaussien suivant les données du théorème 1, alors le champ a posteriori est gaussien. En particulier les lois unidimensionnelles sont gaussiennes de moyenne*

$$\mathbb{E}(Y_x) = \mu(x) + \boldsymbol{\rho}(\boldsymbol{x}, (\boldsymbol{x_i}))\boldsymbol{\Sigma}^{-1}(\boldsymbol{E}(\boldsymbol{y}) - \boldsymbol{\mu}((\boldsymbol{x_i})))^T$$

*et de variance*

$$\mathrm{var}(Y_x) = \rho(x,x) - \boldsymbol{\rho}(\boldsymbol{x}, (\boldsymbol{x_i}))\boldsymbol{\Sigma}^{-1}\boldsymbol{\rho}(\boldsymbol{x}, (\boldsymbol{x_i}))^T + \boldsymbol{\rho}(\boldsymbol{x}, (\boldsymbol{x_i}))\boldsymbol{\Sigma}^{-1}\mathbf{cov}(\boldsymbol{y})\boldsymbol{\Sigma}^{-1}\boldsymbol{\rho}(\boldsymbol{x}, (\boldsymbol{x_i}))^T.$$

**Application au cas de Student** On se place sous les hypothèses du théorème 4, dans le cas limite où $m \to \infty$ et $\epsilon \to 0$. Notons $\boldsymbol{E}(\boldsymbol{y})$ et $\mathbf{cov}(\boldsymbol{y})$, respectivement la moyenne et la covariance de $\boldsymbol{y}$. La distribution de probabilité conditionnelle sachant que $Y_{x_i} = y_i, 1 \leq i \leq n$ a pour espérance conditionnelle :

$$\mathbb{E}(Y_x \mid \boldsymbol{y}) = \mu + \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \mu\boldsymbol{1})^T \quad \text{avec} \quad \mu = \frac{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}$$

et pour variance conditionnelle :

$$\mathrm{var}(Y_x \mid \boldsymbol{y}) = \frac{1}{n-4}\left((\boldsymbol{y} - \mu\boldsymbol{1})\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T\right)\left(1 - \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{x})^T + \frac{\left(1 - \boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{x})^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right).$$

On peut exprimer ces deux quantités comme suit :

$$\mathbb{E}(Y_x \mid \boldsymbol{y}) = \boldsymbol{A}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T \quad \text{avec} \quad \boldsymbol{A}(\boldsymbol{x}) := \frac{\left(1 - \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)\boldsymbol{1}}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T} + \boldsymbol{k}(\boldsymbol{x})$$

et

$$\mathrm{var}(Y_x \mid \boldsymbol{y}) = \boldsymbol{y}\boldsymbol{B}(\boldsymbol{x})\boldsymbol{y}^T$$

avec

$$\boldsymbol{B}(\boldsymbol{x}) := \frac{1}{n-4}\left(1 - \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{x})^T + \frac{\left(1 - \boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{x})^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)\left(\boldsymbol{\Sigma}^{-1} - \frac{\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)\left(\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\right)}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right).$$

Dans un premier temps, calculons :

$$\mathbb{E}(Y_x) = \mathbb{E}(\mathbb{E}(Y_x \mid y)) = \boldsymbol{A}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{E}(\boldsymbol{y})^T$$

Puis, d'après la proposition 2, on a :

$$\mathrm{var}(\mathbb{E}(Y_x \mid \boldsymbol{y})) = \mathrm{var}\left(\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^T\right) = \boldsymbol{A}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\mathbf{cov}(\boldsymbol{y})\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\boldsymbol{x})^T.$$

D'après la proposition 3, on a :

$$\mathbb{E}(\mathrm{var}(Y_x \mid \boldsymbol{y})) = \mathbb{E}\left(\boldsymbol{y}\boldsymbol{B}(\boldsymbol{x})\boldsymbol{y}^T\right) = \mathrm{tr}\left(\boldsymbol{B}(\boldsymbol{x})\mathbf{cov}(\boldsymbol{y})^T\right) + \boldsymbol{E}(\boldsymbol{y})\boldsymbol{B}(\boldsymbol{x})\boldsymbol{E}(\boldsymbol{y})^T.$$

On obtient au final :

$$\mathrm{var}(Y_x) = \boldsymbol{A}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\mathbf{cov}(\boldsymbol{y})\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\boldsymbol{x})^T + \mathrm{tr}\left(\boldsymbol{B}(\boldsymbol{x})\mathbf{cov}(\boldsymbol{y})^T\right) + \boldsymbol{E}(\boldsymbol{y})\boldsymbol{B}(\boldsymbol{x})\boldsymbol{E}(\boldsymbol{y})^T.$$

Ces formules ne nous donnent pas accès à la loi du champ $(Y_x)$. Une étude reste à mener pour connaître la nature de cette loi et préciser les paramètres qui la caractérisent.

#### 1.8.1.2 Sur la probabilité d'accident

Sans hypothèse particulière sur le champ aléatoire, et en notant $Q$ la loi de la variable aléatoire $\boldsymbol{y}$, la formule

$$\mathbb{P}\left(Y_x \in A\right) = \int_{\mathbb{R}^n} \mathbb{P}\left(Y_x \in A \mid \boldsymbol{y}\right) \, \mathrm{d}Q(y).$$

est toujours exploitable numériquement.

### 1.8.2 Transformation de la réponse

Lorsque l'échantillon composé de $n$ données comporte quelques points dont la réponse est hors-norme par rapport à la moyenne des $n$ données, il peut être plus approprié de modéliser la réponse après lui avoir appliqué une transformation afin d'avoir à traiter un échantillon avec une variance raisonnable, voir Snelson et al. (2003).

Pour appliquer une transformation monotone aux modèles précédemment présentés, la méthode est la suivante : Si $Y$ est une variable aléatoire réelle et si $f$ est une fonction dérivable strictement monotone sur $\mathbb{R}$ telle que la variable aléatoire $f\left(Y\right)$ suit la loi de densité $g$, alors $Y$ suit la loi de densité $h := |f'| \cdot gof$.

Cependant, comme les quantités d'intérêt pour l'étude du risque de défaillance sont les probabilités d'accident $\mathbb{P}\left(Y_x \in A\right)$ qui sont égales à $\mathbb{P}\left(f\left(Y_x\right) \in f\left(A\right)\right)$ (par un simple changement de variable), il suffit de composer la fonction inconnue $y$ avec la transformation $f$ ainsi que de considérer comme espace de défaillance $f\left(A\right)$ au lieu $A$. Le formalisme reste alors le même et la construction du champ aléatoire se déroule de la façon décrite dans les précédentes sections de ce chapitre.

La principale difficulté est ici de sélectionner la fonction de transformation adéquate. Nous avons à notre disposition la méthode "Leave-one-out cross validation", décrite dans la section 1.6. Cependant, il n'est pas facile de trouver un critère de sélection approprié. En effet, habituellement, les critères des méthodes par validation croisée impliquent les moments des distributions étudiées. Dans notre cas, les moments de $h$ peuvent s'avérer difficiles à expliciter. Il nous faut ainsi trouver un critère basé sur les moments de $g$ et ce pour toute transformation $f$. La problématique se résume alors à essayer de comparer la distribution $g$ de la variable aléatoire $f\left(Y\right)$ à la distribution $g'$ de la variable aléatoire $f'\left(Y\right)$. Or, ces distributions peuvent avoir des supports différents, ce qui rend toute comparaison difficile voir impossible.

Pour répondre à cette problématique, il est souvent utilisé une famille paramétrique de transformations proposée par Box et Cox (1964) de la forme

$$y \mapsto \frac{y^\lambda - 1}{\lambda \dot{y}^{\lambda-1}} \quad (\text{pour } \lambda \neq 0) \quad \text{ou} \quad y \mapsto \dot{y} \log\left(y\right) \quad (\text{pour } \lambda = 0)$$

avec $\dot{y}$ la moyenne géométrique de l'échantillon $(y_1, y_2, \ldots, y_n)$. Les transformations de cette famille ont la propriété d'avoir leur erreur quadratique comparable.

Mais, pour pouvoir comparer n'importe quels types de transformation, il nous faut trouver un critère normalisé. Nous proposons la quantité suivante, nommée "normalized

root mean squared" :

$$NRMS_f = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(f\left(y_i\right) - \mathbb{E}\left(f\left(Y_{x_i}\right) \mid \left(Y_{x_j} = y_j\right)_{1\leq j\leq n, j\neq i}\right)\right)^2}}{\max_{1\leq i\leq n}\left(f\left(y_i\right)\right) - \min_{1\leq i\leq n}\left(f\left(y_i\right)\right)}.$$

Ainsi les quantités $NRMS_f$ et $NRMS_{f'}$ sont comparables pour toutes fonctions $f$ et $f'$. Un exemple d'utilisation de ce critère est décrit dans la section 3.2 du troisième chapitre de cette thèse.

### 1.8.3 Réponse multidimensionnelle

Dans la plupart des situations, la spécification imposée au produit industriel ne porte pas sur une valeur numérique $y$, mais sur une famille de valeurs $\left(y^{(1)}, y^{(2)}, \ldots, y^{(d)}\right)$ toutes déterminées par le jeu de facteurs $x \in X$.

Le jeu de facteurs $x$ conduit à une pièce défectueuse si

$$y = \left(y^{(1)}, y^{(2)}, \ldots, y^{(d)}\right) \in A$$

où $A$ est l'espace hors-spécification, sous-ensemble de $\mathbb{R}^d$.

D'un point de vue formel, la méthode développée dans les sections précédentes peut continuer à s'appliquer avec la construction d'un champ aléatoire multidimensionnel

$$\left(Y_x\right)_{x\in X} = \left(Y_x^{(1)}, Y_x^{(2)}, \ldots, Y_x^{(d)}\right)_{x\in X},$$

où chaque $\left(Y_x^{(i)}\right)_{x\in X}$ est un champ aléatoire réel.

Du point de vue pratique, cela introduit une complexité considérable puisque les corrélations entre les réponses $y^{(i)}$ doivent être déterminées pour la définition des modèles a priori et a posteriori.

Dans le cas particulier où les spécifications portent sur chaque réponse $y^{(1)}, y^{(2)}, \ldots, y^{(d)}$ individuellement, c'est-à-dire quand

$$y \in A \quad \text{si et seulement si} \quad \left(y^{(1)} \in A_1 \text{ ou } y^{(2)} \in A_2 \text{ ou } \ldots \text{ ou } y^{(d)} \in A_d\right),$$

on se contentera souvent de l'inégalité

$$\mathbb{P}\left(Y_x \in A\right) \leq \min\left(1, \sum_{i=1}^{d} \mathbb{P}\left(Y_x^{(i)} \in A_i\right)\right). \tag{1.5}$$

On note que si les $d$ modèles $Y_x^{(i)}$ sont bien renseignés, les $\mathbb{P}\left(Y_x^{(i)} \in A_i\right)$ sont proches de 0 ou 1. Dans ce cas, la majoration précédente, proche d'une égalité, n'induit pas d'erreur importante.

| Prior | Distributions conditionnelles | Méthode EMV | Paramètre de position | Carré du paramètre d'échelle |
|---|---|---|---|---|
| $Y_x = W_x$ | de $Y_x$,<br>sous les hypothèses<br>du corollaire 1<br>(distribution gaussienne) | Estimateurs :<br>$\mu = \frac{\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}$ et $\sigma^2 = \frac{s}{n}$<br>$\boldsymbol{\theta} \rightarrow_{\min} n\ln(s) + \ln(\lvert\boldsymbol{\Sigma}\rvert)$<br>en posant<br>$s := (\boldsymbol{y} - \mu\boldsymbol{1})\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{y} - \mu\boldsymbol{1})^T$ | $\mu + \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \mu\boldsymbol{1})^T$ | $\sigma^2\left(1 - \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{x})^T\right)$ |
| $Y_x = U + W_x$ | de $Y_x$ quand $m \rightarrow \infty$,<br>sous les hypothèses<br>du théorème 2<br>(distribution gaussienne) | Estimateurs, pour $n > 1$ :<br>$\sigma^2 = \frac{s}{n-1}$<br>$\boldsymbol{\theta} \rightarrow_{\min} (n-1)\ln(s)$<br>$+ \ln(\lvert\boldsymbol{\Sigma}\rvert) + \ln\left(\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)$<br>en posant<br>$s := (\boldsymbol{y} - \mu\boldsymbol{1})\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{y} - \mu\boldsymbol{1})^T$<br>et $\mu := \frac{\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}$ | $\mu + \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \mu\boldsymbol{1})^T$ | $\sigma^2\left(1 - \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{x})^T + \frac{\left(1 - \boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{x})^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)$ |
| | de $U$ quand $m \rightarrow \infty$,<br>sous les hypothèses<br>du théorème 3<br>(distribution gaussienne) | | $\mu$ | $\frac{\sigma^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}$ |
| $Y_x = U + VW_x$ | de $Y_x$ quand $m \rightarrow \infty$ et $\epsilon \rightarrow 0$,<br>pour $n > 2$ et sous les hypothèses<br>du théorème 4<br>(distribution de Student) | Estimateurs :<br>$\boldsymbol{\theta} \rightarrow_{\min} (n-2)\ln(s)$<br>$+ \ln(\lvert\boldsymbol{\Sigma}\rvert) + \ln\left(\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T\right)$<br>en posant<br>$s := (\boldsymbol{y} - \mu\boldsymbol{1})\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{y} - \mu\boldsymbol{1})^T$<br>et $\mu := \frac{\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}$ | $\mu + \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \mu\boldsymbol{1})^T$ | $\sigma^2\left(1 - \boldsymbol{k}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{x})^T + \frac{\left(1 - \boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{k}(\boldsymbol{x})^T\right)^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}\right)$<br>où $\sigma^2 := \frac{s}{n-2}$ |
| | de $U$ quand $m \rightarrow \infty$ et $\epsilon \rightarrow 0$,<br>pour $n > 2$ et sous les hypothèses<br>du théorème 5<br>(distribution de Student) | | $\mu$ | $\frac{\sigma^2}{\boldsymbol{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{1}^T}$ où $\sigma^2 := \frac{s}{n-2}$ |
| | de $V^2$ quand $m \rightarrow \infty$ et $\epsilon \rightarrow 0$,<br>pour $n > 2$ et sous les hypothèses<br>du théorème 6<br>(distribution inverse du $\chi^2$ réduite) | | | $\frac{s}{n-2}$ |

TABLE 1.5 – Récapitulatif des distributions conditionnelles univariées décrites dans ce chapitre, en supposant que la méthode du maximum de vraisemblance ait été utilisée pour régler les paramètres des modèles.

# Chapitre 2

# Exploitation d'un modèle aléatoire

Dans ce chapitre nous proposons des outils nouveaux d'aide à la décision et de validation de modèle. Ces outils sont des indicateurs aléatoires, autrement dit des distributions de probabilité, déduits de la donnée en chaque point de l'espace des facteurs d'une probabilité de défaillance.

Comme nous l'expliquons dans la section 2.1, notre approche est basée sur l'interprétation d'une probabilité comme un degré d'appartenance, nous rapprochant ainsi du formalisme des ensembles flous (Zadeh (1965)).

Supposons que nous disposons d'un champ aléatoire $(Y_x)_{x \in X}$ conditionné par les $n$ observations à notre disposition. Il représente l'information aussi bien que l'incertitude au regard des valeurs observées de la fonction réponse $x \mapsto y(x)$.

Ce champ permet de calculer, pour chaque point $x$ de l'espace des facteurs $X$, la *probabilité de défaillance*

$$\mathbb{P}(Y_x \in A)$$

où nous rappelons que $A$ désigne l'espace hors-spécification (le produit, caractérisé par le jeu de facteurs $x$, ne satisfait pas les spécifications si et seulement si $y(x) \in A$).

Cette probabilité de défaillance, ainsi que la mesure de probabilité $P$ assignée à l'espace des facteurs $X$ vont constituer la base des raisonnements amenant les définitions du risque de défaillance et du risque d'insuffisance décrits dans les sections 2.1 et 2.2 de ce chapitre.

Le risque de défaillance fournit au décideur une indication qui peut lui permettre de statuer quant à la robustesse du produit, relative au processus de fabrication.

Un intervalle de confiance permet de quantifier l'incertitude sur cet indicateur et le risque d'insuffisance du modèle informe quant à la capacité prédictive de celui-ci.

Dans la situation où cet intervalle de confiance est jugé trop grand pour que le risque de défaillance puisse être un estimateur fiable ou dans le cas où le risque d'insuffisance est jugé trop élevé, le modèle aléatoire est considéré comme insuffisamment informé. Pour améliorer la prédiction du risque de défaillance, il est alors nécessaire de rajouter de l'information au modèle, c'est-à-dire de fournir de nouvelles données. Dans ce but, des méthodes d'optimisation sont décrites dans la section 2.3.

Pour valider notre approche pour le calcul du risque de défaillance, nous proposons

trois exemples théoriques, de différents niveaux de complexité, décrits dans la section 2.4. Elle peut alors être comparée avec une méthode statistique classique de calcul de risque.

Dans la section 2.5, une méthode pour calculer l'influence des facteurs sur le risque de défaillance est présentée. Cela permet d'indiquer au concepteur du produit quel facteur nécessite le plus d'attention lors de la mise en production.

## 2.1 Probabilité du risque de défaillance

### 2.1.1 La probabilité de défaillance vue comme une variable aléatoire

La probabilité du risque de défaillance, ou *risque de défaillance* (par abrévation), est la mesure du *domaine de défaillance* $\{x \in X \mid y(x) \in A\}$ où $X \subseteq \mathbb{R}^D$ est l'espace des facteurs muni de la distribution de probabilité $P$. Dans beaucoup de cas pratiques, la fonction $x \mapsto y(x)$ n'est pas disponible ; elle peut seulement être échantillonnée en quelques points $x$. Nous avons vu dans le premier chapitre de cette thèse comment construire un champ aléatoire $(Y_x)_{x \in X}$, défini sur un espace probabiliste $(\Omega, \mathbb{P})$, à partir d'un jeu d'observations dans le but d'approcher la fonction $y$.

Cette étape de modélisation n'est pas sans conséquence sur l'approche à retenir pour calculer le risque de défaillance. Puisque la connaissance de la fonction $y$ est partielle, il semble raisonnable de munir la prédiction du risque de défaillance d'un indice de confiance. Ainsi, un décideur peut aisément évaluer la qualité du modèle et statuer sur la capacité du processus industriel à réaliser le produit étudié. Ce problème se résume à la question suivante : Comment propager l'incertitude inhérente à chaque prédiction, donnée par un modèle aléatoire, à l'estimateur du risque de défaillance ?

Si le phénomène sous-jacent était un champ aléatoire réel $(Y_x)_{x \in X}$, nous pourrions considérer la variable aléatoire

$$\omega \mapsto P(\{x \in X \mid Y_x(\omega) \in A\}) \quad \text{sur} \quad (\Omega, \mathbb{P}) \tag{2.1}$$

comme le *risque de défaillance*. Expliquons pourquoi ce n'est pas une bonne approche dans le cas où le phénomène étudié est déterministe.

À strictement parler, le modèle $(Y_x)_{x \in X}$ ne peut pas décrire la réalité, qui est parfaitement déterminée (mais inconnue). Il représente seulement notre connaissance aux points d'observations. D'ailleurs, nous avons à donner une interprétation à l'aléa du processus. Si nous retenons la conception subjective de la théorie des probabilités, l'aléa provient de la connaissance incomplète de la quantité $y(x)$. Dans ce cadre, la mesure $\mathbb{P}$ quantifie le degré de confiance sur les prédictions du modèle. L'aléa devient un moyen d'évaluer la qualité du modèle, c'est-à-dire la capacité prédictive du modèle. En ce sens, l'expression $\mathbb{P}(Y_x \in A) = 0.5$ ne signifie pas que le point $x$ a une chance sur deux d'appartenir au domaine de défaillance ; elle signifie, au regard de l'information disponible, qu'un décideur sur deux considérera que $x$ appartient au domaine de défaillance. Le modèle est donc inadéquat au point $x$.

Si la définition (2.1) du risque de défaillance est absolument correcte mathématiquement, elle ne peut être utilisée dans le cas particulier de la modélisation de connaissances incomplètes. Un exemple est nécessaire pour clarifier ce point.

Nous affirmons (voir remarque en fin de section) que la condition $\mathbb{P}(Y_x \in A) = 0.5$ pour tout $x \in X$, qui est le signe d'un mauvais modèle, n'est pas incompatible avec le fait que $P(Y_\cdot(\omega) \in A) = 0.5$ pour tout $\omega \in \Omega$. Cependant, une distribution de Dirac, ou une distribution proche d'une masse de Dirac, traduit une situation pour laquelle l'information est sûre, ou près d'être sûre. Cela ne reflète pas le degré d'insécurité associé à un modèle inadéquat. Par conséquent, nous concluons que la distribution de la variable aléatoire $P(Y_\cdot(\omega) \in A)$ ne fournit pas les informations pertinentes concernant la qualité du modèle.

Le champ aléatoire seul n'est pas une représentation de la connaissance ; il ne peut pas se substituer aux procédures mentales impliquées dans la construction d'une opinion. Une analyse complémentaire est donc nécessaire. Nous n'avons pas à notre disposition une stricte condition d'appartenance au domaine de défaillance. Néanmoins, nous pouvons aisément calculer la fonction d'appartenance $\mathcal{M}$ suivante :

$$\begin{cases} (X, P) \to [0,1] \\ x \mapsto \mathcal{M}(x) := \mathbb{P}(Y_x \in A). \end{cases}$$

Cette fonction associe à chaque point $x \in X$ un nombre réel de l'intervalle $[0,1]$ mesurant le degré d'appartenance de $x$ au domaine de défaillance. Ainsi, le domaine de défaillance, inconnu, peut seulement être caractérisé de façon imprécise par l'entremise de la fonction d'appartenance $\mathcal{M}$. Dû à une connaissance incomplète de la fonction $y$, nous n'avons accès qu'à une version floue du domaine de défaillance. D'après Zadeh (1965), "such a framework provides a natural way of dealing with problems in which the sources of imprecision is the absence of sharply defined criteria of class membership rather than the presence of random variables".

Par définition, la notion d'appartenance n'est pas bien définie pour un ensemble flou. Pour un degré d'appartenance $\mathcal{M}(x)$ donné, un décideur peut considérer que $x$ appartient au domaine de défaillance alors qu'un autre non, dépendant de sa propre tolérance au risque. Afin de prendre en compte la nature subjective de cette décision, nous introduisons un seuil $\alpha$, $0 \leq \alpha \leq 1$, et acceptons de dire que "$x$ appartient au domaine de défaillance" si $\mathcal{M}(x) > \alpha$. Le seuil $\alpha$ détermine le statut à donner aux prédictions incertaines. Les données sont ainsi agrégées en deux classes : "bonnes" et "mauvaises". Le domaine de défaillance au seuil $\alpha$ est alors :

$$\{x \in X \mid \mathcal{M}(x) > \alpha\}.$$

Il s'agit d'un sous-ensemble de $X$ et peut être mesuré par

$$R(\alpha) := P(\{x \in X \mid \mathcal{M}(x) > \alpha\}).$$

C'est l'estimation du risque de défaillance par le décideur pour qui le seuil de tolérance au risque est $\alpha$. Sa distribution est une masse de Dirac au point $R(\alpha)$, que nous noterons $\delta_{R(\alpha)}$. Nous définissons implicitement ici un modèle causal : le risque est conditionné au seuil de tolérance au risque (Pearl (1988)). N'étant pas sûrs du seuil de tolérance au risque du décideur, nous considérons $\alpha$ comme une variable aléatoire à valeurs dans $[0,1]$. Soit $\eta$ sa distribution. **Dans ce cadre, le risque de défaillance est finalement défini comme une variable aléatoire dont la distribution est**

$$\boldsymbol{\mathcal{R} := \int_0^1 \delta_{R(\alpha)}\, \eta(\mathrm{d}\alpha).}$$

Autrement dit, $\mathcal{R}$ est la mesure image de la loi $\eta$ par la fonction $R$. Cette distribution est bien définie dès lors que la distribution $\eta$ du seuil $\alpha$ de tolérance au risque est choisie.

*Remarque.* Justifions l'affirmation page 73 : "la condition $\mathbb{P}(Y_x \in A) = 0.5$ pour tout $x \in X$ n'est pas incompatible avec le fait que $P(Y_\cdot(\omega) \in A) = 0.5$ pour tout $\omega \in \Omega$". Le calcul ci-dessous montre que si les variables aléatoires $Y_x$ sont deux à deux indépendantes et si la mesure $P$ est continue, alors la distribution de $P(Y_\cdot(\omega) \in A)$ est une masse de Dirac. Pour justifier l'affirmation, il suffit donc de considérer un modèle remplissant cette condition d'indépendance et tel que $\mathbb{P}(Y_x \in A) = 0.5$ pour tout $x \in X$.

Voici le calcul. Écrivons

$$\begin{aligned}\mathbb{E}\left[(P(Y_\cdot \in A))^2\right] &= \mathbb{E}\left[\int_{X\times X} \mathbb{1}_A(Y_{x_1})\,\mathbb{1}_A(Y_{x_2})\ \mathrm{d}P(x_1)\ \mathrm{d}P(x_2)\right] \\ &= \int_{X\times X} \mathbb{E}\left[\mathbb{1}_A(Y_{x_1})\,\mathbb{1}_A(Y_{x_2})\right]\ \mathrm{d}P(x_1)\ \mathrm{d}P(x_2).\end{aligned}$$

Puisque $P$ est continue, on a $x_1 \neq x_2$ pour $P \otimes P$ presque tout couple $(x_1, x_2)$. D'où, par indépendance de $Y_{x_1}$ et $Y_{x_2}$,

$$\mathbb{E}\left[(P(Y_\cdot \in A))^2\right] = \int_{X\times X} \mathbb{E}\left[\mathbb{1}_A(Y_{x_1})\right]\mathbb{E}\left[\mathbb{1}_A(Y_{x_2})\right]\ \mathrm{d}P(x_1)\ \mathrm{d}P(x_2) = \left(\mathbb{E}\left[P(Y_\cdot \in A)\right]\right)^2.$$

### 2.1.2 Choix d'une distribution uniforme pour le seuil

Dans cette section, nous discutons du choix de la distribution $\eta$ du seuil $\alpha$ de tolérance au risque.

Un résultat attendu de l'évaluation du risque est l'espérance de la probabilité de défaillance, $E(\mathbb{P}(Y_x \in A))$. C'est un estimateur de la $P$-mesure du domaine de défaillance basé sur la connaissance des valeurs $(y_i = y(x_i))_{1\leq i\leq n}$. Si nous supposons maintenant que le jeu d'observations $(x_i)_{1\leq i\leq n}$ couvre $X$, alors $\mathcal{M}(x)$ est la fonction indicatrice de $\{x \in X \mid y(x) \in A\}$ et par conséquent

$$E(\mathcal{M}) = P(y \in A).$$

Sous des hypothèses raisonnables (régularité de la fonction $y$, par exemple), nous attendons un estimateur convergent si l'espace des facteurs $X$ est échantillonné de façon appropriée. Comme nous le verrons par la suite, le modèle causal, introduit dans la section 2.1.1, peut être interprété comme un opérateur agissant sur l'espace des mesures de probabilité de $[0, 1]$, opérateur liant les distributions du risque de défaillance et de la probabilité de défaillance. Ainsi, il semble judicieux de conserver le premier moment de cette dernière quand la transformation est appliquée. De cette façon, la moyenne du risque de défaillance demeure un estimateur de la mesure du domaine de défaillance. Cette remarque nous conduit à la proposition 5.

Si le modèle $(Y_x)$ n'est pas disponible, alors la probabilité de défaillance est inconnue. Un prior uniforme (non-informatif) apparaît intuitivement comme un choix raisonnable pour représenter l'ignorance, compte tenu du principe d'indifférence de Laplace (voir Kass et Wasserman (1996)). Dans cette situation, la distribution du risque de défaillance devrait

également être uniforme, signifiant que l'état d'incertitude est préservé. Motivé par cette propriété d'invariance, nous affirmons la proposition 6.

Dans un premier temps, introduisons quelques notations et définitions utiles aux prochaines démonstrations. Comme précédemment, pour chaque $\alpha \in [0,1]$, nous définissons

$$R(\alpha) := P(\mathcal{M}(x) > \alpha).$$

La fonction $R$ est le complémentaire de la fonction de répartition (ou fonction de queue) de la variable aléatoire $\mathcal{M}(x)$, définie sur l'espace probabilisé $(X, P)$. Nous pouvons alors voir $R$ comme une variable aléatoire définie sur l'espace probabilisé $([0,1], \eta)$, et nous définissons, pour chaque $t \in [0,1]$ :

$$G(t) := \eta(\{\alpha \in [0,1] \mid R(\alpha) > t\}) = \eta(R > t).$$

Ainsi, la fonction $G$ est la fonction de queue de la variable aléatoire $R$.

De plus, notons $K$ la fonction de répartition de la distribution $\eta$ :

$$K(u) = \eta([0,u]) \quad \text{pour tout} \quad u \in [0,1].$$

Rappelons la définition de l'inverse généralisé de la fonction décroissante $R$.

**Définition** (Inverse généralisé)**.** Soit $S$ une fonction décroissante et continue à droite sur l'intervalle $[0,1]$ avec $S(0) \leq 1$ et $S(1) = 0$. Pour chaque $t \in [0,1]$, on définit

$$S^{-1}(t) = \sup\{\alpha \in [0,1] : S(\alpha) > t\}.$$

Il est bien connu que $S^{-1}$ est elle-même une fonction décroissante et continue à droite sur l'intervalle $[0,1]$ avec $S^{-1}(0) \leq 1$ et $S^{-1}(1) = 0$. De plus, $\left(S^{-1}\right)^{-1} = S$.

En revenant aux notations précédant cette définition, il n'est pas difficile de vérifier que $G = K \circ R^{-1}$.
Considérons maintenant cet opérateur, qui associe $G$ à $R$, dans un cadre plus général. La mesure de probabilité $\eta$ est supposée fixée. Pour chaque mesure de probabilité $m$ sur l'intervalle $[0,1]$, il est possible d'associer une autre mesure de probabilité $m'$ sur $[0,1]$ de la manière suivante :

1. notons $F_m$ la fonction $F_m(\alpha) = m(]\alpha, 1])$,
2. définissons $F_{m'} = K \circ F_m^{-1}$.

Notons $L_\eta$ l'opérateur, agissant sur l'espace des mesures de probabilité sur $[0,1]$, qui associe $m'$ à $m$.

**Proposition 5.**

1. *L'opérateur $L_\eta$ préserve le premier moment de probabilité (i.e. $\int_0^1 t\ m(\mathrm{d}t) = \int_0^1 t\ m'(\mathrm{d}t)$ pour chaque $m$) si et seulement si la probabilité $\eta$ est uniforme sur $[0,1]$.*
2. *Si la probabilité $\eta$ est uniforme, alors l'opérateur $L_\eta$ est son propre inverse, signifiant que $(m')' = m$.*

*Démonstration.* Le premier moment de la mesure de probabilité $m$ associée à $F_m$ est égal à :

$$\int_0^1 F_m(\alpha)\ \mathrm{d}\alpha.$$

L'identité suivante peut être prouvée :

$$\int_0^1 F_m(\alpha)\ \mathrm{d}\alpha = \int_0^1 F_m^{-1}(\alpha)\ \mathrm{d}\alpha.$$

(Pour l'établir, on peut considérer une mesure $m$ discrète et un argument général d'approximation.)
Ainsi, le premier moment de $m$ est égal au premier moment de $m'$ si et seulement si :

$$\int_0^1 K\left(F_m^{-1}(\alpha)\right)\ \mathrm{d}\alpha = \int_0^1 F_m^{-1}(\alpha)\ \mathrm{d}\alpha.$$

La fonction identité est l'unique fonction $K$ qui satisfait cette égalité pour tout choix de $m$. Cela signifie que $\eta$ doit être une distribution uniforme. Dans ce cas, nous avons $F_{(m')'} = \mathrm{Id} \circ F_{m'}^{-1} = F_{m'}^{-1} = \left(\mathrm{Id} \circ F_m^{-1}\right)^{-1} = \left(F_m^{-1}\right)^{-1} = F_m$. On vérifie ainsi le deuxième point de la proposition. □

**Proposition 6.** *La distribution uniforme $\eta$ est l'unique distribution pour laquelle l'opérateur $L_\eta$ appliqué à une distribution uniforme donne une distribution uniforme.*

*Démonstration.* Si $m$ et $m'$ sont deux distributions uniformes, alors $F_m = F_{m'} = 1 - \mathrm{Id}$. Et l'équation $1 - \mathrm{Id} = K \circ (1 - \mathrm{Id})$ donne bien sûr $K = \mathrm{Id}$. □

Par la suite, nous choisissons $\eta$ uniforme sur $[0, 1]$. **À cette étape de notre présentation, la distribution du risque de défaillance est définie par**

$$\boldsymbol{\mathcal{R} := \int_0^1 \delta_{R(\alpha)}\ \mathrm{d}\alpha} \tag{2.2}$$

### 2.1.3 Monte-Carlo préférentiel

Expliquons à présent pourquoi, du fait de la démarche pratique que nous suivons, la formule 2.2 doit être modifiée.

Comme annoncé dans l'introduction, la quantité $R(\alpha)$ a une expression intégrale et ne peut pas être exprimée sous forme analytique en pratique :

$$R(\alpha) = \int_{x \in X} \mathbb{1}_{\mathcal{M}(x) > \alpha}\ P(\mathrm{d}x).$$

Pour obtenir une approximation numérique de cette intégrale, nous utilisons une méthode Monte-Carlo préférentiel. Pour plus de détails sur les méthodes Monte-Carlo, voir Caflish (1998). Des tirages indépendants $(x_m)_{1 \leq m \leq M}$, dans l'espace des facteurs $X$, sont générés à partir de la distribution $P$. Pour chaque point $x_m$, la probabilité de défaillance

$\mathcal{M}(x_m) = \mathbb{P}(Y_{x_m} \in A)$ est ensuite calculée. Notons $n(\alpha)$ le nombre de tirages Monte-Carlo $x_m$ tels que $\mathcal{M}(x_m) > \alpha$. L'estimateur de $R(\alpha)$ donné par la simulation Monte-Carlo est le quotient $n(\alpha)/M$.

Notons qu'un raisonnement bayésien classique est utilisé pour modéliser l'incertitude due au tirage fini. Dans cette approche, le tirage Monte-Carlo est considéré comme une expérience binomiale. Cette hypothèse est tout à fait justifiée pour un générateur de nombres pseudo-aléatoires de bonne qualité. En conséquence, $R(\alpha)$ est une variable aléatoire dont la distribution est la distribution bêta définie par la densité

$$\beta_{n(\alpha)}(\lambda) := \frac{\Gamma(M+2)}{\Gamma(n(\alpha)+1)\,\Gamma(M-n(\alpha)+1)}\lambda^{n(\alpha)}(1-\lambda)^{M-n(\alpha)} \quad \text{pour} \quad 0 \leq \lambda \leq 1.$$

La prise en compte de ce nouvel aléa peut paraître un peu excessive. Néanmoins, cela permet d'éviter l'erreur classique (mais courante) qui consiste à considérer que le risque de défaillance est nul quand aucun accident n'a été observé.

**Finalement, nous définissons la densité du risque de défaillance comme suit :**

$$\boldsymbol{\mathcal{R} := \int_0^1 \beta_{n(\alpha)}\, d\alpha.} \tag{2.3}$$

### 2.1.4 Conclusion sur le modèle d'évaluation du risque de défaillance

Dans les raisonnements qui ont précédés, quatre aléas successifs ont été identifiés et considérés pour établir l'expression (2.3) du risque de défaillance, dus

- à la distribution de probabilité $P$ des facteurs sur l'espace $X$,
- à la modélisation de la fonction inconnue $y$ par un champ aléatoire $(Y_x)$,
- au seuil aléatoire $\alpha$ de tolérance au risque,
- à la distribution bêta.

Cependant, le fait que les facteurs soient des variables aléatoires et le fait que nous modélisions la fonction inconnue $y$ par un champ aléatoire n'introduisent pas d'aléa sur le risque de défaillance.

Seuls les deux derniers aléas constituent la source de l'aléa sur le risque de défaillance. En considérant le seuil $\alpha$ de tolérance au risque comme aléatoire, on prend en compte l'incertitude sur les prédictions du modèle. L'incertitude qui résulte de l'utilisation d'un tirage Monte-Carlo est prise en compte par la distribution bêta.

### 2.1.5 Mise en œuvre numérique du calcul du risque de défaillance

Dans cette section, nous précisons les formules numériques pour le calcul des deux premiers moments et d'un intervalle de confiance de la distribution du risque de défaillance à partir de la simulation Monte-Carlo, réalisés avec Matlab (version 2012.a). Nous nous référons à la documentation fourni par la société MathWorks pour les fonctions Matlab utilisées dans cette section. Le but est ici de limiter la manipulation des nombreuses données obtenues par le tirage Monte-Carlo ainsi que les calculs intermédiaires sur celles-ci afin d'avoir une exécution optimale, c'est-à-dire la moins consommatrice de temps et de

mémoire. Notons que l'utilisation d'un autre langage que Matlab, peut amener à considérer d'autres expressions des grandeurs d'intérêt.

Soient $(x_m)_{1\leq m\leq M}$ les tirages réalisés. Pour chaque $x_m$, la probabilité de défaillance $\mathcal{M}(x_m) = \mathbb{P}(Y_{x_m} \in A)$ est calculée numériquement, les distributions gaussienne et de Student étant tabulées. Supposons que parmi les $M$ probabilités de défaillance, il y en ait $M'$ distinctes qu'on peut classer par ordre croissant :

$$0 =: \mu_0 \leq \mu_1 < \mu_2 < \ldots < \mu_{M'} \leq \mu_{M'+1} := 1.$$

On définit $l_m$, comme étant le nombre de probabilités de défaillance égales à $\mu_m$. On a $l_1 + l_2 + \ldots + l_{M'} = M$. On souhaite calculer la distribution du risque de défaillance de densité :

$$\mathcal{R}(\lambda) = \int_0^1 \beta_{n(\alpha)}(\lambda)\, \mathrm{d}\alpha \quad \text{pour} \quad 0 \leq \lambda \leq 1$$

où $n(\alpha)$ est le nombre de tirages Monte-Carlo dont la probabilité de défaillance est strictement supérieure à $\alpha$. On remarque que, pour tous les seuils $\alpha$ tels que $\mu_m \leq \alpha < \mu_{m+1}$, la quantité $n(\alpha)$ est la même. On en déduit, pour $0 \leq \lambda \leq 1$ :

$$\mathcal{R}(\lambda) = \sum_{m=1}^{M'+1} \beta_{n_m}(\lambda)(\mu_m - \mu_{m-1})$$

où $n_m := \mathrm{card}\{k \in [1, 2, \ldots, M] \mid \mathcal{M}(x_k) > \mu_m\}$.
Nous avons $n_m = M - l_0 - l_1 - \ldots - l_m$ pour $1 \leq m \leq M'$ avec $n_0 := M$ et $n_{M'+1} := 0$ qui conduisent à $l_i = n_{i-1} - n_i$.

Nous obtenons les vecteurs $(\mu_m)$ et $(l_m)$ en appliquant la fonction Matlab "unique" aux probabilités de défaillance, calculées pour tous les points du tirage Monte-Carlo. Ces vecteurs peuvent être de très grande longueur. Cherchons donc des expressions pour la moyenne et la variance du risque de défaillance à partir de ces deux vecteurs $(\mu_m)$ et $(l_m)$ en limitant le plus possible leur manipulation. Dans ce but, rappelons le lemme de sommation d'Abel qui sera utile par la suite.

**Lemme** (Lemme de sommation d'Abel)**.** *Pour deux suites numériques $(f_k)_{k\in\mathbb{Z}}$ et $(g_k)_{k\in\mathbb{Z}}$, on a :*

$$\sum_{k=m}^{n} f_k(g_{k+1} - g_k) = (f_{n+1}g_{n+1} - f_m g_m) - \sum_{k=m}^{n} g_{k+1}(f_{k+1} - f_k).$$

#### 2.1.5.1 Moyenne

Rappelons que pour une distribution bêta de paramètres $(n(\alpha) + 1, M - n(\alpha) + 1)$, la moyenne vaut

$$\frac{n(\alpha) + 1}{M + 2}.$$

Ainsi la moyenne du risque de défaillance vaut :

$$\mathbb{E}(R) = \sum_{m=1}^{M'+1} \frac{n_m + 1}{M + 2}(\mu_m - \mu_{m-1}).$$

En développant puis en appliquant le lemme d'Abel, on obtient l'expression suivante :

$$\mathbb{E}(R) = \frac{1}{M+2}\left(1 + \sum_{m=1}^{M'} \mu_m l_{m+1}\right).$$

Le calcul se résume donc à un simple produit scalaire entre $(\mu_m)$ et $(l_m)$ qui est une opération particulièrement optimisée dans Matlab.

#### 2.1.5.2 Variance

Rappelons que l'espérance du carré d'une variable aléatoire suivant une loi bêta de paramètres $(n(\alpha)+1, M-n(\alpha)+1)$ est égale à

$$\frac{(n(\alpha)+1)(n(\alpha)+2)}{(M+2)(M+3)}.$$

La variance du risque de défaillance vaut alors :

$$\operatorname{var}(R) = \sum_{m=1}^{M'+1} \frac{(n_m+1)(n_m+2)}{(M+2)(M+3)}(\mu_m - \mu_{m-1}) - \mathbb{E}(\mathcal{R})^2.$$

C'est-à-dire, en développant :

$$\operatorname{var}(R) = \frac{1}{(M+2)(M+3)}\left(\sum_{m=1}^{M'+1} n_m^2(\mu_m - \mu_{m-1}) + 3\sum_{m=1}^{M'+1} n_m(\mu_m - \mu_{m-1}) + 2\sum_{m=1}^{M'+1}(\mu_m - \mu_{m-1})\right) - \mathbb{E}(\mathcal{R})^2.$$

En appliquant deux fois le lemme d'Abel, on trouve :

$$\operatorname{var}(R) = \frac{1}{(M+2)(M+3)}\left(2 + 3\sum_{m=1}^{M'} \mu_m l_{m+1} + \sum_{m=1}^{M'} \mu_m l_{m+1}(n_m + n_{m+1})\right) - \mathbb{E}(\mathcal{R})^2.$$

La complexité du calcul de la variance n'est ici pas améliorée. Néanmoins, sous cette forme, les fonctions de Matlab sont utilisées dans des conditions optimales et amènent un gain de temps d'exécution et de consommation mémoire non négligeable.

#### 2.1.5.3 Intervalle de confiance

Soit $c \in [0,1]$. L'intervalle $[a,b]$ de confiance $c$ est tel que :

$$\int_a^b \mathcal{R}(\lambda)\ \mathrm{d}\lambda = c.$$

Parmi tous les intervalles $[a,b]$ possibles, nous décidons de prendre le plus petit. Nous avons accès numériquement à la fonction Matlab "betainc" qui donne la fonction de répartition :

$$\operatorname{betainc}(l) := \int_0^l \mathcal{R}(\lambda)\ \mathrm{d}\lambda \quad \text{pour} \quad 0 \le l \le 1.$$

On recherche donc numériquement $a$ et $b$ tels que :

– $\text{betainc}(b) - \text{betainc}(a) = c$
– $b - a$ soit minimal.

Pour cela, nous utilisons conjointement les fonctions numériques Matlab "fzero" et "fminbnd".

## 2.2 Probabilité du risque d'insuffisance du modèle

### 2.2.1 Description de l'approche

Dans cette section, nous introduisons un indicateur mesurant la capacité du modèle à évaluer le risque de défaillance à partir des $n$ observations à disposition.

Considérons dans un premier temps, le cas d'un modèle parfait, c'est-à-dire un modèle qui prédit sans la moindre erreur la fonction $x \mapsto y(x)$. Il n'y a alors que deux valeurs possibles pour les probabilités de défaillance correspondantes : 0 (les "bonnes") ou 1 (les "mauvaises"). La distribution de probabilité du risque de défaillance, dont la construction est détaillée dans la section 2.1, est alors une masse de Dirac.

Considérons, maintenant, un modèle imparfait. Dans une première approche intuitive, on peut reconnaître trois sous-ensembles parmi les probabilités de défaillance : les "bonnes" (proches de 0), les "mauvaises" (proches de 1) et les "indécidables" (les autres). La distribution de probabilité du risque de défaillance est ici plus dispersée et peut ne pas permettre de conclure de façon totalement assurée quant à l'adéquation entre le processus industriel et le produit réalisé.

Nous introduisons deux seuils $\alpha$ et $\gamma$ (avec $0 \leq \alpha \leq \gamma \leq 1$) pour délimiter les trois classes : "bon", "indécidable" et "mauvais". En associant à l'insuffisance du modèle la mesure de la classe des indécidables, nous pouvons définir le risque d'insuffisance du modèle aux seuils $(\alpha, \gamma)$ de la façon suivante :

$$R_I(\alpha, \gamma) := P(\{x \in X \mid \alpha < \mathcal{M}(x) \leq \gamma\}) = R(\alpha) - R(\gamma).$$

Comme pour le risque de défaillance et pour des raisons similaires, nous allons ici considérer que les seuils $\alpha$ et $\gamma$ sont aléatoires suivant une loi de densité $h(\alpha, \gamma)$ définie sur $[0,1]^2$ et telle que $h(\alpha, \gamma)$ est nulle pour $\gamma < \alpha$. **La distribution du risque d'insuffisance est alors :**

$$\boldsymbol{\mathcal{R}_I := \int_0^1 \int_0^1 \delta_{R_I(\alpha,\gamma)} h(\alpha, \gamma) \ \mathrm{d}\alpha \ \mathrm{d}\gamma.} \tag{2.4}$$

Cette distribution est bien définie dès lors que la densité $h$ est fixée.

Le complémentaire de la fonction de répartition du risque d'insuffisance :

$$\mathcal{R}_I(]t, 1]) = \int_0^1 \int_0^1 \mathbb{1}_{[R(\alpha)-R(\gamma)>t]} h(\alpha, \gamma) \ \mathrm{d}\alpha \ \mathrm{d}\gamma$$

peut être exprimé en fonction de $R$, la fonction de queue de $\mathcal{M}$, comme suit :

$$\mathcal{R}_I(]t, 1]) = \int_0^1 \left( \int_0^{R^{-1}(t+R(\gamma))} h(\alpha, \gamma) \ \mathrm{d}\alpha \right) \mathrm{d}\gamma.$$

Avec cette définition, s'il est observé un risque d'insuffisance élevé, il est possible de conclure que le modèle n'est pas assez informé pour évaluer pertinemment le risque de défaillance. Notons que la réciproque est fausse. Un risque d'insuffisance faible est nécessaire mais pas suffisant pour juger de la qualité du modèle, en particulier quand les données utilisées pour le conditionnement sont peu représentatives du phénomène étudié.

### 2.2.2 Choisir la distribution des seuils

Une première approche non-informative, s'inspirant de la construction de la densité du risque de défaillance, consiste à choisir une densité $h$ uniforme :

$$h(\alpha, \gamma) = 2\ \mathbb{1}_{[0 \leq \alpha \leq \gamma \leq 1]}.$$

Dans cette section, nous allons présenter les conditions que nous jugeons nécessaires pour que $h$ soit satisfaisante et ainsi montrer que la densité uniforme ne convient pas.

#### 2.2.2.1 Risque d'insuffisance dans le cas d'un modèle mal informé.

Nous souhaitons que la densité $h$ permette une caractérisation sans ambiguïté des modèles insuffisants. Le modèle le plus insuffisant qui nous vienne intuitivement à l'esprit, est celui qui donne des probabilités de défaillance toutes égales à 0.5.

Dans ce cas, la distribution du risque d'insuffisance (2.4) est alors :

$$\mathcal{R}_I = \int_0^1 \int_0^1 \delta_{\mathbb{1}_{\left[\alpha < \frac{1}{2} \leq \gamma\right]}} h(\alpha, \gamma)\ \mathrm{d}\alpha\ \mathrm{d}\gamma$$

qui vaut

$$\mathcal{R}_I = \left( \int_0^{\frac{1}{2}} \int_{\frac{1}{2}}^1 h(\alpha, \gamma)\ \mathrm{d}\alpha\ \mathrm{d}\gamma \right) \delta_1 + \left( 1 - \int_0^{\frac{1}{2}} \int_{\frac{1}{2}}^1 h(\alpha, \gamma)\ \mathrm{d}\alpha\ \mathrm{d}\gamma \right) \delta_0.$$

Nous nous attendons à ce que ce modèle insuffisant conduise un risque de défaillance égale à 1. Pour remplir cette condition, il est alors nécessaire que la densité $h$ soit partout nulle sauf sur le carré :

$$\left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right]. \tag{2.5}$$

#### 2.2.2.2 Invariance de la distribution uniforme.

Nous souhaitons également que la transformation, qui nous permet d'obtenir la loi de probabilité du risque d'insuffisance d'après la loi de la probabilité de défaillance préserve l'état d'incertitude, tel que décrit dans la proposition 6 de la section précédente. Sous la condition que $\mathcal{M}(x)$ suive une distribution uniforme, ce qui équivaut à écrire

$$P(\{x \in X \mid \alpha < \mathcal{M}(x) \leq \gamma\}) = \gamma - \alpha,$$

on veut $\mathcal{R}_I = 1$.
Cela nous conduit à vouloir que $h$ satisfasse la condition suivante :

$$\int_0^1 \int_0^1 \delta_{\gamma-\alpha} h(\alpha, \gamma) \ \mathrm{d}\alpha \ \mathrm{d}\gamma = 1.$$

En intégrant une fonction $f$ continue contre cette mesure de probabilité, cette équation devient :

$$\int_0^1 \int_0^1 f(\gamma - \alpha) h(\alpha, \gamma) \ \mathrm{d}\alpha \ \mathrm{d}\gamma = \int_0^1 f(u) \ \mathrm{d}u$$

(pour toute fonction $f$ continue sur $[0, 1]$).
Avec le changement de variable $u = \gamma - \alpha$, nous avons :

$$\int_0^1 \left( \int_0^{1-u} h(\alpha, \alpha + u) \ \mathrm{d}\alpha \right) f(u) \ \mathrm{d}u = \int_0^1 f(u) \ \mathrm{d}u,$$

ce qui donne la condition suivante :

$$\int_0^{1-u} h(\alpha, \alpha + u) \ \mathrm{d}\alpha = 1. \tag{2.6}$$

#### 2.2.2.3 Association des deux conditions.

En associant les deux conditions (2.5) et (2.6), nous obtenons comme condition :

$$\int_{\max\left(\frac{1}{2}-u, 0\right)}^{\min\left(1-u, \frac{1}{2}\right)} h(\alpha, \alpha + u) \ \mathrm{d}\alpha = 1 \quad \text{pour} \quad u \in [0, 1].$$

Ainsi, la densité de probabilité $h$ satisfait cette condition si elle est nulle en dehors de $\left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right]$ et vérifie :

$$\begin{cases} \int_{\frac{1}{2}-u}^{\frac{1}{2}} h(\alpha, \alpha + u) \ \mathrm{d}\alpha = 1 & \text{si} \quad 0 \leq u \leq \frac{1}{2} \\ \int_0^{1-u} h(\alpha, \alpha + u) \ \mathrm{d}\alpha = 1 & \text{si} \quad \frac{1}{2} \leq u \leq 1 \end{cases} \tag{2.7}$$

Notons que ces deux conditions sont indépendantes. On observe également que nécessairement, $h(\alpha, \gamma)$ tend vers l'infini quand $\gamma - \alpha$ tend vers 0 ou vers 1. La densité uniforme n'est donc pas solution.

On remarque qu'il existe une unique densité $h$, définie sur $\left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right]$, solution de (2.7) et ne dépendant que de la quantité $\gamma - \alpha$ (voir figure 2.1) :

$$h(\alpha, \gamma) = \begin{cases} \frac{1}{\gamma - \alpha} & \text{si} \quad 0 < \gamma - \alpha \leq \frac{1}{2} \\ \frac{1}{1-(\gamma-\alpha)} & \text{si} \quad \frac{1}{2} \leq \gamma - \alpha < 1 \end{cases}$$

En fait, une infinité de densités satisfont la condition (2.7) qui ne s'avère pas être très restrictive. Par exemple, en considérant $h(\alpha, \gamma)$ sous la forme particulière du produit d'une fonction de $\alpha$ par une fonction de $\gamma - \alpha$, nous obtenons la famille de densités de la forme

$$h(\alpha, \gamma) = \frac{f(\alpha)}{\int_{\frac{1}{2}-(\gamma-\alpha)}^{\frac{1}{2}} f(t) \ \mathrm{d}t} \quad \text{si} \quad 0 \leq \gamma - \alpha \leq \frac{1}{2}$$

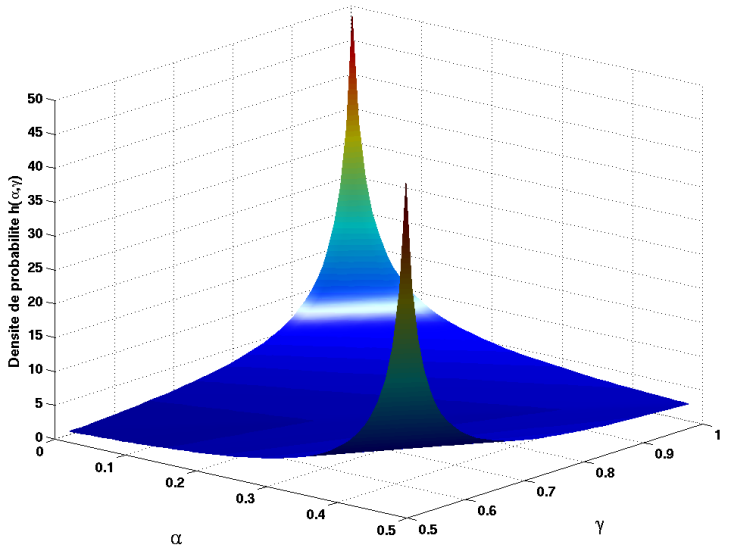



FIGURE 2.1 – Illustration de la fonction $h$, dépendant seulement de la quantité $\gamma - \alpha$.

et

$$h(\alpha, \gamma) = \frac{g(\alpha)}{\int_0^{1-(\gamma-\alpha)} g(t)\ \mathrm{d}t} \quad \text{si} \quad \frac{1}{2} \leq \gamma - \alpha \leq 1$$

pour $f$ et $g$ positives et intégrables sur l'intervalle $\left[0, \frac{1}{2}\right]$.

### 2.2.3 Monte-Carlo préférentiel

Comme pour le risque de défaillance, une fois fixée la densité $h$ pour les seuils $\alpha$ et $\gamma$, nous devons effectuer une simulation Monte-Carlo, suivant la distribution $P$, afin d'approcher numériquement la quantité $R_I(\alpha, \gamma)$. Nous calculons ensuite, pour chaque tirage $x_m$, la probabilité de défaillance $\mathcal{M}(x_m) = \mathbb{P}(Y_{x_m} \in A)$ et notons $n(\alpha, \gamma)$ le nombre de tirages Monte-Carlo $x_m$ parmi les $M$ réalisés tels que $\alpha < \mathcal{M}(x_m) \leq \gamma$.

En appliquant le même raisonnement que celui développé pour la définition du risque de défaillance (section 2.1.3), nous considérons que $R_I(\alpha, \gamma)$ est une variable aléatoire de distribution bêta définie par la densité suivante, pour $0 \leq \lambda \leq 1$ :

$$\beta_{n(\alpha,\gamma)}(\lambda) := \frac{\Gamma(M+2)}{\Gamma(n(\alpha,\gamma)+1)\,\Gamma(M-n(\alpha,\gamma)+1)} \lambda^{n(\alpha,\gamma)} (1-\lambda)^{M-n(\alpha,\gamma)}.$$

**La distribution du risque d'insuffisance du modèle est donc finalement définie par la densité**

$$\boldsymbol{\mathcal{R}_I := \int_0^1 \int_0^1 \beta_{n(\alpha,\gamma)} h(\alpha, \gamma)\ \mathrm{d}\alpha\ \mathrm{d}\gamma.} \tag{2.8}$$

Ainsi, le décideur dispose d'un indicateur lui permettant de mesurer la qualité de la prédiction du risque de défaillance. Dans le cas où le risque d'insuffisance est jugé trop élevé, cela signifie que le modèle n'est pas suffisamment renseigné. La prochaine section de ce chapitre présente des méthodes afin de mieux informer le modèle.

## 2.3 Calcul du risque de défaillance : synoptique

Le premier chapitre de cette thèse traite de la construction d'un champ aléatoire à partir d'un jeu restreint de $n$ données. Nous avons introduit dans ce chapitre, les distributions du risque de défaillance et du risque d'insuffisance.

Le décideur a donc accès à l'espérance du risque de défaillance qui peut lui permettre de statuer quant à la stabilité des performances du produit vis-à-vis des variations inhérentes au processus industriel de fabrication. De plus, il a accès à un intervalle de confiance, qui indique la fiabilité de cet estimateur, les sources d'incertitudes étant la connaissance partielle de la fonction d'origine $y$ et un échantillonnage Monte-Carlo fini. Le risque d'insuffisance du modèle informe également sur la pertinence de cet estimateur.

Ainsi, dans le cas où la qualité de la prédiction est jugée trop faible pour permettre une prise de décision, il est nécessaire d'ajouter de nouvelles données au jeu de données initial. Comme nous l'avons déjà indiqué, il est coûteux d'obtenir ces données. C'est pourquoi, nous souhaitons construire un plan d'expériences. Le but des plans d'expériences est de déterminer une liste de points de $X$, les expériences, optimale au sens où elle permet de minimiser l'incertitude sur les paramètres du modèle. Ainsi, un plan d'expériences est spécifiquement réalisé pour le modèle qu'il renseigne. Voir Atkinson et Donev (1992) pour une présentation plus détaillée de la théorie des plans d'expériences. Dans notre cas, nous cherchons à établir un plan d'expériences tel que le modèle aléatoire qui en résulte permette une meilleure classification des points de l'espace des facteurs.

L'échantillonnage Monte-Carlo est basé sur la donnée de la probabilité $P$ portée par l'espace des facteurs $X$. Cependant, le choix des points d'observation permettant la construction du modèle aléatoire n'est en rien contraint par cette donnée.

Nous choisissons d'ajouter des points d'observation dans les zones de l'espace des facteurs $X$ où le modèle aléatoire nous apparaît comme mal informé. Nous retenons pour cela l'entropie différentielle

$$h\left(Y_x\right) := \mathbb{E}\left(-\log f\left(Y_x\right)\right)$$

où $f$ est la densité de la variable aléatoire $Y_x$, comme une mesure du manque d'information (Cover et Thomas (2006)). L'information fournie au modèle aléatoire sera d'autant plus importante que les observations seront effectuées aux points de $X$ où l'entropie est maximum.

La construction du plan d'expériences se résume alors à un problème d'optimisation multi-modal : l'objectif est de trouver un ensemble de maxima locaux pour $h\left(Y_x\right)$.

La figure 2.2, illustre deux exemples en dimension 2. Dans celui de gauche, il y a eu 9 points initiaux répartis de façon régulière dans l'espace $[0, 1]$. On y observe 4 maxima globaux pour la fonction $h\left(Y_x\right)$. Dans l'exemple de droite, composé de 5 points initiaux non régulièrement disposés, on observe sur la fonction entropie, plusieurs maxima, locaux et globaux, traduisant une plus grande complexité que dans le premier exemple. Il est intéressant de noter que les minima de la fonction $h\left(Y_x\right)$ sont toujours connus, ce sont les points déjà observés $x_1, x_2, \ldots, x_n$.

Ce n'est pas un problème d'optimisation classique car en général seul l'optimum global est recherché. Cependant, plusieurs approches sont possibles pour résoudre cette probléma-

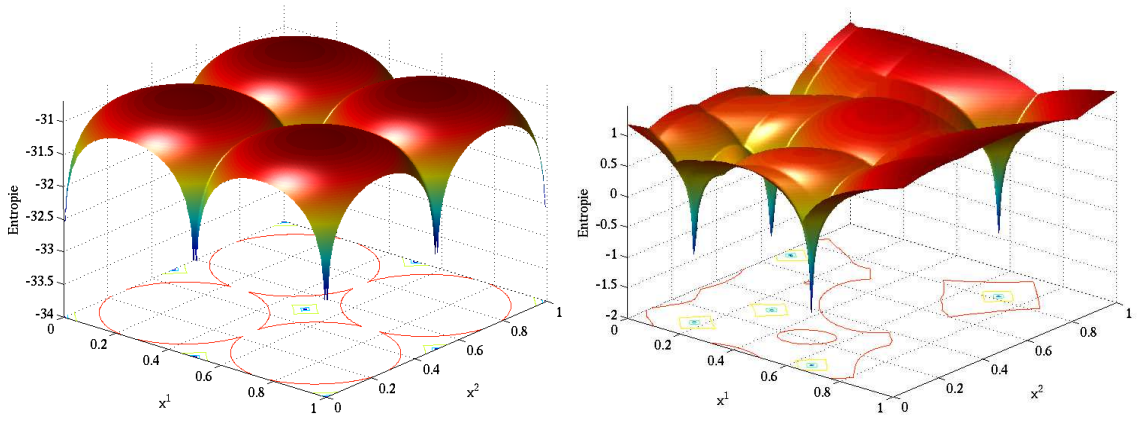



FIGURE 2.2 – Surfaces d'entropie en 2D

tique. Les méthodes conventionnelles comme, par exemple, la descente de gradient peuvent être appliquées. Ces méthodes, généralement développées pour trouver le maximum global, peuvent être "détériorées" afin qu'elles tombent dans le "piège" que constituent les maxima locaux. En itérant ces méthodes, nous collectionnons en quelque sorte les maxima locaux.

Il est possible également d'utiliser des méthodes d'optimisations plus sophistiquées comme les méta-heuristiques. Nous nous référons à Talbi (2009) pour une présentation complète des méta-heuristiques. Les algorithmes évolutionnaires, comme par exemple les algorithmes génétiques, en font partis. On pourra consulter Le Riche et al. (2006) pour une description détaillée de ces méthodes stochastiques, qui s'inspirent, à l'origine, de la théorie de Darwin. Il existe des algorithmes mimétiques développés en imitant le comportement social de certains animaux, comme les fourmis (Monmarché et al. (2010)), les abeilles (Clerc (2003)) ou encore les lucioles (Yang (2009)). Notons qu'il n'existe aucun algorithme qui garantisse de trouver tous les maxima locaux.

Plus précisément, dans le logiciel GoNoGo que nous présenterons dans le troisième chapitre, nous avons utilisé l'algorithme génétique fourni par MathWorks (dans la boîte à outil "Global Optimization Toolbox") et avons implémenté l'algorithme TRIBES de Clerc (2003). Nous avons également testé, à titre expérimental, l'algorithme de Yang (2009).

Une fois que des points candidats sont identifiés, on recommence l'étude schématisée par la figure 2.3. Cette procédure itérative est répétée jusqu'à ce que les exigences de prédictions au regard du risque de défaillance soient validées ou que le temps alloué à cette étude soit consommé.

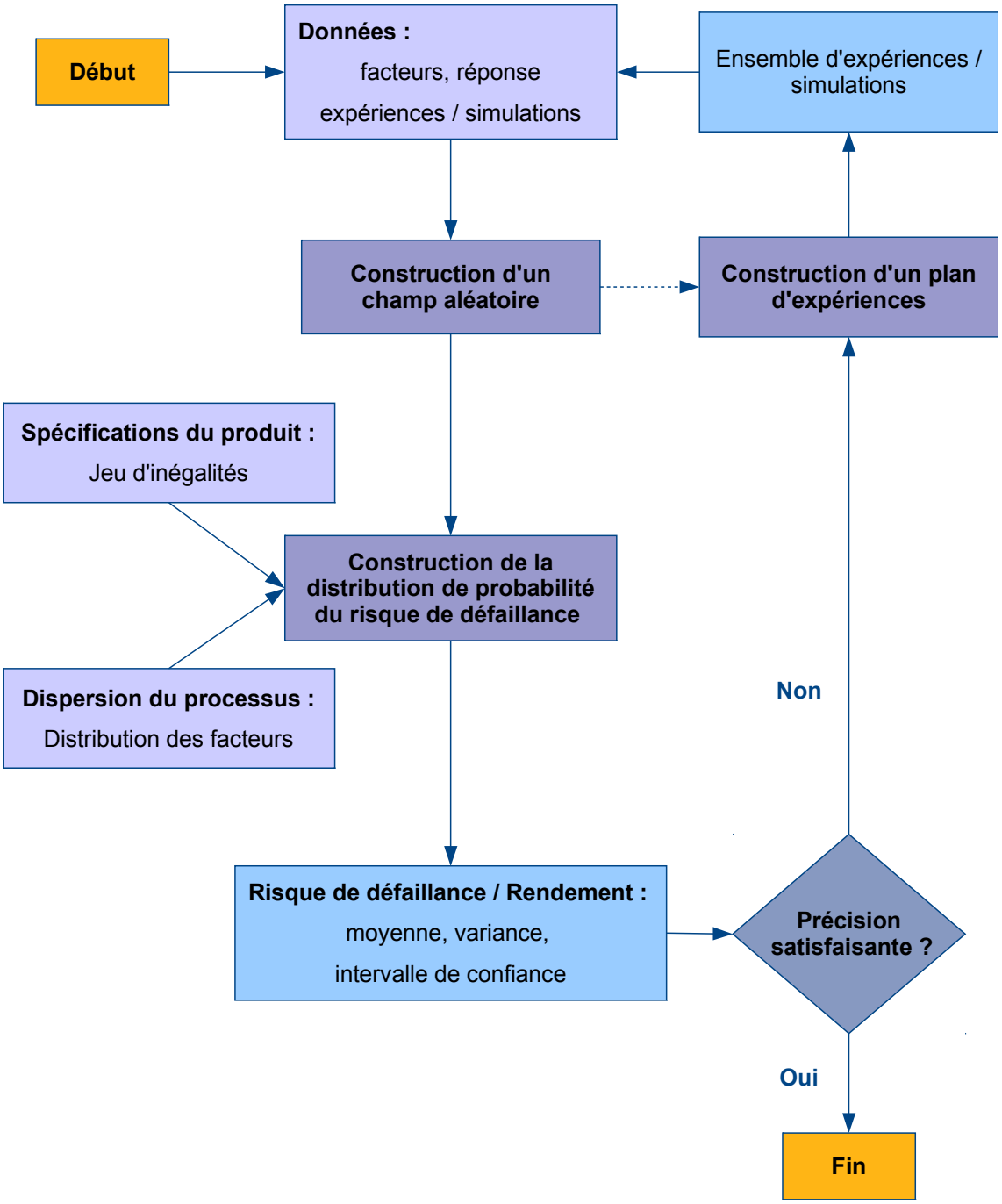



FIGURE 2.3 – Synoptique de l'algorithme de calcul de la distribution du risque de défaillance.

## 2.4 Calcul de rendement sur des exemples théoriques

Dans le but de démontrer l'intérêt de la méthode d'évaluation du risque de défaillance présentée dans la section 2.1, nous considérons trois exemples théoriques. Ces exemples ont été retenus car il est possible d'exprimer analytiquement le risque de défaillance, ce qui nous permet de le calculer exactement et de comparer notre estimateur à celui-ci.

Nous utilisons le modèle de Student (section 1.4.2) associé à la fonction de corrélation anisotrope $\gamma$-exponentielle (section 1.6.1.1) paramétrée par la méthode du maximum de vraisemblance (section 1.6.2). Le modèle a priori a $D+1$ paramètres, un niveau raisonnable de complexité. La distribution des facteurs $P$ est supposée uniforme.

Comme alternative au tirage Monte-Carlo basé sur notre modèle aléatoire (Model-based Monte Carlo - MMC), nous envisageons un Monte-Carlo brut (Brute-force Monte Carlo - BMC) comme décrit dans la section 2.1.3. Soit $k$ le nombre de tirages défectueux, i.e. tels que $y(x) \in A$, parmi les $M$ tirages. Le risque de défaillance défini par la méthode BMC suit une distribution bêta de paramètres $(k+1, M-k+1)$. Notons que ce modèle basique suppose un tirage dans l'espace des facteurs suivant la distribution $P$ (uniforme). Les échantillons sont tirés de façon quasi-aléatoire dans l'espace des facteurs : $(x_m)_{1 \leq m \leq M}$ est une suite à discrépance faible, la séquence de Sobol. Un tel tirage couvre l'espace des facteurs uniformément, une propriété souhaitée pour obtenir une population statistique représentative. Pour chacune des méthodes MMC et BMC, la moyenne et les bornes du plus petit intervalle calculé avec 90% de confiance sont tracées en fonction du nombre $M$ d'évaluations, $10 \leq M \leq 1000$.

Avant d'aller plus loin, il est important d'avoir à l'esprit que l'intervalle de confiance fournit seulement une estimation statistique quant à l'erreur sur le résultat. Il n'implique évidemment pas une condition stricte d'appartenance à l'intervalle. Quelle que soit la méthode d'échantillonnage retenue, celle-ci ne peut pas à coup sûr extraire toutes les caractéristiques essentielles de la fonction $y$. En conséquence, l'incertitude peut être sous-estimée, une situation dangereuse pour l'évaluation de risque. C'est particulièrement vrai pour la méthode BMC qui ne prend pas en compte la distribution spatiale des données (structure géométrique de l'espace des facteurs $X$). C'est la différence clé avec la méthode MMC que nous proposons. C'est pourquoi, la comparaison des propriétés de convergence de ces deux méthodes ne devrait pas seulement porter sur les intervalles de confiance.

Dans chacun de ces trois exemples, l'espace des facteurs $X$ est un pavé de $\mathbb{R}^D$ de la forme $X = \prod_{i=1}^{D} [X_{\min,i}, X_{\max,i}]$. Un nombre réel $m$ est fixé et détermine l'intervalle *hors spécification* $A := ]m, +\infty[$. En conséquence, le domaine de défaillance est :

$$X_A := \{x \in X \mid y(x) > m\}.$$

Le risque de défaillance réel est donné par $r_t = P(X_A)$.

### 2.4.1 Exemple de la forme quadratique.

#### 2.4.1.1 Description de l'exemple

Soient $m > 0, a_1, a_2, \ldots, a_D > m^2$, $X := [-1,1]^D$ et pour $\boldsymbol{x} := \left(x_{(1)}, x_{(2)}, \ldots, x_{(D)}\right) \in X$ :

$$y\left(\boldsymbol{x}\right) := \sqrt{\sum_{i=1}^{D} a_i x_{(i)}^2}.$$

La frontière du domaine de défaillance $X_A$ est une généralisation d'une ellipse en dimension $D$. Le volume de ce domaine de défaillance est :

$$\frac{1}{\mathrm{vol}\left(X\right)} \int_X \mathbb{1}_{[y(\boldsymbol{x})>m]} \, \mathrm{d}\boldsymbol{x} = 1 - \frac{1}{2^D} \frac{\pi^{\frac{D}{2}} m^D}{\Gamma\left(\frac{D}{2}+1\right)} \prod_{i=1}^{D} a_i^{-\frac{1}{2}}.$$

Plus précisément, nous considérons un exemple en dimension $D = 5$ avec $(a_1 = 4.0, a_2 = 4.2, a_3 = 4.4, a_4 = 4.6, a_5 = 4.8)$ et $m = 1.9$. Le risque théorique est alors $r_t = 1.008 \times 10^{-1}$.

#### 2.4.1.2 Analyse des résultats

Regardons la figure 2.4 pour une comparaison des convergences des deux tirages Monte-Carlo effectués. Il s'avère que, dans cet exemple, la méthode MMC surpasse incontestablement la méthode BMC puisque qu'elle permet l'estimation du risque de défaillance à précision équivalente avec deux tiers d'observations en moins.

En effet, bien que la méthode MMC intègre plus de sources d'incertitude, l'intervalle de confiance qu'elle propose est plus petit. Un échantillon de 60 données est nécessaire pour obtenir un intervalle de confiance de taille inférieur à 0.05. Pour obtenir ce résultat avec la méthode BMC, il nous faut 400 points. Par ailleurs, notons que l'intervalle de confiance donné par la méthode MMC contient toujours la vraie valeur du risque de défaillance.

Tout cela suggère que les hypothèses a priori du modèle sont appropriées et améliorent significativement la précision de la prédiction sur le risque de défaillance, ce qui n'est pas surprenant puisque la fonction "test" de cet exemple est relativement simple.

### 2.4.2 Exemple du sinus

#### 2.4.2.1 Description de l'exemple

Soient $a_1, a_2, \ldots, a_D$ des entiers non nuls, $X := [0,1]^D$ et pour $\boldsymbol{x} := \left(x_{(1)}, x_{(2)}, \ldots, x_{(D)}\right) \in X$ :

$$y\left(\boldsymbol{x}\right) := \sin\left(2\pi \sum_{i=1}^{D} a_i x_{(i)}\right).$$

Il peut être démontré que pour tout $m \in [-1,1]$ :

$$\int_X \mathbb{1}_{[y(\boldsymbol{x})>m]} \, \mathrm{d}\boldsymbol{x} = \int_0^1 \mathbb{1}_{[\sin(2\pi t)>m]} \, \mathrm{d}t = \frac{1}{2} - \frac{\arcsin\left(m\right)}{\pi}.$$

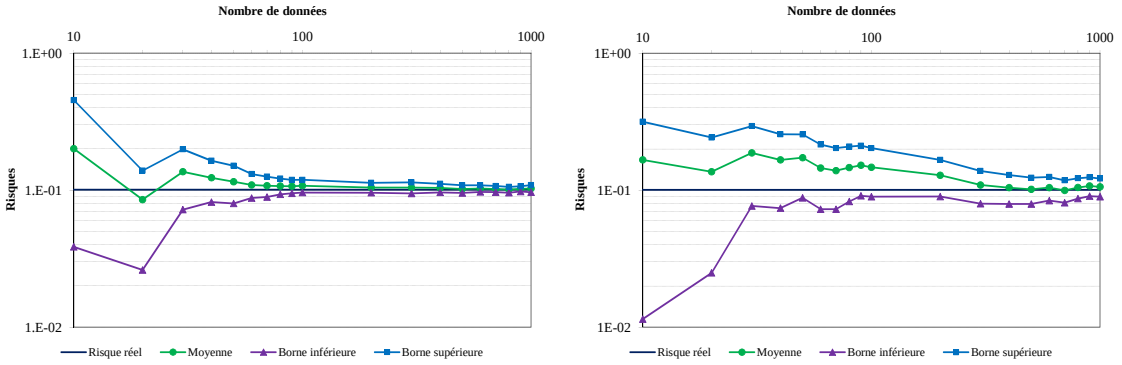



FIGURE 2.4 – Exemple de la forme quadratique : convergence pour la méthode MMC (à gauche) et la méthode BMC (à droite)

En effet, la formule

$$\int_{[0,1]^D} f\left(a_1x_1 + a_2x_2 + \ldots + a_Dx_D\right)\ \mathrm{d}x_1\ \mathrm{d}x_2 \ldots\ \mathrm{d}x_D = \int_0^1 f\left(x\right)\ \mathrm{d}x.$$

se vérifie successivement :

- pour les polynômes trigonométriques 1-périodiques par un calcul direct.
- pour les fonctions continues 1-périodiques sur $\mathbb{R}$ par un argument d'approximation uniforme.
- pour les fonctions en escalier 1-périodiques sur $\mathbb{R}$ par un argument d'encadrement.

Plus précisément, nous considérons un exemple en dimension $D = 3$, avec $(a_1 = 1, a_2 = -1, a_3 = 2)$ et $m$ de valeur 0.8. Le risque théorique est ici de $r_t = 2.048 \times 10^{-1}$.

#### 2.4.2.2 Analyse des résultats

À cet exemple, correspond un problème plus délicat que le précédent puisque la fonction à modéliser est oscillante. Cependant, nous nous attendons à ce que la méthode BMC donne des résultats satisfaisants puisque le risque réel est relativement grand. En première approche, la taille de l'échantillon requis lors du tirage MC peut être assez simplement estimée : $M >> 1/r_t$. Cette condition est facilement remplie dans cet exemple.

Les deux méthodes Monte-Carlo donnent des résultats similaires, représentés dans la figure 2.5. L'erreur relative de la prédiction moyenne du risque de défaillance obtenue par la méthode MMC est inférieure à 5% au delà de 300 données, à comparer aux 400 données nécessaire à la méthode BMC. Les tailles des intervalles de confiance obtenus par les méthodes MMC et BMC sont inférieures à 0.05 pour des échantillons respectivement de 600 et 800 points.

Nous remarquons, que pour des échantillons de moins de 200 points, l'intervalle de confiance donné par la méthode MMC est large, ce qui implique que le conditionnement du modèle a priori ne restreint pas assez l'ensemble des fonctions d'interpolation (défini

par la fonction de corrélation). La propagation de l'information est ainsi faible et nous en concluons que la fonction de corrélation $\gamma$-exponentielle n'est pas la plus appropriée. De plus, on observe un pincement de l'intervalle de confiance avec 30 observations, dû à de fortes variations des paramètres de la fonction de corrélation. La méthode MLE atteint ses limites quand l'information est rare, d'abord parce qu'elle devient très sensible à l'échantillon d'apprentissage et d'autre part parce que le posterior bayésien peut difficilement être approché par son mode.

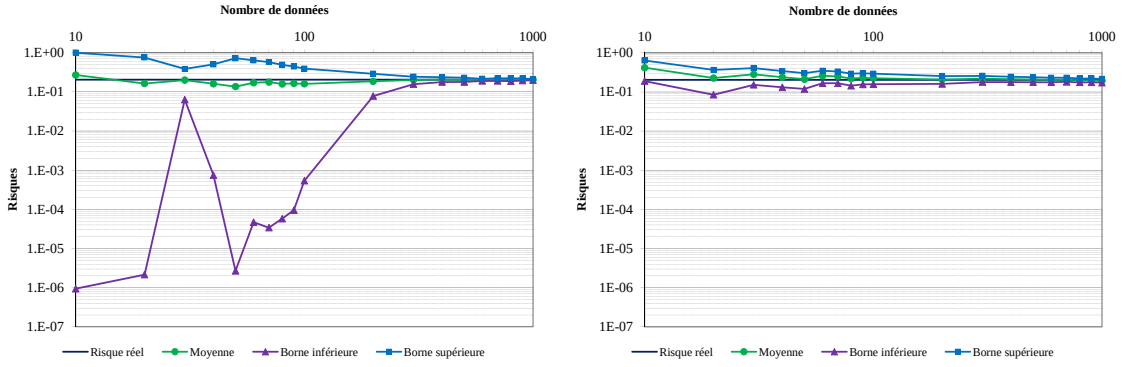



FIGURE 2.5 – Exemple du sinus : convergence pour la méthode MMC (à gauche) et la méthode BMC (à droite)

### 2.4.3 Exemple des gaussiennes

#### 2.4.3.1 Description de l'exemple

Soient $R$ un entier positif, $X = [0,1]^D$ et pour $\boldsymbol{x} \in X$ :

$$y(\boldsymbol{x}) := \max_{1 \leq i \leq R} f_i(\boldsymbol{x}, \boldsymbol{\mu}_i, \sigma_i)$$

où, pour tout $1 \leq i \leq R$, $\boldsymbol{\mu}_i \in \mathbb{R}^D$, $\sigma_i \in \mathbb{R}$, $\sigma_i > 0$ et

$$f_i(\boldsymbol{x}, \boldsymbol{\mu}_i, \sigma_i) = \frac{1}{\left(\sqrt{2\pi}\sigma_i\right)^D} \exp\left(-\frac{1}{2\sigma_i^2} \|\boldsymbol{x} - \boldsymbol{\mu}_i\|^2\right).$$

On appelle domaine de défaillance associé à $f_i$ l'ensemble $\{x \in X \mid f_i(\boldsymbol{x}, \boldsymbol{\mu}_i, \sigma_i) > m\}$. Plus précisément, pour obtenir une expression analytique de $r_t$, nous retenons les conditions suivantes, impliquant les quantités :

$$r_i := \sqrt{-2\sigma_i^2 \log\left(m (2\pi)^{\frac{D}{2}} \sigma_i^D\right)}, \quad 1 \leq i \leq R$$

1. Les domaines de défaillance ne sont pas vides si :

$$0 < \sigma_i < \frac{1}{\sqrt{2\pi} m^{\frac{1}{D}}} \quad \text{pour } 1 \leq i \leq R$$

2. Les domaines de défaillance sont complètement inclus dans l'espace $X$ si :

$$0 \leq \mu_{ij} \pm r_i \leq 1 \quad \text{pour } 1 \leq i \leq R \text{ et } 1 \leq j \leq D$$

3. Les domaines de défaillance sont disjoints si :

$$\left\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right\| > r_i + r_j \quad \text{pour } j \neq i.$$

Ainsi, pour chaque $1 \leq i \leq R$, le domaine de défaillance associé à $f_i$ est une hyper-sphère de dimension $D$ centrée en $\boldsymbol{\mu}_i$ et de rayon $r_i$. Son volume est :

$$\frac{\pi^{\frac{D}{2}} r_i^D}{\Gamma\left(\frac{D}{2}+1\right)}.$$

Pour un seuil fixé $m \in \mathbb{R}$, les domaines de défaillance associés à chaque $f_i$ ne se chevauchent pas. Nous avons alors :

$$\int_X \mathbb{1}_{[y(\boldsymbol{x})>m]} \, \mathrm{d}\boldsymbol{x} = \sum_{i=1}^{R} \int_X \mathbb{1}_{[f_i(\boldsymbol{x},\boldsymbol{\mu}_i,\sigma_i)>m]} \, \mathrm{d}\boldsymbol{x}.$$

Finalement, le risque de défaillance est simplement la somme des volumes de $R$ hyper-sphères disjointes de dimension $D$ :

$$\int_X \mathbb{1}_{[y(\boldsymbol{x})>m]} \, \mathrm{d}\boldsymbol{x} = \frac{\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}+1\right)} \sum_{i=1}^{R} r_i^D.$$

En particulier, nous étudions un exemple de dimension $D = 5$, avec $R = 10$ et $m$ qui vaut 0.1. La fonction $y$ est composée de 10 fonctions gaussiennes multi-dimensionnelles définies dans le tableau 2.1.

Le risque théorique est alors de $r_t = 4.341 \times 10^{-3}$.

#### 2.4.3.2 Analyse des résultats

En dessous de 700 observations, la moyenne statistique décroit en $O\left(n^{-1}\right)$, ce qui est typique d'un sous-échantillonnage du domaine de défaillance. En effet, le premier accident est observé à partir de 800 observations. À la fin du processus d'échantillonnage ($M = 1000$), l'erreur relative moyenne du risque de défaillance est encore de 23% pour la méthode MMC et de 31% pour la méthode BMC. La méthode BMC conduit à des résultats trop optimistes : la borne supérieure de l'intervalle de confiance à 90% est proche du risque réel. La prédiction de la méthode MMC est plus robuste : $r_t$ est correctement encadré par l'intervalle de confiance et le ratio de la prédiction moyenne à la taille de l'intervalle de confiance indique sans ambiguïté que des données additionnelles sont requises.

Il n'est pas facile d'arriver à cette conclusion, en se basant sur l'analyse brute de l'erreur de prédiction (calculée dans la section 1.7.2). Nous notons que $\epsilon_r \leq 3.7\%$ avec une probabilité de 0.9 et que sa valeur moyenne est environ égale à 1.7%, ce qui n'est pas exactement la signature d'un mauvais modèle. Contrairement à l'analyse de sensibilité globale (voir, par exemple, Oakley et O'Hagan (2004)), les conditions requises pour que le

| | $\mu_{i1}$ | $\mu_{i2}$ | $\mu_{i3}$ | $\mu_{i4}$ | $\mu_{i5}$ | $\sigma_i$ |
|---|---|---|---|---|---|---|
| $f_1$ | $4.889 \times 10^{-1}$ | $6.241 \times 10^{-1}$ | $6.791 \times 10^{-1}$ | $3.955 \times 10^{-1}$ | $3.674 \times 10^{-1}$ | $6.259 \times 10^{-1}$ |
| $f_2$ | $5.870 \times 10^{-1}$ | $4.139 \times 10^{-1}$ | $3.091 \times 10^{-1}$ | $2.638 \times 10^{-1}$ | $7.588 \times 10^{-1}$ | $6.297 \times 10^{-1}$ |
| $f_3$ | $7.637 \times 10^{-1}$ | $5.588 \times 10^{-1}$ | $1.838 \times 10^{-1}$ | $4.980 \times 10^{-1}$ | $5.179 \times 10^{-1}$ | $6.292 \times 10^{-1}$ |
| $f_4$ | $2.318 \times 10^{-1}$ | $3.963 \times 10^{-1}$ | $7.051 \times 10^{-1}$ | $5.586 \times 10^{-1}$ | $7.566 \times 10^{-1}$ | $6.299 \times 10^{-1}$ |
| $f_5$ | $5.806 \times 10^{-1}$ | $2.989 \times 10^{-1}$ | $7.137 \times 10^{-1}$ | $3.605 \times 10^{-1}$ | $7.185 \times 10^{-1}$ | $6.298 \times 10^{-1}$ |
| $f_6$ | $3.805 \times 10^{-1}$ | $3.022 \times 10^{-1}$ | $2.945 \times 10^{-1}$ | $2.598 \times 10^{-1}$ | $4.635 \times 10^{-1}$ | $6.258 \times 10^{-1}$ |
| $f_7$ | $1.684 \times 10^{-1}$ | $7.401 \times 10^{-1}$ | $8.728 \times 10^{-1}$ | $3.555 \times 10^{-1}$ | $1.788 \times 10^{-1}$ | $6.304 \times 10^{-1}$ |
| $f_8$ | $7.885 \times 10^{-1}$ | $5.891 \times 10^{-1}$ | $4.363 \times 10^{-1}$ | $5.497 \times 10^{-1}$ | $3.530 \times 10^{-1}$ | $6.300 \times 10^{-1}$ |
| $f_9$ | $8.601 \times 10^{-1}$ | $3.204 \times 10^{-1}$ | $6.636 \times 10^{-1}$ | $9.056 \times 10^{-1}$ | $6.854 \times 10^{-1}$ | $6.318 \times 10^{-1}$ |
| $f_{10}$ | $2.995 \times 10^{-1}$ | $8.727 \times 10^{-1}$ | $3.751 \times 10^{-1}$ | $7.984 \times 10^{-1}$ | $8.168 \times 10^{-1}$ | $6.318 \times 10^{-1}$ |

TABLE 2.1 – Valeurs des paramètres de l'exemple des gaussiennes.

modèle soit adéquat ne dépendent pas uniquement de la distribution $P$. En effet, dans ce type d'analyse, le modèle doit approximer la fonction référence $y$ sur un sous-espace de $X$ d'autant plus précisément que la probabilité de visite de ce domaine (déterminée par $P$) est grande. Dans notre cas, l'ensemble hors-spécifications $A$ joue un rôle tout aussi important. Ainsi, sur cet exemple, une bonne précision est également requise aux frontières du domaine de défaillance $X_A$. En conséquence, la décision relative à la qualité du modèle semble difficilement pouvoir reposer sur un critère global.

D'autre part, notons que se concentrer sur les performances intrinsèques du modèle pour en tirer des conclusions sur la qualité de la mesure du domaine de défaillance peut mener à de la "sur-qualité". En effet, le nombre d'observations nécessaires à l'obtention d'un modèle correct sur $X$ peut être largement supérieur au nombre de points conduisant à une estimation satisfaisante de $r_t$.

L'évaluation probabiliste du risque de défaillance que nous proposons propage l'incertitude sur les prédictions du modèle jusqu'à la grandeur d'intérêt, la mesure de $X_A$. Elle surmonte naturellement les difficultés précédemment évoquées.

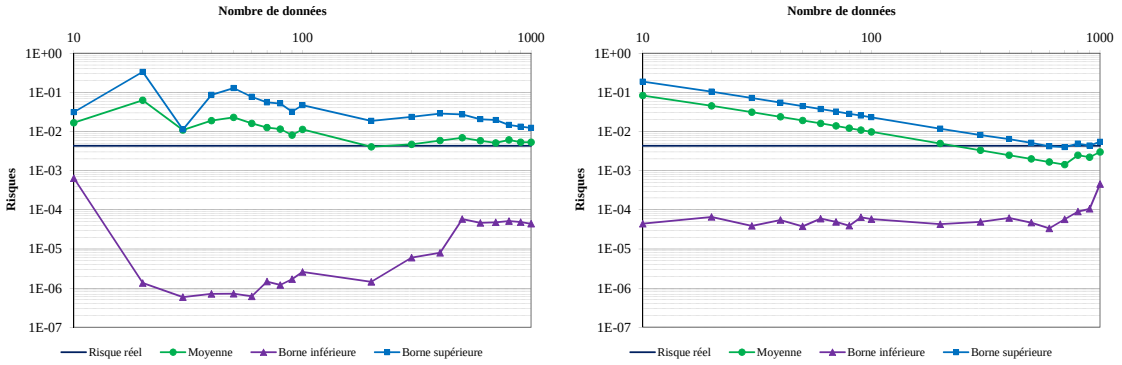



FIGURE 2.6 – Exemple des gaussiennes : convergence pour la méthode MMC (à gauche) et la méthode BMC (à droite)

## 2.5 Influence des facteurs

Il n'est pas rare que l'ingénieur en charge du développement d'un produit souhaite savoir quels sont les facteurs qui influent le plus sur les caractéristiques du produit, surtout quand celui-ci est complexe. En effet, un facteur peu influent peut être omis sans trop de risque dans l'étude des performances du produit et ainsi simplifier cette dernière. À l'inverse, un facteur très influent doit être traité avec plus d'attention, plus de moyens. Nous nous intéressons, dans cette partie, à l'impact des facteurs sur le risque de défaillance. On notera que ce n'est pas la même problématique que l'étude directe de l'influence des facteurs sur la fonction réponse $y$.

### 2.5.1 Méthodologie

Pour étudier l'influence sur le risque de défaillance du $i$-ème facteur d'un produit comportant $D$ facteurs, une méthode intuitive consiste à calculer la distribution du risque de défaillance de densité $\mathcal{R}$ du problème initial ($n$ données, de dimension $D$) ainsi que la distribution du risque de défaillance de densité $\tilde{\mathcal{R}}$ du problème réduit ($n$ données, de dimension $D-1$) où on a omis le $i$-ème facteur. On compare ensuite ces deux distributions : si elles sont proches, voire égales, alors on conclut que le $i$-ème facteur est peu ou n'est pas influent ; à l'inverse, si elles diffèrent de façon plus importante, on conclut que ce facteur est influent.

Le défaut de cette méthode est que le problème réduit comporte nécessairement plus d'information que le problème initial, en proportion de la dimension de l'espace. Le problème réduit conduit ainsi à un modèle aléatoire ayant une variance conditionnelle de prédiction probablement plus faible que celle obtenue pour le problème initial. Ainsi, les distributions $\mathcal{R}$ et $\tilde{\mathcal{R}}$ peuvent être plus ou moins proches, indépendamment de l'influence du facteur étudié.

Il nous faut donc travailler avec des problèmes de même dimension, pour avoir des

modèles renseignés avec la même quantité d'information. C'est pourquoi, nous proposons l'approche suivante. On calcule dans un premier temps la distribution du risque de défaillance de densité $\mathcal{R}$ du problème initial ($n$ données, de dimension $D$). Puis on construit un problème modifié, à partir du modèle initial, où les valeurs du $i$-ème facteur des $n$ données sont tirées aléatoirement et uniformément dans l'espace $X_i$ dans lequel varie le $i$-ème facteur $x_{(i)}$. On calcule ensuite la distribution du risque de défaillance de densité $\tilde{\mathcal{R}}$ de ce problème modifié.

On compare ces deux distributions à l'aide d'un critère de comparaison comme ceux décrit dans la section qui suit.

### 2.5.2 Critère de comparaison

Dans cette section, sont décrits deux indicateurs pouvant être employés pour comparer deux distributions de probabilités, notées $F_1$ et $F_2$. Il est également possible de construire son propre critère (section 2.5.2.3). Ces indicateurs prennent des valeurs réelles comprises entre 0 et 1. Dans l'étude de l'influence d'un facteur donné, l'interprétation est a priori simple : plus un indicateur est proche de zéro, plus les distributions coïncident et moins le facteur étudié est influent. À l'inverse, plus il est proche de 1, plus le facteur est influent.

#### 2.5.2.1 Distance de Kolmogorov-Smirnov

La distance de Kolmogorov-Smirnov entre les distributions $F_1$ et $F_2$ est la distance maximale entre leurs fonctions de répartition $R_1$ et $R_2$ :

$$D_{\mathrm{KS}}\left(F_1, F_2\right) = \max_{\lambda \in [0,1]} \left|R_1\left(\lambda\right) - R_2\left(\lambda\right)\right|.$$

Cette distance est habituellement utilisée pour tester si une distribution empirique est compatible avec une distribution théorique supposée (test de Kolmogorov-Smirnov).

#### 2.5.2.2 Distance de Hellinger

Le carré de la distance de Hellinger entre deux distributions $F_1$ et $F_2$ de densités respectives $f_1$ et $f_2$ vaut :

$$H^2\left(F_1, F_2\right) := 1 - \int_0^1 \sqrt{f_1\left(\lambda\right) f_2\left(\lambda\right)}\, \mathrm{d}\lambda.$$

#### 2.5.2.3 Comparaison des intervalles de confiances

Comme l'intervalle de confiance est une quantité importante pour juger de la pertinence de la prédiction du modèle, nous proposons le critère ci-dessous, qui compare deux intervalles de confiance $I_1$ et $I_2$ des distributions de probabilité $F_1$ et $F_2$. Il est basé sur les longueurs relatives de l'intersection des deux intervalles et des intervalles eux-mêmes et défini ainsi :

$$D\left(I_1, I_2\right) := 1 - \frac{\left|I_1 \cap I_2\right|}{\max\left(\left|I_1\right|, \left|I_2\right|\right)}$$

où $|I|$ désigne la longueur de l'intervalle $I$.

### 2.5.3 Exemple théorique

Dans cette section, nous étudions l'application de ces trois critères à un cas concret pour mettre en évidence leurs différences. Pour cela, considérons l'exemple théorique de dimension 2 suivant : $y(x_1, x_2) = x_2^2$. Les facteurs $x_1$ et $x_2$ sont indépendants et identiquement distribués selon la loi normale centrée réduite restreinte à l'intervalle $X := [-2, 2]$. La contrainte est que la réponse doit être inférieure à 2, c'est-à-dire que $A := ]2, +\infty]$. Nous observons 100 points tirés uniformément dans l'espace des facteurs. Le risque de défaillance réel vaut approximativement $1.171 \times 10^{-1}$. Les intervalles de confiance calculés dans cet exemple sont de confiance égale à 90%.

Plaçons nous sous les hypothèses du modèle de Student décrites dans la section 1.4.2, avec la fonction de corrélation de Mockus (section 1.6.1.1). Nous avons retenu ici une fonction de corrélation anisotrope, caractéristique nécessaire pour pouvoir étudier l'influence des facteurs. De plus, cette fonction n'a pas de paramètre à fixer a priori. Les modèles conditionnés étudiés dans cette section partagent donc le même prior. Pour chaque étude d'influence, les modèles a posteriori sont identiques pour chacun des trois critères.

Dans un premier temps, regardons la distribution du risque de défaillance, $\mathcal{R}$ :

| Moyenne | Variance | Borne minimale | Borne maximale |
|---|---|---|---|
| $1.211 \times 10^{-1}$ | $2.430 \times 10^{-4}$ | $9.619 \times 10^{-2}$ | $1.467 \times 10^{-1}$ |

Ensuite, calculons les trois critères pour chacun des facteurs :

| Critères de comparaison | $x_1$ | $x_2$ |
|---|---|---|
| Distance de Kolmogorov-Smirnov | $2.93 \times 10^{-2}$ | $5.51 \times 10^{-1}$ |
| Distance de Hellinger | $6.93 \times 10^{-2}$ | $7.85 \times 10^{-1}$ |
| Comparaison des intervalles de confiances | $5.55 \times 10^{-2}$ | $9.31 \times 10^{-1}$ |

Comme attendu, leurs valeurs sont plus importantes pour le deuxième facteur que pour le premier. Cette différence significative entre les deux facteurs peut permettre de conclure que le premier facteur n'est pas influent. Dans cet exemple, la méthode de comparaison des intervalles de confiance est la plus discriminante.

# Chapitre 3

# Applications pratiques

La mise en œuvre des méthodes, présentées dans les précédents chapitres, peut s'avérer complexe pour une population d'utilisateurs non initiés aux méthodes statistiques avancées. Cette complexité peut notamment conduire à des erreurs d'implémentations qui, si elles ne sont pas remarquées, peuvent mener à de mauvaises conclusions.

Une automatisation de la mise en pratique de ces méthodes, via un logiciel, rend celles-ci accessibles à des utilisateurs non spécialement formés aux mathématiques appliquées et de limiter les sources d'erreurs.

Ainsi, dans ce chapitre, est décrit le logiciel GoNoGo (section 3.1). Réalisé durant la préparation de cette thèse, il intègre toutes les méthodes mathématiques décrites dans ce mémoire, permettant ainsi à son utilisateur le calcul des risques de défaillance et d'insuffisance pour un produit déterminé. Il met en oeuvre l'algorithme décrit dans la section 2.3.

Nous illustrons plusieurs applications possibles de ce logiciel sur un cas réel provenant de la société STMicroelectronics (section 3.2) : *un diplexeur*. Son risque réel n'étant pas connu, ce produit ne nous permet pas de valider nos méthodes. Néanmoins, cet exemple met en évidence un vaste champ d'applications pour les différents acteurs industriels qui interviennent dans la conception d'un produit.

## 3.1 Description du logiciel GoNoGo

Il est judicieux dans un contexte industriel, d'automatiser autant que faire se peut le processus, relativement complexe, de calcul des risques afin de limiter les sources d'erreurs. Nous avons ainsi mis en œuvre l'ensemble des méthodes et modèles, décrits dans les chapitres précédents, dans un logiciel nommé GoNoGo. Comme celui-ci a pour but de fournir des outils nécessaires à la prise de décision, son nom fait référence à l'étape de réflexion précédant une décision de validation ("Go !") ou d'invalidation ("No go !") de la mise en production.

### 3.1.1 Cœur de calcul

La programmation des fonctionnalités mathématiques de ce logiciel a été réalisée dans le langage Matlab (version R2012.a) pour deux raisons principales. Premièrement, les données peuvent le plus souvent être représentées par des vecteurs. Or le langage Matlab permet la manipulation directe de grandeurs vectorielles, ce qui permet d'optimiser le traitement numérique (en mémoire et en temps de calcul). Deuxièmement, ce langage de programmation est muni d'un nombre important de fonctions mathématiques déjà implémentées, autorisant la une mise en œuvre rapide de nos méthodes. En effet, nous avons utilisé en complément du logiciel (de base) Matlab, trois boîtes à outils, l'*optimization toolbox*, la *statistics toolbox* et la *Global Optimization Toolbox* fournies par MathWorks.

### 3.1.2 Interface utilisateur

Pour l'interface du logiciel, nous utilisons le langage Tcl (*Tool command language*) couplé avec l'extension Tk qui fournit une bibliothèque de fonctionnalités graphiques.

L'utilisateur y renseigne les données de son problème, en définissant les facteurs et la réponse (voir figure 3.1). Pour cela, il saisit les intervalles de définition $X_i$ de chacun des $D$ facteurs, permettant ainsi de considérer

$$X := X_1 \times X_2 \times \ldots \times X_D.$$

En pratique, on note que seules les distributions marginales $P_i$ des facteurs sont accessibles via le suivi du processus industriel de fabrication, définissant donc :

$$P := P_1 \otimes P_2 \otimes \ldots \otimes P_D.$$

Dans GoNoGo, la modélisation d'une réponse multidimensionnelle $y = \left(y^{(1)}, y^{(2)}, \ldots, y^{(d)}\right)$ ne s'effectue pas par un unique modèle aléatoire, mais par autant de modèles aléatoires (construits indépendamment les uns des autres) qu'il y a de composantes dans la réponse. Dans la suite de ce chapitre, nous nommerons *réponses* les composantes de la réponse. Les spécifications du produit portant sur chacune des réponses, cela nous amène à considérer que :

$$y \in A \quad \text{si et seulement si} \quad \left(y^{(1)} \in A_1 \text{ ou } y^{(2)} \in A_2 \text{ ou } \ldots \text{ ou } y^{(d)} \in A_d\right).$$

Toutes ces données (distributions marginales des facteurs et spécifications) sont également saisies par l'utilisateur. Pour chaque réponse, un modèle aléatoire peut ensuite être sélectionné ainsi qu'une fonction de corrélation et une transformation.

L'utilisateur importe ensuite les valeurs des facteurs et des réponses pour les $n$ données du problème dans un tableau (voir figure 3.2). En pratique, on note que la méthode Monte-Carlo utilisée est itérative, sa convergence étant testée sur les deux premiers moments de la distribution de la probabilité de défaillance, c'est-à-dire la variable aléatoire $x \mapsto \mathbb{P}(Y_x \in A)$ définie sur $(X, P)$. Cette méthode requiert ainsi un certain nombre de paramètres comme la taille des échantillons tirés, un seuil en deçà duquel le tirage est considéré convergent ou encore le nombre d'itérations maximal (garde fou). Une fois ces données saisies, l'utilisateur peut lancer le calcul des risques de défaillance et d'insuffisance.

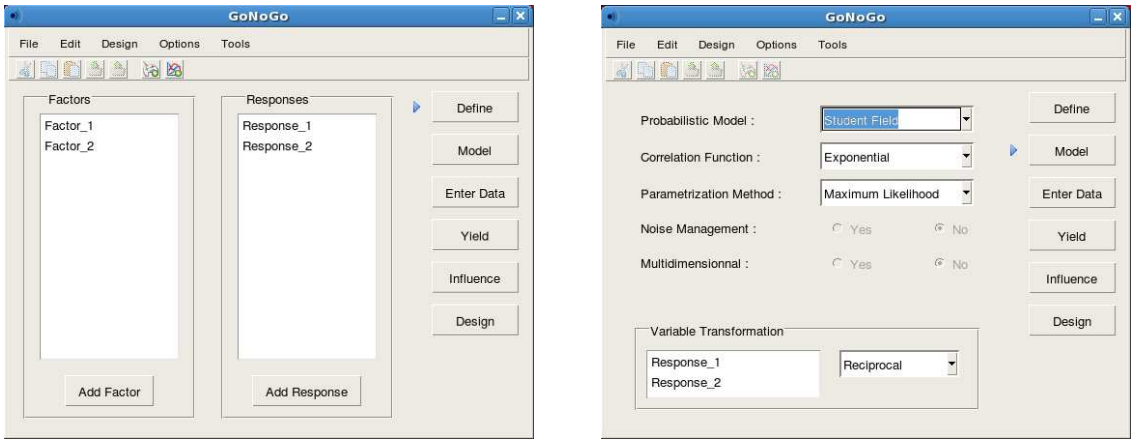



FIGURE 3.1 – Définition des facteurs et des réponses & sélection d'un modèle aléatoire.

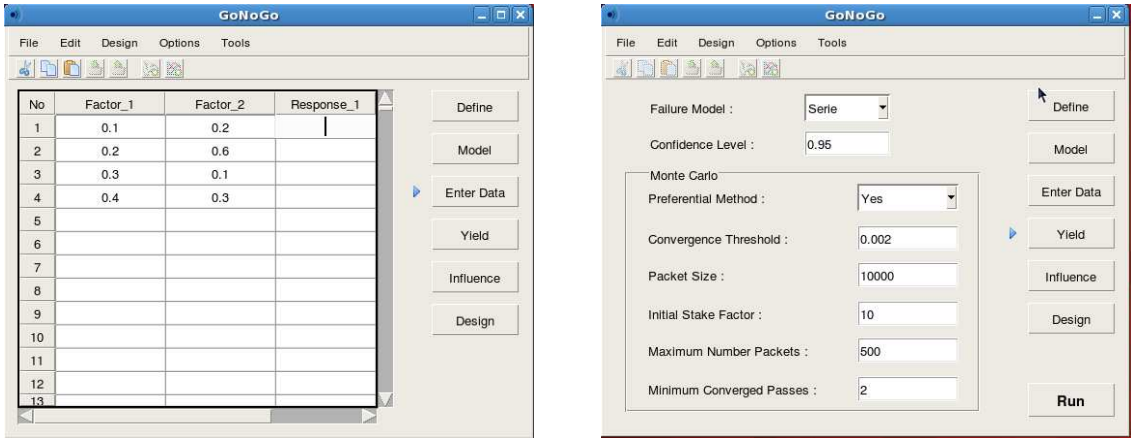



FIGURE 3.2 – Saisie des données du problème & lancement du calcul de rendement.

L'utilisateur peut ensuite charger les résultats dans le logiciel de visualisation de données EZwave (de l'éditeur Mentor Graphics), voir figure 3.3. Il peut ainsi visualiser les distributions marginales des facteurs issues du tirage Monte-Carlo, la convergence de la simulation sur la moyenne et la variance de la probabilité de défaillance, ainsi que les distributions des risques de défaillance et d'insuffisance du modèle.

L'utilisateur peut également construire un plan d'expériences pour l'acquisition de nouvelles données ou encore étudier l'influence des facteurs (figure 3.4).

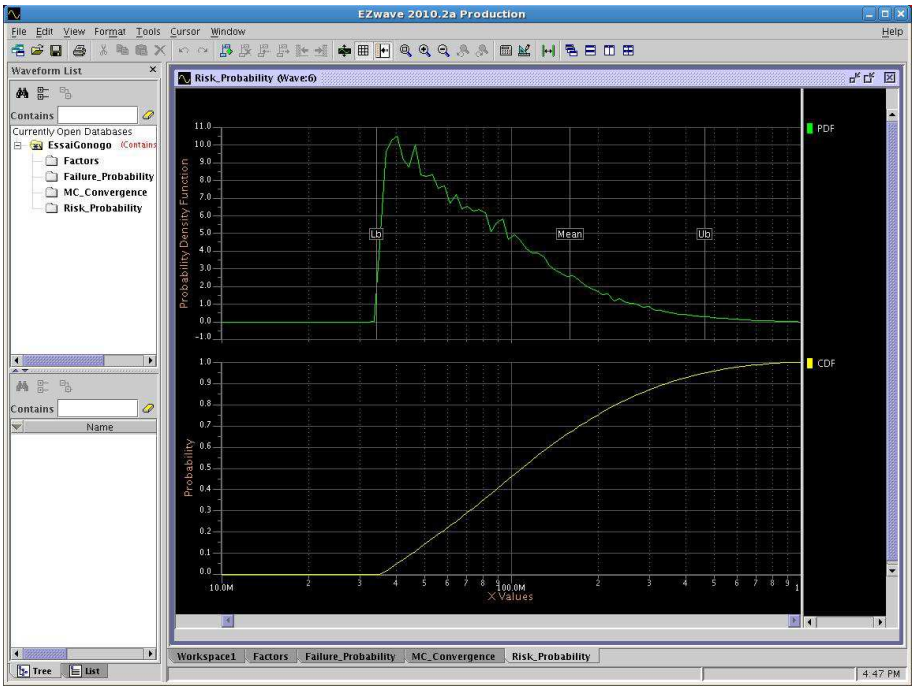



FIGURE 3.3 – Visualisation des résultats via EZwave (ici la densité de probabilité du risque de défaillance).

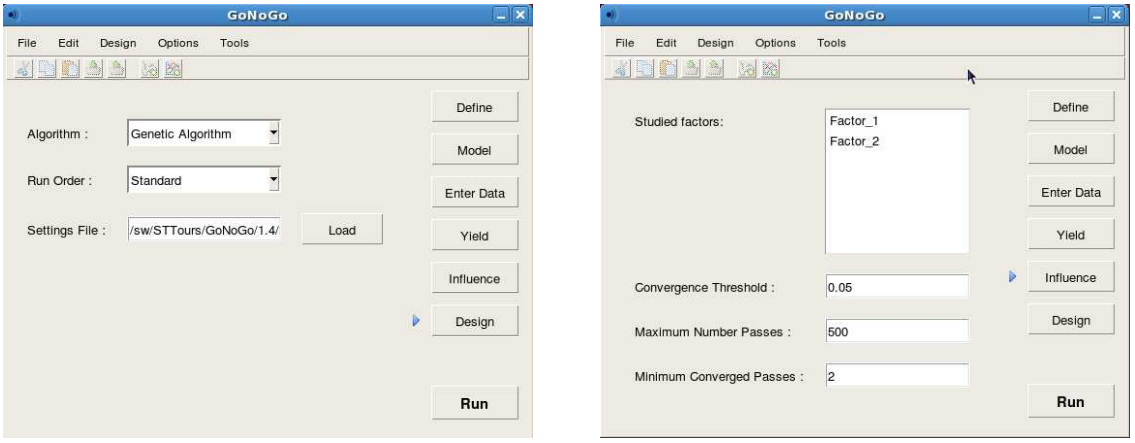



FIGURE 3.4 – Création d'un plan d'expériences & étude de l'influence des facteurs.

## 3.2 Étude d'un cas réel de la société STMicroelectronics

Avant de présenter certaines applications possibles du logiciel GoNoGo, il est important d'identifier les principaux acteurs qui participent à l'élaboration d'un produit. Pour commencer, il y a l'ingénieur commercial ("marketing engineer") qui est en contact avec le client potentiel. Il connaît les besoins de celui-ci, les exigences non-discutables et les souhaits négociables. Il détermine le cahier des charges (les spécifications) du produit. Puis, intervient l'ingénieur en charge de la conception du produit ("design engineer"). En s'appuyant sur les informations fournies par le service Recherche et Développement, il sait quelles sont les caractéristiques physiques réalisables, ajustables ou non. A l'autre bout de la chaîne, il y a l'ingénieur de production ("manufacturing engineer"), qui connaît l'état actuel des moyens (humains, machines...) de production et leurs possibilités d'adaptation aux besoins de production. Nous allons montrer qu'il est possible à ces trois acteurs d'utiliser la prédiction du risque de défaillance comme un outil supplémentaire à ceux déjà en leur possession pour optimiser la fabrication d'un produit.

### 3.2.1 Diplexeur

Un diplexeur (de l'anglais "diplexer") est un dispositif passif qui combine ou sépare des signaux occupant des bandes de fréquences distinctes. Chaque voie (voir figure 3.5) comporte un filtre passe-bande, les bandes passantes ne se recouvrant pas.

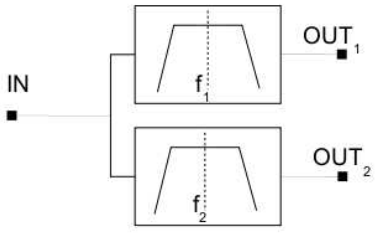



FIGURE 3.5 – Schéma d'un diplexeur.

C'est une fonction électronique que l'ont trouve très souvent dans les circuits de télécommunications. Elle permet par exemple à des modules d'émission et de réception de partager le même canal de communication. STMicroelectronics couvrant le marché de la téléphonie mobile, le diplexeur étudié fonctionne aux radiofréquences (f $\sim$ 1 GHz).

Le filtre est réalisé en utilisant les procédés de fabrication propres à la micro-électronique. C'est une suite d'opérations nombreuses et complexes : croissance de couches métalliques par électrolyse, dépôt de diélectriques organiques, gravure par plasma, photolithographie... La technologie qui en résulte (nommée RLC06A) peut-être vue comme un empilement de couches isolantes et conductrices (voir figure 3.6).

Les paramètres sujets à des variations qui sont susceptibles d'impacter la réponse du filtre sont typiquement les épaisseurs des dépôts. Considérons par exemple la couche de cuivre (notée Meta2) qui est obtenue par une réaction électrochimique : la composition du bain n'est pas totalement maîtrisée et varie dans le temps (du fait des réactions dont il est le siège). Localement, les espèces réactives peuvent venir à manquer si le brassage est

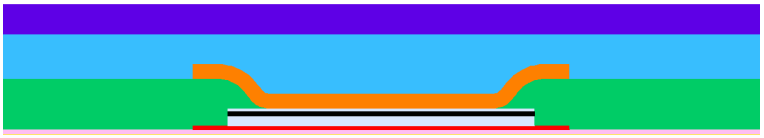


FIGURE 3.6 – Vue en coupe de la technologie RLC06A dont les paramètres d'intérêt sont : Capa2, Meta2, BCB1, Meta1.

insuffisant. Tout ceci impacte l'épaisseur finale du dépôt. On comprend bien, vu le nombre de phénomènes influents, que l'épaisseur puisse avoir un caractère aléatoire.

Nous avons retenu 4 paramètres (4 épaisseurs) qui ont un impact plus ou moins important sur les performances du diplexeur.

Dans la figure 3.7 est représentée la vue de dessus de ce diplexeur, telle qu'elle apparaît dans l'environnement de conception micro-électronique Cadence qui propose, entre autres, un logiciel de dessin technique (Virtuoso).

Les spécifications officielles du filtre, au nombre de 12, sont indiquées dans la figure 3.8. Nous étudions ainsi 12 réponses.

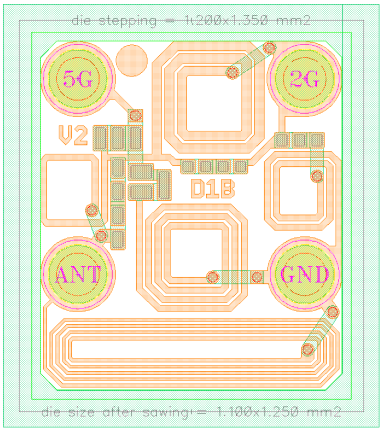



FIGURE 3.7 – Modélisation 2D du diplexeur.

| Symbol | Parameter | Test condition | Value | | | Unit |
|---|---|---|---|---|---|---|
| | | | Min. | Typ. | Max. | |
| **Pass band** | | | | | | |
| f | 2 G band pass | | 2400 | | 2483.5 | MHz |
| | 5 G band pass | | 4900 | | 5850 | MHz |
| Z | Nominal impedance | | | 50 | | Ω |
| Return loss | | All ports | | | -17 | dB |
| S21 | 2 G to antenna insertion loss | 2400 to 2483.5 MHz | | 0.6 | 0.7 | dB |
| S31 | 5 G to antenna insertion loss | 4900 to 5850 MHz | | 0.6 | 0.7 | dB |
| **Attenuation** | | | | | | |
| S21 | 2 G to antenna attenuation | 4900 to 5850 MHz | 20 | | | dB |
| S31 | 5 G to antenna attenuation | 2400 to 2483.5 MHz | 18 | | | dB |
| **Out of band attenuation** | | | | | | |
| S21 | 2 G to antenna attenuation | 5850 to 7000 MHz | 15 | | | dB |
| | | 7000 to 9500 MHz | 9 | | | |
| | | 9800 to 10500 MHz | 16 | | | |
| S31 | 5 G to antenna attenuation | 9800 to 11650 MHz | 11 | | | dB |

FIGURE 3.8 – Spécifications du diplexeur.

### 3.2.2 HFSS

Nous établissons une relation entre les paramètres "process" et les caractéristiques électriques du diplexeur à l'aide d'un simulateur physique. Le logiciel commercial HFSS ("High Frequency Structural Simulator"), de la société Ansys, permet la résolution d'un système d'équations intégro-différentielles, les équations de Maxwell qui régissent l'interaction entre les charges électriques (fixes, mobiles) et le champ électromagnétique. Ce type d'outil est extrêmement complexe, il 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 dispose d'un module de représentation 3D qui permet la génération d'un modèle géométrique fidèle (figure 3.9). Notez que ce modèle est paramétré, les variables étant les épaisseurs, les conductivités et permittivités 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. Il s'agit donc, en quelque sorte, d'une usine virtuelle.

La simulation physique de ce modèle nécessite 60Go de mémoire, 15 heures d'exécution avec 6 processeurs ("threads") soit deux jours de temps CPU. Une étude Monte-Carlo brute est, par conséquent, si ce n'est inconcevable, du moins peu réaliste sans des moyens informatiques de traitement très importants.

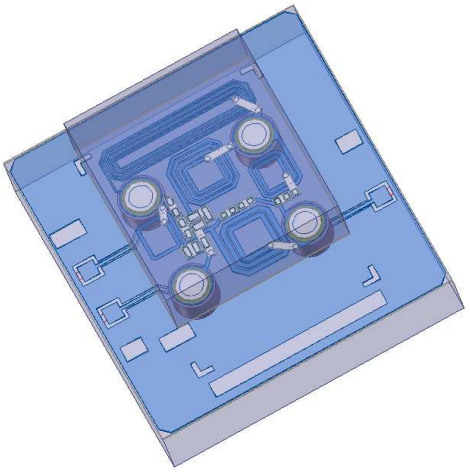


FIGURE 3.9 – Modélisation 3D du diplexeur.

### 3.2.3 CornerStone et GoNoGo

Le logiciel habituellement utilisé par les ingénieurs de STMicroelectronics pour construire des modèles de régression se nomme CornerStone (version 5.3), de la société Domain Manufacturing Corporation. Ce logiciel génère à partir d'un ensemble de données, un modèle polynomial par la méthode dite des surfaces de réponses, intimement liée à la théorie des plans d'expériences (voir Atkinson et Donev (1992)). Comme nous l'avons indiqué dans la section 0.2, l'ordre maximal du polynôme accessible dépend fortement du nombre de données fournies. Les résidus (écarts entre les prédictions et les observations) y sont supposés être indépendants les uns des autres et suivre une distribution normale de moyenne nulle et de variance constante inconnue (que le logiciel va se charger d'estimer à partir des données avec la méthode MLE). Cette supposition est difficilement justifiable dans les cas, comme le notre, où les données étudiées proviennent de simulations numériques, reproductibles par nature. Il est possible dans GoNoGo d'importer un modèle polynomial provenant de CornerStone pour calculer ensuite le risque de défaillance. Il est donc aisé de comparer les résultats obtenus d'après notre modèle basé sur un champ aléatoire à ceux obtenus par cette méthode qui fait lieu de référence dans le monde industriel.

Ainsi, dans la section 3.2.4, nous comparons un modèle polynomial (avec une composante aléatoire sur les résidus) provenant de CornerStone et le modèle aléatoire de Student (section 1.4.2) couplé avec la fonction de corrélation $\gamma$-exponentielle anisotrope (section 1.6.1.1) paramétrée avec la méthode du maximum de vraisemblance (section 1.6.2). Ce modèle aléatoire est également utilisé pour les études menées dans les sections 3.2.6 et 3.2.7.

Comme les spécifications portent sur chacune des 12 réponses individuellement, nous retenons l'estimation de la probabilité de défaillance de l'équation (1.5) de la section 1.8.3.

De plus, pour la réalisation des trois études qui vont suivre, les données provenant d'HFSS ont été choisies en fonction de plusieurs plans d'expériences explicités dans la figure 3.1. Remarquons que ces plans sont ordonnés. On parle de plans fractionnaires. Par exemple, le plan composite central peut être vu comme l'union du plan en étoile et du plan

factoriel à deux niveaux. Les expériences déjà réalisées sont disponibles pour un plan plus complet, permettant de renseigner un modèle plus complexe.

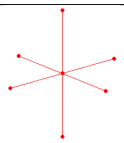
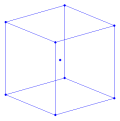
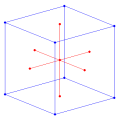
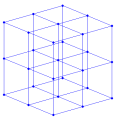
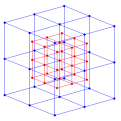
| Plan | étoile | factoriel à 2 niveau | composite centré | factoriel à 3 niveau | factoriels imbriqués |
|---|---|---|---|---|---|
| Structure |  |  |  |  |  |
| Nombre de données | 9 | 17 | 25 | 81 | 162 |
| Modèles polynomiaux | linéaire sans interaction | linéaire | quadratique | quadratique | cubique |
| Nombre de termes | 5 | 11 | 15 | 15 | 127 |

TABLE 3.1 – Séquence de plans d'expériences.

### 3.2.4 Étude "marketing/design" : choix des spécifications

GoNoGo permet de mettre en adéquation les spécifications annoncées du produit avec un rendement de production considéré comme acceptable. Ce genre d'étude est intéressant car il peut servir de base de négociation dans un cadre commercial. Lors de la phase de conception, on peut ajuster au mieux les performances annoncées du produit développé en intégrant les contraintes de fabrication. On évite ainsi une sous-évaluation des performances réelles, cette sous-évaluation résultant des marges prises par les équipes de développement afin de pallier au manque d'information.

Pour illustrer cette application, considérons, par exemple, le risque de défaillance du diplexeur en fonction des spécifications de l'atténuation (voir tableau 3.8) dans la bande opposée (2G et 5G). Les spécifications des autres réponses restent fixées à leurs valeurs initiales.

Ainsi, sur la figure 3.10, on observe que pour la réponse *5G attenuation @ 2G* (en abscisse), la limite inférieure associée peut être augmentée de 18dB à 20dB sans dégrader le risque de défaillance (détérioration de l'ordre de 500ppm). En revanche, la réponse *2G attenuation @ 5G* (en ordonnée) est beaucoup plus sensible ; modifier la spécification, ne serait-ce que d'un dB, détériore significativement le rendement, augmentant le risque de défaillance de près de 70%. L'augmenter de 0.1dB se traduit par une perte d'environ 1.5%, ce qui trop important.

La démarche présentée peut facilement être généralisée. Ainsi, il est possible de proposer un nouveau jeu de spécifications (voir figure 3.11) respectant les conditions suivantes :

- un risque de défaillance d'environ 2.5%,
- un intervalle de confiance (à 95%) de longueur 0.05.

On observe que trois spécifications (en vert) peuvent être améliorées sans trop dégrader le risque de défaillance, mais qu'il est nécessaire que la spécification relative à l'atténuation dans la bande de fréquence $[7, 9.5]$ GHz (en orange) soit moins contraignante. En effet,

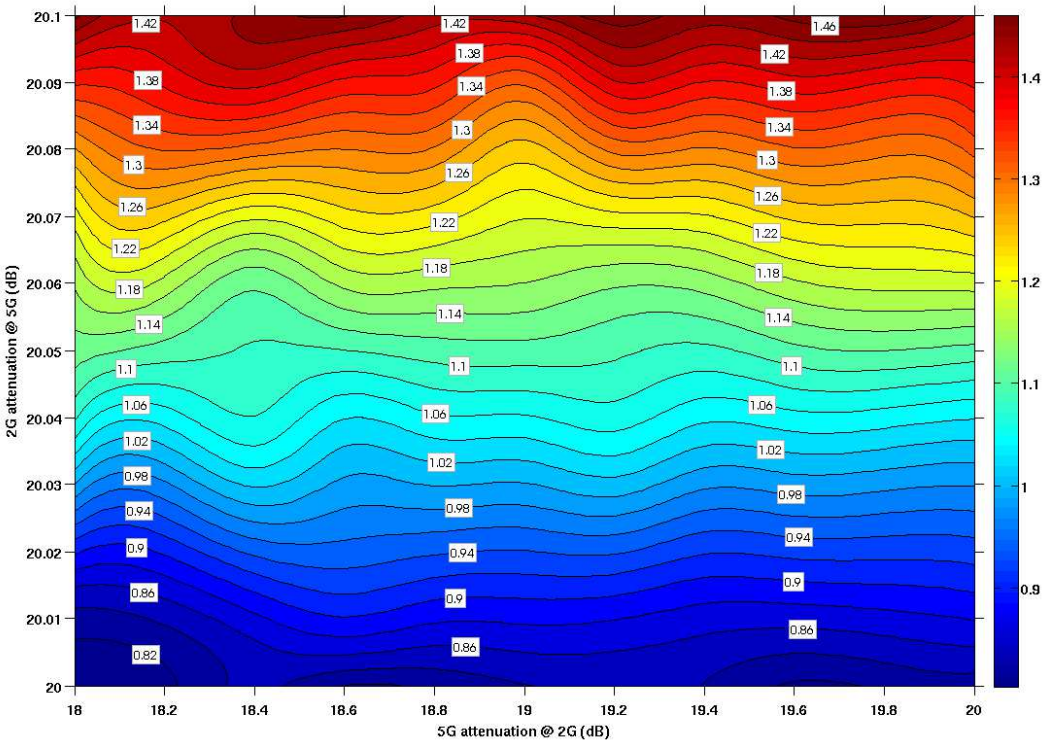



FIGURE 3.10 – Moyenne du risque de défaillance (en %) en fonction des deux spécifications étudiées.

ne serait-ce qu'en considérant les données brutes obtenues via le logiciel HFSS, il apparaît qu'aucune d'entre elles ne satisfait cette spécification à sa valeur initiale (9dB).

| Frequency range | Response | | Data Sheet | |
|---|---|---|---|---|
| | | | Old | New |
| [2400,2483.5] MHz (2G) | Return Loss | $S_{11}$ | 17 | 17 |
| | Return Loss | $S_{22}$ | 17 | 17 |
| | Insertion Loss | $S_{21}$ | 0.7 | 0.6 |
| | 5G Attenuation | $S_{31}$ | 18 | 20 |
| [4900,5850] MHz (5G) | Return Loss | $S_{11}$ | 17 | 17 |
| | Return Loss | $S_{33}$ | 17 | 17 |
| | Insertion Loss | $S_{31}$ | 0.7 | 0.6 |
| | 2G Attenuation | $S_{21}$ | 20 | 20 |
| [5850,7000] MHz | 2G Attenuation | $S_{21}$ | 15 | 15 |
| [7000,9500] MHz | 2G Attenuation | $S_{21}$ | 9 | 7 |
| [9800, 11650] MHz | 2G Attenuation | $S_{21}$ | 16 | 16 |
| | 5G Attenuation | $S_{31}$ | 11 | 11 |

FIGURE 3.11 – Nouveau jeu de spécifications pour le diplexeur.

Il est également possible sur cet exemple d'analyser la convergence du risque de défaillance selon que le modèle soit de Student ou quadratique (introduit dans la section précédente). Dans la figure 3.13, on observe un intervalle de confiance pour le risque de

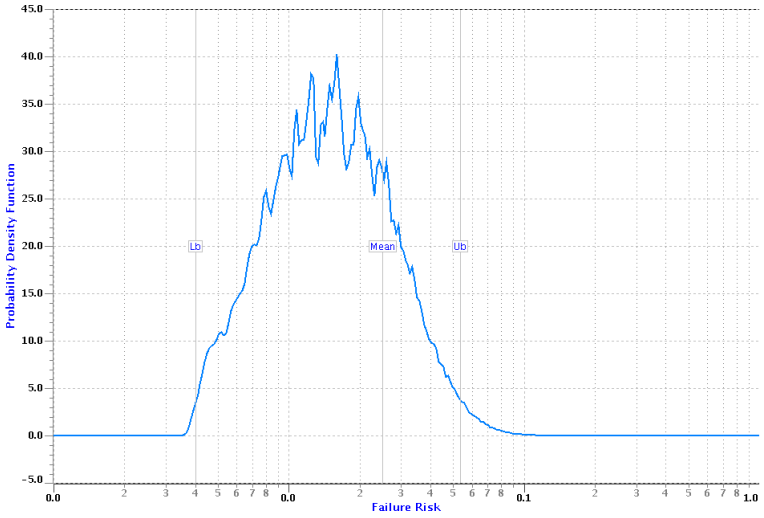



FIGURE 3.12 – Risque de défaillance, pour le nouveau jeu de spécifications, calculé avec le modèle de Student.

défaillance plus resserré avec le modèle polynomial qu'avec le modèle de Student. Cette observation est cohérente avec le fait qu'il y a plus d'hypothèses effectuées avec le premier modèle qu'avec le second. Cependant, ces hypothèses sont difficiles à vérifier en pratique. C'est la raison pour laquelle, on préfère utiliser un modèle aussi simple que possible. En effet, il y a 14 coefficients aléatoires pour le modèle polynomial quadratique ainsi que la variance des résidus qui est inconnue. Le modèle de Student, quant à lui, ne compte que 2 termes aléatoires $U$ et $V$ et 2 coefficients à identifier a priori (ceux de la fonction de corrélation).

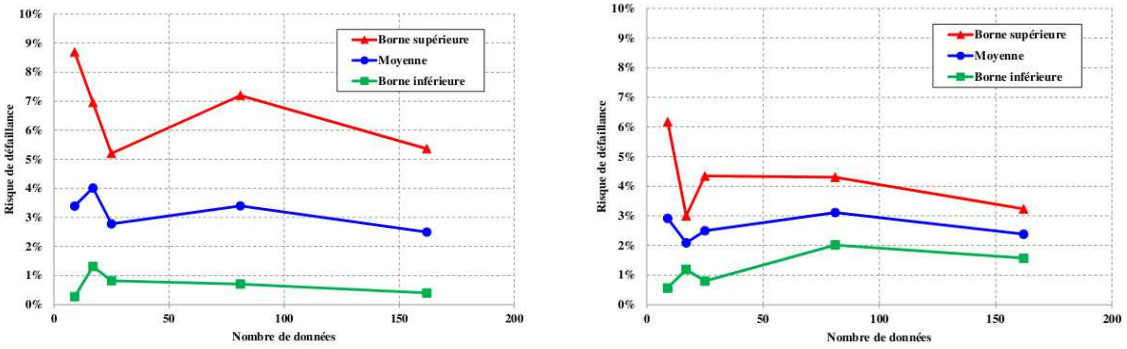



FIGURE 3.13 – Convergence du risque de défaillance avec le modèle de Student (à gauche) et polynomial quadratique (à droite).

### 3.2.5 Sélection de transformations pour les réponses

Nous illustrons, dans cette section, l'utilisation de l'indicateur NRMS ("Normalized Root Mean Squared"), introduit, dans la section 1.8.2, en sélectionnant des transformations pour les 12 réponses caractérisant le diplexeur. Pour cela, nous considérons les 9 données correspondant au plan d'expériences en étoile ainsi que la liste de transformations suivante : $\log(y)$, $y^{-1}$, $\sqrt{y}$, $\sinh(y)$, $y^2$, $10^y$, $1/\sqrt{y}$. Nous retenons les transformations présentant le meilleur NRMS, voir le tableau 3.2, le modèle considéré étant le champ aléatoire de Student introduit dans la section 3.2.3.

| réponses | transformations retenues |
|---|---|
| S11_2G | $\log(y)$ |
| S21_2G | $\sinh(y)$ |
| S22_2G | $1/\sqrt{y}$ |
| S31_2G | $y^2$ |
| S11_5G | $\log_{10}(y)$ |
| S21_5G | $y^{-1}$ |
| S31_5G | $1/\sqrt{y}$ |
| S33_5G | $\sinh(y)$ |
| S21_oB1 | $y$ |
| S21_oB2 | $y$ |
| S21_oB3 | $y^2$ |
| S31_oB3 | $\log(y)$ |

TABLE 3.2 – Transformations sélectionnées à l'aide de l'indicateur NRMS.

Calculons ensuite les risques de défaillance dans les deux cas suivants : réponses transformées ou non. Pour cela, nous avons retenu les nouvelles spécifications, définies dans la section précédente. Dans le tableau 3.3, sont inscrites les valeurs des moyennes et des bornes des intervalles de confiance à 0.9.

| | moyenne | borne inférieure | borne supérieure |
|---|---|---|---|
| avec transformations | $4.42 \times 10^{-2}$ | $2.66 \times 10^{-3}$ | $1.50 \times 10^{-1}$ |
| sans transformations | $3.39 \times 10^{-2}$ | $2.73 \times 10^{-3}$ | $8.69 \times 10^{-2}$ |

TABLE 3.3 – Moyennes et bornes des intervalles de confiance à 0.9 des risques de défaillances avec ou sans transformation sur les réponses.

Bien que l'estimation de chaque composante de la fonction $y$ ait été améliorée par l'utilisation de transformations, nous n'observons pas d'amélioration significative de la qualité de l'estimation du risque de défaillance (l'intervalle de confiance ne s'est pas réduit). De fait, nous pourrions estimer qu'il faut plutôt choisir des transformations telles que le risque de défaillance soit mieux estimé c'est-à-dire telles que l'intervalle de confiance soit petit. Mais cela n'est pas forcément pertinent car il faut alors calculer le risque de

défaillance pour chaque combinaison de transformations possibles, ce qui entraîne une plus grande complexité.

### 3.2.6 Étude ingénierie de test : influence des réponses

À la fin de la chaîne de production, des tests sont effectués pour détecter des anomalies de fabrication sur chacun des lots de pièces produites. De part leur coût, il est important d'optimiser ces tests, c'est-à-dire d'en limiter le nombre.

Évaluer la contribution de chacune des réponses au risque de défaillance peut être très utile à cette optimisation. Pour cela, il suffit de calculer le risque de défaillance pour chacune des réponses, puis pour l'ensemble des 12 réponses. On obtient la figure 3.14. Le terme d'exclusion est la différence entre le risque global et la somme des risques individuels. Il fait référence à la formule du crible de Poincaré ("Inclusion-Exclusion Principle" en anglais). Il correspond aux défaillances simultanées de plusieurs réponses. Si les domaines de défaillance associés à chacune des réponses sont disjoints, alors ce terme est nul. On observe ainsi que la réponse S21_oB1 contribue pour environ 60% au risque de défaillance global. Il est donc pertinent de tester cette réponse en priorité.

Le classement des réponses selon leur contribution au risque de défaillance peut également être utile pour améliorer le produit. On identifie, en effet, la réponse à améliorer en priorité.

On pourrait avoir la même démarche de décomposition pour le risque d'insuffisance ce qui permettrait de déterminer le modèle à améliorer en priorité.

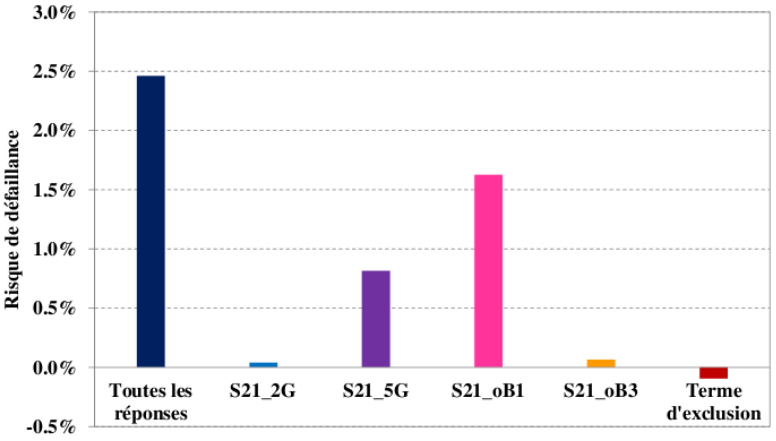



FIGURE 3.14 – Décomposition en contributions au risque de défaillance de chacune des réponses.

### 3.2.7 Étude "manufacturing" : dérive de la distribution des facteurs

La capabilité d'un processus de fabrication est un indicateur mesurant la capacité de celui-ci à réaliser des pièces identiques. Ainsi, seules les distributions marginales des facteurs sont prises en compte ; les réponses n'interviennent pas du tout ici.

Le $C_{pk}$ est un indice de performance standard dans l'industrie. Le $C_{pk}$ d'un facteur vaut :

$$C_{pk} := \min\left(\frac{\mathrm{USL} - \hat{\mu}}{3\hat{\sigma}}, \frac{\hat{\mu} - \mathrm{LSL}}{3\hat{\sigma}}\right)$$

où $\hat{\mu}$ et $\hat{\sigma}$ désignent la moyenne et l'écart-type de la distribution du facteur et l'intervalle $[\mathrm{LSL}, \mathrm{USL}]$ est la plage de valeurs du facteur acceptée par l'usine. Il y a ainsi une valeur de $C_{pk}$ pour la distribution de chacun des facteurs. Les valeurs standards de cet indice, retenues par STMicroelectronics, sont 1.33 et 1.66, correspondant respectivement à 63ppm et 1ppm de pièces rejetées, si l'on suppose que le facteur admet une loi normale.

Durant la production, pour chaque lot produit, on relève les distributions marginales des facteurs qui permettent de calculer les $C_{pk}$. Les rêgles d'acceptation d'un lot sont complexes, mais l'on peut, à titre d'illustration, retenir les critères suivants :

| | |
|---|---|
| $1.66 \leq C_{pk}$ | lot accepté |
| $1.33 \leq C_{pk} \leq 1.66$ | lot soumis à une étude complémentaire |
| $C_{pk} \leq 1.33$ | lot refusé |

L'étude complémentaire nécessite la réunion de tous les acteurs intervenant dans la fabrication du produit : "product", "device", "quality" et "manufacturing". Le calcul du risque de défaillance peut, dans cette situation, aider à statuer rapidement sur le lot anormal.

Pour illustrer cette application, représentons le rendement, complémentaire du risque de défaillance, en fonction des paramètres de position (en abscisse) et d'échelle (en ordonnée) de la distribution gaussienne du paramètre Meta2 restreinte à l'intervalle $[\mathrm{LSL}, \mathrm{USL}]$, voir figure 3.15. Nous représentons également les $C_{pk}$ standards retenus par STMicroelectronics (les deux courbes d'iso-valeurs).

Le $C_{pk}$ se veut être un garant du rendement de fabrication. Cependant, il n'intègre pas les spécifications du produit réalisé : on suppose que toutes les pièces qui sont dans l'intervalle $[\mathrm{LSL}, \mathrm{USL}]$, sont bonnes. L'abaque montre que la situation est un peu plus compliquée que cela. Ainsi, si $\hat{\mu}$ diminue de $3\mathrm{m}\Omega/\square$ à $2.9\mathrm{m}\Omega/\square$, on voit que le rendement augmente bien que le $C_{pk}$ diminue.

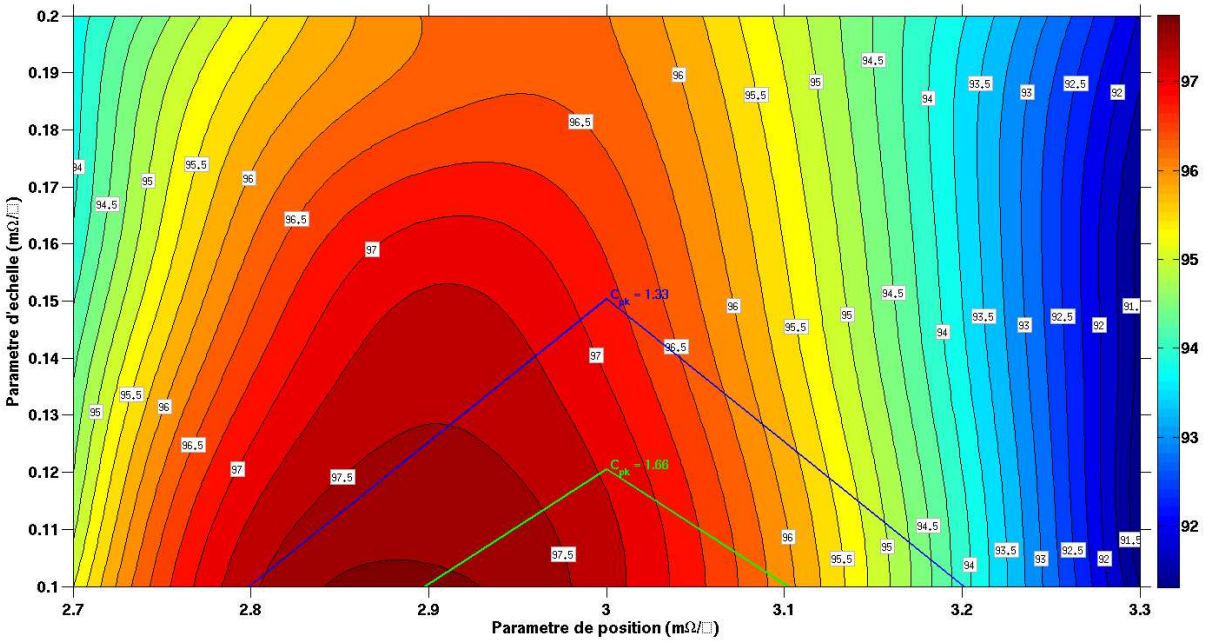



FIGURE 3.15 – Rendement en fonction des paramètres de la distribution de Meta2 ainsi que les standards $C_{pk}$.

# Conclusion

Les travaux menés durant cette thèse ont eu pour objectif de fournir au décideur un outil lui permettant de mesurer la capacité de son processus de fabrication à réaliser un produit satisfaisant une liste de spécifications. La variabilité inhérente à tout processus peut effectivement entraîner des pièces manufacturées non-conformes à un cahier des charges. C'est pourquoi, il est important de prendre en compte cette variabilité et d'en quantifier les conséquences.

Ces travaux sont particulièrement adaptés à "l'expérimentation virtuelle" pour laquelle des simulations numériques remplacent la réalisation de prototypes pour prévoir les performances du produit. Les durées de ces simulations sont souvent prohibitives ; le simulateur numérique est alors considéré comme une fonction "boîte noire" et est approché par un émulateur analytique.

Cette thèse est structurée autour de trois axes :

1. Construction d'un modèle aléatoire,
2. Définition des risques de défaillance du produit et d'insuffisance du modèle,
3. Mise en œuvre pratique de l'ensemble de ces méthodes dans un logiciel nommé GoNoGo.

Sa réalisation s'est effectuée sous la contrainte suivante : peu de données sont disponibles pour représenter le phénomène physique étudié en raison de la complexité du processus d'acquisition de données ou en raison de la dimension de l'espace des facteurs étudiés.

Afin de tenir compte de l'incertitude due à cette connaissance incomplète de la fonction d'étude, quatre modèles ont été décrits, illustrant l'application de l'inférence bayésienne à la construction d'un champ aléatoire.

Ainsi, dans le but de pallier au manque d'information, différents niveaux d'aléa (subis ou choisis) on été pris en compte, nous menant à des concepts originaux : les risques de défaillance du produit (2.3) d'insuffisance du modèle (2.8). La distribution du risque de défaillance est une mesure directe de l'impact de la variabilité du procédé de fabrication sur les performances des produits, d'une manière qui reflète les diverses sources d'incertitude dont la connaissance incomplète de la fonction "boîte noire", les valeurs des variables du modèle et les approximations numériques. La distribution d'insuffisance du modèle permet d'apprécier la pertinence de l'estimateur du risque de défaillance vu le niveau de connaissance de la fonction étudiée.

Le logiciel GoNoGo, où sont implémentées les diverses méthodes mathématiques décrites dans cette thèse, a été validé sur des exemples théoriques et sur un exemple réel

provenant de la société STMicroelectronics.

Cette thèse ouvre de nouvelles perspectives pour l'application des champs aléatoires à l'analyse des risques d'ingénierie et nous pouvons d'ores et déjà donner des pistes pour des recherches ultérieures :

- Une justification théorique complète pour l'application de la méthode du maximum de vraisemblance dans le cadre de la régression bayésienne doit être apportée.
- Dans notre étude théorique des champs aléatoires, nous avons systématiquement exclu les données conduisant à des distributions dégénérées. Il y a là un champ d'investigation possible.
- Il serait intéressant de comprendre les liens entre les résultats obtenus par la méthode du maximum de vraisemblance et les modes des distributions a posteriori des paramètres aléatoire des modèles (voir le tableau 1.5).
- Des critères plus restrictifs permettant de fixer la distribution $h$ des seuils $\alpha$ et $\gamma$ dans l'étude du risque d'insuffisance du modèle (section 2.2) seraient utiles.
- Observons que, sous le formalisme de la section 2.2, la distribution bêta constituant le risque d'insuffisance est une des marginales d'une distribution de Dirichlet de paramètres $(n(\alpha)+1, n(\alpha,\gamma)+1, n(\gamma)+1)$ avec

$$\begin{aligned} n(\alpha) &:= \operatorname{card}\{m \in [1,2,\ldots,M] \mid \mathcal{M}(x) \leq \alpha\}, \\ n(\alpha,\gamma) &:= \operatorname{card}\{m \in [1,2,\ldots,M] \mid \alpha < \mathcal{M}(x) \leq \gamma\}, \\ n(\gamma) &:= \operatorname{card}\{m \in [1,2,\ldots,M] \mid \gamma < \mathcal{M}(x)\}. \end{aligned}$$

  Dans le cadre d'un modèle trinomial, il devrait être possible d'envisager une unique distribution décrivant le risque de défaillance et d'insuffisance.
- Pour la construction d'un plan d'expériences, il nous faut trouver une méthode d'optimisation multi-modale efficace afin d'identifier les maxima locaux de l'entropie. Les minima étant connus, cela peut servir de piste de recherche.
- Dans la pratique, il est nécessaire de maîtriser tous les calculs et approximations numériques réalisés dans GoNoGo. Un problème particulièrement difficile est l'inversion d'une matrice définie positive lorsqu'elle est proche d'être singulière. Ce point doit être relié à l'étude des situations dégénérées.
- Il pourrait être intéressant d'étudier la complexité des calculs réalisés dans GoNoGo et d'optimiser leur implémentation via, par exemple, des méthodes de calcul parallèle.

# Annexes

# Annexe A

# Démonstration du lemme 4

*Démonstration.* Premièrement, nous notons que l'égalité

$$\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1} - \frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1} - \boldsymbol{a}^{-1} = 0 \tag{A.1}$$

est vraie puisqu'en multipliant chacun de ses membres par $\boldsymbol{a}$, on trouve l'égalité, évidente :

$$\begin{aligned}\boldsymbol{a}\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1} &= \frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1} + \mathrm{Id}\\ &= \left(\frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right) + \boldsymbol{S}^T\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}.\end{aligned}$$

En multipliant les termes de l'égalité (A.1) par $\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)$ à gauche et par $\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)$ à droite, on obtient l'égalité suivante :

$$\begin{aligned}&\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\\ &\quad - \frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\\ &\qquad - \left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right) = 0.\end{aligned}$$

Les expressions suivantes sont équivalentes à la précédente.

$$\begin{aligned}&\frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\\ &\quad - \frac{1}{\left(1-\boldsymbol{v}\boldsymbol{v}^T\right)^2}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\\ &\qquad - \frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right) = 0,\end{aligned}$$

$$\begin{aligned}&\left(1 - \frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\right)\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\\ &\qquad \left(\frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right) + \boldsymbol{S}^T\boldsymbol{S}\right) = \boldsymbol{s}-\boldsymbol{v}\boldsymbol{S},\end{aligned}$$

$$\left(1-\frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\right)\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}=\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1},$$

$$\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}=\frac{\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}}{1-\frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T}.$$

Nous obtenons ainsi le premier résultat du lemme 4.

Considérons maintenant l'expression :

$$d:=\frac{\boldsymbol{w}\boldsymbol{w}^T-\left(\boldsymbol{w}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{w}\boldsymbol{S}\right)^T}{\frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(1-\frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\right)}-\left(\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{w}\boldsymbol{S}\right)^T\right)^2.$$

Nous avons

$$\begin{aligned}\left(\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{w}\boldsymbol{S}\right)^T\right)^2 &\\ &=\left(\boldsymbol{w}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{w}\boldsymbol{S}\right)^T\\ &=\left(\boldsymbol{w}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\frac{\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}}{1-\frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T}\left(\boldsymbol{w}\boldsymbol{S}\right)^T\end{aligned}$$

en utilisant le premier résultat du lemme 4. Nous obtenons :

$$d=\frac{\left(1-\boldsymbol{v}\boldsymbol{v}^T\right)\left(\boldsymbol{w}\boldsymbol{w}^T-\left(\boldsymbol{w}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{w}\boldsymbol{S}\right)^T\right)-\left(\boldsymbol{w}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{w}\boldsymbol{S}\right)^T}{1-\frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T}.$$

Rappellons la définition de $\boldsymbol{a}$ :

$$\boldsymbol{a}=\frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)+\boldsymbol{S}^T\boldsymbol{S}$$

qui donne :

$$\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)=\left(1-\boldsymbol{v}\boldsymbol{v}^T\right)\left(\boldsymbol{a}-\boldsymbol{S}^T\boldsymbol{S}\right).$$

Nous avons alors :

$$\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}=\left(1-\boldsymbol{v}\boldsymbol{v}^T\right)\left(\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}-\boldsymbol{a}^{-1}\right)$$

et nous obtenons :

$$d=\frac{\left(1-\boldsymbol{v}\boldsymbol{v}^T\right)\left(\boldsymbol{w}\boldsymbol{w}^T-\left(\boldsymbol{w}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{w}\boldsymbol{S}\right)^T\right)-\left(1-\boldsymbol{v}\boldsymbol{v}^T\right)\left(\left(\boldsymbol{w}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{w}\boldsymbol{S}\right)^T-\left(\boldsymbol{w}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{w}\boldsymbol{S}\right)^T\right)}{1-\frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T},$$

c'est-à-dire :

$$d=\frac{\left(1-\boldsymbol{v}\boldsymbol{v}^T\right)\left(\boldsymbol{w}\boldsymbol{w}^T-\left(\boldsymbol{w}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{w}\boldsymbol{S}\right)^T\right)}{1-\frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T}.$$

En posant $e=\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T$, nous avons :

$$\frac{1}{1-\frac{e}{1-\boldsymbol{v}\boldsymbol{v}^T}}=\frac{1-\boldsymbol{v}\boldsymbol{v}^T}{1-\boldsymbol{v}\boldsymbol{v}^T-e}=1+\frac{e}{1-\boldsymbol{v}\boldsymbol{v}^T-e},$$

ce qui donne

$$\begin{aligned}
&\frac{\left(1-\boldsymbol{v}\boldsymbol{v}^T\right)}{1-\frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T} \\
&\qquad = 1-\boldsymbol{v}\boldsymbol{v}^T+\frac{\left(1-\boldsymbol{v}\boldsymbol{v}^T\right)\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T}{1-\boldsymbol{v}\boldsymbol{v}^T-\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T} \\
&\qquad\qquad = 1-\boldsymbol{v}\boldsymbol{v}^T+\frac{\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T}{1-\frac{1}{1-\boldsymbol{v}\boldsymbol{v}^T}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\boldsymbol{a}^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T}.
\end{aligned}$$

En utilisant le premier résultat du lemme 4, nous obtenons :

$$d=\left(\boldsymbol{w}\boldsymbol{w}^T-\left(\boldsymbol{w}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{w}\boldsymbol{S}\right)^T\right)\left(1-\boldsymbol{v}\boldsymbol{v}^T+\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)\left(\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\left(\boldsymbol{s}-\boldsymbol{v}\boldsymbol{S}\right)^T\right).$$

□

# Bibliographie


Abrahamsen, P. 1997, *A review of Gaussian random fields and correlation functions - 2nd edition*, Norwegian Computing Center.

Akaike, H. 1980, «The interpretation of improper prior distributions as limits of data dependent proper prior distributions», *Journal of the Royal Statistical Society*, vol. 42, p. 46–52.

Atkinson, A. et Donev, A. 1992, *Optimum Experimental Designs*, Oxford Science Publications.

Berger, B., De Oliviera, V et Sansó, B. 2000, «Objective bayesian analysis of spatially correlated data», *Journal of the American Statistical Association*, vol. 96, p. 1361–1374.

Box, G. et Cox, D. 1964, «An analysis of transformations», *Journal of the Royal Statistical Society*, vol. 26, p. 211–252.

Caflish, R. 1998, «Monte carlo and quasi-monte carlo methods», *Acta Numerica*, vol. 7, p. 1–49.

Clerc, M. 2003, «Tribes, ou la coopération de tribus», *Particle Swarm Optimization Seminar*.

Cornford, D., Nabney, I. et Williams, C. 2002, «Modelling frontal discontinuities in wind fields», *Journal of nonparametric statistics*, vol. 14(1-2), p. 43–58.

Cover, T. et Thomas, J. 2006, *Elements of information theory, 2nd edition*, Wiley-Interscience.

Currin, C., Mitchell, T., Morris, M. et Ylvisaker, D. 1988, «A bayesian approach to the design and analysis of computer experiments», Technical Report ORNL-6498.

DeGroot, M. 1970, *Optimal Statistical Decisions*, McGraw-Hill.

DiMatteo, I., Genovese, C. et Kass, R. 2001, «Bayesian curve-fitting with free-knot splines», *Biometrika*, vol. 88, p. 1055–1071.

Emmerich, M., Giannakoglou, k. et Naujoks, B. 2006, «Single- and multiobjective evolutionary optimization assisted by gaussian random field metamodels», *IEEE Transactions on Evolutionary Computation*, vol. 10, p. 421–439.

Gelman, A. 2006, «Prior distributions for variance parameters in hierarchical models», *Bayesian Analysis*, vol. 1, p. 515–533.

Gelman, A., Carlin, J., Stern, H. et Rubin, D. 2004, *Bayesian Data Analysis - Second Edition*, Chapman & Hall/CRC.

Goldberg, P., Williams, C. et Bishop, C. 1997, «Regression with input-dependent noise : a gaussian process treatment», *Advances in Neural information processing systems*, vol. 10, The MIT Press, p. 493–499.

Jones, D., Schonlau, M. et Welch, W. 1998, «Efficient global optimization of expensive black box function», *Journal of Global Optimization*, vol. 13, p. 455–492.

Kass, E. et Wasserman, L. 1996, «The selection of prior distributions by formal rules», *Journal of the American Statistical Association*, vol. 91, p. 1343–1370.

Kato, K. 2009, «Improved prediction for a multivariate normal distribution with unknown mean and variance», *Annals of the Institute of Statistical Mathematics*, vol. 61, p. 531–542.

Kotz, S. et Nadarajah, S. 2004, *Multivariate t-distributions and their applications*, Cambridge University Press.

Le Riche, R., Schoenauer, M. et Sebag, M. 2006, «Un état des lieux de l'optimisation évolutionnaire et de ses implications en sciences pour l'ingénieur.», *Modélisation Numérique : défis et perspectives, Traité Mécanique et Ingénierie des Matériaux*, vol. 2, p. 187–259.

Matheron, G. 1989, *Estimating and choosing. An essay on probability in practice. Translated from the French and with a preface by A. M. Hasofer*, Springer-Verlag.

MathWorks. URL `http://www.mathworks.fr/fr/help/documentation-center.html`.

Mockus, J., Eddy, W., Mockus, A., Mockus, L. et Reklaitis, G. 1997, *Bayesian Heuristic Approach to Discrete and Global Optimization*, Kluwer Academic Publishers.

Monmarché, N., Guinand, F. et Siarry, P. 2010, *Artificial Ants : from collective intelligence to real life optimization and beyond*, ISTE - Wiley.

Neal, R. 1996, *Bayesian Learning for neural Networks*, vol. 118, Springer, New York.

Oakley, J. et O'Hagan, A. 2004, «Probabilistic sensitivity analysis of complex models : a bayesian approach», *Journal of the Royal Statistical Society*, vol. 66, p. 751–769.

Pearl, J. 1988, «Do we need higher-order probabilities, and if so, what do they mean ?», *Conference on Uncertainty in Artificial Intelligence.*

Pfingsten, T., Herrmann, D. et Rasmussen, C. 2006, «Model-based design analysis and yield optimization», *IEEE Transactions on semiconductor manufacturing*, vol. 19, p. 475–486.

Rasmussen, C. et Williams, C. 2006, *Gaussian processes for machine learning*, The MIT Press.

Robert, C. 2007, *The Bayesian choice : from decision-theoretic foundations to computational implementation - 2nd edition*, Springer-Verlag New York Inc.

Sacks, J., Welch, W., Mitchell, T. et Wynn, H. 1989, «Design and analysis of computer experiments», *Statistical Science*, vol. 4, p. 409–435.

Santner, T., Williams, B. et Notz, W. 2003, *The design and analysis of computer experiments*, Springer.

Snelson, E., Rasmussen, C. et Ghahramani, Z. 2003, «Warped gaussian processes», *Neural Information Processing Systems (NIPS).*

Stein, M. 1999, *Interpolation of spatial data : some theory for kriging*, Springer.

Talbi, E. 2009, *Metaheuristics : from design to implementation*, vol. 74, Wiley.

Yang, X. 2009, «Firefly algorithms for multimodal optimization», *Stochastic algorithms : Foundation and applications - SAGA 2009*, vol. 5792, p. 169–178.

Zadeh, L. 1965, «Fuzzy sets», *Information and control*, vol. 8, p. 338–353.

**Résumé :** Dans un contexte industriel compétitif, une prévision fiable du rendement est une information primordiale pour déterminer avec précision les coûts de production et donc assurer la rentabilité d'un projet. La quantification des risques en amont du démarrage d'un processus de fabrication permet des prises de décision efficaces. Durant la phase de conception d'un produit, les efforts de développement peuvent être alors identifiés et ordonnés par priorité. Afin de mesurer l'impact des fluctuations des procédés industriels sur les performances d'un produit donné, la construction de *la probabilité du risque défaillance* est développée dans cette thèse. La relation complexe entre le processus de fabrication et le produit conçu (non linéaire, caractéristiques multi-modales...) est assurée par une méthode de régression bayésienne. Un champ aléatoire représente ainsi, pour chaque configuration du produit, l'information disponible concernant la probabilité de défaillance. Après une présentation du modèle gaussien, nous décrivons un raisonnement bayésien évitant le choix a priori des paramètres de position et d'échelle. Dans notre modèle, le mélange gaussien a priori, conditionné par des données mesurées (ou calculées), conduit à un posterior caractérisé par une distribution de Student multivariée.

La nature probabiliste du modèle est alors exploitée pour construire une probabilité de risque de défaillance, définie comme une variable aléatoire. Pour ce faire, notre approche consiste à considérer comme aléatoire toutes les données inconnues, inaccessibles ou fluctuantes. Afin de propager les incertitudes, une approche basée sur les ensembles flous fournit un cadre approprié pour la mise en œuvre d'un modèle bayésien imitant le raisonnement d'expert. L'idée sous-jacente est d'ajouter un minimum d'information a priori dans le modèle du risque de défaillance.

Notre méthodologie a été mise en œuvre dans un logiciel nommé GoNoGo. La pertinence de cette approche est illustrée par des exemples théoriques ainsi que sur un exemple réel provenant de la société STMicroelectronics.

**Mots clés :** Krigeage, inférence bayésienne, mélange de processus gaussiens, distribution de Student multivariée, analyse d'incertitude, évaluation de rendement industriel

**Abstract :** In competitive industries, a reliable yield forecasting is a prime factor to accurately determine the production costs and therefore ensure profitability. Indeed, quantifying the risks long before the effective manufacturing process enables fact-based decision-making. From the development stage, improvement efforts can be early identified and prioritized. In order to measure the impact of industrial process fluctuations on the product performances, the construction of a failure risk probability estimator is developed in this thesis. The complex relationship between the process technology and the product design (non linearities, multi-modal features...) is handled via random process regression. A random field encodes, for each product configuration, the available information regarding the risk of non-compliance. After a presentation of the Gaussian model approach, we describe a Bayesian reasoning avoiding a priori choices of location and scale parameters. The Gaussian mixture prior, conditioned by measured (or calculated) data, yields a posterior characterized by a multivariate Student distribution.

The probabilistic nature of the model is then operated to derive a failure risk probability, defined as a random variable. To do this, our approach is to consider as random all unknown, inaccessible or fluctuating data. In order to propagate uncertainties, a fuzzy set approach provides an appropriate framework for the implementation of a Bayesian model mimicking expert elicitation. The underlying leitmotiv is to insert minimal a priori information in the failure risk model.

Our reasoning has been implemented in a software called GoNoGo. The relevancy of this concept is illustrated with theoretical examples and on real-data example coming from the company STMicroelectronics.

**Keywords :** Kriging, Bayesian inference, Gaussian processes mixture prior, multivariate t-distribution, uncertainty analysis, manufacturing yield evaluation