



**UNIVERSITE DE LILLE 1  
SCIENCES ET TECHNOLOGIES**



**UNIVERSITE DES SCIENCES  
ET TECHNIQUES DE MASUKU**

-----  
**ECOLE DOCTORALE  
SCIENCES POUR  
L'INGENIEUR (EDSPI)**

-----  
**ECOLE DOCTORALE  
SCIENCES  
FONDAMENTALES ET  
APPLIQUEES (EDSFA)**

N° d'ordre : 42 414

\*\*\*

## **THESE**

**Présentée par :  
MBINA MBINA Alban**

**Pour obtenir le grade de  
Docteur de l'Université de Lille 1 Sciences et Technologies  
et de  
l'Université des Sciences et Techniques de Masuku**

**Discipline : *Mathématiques* - Spécialité: *Statistique***

### **Contributions à la Sélection des Variables en Statistique Multidimensionnelle et Fonctionnelle**

Soutenue publiquement le **28 / 10 /2017**, devant le Jury composé de :

**Sophie DABO-NIANG**, Professeur, Université de Lille 3, France, Président  
**Bezza HAFIDI**, MCF-HDR, Université d'Agadir, Maroc, Examineur  
**Célestin KOKONENDJI**, Professeur, Université Franche-Comté, France, Examineur  
**Sana LOUHICHI**, Professeur, Université de Grenoble, France, Rapporteur  
**Abdallah MKHADRI**, Professeur, Université Cadi Ayyad, Maroc, Rapporteur  
**Octave MOUTSINGA**, MCF-CAMES, USTM, Gabon, Examineur  
**Assi N'GUESSAN**, MCF-HDR, Université de Lille 1, France, Directeur de Thèse  
**Guy Martial NKIET**, Professeur CAMES, USTM, Gabon, Co-directeur de Thèse

---

## Table des matières

Table des matières . . . . .	ii
<b>Dédicaces</b>	<b>viii</b>
<b>Remerciements</b>	<b>ix</b>
<b>Résumé</b>	<b>x</b>
<b>Abstract</b>	<b>xi</b>
<b>Notations</b>	<b>xii</b>
<b>Abréviations</b>	<b>xiii</b>
<b>Présentation générale de la thèse</b>	<b>xiv</b>
<b>Introduction générale</b>	<b>1</b>
<b>I Sélection des variables en régression linéaire</b>	<b>15</b>
<b>1 Etat de l'art de quelques méthodes de sélection des variables en régression linéaire</b>	<b>16</b>
1.1 Le problème d'invariance et sélection des variables . . . . .	17
1.2 Approche de Nkiet (2001) : Méthode directe . . . . .	19
1.2.1 Le modèle . . . . .	19
1.2.2 Critère de sélection et propriétés . . . . .	19
1.2.3 Sélection des variables . . . . .	23
1.2.4 Estimation des paramètres $\sigma$ et $p_0$ . . . . .	24
1.3 Approche de An et al (2013) : Méthode ASCCA . . . . .	25
1.3.1 Définition, modèle et caractérisation . . . . .	26
1.3.2 L'analyse canonique des corrélations (ACC) . . . . .	27
1.3.3 Procédure de sélection : Formulation par les moindres carrés . . . . .	28
1.4 Conclusion . . . . .	30
<b>2 Sélection des variables en régression linéaire multidimensionnelle</b>	<b>32</b>
2.1 Modèle et critère de sélection . . . . .	33
2.1.1 Modèle et énoncé du problème . . . . .	33

---

## TABLE DES MATIÈRES

2.1.2	Caractérisation du sous-ensemble $I_1$ . . . . .	35
2.1.3	Estimation du critère . . . . .	41
2.2	Sélection des variables . . . . .	45
2.2.1	Estimation des paramètres $\sigma$ et $s$ . . . . .	45
2.2.2	Propriétés de convergence . . . . .	46
2.3	Etude numérique . . . . .	49
2.3.1	Simulations . . . . .	50
2.3.2	Influence des fonctions de pénalité . . . . .	51
2.3.3	Influence des paramètres $\alpha$ et $\beta$ . . . . .	52
2.3.4	Choix optimal des paramètres $\alpha$ et $\beta$ . . . . .	52
2.3.5	Comparaison avec la méthode An et al. (2013) . . . . .	53
2.3.6	Application aux données réelles . . . . .	55
2.4	Conclusion . . . . .	56
<b>II Sélection des variables en analyse discriminante mixte</b>		<b>65</b>
<b>3</b>	<b>Etat de l'art de quelques méthodes de sélection des variables en analyse discriminante</b>	<b>66</b>
3.1	Approche de NKIET (2012) . . . . .	67
3.1.1	Notations et caractérisation . . . . .	67
3.1.2	Critère de sélection et estimation . . . . .	68
3.2	Approche de Mahat et al (2007) . . . . .	69
3.2.1	Le modèle LoC . . . . .	70
3.2.2	Critère de sélection . . . . .	71
3.2.3	Procédure de sélection : Méthode de Mahat et al. (2007) . . . . .	71
<b>4</b>	<b>Sélection des variables en analyse discriminante mixte sur plusieurs groupes</b>	<b>74</b>
4.1	Critère de sélection pour les variables mixtes dans plusieurs groupes . . . . .	76
4.1.1	Notations et énoncé du problème . . . . .	76
4.1.2	Critère de sélection pour les variables mixtes . . . . .	77
4.2	Estimation et convergence du critère . . . . .	80
4.2.1	Estimation empirique du critère . . . . .	81
4.2.2	Estimation du critère par une approche non paramétrique . . . . .	82
4.2.3	Convergence du critère . . . . .	83
4.2.4	Principaux résultats . . . . .	87
4.3	Sélection des variables . . . . .	100

---

**TABLE DES MATIÈRES**


---

4.3.1	Estimation et convergence des paramètres $\sigma$ et $s$ . . . . .	101
4.4	Etude numérique . . . . .	107
4.4.1	Simulations des données . . . . .	107
4.4.2	Influence des fonctions de pénalité . . . . .	109
4.4.3	Influence des paramètres $\alpha$ et $\beta$ . . . . .	110
4.4.4	Choix optimal des paramètres $\alpha$ et $\beta$ : Algorithme . . . . .	110
4.4.5	Comparaison avec la méthode Mahat et al. (2007) . . . . .	112
4.4.6	Exemple sur des données réelles . . . . .	112
4.5	Conclusion . . . . .	113
 <b>III Sélection des variables sur un modèle linéaire fonctionnel additif</b>		<b>122</b>
<b>5</b>	<b>Etat de l'art sur les modèles linéaires fonctionnels : Estimation et tests</b>	<b>123</b>
5.1	Présentation du modèle et notations . . . . .	124
5.2	Estimation . . . . .	126
5.2.1	Analyse en composantes principales . . . . .	126
5.2.2	Modèle linéaire fonctionnel additif . . . . .	128
5.3	Tests . . . . .	129
5.3.1	Tests dans les modèles linéaires fonctionnels . . . . .	129
5.3.2	Tests dans les modèles linéaires additifs . . . . .	130
<b>6</b>	<b>Sélection des variable sur un modèle linéaire fonctionnel additif</b>	<b>132</b>
6.1	Modèle et critère de sélection . . . . .	134
6.1.1	Modèle, définition et motivation . . . . .	134
6.1.2	Critère de sélection et propriétés . . . . .	135
6.1.3	Estimation et convergence du critère . . . . .	140
6.1.4	Sélection des variables . . . . .	144
6.2	Sélection des variables dans un modèle linéaire fonctionnel additif . . . . .	147
6.2.1	Approximation du modèle linéaire fonctionnel additif : critère de sélection pour les variables fonctionnelles . . . . .	147
6.3	Conclusion . . . . .	151

*TABLE DES MATIÈRES*

---

**Conclusion générale et perspectives** **153**

**Bibliographie** **159**

---

## Liste des tableaux

---

2.1	<i>Perte de prédiction moyenne (PP) sur 1000 réplifications pour différentes fonctions de pénalité, avec <math>\alpha = 0.45</math> et <math>\beta = 0.6</math> . . . . .</i>	60
2.2	<i>Performance moyenne : <i>cptegal</i>, <i>cptinclus</i>, NV par validation croisée (k-fold cross-validation) pour différentes valeurs de <math>\tau</math>, sur 500 réplifications : dimensions fixes . .</i>	61
2.3	<i>Perte de prédiction moyenne (PP) et nombre de variable sélectionnées (NV) par validation croisée (k-fold cross-validation) sur 500 réplifications, dimensions croissantes : <math>n = 50, p = 36, q = 18, n = 100, p = 46, q = 23</math> et <math>n = 200, p = 58, q = 29</math> . . .</i>	62
2.4	<i>Performance moyenne : <i>cptegal</i>, <i>cptinclus</i>, NV par validation croisée (k-fold cross-validation) sur 500 réplifications, dimensions croissantes : <math>n = 50, p = 36, q = 18, n = 100, p = 46, q = 23</math> et <math>n = 200, p = 58, q = 29</math> . . . . .</i>	63
2.5	<i>Perte de prédiction moyenne (PP) et nombre de variables sélectionnées par validation croisée (k-fold cross-validation) sur 1000 réplifications. <math>n = 252, p = q = 6, k = 15</math> . . . . .</i>	64
4.1	<i>Taux de bon classement moyen sur 1000 réplifications avec différentes fonctions de pénalité (<math>n = 100, n_1=n_2=50</math>), <math>\alpha = 0.25, \beta = 0.5, \lambda = 0.5</math>). . . . .</i>	115
4.2	<i>Taux de bon classement moyen sur 1000 réplifications avec différentes fonctions de pénalité (<math>n = 300, n_1=n_2=150</math>), <math>\alpha = 0.25, \beta = 0.5, \lambda = 0.5</math>). . . . .</i>	116
4.3	<i>Taux de bon classement moyen sur 1000 réplifications avec différentes fonctions de pénalité (<math>n = 500, n_1=n_2=250</math>), <math>\alpha = 0.25, \beta = 0.5, \lambda = 0.5</math>). . . . .</i>	117
4.4	<i>Taux de bon classement moyen sur 1000 réplifications de notre méthode contre celle de Mahat. . . . .</i>	120
4.5	<i>Taux de bon classement moyen de notre méthode sur 500 réplifications. . . . .</i>	120
4.6	<i>Taux de bon classement moyen sur 1000 réplifications avec <math>f_7 = n^{-1/4}/h_7, g_7 = n^{-1/4}h_7, \alpha = 0.25</math> et <math>\beta = 0.5</math>. Cas du Modèle 2. . . . .</i>	121
4.7	<i>Taux de bon classement moyen et nombre de variables sélectionnées (NV) par la méthode 10-fold cross-validation. . . . .</i>	121

---

---

## Table des figures

---

2.1	<i>Perte de Prédiction moyenne sur 1000 réplifications des différentes valeurs de <math>\alpha</math> en fonction de <math>\beta</math>. 1 – (a)<math>n = 100</math>, 1 – (b)<math>n = 300</math> . . . . .</i>	58
2.2	<i>Perte de Prédiction moyenne sur 1000 réplifications de notre méthode contre la méthode ASCCA en fonction du paramètre <math>\rho</math>. 2 – (a)<math>n = 100</math>, 2 – (b)<math>n = 300</math> : dimensions fixes . . . . .</i>	58
2.3	<i>Perte de Prédiction moyenne sur 1000 réplifications de notre méthode contre la méthode ASCCA en fonction du paramètre <math>\tau</math>. 3 – (a)<math>n = 100</math>, 3 – (b)<math>n = 300</math> : dimensions fixes . . . . .</i>	59
4.1	<i>Taux de bon classement moyen sur 1000 réplifications pour différentes valeurs de <math>\alpha</math> en fonction de <math>\beta</math>. <math>n = 100</math> . . . . .</i>	118
4.2	<i>Taux de bon classement moyen sur 1000 réplifications pour différentes valeurs de <math>\alpha</math> en fonction de <math>\beta</math>. <math>n=300</math> . . . . .</i>	119
4.3	<i>Taux de bon classement moyen sur 1000 réplifications pour différentes valeurs de <math>\alpha</math> en fonction de <math>\beta</math>. <math>n=500</math> . . . . .</i>	119

---

---

# Dédicaces

---

*A mes parents*

*Joachim MBINA MAPANGOUI et Geneviève TOBA.*

*A mes fils*

*Christ Georges et Pierre Lucas.*

*A ma chérie*

*Elsa OTAYI.*



---

# Remerciements

---

Je tiens à remercier le Professeur Guy Martial NKIET de l'Unité de Recherche de Mathématiques et Informatique (URMI) de l'Université des Sciences et Techniques de Masuku (USTM - Gabon), mon directeur de thèse, qui a bien voulu diriger ce travail. Sa rigueur scientifique, ses conseils et les différentes discussions m'ont beaucoup aidé à mener ce travail à son terme.

Mes remerciements vont également à Assi N'GUESSAN du Laboratoire Paul Painlevé de l'Université de Lille 1 Sciences et Technologies (USTL - France), mon co-directeur de thèse, dont la collaboration, tant sur le plan scientifique qu'administratif, a largement contribué à l'aboutissement de cette thèse. A travers lui, ce sont tous les enseignants qui ont contribué à ma formation, à qui j'exprime ma profonde reconnaissance.

Je remercie le professeur Sophie DABO-NIANG, d'avoir accepté de présider le jury de cette soutenance, à travers elle, je remercie aussi les professeurs Abdallah MKHADRI et Sana LOUHICHI qui ont accepté de rapporter cette thèse.

Une mention spéciale à l'endroit du Professeur Octave MOUTSINGA, chef de département de Mathématiques et Informatique à l'Université des Sciences et Techniques de Masuku, pour son implication personnelle dans l'élaboration des différentes conventions qui ont facilité la co-tutelle entre USTM et USTL d'une part, et d'autre part, l'URMI et le Laboratoire Paul Painlevé au sein desquels j'ai travaillé.

Merci aux partenaires financiers qui ont contribué au financement de cette thèse, notamment l'Agence Nationale des Bourses du Gabon (ANBG) et l'Agence Universitaire de la Francophonie (AUF) et le Programme d'Appui Institutionnel et Développement des Ressources Humaines (PAI-DRH).

Je tiens à louer le soutien perpétuel qui m'a été apporté par ma famille. Merci à mes frères et soeurs pour le soutien moral. Mes remerciements vont aussi à mes amis, pour m'avoir souvent été de bon conseil. Ils sont nombreux, et je sais qu'ils m'excusent déjà de ne pouvoir les citer. Que tous trouvent, ici, l'expression de ma profonde gratitude et de ma sincère reconnaissance pour tous les services dont j'ai bénéficié.

---

# Résumé

---

Cette thèse aborde le problème de la sélection des variables dans les modèles de régression linéaires multidimensionnels et les modèles de régression linéaires fonctionnels. Plus précisément, nous proposons trois nouvelles approches de sélection des variables qui généralisent des méthodes existantes dans la littérature. La première méthode permet de sélectionner des variables aléatoires continues dans un modèle linéaire multidimensionnel. Cette approche généralise celle de *Nkiet* (2001) obtenue dans le cas d'un modèle linéaire unidimensionnel. Une étude comparative effectuée par simulation et une application aux données réelles sont proposées. Le critère de performance repose sur le calcul de la perte de prédiction. Sur la base de ce critère de performance, notre méthode présente des résultats très encourageants. La deuxième approche propose une nouvelle méthode de sélection des variables mixtes (mélange de variables aléatoires continues et discrètes) en analyse discriminante, pour plus de deux groupes. C'est une extension de la première contribution dans le cadre mixte. Une étude numérique et une application aux données réelles sont également proposées. Les résultats obtenus montrent à partir du calcul du taux de bon classement que cette méthode a des performances comparables à celle de *Mahat et al.* (2007), dans le cas de deux groupes. Dans la troisième et dernière partie de cette thèse, nous proposons d'étendre dans le cadre fonctionnel, les résultats obtenus dans les deux premières contributions. Nous établissons un critère de sélection pour le cas fonctionnel en nous appuyant sur un modèle linéaire multivarié. Ses trois méthodes reposent sur des critères qui mesurent la perte d'informations, d'un sous-ensemble de variables sélectionné. Les critères permettent ensuite de ranger les variables dans l'ordre décroissant, de tel sorte que le problème de sélection des variables se ramène à un problème d'estimation d'une permutation aléatoire et d'un paramètre appelé dimension.

**Mots clés :** *Modèle linéaire multidimensionnel - Modèle linéaire fonctionnel additif - Sélection des variables - Critère de sélection - Variables aléatoires mixtes - Analyse discriminante mixte - Convergence - Simulations - Données réelles - Perte de prédiction - Taux de bon classement.*

---

# Abstract

---

This thesis studies variable selection in multidimensional linear regression models and functional linear regression models. More precisely, we provide three new approaches on variable selection. The first one, allows to select continuous random variables in multidimensional linear model. This approach generalizes the method of *Nkiet* (2001), obtained in the case of univariate linear models. A simulation study is proposed. It is based on the calculation of prediction loss and an application to real data. Our method gives encouraging results. The seconde one gives a new way to selection mixed variables (mixture of continuous and discrete variables) in discriminant analysis, when the number of groups is greater than two. It is an extension of the first contribution in the mixed case. A numerical study and an application to real data is also proposed. The results obtained show from the calculation of classification capacity, that our method is equivalente to that of *Mahat et al.* (2007) when we fixe the number of groups at two. In the third part of this thesis, we propose to extend in functional case, the results obtained in the above mentioned contributions by providing a selection criterion for the functional data. These three approaches are based on criteria which measure information loss, when a subset of variables is selected. Criteria permit to arrange the variables in decreasing order. In another words, we transform our variable selectio problem to an estimation problem.

**Keywords :** *Multivariate linear model - Additif fonctional linear model - Variable selection - Selection criterion - Mixed random variable - Mixed discriminant analysis - Consistency - Simulations - Real data set - Prediction loss - Rate of good classification.*

---

# Notations

---

$(\Omega, \mathcal{A}, P)$	Espace probabilisé
$I_0$	Sous-ensemble des indices des variables non sélectionnées
$I_1$	Sous-ensemble des indices des variables sélectionnées
$\mathbb{R}^p$	Vecteurs des réels de tailles $p$
$\text{card}(K)$	Cardinal du sous-ensemble $K$
$\mathbb{P}$	Probabilité
$\mathbb{P}\{W=w\}$	Probabilité sachant l'évènement $\{W = w\}$
$\mathbb{E}$	Espérance mathématique
$\approx$	Approximation
$ \cdot $	Valeur absolue
$[\cdot]$	Partie entière
$\ \cdot\ _{\mathbb{R}^p}$	Norme euclidienne dans $\mathbb{R}^p$
$\langle \cdot, \cdot \rangle_{\mathbb{R}^p}$	Produit scalaire euclidien dans $\mathbb{R}^p$
$\otimes$	Produit tensoriel vectoriel
$\mathcal{L}(\mathbb{R}^p)$	Espace vectoriel d'opérateurs linéaires définis de $\mathbb{R}^p$ vers $\mathbb{R}^p$
$\mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$	Espace vectoriel d'opérateurs linéaires définis de $\mathbb{R}^p$ vers $\mathbb{R}^q$
$\mathcal{L}(\mathbb{R}^p)^M$	Espace vectoriel d'opérateurs linéaires définis de $\mathbb{R}^p$ vers $\mathbb{R}^p$ de taille $M$
$\mathcal{L}(\mathcal{L}(\mathbb{R}^p), \mathbb{R}^q)$	Espace vectoriel d'opérateurs linéaires définis de $\mathcal{L}(\mathbb{R}^p)$ vers $\mathbb{R}^q$
$\mathcal{M}_{q,M}(\mathbb{R})$	Espace des matrices à coefficients réelles à $q$ lignes et $M$ colonnes
$L^2_{\mathcal{C}}$	Espace des fonctions de carré intégrables dans le compact $\mathcal{C}$
$L^2((\Omega, \mathcal{A}, P))$	Espace des variables aléatoires euclidiennes de carré intégrable

---

# Abréviations

---

ACP	Analyse en Composante Principale
ACL	Analyse Canonique Linéaire
AIC	Akaike Information Criterion
ASCCA	Adaptive Sperse Canonical Correlation Analysis
BIC	Bayesian Information Criterion
CMV	Cross Model Validation
ERC	Error Rate of Misclassification
FP	Fonction de Pénalité
FPE	Final Predictor Error
LASSO	Least Absolute Shrinkage and Selection Operator
LoC	Location Model
NV	Nombre de variables Sélectionnées
PFM	Predictor Full Model
PP	Perte de Prédiction
PTM	Predictor True Model
RFM	Response Full Model
RTM	Response True Model
SRM	Sufficient Response Model
VC	Validation-Croisée

---

# Présentation générale de la thèse

---

## Nom de l'auteur

Alban MBINA MBINA

## Intitulé de la thèse

Contributions à la Sélection des Variables en Statistique Multidimensionnelle et Fonctionnelle

## Encadrants (Nom, grade et institution )

- Assi N'GUESSAN, MCF-HDR, Lille1, USTL, Co-directeur de thèse
- Guy Martial NKIET, Professeur (CAMES), USTM, Directeur de thèse

**Période de réalisation de la thèse** Janvier 2012 - Mai 2016

## Lieux

- Unité de Recherche de Mathématiques et Informatique (URMI), USTM-Gabon
- Laboratoire Paul Painlevé (PPL), USTL-France

## Rapporteurs (Nom, grade, institution)

- Sana LOUHICHI, Professeur, Université de Grenoble Alpes, France
- Abdallah MKHADRI, Professeur, Université Cadi Ayyad, Maroc

## Examineurs (Nom, grade, institution)

- Sophie DABO-NIANG, Professeur, Université Lille 3, France
- Bezza HAFIDI, MCF-HDR, Université d'Agadir, Maroc
- Célestin KOKONENDJI, Professeur, Université Franche-Comté, France
- Octave MOUTSINGA, MCF-CAMES, USTM, Gabon

---

---

---

## **Principales publications auxquelles ce travail a donné lieu**

1. Variable selection in multiple regression with random design, to appear in *Statistical Methods and Applications*.
2. Variable selection in discriminant analysis for mixed variables and several groups, to appear in *Advances in Data Analysis and Classification*.

## **Principales communications auxquelles ce travail a donné lieu**

1. Variable selection in discriminant analysis for mixed variables and several groups, présentée à la Journée des Doctorants, 12-13 février 2014, USTM Gabon.
2. Variable selection in multiple regression with random design, présentée au regroupement sous Régional (AUF-Projet Horizon Francophone), Yaoundé 2012.

---

# Introduction générale

---

Cette thèse, aborde le problème de la sélection de variables dans les modèles de régression linéaires multidimensionnels et les modèles linéaires fonctionnels additifs. Plus précisément, les objectifs visés sont les suivants :

1. Proposer une méthode de sélection des variables, pour des variables aléatoires continues, dans un modèle linéaire multidimensionnel ;
2. Proposer une méthode de sélection des variables, pour des variables aléatoires mixtes en analyse discriminante ;
3. Proposer une méthode de sélection des variables, dans un modèle linéaire fonctionnel additif.

Cette thèse est principalement motivée par les travaux de *Nkiet* (2001); *Nkiet* (2012), qui ont mis en lumière une méthode de sélection des variables à partir d'un critère de sélection (Chapitre 1, Section 1.2.2, Equation (1.2)), dont l'un des avantages vient du fait que les variables sélectionnées ne font pas l'objet d'une quelconque hypothèse de loi de probabilité. Cette condition, toujours présente dans les méthodes classiques (littérature) de sélection des variables, peut être très restrictive en pratique. Aujourd'hui levée, elle offre de perspectives théoriques et pratiques dans le domaine de la sélection des variables. Le critère proposé permet de réduire le nombre d'étapes, dans le processus de sélection du sous-ensemble des variables sélectionnées, lorsque la taille de l'ensemble des variables à sélectionner est grande. C'est dans ce cadre que nous nous proposons d'étendre cette méthode dans des modèles statistiques plus complexes que le modèle linéaire unidimensionnel (*Nkiet* (2001)) et l'analyse discriminante des variables aléatoires continues (*Nkiet* (2012)). Dans chaque généralisation, nous proposons une étude de simulations détaillée et un exemple sur des données réelles.



---

---

---

Ainsi, le présent manuscrit, constitué de six chapitres, se divise en trois principales parties de deux chapitres chacune. La première partie traite du problème de sélection des variables continues dans le contexte d'un modèle linéaire multidimensionnel. La deuxième partie traite du problème de sélection des variables en analyse discriminante mixte et la troisième partie porte sur la sélection des variables dans un modèle linéaire fonctionnel additif.

## Sélection des variables continues en régression linéaire

La sélection des variables dans les modèles de régression linéaires est un problème relativement ancien en statistique. Son objectif est de déterminer un sous-ensemble de variables (prédicteurs ou variables explicatives) qui expliquent et prédisent mieux la variable réponse (variable à expliquer ou variable d'intérêt). Les domaines d'application sont multiples et variés. On peut citer sans être exhaustif l'épidémiologie, les finances et banques, la génétique, la biologie, etc. Pour apporter des réponses théoriques et pratiques liées aux problèmes des domaines d'application susmentionnés, plusieurs approches ont été développées à travers diverses méthodes d'analyse statistique. En régression linéaire par exemple, de nombreux critères de sélection ont été proposés dans la littérature. Des travaux dans ce domaine peuvent être trouvés dans *Hocking* (1976) et *Thompson* (1978), alors que certaines monographies à ce sujet sont disponibles dans *Linhart et Zucchini* (1986) et *Miller* (1990). Dans le cas des variables continues, des méthodes de sélection des variables ont fait l'objet de plusieurs recherches, c'est le cas des méthodes de *McKay* (1976, 1977), *McLachlan* (1980), *Fujikoshi* (1985), *Le Roux et al.* (1997).

---

---

---

Dans ce contexte, deux types de variables sont généralement utilisés : les variables non aléatoires et les variables aléatoires. Le cas des variables non aléatoires a fait l'objet de plusieurs critères de sélection. C'est le cas du critère "Final Predictor Error" (FPE) de *Thompson* (1978), *Shibata* (1984), *Zhang* (1992), le critère de validation croisée de *Shao* (1993), *Zhang* (1993), le critère d'information d'Akaike "Akaike Information Criterion" (AIC), le critère  $C_p$  de *Fujikoshi et Sato* (1997), le critère d'erreur de prédiction de *Fujikoshi et al.* (2011) etc. Par contre, on trouve juste quelques travaux dans le cas où les variables sont supposées aléatoires. C'est sous cette hypothèse que *Zheng et Low* (1997) ont proposé des méthodes de sélection des variables pour des modèles linéaires univariés (la variable réponse est une variable aléatoire réelle). Bien que son importance a été reconnue par *Breiman et Spector* (1992), cette approche donne généralement des erreurs de prédiction plus élevées que celles obtenues lorsque les variables sont supposées fixes.

La plupart des méthodes de sélection citées précédemment sont généralement obtenues sous l'hypothèse de normalité des variables sélectionnées. Cette hypothèse pourrait être très restrictive dans la mesure où en pratique, toutes les variables ne suivent pas nécessairement une loi gaussienne. C'est ainsi que *Nkiet* (2001) (Chapitre 1, Section 1.2) a introduit une méthode de sélection sur un modèle linéaire univarié, dont les propriétés ne sont pas assujetties à une quelconque hypothèse de loi. Cette méthode repose sur le critère d'invariance de l'Analyse Canonique Linéaire (ACL) de *Dauxois et Nkiet* (1997a), et donne de bonnes performances comparativement aux critères des méthodes classiques.

Récemment, *An et al.* (2013) (Chapitre 1, Section 1.3) ont proposé une méthode de sélection appelée (ASCCA) sur un modèle linéaire multidimensionnel. Cette méthode traite du problème de la régression "shrinkage" et de la sélection des variables en régression linéaire multiple.

---

Elle est basée sur la redéfinition d'un problème de régression linéaire multivarié, en un problème classique d'analyse canonique corrélé, pour lequel des formulations des méthodes des moindres carrés, méthode du LASSO sont proposées, tout en utilisant un critère BIC.

De ce fait, il y a donc intérêt à étendre la méthode de sélection de *Nkiet (2001)* à un modèle linéaire multidimensionnel afin de la comparer à celle de *An et al. (2013)*.

Nous revenons sur les grandes lignes de ces deux méthodes (*Nkiet (2001)* et *An et al. (2013)*) au Chapitre 1. Dans le Chapitre 2, nous généralisons la méthode de sélection de *Nkiet (2001)* en considérant un modèle linéaire multidimensionnel. Ce chapitre contient la première innovation méthodologique et théorique de cette thèse. Nous nous plaçons dans le cadre d'un modèle linéaire multidimensionnel de la forme :

$$Y = BX + \varepsilon, \quad (1)$$

où la variable aléatoire  $Y$  n'est plus cette fois-ci une variable aléatoire réelle, mais plutôt un vecteur aléatoire dans  $\mathbb{R}^q$  ( $q \geq 2$ ),  $X$  est un vecteur aléatoire à valeurs dans  $\mathbb{R}^p$  ( $p \geq 2$ ),  $B$  est une matrice dans  $\mathcal{M}_{q,p}(\mathbb{R})$  et  $\varepsilon$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^q$ . C'est le vecteur aléatoire qui caractérise l'erreur du modèle. Il est supposé indépendant de  $X$ . Ainsi, pour tout  $j = 1, \dots, q$ , l'équation (1) devient :

$$Y_j = \sum_{i=1}^p X_i b_{ji} + \varepsilon_j \quad \text{où } j = 1, \dots, q. \quad (2)$$

Ce modèle (Equation (2)) est le modèle utilisé dans *Nkiet (2001)*.

---



---

On suppose que les vecteurs  $X, Y$  et  $\varepsilon$  sont centrés. On définit les opérateurs de covariances associés aux variables  $X$  et  $Y$  par :

$$V_1 = \mathbb{E}(X \otimes X) \quad \text{et} \quad V_{12} = \mathbb{E}(Y \otimes X),$$

où  $\otimes$  est le produit tensoriel vectoriel. L'opérateur  $V_1$  est supposé inversible. On considère l'opérateur  $A_K$  défini par :

$$x = (x_i)_{i \in I} \in \mathbb{R}^p \mapsto x_K = (x_i)_{i \in K} \in \mathbb{R}^{\text{card}(K)},$$

où  $K$  est un sous-ensemble de  $I$ , et  $I$  un ensemble d'indices (de taille  $p$ ) des variables candidates à la sélection. On définit l'opérateur  $\Pi_K$  par  $\Pi_K := A_K^* (A_K V_1 A_K^*)^{-1} A_K$ , c'est un projecteur défini de  $\mathbb{R}^p$  vers le sous ensemble  $K$ ,  $A_K^*$  est l'adjoint de l'opérateur  $A_K$ . On utilise le critère de sélection suivant :

$$\xi_K = \|V_{12} - V_1 \Pi_K V_{12}\| \tag{3}$$

où  $\|\cdot\|$  est la norme usuelle d'opérateurs définie par  $\|A\| = \sqrt{\text{tr}(A^* A)}$ . Le critère (3) diffère de celui de [Nkiet \(2001\)](#) au niveau de l'opérateur  $V_{12}$ , qui est maintenant l'espérance d'un produit tensoriel entre les vecteurs  $Y$  et  $X$ . Nous définissons à partir de l'équation (1), les sous-ensembles  $I_0$  et  $I_1$ , respectivement des indices des variables liées au vecteur aléatoire  $Y$ , et ceux non liées au vecteur aléatoire  $Y$  :

---



---



---


$$I_0 = \{j \in I / \|b_{\bullet j}\|_{\mathbb{R}^q} = 0\}, \quad I_1 = \{\exists j \in I / \|b_{\bullet j}\|_{\mathbb{R}^q} \neq 0\}, \quad \text{et } \mathbf{b}_{\bullet j} = \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{qj} \end{pmatrix},$$

où  $I = \{1, \dots, p\}$  est l'ensemble d'indices de variables candidates à la sélection.

A partir de la définition du sous-ensemble  $I_1$ , nous établissons dans le Lemme 3, le premier résultat de notre approche en donnant une caractérisation. En nous appuyant sur ce résultat fondamental, nous proposons au Chapitre 2 une nouvelle méthode de sélection des variables continues sur un modèle linéaire multidimensionnel, valable quelle que soit la loi de probabilité suivie par les vecteurs  $X$  et  $Y$ . Les résultats théoriques, renfermant les estimations, la convergence du critère et la procédure de sélection des variables sont établis dans la Proposition 2.1.1 et le Théorème 7. Sur la base d'une étude numérique, nous comparons notre approche à celle de [An et al. \(2013\)](#). Les principaux résultats sont reportés dans les Tableaux 2.1-2.4, Section 2.3 et les Figures 2.1-2.3. Ils montrent sur la base de la perte de prédiction, que notre méthode donne de bon résultats comparativement à celle de [An et al. \(2013\)](#).

Pour boucler notre étude, nous donnons un exemple sur des données réelles. Nous utilisons des données sur les performances des portefeuilles d'actions à notation pondérée, obtenues sur la base de données historiques des marchés boursiers américains (voir [Liu et Yeh \(2015\)](#)). Les résultats obtenus viennent confirmer ceux établis par simulations (voir Tableau 2.5).

---

---

## Sélection des variables en analyse discriminante mixte

Lorsque l'on considère un mélange de variables aléatoires continues et de variables aléatoires discrètes, on parle de variables aléatoires mixtes. L'analyse discriminante devient l'analyse discriminante mixtes. Cette branche de la statistique a été développée pour la première fois par *Olkin et Tate* (1961) et *Krzanowski* (1975) à travers l'utilisation du "Location Model" (LoC). Plusieurs approches de sélection de variables ont ainsi été proposées (voir *Krzanowski* (1983), *Bar-Hen et Daudin* (1995), *Daudin* (1986), *Daudin et Bar-Hen* (1999)). Tout comme dans le cas des variables continues, les méthodes de sélection des variables mixtes sont généralement obtenues sous l'hypothèses de normalité des variables étudiées, mais en plus, elles sont en général des extensions dans le cas mixte, des méthodes où les variables sont continues. C'est dans ce contexte que *Mahat et al.* (2007) (Chapitre 3, Section 3.2) proposent une méthode de sélection des variables dont le critère repose sur la divergence de Kullback-Leibler. En théorie, la méthode résulte du problème de classement en analyse discriminante mixte sur deux groupes de population. *Nkiet* (2012) (Chapitre 3, Section 3.1) propose dans le cas des variables aléatoires continues une méthode de sélection des variables en supposant cette fois-ci  $q$  groupes de population ( $q > 2$ ) au lieu de deux ( $q = 2$ ). En généralisant le critère (1.2) (Chapitre 1, Section 1.2.2) dans le cadre de l'analyse discriminante, il obtient une méthode plus performante comparativement aux méthodes classiques issues des critères AIC, ERC et CMV.

Ainsi, il y a donc intérêt à étendre la méthode de *Nkiet* (2012) dans le cas des variables aléatoires mixtes, et de la comparer à celle de *Mahat et al.* (2007).

---



---

Après un rappel des méthodes *Nkiet* (2012) et *Mahat et al.* (2007) au Chapitre 3, dans le Chapitre 4, nous proposons une méthode de sélection de variables dans le cadre de l'analyse discriminante pour des variables aléatoires mixtes sur plus de deux groupes. Comme notre méthode de sélection tient compte de la variable discrète, nous définissons dans ces conditions, un vecteur multinomial  $U$  contenant  $M$  états ou cellules, tels que  $M = 2^{k-1}$  si  $k$  est le nombre de catégories de la variable discrète

et  $M = 2^d$  si la variable discrète est une variable binaire. Nous établissons sur la base du critère (3.1) (Chapitre 3, Section 3.1.2), le critère de sélection pour les variables aléatoires mixtes :

$$\xi'_K = \sum_{m=1}^M \sum_{\ell=1}^q p_m^2 p_{\ell|m}^2 \| (I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m}) (\mu_{\ell,m} - \mu_m) \|_{\mathbb{R}^p}^2, \quad (4)$$

où  $p_m = P(U = m)$ ,  $p_{\ell|m} = P(Z = \ell | U = m)$ ,  $\mu_m = \mathbb{E}(X | U = m)$ ,  $\mu_{\ell,m} = \mathbb{E}(X | Z = \ell, U = m)$ ,  $\mathbb{V}_m = \mathbb{E}[(X - \mu_m) \otimes (X - \mu_m) | U = m]$  et  $Q_{K|m} = A_K^* (A_K \mathbb{V}_m A_K^*)^{-1} A_K$ , (l'opérateur  $A_K$  et son adjoint  $A_K^*$  sont définis au Chapitre 1).

Dans le Théorème 8 du Chapitre 4, nous donnons une caractérisation du sous-ensemble  $I_1$  des variables à sélectionner. Partant de ce résultat, nous proposons une méthode de sélection des variables dans le cadre mixte, pour plus de deux groupes et reste valable quelle que soit la loi de probabilité suivie par les variables du modèle. Nous proposons un estimateur du critère (4), et à travers les Lemmes 6 - 9 et les Théorèmes 9 - 11, nous étudions les propriétés de convergence de notre approche. Sur une étude numérique, nous comparons notre approche à celle de *Mahat et al.* (2007). Les principaux résultats sont reportés dans les Tableaux 4.4 et 4.5 de la Section 4.4.2.

---

---

---

En utilisant le taux de bon classement, nous montrons que les deux approches sont équivalentes. Ce résultat est encore vérifié sur un exemple avec des données réelles. Nous utilisons dans cet exemple des données de demande de carte de crédit australienne que l'on peut retrouver dans la base de données de l'Université de Californie à Irvine (voir [Lichman M \(2013\)](#)). Les deux méthodes donnent des résultats semblables en ce qui concerne le taux de bon classement. Par contre, la méthode de [Mahat et al. \(2007\)](#) sélectionne plus de variables que la notre (voir Tableau 4.7).

## Sélection des variables en statistique fonctionnelle

La statistique fonctionnelle a connu un développement important ces dernières années. Cette partie de la statistique consiste à étudier des données qui peuvent être assimilées à des courbes ou à des surfaces. Ces données sont habituellement fonction du temps ou de l'espace. Dans la littérature, les données fonctionnelles ont été introduites avant tout pour des besoins purement pratiques. Les domaines d'application sont multiples et variés. Nous avons par exemple les données de température, les données d'images observées par satellite, les données de chimiothérapie, les données de courbes de croissance etc.

Au-delà du cadre pratique, la littérature fournit un cadre théorique pour l'étude des données fonctionnelles, et garde les mêmes objectifs que la statistique classique (analyse de données, inférence, etc). Sauf que cette fois-ci, les données ont simplement la particularité d'être des fonctions.



---

---

---

C'est dans ce contexte que la régression du signal fût largement étudiée à partir des données fonctionnelles. Un excellent résumé des méthodes et techniques a été donné par *Frank et Friedman* (1993). Récemment, *Ramsay et Silverman* (2005) ont décrit un exemple où les données fonctionnelles sont utilisées pour étudier la régression du signal. Cette étude est une excellente application dans le domaine de la chirurgie. Si les données fonctionnelles comme les signaux sont utilisées comme régresseurs, le principal problème est le grand nombre de prédicteurs rendant inapplicables des techniques de la statistique multivariée telles que les moindres carrés. Elle peut donner de mauvais résultats à cause du problème d'inverse qui peut se révéler difficile, voire impossible à calculer pour deux raisons. La première raison vient du fait que dans le cas d'une matrice de taille  $n \times p$ , le nombre de colonnes  $p$  de la matrice est généralement plus élevé que la taille de l'échantillon  $n$  (on parle de la statistique en haute dimension). La seconde raison vient du fait qu'il y a de fortes chances que les  $p$  prédicteurs soient fortement corrélés du fait qu'ils sont tous des points d'une même fonction.

Pour contourner la difficulté du calcul d'inverse de l'opérateur de covariance associé à la fonction aléatoire  $X$ , des solutions ont été proposées, on peut relever :

- La "ridge regression". Elle a été initiée par *Hoerl et Kennard* (1980). Elle consiste à ajouter un terme de pénalité dans le critère des moindres carrés.
- La régression par composantes principales *Cardot et al.* (1999). Elle consiste à réduire la dimension  $p$  en utilisant les  $k$  premières composantes principales de l'opérateur de covariance associé à la variable  $X$ .

- 
- 
- 
- La régression "partial least squares". Elle a été proposée par [Helland \(1990\)](#). C'est une méthode algorithmique basée à chaque étape sur la régression par moindres carrés ordinaires, sur les résidus de l'étape précédente.

En plus de ces méthodes, [Deville \(1974\)](#) a introduit une analyse en composantes principales des courbes pour généraliser au cadre fonctionnel les méthodes statistiques multivariées. Tandis que dans la thèse de [Dauxois et Pousse \(1976\)](#), on traite d'un certain nombre de méthodes regroupées sous le nom d'analyse factorielle. [Dauxois et al. \(1982\)](#) abordent en plus les aspects asymptotiques fournissant des résultats de convergence pour les opérateurs de covariances empiriques. Récemment, [Crambes \(2006\)](#), propose plusieurs solutions dans le cas de la régression linéaire pour des variables explicatives fonctionnelles. Pour plus de lecture, nous pouvons aussi consulter [Gareth et al. \(2009\)](#) et [Gerhard Tutz et Jan Gertheiss \(2007\)](#).

Dans cette partie, nous nous intéressons aux modèles linéaires fonctionnels de la forme :

$$Y = \int_{\mathcal{C}} \phi(t)X(t)dt + \varepsilon, \quad (5)$$

où  $\phi$  est une fonction de carré intégrable sur un compact  $\mathcal{C}$ ,  $\varepsilon$  est une variable aléatoire réelle indépendante de  $X$  de moyenne nulle et de variance  $\sigma^2$  et  $X = \{X(t), t \in \mathcal{C}\}$  est un processus continu et défini sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ . L'équation (5) peut aussi se mettre sous la forme :

$$Y = \Phi(X) + \varepsilon, \quad (6)$$

(voir [Cardot et al. \(1999\)](#)), où  $X$  est une variable aléatoire à valeurs dans un espace de Hilbert  $\mathcal{H}$  et  $\Phi$  est un opérateur linéaire continu sur  $\mathcal{H}$  ([Cardot et al. \(1999, 2003b\)](#)). L'intérêt de ce modèle est qu'il permet d'expliquer la relation entre une variable explicative de type fonctionnelle et une variable aléatoire réelle.

---



---

Cependant, l'équation (6) (ou l'équation (5)) ne prend pas suffisamment en compte le cas où on voudrait expliquer la variable aléatoire réelle  $Y$  en fonction de plusieurs autres variables explicatives fonctionnelles  $X_1, \dots, X_p$ , à valeurs dans des espaces de Hilbert  $\mathcal{H}_1, \dots, \mathcal{H}_p$ , respectivement. Nous considérons ainsi dans cette thèse, un modèle linéaire fonctionnel additif de la forme :

$$Y = \sum_{k=1}^p \int_{\mathcal{C}_k} \phi_k(t) X_k(t) dt + \varepsilon, \quad (7)$$

où  $Y$  est une variable aléatoire réelle, les  $\phi_k$  sont des fonctions de carrés intégrables sur le compact  $\mathcal{C}_k$ ,  $\varepsilon$  est une variable aléatoire réelle indépendante de  $X_k$  et  $X_k = \{X_k(t), t \in \mathcal{C}_k\}$  est un processus continu défini sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Tout comme le modèle fonctionnel de l'équation (5), le modèle fonctionnel additif se met facilement sous la forme

$$Y = \sum_{k=1}^p \Phi_k(X_k) + \varepsilon, \quad (8)$$

Les solutions aux problèmes d'estimations et de tests des modèles linéaires fonctionnels classiques (Equations (5) et (6)) ne permettent pas de répondre totalement à la question qui arrive naturellement lorsque l'on veut modéliser une relation entre une variable aléatoire réelle et plusieurs autres variables explicatives fonctionnelles. Cette problématique a permis à [Cardot et al. \(2004\)](#) de proposer une solution en approchant le modèle (7) en un modèle linéaire multivarié. Dans le Chapitre 5 de cette thèse, nous revenons sur quelques techniques d'inférence (estimation et test) des modèles linéaires fonctionnels, en nous appuyant sur les travaux de [Cardot et al. \(1999, 2004\)](#).

---



---

La problématique de la sélection de variables fonctionnelles dans le modèle (7) reste encore très peu abordée dans la littérature. Ainsi, dans le Chapitre 6, nous proposons une méthode théorique de sélection des variables sur un modèle linéaire fonctionnel additif. Plus précisément, nous proposons une extension dans le cadre fonctionnel du critère de sélection proposé aux Chapitres 2 et 3.

Cette partie constitue une introduction de notre méthode de sélection des variables dans le cadre fonctionnel. Pour cela, nous procédons de la façon suivante :

1. Nous considérons un modèle linéaire de la forme :

$$Y = \sum_{k=1}^p b_k^T X_k + \varepsilon, \quad (9)$$

où  $Y$  et  $\varepsilon$  sont des variables aléatoires réelles,  $b_k \in \mathbb{R}^q$  et  $X_k$  est un vecteur aléatoire dans  $\mathbb{R}^q$ . Partant de ce modèle, nous proposons une généralisation de la méthode de [Nkiet \(2001\)](#), dans le but de sélectionner les vecteurs  $X_k$ . Pour cela, on définit

$$I_0 = \{k \in \mathcal{I} | b_k = 0\} \text{ et } I_1 = I - I_0 = \{k \in \mathcal{I} | b_k \neq 0\},$$

respectivement le sous-ensemble d'indices des vecteurs qui n'expliquent pas la variable aléatoire réelle  $Y$  et le sous-ensemble d'indices des vecteurs qui expliquent la variable  $Y$  dans le modèle (9), où  $I = \{1, \dots, p\}$  est l'ensemble des indices des vecteurs à sélectionner. Nous définissons le vecteur aléatoire  $X$  de la façon suivante :

---



---



---


$$X = (X_1^T | X_2^T | \dots | X_p^T)^T \in \mathbb{R}^{pq},$$

Puis, nous définissons les opérateurs de covariances  $V_X$  et  $V_{XY}$  tels que :

$$V_X = \mathbb{E}(X \otimes X) \text{ et } V_{XY} = \mathbb{E}(YX).$$

On suppose dans la suite que l'opérateur  $V_X$  est inversible. Nous définissons l'opérateur linéaire  $A_K$  par :

$$A_K : x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \in \mathbb{R}^{pq} \mapsto (x_i)_{i \in K} \in \mathbb{R}^{\text{card}(K)q},$$

où  $\text{card}(K)$  est le cardinal du sous-ensemble  $K$ . Nous prenons  $\Pi_K = A_K^* (A_K V_X A_K^*)^{-1} A_K$ , puis nous considérons le critère de sélection suivant :

$$\mathcal{C}_K = \|V_{XY} - V_X \Pi_K V_{XY}\|.$$

Nous donnons une caractérisation du sous-ensemble  $I_1$ , puis un estimateur du critère  $\mathcal{C}_K$ . Dans le Théorème 14, nous établissons la convergence du critère, puis dans le Théorème 15, nous établissons la convergence de notre méthode de sélection.

2. A la suite de cette étude, nous approchons le modèle (7) comme un modèle linéaire multivarié semblable au modèle (9). Partant du modèle approché, nous proposons un critère de sélection pour les variables aléatoires fonctionnelles.

# Première partie

## Sélection des variables en régression linéaire

---

# ÉTAT DE L'ART DE QUELQUES MÉTHODES DE SÉLECTION DES VARIABLES EN RÉGRESSION LINÉAIRE

---

## Sommaire

---

<b>1.1</b>	<b>Le problème d'invariance et sélection des variables</b>	<b>17</b>
<b>1.2</b>	<b>Approche de Nkiet (2001) : Méthode directe</b>	<b>19</b>
1.2.1	Le modèle	19
1.2.2	Critère de sélection et propriétés	19
1.2.3	Sélection des variables	23
1.2.4	Estimation des paramètres $\sigma$ et $p_0$	24
<b>1.3</b>	<b>Approche de An et al (2013) : Méthode ASCCA</b>	<b>25</b>
1.3.1	Définition, modèle et caractérisation	26
1.3.2	L'analyse canonique des corrélations (ACC)	27
1.3.3	Procédure de sélection : Formulation par les moindres carrés	28
<b>1.4</b>	<b>Conclusion</b>	<b>30</b>

---

## 1.1 Le problème d'invariance et sélection des variables

---

### Introduction

La sélection des variables est un problème largement abordé dans la littérature. On retrouve un grand nombre de méthodes dans les modèles de régression linéaires. Parmi toutes ses méthodes, il y a celle de *Nkiet* (2001), qui propose une méthode de sélection dans le cas d'un modèle linéaire unidimensionnel. La particularité de cette méthode vient du fait que les variables candidates à la sélection, ne sont soumises à aucune hypothèse de loi de probabilité. Récemment, *An et al.* (2013) ont proposé une méthode (ASCCA) dans le cas d'un modèle linéaire multidimensionnel, qui traite du problème de la régression "shrinkage" et de la sélection des variables en régression linéaire multiple. Les résultats théoriques et les simulations montrent que les performances de ces méthodes (*Nkiet* (2001) et *An et al.* (2013)) sont bien meilleurs comparées aux méthodes classiques.

L'objectif de ce chapitre n'est pas de faire une revue de littérature exhaustive des méthodes classiques de sélection des variables en régression linéaire, nous nous limiterons à présenter les résultats des travaux de *Nkiet* (2001) et de *An et al.* (2013) que nous utiliserons dans cette partie. Mais avant, revenons sur la notion d'invariance qui est à la base du critère que nous allons étendre tout au long de cette thèse.

## 1.1 Le problème d'invariance et sélection des variables

Dans les travaux de *Dauxois et Nkiet* (1997a), la notion d'invariance consiste à préciser le comportement des coefficients canoniques non nuls, obtenus par l'analyse canonique linéaire (ACL) de deux variables aléatoires euclidiennes  $X_1$  et  $X_2$ , lorsque ses derniers sont transformées séparément par des applications linéaires. Ce problème a été abordé précisément dans la Proposition 4.2 (p. 380) et le Corollaire 4.1 (p. 382) de *Dauxois et Nkiet* (1997a). Ils montrent que pour tout opérateur linéaire inversible  $A_1$  et  $A_2$ , les coefficients canoniques non nuls du couple  $(X_1, X_2)$  sont invariants par la transformation  $(X_1, X_2) \mapsto (A_1 X_1, A_2 X_2)$ . Dans la continuité de leur travaux, *Dauxois et al.* (2004) proposent une étude asymptotique et quelques applications permettant ainsi de donner tous les outils à l'étude du problème d'invariance. Pour plus de lecture, nous pouvons aussi consulter *Nkiet* (2003, 2005).



## 1.1 Le problème d'invariance et sélection des variables

---

Cette notion d'invariance serait encore plus intéressante si l'on montrait que la transformation appliquée au couple  $(X_1, X_2)$  conduise par exemple, à un couple de variables aléatoires euclidiennes à valeurs dans des espaces de dimensions plus petites que celles des espaces initiaux. Ainsi, il est question de savoir si l'on peut donner une condition nécessaire et suffisante pour que l'ACL appliquée à la transformation du couple  $(X_1, X_2)$  fournisse autant "d'informations" que celle de  $(X_1, X_2)$ .

Vue de cette façon, ce problème relève de la sélection des variables, qui est une conséquence du problème d'invariance. La question posée précédemment reviendrait alors à se demander dans quelle condition on ajoute aucune information supplémentaire à l'ACL de  $A_1X_1$  et  $A_2X_2$ . Une condition nécessaire et suffisante se déduit de la Proposition 4.2 et du Corollaire 4.1 de *Dauxois et Nkiet* (1997a). *Fujikoshi et Yasunori* (1982) posaient et résolvaient déjà cette problématique sous le nom "effet additionnel de variables" et un peu plus tard, *Tamoudi* (1987) l'aborde dans le cadre hilbertien. On suppose en général que les opérateurs de covariance associés aux variables aléatoires  $A_1X_1$  et  $A_2X_2$  notés respectivement  $V_1$  et  $V_2$  sont inversibles, c'est-à-dire que les composantes de  $A_1X_1$  et  $A_2X_2$  sont linéairement indépendantes.

*Nkiet* (2001) aborde cette problématique à partir d'un critère de sélection qui repose sur la notion d'invariance et dont les hypothèses sont celles énoncées dans le cadre de l'ACL. L'intérêt que nous portons sur ce critère comme nous le verrons dans la section 1.2, vient du fait que :

- théoriquement, le critère est facile à calculer, car défini à partir de la norme de la différence de deux opérateurs,
- par définition, le critère mesure la perte d'information lorsqu'un sous-ensemble de variable est sélectionné,
- le moment d'ordre quatre des variables du modèle doit être fini. Ce qui permet bien-sûr de le programmer aisément dans le cas d'une étude numérique,
- plus intéressant c'est que le critère est valable quelque soit la loi de probabilité suivie par la variable à sélectionner.

Dans la section 1.2, nous présentons le critère et ses avantages puis, nous donnons quelques propriétés et la méthode de sélection.

## 1.2 Approche de Nkiet (2001) : Méthode directe

### 1.2.1 Le modèle

On considère le modèle linéaire suivant :

$$Y = \sum_{i=1}^p \alpha_i X_i + \varepsilon, \quad (1.1)$$

où  $p \geq 2$ ,  $Y$ ,  $\varepsilon$  et les  $X_i$ , ( $i = 1, \dots, p$ ) sont des variables aléatoires réelles définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , et les  $\alpha_i$  ( $i = 1, \dots, p$ ) sont des réels. On suppose que les variables aléatoires  $\varepsilon$  et  $X$  sont indépendantes. Sans perte de généralité, on suppose en plus que les variables aléatoires  $X$  et  $Y$  sont centrées et que les  $X_1, \dots, X_p$  sont linéairement indépendantes. Cette dernière hypothèse assure l'inversibilité de l'opérateur de covariance  $V_1$  associé à la variable  $X$ .

Dans ce genre de modèle, on a souvent utilisé le critère de [Thompson \(1978\)](#) pour effectuer la sélection des variables, mais [Zheng et Low \(1997\)](#) ont montré que ce critère conduit à une méthode qui n'est généralement pas convergente. C'est dans cette optique qu'ils ont proposé une modification de ce critère pour obtenir une méthode convergente. Les propriétés justifiant l'utilisation de ce critère sont obtenues sous les hypothèses de normalité du vecteur aléatoire  $(Y, X_1, \dots, X_p)$ . Le problème de robustesse des méthodes utilisées pour obtenir la convergence du critère se pose également. C'est ainsi que [Nkiet \(2001\)](#) introduit un critère de sélection dont les propriétés ne sont pas assujetties à une quelconque hypothèse de loi.

### 1.2.2 Critère de sélection et propriétés

Notons par  $I$ , l'ensemble des indices des variables aléatoires candidates à la sélection. Soient  $I_0$  un sous-ensemble de  $I$  tel que les coefficients  $\alpha_i$  sont nuls et  $I_1$  le sous-ensemble de  $I$  tel que les coefficients  $\alpha_i$  ( $i \in I$ ) sont non nuls, tels que :

$$I = \{1, \dots, p\}, \quad I_0 = \{i \in I | \alpha_i = 0\} \quad \text{et} \quad I_1 = \{i \in I | \alpha_i \neq 0\}.$$

## 1.2 Approche de Nkiet (2001) : Méthode directe

---

La définition des sous-ensembles  $I_0$  et  $I_1$  montre respectivement que si les coefficients  $\alpha_i$  ( $i = 1, \dots, p$ ) sont nuls, alors les variables aléatoires  $X_i$  associées aux coefficients  $\alpha_i$  ne sont pas liées à la variable aléatoire réelle  $Y$  dans le modèle (1.1). Par contre si les coefficients  $\alpha_i$  ( $i = 1, \dots, p$ ) sont non nuls, alors les variables aléatoires  $X_i$  associées aux coefficients  $\alpha_i$  sont liées et expliquent la variable aléatoire  $Y$ . On suppose dans la suite que le sous-ensemble  $I_1$  est non vide.

On s'intéresse à la sélection des variables dans l'équation (1.1), c'est-à-dire la caractérisation et l'estimation du sous-ensemble  $I_1$  sur la base d'un échantillon i.i.d  $(Y_k, X^k)_{1 \leq k \leq n}$  du couple  $(Y, X)$ , où  $X$  est un vecteur aléatoire continu de la forme :

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}.$$

**Nkiet (2001)** développe sa méthode de sélection autour d'un critère qui repose sur la notion d'invariance, introduit dans le cadre de l'analyse canonique linéaire (ACL) (Section 1.1), à travers la transformation linéaire  $(Y, X) \rightarrow (Y, A_K X)$ , où  $K$  est un sous-ensemble de  $I$  et  $A_K$  un opérateur linéaire encore appelé opérateur de sélection, défini de  $\mathbb{R}^p$  vers  $\mathbb{R}^{\text{card}(K)}$  tel que pour tout  $x = (x_1, \dots, x_p) \in \mathbb{R}^p$ , on a  $A_K x = (x_i)_{i \in K} \in \mathbb{R}^{\text{card}(K)}$ . La seule hypothèse faite sur la loi de  $(Y, X)$  est que  $\mathbb{E}(Y^4) < +\infty$  et  $\mathbb{E}(\|X\|_{\mathbb{R}^p}^4) < +\infty$ . Ces hypothèses traduisent simplement le fait que les variables aléatoires  $Y$  et  $X$  sont dans  $L^4(\Omega, \mathcal{A}, P)$ .

Le critère de sélection introduit par **Nkiet (2001)** est vu comme une mesure de perte d'information, obtenue en sélectionnant un sous-ensemble de variables. C'est-à-dire qu'il permet de mesurer la quantité d'information perdue lorsqu'un sous-ensemble de variables est sélectionné. Il est défini par l'équation suivante :

$$C_K = \|V_{12} - V_1 \Pi_K V_{12}\|, \quad (1.2)$$

où  $V_{12}$  et  $V_1$ , définis par  $V_{12} = \mathbb{E}(YX)$  et  $V_1 = \mathbb{E}(X \otimes X)$  sont respectivement l'opérateur de covariance croisé des variables aléatoires  $X, Y$  et l'opérateur de covariance lié au vecteur aléatoire  $X$ . Le terme  $\Pi_K$  défini par  $\Pi_K = A_K^* (A_K V_1 A_K^*)^{-1} A_K$  est un opérateur linéaire, plus précisément un projecteur dans le sous-ensemble  $K$ . L'opérateur linéaire  $A_K^*$  désigne l'adjoint de l'opérateur  $A_K$  et  $\otimes$  désigne le produit tensoriel vectoriel défini de la façon

---

## 1.2 Approche de Nkiet (2001) : Méthode directe

---

suivante : soit  $E$  et  $F$  deux espaces euclidiens et  $(u, v)$  un couple de  $E \times F$ , le produit tensoriel  $u \otimes v$  est une application linéaire définie de  $E$  vers  $F$  tel que

$$\forall h \in E, (u \otimes v)(h) = \langle u, h \rangle_E v, \text{ où } \langle \cdot, \cdot \rangle_E \text{ désigne le produit scalaire dans } E.$$

L'idée séduisante de ce critère est qu'il est mathématiquement simple et peut être facilement programmé. Il est défini à partir de la trace de la différence entre les opérateurs de covariance  $V_{12}$  d'un coté, et le produit entre  $V_1$ , le projecteur  $\Pi_K$  et  $V_{12}$ . De plus, l'utilisation de ce dernier nécessite deux hypothèses fondamentales et très réaliste à savoir :

1. Les moments d'ordre quatre des variables  $X$  et  $Y$  sont supposés finis ;
2. Les  $p$  variables (ou colonnes du tableau  $X$ ) sont supposées linéairement indépendantes, pour assurer l'inverse de l'opérateur  $V_1$ .

L'importance de ce critère réside dans sa propriété de caractérisation. Nous l'énonçons dans le Lemme 1 qui suit.

**Lemme 1**  $I_1 \subset K \Leftrightarrow C_K = 0$ .

Ce lemme traduit le fait que  $I_1$  est le sous-ensemble minimal de  $I$  au sens de l'inclusion qui vérifie  $C_K = 0$ . Cette caractérisation nous laisse penser qu'il est possible d'utiliser ce critère pour effectuer la sélection des variables. En effet, il nous suffit de déterminer le plus petit sous-ensemble de  $I$  contenu dans  $K$  qui vérifie le Lemme 1. Cependant,  $C_K$  est inconnu en pratique, il est donc nécessaire de l'estimer.

Considérons un échantillon i.i.d  $(Y_{(k)}, X^{(k)})_{1 \leq k \leq n}$  de  $(Y, X)$ , on a :

$$\bar{X}^{(n)} = n^{-1} \sum_{k=1}^n X^{(k)}, \quad \bar{Y}^{(n)} = n^{-1} \sum_{k=1}^n Y_{(k)}.$$

On définit les opérateurs de covariances empiriques suivants :

$$\hat{V}_1^{(n)} = n^{-1} \sum_{k=1}^n \left( X^{(k)} - \bar{X}^{(n)} \right) \otimes \left( X^{(k)} - \bar{X}^{(n)} \right),$$


---

---

1.2 Approche de Nkiet (2001) : Méthode directe

---

et

$$\widehat{V}_{12}^{(n)} = n^{-1} \sum_{k=1}^n \left( Y^{(k)} - \overline{Y}^{(n)} \right) \left( X^{(k)} - \overline{X}^{(n)} \right).$$

Ainsi, pour tout  $K \subset I$ , un estimateur du critère  $C_K$  est donné par :

$$\widehat{C}_K^{(n)} = \|\widehat{V}_{12}^{(n)} - \widehat{V}_1^{(n)} \widehat{\Pi}_K^{(n)} \widehat{V}_{12}^{(n)}\|, \quad (1.3)$$

où  $\widehat{\Pi}_K^{(n)} = A_K^* (A_K \widehat{V}_1^{(n)} A_K^*)^{-1} A_K$ . Le Lemme 2 donne les propriétés de convergence de l'estimateur  $\widehat{C}_K^{(n)}$ .

**Lemme 2** Pour tout  $K \subset I$ , on a :

$$\sqrt{n} \widehat{C}_K^{(n)} = \|\widehat{\Psi}_K^{(n)}(\widehat{H}^{(n)}) + \sqrt{n} \xi_K\|,$$

où  $(\widehat{\Psi}_K^{(n)})_{n \in \mathbb{N}^*}$  est une suite de variables aléatoires dans  $\mathcal{L}(\mathcal{L}(\mathbb{R}^{p+q}), \mathbb{R}^p)$  qui converge presque sûrement vers l'opérateur  $\Psi_K$  à valeur dans  $\mathcal{L}(\mathcal{L}(\mathbb{R}^{p+q}), \mathbb{R}^p)$  et définie par :

$$\forall T \in \mathcal{L}(\mathbb{R}^{p+q}), \Psi_K(T) = \pi_2(T) - \pi_1(T) \Pi_K V_{12} + V_1 \Pi_K \pi_1(T) \Pi_K V_{12} - V_1 \Pi_K \pi_2(T)$$

et  $(\widehat{H}^{(n)})_{n \in \mathbb{N}^*}$  une suite de variables aléatoires définie dans  $\mathcal{L}(\mathbb{R}^{p+q})$  qui converge en loi vers  $H$ , de loi normale centrée et d'opérateur de covariance

$$\Lambda = \mathbb{E} \left( \left( Z^{\otimes 2} - \mathbb{E}(Z^{\otimes 2}) \right)^{\tilde{\otimes} 2} \right),$$

où  $Z$  est une variable aléatoire dans  $\mathbb{R}^{p+q}$  telle que

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix},$$

et  $\tilde{\otimes}$  est le produit tensoriel d'éléments de  $\mathcal{L}(\mathbb{R}^{p+q})$ , relatif au produit scalaire  $\langle A, B \rangle_2 = \text{tr}(AB^*)$ ,  $\xi_K = V_{12} - V_1 \Pi_K V_{12}$ ,  $\pi_1$  (resp.  $\pi_2$ ) est un projecteur orthogonal défini  $\mathcal{L}(\mathbb{R}^{p+q})$  vers  $\mathcal{L}(\mathbb{R}^p)$  (resp.  $\mathcal{L}(\mathbb{R}^{p+q})$  vers  $\mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$ ).

### 1.2.3 Sélection des variables

Partant du modèle (1.1) et du sous-ensemble  $I_1$ , nous avons  $2^p - 1$  sous-ensembles de variables sont potentiellement candidates à la sélection. Si  $p$  est grand, la détermination du sous-ensemble  $I_1$  devient très laborieux, car le nombre de sous-ensemble à tester serait dans ce cas élevé. La procédure de sélection proposée par [Nkiet \(2001\)](#) permet de réduire significativement le nombre d'étapes pour déterminer le sous-ensemble  $I_1$ . Elle se présente de la façon suivante :

Soit  $i$  un indice de l'ensemble  $I$ , et on pose  $K_i = I - \{i\}$ . A partir du Lemme (1), il vient que  $C_{K_i} = 0$ , si et seulement si  $i \notin I_1$ . Alors une variable  $X_i$  est adéquate à la sélection si  $C_{K_i} > 0$ , et une autre variable  $X_j$  ( $j \neq i$ ) est plus adéquate à la sélection que  $X_i$  si  $C_{K_j} > C_{K_i}$ . En rangeant les  $C_{K_i}$  dans l'ordre décroissant, on considère la permutation aléatoire  $\sigma$  de  $I$  telles que :

$$\begin{cases} (A_1) : C_{K_{\sigma(1)}} \geq C_{K_{\sigma(2)}} \geq \dots \geq C_{K_{\sigma(p)}}, \\ (A_2) : \text{si } C_{K_{\sigma(i)}} = C_{K_{\sigma(j)}}, \text{ et si } i < j \text{ alors } \sigma(i) < \sigma(j). \end{cases}$$

Comme le sous-ensemble  $I_1$  est non vide, on considère l'entier  $p_0 \in I$ , qui satisfait :

$$\begin{cases} (A_3) : p_0 = p \text{ si } I_1 = I, \\ (A_4) : C_{K_{\sigma(1)}} \geq C_{K_{\sigma(2)}} \geq \dots \geq C_{K_{\sigma(p_0)}} > 0 = C_{K_{\sigma(p_0+1)}} = \dots = C_{K_{\sigma(p)}} \text{ si } I_1 \neq I. \end{cases}$$

Ainsi, il résulte que :

$$I_1 = \{\sigma(k) / 1 \leq k \leq p_0\}.$$

La sélection des variables se réduit ainsi à l'estimation de la permutation aléatoire  $\sigma$  et de l'entier  $p_0$ .

### 1.2.4 Estimation des paramètres $\sigma$ et $p_0$

Pour estimer  $\sigma$ , on considère une suite de fonction  $(f_n)_{n \in \mathbb{N}^*}$  définie de  $I$  vers  $\mathbb{R}_+$  et vérifiant les hypothèses suivantes :

$$\begin{cases} (H_1) : \forall \ell \in I, \lim_{n \rightarrow \infty} f_n(\ell) = 0, \\ (H_2) : \exists \alpha > 0 \text{ tel que } \alpha < \frac{1}{2}, \exists f : I \rightarrow \mathbb{R}_+ \text{ strictement décroissante tel que} \end{cases}$$

$$\forall \ell \in I, \lim_{n \rightarrow \infty} n^\alpha f_n(\ell) = f(\ell).$$

et pour tout  $\ell \in I$ , on pose :

$$\widehat{\psi}_\ell^{(n)} = \widehat{C}_{K_\ell}^{(n)} + f_n(\ell).$$

Un estimateur de la permutation aléatoire  $\sigma$  est obtenu en rangeant les opérateurs  $(\widehat{\psi}_\ell^{(n)})_{\ell \in I}$  dans l'ordre décroissant :

$$\widehat{\psi}_{\widehat{\sigma}^{(n)}(1)}^{(n)} \geq \widehat{\psi}_{\widehat{\sigma}^{(n)}(2)}^{(n)} \geq \dots \geq \widehat{\psi}_{\widehat{\sigma}^{(n)}(p)}^{(n)},$$

de telle sorte que les ex-aequo sont rangés dans l'ordre croissant des indices correspondants, c'est-à-dire, si  $\widehat{\psi}_{\widehat{\sigma}^{(n)}(i)}^{(n)} = \widehat{\psi}_{\widehat{\sigma}^{(n)}(j)}^{(n)}$  avec  $i < j$  alors  $\widehat{\sigma}^{(n)}(i) < \widehat{\sigma}^{(n)}(j)$ .

**Théorème 1** *Sous les hypothèses  $(H_1)$  et  $(H_2)$ , on a :*

$$\lim_{n \rightarrow +\infty} \mathbb{P}(\widehat{\sigma}^{(n)} = \sigma) = 1.$$

Pour estimer  $p_0$ , on considère une suite de fonction  $(g_n)_{n \in \mathbb{N}^*}$  définie de  $I$  vers  $\mathbb{R}_+$  telle que  $\forall \ell \in I$ , il existe un réel  $\beta \in ]0, 1[$ , et une fonction strictement croissante  $g$  définie de  $I$  vers  $\mathbb{R}_+$  telle que :

$$\forall \ell \in I, \lim_{n \rightarrow +\infty} (n^\beta g_n(\ell)) = g(\ell).$$

On pose :

$$\widehat{\phi}_\ell^{(n)} = \widehat{C}_{\widehat{I}_\ell^{(n)}}^{(n)} + g_n(\widehat{\sigma}^{(n)}(\ell)),$$

où

$$\widehat{I}_\ell^{(n)} = \{\widehat{\sigma}^{(n)}(k); 1 \leq k \leq \ell\}.$$

On obtient un estimateur de  $p_0$  par :

### 1.3 Approche de An et al (2013) : Méthode ASCCA

---

$$\hat{p}_0^{(n)} = \min \left\{ \ell \in I \mid \hat{\phi}_\ell^{(n)} = \min_{\ell' \in I} \left( \hat{\phi}_{\ell'}^{(n)} \right) \right\}.$$

**Théorème 2** La suite de variables aléatoires  $\left( \hat{p}_0^{(n)} \right)_{n \in \mathbb{N}^*}$  converge en probabilité vers  $p_0$  quand  $n \rightarrow +\infty$ .

La sélection des variables est ainsi effectuée en considérant le sous-ensemble

$$\hat{I}_1^{(n)} = \left\{ \hat{\sigma}^{(n)}(k); 1 \leq k \leq \hat{p}_0^{(n)} \right\}.$$

A partir des Théorèmes 1 et 2, on établit la convergence en probabilité du sous-ensemble  $\hat{I}_1^{(n)}$  par :

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left( \hat{I}_1^{(n)} = I_1 \right) = 1.$$

Ce qui établit la convergence en probabilité de la méthode.

### 1.3 Approche de An et al (2013) : Méthode ASCCA

La méthode ASCCA (Adaptive Sparse Canonical Correlation Analysis) est une approche développée dans [An et al. \(2013\)](#). Elle traite du problème de la régression "shrinkage" et de la sélection des variables en régression linéaire multiple. La méthode est basée sur la redéfinition d'un problème de régression linéaire multivarié en un problème classique d'analyse canonique, pour laquelle une formulation de la méthode des moindres carrés est proposée et l'application de la méthode du LASSO adaptatif pénalisé, tout en utilisant un critère de sélection BIC. Nous rappelons dans cette section quelques outils et résultats de cette méthode.



### 1.3.1 Définition, modèle et caractérisation

Soit  $(X^k, Y^k)_{1 \leq k \leq n}$  un échantillon i.i.d de  $(X, Y)$ , où le vecteur aléatoire  $X^k = (X_1^k, \dots, X_p^k) \in \mathbb{R}^p$  est le prédicteur et  $Y^k = (Y_1^k, \dots, Y_q^k) \in \mathbb{R}^q$  la variable réponse multivariée. On suppose que  $\mathbb{E}(X^k) = 0$  et  $\mathbb{E}(Y^k) = 0$ . On définit les opérateurs de covariance par  $\text{Cov}(X^k) = \sum_{xx} \in \mathbb{R}^{p \times p}$ ,  $\text{Cov}(Y^k) = \sum_{yy} \in \mathbb{R}^{q \times q}$  et  $\text{Cov}(X^k, Y^k) = \sum_{xy} \in \mathbb{R}^{p \times q}$ . De plus, on suppose que les matrices  $\sum_{xx}$  et  $\sum_{yy}$  sont définies positives. La relation entre les variables aléatoires  $X^k$  et  $Y^k$  est donnée par le modèle linéaire suivant :

$$Y^k = B^T X^k + \varepsilon^k, \quad (1.4)$$

où  $\varepsilon^k = (\varepsilon_1^k, \dots, \varepsilon_q^k)^T \in \mathbb{R}^q$  est la variable aléatoire qui représente l'erreur du modèle et  $B = (b_{jk}) \in \mathbb{R}^{p \times q}$  la matrice de coefficients réels.

Notons par  $\alpha_j = (b_{j1}, \dots, b_{jq})^T \in \mathbb{R}^q$  la  $j^{\text{ième}}$  ligne de la matrice  $B$  et  $\beta_k = (b_{1k}, \dots, b_{pk})^T \in \mathbb{R}^p$  la  $k^{\text{ième}}$  colonne de  $B$ . On peut alors écrire  $B$  sous la forme  $B = (\beta_1, \dots, \beta_q)^T = (\alpha_1, \dots, \alpha_p)^T \in \mathbb{R}^{p \times q}$ . Notons par  $B_0$ ,  $\alpha_{0j}$  et  $\beta_{0k}$  les vraies valeurs de  $B$ ,  $\alpha_j$  et  $\beta_k$ , respectivement. On sait que seuls les prédicteurs  $X^k$  dont les coefficients  $\|\alpha_{0j}\|$  sont non nuls, sont liés à la variable aléatoire réelle  $Y^k$ , où  $\|\cdot\|$  désigne la norme euclidienne classique. On définit par  $\mathcal{M}_F$  "Predictor Full Model" (PFM), l'ensemble d'indices du vecteur aléatoire  $X^k$ , et par  $\mathcal{M}_T$  "Predictor True Model" (PTM), le sous-ensemble des vecteurs dont les coefficients  $\|\alpha_{0j}\|$  sont non nuls :

$$\mathcal{M}_F = \{1, \dots, p\} \quad \text{et} \quad \mathcal{M}_T = \{1 \leq j \leq p; \|\alpha_{0j}\| > 0\}.$$

Nous donnons maintenant dans ce qui suit, une caractérisation du sous-ensemble du vecteur  $Y^k$ . Notons par  $\mathcal{N}_F$  (RFM), l'ensemble d'indices des vecteurs  $Y^k$  avec :

$$\mathcal{N}_F = \{1, \dots, q\}.$$

Soit  $\mathcal{N}$  un sous-ensemble de  $\mathcal{N}_F$ , et on note par  $Y_{(\mathcal{N})}^k$  le sous-vecteur de  $Y^k$  correspondant à  $\mathcal{N}$ , c'est-à-dire le vecteur  $Y^k$  défini sur la base des indices du sous ensemble  $\mathcal{N}$ . On note par  $\mathcal{N}^c = \mathcal{N}_F \setminus \mathcal{N}$ , l'ensemble des indices de  $\mathcal{N}_F$  privé de  $\mathcal{N}$ . Le sous-ensemble  $\mathcal{N}$  est appelé "Sufficient Response Model" (SRM) si les conditions de distribution de  $Y_{(\mathcal{N}^c)}^k | (Y_{(\mathcal{N})}^k, X^k)$

### 1.3 Approche de An et al (2013) : Méthode ASCCA

sont les mêmes que  $Y_{(\mathcal{N}^c)}^k | Y_{(\mathcal{N})}^k$ . Remarquons qu'il existe toujours un (SRM) car l'ensemble  $\mathcal{N}_F$  est aussi un (SRM). On définit par  $\mathcal{N}_T$  l'intersection de tous les (SRM), et sous certaines conditions de régularité, on montre que  $\mathcal{N}_T$  est aussi un (SRM) et est le plus petit de (SRM), il est appelé "Response True Model" (RTM). Ainsi, l'objectif de la méthode ASCCA est d'identifier non seulement le sous-ensemble  $\mathcal{M}_T$  "Predictor True Model" (PTM) mais aussi le sous-ensemble  $\mathcal{N}_T$  (RTM).

#### 1.3.2 L'analyse canonique des corrélations (ACC)

Comme on peut le constater, la définition du sous-ensemble  $\mathcal{N}_T$  est très intuitive. Ceci s'explique par le fait qu'une variable réponse adéquate ou non à la sélection n'est pas explicitement définie à travers un paramètre (par exemple le coefficient de régression  $B_0$ ). Alors, identifier le sous-ensemble  $\mathcal{N}_T$  paraît plus difficile qu'identifier le sous-ensemble  $\mathcal{M}_T$ . Ainsi [An et al. \(2013\)](#) propose une approche pour pallier à ce problème.

Une variable réponse  $Y^k$  ou un prédicteur  $X^k$  est pertinent (adéquat à la sélection), si cette variable est étroitement liée au vecteur propre, issu de l'opérateur défini à partir des matrices de corrélation des vecteurs aléatoires  $X^k$  et  $Y^k$ .

Soit  $K = \min \{p, q\}$ , alors  $(\mu_k^T X^i, v_k^T Y^i)$ , où  $\mu_k = (\mu_{k1}, \dots, \mu_{kp})^T \in \mathbb{R}^p$  et  $v_k = (v_{k1}, \dots, v_{kq})^T \in \mathbb{R}^q$  forme la  $k^{ième}$  paire canonique ( $1 \leq k \leq K, 1 \leq i \leq n$ ).  $\mu_k$  et  $v_k$  sont définis dans [Johnson et Wichern \(2003\)](#), à partir des matrices de corrélation des vecteurs  $X^k, Y^k$ . Ces vecteurs propres maximisent la valeur propre  $\lambda_k = \mu_k^T \Sigma_{xy} v_k$ . On s'intéresse uniquement aux coefficients de corrélations canoniques positifs (c'est-à-dire  $\lambda_k > 0$ ). On pose

$$K_0 = \{1 \leq k \leq K; \lambda_k > 0\}.$$

Les Théorèmes 3 et 4 donnent les caractéristiques de  $K_0$  et des vecteurs propres  $\mu_k$  et  $v_k$ .

**Théorème 3** *Sous le modèle (1.4), supposons que  $X^k$  et  $\varepsilon^k$  sont indépendantes et de loi normale. On a  $K_0 = \text{rang}(\Sigma_{xy}) \leq \min \{|\mathcal{M}_T|, |\mathcal{N}_T|\}$ , où  $\text{rang}(\Sigma_{xy})$  désigne le rang de la matrice de covariance  $\Sigma_{xy}$ , et  $|\cdot|$  le cardinal des sous-ensembles  $\mathcal{M}_T$  et  $\mathcal{N}_T$ .*

**Théorème 4** *Posons  $\tilde{\mu}_j = (\mu_{1j}, \dots, \mu_{K_0j})^T \in \mathbb{R}^{K_0}$  et  $\tilde{v}_j = (v_{1j}, \dots, v_{K_0j})^T \in \mathbb{R}^{K_0}$ . Sous le*

### 1.3 Approche de An et al (2013) : Méthode ASCCA

modèle (1.4), on suppose que les vecteurs aléatoires  $X^k$  et  $\varepsilon^k$  sont indépendants et de loi normale, et on a :

1.  $\|\tilde{\mu}_j\| = 0$  pour tout  $j \notin \mathcal{M}_T$ ,  $\|\tilde{v}_j\| = 0$  pour tout  $j \notin \mathcal{N}_T$  ;
2.  $\|\tilde{\mu}_j\| > 0$  pour tout  $j \in \mathcal{M}_T$ ,  $\|\tilde{v}_j\| > 0$  pour tout  $j \in \mathcal{N}_T$ .

#### 1.3.3 Procédure de sélection : Formulation par les moindres carrés

Le Théorème 4 montre que les sous-ensembles  $\mathcal{M}_T$  et  $\mathcal{N}_T$  sont clairement identifiés par inférence à partir des charges canoniques  $\mu_k$  et  $v_k$  ( $1 \leq k \leq K$ ). Soit  $(X^i, Y^i)_{1 \leq i \leq n}$  un échantillon i.i.d de  $(X, Y)$ . On a :

$$\widehat{\Sigma}_{xx} = n^{-1} \sum_{i=1}^n X^i X^{iT}, \quad \widehat{\Sigma}_{xy} = n^{-1} \sum_{i=1}^n X^i Y^{iT}, \quad \widehat{\Sigma}_{yy} = n^{-1} \sum_{i=1}^n Y^i Y^{iT}.$$

Posons  $\widehat{T} = \widehat{\Sigma}_{xx}^{-1/2} \widehat{\Sigma}_{xy} \widehat{\Sigma}_{yy}^{-1/2}$ , et on note par  $\widehat{T} = \widehat{U} \widehat{D} \widehat{V}^T$  sa décomposition en valeurs singulières, où  $\widehat{U} = (\widehat{u}_1, \dots, \widehat{u}_k) \in \mathbb{R}^{p \times K}$  et  $\widehat{V} = (\widehat{v}_1, \dots, \widehat{v}_k) \in \mathbb{R}^{q \times K}$  sont des matrices colonnes orthogonales et  $\widehat{D} = \text{diag} \{ \widehat{\lambda}_1, \dots, \widehat{\lambda}_k \}$  une matrice diagonale, où  $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \dots \geq \widehat{\lambda}_k$  sont des valeurs singulières de  $\widehat{T}$ . On estime les vecteurs propres  $\mu_k$  et  $v_k$  par :

$$\widehat{\mu}_k = \widehat{\Sigma}_{xx} \widehat{u}_k \quad \text{et} \quad \widehat{v}_k = \widehat{\Sigma}_{yy} \widehat{v}_k.$$

En utilisant la "méthode shrinkage", puis grâce aux fonctions de moindres carrés, [An et al. \(2013\)](#) déterminent les vecteurs propres  $\mu_k$  et  $v_k$  par :

- (1)  $\mu_k \approx \mu_k^* = \arg \min_{\mu} \mathbb{E} \left( \mu^T X - v_k^T Y \right)^2$  ;
- (2)  $v_k \approx v_k^* = \arg \min_{v} \mathbb{E} \left( \mu_k^T X - v^T Y \right)^2$ .

Ainsi les relations (1) et (2) montrent que les vecteurs propres  $\mu_k$  et  $v_k$  sont proportionnels à  $\mu_k^*$  et  $v_k^*$ . En dimension finie, on estime  $\mu_k^*$  et  $v_k^*$  par :

- $\widehat{\mu}_k^* = \arg \min_{\mu} \sum_{i=1}^n \left( X_i^T \mu - Y_i^T \widehat{v}_k \right)^2$  ;
- $\widehat{v}_k^* = \arg \min_{v} \sum_{i=1}^n \left( X_i^T \widehat{\mu}_k - Y_i^T v \right)^2$ .

### 1.3 Approche de An et al (2013) : Méthode ASCCA

Une autre solution des relations (1) et (2) (des vecteurs propres  $\mu_k^*$  et  $\nu_k^*$ ) est obtenue grâce aux fonctions de moindres carrés pénalisées :

$$(3) \quad Q_\lambda^a(\mu) = \sum_{i=1}^n \left( X_i^T \mu - Y_i^T \hat{\nu}_k \right)^2 + \lambda \sum_{j=1}^p |\mu_j| / |\hat{\mu}_{kj}|;$$

$$(4) \quad Q_\tau^b(\nu) = \sum_{i=1}^n \left( X_i^T \hat{\mu}_k - Y_i^T \nu \right)^2 + \tau \sum_{j=1}^q |\nu_j| / |\hat{\nu}_{kj}|;$$

où  $\mu_j$  ( $1 \leq j \leq p$ ) et  $\nu_j$  ( $1 \leq j \leq q$ ) sont respectivement la  $j^{\text{ième}}$  composante de  $\mu$  et  $\nu$ . En regardant attentivement les relations (3) et (4), on voit des fonctions adaptées de type LASSO. Un estimateur par la méthode "shrinkage" des vecteurs  $\mu_k^*$  et  $\nu_k^*$  est donné par les équations suivantes :

$$(5) \quad \hat{\mu}_{\lambda,k}^* = \arg \min_{\mu} Q_\lambda^a(\mu);$$

$$(6) \quad \hat{\nu}_{\tau,k}^* = \arg \min_{\nu} Q_\tau^b(\nu).$$

Enfin, les estimateurs par la méthode "shrinkage" des vecteurs propres  $\mu_k$  et  $\nu_k$  sont donnés par :

$$(7) \quad \hat{\mu}_{\lambda,k} = \hat{\mu}_{\lambda,k}^* / \sqrt{\hat{\mu}_{\lambda,k}^{*T} \widehat{\Sigma}_{xx} \hat{\mu}_{\lambda,k}^*};$$

$$(8) \quad \hat{\nu}_{\tau,k} = \hat{\nu}_{\tau,k}^* / \sqrt{\hat{\nu}_{\tau,k}^{*T} \widehat{\Sigma}_{yy} \hat{\nu}_{\tau,k}^*}.$$

On obtient ainsi le meilleur modèle de sélection en minimisant le critère BIC suivant :

$$(9) \quad BIC_{\lambda,k}^a = \log \left\{ n^{-1} \sum_{i=1}^n \left( X_i^T \hat{\mu}_{\lambda,k}^* - Y_i^T \hat{\nu}_k \right)^2 \right\} + \hat{d}f_{\lambda,k}^a \times \frac{\log n}{n}, \quad \forall \hat{\mu}_{\lambda,k}^*;$$

$$(10) \quad BIC_{\tau,k}^b = \log \left\{ n^{-1} \sum_{i=1}^n \left( X_i^T \hat{\mu}_k - Y_i^T \hat{\nu}_{\tau,k}^* \right)^2 \right\} + \hat{d}f_{\tau,k}^b \times \frac{\log n}{n}, \quad \forall \hat{\nu}_{\tau,k}^*;$$

où  $\hat{d}f_{\lambda,k}^a$  et  $\hat{d}f_{\tau,k}^b$  est le nombre de charges non nulles respectivement dans  $\hat{\mu}_{\lambda,k}^*$  et  $\hat{\nu}_{\tau,k}^*$ . Ainsi, les sous-ensembles sélectionnés via la méthode "shrinkage" sont donnés par :

$$\widehat{\mathcal{M}} = \left\{ 1 \leq j \leq p; \hat{\mu}_{kj}^* \neq 0, 1 \leq k \leq K_0 \right\} \quad \text{et} \quad \widehat{\mathcal{N}} = \left\{ 1 \leq j \leq q; \hat{\nu}_{kj}^* \neq 0, 1 \leq k \leq K_0 \right\}.$$

$\forall 1 \leq k \leq K_0$ , posons  $A_k = \{j; \mu_{kj}^* \neq 0\}$  et  $\hat{A}_{\lambda,k} = \{j; \hat{\mu}_{kj}^* \neq 0\}$ . Les Théorèmes 5 et 6 établissent la convergence en probabilité des estimateurs  $\hat{\mu}_{kj}^*$  et  $\hat{A}_k$ .

**Théorème 5** Si  $\lambda \rightarrow +\infty$ , et que  $n^{-1/2}\lambda \rightarrow 0$ , alors  $\forall k \leq K_0$ , on a :

- (1)  $\hat{\mu}_{kj}^* - \mu_{kj}^* = O_p(n^{-1/2})$ ;
- (2)  $P(\hat{A}_{\lambda,k} = A_k) \rightarrow 1$ .

**Théorème 6**  $\forall k \leq K_0$ , on a  $P(\hat{A}_k = A_k) \rightarrow 1$ .

## 1.4 Conclusion

La méthode de [Nkiet \(2001\)](#) traite du problème de sélection pour des variables aléatoires continues. Cette méthode repose sur un critère issu du problème d'invariance de l'ACL ([Dauxois et Nkiet \(1997a\)](#)), et peut être vue comme une mesure de perte d'informations, obtenue en sélectionnant un sous-ensemble de variables. Le critère est aussi utilisé pour ranger les variables dans l'ordre décroissant en fonction de l'information qu'elle apporte, afin que le problème de sélection de variables se réduise à l'estimation d'une permutation et d'une dimension appropriées. C'est-à-dire qu'il est question ici d'estimer les paramètres  $\sigma$  et  $p_0$  du sous-ensemble  $I_1$  (Section 1.2.3). De cette approche, on peut ainsi remarquer les avantages suivants :

1. Les seules hypothèses utilisées dans cette méthode sont :  $\mathbb{E}(Y^4) < +\infty$ ,  $\mathbb{E}(\|X\|^4) < +\infty$  et  $V_1$  est supposé inversible.
2. La sélection des variables se réduit simplement à un problème d'estimation.
3. Du point 2, il en résulte que théoriquement, il y a une réduction significative du nombre d'étapes pour déterminer le sous-ensemble  $I_1$ .

En dépit des avantages relevés, il faut tout de même signaler qu'elle n'est valable que pour des modèles linéaires unidimensionnel, c'est-à-dire que  $Y$  dans le modèle (1.1) est une variable aléatoire réelle. Or, de nos jours, on utilise de plus en plus des modèles linéaires avec un vecteur aléatoire comme variable d'intérêt. Ajouter à cela, aucune étude sur les fonctions de pénalité  $f_n$ ,  $g_n$  et des paramètres  $\alpha$  et  $\beta$  n'a été proposée, afin d'établir leur influence si elle existe sur la méthode de sélection. Ainsi, dans le Chapitre 2 de cette thèse,

## 1.4 Conclusion

---

il sera question de généraliser cette méthode en utilisant un modèle linéaire multidimensionnel comme dans *An et al. (2013)*, faire une étude sur l'influence des paramètres de contrôle  $f_n$ ,  $g_n$ ,  $\alpha$  et  $\beta$  et la comparer à celle de *An et al. (2013)*.

# SÉLECTION DES VARIABLES EN RÉGRESSION LINÉAIRE MULTIDIMENSIONNELLE

## Sommaire

<b>2.1</b>	<b>Modèle et critère de sélection . . . . .</b>	<b>33</b>
2.1.1	Modèle et énoncé du problème . . . . .	33
2.1.2	Caractérisation du sous-ensemble $I_1$ . . . . .	35
2.1.3	Estimation du critère . . . . .	41
<b>2.2</b>	<b>Sélection des variables . . . . .</b>	<b>45</b>
2.2.1	Estimation des paramètres $\sigma$ et $s$ . . . . .	45
2.2.2	Propriétés de convergence . . . . .	46
<b>2.3</b>	<b>Etude numérique . . . . .</b>	<b>49</b>
2.3.1	Simulations . . . . .	50
2.3.2	Influence des fonctions de pénalité . . . . .	51
2.3.3	Influence des paramètres $\alpha$ et $\beta$ . . . . .	52
2.3.4	Choix optimal des paramètres $\alpha$ et $\beta$ . . . . .	52
2.3.5	Comparaison avec la méthode An et al. (2013) . . . . .	53
2.3.6	Application aux données réelles . . . . .	55
<b>2.4</b>	<b>Conclusion . . . . .</b>	<b>56</b>

## Introduction

Dans les études statistiques théoriques et pratiques, les méthodes de sélection des variables ont souvent considéré des modèles linéaires avec un vecteur aléatoire comme variable réponse. Dans ce contexte, la méthode de sélection de *Nkiet* (2001) n'est plus valable, car son modèle (Equation (1.1)) ne tient compte que d'une seule variable aléatoire réelle comme variable réponse. Compte tenu des qualités et des avantages de cette méthode, il y a intérêt à l'étendre en considérant un modèle plus général : le modèle linéaire multidimensionnel. Nous proposons une extension de la méthode de sélection des variables de *Nkiet* (2001) dans le cas d'un modèle linéaire multidimensionnel. Cette généralisation repose sur un critère qui permet de réduire le problème de sélection des variables à un problème d'estimation d'une permutation et d'un paramètre appelé dimension. Nous donnons un estimateur du critère de sélection et ses propriétés, puis nous établissons la convergence de notre méthode. Des simulations permettent de mettre en évidence les performances de notre approche. Plus précisément, nous étudions l'influence des paramètres de contrôles  $f_n$ ,  $g_n$ ,  $\alpha$  et  $\beta$  (Section 2.2) sur la méthode de sélection en utilisant comme critère la perte de prédiction. Nous donnons un exemple avec des données réelles et nous comparons notre méthode à celle de *An et al.* (2013).

## 2.1 Modèle et critère de sélection

Dans cette section, nous présentons le modèle linéaire multidimensionnel considéré, puis nous proposons un critère de sélection et nous établissons sa convergence.

### 2.1.1 Modèle et énoncé du problème

Considérons le modèle linéaire suivant :

$$Y = BX + \varepsilon, \tag{2.1}$$



## 2.1 Modèle et critère de sélection

---

où  $X$  et  $Y$  sont des vecteurs aléatoires réels respectivement dans  $\mathbb{R}^p$  et  $\mathbb{R}^q$  avec ( $p \geq 2, q \geq 2$ ),  $B$  est une matrice à coefficients réels dans  $\mathcal{M}_{qp}(\mathbb{R})$  et  $\varepsilon$  est un vecteur aléatoire dans  $\mathbb{R}^q$  et indépendant de  $X$  :

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}, Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_q \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_q \end{pmatrix}$$

et

$$B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ b_{q1} & b_{q2} & \cdots & b_{qp} \end{pmatrix}.$$

On suppose que les vecteurs aléatoires  $X, Y$  et  $\varepsilon$  sont centrés et que les variables  $X_1, \dots, X_p$  sont linéairement indépendantes. On peut aisément écrire l'équation (2.1) comme  $q$  modèles de régression univariés de la forme :

$$Y_j = \sum_{i=1}^p b_{ji} X_i + \varepsilon_j, \quad j = 1, \dots, q, \quad (2.2)$$

ou encore

$$Y = \sum_{i=1}^p X_i \mathbf{b}_{\bullet i} + \varepsilon, \quad (2.3)$$

avec

$$\mathbf{b}_{\bullet i} = \begin{pmatrix} b_{1i} \\ b_{2i} \\ \vdots \\ b_{qi} \end{pmatrix}.$$

Nous nous intéressons au problème de sélection des variables dans le modèle (2.1). Cela est équivalent à identifier les variables aléatoires  $X_i$  ( $i = 1, \dots, p$ ) qui ne contribuent pas à expliquer le vecteur aléatoire  $Y$ , à partir d'un échantillon i.i.d  $(X^{(k)}, Y^{(k)})_{1 \leq k \leq n}$  du couple  $(X, Y)$ . Nous prenons  $p < n$  et  $q < n$ . Nous savons qu'une variable  $X_i$  n'explique pas le vecteur  $Y$  ou n'est pas liée au vecteur  $Y$  si son vecteur de coefficients correspondants  $\mathbf{b}_{\bullet i}$  est nul. Ainsi, Posons :

$$I = \{1, \dots, p\},$$

## 2.1 Modèle et critère de sélection

---

l'ensemble des indices des variables  $X_i$  ( $i = 1, \dots, p$ ) à sélectionner. On note par  $I_0$  le sous-ensemble non vide de  $I$ , représentant les indices des variables dont les coefficients  $\mathbf{b}_{\bullet i}$  sont nuls. Soit  $\|\cdot\|_{\mathbb{R}^k}$  la norme euclidienne dans  $\mathbb{R}^k$ , on a :

$$I_0 = \{i \in I / \|\mathbf{b}_{\bullet i}\|_{\mathbb{R}^q} = 0\}.$$

Ainsi, nous considérons le problème de sélection des variables dans le modèle (2.1) comme un problème d'estimation du sous-ensemble  $I_0$ . Cela est également équivalent à estimer le sous-ensemble  $I_1 = I - I_0$ , avec

$$I_1 = \{\exists i \in I / \|\mathbf{b}_{\bullet i}\|_{\mathbb{R}^q} \neq 0\}.$$

Nous donnons maintenant, dans la section qui suit, une caractérisation du sous-ensemble  $I_1$  par le biais d'un critère de sélection.

### 2.1.2 Caractérisation du sous-ensemble $I_1$

Soit  $\otimes$  le produit tensoriel vectoriel défini comme au Chapitre 1 (Section 1.2.2). Nous supposons que  $\mathbb{E}(\|X\|_{\mathbb{R}^p}^4) < +\infty$  et  $\mathbb{E}(\|Y\|_{\mathbb{R}^q}^4) < +\infty$ , c'est-à-dire que les vecteurs aléatoires  $X$  et  $Y$  sont dans  $L^4(\Omega, \mathcal{A}, P)$ . On définit respectivement par  $V_1$  et  $V_{12}$  l'opérateur de covariance associé au vecteur aléatoire  $X$  et l'opérateur de covariance croisé associé à  $X$  et  $Y$  par :

$$V_1 = \mathbb{E}(X \otimes X) \quad \text{et} \quad V_{12} = \mathbb{E}(Y \otimes X). \quad (2.4)$$

On suppose que l'opérateur  $V_1$  est inversible.

**Remarque 2.1.1** *Tout au long de cette thèse, nous utilisons essentiellement les opérateurs de covariances, mais leur matrice relative est détaillée dans [Dauxois et al. \(1994\)](#). Soient  $u$  et  $v$  deux vecteurs respectivement dans  $\mathbb{R}^p$  et  $\mathbb{R}^q$ , la matrice relative à la base canonique de l'opérateur  $u \otimes v$  est notée  $uv^T$ , où  $v^T$  est la transposée du vecteur  $v$ . Si les expressions matricielles sont préférées aux opérateurs, alors les opérateurs de l'équation (2.4) deviennent  $V_1 = \mathbb{E}(XX^T)$  et  $V_{12} = \mathbb{E}(YX^T)$ .*

## 2.1 Modèle et critère de sélection

---

Soit  $K$ , un sous-ensemble de  $I$ . On note par  $A_K$  un opérateur encore appelé opérateur de sélection (Chapitre 1 Section 1.2.2). Notons par  $\Pi_K$  un opérateur défini par  $\Pi_K := A_K^* (A_K V_1 A_K^*)^{-1} A_K$ , où  $A_K^*$  désigne l'adjoint de l'opérateur  $A_K$ . Ainsi, on considère le critère de sélection suivant :

$$\xi_K = \|V_{12} - V_1 \Pi_K V_{12}\|, \quad (2.5)$$

où  $\|\cdot\|$  est la norme usuelle d'un opérateur donnée par  $\|A\| = \sqrt{\text{tr}(A^*A)}$ . Ce critère permet de mesurer la perte d'information lorsqu'un sous-ensemble de variables  $K$  est sélectionné. Il permet aussi de donner une caractérisation du sous-ensemble  $I_1$ , comme indiqué dans le Lemme 3.

**Lemme 3** *Pour tout sous-ensemble  $K$  de  $I$ , on a  $I_1 \subset K$  si et seulement si,  $\xi_K = 0$ .*

**Preuve:**

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé. On définit les opérateurs linéaires suivants :

$$L_1 : x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \in \mathbb{R}^p \longmapsto \sum_{i=1}^p x_i X_i \in L^2(\Omega, \mathcal{A}, P)$$

et

$$L_2 : y = \begin{pmatrix} y_1 \\ \vdots \\ y_q \end{pmatrix} \in \mathbb{R}^q \longmapsto \sum_{j=1}^q y_j Y_j \in L^2(\Omega, \mathcal{A}, P).$$

En plus,  $\forall Z \in L^2(\Omega, \mathcal{A}, P)$ , on a :

$$\begin{aligned} \langle L_1 x, Z \rangle_{L^2(\Omega, \mathcal{A}, P)} &= \mathbb{E}(Z L_1 x) \\ &= \mathbb{E}(Z \langle x, X \rangle_{\mathbb{R}^p}) \\ &= \mathbb{E}(\langle x, Z X \rangle_{\mathbb{R}^p}) \\ &= \langle x, \mathbb{E}(Z X) \rangle_{\mathbb{R}^p} \\ &= \langle x, L_1^*(Z) \rangle_{\mathbb{R}^p}. \end{aligned}$$

---

2.1 Modèle et critère de sélection

---

D'où l'on note par  $L_1^*$  l'adjoint de l'opérateur  $L_1$  défini par :

$$L_1^* : Z \in L^2(\Omega, \mathcal{A}, P) \longmapsto \mathbb{E}(ZX) \in \mathbb{R}^p.$$

De la même façon, on montre que  $L_2^*$  est l'adjoint de l'opérateur  $L_2$  défini par :

$$L_2^* : Z \in L^2(\Omega, \mathcal{A}, P) \longmapsto \mathbb{E}(ZY) \in \mathbb{R}^q.$$

Ainsi, il vient que :

$$\begin{cases} V_1 = L_1^* L_1 \\ V_{12} = L_1^* L_2. \end{cases}$$

En effet,  $\forall u \in \mathbb{R}^p$ ,

$$\begin{aligned} V_1(u) &= \mathbb{E}[(X \otimes X)(u)] \\ &= \mathbb{E}(\langle X, u \rangle X) \\ &= L_1^*(\langle X, u \rangle) \\ &= L_1^*\left(\sum_{i=1}^p u_i X_i\right) \\ &= L_1^* L_1(u), \end{aligned}$$

d'où  $V_1 = L_1^* L_1$ ,  $\forall u \in \mathbb{R}^p$ . De même  $v \in \mathbb{R}^q$ ,

$$\begin{aligned} V_{12}(v) &= \mathbb{E}[(Y \otimes X)(v)] \\ &= \mathbb{E}(\langle Y, v \rangle X) \\ &= L_1^*(\langle Y, v \rangle) \\ &= L_1^*\left(\sum_{j=1}^q v_j Y_j\right) \\ &= L_1^* L_2(v), \end{aligned}$$

d'où  $V_{12} = L_1^* L_2$ ,  $\forall v \in \mathbb{R}^q$ . Ainsi, on a :

$$V_{12} - V_1 \Pi_K V_{12} = L_1^* L_1 - L_1^* L_1 A_K^* (A_K L_1^* L_1 A_K^*)^{-1} A_K L_1^*.$$

Notons par  $R(A)$  l'image de l'opérateur  $A$ , et  $\Pi_{R(A)}$  le projecteur orthogonal sur  $R(A)$  défini par

$$\Pi_{R(A)} = A(A^* A)^{-1} A^*.$$


---

## 2.1 Modèle et critère de sélection

---

L'équation (2.5) devient :

$$\xi_K = \|L_1^* L_2 - L_1^* \Pi_{R(L_1 A_K^*)} L_2\| = \|L_1^* \Pi_{R(L_1 A_K^*)^\perp} L_2\|, \quad (2.6)$$

où  $\Pi_{R(L_1 A_K^*)} = L_1 A_1^* ((L_1 A_1^*)^* (L_1 A_1^*))^{-1} (L_1 A_1^*)^*$ , et  $R(L_1 A_K^*)^\perp$  désigne l'espace orthogonal de l'espace vectoriel  $R(L_1 A_K^*)$ . D'où l'on a :

$$\begin{aligned} \xi_K = 0 &\Leftrightarrow \|L_1^* \Pi_{R(L_1 A_K^*)^\perp} L_2\| = 0 \\ &\Leftrightarrow L_1^* \Pi_{R(L_1 A_K^*)^\perp} L_2 = 0 \end{aligned}$$

Pour tout vecteur  $\alpha = (\alpha_1, \dots, \alpha_q)^T \in \mathbb{R}^q$ , on a

$$\begin{aligned} L_2(\alpha) &= \sum_{j=1}^q \alpha_j Y_j = \sum_{j=1}^q \alpha_j \left( \sum_{i=1}^p b_{ji} X_i + \varepsilon_j \right) \\ &= \sum_{j=1}^q \sum_{i=1}^p \alpha_j b_{ji} X_i + \sum_{j=1}^q \alpha_j \varepsilon_j. \end{aligned}$$

D'où

$$L_1^* \Pi_{R(L_1 A_K^*)^\perp} L_2(\alpha) = L_1^* \Pi_{R(L_1 A_K^*)^\perp} \sum_{j=1}^q \sum_{i=1}^p \alpha_j b_{ji} X_i + \sum_{j=1}^q L_1^* \Pi_{R(L_1 A_K^*)^\perp} \alpha_j \varepsilon_j.$$

Pour tout vecteur  $u = (u_1, \dots, u_p)^T \in \mathbb{R}^p$ , on a :

$$\begin{aligned} \langle L_1(u), \alpha_j \varepsilon_j \rangle &= \sum_{i=1}^p u_i \langle X_i, \alpha_j \varepsilon_j \rangle = \sum_{i=1}^p u_i \alpha_j \mathbb{E}(X_i \varepsilon_j) \\ &= \sum_{i=1}^p u_i \alpha_j \mathbb{E}(X_i) \mathbb{E}(\varepsilon_j) = 0. \end{aligned}$$

Cela implique que  $\alpha_j \varepsilon_j \in R(L_1)^\perp$  et comme  $R(L_1)^\perp \subset R(L_1 A_K^*)^\perp$ , on a  $\alpha_j \varepsilon_j \in R(L_1 A_K^*)^\perp$ .

D'où

$$\begin{aligned} L_1^* \Pi_{R(L_1 A_K^*)^\perp} \alpha_j \varepsilon_j &= L_1^* \alpha_j \varepsilon_j = \mathbb{E}(\alpha_j \varepsilon_j X) \\ &= \alpha_j \mathbb{E}(\varepsilon_j) \mathbb{E}(X) \\ &= 0. \end{aligned}$$

## 2.1 Modèle et critère de sélection

---

Alors

$$L_1^* \Pi_{R(L_1 A_K^*)^\perp} L_2(\alpha) = L_1^* \Pi_{R(L_1 A_K^*)^\perp} \sum_{j=1}^q \sum_{i=1}^p \alpha_j b_{ji} X_i = \sum_{j=1}^q \alpha_j L_1^* \Pi_{R(L_1 A_K^*)^\perp} L_1(\mathbf{b}_{j\bullet}), \quad (2.7)$$

où

$$\mathbf{b}_{j\bullet} = \begin{pmatrix} b_{j1} \\ b_{j2} \\ \vdots \\ b_{jp} \end{pmatrix}.$$

Si  $\xi_K = 0$ , alors en considérant, pour tout  $j = 1, \dots, q$ , le vecteur  $\alpha = (0, \dots, 0, 1, 0, \dots, 0)$  de  $\mathbb{R}^q$  dont toutes les coordonnées sont nulles exceptée la  $j^{\text{ième}}$  coordonnée qui est égale 1, nous déduisons de l'équation (2.7) que  $L_1^* \Pi_{R(L_1 A_K^*)^\perp} L_1(\mathbf{b}_{j\bullet}) = 0$ . Comme pour tout opérateur  $A$ ,  $\ker(A^* A) = \ker(A)$ , alors  $\Pi_{R(L_1 A_K^*)^\perp} L_1(\mathbf{b}_{j\bullet}) = 0$ , ce qui implique que

$$L_1(\mathbf{b}_{j\bullet}) \in R(L_1 A_K^*). \quad (2.8)$$

Notons par  $\text{card}(K)$ , le cardinal de l'ensemble  $K$  et posons  $K = \{k_1, k_2, \dots, k_{\text{card}(K)}\}$ , nous déduisons de l'équation (2.8) qu'il existe un vecteur  $\beta = (\beta_1, \dots, \beta_{\text{card}(K)})^T \in \mathbb{R}^{\text{card}(K)}$  tel que  $L_1(\mathbf{b}_{i\bullet}) = L_1 A_K^* \beta$ , c'est-à-dire

$$\sum_{i=1}^p b_{ji} X_i = \sum_{\ell=1}^{\text{card}(K)} \beta_\ell X_{k_\ell},$$

ce qui est équivalent à

$$\sum_{\ell=1}^{\text{card}(K)} (b_{jk_\ell} - \beta_\ell) X_{k_\ell} + \sum_{i \in I-K} b_{ji} X_i = 0. \quad (2.9)$$

Puisque l'opérateur  $V_1$  est inversible, alors nous avons  $\ker(L_1) = \ker(L_1^* L_1) = \ker(V_1) = \{0\}$ . Comme  $X_1, \dots, X_p$  sont linéairement indépendants, l'équation (2.9) implique que pour tout  $i \in I - K$ ,  $b_{ji} = 0$ . Cette propriété étant valable pour tout  $j \in \{1, \dots, q\}$ , alors nous déduisons que  $I - K \subset I_0$ , ce qui est équivalent à  $I_1 \subset K$ . Réciproquement, nous

## 2.1 Modèle et critère de sélection

avons

$$\begin{aligned}
L_1^* \Pi_{R(L_1 A_K^*)^\perp} L_1(\mathbf{b}_{i\bullet}) &= L_1^* \Pi_{R(L_1 A_K^*)^\perp} \sum_{i=1}^p b_{ji} X_i \\
&= L_1^* \Pi_{R(L_1 A_K^*)^\perp} \left( \sum_{i \in K} b_{ji} X_i + \sum_{i \in I-K} b_{ji} X_i \right) \\
&= L_1^* \Pi_{R(L_1 A_K^*)^\perp} \left( \sum_{\ell=1}^{\text{card}(K)} b_{jk_\ell} X_{k_\ell} + \sum_{i \in I-K} b_{ji} X_i \right).
\end{aligned}$$

Si  $I_1 \subset K$ , alors  $I - K \subset I_0$  et par conséquent, pour tout  $i \in I - K$ ,  $b_{ji} = 0$ . Alors

$$\begin{aligned}
L_1^* \Pi_{R(L_1 A_K^*)^\perp} L_1(\mathbf{b}_{j\bullet}) &= L_1^* \Pi_{R(L_1 A_K^*)^\perp} \left( \sum_{\ell=1}^{\text{card}(K)} b_{jk_\ell} X_{k_\ell} \right) \\
&= L_1^* \Pi_{R(L_1 A_K^*)^\perp} L_1 A_K^*(\mathbf{b}_{j\bullet}) \\
&= 0,
\end{aligned}$$

car  $L_1 A_K^*(\mathbf{b}_{j\bullet}) \in R(L_1 A_K^*)$ . Des équations (2.6) et (2.7), nous déduisons que  $\xi_K = 0$ .  $\square$

Ce lemme permet de caractériser l'ensemble des entiers  $i$  qui appartiennent à  $I_0$ . Ainsi, si  $i \in I_0$ , cela est équivalent à  $I_1 \subset I - \{i\}$ . On déduit ainsi que l'on a  $i \in I_0$  si et seulement si,  $\xi_{K_i} = 0$  où  $K_i = I - \{i\}$ . Alors  $I_1$  est l'ensemble des éléments de  $I$  tels que  $\xi_{K_i} \neq 0$ . Nous rangeons les variables  $\xi_{K_i}$  dans l'ordre décroissant. Pour cela, on considère la permutation aléatoire  $\sigma$  de  $I$  telle que :

$$\begin{cases}
(i) \ \xi_{K\sigma(1)} \geq \xi_{K\sigma(2)} \geq \dots \geq \xi_{K\sigma(p)}; \\
(ii) \ \xi_{K\sigma(i)} = \xi_{K\sigma(j)} \text{ et } i < j; \text{ implique } \sigma(i) < \sigma(j).
\end{cases}$$

Comme le sous-ensemble  $I_1$  est supposé non vide, il va exister un entier  $s \in I$  que nous appelons *la dimension*, qui satisfait

$$\xi_{K\sigma(1)} \geq \xi_{K\sigma(2)} \geq \dots \geq \xi_{K\sigma(s)} > 0 = \xi_{K\sigma(s+1)} = \dots = \xi_{K\sigma(p)}.$$

Nous donnons dans le Lemme 4, une forme explicite du sous-ensemble  $I_1$ .

**Lemme 4**  $I_1 = \{\sigma(k) / 1 \leq k \leq s\}$ .

Ce résultat montre que le sous-ensemble  $I_1$  est entièrement caractérisé par les paramètres  $\sigma$  et  $s$ . Ainsi, l'estimation de  $I_1$  se déduit directement par l'estimation de ces paramètres.

### 2.1.3 Estimation du critère

Considérons un échantillon i.i.d  $(X^{(k)}, Y^{(k)})_{1 \leq k \leq n}$  de  $(X, Y)$ . On a :

$$\bar{X}^{(n)} = n^{-1} \sum_{k=1}^n X^{(k)}, \quad \bar{Y}^{(n)} = n^{-1} \sum_{k=1}^n Y^{(k)}.$$

On déduit ainsi les estimateurs des opérateurs de covariance  $V_1$  et  $V_{12}$  par :

$$\hat{V}_1^{(n)} = n^{-1} \sum_{k=1}^n (X^{(k)} - \bar{X}^{(n)}) \otimes (X^{(k)} - \bar{X}^{(n)}),$$

et

$$\hat{V}_{12}^{(n)} = n^{-1} \sum_{k=1}^n (Y^{(k)} - \bar{Y}^{(n)}) \otimes (X^{(k)} - \bar{X}^{(n)}).$$

Pour tout sous-ensemble  $K \subset I$ , un estimateur du critère  $\xi_K$  est donné par

$$\hat{\xi}_K^{(n)} = \|\hat{V}_{12}^{(n)} - \hat{V}_1^{(n)} \hat{\Pi}_K^{(n)} \hat{V}_{12}^{(n)}\|, \quad (2.10)$$

où  $\hat{\Pi}_K^{(n)} = A_K^* (A_K \hat{V}_1^{(n)} A_K^*)^{-1} A_K$ . Une application de la loi forte de grand nombre sur les opérateurs  $\hat{V}_{12}^{(n)}$ ,  $\hat{V}_1^{(n)}$  et  $\hat{\Pi}_K^{(n)}$  établit la convergence presque sûre du critère  $\hat{\xi}_K^{(n)}$  vers  $\xi_K$  quand  $n \rightarrow +\infty$ .

Soient  $E$  et  $F$  deux espaces vectoriels euclidiens, on note par  $\mathcal{L}(E, F)$  l'espace vectoriel des opérateurs définis de  $E$  vers  $F$ . Si  $E = F$ , on note simplement  $\mathcal{L}(E)$  au lieu de  $\mathcal{L}(E, F)$ . Posons  $E = \mathbb{R}^p$  et  $F = \mathbb{R}^q$ . On sait que tout élément  $A$  de  $\mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$  peut s'écrire sous la forme :

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

où  $A_{11} \in \mathcal{L}(\mathbb{R}^p)$ ,  $A_{12} \in \mathcal{L}(\mathbb{R}^q, \mathbb{R}^p)$ ,  $A_{21} \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$  et  $A_{22} \in \mathcal{L}(\mathbb{R}^q)$ . Puis on considère les



---

## 2.1 Modèle et critère de sélection

---

projecteurs orthogonaux suivants :

$$P_1 : A \in \mathcal{L}(\mathbb{R}^{p+q}) \mapsto A_{11} \in \mathcal{L}(\mathbb{R}^p) \text{ et } P_2 : A \in \mathcal{L}(\mathbb{R}^{p+q}) \mapsto A_{12} \in \mathcal{L}(\mathbb{R}^q, \mathbb{R}^p).$$

Dans la Proposition (2.1.1) qui suit, nous donnons les propriétés de convergence de l'estimateur  $\widehat{\xi}_K^{(n)}$ .

### Proposition 2.1.1

$$\sqrt{n} \widehat{\xi}_K^{(n)} = \|\widehat{\Psi}_K^{(n)}(\widehat{H}^{(n)}) + \sqrt{n} \delta_K\|,$$

où  $\delta_K = V_{12} - V_1 \Pi_K V_{12}$ ,  $(\widehat{\Psi}_K^{(n)})_{n \in \mathbb{N}^*}$  est une suite d'opérateurs aléatoires qui converge presque sûrement vers l'opérateur  $\Psi_K$  dans  $\mathcal{L}(\mathcal{L}(\mathbb{R}^{p+q}), \mathcal{L}(\mathbb{R}^q, \mathbb{R}^p))$  quand  $n \rightarrow +\infty$  avec

$$\Psi_K(A) = P_2(A) - P_1(A) \Pi_K V_{12} + V_1 \Pi_K P_1(A) \Pi_K V_{12} - V_1 \Pi_K P_2(A),$$

où  $P_1$  (resp.  $P_2$ ) est un projecteur orthogonal défini de  $\mathcal{L}(\mathbb{R}^{p+q})$  vers  $\mathcal{L}(\mathbb{R}^p)$

(resp.  $\mathcal{L}(\mathbb{R}^{p+q})$  vers  $\mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$ ).  $(\widehat{H}^{(n)})_{n \in \mathbb{N}^*}$  est une suite de variables aléatoires à valeurs dans  $\mathcal{L}(\mathbb{R}^{p+q})$  qui converge en loi vers la variable aléatoire  $H$  de loi normale de moyenne nulle et d'opérateur de covariance  $\Gamma$  tel que :

$$\Gamma = \mathbb{E} \left( (Z \otimes Z - V) \widetilde{\otimes} (Z \otimes Z - V) \right),$$

où  $Z$  est un vecteur aléatoire à valeurs dans  $\mathbb{R}^{p+q}$  avec

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix} \text{ et } V = \mathbb{E}(Z \otimes Z).$$

**Preuve:** On a :

$$\begin{aligned} \sqrt{n} \widehat{\xi}_K^{(n)} &= \|\sqrt{n}(\widehat{V}_{12}^{(n)} - V_{12}) - \sqrt{n}(\widehat{V}_1^{(n)} - V_1) \widehat{\Pi}_K^{(n)} \widehat{V}_{12}^{(n)} - V_1 \left( \sqrt{n}(\widehat{\Pi}_K^{(n)} - \Pi_K) \right) \widehat{V}_{12}^{(n)} \\ &\quad - V_1 \Pi_K \left( \sqrt{n}(\widehat{V}_{12}^{(n)} - V_{12}) \right) + \sqrt{n} \delta_K\|, \end{aligned}$$

## 2.1 Modèle et critère de sélection

puisque

$$\begin{aligned}
\widehat{\Pi}_K^{(n)} - \Pi_K &= A_K^* \left( (A_K \widehat{V}_1^{(n)} A_K^*)^{-1} - (A_K V_1 A_K^*)^{-1} \right) A_K \\
&= A_K^* \left( -(A_K \widehat{V}_1^{(n)} A_K^*)^{-1} (A_K \widehat{V}_1^{(n)} A_K^* - A_K V_1 A_K^*) (A_K V_1 A_K^*)^{-1} \right) A_K \\
&= -A_K^* (A_K \widehat{V}_1^{(n)} A_K^*)^{-1} (A_K \widehat{V}_1^{(n)} A_K^*) (A_K V_1 A_K^*)^{-1} A_K \\
&\quad + A_K^* (A_K \widehat{V}_1^{(n)} A_K^*)^{-1} (A_K V_1 A_K^*) (A_K V_1 A_K^*)^{-1} A_K \\
&= -\widehat{\Pi}_K^{(n)} \widehat{V}_1^{(n)} \Pi_K + \widehat{\Pi}_K^{(n)} V_1 \Pi_K \\
&= -\widehat{\Pi}_K^{(n)} (\widehat{V}_1^{(n)} - V_1) \Pi_K,
\end{aligned}$$

alors

$$\begin{aligned}
\sqrt{n} \widehat{\xi}_K^{(n)} &= \|\sqrt{n}(\widehat{V}_{12}^{(n)} - V_{12}) - \sqrt{n}(\widehat{V}_1^{(n)} - V_1) \widehat{\Pi}_K^{(n)} \widehat{V}_{12}^{(n)} \\
&\quad + V_1 \widehat{\Pi}_K^{(n)} (\sqrt{n}(\widehat{V}_1^{(n)} - V_1)) \Pi_K \widehat{V}_{12}^{(n)} \\
&\quad - V_1 \Pi_K (\sqrt{n}(\widehat{V}_{12}^{(n)} - V_{12})) + \sqrt{n} \delta_K\|.
\end{aligned} \tag{2.11}$$

Considérons le vecteur aléatoire  $Z$  à valeurs dans  $\mathbb{R}^{p+q}$  telle que :

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix}, \quad Z^{(k)} = \begin{pmatrix} X^{(k)} \\ Y^{(k)} \end{pmatrix}, \quad k = 1, \dots, n;$$

l'opérateur de covariance de  $Z$  est donné par  $V = \mathbb{E}(Z \otimes Z)$  et peut être encore écrit comme

$$V = \begin{pmatrix} V_1 & V_{12} \\ V_{21} & V_2 \end{pmatrix}, \tag{2.12}$$

où  $V_2 = \mathbb{E}(Y \otimes Y)$  et  $V_{21} = V_{12}^*$ . De plus, posons :

$$\overline{Z}^{(n)} = n^{-1} \sum_{k=1}^n Z^{(k)} \quad \text{et} \quad \widehat{V}^{(n)} = n^{-1} \sum_{k=1}^n (Z^{(k)} - \overline{Z}^{(n)}) \otimes (Z^{(k)} - \overline{Z}^{(n)}),$$

avec

$$\widehat{V}^{(n)} = \begin{pmatrix} \widehat{V}_1^{(n)} & \widehat{V}_{12}^{(n)} \\ \widehat{V}_{21}^{(n)} & \widehat{V}_2^{(n)} \end{pmatrix}, \tag{2.13}$$

## 2.1 Modèle et critère de sélection

où  $\widehat{V}_2^{(n)} = n^{-1} \sum_{k=1}^n (Y^{(k)} - \overline{Y}^{(n)}) \otimes (Y^{(k)} - \overline{Y}^{(n)})$  et  $\widehat{V}_{21}^{(n)} = (\widehat{V}_{12}^{(n)})^*$ . Nous déduisons des équations (2.11) - (2.13) que  $\sqrt{n}\widehat{\xi}_K^{(n)} = \|\widehat{\Psi}_K^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\delta_K\|$ , où  $\widehat{H}^{(n)} = \sqrt{n}(\widehat{V}^{(n)} - V)$  et  $\widehat{\Psi}_K^{(n)}$  est un opérateur aléatoire défini de  $\mathcal{L}(\mathbb{R}^{p+q})$  vers  $\mathcal{L}(\mathbb{R}^p)$  tel que  $\forall A \in \mathcal{L}(\mathbb{R}^{p+q})$ ,

$$\widehat{\Psi}_K^{(n)}(A) = P_2(A) - P_1(A)\widehat{\Pi}_K^{(n)}\widehat{V}_{12}^{(n)} + V_1\widehat{\Pi}_K^{(n)}P_1(A)\Pi_A\widehat{V}_{12}^{(n)} - V_1\Pi_K P_2(A).$$

### Convergence de l'opérateur $\widehat{\Psi}_K^{(n)}$

Considérons la norme usuelle d'opérateurs  $\|\cdot\|_\infty$  définie dans  $\mathcal{L}(E, F)$  par

$$\|A\|_\infty = \sup_{x \in E - \{0\}} \|Ax\|_F / \|x\|_E.$$

Rappelons que pour tout opérateurs  $A$  et  $B$ , on a  $\|AB\|_\infty \leq \|A\|_\infty \|B\|_\infty$ , on obtient ainsi

$$\begin{aligned} \|\widehat{\Psi}_K^{(n)}(A) - \Psi_K(A)\|_\infty &= \left\| -P_1(A) \left( \widehat{\Pi}_K^{(n)} - \Pi_K \right) \widehat{V}_{12}^{(n)} - P_1(A)\Pi_K \left( \widehat{V}_{12}^{(n)} - V_{12} \right) \right. \\ &\quad \left. + V_1 \left( \widehat{\Pi}_K^{(n)} - \Pi_K \right) P_1(A)\Pi_K \widehat{V}_{12}^{(n)} + V_1\Pi_K P_1(A)\Pi_K \left( \widehat{V}_{12}^{(n)} - V_{12} \right) \right\|_\infty \\ &\leq \|P_1(A)\|_\infty \left\{ \|\widehat{\Pi}_K^{(n)} - \Pi_K\|_\infty \|\widehat{V}_{12}^{(n)}\|_\infty + \|\Pi_K\|_\infty \|\widehat{V}_{12}^{(n)} - V_{12}\|_\infty \right. \\ &\quad \left. + \|V_1\|_\infty \|\Pi_K\|_\infty \|\widehat{\Pi}_K^{(n)} - \Pi_K\|_\infty \|\widehat{V}_{12}^{(n)}\|_\infty \right. \\ &\quad \left. + \|V_1\|_\infty \|\Pi_K\|_\infty^2 \|\widehat{V}_{12}^{(n)} - V_{12}\|_\infty \right\} \\ &\leq \left\{ \|\widehat{\Pi}_K^{(n)} - \Pi_K\|_\infty \|\widehat{V}_{12}^{(n)}\|_\infty + \|\Pi_K\|_\infty \|\widehat{V}_{12}^{(n)} - V_{12}\|_\infty \right. \\ &\quad \left. + \|V_1\|_\infty \|\Pi_K\|_\infty \|\widehat{\Pi}_K^{(n)} - \Pi_K\|_\infty \|\widehat{V}_{12}^{(n)}\|_\infty \right. \\ &\quad \left. + \|V_1\|_\infty \|\Pi_K\|_\infty^2 \|\widehat{V}_{12}^{(n)} - V_{12}\|_\infty \right\} \|P_1\|_{\infty, \infty} \|A\|_\infty, \end{aligned}$$

où  $\|T\|_{\infty, \infty} := \sup_{A \in \mathcal{L}(\mathbb{R}^{p+q}) - \{0\}} \|T(A)\|_\infty / \|A\|_\infty$ . Puisque

$$\begin{aligned} \|\widehat{\Psi}_K^{(n)} - \Psi_K\|_{\infty, \infty} &\leq \left\{ \|1 + \|V_1\|_\infty \|\Pi_K\|_\infty\| \right\} \|\widehat{V}_{12}^{(n)}\|_\infty \|\widehat{\Pi}_K^{(n)} - \Pi_K\|_\infty \|P_1\|_{\infty, \infty} \quad (2.14) \\ &\quad + \left\{ 1 + \|V_1\|_\infty \|\Pi_K\|_\infty \right\} \|\Pi_K\|_\infty \|\widehat{V}_{12}^{(n)} - V_{12}\|_\infty \|P_1\|_{\infty, \infty}, \end{aligned}$$

à partir de la loi forte des grands nombres, on a  $\widehat{V}_1^{(n)}$  (resp.  $\widehat{V}_{12}^{(n)}$ ) converge presque sûrement vers  $V_1$  (resp. vers  $V_{12}$ ), quand  $n \rightarrow +\infty$ . Par conséquent,  $\widehat{\Pi}_K^{(n)}$  converge presque sûrement vers  $\Pi_K$ , quand  $n \rightarrow +\infty$ . De l'équation (2.14), on déduit que  $\widehat{\Psi}_K^{(n)}$  converge presque sûrement vers  $\Psi_K$ , quand  $n \rightarrow +\infty$ .

### Convergence de l'opérateur $\widehat{H}^{(n)}$

On a  $\widehat{H}^{(n)} = \sqrt{n} (\widehat{V}^{(n)} - V)$ . En le développant, on peut l'écrire sous la forme

$$\widehat{H}^{(n)} = \widehat{H}_1^{(n)} + \widehat{H}_2^{(n)}$$

tels que :

$$\widehat{H}_1^{(n)} = \sqrt{n} \left( \frac{1}{n} \sum_{k=1}^n Z_k \otimes Z_k - V \right) \text{ et } \widehat{H}_2^{(n)} = \frac{1}{\sqrt{n}} \left( (\sqrt{n} \overline{Z}^{(n)}) \otimes (\sqrt{n} \overline{Z}^{(n)}) \right).$$

Le théorème limite central montre que  $\widehat{H}_1^{(n)}$  (resp.  $\sqrt{n} \overline{Z}^{(n)}$ ) converge en loi vers la variable aléatoire  $H$  (resp. vers  $U$ ), quand  $n \rightarrow +\infty$ , de loi normale centrée et d'opérateur de covariance  $\Gamma$  (resp.  $\Gamma'$ ) donné par :

$$\Gamma = \mathbb{E} \left( (Z \otimes Z - V) \tilde{\otimes} (Z \otimes Z - V) \right) \quad (\text{resp. } \Gamma' = \mathbb{E} (Z \otimes Z) ).$$

Ainsi,  $\widehat{H}_2^{(n)}$  converge en probabilité vers 0 quand  $n \rightarrow +\infty$ , et le théorème de Slutsky permet de conclure que  $\widehat{H}^{(n)}$  converge en loi vers  $H$  quand  $n \rightarrow +\infty$   $\square$

## 2.2 Sélection des variables

Le Lemme 4 montre que l'estimation du sous-ensemble  $I_1$  se déduit directement de celle des paramètres aléatoires  $\sigma$  et  $s$ . Dans cette section, nous proposons une estimation de ces paramètres et nous établissons leur propriété de convergence.

### 2.2.1 Estimation des paramètres $\sigma$ et $s$

Considérons la suite de fonctions  $(f_n)_{n \in \mathbb{N}^*}$  définies de  $I$  vers  $\mathbb{R}_+$  tels qu'il existe un réel  $\alpha \in ]0, 1/2[$  et une fonction strictement décroissante  $f : I \rightarrow \mathbb{R}_+$  tel que :

$$\forall i \in I, \quad \lim_{n \rightarrow +\infty} f_n(i) = 0 \text{ et } \lim_{n \rightarrow +\infty} (n^\alpha f_n(i)) = f(i).$$

Or, nous savons que  $K_i = I - \{i\}$ , on pose

$$\widehat{\phi}_i^{(n)} = \widehat{\xi}_{K_i}^{(n)} + f_n(i) \quad (i \in I).$$

Nous prenons comme estimateur de la permutation  $\sigma$ , la permutation aléatoire  $\widehat{\sigma}^{(n)}$  de  $I$  en rangeant l'opérateur  $\widehat{\phi}_i^{(n)}$  dans l'ordre décroissant telle que :

$$\widehat{\phi}_{\widehat{\sigma}^{(n)}(1)}^{(n)} \geq \widehat{\phi}_{\widehat{\sigma}^{(n)}(2)}^{(n)} \geq \cdots \geq \widehat{\phi}_{\widehat{\sigma}^{(n)}(p)}^{(n)}$$

et si  $\widehat{\phi}_{\widehat{\sigma}^{(n)}(i)}^{(n)} = \widehat{\phi}_{\widehat{\sigma}^{(n)}(j)}^{(n)}$  avec  $i < j$ , alors  $\widehat{\sigma}^{(n)}(i) < \widehat{\sigma}^{(n)}(j)$ .

Soit  $(g_n)_{n \in \mathbb{N}^*}$  une suite de fonctions définies de  $I$  vers  $\mathbb{R}_+$  tels qu'il existe un réel  $\beta \in ]0, 1[$  et une fonction  $g$  strictement croissante définie  $I$  vers  $\mathbb{R}_+$  telle que

$$\forall i \in I, \lim_{n \rightarrow +\infty} g_n(i) = 0, \text{ et } \lim_{n \rightarrow +\infty} (n^\beta g_n(i)) = g(i).$$

Considérons le sous-ensemble aléatoire  $\widehat{J}_i^{(n)} = \{\widehat{\sigma}^{(n)}(j); 1 \leq j \leq i\}$  et la variable aléatoire  $\widehat{\psi}_i$  telle que :

$$\widehat{\psi}_i^{(n)} = \widehat{\xi}_{\widehat{J}_i^{(n)}}^{(n)} + g_n(\widehat{\sigma}^{(n)}(i)) \quad (i \in I)$$

Nous prenons l'entier  $\widehat{s}^{(n)}$  comme estimateur de l'entier  $s$  tel que :

$$\widehat{s}^{(n)} = \min \left\{ i \in I / \widehat{\psi}_i^{(n)} = \min_{j \in I} (\widehat{\psi}_j^{(n)}) \right\}.$$

Ainsi, la sélection des variables est effectuée en considérant le sous-ensemble aléatoire  $\widehat{I}_1^{(n)}$  comme estimateur de  $I_1$ , où

$$\widehat{I}_1^{(n)} = \{\widehat{\sigma}^{(n)}(i) ; 1 \leq i \leq \widehat{s}^{(n)}\}.$$

## 2.2.2 Propriétés de convergence

Le Théorème 7 qui suit, établit la convergence en probabilité des estimateurs  $\widehat{\sigma}^{(n)}$  et  $\widehat{s}^{(n)}$ .

**Théorème 7** On a :

## 2.2 Sélection des variables

$$(i) \lim_{n \rightarrow +\infty} P\left(\widehat{\sigma}^{(n)} = \sigma\right) = 1;$$

(ii)  $\widehat{s}^{(n)}$  converge en probabilité vers  $s$ , quand  $n \rightarrow +\infty$ .

Pour la preuve, nous allons juste prouver le Lemme 5, car la preuve du Théorème 7 est similaire au Théorème 3.1 dans *Nkiet (2012)*. Soit  $r \in \mathbb{N}^*$  et  $(m_1, \dots, m_r) \in (\mathbb{N}^*)^r$  telles que  $\sum_{\ell=1}^r m_\ell = p$  et

$$\begin{aligned} \xi_{K_{\sigma(1)}} = \dots = \xi_{K_{\sigma(m_1)}} > \xi_{K_{\sigma(m_1+1)}} = \dots = \xi_{K_{\sigma(m_1+m_2)}} > \dots > \\ \xi_{K_{\sigma(m_1+m_2+\dots+m_{r-1}+1)}} = \dots = \xi_{K_{\sigma(m_1+m_2+\dots+m_r)}}. \end{aligned}$$

Posons

$$E = \{\ell \in \mathbb{N}^* / 1 \leq \ell \leq r, m_\ell \geq 2\}$$

et

$$F_\ell := \left\{ \left( \sum_{k=0}^{\ell-1} m_k \right) + 1, \dots, \left( \sum_{k=0}^{\ell} m_k \right) - 1 \right\}, \text{ avec } m_0 = 0.$$

**Lemme 5** Si  $E \neq \emptyset$ , alors pour tout  $\ell \in E$  et pour tout  $i \in F_\ell$ , la suite  $n^\alpha \left( \widehat{\xi}_{K_{\sigma(i)}}^{(n)} - \widehat{\xi}_{K_{\sigma(i+1)}}^{(n)} \right)$  converge en probabilité vers 0 quand  $n \rightarrow +\infty$ .

**Preuve:** Posons  $\gamma_\ell = \delta_{K_{\sigma(i)}} = \delta_{K_{\sigma(i+1)}}$ ; si  $\gamma_\ell = 0$  alors

$$\begin{aligned} \left| n^\alpha \left( \widehat{\xi}_{K_{\sigma(i)}}^{(n)} - \widehat{\xi}_{K_{\sigma(i+1)}}^{(n)} \right) \right| &= n^{\alpha-\frac{1}{2}} \left| \left\| \widehat{\Psi}_{K_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) \right\| - \left\| \widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) \right\| \right| \\ &\leq n^{\alpha-\frac{1}{2}} \left\| \left( \widehat{\Psi}_{K_{\sigma(i)}}^{(n)} - \widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)} \right) (\widehat{H}^{(n)}) \right\| \\ &\leq n^{\alpha-\frac{1}{2}} \left\| \widehat{\Psi}_{K_{\sigma(i)}}^{(n)} - \widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)} \right\|_\infty \left\| \widehat{H}^{(n)} \right\|. \end{aligned}$$

Puisque  $\widehat{\Psi}_{K_{\sigma(i)}}^{(n)}$  et  $\widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}$  convergent presque sûrement vers  $\Psi_{K_{\sigma(i)}}$  et  $\Psi_{K_{\sigma(i+1)}}$ , respectivement quand  $n \rightarrow +\infty$ , et comme  $\widehat{H}^{(n)}$  converge en loi vers  $H$  quand  $n \rightarrow +\infty$ , de la dernière inégalité et du fait que  $\alpha < 1/2$ ,  $n^\alpha \left( \widehat{\xi}_{K_{\sigma(i)}}^{(n)} - \widehat{\xi}_{K_{\sigma(i+1)}}^{(n)} \right)$  converge en probabilité vers 0 quand  $n \rightarrow +\infty$ .

Si  $\gamma_\ell \neq 0$ , nous avons

$$\begin{aligned}
n^\alpha \left( \widehat{\xi}_{K_{\sigma(i)}}^{(n)} - \widehat{\xi}_{K_{\sigma(i+1)}}^{(n)} \right) &= n^{\alpha-\frac{1}{2}} \left( \|\widehat{\Psi}_{K_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\delta_{K_{\sigma(i)}}\| - \|\widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\delta_{K_{\sigma(i+1)}}\| \right) \\
&= \frac{n^{\alpha-\frac{1}{2}} \left( \|\widehat{\Psi}_{K_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)})\|^2 - \|\widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)})\|^2 \right)}{\|\widehat{\Psi}_{K_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\delta_{K_{\sigma(i)}}\| + \|\widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\delta_{K_{\sigma(i+1)}}\|} \\
&+ \frac{2n^\alpha \left( \langle \delta_{K_{\sigma(i)}}, \widehat{\Psi}_{K_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) \rangle - \langle \delta_{K_{\sigma(i+1)}}, \widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) \rangle \right)}{\|\widehat{\Psi}_{K_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\delta_{K_{\sigma(i)}}\| + \|\widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\delta_{K_{\sigma(i+1)}}\|} \\
&= \frac{n^{\alpha-1} \left( \|\widehat{\Psi}_{K_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)})\|^2 - \|\widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)})\|^2 \right)}{\|n^{-\frac{1}{2}}\widehat{\Psi}_{K_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) + \delta_{K_{\sigma(i)}}\| + \|n^{-\frac{1}{2}}\widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) + \delta_{K_{\sigma(i+1)}}\|} \\
&+ \frac{2n^{\alpha-\frac{1}{2}} \left( \langle \delta_{K_{\sigma(i)}}, \widehat{\Psi}_{K_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) \rangle - \langle \delta_{K_{\sigma(i+1)}}, \widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) \rangle \right)}{\|n^{-\frac{1}{2}}\widehat{\Psi}_{K_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) + \delta_{K_{\sigma(i)}}\| + \|n^{-\frac{1}{2}}\widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) + \delta_{K_{\sigma(i+1)}}\|},
\end{aligned}$$

où  $\langle \cdot, \cdot \rangle$  est le produit scalaire défini par  $\langle A, B \rangle = \text{tr}(A^*B)$ . Nous avons

$$\begin{aligned}
&\left| n^{\alpha-1} \left( \|\widehat{\Psi}_{K_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)})\|^2 - \|\widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)})\|^2 \right) \right| \\
&\leq n^{\alpha-1} \left( \|\widehat{\Psi}_{K_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)})\|^2 + \|\widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)})\|^2 \right) \\
&\leq n^{\alpha-1} \left( \|\widehat{\Psi}_{K_{\sigma(i)}}^{(n)}\|_\infty^2 + \|\widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}\|_\infty^2 \right) \|\widehat{H}^{(n)}\|^2
\end{aligned} \tag{2.15}$$

ensuite,

$$\begin{aligned}
&\left| 2n^{\alpha-\frac{1}{2}} \left( \langle \delta_{K_{\sigma(i)}}, \widehat{\Psi}_{K_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) \rangle - \langle \delta_{K_{\sigma(i+1)}}, \widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) \rangle \right) \right| \\
&\leq 2n^{\alpha-\frac{1}{2}} \left( \left| \langle \delta_{K_{\sigma(i)}}, \widehat{\Psi}_{K_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) \rangle \right| + \left| \langle \delta_{K_{\sigma(i+1)}}, \widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) \rangle \right| \right) \\
&\leq 2n^{\alpha-\frac{1}{2}} \left( \|\delta_{K_{\sigma(i)}}\| \|\widehat{\Psi}_{K_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)})\| + \|\delta_{K_{\sigma(i+1)}}\| \|\widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)})\| \right) \\
&\leq 2n^{\alpha-\frac{1}{2}} \gamma_\ell \left( \|\widehat{\Psi}_{K_{\sigma(i)}}^{(n)}\|_\infty + \|\widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}\|_\infty \right) \|\widehat{H}^{(n)}\|.
\end{aligned} \tag{2.16}$$

Des équations (2.15) et (2.16) et leur propriété de convergence des opérateurs  $(\widehat{\Psi}_{K_{\sigma(i)}}^{(n)})$  (resp.  $\widehat{\Psi}_{K_{\sigma(i+1)}}^{(n)}$ ) et  $\widehat{H}^{(n)}$ , on conclut que la suite  $n^\alpha \left( \widehat{\xi}_{K_{\sigma(i)}}^{(n)} - \widehat{\xi}_{K_{\sigma(i+1)}}^{(n)} \right)$  converge en probabilité vers 0 quand  $n \rightarrow +\infty$ .  $\square$

Comme conséquence du Théorème 7, on obtient directement

$$\lim_{n \rightarrow +\infty} P \left( \widehat{I}_1^{(n)} = I_1 \right) = 1.$$

Ce dernier résultat établit la convergence en probabilité de notre méthode de sélection des variables dans le modèle (2.1).

## 2.3 Etude numérique

Dans cette section, nous illustrons notre méthode par une étude numérique. Plus précisément nous étudions l'influence des fonctions de pénalité  $f_n$  et  $g_n$  et des paramètres  $\alpha$ ,  $\beta$ , puis nous montrons l'efficacité de notre approche en la comparant à celle de *An et al.* (2013). Bien avant, nous définissons la méthode de validation croisée qui sera utilisée tout au long de nos simulations.

### Validation croisée

La validation croisée («*cross-validation*») est une méthode d'estimation fondée sur une technique d'échantillonnage. Trois techniques sont généralement utilisées dans la littérature : «*holdout method*», «*k-fold cross-validation*» et «*leave-one-out cross validation*».

- La première méthode est très simple, il suffit de diviser l'échantillon d'origine (de taille  $n$ ) en deux sous-échantillons. Le premier d'apprentissage (communément supérieur à 60% de l'échantillon) et le second de test. Le modèle est bâti sur l'échantillon d'apprentissage et la performance de la méthode est calculée sur l'échantillon test.
- Dans la seconde méthode, on divise de manière aléatoire l'échantillon original en  $k$  sous-échantillons de même taille. On sélectionne un des  $k$  sous-échantillons comme échantillon test et les  $(k - 1)$  autres sous-échantillons constituent l'échantillon d'apprentissage. Le modèle est bâti comme dans la première méthode sur l'échantillon d'apprentissage et la performance de la méthode est calculée sur l'échantillon test. On répète l'opération en sélectionnant un autre échantillon de validation parmi les  $(k - 1)$  échantillon qui n'ont pas encore été utilisés pour la validation du modèle. L'opération se répète  $k$ -fois pour qu'en fin de compte chaque sous-échantillon ait été utilisé exactement une fois comme ensemble de validation. A la fin, la performance de la méthode est calculée en prenant la moyenne des  $k$  sous-échantillon.
- La troisième méthode est un cas particulier de la deuxième méthode en posant  $k = n$ .



Ici, on bâti le modèle sur  $(n - 1)$  observation, puis on calcule la performance de la méthode sur la  $i - i\grave{e}me$  observation et on répète l'opération  $n$ -fois.

### 2.3.1 Simulations

Nous simulons un échantillon d'origine, de taille 50, 100, 500, que l'on divise à chaque fois en deux sous-échantillons : un échantillon d'apprentissage et un échantillon test (voir méthode de validation croisée). Sur l'échantillon d'apprentissage, nous appliquons notre méthode de sélection pour étudier l'influence des paramètres de contrôle  $f_n$ ,  $g_n$ ,  $\alpha$  et  $\beta$  sur notre méthode de sélection. L'échantillon test quant à lui, sert à comparer notre méthode à celle de [An et al. \(2013\)](#) en utilisant comme critère de performance la perte de prédiction. C'est l'écart entre le modèle initial et l'estimation du même modèle en utilisant uniquement les variables sélectionnées. Elle est donnée par la relation suivante :

$$e = \frac{1}{n} \sum_{k=1}^n \|Y^{(k)} - \hat{Y}_{\hat{I}_1^{(n)}}^{(k)}\|_{\mathbb{R}^q}^2,$$

où

$$\hat{Y}_{\hat{I}_1^{(n)}}^{(k)} = \mathbb{X}_{\hat{I}_1^{(n)}} \left( \mathbb{X}_{\hat{I}_1^{(n)}}^T \mathbb{X}_{\hat{I}_1^{(n)}} \right)^{-1} \mathbb{X}_{\hat{I}_1^{(n)}}^T Y^{(k)},$$

où  $\mathbb{X}_{\hat{I}_1^{(n)}}$  désigne une matrice de taille  $n \times p$ , contenant les variables sélectionnées.

Nous générons les données de la façon suivante : la variable aléatoire  $X^{(k)}$  est générée comme une variable aléatoire normale multivariée dans  $\mathbb{R}^7$ , de moyenne nulle et de matrice de covariance  $\text{cov}(X_i^{(k)}, X_j^{(k)}) = \rho^{|i-j|}$  pour tout  $1 \leq i, j \leq 7$ , où  $\rho$  est le paramètre de corrélation à valeurs dans l'intervalle  $[0, 1]$ . La variable réponse  $Y^{(k)}$  quant à elle est générée comme indiquée dans le modèle (2.1) avec :

$$B = \begin{pmatrix} 3 & 0 & 0 & 1.5 & 0 & 0 & 2 \\ 4 & 0 & 0 & 2.5 & 0 & 0 & -1 \\ 5 & 0 & 0 & 0.5 & 0 & 0 & 3 \\ 6 & 0 & 0 & 3.5 & 0 & 0 & 1 \\ 7 & 0 & 0 & 6.5 & 0 & 0 & 4 \end{pmatrix}.$$

L'erreur  $\varepsilon^{(k)}$  est générée comme une variable aléatoire normale multivariée dans  $\mathbb{R}^5$  de

moyenne nulle et de matrice de covariance  $\sigma^2 I_5$ , où  $I_5$  désigne la matrice identité de dimension 5.

### 2.3.2 Influence des fonctions de pénalité

On utilise un ensemble de fonctions de pénalité  $f_n$  et  $g_n$  pour étudier leur influence sur notre méthode de sélection, avec :

$$f_n(i) = n^{-0.45}f(i), \quad g_n(i) = n^{-0.6}g(i) \quad \text{avec} \quad f(i) = 1/g(i).$$

Nous prenons comme dans *Kundu et Murali (1996)* :  $g_1(x) = x$ ,  $g_2(x) = x^{0.1}$ ,  $g_3(x) = x^{0.5}$ ,  $g_4(x) = x^{0.9}$ ,  $g_5(x) = x^{10}$ ,  $g_6(x) = \ln(x)$ ,  $g_7(x) = (\ln(x))^{0.1}$ ,  $g_8(x) = (\ln(x))^{0.5}$ ,  $g_9(x) = (\ln(x))^{0.9}$ ,  $g_{10}(x) = x \ln(x)$ ,  $g_{11}(x) = (x \ln(x))^{0.1}$ ,  $g_{12}(x) = (x \ln(x))^{0.5}$ ,  $g_{13}(x) = (x \ln(x))^{0.9}$ .

Pour chaque fonction de pénalité  $g_k$ , ( $k = 1, \dots, 13$ ), on calcule la perte de prédiction liée à notre méthode, avec  $\rho = 0$  (les vecteurs colonnes de la matrice  $\mathbb{X}$  de taille  $n \times p$  ne sont pas corrélés),  $\rho = 0.5$  (les vecteurs colonnes de la matrice  $\mathbb{X}$  sont moyennement corrélés) et  $\rho = 0.95$  (les vecteurs colonnes de la matrice  $\mathbb{X}$  sont fortement corrélés), et on fixe  $\alpha = 0.45$ ,  $\beta = 0.6$  et  $\sigma = 1$ .

Les résultats sont présentés dans les Tableaux 2.1. Pour chaque valeur de  $n = 100$  ( $k = 15$ ),  $n = 300$  ( $k = 40$ ), chaque fonction  $g_k$  ( $k = 1, \dots, 13$ ) et  $\rho = 0, 0.5, 0.95$ , on remarque de façon générale que la perte de prédiction ne varie pas, sauf pour les fonctions de pénalité  $g_5$ ,  $g_{10}$  et  $g_{13}$  qui sont très élevés. Par contre, plus la corrélation  $\rho$  augmente, on observe une légère augmentation la perte de prédiction, à l'exception des fonctions  $g_5$  et  $g_{10}$ .

Ainsi, en faisant varier les fonctions de pénalités  $f_n$  et  $g_n$ , il n'y a pas de variation significative de la perte de prédiction à l'exception des fonctions  $g_5$ ,  $g_{10}$  et  $g_{13}$ . On peut aussi observer une légère augmentation de la perte de prédiction lorsque la corrélation  $\rho$  augmente excepté les fonctions  $g_5$  et  $g_{10}$ . Dans la suite, nous déterminons par la méthode de validation croisée les fonctions  $f_n$  et  $g_n$  qui minimisent la perte de prédiction de notre méthode.

### 2.3.3 Influence des paramètres $\alpha$ et $\beta$

Nous étudions maintenant l'influence des paramètres  $\alpha$  et  $\beta$  sur notre méthode de sélection. Pour cela, nous fixons :

$$g_7(i) = (\ln(i))^{0.1}, f_n(i) = 1/n^\alpha (\ln(i))^{0.1}, g_n(i) = (\ln(i))^{0.1} / n^\beta, \text{ où } f(i) = 1/(\ln(i))^{0.1}.$$

On fait varier  $\alpha$  dans  $]0, 1/2[$ , et  $\beta$  dans  $]0, 1[$ . Les résultats sont reportés dans le graphe 2.1. Pour différentes valeurs du paramètre  $\alpha$ , on observe une variation de la perte de prédiction lorsque  $\beta$  varie dans  $[0, 1]$ , cela pour tout  $n = 100, 300$ .

Ainsi, en faisant varier les paramètres  $(\alpha, \beta)$  dans  $]0, 1/2[ \times ]0, 1[$ , on observe une variation de la perte de prédiction. Il y a donc une influence des paramètres  $\alpha$  et  $\beta$  sur notre méthode de sélection. D'où intérêt à calculer les paramètres  $\alpha$  et  $\beta$  qui minimisent la perte de prédiction de notre méthode.

### 2.3.4 Choix optimal des paramètres $\alpha$ et $\beta$

Nous avons montré dans la Section (2.3.3) qu'il y a une influence des paramètres  $\alpha$  et  $\beta$  sur la perte de prédiction liée à notre méthode. Dans cette section, nous expliquons comment obtenir les paramètres  $\alpha$  et  $\beta$  qui minimisent la perte de prédiction. Pour cela, nous utilisons la méthode de validation croisée comme il suit :

1. Nous utilisons l'échantillon d'apprentissage qui conduit aux matrices  $\mathbb{X}$  et  $\mathbb{Y}$  de tailles respectives  $n \times p$  et  $n \times q$ ;
2. On fait varier  $(\alpha, \beta)$  en discrétisant les intervalles  $]0, 1/2[ \times ]0, 1[$ ;
3. pour tout  $k = 1, \dots, n$ , on détermine  $\hat{I}_1^{(n)}$  par notre méthode sur la base des matrices  $\mathbb{X}^{-k}$  et  $\mathbb{Y}^{-k}$  obtenues à partir des matrices  $\mathbb{X}$  et  $\mathbb{Y}$  en enlevant les  $k^{\text{ième}}$  lignes ;
4. On calcule la perte de prédiction  $e_k(\alpha, \beta)$  de la  $k^{\text{ième}}$  observation par :

$$e_k(\alpha, \beta) = \|X^{(k)} \left[ (\mathbb{X}^{-k})^T \mathbb{X}^{-k} \right]^{-1} (\mathbb{X}^{-k})^T \mathbb{Y}^{-k} - X_{\hat{I}_1^{(n)}}^{(k)} \left[ (\mathbb{X}_{\hat{I}_1^{(n)}}^{-k})^T \mathbb{X}_{\hat{I}_1^{(n)}}^{-k} \right]^{-1} (\mathbb{X}_{\hat{I}_1^{(n)}}^{-k})^T \mathbb{Y}^{-k} \|_{\mathbb{R}^q}^2$$

où  $\mathbb{X}_{\hat{I}_1^{(n)}}^{-k}$  est la sous-matrice de  $\mathbb{X}^{-k}$  obtenue en utilisant uniquement les variables du sous-ensemble  $\hat{I}_1^{(n)}$ , c'est-à-dire les variables sélectionnées à l'étape précédente,  $X_{\hat{I}_1^{(n)}}^{(k)}$  est la  $k^{\text{ième}}$  ligne de la matrice  $\mathbb{X}_{\hat{I}_1^{(n)}}$ . Les inverses  $[(\mathbb{X}^{-k})^T \mathbb{X}^{-k}]^{-1}$  et  $[(\mathbb{X}_{\hat{I}_1^{(n)}}^{-k})^T \mathbb{X}_{\hat{I}_1^{(n)}}^{-k}]^{-1}$  sont calculés à partir de l'inverse généralisé de Moore-Penrose. Finalement on calcule la perte de prédiction finale par :

$$e(\alpha, \beta) = \frac{1}{n} \sum_{k=1}^n e_k(\alpha, \beta).$$

On obtient ainsi autant de valeurs de  $e(\alpha, \beta)$  que de valeurs dans la discrétisation  $[0, 1/2] \times [0, 1]$ .

5. On choisit  $\hat{\alpha}$  et  $\hat{\beta}$  pour lesquels  $e(\alpha, \beta)$  est minimal

$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{(\alpha, \beta) \in [0, 1/2] \times [0, 1]} e(\alpha, \beta).$$

### 2.3.5 Comparaison avec la méthode An et al. (2013)

La comparaison se fait en deux parties. La première comparaison se fait en considérant le cas de dimensions fixes et la seconde comparaison se fait en considérant le cas de dimensions croissantes. Pour la suite, nous remplaçons la permutation aléatoire  $\sigma$  par le paramètre  $\tau$ , pour ne pas confondre avec la variance du vecteur aléatoire  $\varepsilon$ .

- Dans le cas de dimensions fixes, on prend  $n = 100, 300$ . On applique notre méthode et celle de [An et al. \(2013\)](#) en faisant varier  $\rho$  dans  $[0, 1]$  et la variance  $\tau$  dans  $[0, 1/2]$ .
- Dans le cas de dimensions croissantes, nous reprenons l'exemple 4 de [An et al. \(2013\)](#) où la taille de  $p$  et  $q$  augmentent avec celle de  $n$ , avec  $p = \lceil 10n^{1/3} \rceil$  et  $q = \lfloor p/2 \rfloor$  et on fixe  $\sigma = 1$  et  $\rho = 0.5$ .

Dans les deux cas :

1. On simule l'échantillon d'apprentissage sur lequel on applique notre méthode pour estimer  $\hat{\alpha}, \hat{\beta}$  et déterminer les fonctions  $f_n$  et  $g_n$  qui minimisent la perte de prédiction, en utilisant la méthode de validation croisée (leave-one-cross validation). Puis, on estime le sous-ensemble  $\hat{I}_1^{(n)}$ . On applique également la méthode de [An et al. \(2013\)](#);
2. On simule ensuite l'échantillon test sur lequel on calcule la perte de prédiction sur la base des variables sélectionnées à l'étape précédente.

Les résultats sont reportés dans les Figures 2.2 et 2.3 et les Tableaux 2.2 - 2.4.

### Cas de dimensions fixes

- En faisant varier les paramètres  $\rho$  et  $\tau$  respectivement dans  $[0, 1]$  et  $[0, 1/2]$ , nous voyons à travers les figures Figures 2.2 et 2.3 que la perte de prédiction de notre méthode reste largement inférieure à celle de *An et al.* (2013) quel que soit la méthode de validation croisée (leave-one-cross validation et k-fold cross-validation). Notre approche donne une perte de prédiction quasiment nulle (de l'ordre de  $10^{-10}$ ) avec la méthode du k-fold cross-validation, tandis qu'avec le leave-one-cross validation, elle reste inférieure à celle de *An et al.* (2013) même quand elle paraît élevée. Pour cette raison nous retiendrons la méthode du k-fold cross-validation pour la suite des comparaisons.
- Lorsqu'on s'intéresse à d'autres paramètres de performance tels que 'cptegal' (taux qui représente le nombre de fois où le sous-ensemble  $I_1$  correspond au vrai sous-ensemble), 'cptinclus' (taux qui représente le nombre de fois où les éléments du sous-ensemble  $I_1$  sont ceux du vrai sous-ensemble) et 'NV' (nombre de variables sélectionnées) (voir Tableau 2.2), on remarque que pour  $n = 100(k = 15)$  et pour différentes valeurs de  $\sigma \in [0.1, 0.5]$ , le paramètre 'cptinclus' donne la même valeur 1 pour les deux méthodes, tandis que le paramètre 'cptegal' de notre méthode donne des taux supérieurs à celle de la méthode ASCCA pour  $\tau = 0.1, 0.2, 0.4$ . Le nombre de variables sélectionnées (NV) des deux méthodes est sensiblement le même avec 4 variables sélectionnées en moyenne pour notre méthode et 3 pour celle de la méthode ASCCA. Lorsque la taille de l'échantillon passe à  $n = 300(k = 20)$ , les valeurs des paramètres 'cptinclus' et 'NV' restent quasiment inchangées, alors que les valeurs du paramètre 'cptegal' sont supérieures avec la méthode ASCCA.

Ainsi, lorsque l'on considère le cas de dimensions fixes, la perte de prédiction de notre méthode reste bien inférieure à celle de la méthode de *An et al.* (2013) (Figures 2.2 et 2.3). L'information contenue de le sous-ensemble sélectionnée est quasiment la même que celle de l'ensemble initial.

### Cas de dimensions croissantes

On se met dans le cas où les paramètres  $p$  et  $q$  augmentent en fonction de la taille de l'échantillon  $n$  tels que :

$$p = \mathbb{E}[10 \times (n)^{1/3}] \quad \text{et} \quad q = \mathbb{E}[p/2].$$

La perte de prédiction (PP) et le nombre de variables sélectionnées (NV) de notre méthode reste bien inférieur à celle de la méthode ASCCA, cela quelque soit la taille de l'échantillon  $n = 50, 100, 200$  (Tableau 2.3). Lorsqu'on regarde les paramètres 'cptegal' et 'cptinclus', on fait le même constat. En effet, le paramètre 'cptegal' de notre méthode donne un pourcentage bien supérieur à celui de la méthode ASCCA, par contre, le paramètre 'cptinclus' donne la même valeur pour les deux méthodes (Tableau 2.4).

Ainsi, lorsque les dimensions des paramètres  $p$  et  $q$  augmente avec la taille de l'échantillon  $n$ , notre méthode est bien meilleur à celle de [An et al. \(2013\)](#). Là encore, l'information contenue de le sous-ensemble sélectionnée est quasiment la même que celle de l'ensemble initial. Notons dans les deux cas que la méthode de [An et al. \(2013\)](#) sélection plus de variables que a notre.

### 2.3.6 Application aux données réelles

Nous utilisons des données sur les performances des portefeuilles d'actions à notation pondérée, obtenues sur la base de données historiques des marchés boursiers américains (voir [Liu et Yeh \(2015\)](#)).

La base de données est de taille  $n = 315$ , avec six (6) variables explicatives et six variables réponses. Nous appliquons notre méthode et celle de [An et al. \(2013\)](#). Nous les comparons en calculant la perte de prédiction (PP) et nombre de variables sélectionnées (NV) en utilisant 10-fold cross-validation. Les résultats sont reportés dans le Tableau 2.5.

Tout comme dans la Section 2.3.5, la perte de prédiction (PP) de notre méthode est bien inférieur à celle de [An et al. \(2013\)](#), respectivement de  $1.66055e^{-33}$  contre 0.07640938. Le nombre de variables sélectionnées de notre méthode est en moyenne de 2.61 contre celle de [An et al. \(2013\)](#) qui est de 6.

---

## Informations sur les variables (voir *Liu et Yeh (2015)*) pour plus de détails

---

### Pondération des concepts

---

$X_1$  : poids du concept Grand B/P

$X_2$  : poids du concept Grand RE

$X_3$  : poids du grand concept S/P

$X_4$  : poids du taux de rendement élevé dans le concept du dernier trimestre

$X_5$  : poids du concept de grande valeur marchande

$X_6$  : poids du concept de petit risque symétrique

---

### Les rendements (indicateurs de performance des placements)

---

$Y_1$  : rendement annuel

$Y_2$  : rendement excédentaire

$Y_3$  : risque systématique

$Y_4$  : risque total

$Y_5$  : taux de victoire

$Y_6$  : taux de réussite

---

## 2.4 Conclusion

Dans ce chapitre, nous avons généralisé la méthode de sélection de variables de *Nkiet (2001)* dans le cas d'un modèle linéaire multidimensionnel (Equation (2.1)). Notre approche repose sur un critère (Equation (2.5)) qui est une extension du critère de *Nkiet (2001)* dans le cas multidimensionnel. Nous supposons seulement que les moments d'ordre quatre des vecteurs aléatoires  $X$  et  $Y$  sont finis. Dans notre approche, le critère est utilisé pour ranger les variables dans l'ordre décroissant, de sorte à ramener le problème de sélection à un problème d'estimation d'une permutation et d'un paramètre appelé dimension. A partir des estimateurs empiriques, nous avons établi un estimateur du critère de sélection et les propriétés de convergence, puis la convergence en probabilité de notre méthode

de sélection de variables.

Nous avons illustré à partir d'un exemple de simulations notre méthode de sélection dans le but d'étudier principalement l'influence des paramètres de contrôle  $f_n$ ,  $g_n$ ,  $\alpha$ ,  $\beta$ , puis la comparer à celle de [An et al. \(2013\)](#). Cette étude relève qu'il n'y a pas une influence significative des fonctions de pénalité  $f_n$  et  $g_n$  sur notre méthode de sélection, à l'exception des fonctions  $g_5$ ,  $g_{10}$  et  $g_{10}$  dont les valeurs de la perte de prédiction reste élevés (Tableaux 2.1). Par contre, nous observons une influence des paramètres  $\alpha$  et  $\beta$  sur notre méthode de sélection (Figure 2.1). Ce constat nous a amené à utiliser la méthode de validation croisée dans le but de calculer automatiquement les paramètres optimaux  $\alpha$ ,  $\beta$ ,  $f_n$  et  $g_n$  qui minimisent la perte de prédiction. L'étude comparative avec la méthode de [An et al. \(2013\)](#) montre que notre méthode de sélection à un certain avantage à celle de [An et al. \(2013\)](#), sur la base de la perte de prédiction, cela quelque soit le mode de comparaison considéré (dimensions fixes ou dimensions croissantes) (Figures 2.2 et 2.3, et Tableaux 2.2-2.4).

Une application aux données réelles vient confirmé les résultats obtenus par simulations. En effet, la perte de prédiction (PP) de notre méthode est restée inférieur à celle de [An et al. \(2013\)](#), y compris le nombre de variables sélectionnées (NV) (Tableau 2.5).



## 2.4 Conclusion

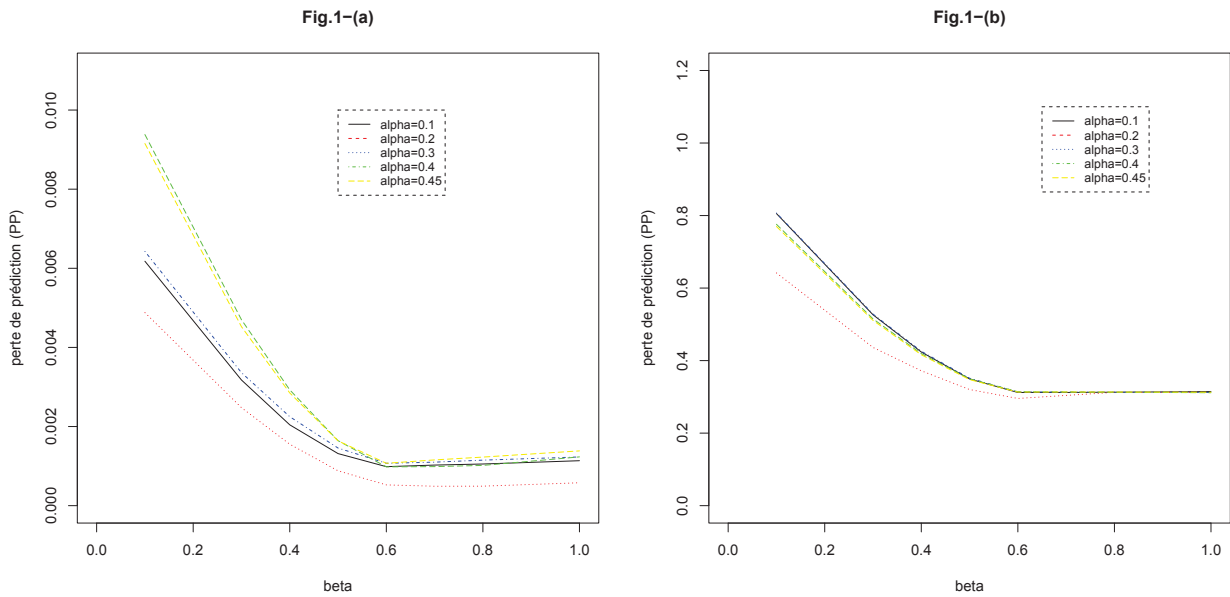


FIGURE 2.1 – *Perte de Prédiction moyenne sur 1000 réplifications des différentes valeurs de  $\alpha$  en fonction  $\beta$ . 1 – (a) $n = 100$ , 1 – (b) $n = 300$*

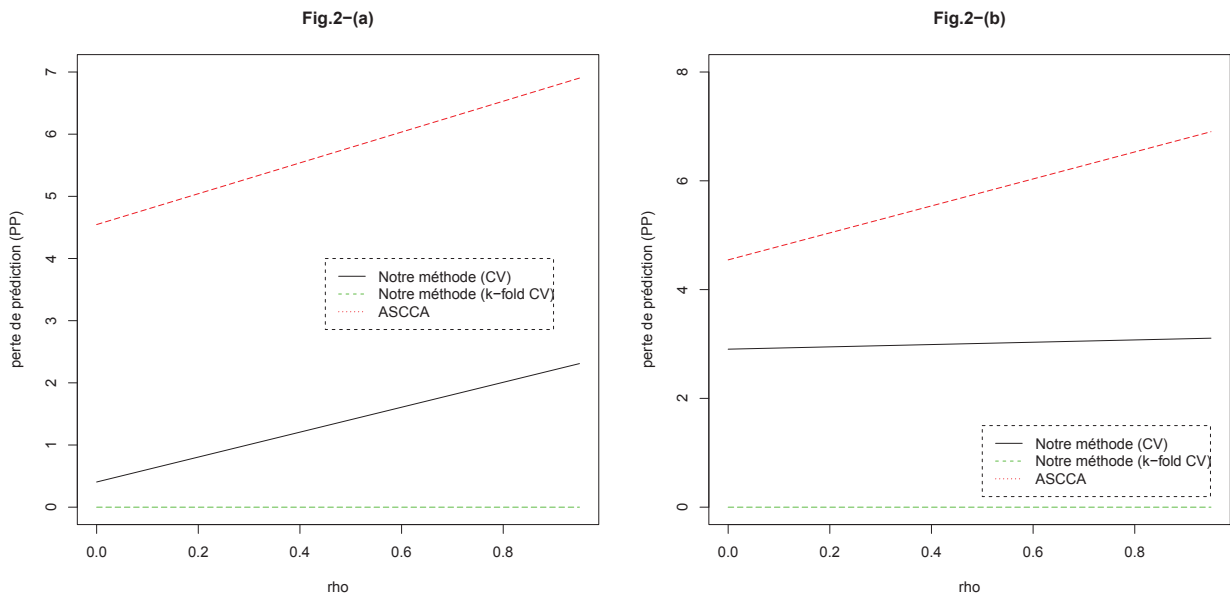


FIGURE 2.2 – *Perte de Prédiction moyenne sur 1000 réplifications de notre méthode contre la méthode ASCCA en fonction du paramètre  $\rho$ . 2 – (a) $n = 100$ , 2 – (b) $n = 300$  : dimensions fixes*

## 2.4 Conclusion

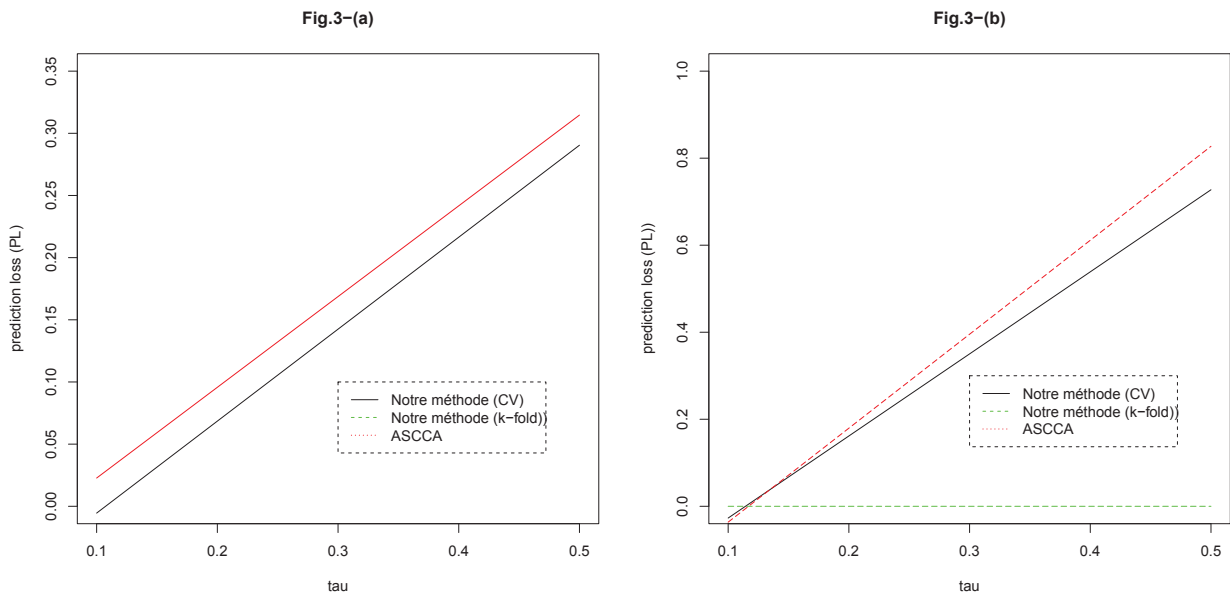


FIGURE 2.3 – Perte de Prédiction moyenne sur 1000 réplifications de notre méthode contre la méthode ASCCA en fonction du paramètre  $\tau$ . 3 – (a)  $n = 100$ , 3 – (b)  $n = 300$  : dimensions fixes

## 2.4 Conclusion

TABLE 2.1 – Perte de prédiction moyenne (PP) sur 1000 répliquions pour différentes fonctions de pénalité, avec  $\alpha = 0.45$  et  $\beta = 0.6$ 

Fonctions	n = 100 (k=15)		
	PP		
	$\rho = 0.0$	$\rho = 0.5$	$\rho = 0.95$
$g_1$	0.02657191	0.1990921	2.685934
$g_2$	0.0006858685	0.0009572785	0.002988558
$g_3$	0.0003579082	0.0007596651	0.009403903
$g_4$	0.005338642	0.01949485	2.653969
$g_5$	114.9085	117.4871	54.88406
$g_6$	0.0009480779	0.0007012231	0.005244042
$g_7$	0.000592644	0.001071362	0.004137584
$g_8$	0.000610327	0.0006652832	0.00396092
$g_9$	0.0005548864	0.001151073	0.004580952
$g_{10}$	2.588115	5.635626	2.719288
$g_{11}$	0.000725634	0.0005990683	0.004515614
$g_{12}$	0.0005860259	0.00205302	1.685641
$g_{13}$	1.954152	2.382472	2.717708
n = 300 (k=40)			
$g_1$	0.3257221	0.3742102	2.972227
$g_2$	0.3119884	0.3119033	0.3110775
$g_3$	0.3151506	0.3097019	0.3239218
$g_4$	0.3142556	0.3206701	2.916341
$g_5$	113.4268	119.7323	56.17824
$g_6$	0.3125098	0.3131748	0.3151784
$g_7$	0.3096639	0.3111037	0.3162395
$g_8$	0.3155095	0.3142341	0.3135263
$g_9$	0.3104943	0.3117994	0.3138814
$g_{10}$	2.73591	2.866421	2.98339
$g_{11}$	0.3144191	0.3103335	0.3101094
$g_{12}$	0.3088191	0.3162726	1.345225
$g_{13}$	1.607759	2.411564	2.985573

TABLE 2.2 – Performance moyenne : *cptegal*, *cptinclus*, NV par validation croisée (*k*-fold cross-validation) pour différentes valeurs de  $\tau$ , sur 500 réplifications : dimensions fixes

		Notre méthode			Méthode de An et al.		
		Directe			ASCCA		
		<i>cptegal</i>	<i>cptinclus</i>	NV	<i>cptegal</i>	<i>cptinclus</i>	NV
$\tau$		$n = 100(k = 15)$					
0.1	1	1		3	0.9066667	1	3.146667
0.2	0.6533333	1		4.066667	0.5466667	1	3.546667
0.3	0.51	1		4.506667	0.6133333	1	3.493333
0.4	0.46666667	1		4.026667	0.2933333	1	3.81333
0.5	0.4533333	1		4.56	0.52	1	3.76
$\tau$		$n = 300(k = 20)$					
0.1	0.88	1		3.23	0.94	1	3.06
0.2	0.51	1		4.03	0.83	1	3.17
0.3	0.48	1		4.39	0.79	1	3.21
0.4	0.47	1		4.82	0.93	1	3.07
0.5	0.50	1		4.51	0.97	1	0.03

TABLE 2.3 – Perte de prédiction moyenne (PP) et nombre de variable sélectionnées (NV) par validation croisée ( $k$ -fold cross-validation) sur 500 réplifications, dimensions croissantes :  $n = 50, p = 36, q = 18, n = 100, p = 46, q = 23$  et  $n = 200, p = 58, q = 29$

	Notre méthode		Méthode de An et al.	
	Directe		ASCCA	
$n(k)$	PP	NV	PP	NV
50(15)	6.881943e-28	4.6	444.5936	36
100(15)	4.319334e-28	7.088889	2.975767	23.55556
200(15)	8.749825e-18	3	4.926787	5.6

TABLE 2.4 – Performance moyenne : *cptegal*, *cptinclus*, NV par validation croisée (*k*-fold cross-validation) sur 500 réplifications, dimensions croissantes :  $n = 50, p = 36, q = 18$ ,  $n = 100, p = 46, q = 23$  et  $n = 200, p = 58, q = 29$

$n(k)$	Notre méthode		Méthode de An et al.	
	Directe		ASCCA	
	<i>cptegal</i>	<i>cptinclus</i>	<i>cptegal</i>	<i>cptinclus</i>
50(15)	0.6	1	0	1
100(15)	0.4	1	0	1
200(15)	1	1	0.06666667	1

TABLE 2.5 – Perte de prédiction moyenne (PP) et nombre de variables sélectionnées par validation croisée (*k*-fold cross-validation) sur 1000 réplifications.  $n = 252$ ,  $p = q = 6$ ,  $k = 15$

Notre méthode		Méthode de An et al.	
Directe		ASCCA	
PP	NV	PP	NV
$1.660551e^{-33}$	2.61	0.07640938	6

## **Deuxième partie**

# **Sélection des variables en analyse discriminante mixte**



---

# ÉTAT DE L'ART DE QUELQUES MÉTHODES DE SÉLECTION DES VARIABLES EN ANALYSE DISCRIMINANTE

---

## Sommaire

---

<b>3.1</b>	<b>Approche de NKIET (2012)</b> . . . . .	<b>67</b>
3.1.1	Notations et caractérisation . . . . .	67
3.1.2	Critère de sélection et estimation . . . . .	68
<b>3.2</b>	<b>Approche de Mahat et al (2007)</b> . . . . .	<b>69</b>
3.2.1	Le modèle LoC . . . . .	70
3.2.2	Critère de sélection . . . . .	71
3.2.3	Procédure de sélection : Méthode de Mahat et al. (2007) . . . . .	71

---

## Introduction

La sélection des variables en analyse discriminante pour des variables aléatoires continues est bien développée dans la littérature. Plusieurs approches et méthodes ont été proposées. Parmi ces méthodes, il y a celle de *Nkiet* (2012). Par contre, quand on s'intéresse à l'analyse discriminante pour des variables aléatoires mixtes, on y trouve peu de travaux. L'une des méthodes les plus répandues dans ce contexte est celle de *Mahat et al.* (2007). Dans le cas de variables aléatoires continues, *Nkiet* (2012) propose une approche dont les variables à sélectionner sont issues non pas de deux groupes, mais plutôt de  $q$  groupes ( $q \geq 2$ ) d'une population donnée. *Mahat et al.* (2007) quant à eux proposent une approche de sélection des variables en analyse discriminante mixte sur deux groupes de population, en utilisant un location model (LoC). En plus de ces caractéristiques, ces approches donnent des résultats bien meilleurs comparés aux méthodes équivalentes dans leur domaine respectif.

Dans ce chapitre, nous rappelons quelques résultats essentiels des méthodes de sélection des variables en analyse discriminante, en mettant un accent sur la méthode de *Nkiet* (2012), qui est basée sur une extension du critère (1.2) (Chapitre 1, Section 1.2.2) dans le cadre de l'analyse discriminante linéaire sur des variables aléatoires continues. La seconde méthode est celle de *Mahat et al.* (2007), qui repose sur la sélection des variables mixtes, dans le cadre de l'analyse discriminante, en utilisant la divergence de Kulback-Leibler entre deux groupes.

### 3.1 Approche de NKIET (2012)

Dans cette section, nous rappelons les outils et les principaux résultats de la méthode de *Nkiet* (2012) qui traite de la sélection des variables aléatoires continues en analyse discriminante dans plusieurs groupes.

#### 3.1.1 Notations et caractérisation

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé. Considérons la variable aléatoire  $Z$  à valeurs dans l'ensemble  $\{1, \dots, q\}$ , qui représente  $q$  groupes d'une population disjoint formant une partition de  $\Omega$ . On suppose que  $\forall \ell \in \{1, 2, \dots, q\}$ , la probabilité d'appartenir au groupe  $\ell$

### 3.1 Approche de NKIET (2012)

est notée  $p_\ell = P(Z = \ell) > 0$ . Soit  $X = (X_1, \dots, X_p)^T$ , un vecteur aléatoire défini sur  $(\Omega, \mathcal{A}, P)$  et à valeurs dans  $\mathbb{R}^p$ . On note  $\mathbb{E}$  l'espérance mathématique telle que  $\mu = \mathbb{E}(X)$ , et  $\mathbb{V}$  l'opérateur de covariance associé à la variable  $X$  tel que  $\mathbb{V} = \mathbb{E}((X - \mu)^{\otimes 2})$ . On suppose que l'opérateur  $\mathbb{V}$  est inversible et que  $\mathbb{E}(\|X\|_{\mathbb{R}^p}^4) < +\infty$ . En outre, on note par  $\mu_\ell := \mathbb{E}(X|Z = \ell)$ ,  $\forall \ell \in \{1, \dots, q\}$  et  $\mathbf{B} := \sum_{\ell=1}^q p_\ell (\mu_\ell - \mu)^{\otimes 2}$ , l'opérateur de covariance entre les groupes.

Notons par  $I_0$ , l'ensemble des indices des variables qui ne contribuent pas à la discrimination entre les groupes. [McKay \(1977\)](#), [Fujikoshi et Yasunori \(1982\)](#) et [Krishnaiah \(1982\)](#) définissent le sous-ensemble  $I_0$  par :

$$I_0 = \{k \in I \mid \forall 1 \leq i \leq r, v_{ik} = 0\},$$

où  $I = \{1, \dots, p\}$ ,  $r$  est le rang de l'opérateur  $T = \mathbb{V}^{-1}\mathbf{B}$  et  $v_i = (v_{i1}, \dots, v_{ip})^T$  est le vecteur propre de  $T$  associé à la valeur propre  $\lambda_i$ . Cette définition a été reprise par [Nkiet \(2012\)](#) afin d'élaborer sa méthode de sélection des variables. Ainsi, tout comme au Chapitre 1, il ramène le problème de sélection de variables à un problème d'estimation du sous-ensemble  $I_1 = I - I_0$ , où  $I_1$  est le sous-ensemble des variables qui contribuent à la discrimination entre les groupes, avec

$$I_1 = \{k \in I \mid \exists 1 \leq i \leq r, v_{ik} \neq 0\}.$$

#### 3.1.2 Critère de sélection et estimation

Le critère de sélection utilisé dans [Nkiet \(2012\)](#) est donné par l'équation suivante :

$$\xi_K = \sum_{\ell=1}^q p_\ell^2 \| (I_{\mathbb{R}^p} - \mathbb{V}Q_K) (\mu_\ell - \mu) \|_{\mathbb{R}^p}^2. \quad (3.1)$$

où  $Q_K := A_K^* (A_K \mathbb{V} A_K^*)^{-1} A_K$  est un opérateur linéaire,  $A_K^*$  désigne l'adjoint de  $A_K$ , et  $I_{\mathbb{R}^p}$  l'opérateur identité de taille  $p$ . Ce critère est considéré comme une mesure de perte d'informations lorsqu'un sous-ensemble de variables est sélectionné (Chapitres 1 et 2). Il est utilisé pour ranger toutes les variables dans l'ordre décroissant du pouvoir discriminant, afin que le problème de sélection des variables se réduise à l'estimation d'une permutation

### 3.2 Approche de Mahat et al (2007)

et d'une dimension, c'est-à-dire qu'il est question d'estimer les paramètres  $\sigma$  et  $s$  tel que  $I_1 = \{\sigma(i); 1 \leq i \leq s\}$  (voir Chapitres 1 et 2).

Soit  $(X_i)_{1 \leq i \leq n}$ , un échantillon i.i.d de  $X$ . On a :

$$\bar{X}^{(n)} := \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad \widehat{V}^{(n)} := \frac{1}{n} \sum_{i=1}^n \left( X_i - \bar{X}^{(n)} \right)^{\otimes 2} = \frac{1}{n} \sum_{i=1}^n X_i^{\otimes 2} - \bar{X}^{(n)\otimes 2},$$

et pour tout  $\ell \in \{1, \dots, q\}$ , on note :

$$\widehat{N}_\ell^{(n)} = \sum_{i=1}^n \mathbb{1}_{i \in \mathcal{Z}_\ell}, \quad \text{et} \quad \bar{X}_\ell^{(n)} = \widehat{N}_\ell^{(n)-1} \sum_{i \in \mathcal{Z}_\ell} X_i,$$

où  $\mathcal{Z}_\ell$  désigne l'ensemble d'éléments tombant dans le groupe  $Z = \{\ell\}$  et  $\mathbb{1}_{\{i \in \mathcal{Z}_\ell\}}$  est la fonction indicatrice sur l'évènement  $\{i \in \mathcal{Z}_\ell\}$ . Alors pour tout sous-ensemble  $K \subset I$ , on a :

$$\widehat{Q}_K^{(n)} = A_K^* \left( A_K \widehat{V}^{(n)} A_K^* \right)^{-1} A_K$$

Ainsi, on obtient un estimateur du critère  $\xi_K$  par :

$$\widehat{\xi}_K^{(n)} = \sum_{\ell=1}^q \left( \frac{\widehat{N}_\ell^{(n)}}{n} \right)^2 \left\| \left( I_{\mathbb{R}^p} - \widehat{V}^{(n)} \widehat{Q}_K^{(n)} \right) \left( \bar{X}_\ell^{(n)} - \bar{X}^{(n)} \right) \right\|_{\mathbb{R}^p}^2. \quad (3.2)$$

De l'équation (3.2), la méthode de sélection des variables reste identique à celle présentée au Chapitre 1, Section 1.2.3. Plus précisément, l'estimation du sous-ensemble  $I_1$  et la procédure de sélection suivent les mêmes étapes que celles du Chapitre 1.

## 3.2 Approche de Mahat et al (2007)

La méthode de sélection proposée par *Mahat et al.* (2007) traite du problème de sélection de variables sur la base de la distance entre deux groupes à partir de la divergence de Kullback-Leibler. Elle résulte du problème de classification en analyse discriminante mixte (mélange de variables continues et discrètes) sur deux groupes de populations. Cette méthode a été proposée pour la première fois par *Olkin et Tate* (1961), puis plus

tard par *Krzanowski* (1983) à travers l'utilisation du modèle LoC (Location Model). Un des grands avantages de cette méthode est qu'elle offre plusieurs applications en pratique *Krzanowski* (1975); *Krishnaiah* (1982); *Krzanowski* (1983). L'inconvénient de cette approche est que le nombre de variables discrètes doit être limité, sinon l'ensemble des paramètres à estimer devient excessif. Pour pallier ce problème, *Mahat et al.* (2007) proposent, dans le cadre mixte, une nouvelle méthode de sélection de variables.

Dans cette section, nous revenons sur les grandes étapes de la méthode de *Mahat et al.* (2007), c'est-à-dire résumer la méthode de discrimination entre les groupes développée à partir du modèle LoC, puis nous présentons quelques procédures de sélection liées à cette méthode.

### 3.2.1 Le modèle LoC

Le modèle LoC est un modèle probabiliste proposé par *Olkin et Tate* (1961) et *Krzanowski* (1983). Il se présente comme suit : Soit  $p$  variables continues  $X = (X_1, X_2, \dots, X_p)$  et  $d$  variables binaires  $Y = (Y_1, Y_2, \dots, Y_d)$  mesurées sur chaque individu d'une population  $\pi_1$  ou d'une population  $\pi_2$ . Les  $d$  variables binaires et les  $p$  variables continues définissent un vecteur multinomial  $W^T = (X^T, Y^T)$  sur chaque observation et contenant  $M$  états (où cellules), avec  $M = 2^d$ . La variable aléatoire  $Y$  génère  $1 + \sum_{j=1}^d Y_j 2^{j-1}$  cellules.

La probabilité d'observer une cellule  $m$  dans la population  $\pi_i$  ( $i = 1, 2$ ) est notée  $p_{im} > 0$  ( $m = 1, \dots, M$ ), avec  $\sum_{m=1}^M p_{im} = 1$ . On suppose que la variable aléatoire  $X$  suit une loi normale d'espérance  $\mu_{im}$  dans la cellule  $m$  et la population  $\pi_i$  ( $i = 1, 2, m = 1, \dots, M$ ), avec une matrice de covariance  $\Sigma$  commune à toutes les cellules des deux populations  $\pi_1$  et  $\pi_2$ . On note par  $G : (\Omega, \mathcal{F}) \rightarrow \{1, 2\}$  la variable aléatoire qui caractérise les groupes, on a :

$$(X|W = m, G = i) \sim \mathcal{N}(\mu_{im}, \Sigma).$$

**Remarque 3.2.1** Pour plus de généralité, les variables catégorielles peuvent être aussi incluses dans cette formulation. En effet, une variable nominale avec  $k$  catégories peut être représentée par  $k - 1$  variables binaires (*Krzanowski* (1980)).

### 3.2.2 Critère de sélection

A partir de la théorie générale de classification, *Krzanowski* (1975) établit qu'une observation  $w^T = (x^T, y^T)$  appartient au groupe  $\pi_1$  si  $y$  tombe dans la cellule  $m$  et que  $x$  satisfait la relation suivante :

$$(\mu_{1m} - \mu_{2m})^T (\Sigma)^{-1} \left\{ x - \frac{1}{2} (\mu_{1m} + \mu_{2m}) \right\} \geq \log \left( \frac{p_{2m}}{p_{1m}} \right) + \log(a), \quad (3.3)$$

sinon, on est dans le groupe  $\pi_2$ ,  $a$  est une constante positive.

Plus tard, *Bar-Hen et Daudin* (1995) généralise la distance de Mahalanobis pour obtenir la divergence de Kullback-Leibler afin de mesurer la distance entre deux groupes quand les variables sont mixtes :

$$\Delta_J = \Delta_{J_1} + \Delta_{J_2}. \quad (3.4)$$

où

$$\Delta_{J_1} = \sum_{m=1}^M (p_{1m} - p_{2m}) \log \left( \frac{p_{1m}}{p_{2m}} \right)$$

et

$$\Delta_{J_2} = \frac{1}{2} \sum_{m=1}^M (p_{1m} + p_{2m}) (\mu_{1m} - \mu_{2m})^T (\Sigma)^{-1} (\mu_{1m} - \mu_{2m}).$$

C'est ce critère (Equation (3.4)) qu'a repris *Mahat et al.* (2007) pour effectuer la sélection des variables dans le modèle LoC.

### 3.2.3 Procédure de sélection : Méthode de Mahat et al. (2007)

Comme mentionné à la Section 3.2.2, le critère utilisé pour identifier les variables utiles est la divergence de Kulback-Leibler (Equation (3.4)). Notons par  $\hat{\Delta}_J$ , l'estimateur de  $\Delta_J$  en remplaçant les paramètres  $\mu_{im}$ ,  $p_{im}$  et  $\Sigma$  par leur estimateur  $\hat{\mu}_{im}$ ,  $\hat{p}_{im}$  et  $\hat{\Sigma}$ , en utilisant des estimateurs non paramétriques (*Mahat et al.* (2007)). Les variables qui maximisent

### 3.2 Approche de Mahat et al (2007)

$\widehat{\Delta}_J$  sont identifiées en utilisant la sélection ascendante, descendante et et la sélection pas à pas. Ces méthodes sont basées sur le test d'information non additionnelle ([Rao \(1973\)](#)). L'inclusion ou l'élimination des variables est établie lors du changement de la valeur de  $\widehat{\Delta}_J$ . Dans la suite, notons par  $\mathcal{C}$  l'ensemble des variables sélectionnées.

#### Sélection ascendante

La sélection ascendante consiste à trouver la meilleure des variables dans  $\mathcal{C}$  qui donne la plus grande valeur de  $\widehat{\Delta}_J$ . Si par exemple la variable testée n'accroît pas la séparation entre les groupes, alors toutes les valeurs  $\widehat{\Delta}_J$  sont des estimations de la même quantité  $\Delta$ . Dans ce cas, la différence entre  $\widehat{\Delta}_{J_j}$  et  $\widehat{\Delta}_{J_{j-1}}$  pour  $j = 2, \dots, p + d$  est nulle et suit approximativement la loi de  $\chi^2$  avec les degrés de liberté  $\nu = \nu_j - \nu_{j-1}$  :

$$X_0^f = \frac{n_1 n_2}{n_1 + n_2} \left( \widehat{\Delta}_{J_j} - \widehat{\Delta}_{J_{j-1}} \right) \sim \chi^2(\nu = \nu_j - \nu_{j-1}),$$

où  $n_i$  est la taille de l'échantillon du groupe  $\pi_i$ ,  $\widehat{\Delta}_{j_i}$  et  $\widehat{\Delta}_{J_{j-1}}$  sont respectivement les distances de Kullback-Leibler estimées aux états  $j$  et  $j - 1$ ;  $\nu_j$  et  $\nu_{j-1}$  sont les valeurs de degrés de liberté pour la loi  $\chi^2$  de  $\widehat{\Delta}_{j_i}$  et  $\widehat{\Delta}_{J_{j-1}}$ , respectivement. Cette statistique de test est comparée à  $\chi^2(\nu = 2^d - 1, 1 - \alpha)$  si une variable continue est sélectionnée à l'étape  $j$  ou à  $\chi^2(\nu = 2^d(2 + p), 1 - \alpha)$  si une variable binaire est sélectionnée à l'étape  $j$ ;  $p$  et  $d$  sont respectivement le nombre de variables continues et binaires à l'étape  $1 - j$  et  $\alpha$  est l'erreur de première espèce. La variable testée est sélectionnée si  $X_0^f > \chi^2(\nu, 1 - \alpha)$ , autrement le processus s'arrête.

#### Sélection descendante

La sélection descendante consiste à trouver la plus mauvaise variable dans  $\mathcal{C}$  qui donne la plus grande valeur de  $\widehat{\Delta}_J$ . Pour cela, on compare la statistique de test

$$\chi_0^b = \frac{n_1 n_2}{n_1 + n_2} \left( \widehat{\Delta}_{J_{j-1}} - \widehat{\Delta}_{J_j} \right)$$

### 3.2 Approche de Mahat et al (2007)

---

avec  $\chi^2(\nu = 2^d - 1, 1 - \alpha)$  si une variable continue est sélectionnée à l'étape  $j$  ou à  $\chi^2(\nu = 2^d + p2^{d-1}, 1 - \alpha)$  si une variable binaire est sélectionnée à l'étape  $j$ . Les entiers  $p$  et  $d$  sont respectivement le nombre de variables continues et binaires à l'étape  $j - 1$ . La variable sélectionnée est supprimée si  $\chi_0^b > \chi^2(\nu, 1 - \alpha)$ , autrement le processus s'arrête. Cette stratégie a une possibilité de supprimer toutes les variables et peut aller à l'encontre de la construction de la règle de classification. Alors, la sélection descendante est aussi arrêtée s'il y a seulement une variable continue et une variable binaire.

#### Sélection pas à pas

La sélection pas à pas améliore la sélection ascendante, ainsi la règle d'arrêt décrite dans la sélection ascendante est utilisée. L'ensemble des variables choisies est examiné en utilisant la sélection descendante. La statistique de test de la sélection pas à pas est donnée par :

$$X_0^s = \frac{n_1 n_2}{n_1 + n_2} (\hat{\Delta}_{J_j}^f - \hat{\Delta}_{J_j}^b),$$

où  $\hat{\Delta}_{J_j}^f$  est la distance estimée à l'étape  $j$  obtenue dans la sélection ascendante et  $\hat{\Delta}_{J_j}^b$  est la distance à l'étape  $j$  obtenue dans la sélection descendante.  $X_0^s$  est comparée à  $\chi^2(\nu = 2^d - 1, 1 - \alpha)$  ou à  $\chi^2(\nu = 2^d + p2^{d-1}, 1 - \alpha)$  selon le type de variable qui a été sélectionnée. La recherche s'arrête lorsque les deux tests ascendant et descendant donnent l'instruction arrêt.



# SÉLECTION DES VARIABLES EN ANALYSE DISCRIMINANTE MIXTE SUR PLUSIEURS GROUPES

## Sommaire

<b>4.1</b>	<b>Critère de sélection pour les variables mixtes dans plusieurs groupes . . .</b>	<b>76</b>
4.1.1	Notations et énoncé du problème . . . . .	76
4.1.2	Critère de sélection pour les variables mixtes . . . . .	77
<b>4.2</b>	<b>Estimation et convergence du critère . . . . .</b>	<b>80</b>
4.2.1	Estimation empirique du critère . . . . .	81
4.2.2	Estimation du critère par une approche non paramétrique . . . . .	82
4.2.3	Convergence du critère . . . . .	83
4.2.4	Principaux résultats . . . . .	87
<b>4.3</b>	<b>Sélection des variables . . . . .</b>	<b>100</b>
4.3.1	Estimation et convergence des paramètres $\sigma$ et $s$ . . . . .	101
<b>4.4</b>	<b>Etude numérique . . . . .</b>	<b>107</b>
4.4.1	Simulations des données . . . . .	107
4.4.2	Influence des fonctions de pénalité . . . . .	109
4.4.3	Influence des paramètres $\alpha$ et $\beta$ . . . . .	110
4.4.4	Choix optimal des paramètres $\alpha$ et $\beta$ : Algorithme . . . . .	110
4.4.5	Comparaison avec la méthode Mahat et al. (2007) . . . . .	112
4.4.6	Exemple sur des données réelles . . . . .	112
<b>4.5</b>	<b>Conclusion . . . . .</b>	<b>113</b>

---

## Introduction

---

L'analyse des données mixtes, c'est-à-dire le mélange de variables aléatoires continues et de variables aléatoires discrètes, est un problème largement abordé dans les études statistiques théoriques et pratiques. Dans le domaine de la sélection des variables, contrairement aux variables continues, la sélection des variables mixtes reste moins développée dans la littérature. On y trouve quelques travaux tels que ceux de *Mahat et al.* (2007) qui traitent de cette problématique dans le cas de l'analyse discriminante en considérant deux groupes. Tout comme la plupart des méthodes de sélection des variables continues, celle de *Mahat et al.* (2007) suppose une hypothèse de loi (loi gaussienne) aux variables candidates à la sélection. Au Chapitre 3, nous avons vu que *Nkiet* (2012) propose une approche de sélection où les variables ne sont soumises à aucune hypothèse de loi. En plus elle est valable pour  $q$  groupes ( $q \geq 2$ ) contrairement à celle de *Mahat et al.* (2007) qui est valable sur deux groupes. Ainsi, il serait intéressant de proposer une méthode de sélection de variables dans le cadre de l'analyse discriminante mixte, pour plus de deux groupes, dans le but d'obtenir des résultats analogues à ceux de *Nkiet* (2012) dans le cadre mixte.

Nous proposons ainsi dans ce chapitre, une méthode de sélection des variables pour des variables aléatoires mixtes. C'est la deuxième contribution majeure de cette thèse. Elle est plus théorique que pratique. Cette généralisation repose sur un critère de sélection qui permet de réduire le problème de sélection des variables à un problème d'estimation d'une permutation et d'un paramètre appelé dimension. Nous donnons dans la suite un estimateur du critère de sélection en utilisant deux types d'estimateurs : les estimateurs empiriques et les estimateurs non paramétriques. Nous établissons ses propriétés de convergence et celles de la méthode. Une étude de simulations permettra d'étudier plusieurs propriétés de cette approche. Nous évaluons l'influence des fonctions de pénalité  $f_n, g_n$ , et des paramètres  $\alpha, \beta$  sur la méthode de sélection à partir du taux de bon classement et proposons un algorithme pour le choix optimal de ces paramètres. Nous comparons notre méthode de sélection à celle de *Mahat et al.* (2007) dans le cas de deux groupes, et proposons comme au Chapitre 2, un exemple avec des données réelles.

## 4.1 Critère de sélection pour les variables mixtes dans plusieurs groupes

Dans cette section, nous donnons les principales notations et motivation de ce chapitre, puis nous présentons le critère de sélection pour les variables aléatoires mixtes.

### 4.1.1 Notations et énoncé du problème

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé. Considérons les vecteurs aléatoires  $X = (X_1, \dots, X_p)^T$ , et  $Y = (Y^{(1)}, \dots, Y^{(d)})$  définis dans le même espace de probabilité et à valeurs respectivement dans  $\mathbb{R}^p$  et  $\{0, 1\}^d$ .  $X$  est un vecteur aléatoire continu et  $Y$  un vecteur aléatoire binaire.  $Y$  peut être associé à une variable aléatoire multinomiale  $U = 1 + \sum_{j=1}^d Y^{(j)} 2^{j-1}$ , à valeurs dans  $\{1, \dots, M\}$ , où  $M = 2^d$  ( $d$  est le nombre de variables binaires). On suppose que les observations  $(X, Y)$  sont réparties dans  $q$  groupes  $\mathcal{G}_1, \dots, \mathcal{G}_q$ , caractérisées par la variable aléatoire  $Z$ , à valeurs dans  $\{1, \dots, q\}$ . Cela implique que le couple  $(X, Y)$  est dans la groupe  $\mathcal{G}_\ell$ , de taille  $n_\ell$  si, et seulement si,  $Z = \ell$ .

On se propose d'étendre dans la cas mixte, les résultats obtenus dans [Nkiet \(2012\)](#). C'est-à-dire, proposer une méthode de sélection de variables dans le cadre d'un mélange de variables aléatoires continues et discrètes, pour plus de deux groupes. Plus précisément, on s'intéresse au problème de la discrimination et de la sélection des variables  $X_j$  ( $j = 1, \dots, p$ ) à partir d'un échantillon de  $(X^{(i)}, Y_i, Z_i)_{1 \leq i \leq n}$ . Nous supposons comme au Chapitre 2 que  $\mathbb{E}(\|X\|_{\mathbb{R}^p}^4) < +\infty$ .

On définit les paramètres suivants :

$$\mu = \mathbb{E}(X), \quad \mu_\ell = \mathbb{E}(X|Z = \ell), \quad \text{et} \quad \mu_{\ell,m} = \mathbb{E}(X|Z = \ell, U = m).$$

$$p_m = P(U = m), \quad p_{\ell,m} = P(Z = \ell, U = m), \quad p_{\ell|m} = P(Z = \ell|U = m),$$

$$\mu_m = \mathbb{E}(X|U = m), \quad \mathbb{V}_m = \mathbb{E}[(X - \mu_m) \otimes (X - \mu_m) | U = m],$$

et

$$\mathbf{B}_m = \sum_{\ell=1}^q p_{\ell|m} (\mu_{\ell,m} - \mu_m) \otimes (\mu_{\ell,m} - \mu_m),$$

### 4.1 Critère de sélection pour les variables mixtes dans plusieurs groupes

où  $\mathbf{B}_m$  représente la matrice de covariance entre les groupes dans la cellule  $m$ . Soient  $\lambda_{1,m} \geq \lambda_{2,m} \geq \dots \geq \lambda_{p,m}$  les valeurs propres de l'opérateur  $T_m = \mathbb{V}_m^{-1} \mathbf{B}_m$  associées au vecteur propre  $v_i^m = (v_{i1}^m, \dots, v_{ip}^m)^T$  ( $i = 1, \dots, p$ ). On note  $r_m$  le rang de l'opérateur  $T_m$  dans la cellule  $m$ . Nous choisissons comme fonctions discriminantes, les vecteurs propres de l'opérateurs  $T_m$ , pour différencier le sous-ensemble des variables pertinentes et le sous-ensemble des variables non pertinentes (*Fujikoshi et Yasunori* (1982), *Nkiet* (2012)). Ainsi, Posons :

$$I = \{1, \dots, p\},$$

l'ensemble des indices des variables à sélectionner. Nous définissons l'ensemble des variables qui ne contribuent pas à la discrimination entre les groupes, pour une cellule  $m$  fixée de la variable aléatoire  $U$ , par le sous-ensemble  $I_{0,m}$  tel que :

$$I_{0,m} = \{k \in I \mid \forall 1 \leq i \leq r_m, v_{ik}^m = 0\}, \quad (4.1)$$

d'où

$$I_0 = \bigcap_{m=1}^M I_{0,m}. \quad (4.2)$$

On définit alors par  $I_1$ , le sous-ensemble des indices des variables qui contribuent à la discrimination entre les groupes pour une cellule  $m$  de la variable aléatoire  $U$ . C'est le sous-ensemble des variables dont les coefficients des fonctions discriminantes sont non nulles :

$$I_{1,m} = I_m - I_{0,m},$$

c'est-à-dire :

$$I_{1,m} = \{k \in I \mid \exists 1 \leq i \leq r_m, v_{ik}^m \neq 0\}, \quad (4.3)$$

d'où

$$I_1 = \bigcup_{m=1}^M I_{1,m}. \quad (4.4)$$

#### 4.1.2 Critère de sélection pour les variables mixtes

L'objectif dans cette section est de définir un critère de sélection, dans le but de sélectionner les variables aléatoires  $X_i$  ( $i = 1, \dots, p$ ) sachant son groupe  $\ell$  et sa cellule  $m$ . Pour

### 4.1 Critère de sélection pour les variables mixtes dans plusieurs groupes

---

toute cellule  $m$  fixée de la variable aléatoire  $U$ , la probabilité d'être dans le groupe  $\ell \in \{1, \dots, q\}$  est notée  $p_{\ell|m}$ . Alors le critère  $\xi_K$  (Equation (3.1), Section 3.1.2, Chapitre 3) pour une cellule  $m$  fixée de la variable aléatoire  $U$  devient :

$$\xi_{K|m} = \sum_{\ell=1}^q p_{\ell|m}^2 \| (I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m}) (\mu_{\ell,m} - \mu_m) \|_{\mathbb{R}^p}^2, \quad (4.5)$$

où  $Q_{K|m} = A_K^* (A_K \mathbb{V}_m A_K^*)^{-1} A_K$ .

**Remarques 4.1.1** On a :

1. Dans une cellule fixée  $m \in \{1, \dots, M\}$ , nous avons uniquement des variables aléatoires continues. Alors le critère  $\xi_{K|m}$  (Equation (4.5)) est exactement celui défini dans **Nkiet (2012)** p. 152.
2. En considérant un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , **Nkiet (2012)** montre dans le Théorème 2.1 que le critère  $\xi_K$  (Equation (3.1)) vérifie la relation suivante :

$$\xi_K = 0 \Leftrightarrow I_1 \subset K.$$

Pour tout  $m \in \{1, \dots, M\}$ , notons  $P^{\{U=m\}}$  la probabilité conditionnelle à l'évènement  $\{U = m\}$ , c'est-à-dire  $P^{\{U=m\}}(\cdot) = P(\cdot | \{U = m\})$ . Si on considère l'espace probabilisé  $(\Omega, \mathcal{A}, P^{\{U=m\}})$ , les résultats du Théorème 2.1 de **Nkiet (2012)** restent valables. Nous pouvons ainsi dire que

$$\xi_{K|m} = 0 \Leftrightarrow I_{1,m} \subset K.$$

Comme pour tout  $m \in \{1, \dots, M\}$ , les évènements  $\{U = m\}$  forment un système complet, c'est-à-dire que les évènements  $\{U = m\}$  forment une partition de  $\Omega$  et sont deux à deux disjoints, il vient que :

$$\begin{aligned} P(Z = \ell) &= \sum_{m=1}^M P(U = m) P(Z = \ell | U = m) \\ p_\ell &= \sum_{m=1}^M p_m p_{\ell|m}, \end{aligned} \quad (4.6)$$

où  $p_\ell = P(Z = \ell)$ . En considérant l'équation (4.5), puis en remplaçant  $p_\ell$  (Equation (4.6)) dans l'équation (3.1) (Chapitre 3 Section 3.1.2), on définit le critère de sélection pour l'ana-

### 4.1 Critère de sélection pour les variables mixtes dans plusieurs groupes

---

lyse discriminante mixte entre plusieurs groupes pour l'ensemble des cellules  $m$  du vecteur aléatoire  $U$  par :

$$\xi'_K = \sum_{m=1}^M \sum_{\ell=1}^q p_m^2 p_{\ell|m}^2 \| (I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m}) (\mu_{\ell,m} - \mu_m) \|_{\mathbb{R}^p}^2 = \sum_{m=1}^M p_m^2 \xi_{K|m}. \quad (4.7)$$

Ce critère est défini comme la somme sur toutes les cellules  $m$  du produit des probabilités de la cellule  $m$ , au carré du critère (4.5) (critère de [Nkiet \(2012\)](#) dans la cellule  $m$ ) sachant que l'on est dans la cellule  $m$ . Il est vu comme une mesure de perte d'informations lorsqu'un sous-ensemble de variables  $K$  est sélectionné. Il est aussi utilisé pour ranger toutes les variables dans l'ordre décroissant du pouvoir discriminant (voir Chapitre 3). Dans le Théorème 8 qui suit, nous donnons une caractérisation du sous-ensemble  $I_1$  (Equation (4.4)).

**Théorème 8**  $\forall K \subset I$ , on a  $\xi'_K = 0 \Leftrightarrow I_1 \subset K$ .

**Preuve:**

$$\begin{aligned} \xi'_K = 0 &\Leftrightarrow \sum_{m=1}^M p_m^2 \xi_{K|m} = 0 \\ &\Leftrightarrow \forall m \in \{1, \dots, M\}, \xi_{K|m} = 0. \end{aligned}$$

En appliquant le Théorème 2.1 de [Nkiet \(2012\)](#) à l'espace probabilisé  $(\Omega, \mathcal{A}, P^{\{U=m\}})$  on a  $I_{1,m} \subset K \Leftrightarrow \xi_{K|m} = 0$ . Ainsi

$$\begin{aligned} \xi'_K = 0 &\Leftrightarrow \forall m \in \{1, \dots, M\}, I_{1,m} \subset K \\ &\Leftrightarrow \bigcup_{m=1}^M I_{1,m} \subset K \\ &\Leftrightarrow I_1 \subset K \end{aligned}$$

□

Le Théorème 8, montre que  $\xi'_{I-\{i\}} = 0$  est équivalent à  $I_1 \subset I - \{i\}$ , c'est-à-dire  $i \in I_0$ . Ainsi, nous utilisons les nombres réels de la forme  $\xi'_{K_i}$  avec  $K_i = I - \{i\}$  comme mesure de perte d'information pour la discrimination des variables mixtes. C'est-à-dire qu'une variable  $X_i$  est adéquate pour la discrimination si  $\xi'_{K_i} > 0$ , et une autre variable  $X_j$  est plus adéquate pour la discrimination par rapport à la variable  $X_i$  (avec  $j \neq i$ ) si  $\xi'_{K_j} > \xi'_{K_i}$ .

---

## 4.2 Estimation et convergence du critère

---

En rangeant les  $X_i$  à partir du critère  $\xi'_{K_i}$  dans l'ordre décroissant de pertinence, nous considérons une permutation aléatoire  $\sigma$  de  $I$  qui satisfait les deux propriétés suivantes :

$$\begin{cases} (P_1) : \xi'_{K_{\sigma(1)}} \geq \xi'_{K_{\sigma(2)}} \geq \cdots \geq \xi'_{K_{\sigma(p)}}; \\ (P_2) : \text{si } \xi'_{K_{\sigma(i)}} = \xi'_{K_{\sigma(j)}} \text{ et } i < j \text{ alors } \sigma(i) < \sigma(j). \end{cases}$$

Les variables ayant la même valeur discriminante sont disposées dans l'ordre croissant des indices correspondants. Comme  $I_1$  est un ensemble non vide, il va exister un entier  $s \in I$  tels que :

$$\begin{cases} (i) & s = p, \text{ si } I_1 = I; \\ (ii) & \xi_{K'_{\sigma(1)}} \geq \xi_{K'_{\sigma(2)}} \geq \cdots \geq \xi_{K'_{\sigma(s)}} \geq \xi_{K'_{\sigma(s+1)}} = \cdots = \xi_{K'_{\sigma(p)}} = 0, \text{ si } I_1 \neq I. \end{cases}$$

Alors, on déduit que le sous-ensemble  $I_1$  est entièrement caractérisé par la permutation aléatoire  $\sigma$  et l'entier  $s$  :

$$I_1 = \{\sigma(i) | 1 \leq i \leq s\}. \quad (4.8)$$

D'où l'estimation du sous-ensemble  $I_1$  se déduit directement de l'estimation de la permutation aléatoire  $\sigma$  et de l'entier  $s$ .

## 4.2 Estimation et convergence du critère

Dans cette partie, nous proposons un estimateur du critère de sélection  $\xi'_K$ . Pour cela, nous utilisons principalement deux types d'estimateurs à savoir les estimateurs empiriques et les estimateurs non paramétriques proposés dans [Mahat et al. \(2007\)](#). Avec les estimateurs empiriques, nous donnons un estimateur du critère  $\xi'_K$  et ses propriétés de convergence. Nous estimons également le critère en utilisant les estimateurs non paramétriques obtenus par une procédure de lissage.

### 4.2.1 Estimation empirique du critère

Soit  $(X^{(i)}, Y_i, Z_i)_{1 \leq i \leq n}$  un échantillon i.i.d de  $(X, Y, Z)$ . On a  $X^{(i)} = (X_1^{(i)}, \dots, X_p^{(i)})^T$ ,  $Y_i = (Y_i^{(1)}, \dots, Y_i^{(d)})^T$  et  $U_i = 1 + \sum_{j=1}^d Y_i^{(j)} 2^{j-1}$ , avec  $p < n$ ,  $d < n$  et  $M < n$ . Notons par

$$\widehat{N}_m^{(n)} = \sum_{i=1}^n \mathbb{1}_{\{U_i=m\}}, \quad \text{et} \quad \widehat{N}_{\ell,m}^{(n)} = \sum_{i=1}^n \mathbb{1}_{\{Z_i=\ell, U_i=m\}},$$

respectivement le nombre d'éléments dans la cellule  $m$  et le nombre d'éléments dans le groupe  $\ell$  et la cellule  $m$ . On estime les paramètres  $p_m, p_{\ell|m}, V_m, \mu_m$  et  $\mu_{\ell,m}$  du critère  $\xi'_K$  par leur estimateur empirique de la façon suivante :

$$\widehat{p}_m^{(n)} = \frac{\widehat{N}_m^{(n)}}{n}, \quad \widehat{p}_{\ell,m}^{(n)} = \frac{\widehat{N}_{\ell,m}^{(n)}}{n},$$

$$\widehat{\mu}_{\ell|m}^{(n)} = \frac{\widehat{N}_{\ell,m}^{(n)}}{\widehat{N}_m^{(n)}}, \quad \widehat{\mu}_m^{(n)} = \frac{1}{\widehat{N}_m^{(n)}} \sum_{i=1}^n \mathbb{1}_{\{U_i=m\}} X_i,$$

$$\widehat{\mu}_{\ell,m}^{(n)} = \frac{1}{\widehat{N}_{\ell,m}^{(n)}} \sum_{i=1}^n \mathbb{1}_{\{Z_i=\ell, U_i=m\}} X_i, \quad \text{et} \quad \widehat{V}_m^{(n)} = \frac{1}{\widehat{N}_m^{(n)}} \sum_{i=1}^n \mathbb{1}_{\{U_i=m\}} (X_i - \widehat{\mu}_m^{(n)}) \otimes (X_i - \widehat{\mu}_m^{(n)}),$$

$$\widehat{Q}_{K|m}^{(n)} = A_K^* \left( A_K \widehat{V}_m^{(n)} A_K^* \right)^{-1} A_K$$

et

$$\widehat{\xi}_{K|m}^{(n)} = \sum_{\ell}^q \left( \widehat{p}_{\ell|m}^{(n)} \right)^2 \left\| \left( I_{\mathbb{R}^p} - \widehat{V}_m^{(n)} \widehat{Q}_{K|m}^{(n)} \right) \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \right\|^2.$$

Un estimateur du critère  $\xi'_K$  est donné par :

$$\widehat{\xi}_K^{(n)'} = \sum_{m=1}^M \sum_{\ell=1}^q \left( \widehat{p}_m^{(n)} \right)^2 \left( \widehat{p}_{\ell|m}^{(n)} \right)^2 \left\| \left( I_{\mathbb{R}^p} - \widehat{V}_m^{(n)} \widehat{Q}_{K|m}^{(n)} \right) \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \right\|_{\mathbb{R}^p}^2. \quad (4.9)$$



### 4.2.2 Estimation du critère par une approche non paramétrique

Les estimateurs empiriques utilisés précédemment ont de bonnes propriétés de convergence, mais peuvent rencontrer des problèmes dans certains cas. En effet, il est possible que  $\widehat{N}_m^{(n)}$  (resp.  $\widehat{N}_{\ell,m}^{(n)}$ ), le nombre d'objet dans la cellule  $m$  (resp. le nombre d'objet dans le groupe  $\ell$  et la cellule  $m$ ) soit négligeable ou nul. Dans ce cas, les estimateurs empiriques utilisés dans la Section 4.2.1 ne sont plus valables (*Asparoukhov et Krzanowski (2000)*). Il est donc nécessaire d'utiliser un autre type d'estimateur afin de remédier au problème de cellules vides. Nous proposons dans cette section d'utiliser une méthode non paramétrique pour estimer le critère  $\xi'_K$  (Equation (4.7)). Cette approche est basée sur une procédure de lissage (voir *Mahat et al. (2007)*).

Notons par  $h$  le coefficient de différentiation entre les cellules  $m$  et  $k$  tel que :

$$h(m, k) = \|\mathbf{x}_m - \mathbf{x}_k\|^2,$$

où pour tout  $k \in \{1, \dots, M\}$ ,  $\mathbf{x}_k = (\mathbf{x}_k^{(1)}, \dots, \mathbf{x}_k^{(d)})^T \in \{0, 1\}^d$ , le vecteur de variables binaires tel que  $1 + \sum_{j=1}^d \mathbf{x}_k^{(j)} 2^{j-1} = k$ . Soit  $\lambda \in ]0, 1[$  un paramètre de lissage. On considère la pondération (ou le poids)  $\omega(m, k) = \lambda^{h(m,k)}$ . Plusieurs fonctions de poids  $\omega(m, k)$  peuvent être utilisées, on a par exemple (*Mahat et al. (2007)*) :

$$\begin{cases} (iii) & \omega(m, k) = \lambda^{h(m,k)}, \quad 0 \leq \lambda \leq 1, \\ (iv) & \omega(m, k) = \lambda_{h(m,k)} \quad 0 \leq \lambda_{h(m,k)} \leq 1. \end{cases}$$

La première fonction (iii) est une fonction exponentielle, tandis que la seconde (iv) est une fonction linéaire par morceaux. Dans les deux cas,  $\lambda$  représente le paramètre de lissage identique à toutes les variables. Pour la suite, nous travaillons avec la fonction exponentielle  $\omega(m, k) = \lambda^{h(m,k)}$ . Le choix optimal du paramètre de lissage  $\lambda$  est donné dans *Asparoukhov et Krzanowski (2000)* et *Mahat et al. (2007)*. Ainsi, on estime les paramètres  $p_m$ ,  $p_{\ell|m}$ ,  $\mu_m$ ,  $\mu_{\ell,m}$  et  $V_m$  respectivement par :

$$\widehat{p}_m^{(n)} = \frac{\sum_{j=1}^M w(m, j) \widehat{N}_j^{(n)}}{\sum_{m=1}^M \sum_{j=1}^M w(m, j) \widehat{N}_j^{(n)}}, \quad \widehat{p}_{\ell|m}^{(n)} = \frac{\sum_{j=1}^M w(m, j) \widehat{N}_{\ell,j}^{(n)}}{\widehat{p}_m^{(n)} \sum_{m=1}^M \sum_{j=1}^M w(m, j) \widehat{N}_{\ell,j}^{(n)}}$$

$$\begin{aligned}\tilde{\mu}_m^{(n)} &= \left\{ \sum_{j=1}^M w(m, j) \widehat{N}_j^{(n)} \right\}^{-1} \sum_{j=1}^M \left\{ w(m, j) \sum_{n=1}^n X_i \mathbb{1}_{\{U_i=j\}} \right\}, \\ \tilde{\mu}_{\ell, m}^{(n)} &= \left\{ \sum_{j=1}^M w(m, j) \widehat{N}_{\ell, j}^{(n)} \right\}^{-1} \sum_{j=1}^M \left\{ w(m, j) \sum_{n=1}^n X_i \mathbb{1}_{\{Z_i=\ell, U_i=j\}} \right\}\end{aligned}$$

et

$$\tilde{V}_m^{(n)} = \left\{ \sum_{j=1}^M w(m, j) \widehat{N}_j^{(n)} \right\}^{-1} \sum_{j=1}^M \left\{ w(m, j) \sum_{n=1}^n \mathbb{1}_{\{U_i=j\}} (X_i - \tilde{\mu}_m^{(n)}) \otimes (X_i - \tilde{\mu}_m^{(n)}) \right\}.$$

On obtient ainsi un estimateur du critère  $\xi'_K$  par la méthode de lissage non paramétrique, en remplaçant dans l'équation (4.7) les paramètres  $p_m, p_{\ell|m}, \mu_m, \mu_{\ell, m}$  et  $V_m$  par leur estimateur :

$$\tilde{\xi}_K^{(n)'} = \sum_{m=1}^M \sum_{\ell=1}^q \left( \tilde{p}_m^{(n)} \right)^2 \left( \tilde{p}_{\ell|m}^{(n)} \right)^2 \left\| \left( I_{\mathbb{R}^p} - \tilde{V}_m^{(n)} \tilde{Q}_{K|m}^{(n)} \right) \left( \tilde{\mu}_{\ell, m}^{(n)} - \tilde{\mu}_m^{(n)} \right) \right\|_{\mathbb{R}^p}^2.$$

$$\text{où } \tilde{Q}_{K|m}^{(n)} = A_K^* \left( A_K \tilde{V}_m^{(n)} A_K^* \right)^{-1} A_K.$$

### 4.2.3 Convergence du critère

Dans cette section, nous énonçons dans les Théorèmes 9 - 10 les principaux résultats de convergence du critère  $\tilde{\xi}_K^{(n)'}$  obtenus en utilisant les estimateurs empiriques. Mais auparavant, nous énonçons trois lemmes essentiels à leur démonstration. Considérons les variables aléatoires suivantes :

$$A = \begin{pmatrix} \mathbb{1}_{\{Z=1, U=1\}} X & \dots & \mathbb{1}_{\{Z=1, U=M\}} X \\ \mathbb{1}_{\{Z=2, U=1\}} X & \dots & \mathbb{1}_{\{Z=2, U=M\}} X \\ \vdots & \ddots & \vdots \\ \mathbb{1}_{\{Z=q, U=1\}} X & \dots & \mathbb{1}_{\{Z=q, U=M\}} X \end{pmatrix},$$

## 4.2 Estimation et convergence du critère

et  $(A_i)_{1 \leq i \leq n}$  un échantillon i.i.d de  $A$  tel que :

$$A_i = \begin{pmatrix} \mathbb{1}_{\{Z_i=1, U_i=1\}} X_i & \cdots & \mathbb{1}_{\{Z_i=1, U_i=M\}} X_i \\ \mathbb{1}_{\{Z_i=2, U_i=1\}} X_i & \cdots & \mathbb{1}_{\{Z_i=2, U_i=M\}} X_i \\ \vdots & \ddots & \vdots \\ \mathbb{1}_{\{Z_i=q, U_i=1\}} X_i & \cdots & \mathbb{1}_{\{Z_i=q, U_i=M\}} X_i \end{pmatrix}, i = 1, \dots, n$$

à valeurs dans  $\mathcal{M}_{qp, M}(\mathbb{R})$ , l'espace des matrices réelles a  $q \times p$  lignes et  $M$  colonnes ;

$$B = \begin{pmatrix} \mathbb{1}_{\{U=1\}} X \\ \mathbb{1}_{\{U=2\}} X \\ \vdots \\ \mathbb{1}_{\{U=M\}} X \end{pmatrix},$$

et  $(B_i)_{1 \leq i \leq n}$  un échantillon i.i.d de  $B$  tel que :

$$B_i = \begin{pmatrix} \mathbb{1}_{\{U_i=1\}} X_i \\ \mathbb{1}_{\{U_i=2\}} X_i \\ \vdots \\ \mathbb{1}_{\{U_i=M\}} X_i \end{pmatrix}, i = 1, \dots, n$$

à valeurs dans  $\mathbb{R}^{pM}$  ;

$$C = \begin{pmatrix} \mathbb{1}_{\{Z=1, U=1\}} & \cdots & \mathbb{1}_{\{Z=1, U=M\}} \\ \mathbb{1}_{\{Z=2, U=1\}} & \cdots & \mathbb{1}_{\{Z=2, U=M\}} \\ \vdots & \ddots & \vdots \\ \mathbb{1}_{\{Z=q, U=1\}} & \cdots & \mathbb{1}_{\{Z=q, U=M\}} \end{pmatrix},$$

et  $(C_i)_{1 \leq i \leq n}$  un échantillon i.i.d de  $C$  tel que :

$$C_i = \begin{pmatrix} \mathbb{1}_{\{Z_i=1, U_i=1\}} & \cdots & \mathbb{1}_{\{Z_i=1, U_i=M\}} \\ \mathbb{1}_{\{Z_i=2, U_i=1\}} & \cdots & \mathbb{1}_{\{Z_i=2, U_i=M\}} \\ \vdots & \ddots & \vdots \\ \mathbb{1}_{\{Z_i=q, U_i=1\}} & \cdots & \mathbb{1}_{\{Z_i=q, U_i=M\}} \end{pmatrix}, i = 1, \dots, n$$

à valeurs dans  $\mathcal{M}_{q, M}(\mathbb{R})$  l'espace des matrices réelles a  $q$  lignes et  $M$  colonnes ;

## 4.2 Estimation et convergence du critère

$$D = \begin{pmatrix} \mathbb{1}_{\{U=1\}} \\ \mathbb{1}_{\{U=2\}} \\ \vdots \\ \mathbb{1}_{\{U=M\}} \end{pmatrix},$$

et  $(D_i)_{1 \leq i \leq n}$  un échantillon i.i.d de  $D$  tel que :

$$D_i = \begin{pmatrix} \mathbb{1}_{\{U_i=1\}} \\ \mathbb{1}_{\{U_i=2\}} \\ \vdots \\ \mathbb{1}_{\{U_i=M\}} \end{pmatrix}, i = 1, \dots, n$$

à valeurs dans  $\mathbb{R}^M$  ;

$$E = \begin{pmatrix} \mathbb{1}_{\{U=1\}} X \otimes X \\ \mathbb{1}_{\{U=2\}} X \otimes X \\ \vdots \\ \mathbb{1}_{\{U=M\}} X \otimes X \end{pmatrix},$$

$(E_i)_{1 \leq i \leq n}$  un échantillon i.i.d de  $E$  tel que :

$$E_i = \begin{pmatrix} \mathbb{1}_{\{U_i=1\}} X_i \otimes X_i \\ \mathbb{1}_{\{U_i=2\}} X_i \otimes X_i \\ \vdots \\ \mathbb{1}_{\{U_i=M\}} X_i \otimes X_i \end{pmatrix}, i = 1, \dots, n$$

à valeurs dans  $\mathcal{L}(\mathbb{R}^p)^M$ . Soit la variable aléatoire  $T = (A, B, C, D, E)$  et  $(T_i)_{1 \leq i \leq n}$  un échantillon i.i.d de  $T$  avec  $T_i = (A_i, B_i, C_i, D_i, E_i)$ , à valeurs dans  $\mathcal{M}_{qp,M}(\mathbb{R}) \times \mathbb{R}^{pM} \times \mathcal{M}_{q,M}(\mathbb{R}) \times \mathbb{R}^M \times \mathcal{L}(\mathbb{R}^p)^M$ . Posons :

$$F = \mathcal{M}_{qp,M}(\mathbb{R}) \times \mathbb{R}^{pM} \times \mathcal{M}_{q,M}(\mathbb{R}) \times \mathbb{R}^M \times \mathcal{L}(\mathbb{R}^p)^M,$$

et pour tout  $(a, b, c, d, e) \in F$ , on a :

## 4.2 Estimation et convergence du critère

$$a = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1,M} \\ a_{2,1} & a_{2,2} & \dots & a_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ a_{q,1} & a_{q,2} & \dots & a_{q,M} \end{pmatrix}, \text{ où } a_{\ell,m} \in \mathbb{R}^p, 1 \leq \ell \leq q \text{ et } 1 \leq m \leq M;$$

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{pmatrix}, \text{ où } b_m \in \mathbb{R}^p, 1 \leq m \leq M;$$

$$c = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1,M} \\ c_{2,1} & c_{2,2} & \dots & c_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q,1} & c_{q,2} & \dots & c_{q,M} \end{pmatrix}, \text{ où } c_{\ell,m} \in \mathbb{R}, 1 \leq \ell \leq q \text{ et } 1 \leq m \leq M;$$

$$d = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_M \end{pmatrix}, \text{ où } d_m \in \mathbb{R}, 1 \leq m \leq M;$$

$$e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{pmatrix}, \text{ où } e_m \in \mathcal{L}(\mathbb{R}^p), 1 \leq m \leq M.$$

Considérons en plus les projections canoniques suivantes :

$$\pi_1^{\ell m} : (a, b, c, d, e) \in F \mapsto a_{\ell m} \in \mathbb{R}^p, 1 \leq \ell \leq q \text{ et } 1 \leq m \leq M,$$

$$\pi_2^m : (a, b, c, d, e) \in F \mapsto b_m \in \mathbb{R}^p, 1 \leq m \leq M,$$

$$\pi_3^{\ell m} : (a, b, c, d, e) \in F \mapsto c_{\ell m} \in \mathbb{R}, 1 \leq \ell \leq q \text{ et } 1 \leq m \leq M,$$

$$\pi_4^m : (a, b, c, d, e) \in F \mapsto d_m \in \mathbb{R}, 1 \leq m \leq M,$$

$$\pi_5^m : (a, b, c, d, e) \in F \mapsto e_m \in L(\mathbb{R}^p), 1 \leq m \leq M,$$

et posons  $\widehat{W}^{(n)} = \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n T_i - \mathbb{E}(T) \right)$ .

### 4.2.4 Principaux résultats

**Lemme 6**

$$\sqrt{n} \left( \hat{p}_m^{(n)} - p_m \right) = \pi_4^m \left( \widehat{W}^{(n)} \right),$$

et

$$\sqrt{n} \left( \hat{p}_{\ell|m}^{(n)} - p_{\ell|m} \right) = p_m^{-1} \left( \pi_3^{\ell m} \left( \widehat{W}^{(n)} \right) - \pi_4^m \left( \widehat{W}^{(n)} \right) \hat{p}_{\ell|m}^{(n)} \right).$$

**Preuve:** On a :

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n D_i - \mathbb{E}(D) \right) = \begin{pmatrix} \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{U_i=1\}} - \mathbb{E}(\mathbf{1}_{\{U=1\}}) \right) \\ \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{U_i=2\}} - \mathbb{E}(\mathbf{1}_{\{U=2\}}) \right) \\ \vdots \\ \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{U_i=M\}} - \mathbb{E}(\mathbf{1}_{\{U=M\}}) \right) \end{pmatrix}$$

$$\begin{aligned} \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n D_i - \mathbb{E}(D) \right) &= \begin{pmatrix} \sqrt{n} \left( \hat{p}_1^{(n)} - p_1 \right) \\ \sqrt{n} \left( \hat{p}_2^{(n)} - p_2 \right) \\ \vdots \\ \sqrt{n} \left( \hat{p}_M^{(n)} - p_M \right) \end{pmatrix} \\ &= \left( \sqrt{n} \left( \hat{p}_m^{(n)} - p_m \right) \right)_{1 \leq m \leq M} \\ &= \left( \pi_4^m \left( \widehat{W}^{(n)} \right) \right)_{1 \leq m \leq M}. \end{aligned}$$

Des deux dernières égalités, on en déduit que pour tout  $m \in \{1, \dots, M\}$ ,

$$\sqrt{n} \left( \hat{p}_m^{(n)} - p_m \right) = \pi_4^m \left( \widehat{W}^{(n)} \right).$$

De même

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n C_i - \mathbb{E}(C) \right)$$

## 4.2 Estimation et convergence du critère

$$\begin{aligned}
&= \begin{pmatrix} \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i=1, U_i=1\}} - \mathbb{E}(\mathbb{1}_{\{Z=1, U=1\}}) \right) & \cdots & \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i=1, U_i=M\}} - \mathbb{E}(\mathbb{1}_{\{Z=1, U=M\}}) \right) \\ \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i=2, U_i=1\}} - \mathbb{E}(\mathbb{1}_{\{Z=2, U=1\}}) \right) & \cdots & \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i=2, U_i=M\}} - \mathbb{E}(\mathbb{1}_{\{Z=2, U=M\}}) \right) \\ \vdots & \ddots & \vdots \\ \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i=q, U_i=1\}} - \mathbb{E}(\mathbb{1}_{\{Z=q, U=1\}}) \right) & \cdots & \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i=q, U_i=M\}} - \mathbb{E}(\mathbb{1}_{\{Z=q, U=M\}}) \right) \end{pmatrix} \\
&= \begin{pmatrix} \sqrt{n} (\hat{p}_{11}^{(n)} - p_{11}) & \cdots & \sqrt{n} (\hat{p}_{1M}^{(n)} - p_{1M}) \\ \sqrt{n} (\hat{p}_{21}^{(n)} - p_{21}) & \cdots & \sqrt{n} (\hat{p}_{2M}^{(n)} - p_{2M}) \\ \vdots & \ddots & \vdots \\ \sqrt{n} (\hat{p}_{q1}^{(n)} - p_{q1}) & \cdots & \sqrt{n} (\hat{p}_{qM}^{(n)} - p_{qM}) \end{pmatrix} \\
&= \left( \sqrt{n} (\hat{p}_{\ell,m}^{(n)} - p_{\ell,m}) \right)_{1 \leq \ell \leq q, 1 \leq m \leq M} \\
&= \left( \pi_3^{\ell m} (\widehat{W}^{(n)}) \right)_{1 \leq \ell \leq q, 1 \leq m \leq M}.
\end{aligned}$$

Pour tout  $\ell \in \{1, \dots, q\}$  et  $m \in \{1, \dots, M\}$

$$\sqrt{n} (\hat{p}_{\ell,m}^{(n)} - p_{\ell,m}) = \pi_3^{\ell m} (\widehat{W}^{(n)}).$$

Or,

$$\begin{aligned}
\hat{p}_{\ell|m}^{(n)} &= \frac{\widehat{N}_{\ell,m}^{(n)}}{\widehat{N}_m^{(n)}} = \frac{\widehat{N}_{\ell,m}^{(n)}/n}{\widehat{N}_m^{(n)}/n} \\
&= \frac{\hat{p}_{\ell,m}^{(n)}}{\hat{p}_m^{(n)}},
\end{aligned}$$

d'où

$$\hat{p}_{\ell,m}^{(n)} = \hat{p}_m^{(n)} \hat{p}_{\ell|m}^{(n)}.$$

Ainsi on a :

$$\left( \hat{p}_{\ell,m}^{(n)} - p_{\ell,m} \right) = \left( \hat{p}_m^{(n)} \hat{p}_{\ell|m}^{(n)} - p_m p_{\ell|m} \right).$$

Remarquons que :

$$\left( \hat{p}_m^{(n)} \hat{p}_{\ell|m}^{(n)} - p_m p_{\ell|m} \right) = \left( \hat{p}_m^{(n)} - p_m \right) \hat{p}_{\ell|m}^{(n)} + p_m \left( \hat{p}_{\ell|m}^{(n)} - p_{\ell|m} \right).$$

## 4.2 Estimation et convergence du critère

Il vient que

$$\begin{aligned}\sqrt{n} \left( \widehat{p}_m^{(n)} \widehat{p}_{\ell|m}^{(n)} - p_m p_{\ell|m} \right) &= \sqrt{n} \left( \widehat{p}_m^{(n)} - p_m \right) \widehat{p}_{\ell|m}^{(n)} + p_m \sqrt{n} \left( \widehat{p}_{\ell|m}^{(n)} - p_{\ell|m} \right) \\ &= \pi_3^{\ell m} \left( \widehat{W}^{(n)} \right).\end{aligned}$$

Pour tout  $\ell \in \{1, \dots, q\}$  et  $m \in \{1, \dots, M\}$ ,

$$\sqrt{n} \left( \widehat{p}_{\ell|m}^{(n)} - p_{\ell|m} \right) = p_m^{-1} \left( \pi_3^{\ell m} \left( \widehat{W}^{(n)} \right) - \pi_4^m \left( \widehat{W}^{(n)} \right) \widehat{p}_{\ell|m}^{(n)} \right)$$

□

**Lemme 7**

$$\sqrt{n} \left( \widehat{\mu}_m^{(n)} - \mu_m \right) = p_m^{-1} \left( \pi_2^m \left( \widehat{W}^{(n)} \right) - \pi_4^m \left( \widehat{W}^{(n)} \right) \widehat{\mu}_m^{(n)} \right),$$

et

$$\sqrt{n} \left( \widehat{\mu}_{\ell,m}^{(n)} - \mu_{\ell,m} \right) = p_{\ell,m}^{-1} \left( \pi_1^{\ell m} \left( \widehat{W}^{(n)} \right) - \pi_3^{\ell m} \left( \widehat{W}^{(n)} \right) \widehat{\mu}_{\ell,m}^{(n)} \right).$$

**Preuve:** On a :

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n B_i - \mathbb{E}(B) \right) = \begin{pmatrix} \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{U_i=1\}} X_i - \mathbb{E}(\mathbb{1}_{\{U=1\}} X) \right) \\ \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{U_i=2\}} X_i - \mathbb{E}(\mathbb{1}_{\{U=2\}} X) \right) \\ \vdots \\ \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{U_i=M\}} X_i - \mathbb{E}(\mathbb{1}_{\{U=M\}} X) \right) \end{pmatrix}$$



## 4.2 Estimation et convergence du critère

$$\begin{aligned}
\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n B_i - \mathbb{E}(B) \right) &= \begin{pmatrix} \sqrt{n} \left( \frac{\hat{p}_1^{(n)}}{\hat{N}_1^{(n)}} \sum_{i=1}^n \mathbf{1}_{\{U_i=1\}} X_i - p_1 \mathbb{E}(X|\{U=1\}) \right) \\ \sqrt{n} \left( \frac{\hat{p}_2^{(n)}}{\hat{N}_2^{(n)}} \sum_{i=1}^n \mathbf{1}_{\{U_i=2\}} X_i - p_2 \mathbb{E}(X|\{U=2\}) \right) \\ \vdots \\ \sqrt{n} \left( \frac{\hat{p}_M^{(n)}}{\hat{N}_M^{(n)}} \sum_{i=1}^n \mathbf{1}_{\{U_i=M\}} X_i - p_M \mathbb{E}(X|\{U=M\}) \right) \end{pmatrix} \\
&= \begin{pmatrix} \sqrt{n} \left( \hat{p}_1^{(n)} \hat{\mu}_1^{(n)} - p_1 \mu_1 \right) \\ \sqrt{n} \left( \hat{p}_2^{(n)} \hat{\mu}_2^{(n)} - p_2 \mu_2 \right) \\ \vdots \\ \sqrt{n} \left( \hat{p}_M^{(n)} \hat{\mu}_M^{(n)} - p_M \mu_M \right) \end{pmatrix} \\
&= \left( \sqrt{n} \left( \hat{p}_m^{(n)} \hat{\mu}_m^{(n)} - p_m \mu_m \right) \right)_{1 \leq m \leq M} \\
&= \left( \pi_2^m \left( \widehat{W}^{(n)} \right) \right)_{1 \leq m \leq M}.
\end{aligned}$$

Pour tout  $m \in \{1, \dots, M\}$ ,

$$\sqrt{n} \left( \hat{p}_m^{(n)} \hat{\mu}_m^{(n)} - p_m \mu_m \right) = \pi_2^m \left( \widehat{W}^{(n)} \right).$$

Or,

$$\left( \hat{p}_m^{(n)} \hat{\mu}_m^{(n)} - p_m \mu_m \right) = \left( \hat{p}_m^{(n)} - p_m \right) \hat{\mu}_m^{(n)} + p_m \left( \hat{\mu}_m^{(n)} - \mu_m \right)$$

D'où  $\pi_2^m \left( \widehat{W}^{(n)} \right) = \sqrt{n} \left( \left( \hat{p}_m^{(n)} - p_m \right) \hat{\mu}_m^{(n)} + p_m \left( \hat{\mu}_m^{(n)} - \mu_m \right) \right)$ . On en déduit que

$$\sqrt{n} \left( \hat{\mu}_m^{(n)} - \mu_m \right) = p_m^{-1} \left( \pi_2^m \left( \widehat{W}^{(n)} \right) - \pi_4^m \left( \widehat{W}^{(n)} \right) \hat{\mu}_m^{(n)} \right).$$

De même

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n A_i - \mathbb{E}(A) \right) = \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1M} \\ R_{21} & R_{22} & \cdots & R_{2M} \\ \vdots & \vdots & \cdots & \vdots \\ R_{q1} & R_{q2} & \cdots & R_{qM} \end{pmatrix} = \begin{pmatrix} \tilde{R}_{11} & \tilde{R}_{12} & \cdots & \tilde{R}_{1M} \\ \tilde{R}_{21} & \tilde{R}_{22} & \cdots & \tilde{R}_{2M} \\ \vdots & \vdots & \cdots & \vdots \\ \tilde{R}_{q1} & \tilde{R}_{q2} & \cdots & \tilde{R}_{qM} \end{pmatrix}$$

## 4.2 Estimation et convergence du critère

où  $\forall (\ell, m) \in \{1, \dots, q\} \times \{1, \dots, M\}$ ,

$$R_{\ell,m} = \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i=\ell, U_i=m\}} X_i - \mathbb{E}(\mathbb{1}_{\{Z=\ell, U=m\}} X) \right)$$

et

$$\tilde{R}_{\ell,m} = \sqrt{n} \left( \frac{\hat{p}_{\ell,m}^{(n)}}{\widehat{N}_{\ell,m}^{(n)}} \sum_{i=1}^n \mathbb{1}_{\{Z_i=\ell, U_i=m\}} X_i - p_{\ell,m} \mathbb{E}(X | \{Z = \ell, U = m\}) \right).$$

D'où

$$\begin{aligned} \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n A_i - \mathbb{E}(A) \right) &= \begin{pmatrix} \sqrt{n} \left( \hat{p}_{11}^{(n)} \hat{\mu}_{11}^{(n)} - p_{11} \mu_{11} \right) & \dots & \sqrt{n} \left( \hat{p}_{1M}^{(n)} \hat{\mu}_{1M}^{(n)} - p_{1M} \mu_{1M} \right) \\ \sqrt{n} \left( \hat{p}_{21}^{(n)} \hat{\mu}_{21}^{(n)} - p_{21} \mu_{21} \right) & \dots & \sqrt{n} \left( \hat{p}_{2M}^{(n)} \hat{\mu}_{2M}^{(n)} - p_{2M} \mu_{2M} \right) \\ \vdots & \ddots & \vdots \\ \sqrt{n} \left( \hat{p}_{q1}^{(n)} \hat{\mu}_{q1}^{(n)} - p_{q1} \mu_{q1} \right) & \dots & \sqrt{n} \left( \hat{p}_{qM}^{(n)} \hat{\mu}_{qM}^{(n)} - p_{qM} \mu_{qM} \right) \end{pmatrix} \\ &= \left( \sqrt{n} \left( \hat{p}_{\ell,m}^{(n)} \hat{\mu}_{\ell,m}^{(n)} - p_{\ell,m} \mu_{\ell,m} \right) \right)_{1 \leq \ell \leq q, 1 \leq m \leq M} \\ &= \left( \pi_1^{\ell m} \left( \widehat{W}^{(n)} \right) \right)_{1 \leq \ell \leq q, 1 \leq m \leq M}. \end{aligned}$$

D'où, pour tout  $\ell \in \{1, \dots, q\}$  et  $m \in \{1, \dots, M\}$ ,

$$\sqrt{n} \left( \hat{p}_{\ell,m}^{(n)} \hat{\mu}_{\ell,m}^{(n)} - p_{\ell,m} \mu_{\ell,m} \right) = \pi_1^{\ell m} \left( \widehat{W}^{(n)} \right).$$

Or

$$\left( \hat{p}_{\ell,m}^{(n)} \hat{\mu}_{\ell,m}^{(n)} - p_{\ell,m} \mu_{\ell,m} \right) = \left( \hat{p}_{\ell,m}^{(n)} - p_{\ell,m} \right) \hat{\mu}_{\ell,m}^{(n)} + p_{\ell,m} \left( \hat{\mu}_{\ell,m}^{(n)} - \mu_{\ell,m} \right).$$

Alors  $\pi_1^{\ell m} \left( \widehat{W}^{(n)} \right) = \sqrt{n} \left( \left( \hat{p}_{\ell,m}^{(n)} - p_{\ell,m} \right) \hat{\mu}_{\ell,m}^{(n)} + p_{\ell,m} \left( \hat{\mu}_{\ell,m}^{(n)} - \mu_{\ell,m} \right) \right)$ . On en déduit que :

$$\sqrt{n} \left( \hat{\mu}_{\ell,m}^{(n)} - \mu_{\ell,m} \right) = p_{\ell,m}^{-1} \left( \pi_1^{\ell m} \left( \widehat{W}^{(n)} \right) - \pi_3^{\ell m} \left( \widehat{W}^{(n)} \right) \hat{\mu}_{\ell,m}^{(n)} \right).$$

□

**Lemme 8**

$$\begin{aligned}
\sqrt{n} \left( \widehat{V}_m^{(n)} - V_m \right) &= p_m^{-1} \left[ \pi_5^m \left( \widehat{W}^{(n)} \right) - \pi_4^m \left( \widehat{W}^{(n)} \right) \left( \widehat{V}_m^{(n)} + \widehat{\mu}_m^{(n)} \otimes \widehat{\mu}_m^{(n)} \right) \right] \\
&\quad - p_m^{-1} \left[ \left( \pi_2^m \left( \widehat{W}^{(n)} \right) - \pi_4^m \left( \widehat{W}^{(n)} \right) \widehat{\mu}_m^{(n)} \right) \otimes \widehat{\mu}_m^{(n)} \right. \\
&\quad \left. + \mu_m \otimes \left( \pi_2^m \left( \widehat{W}^{(n)} \right) - \pi_4^m \left( \widehat{W}^{(n)} \right) \widehat{\mu}_m^{(n)} \right) \right].
\end{aligned}$$

**Preuve:** On a :

$$\begin{aligned}
&\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n E_i - \mathbb{E}(E) \right) \\
&= \left( \begin{array}{c} \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{U_i=1\}} X_i \otimes X_i - \mathbb{E}(\mathbb{1}_{\{U=1\}} X \otimes X) \right) \\ \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{U_i=2\}} X_i \otimes X_i - \mathbb{E}(\mathbb{1}_{\{U=2\}} X \otimes X) \right) \\ \vdots \\ \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{U_i=M\}} X_i \otimes X_i - \mathbb{E}(\mathbb{1}_{\{U=M\}} X \otimes X) \right) \end{array} \right) \\
&= \left( \begin{array}{c} \sqrt{n} \left( \frac{\widehat{p}_1^{(n)}}{\widehat{N}_1^{(n)}} \sum_{i=1}^n \mathbb{1}_{\{U_i=1\}} X_i \otimes X_i - p_1 \mathbb{E}(X \otimes X | \{U=1\}) \right) \\ \sqrt{n} \left( \frac{\widehat{p}_2^{(n)}}{\widehat{N}_2^{(n)}} \sum_{i=1}^n \mathbb{1}_{\{U_i=2\}} X_i \otimes X_i - p_2 \mathbb{E}(X \otimes X | \{U=2\}) \right) \\ \vdots \\ \sqrt{n} \left( \frac{\widehat{p}_M^{(n)}}{\widehat{N}_M^{(n)}} \sum_{i=1}^n \mathbb{1}_{\{U_i=M\}} X_i \otimes X_i - p_M \mathbb{E}(X \otimes X | \{U=M\}) \right) \end{array} \right) \\
&= \left( \sqrt{n} \left( \widehat{p}_m^{(n)} \frac{1}{\widehat{N}_m^{(n)}} \sum_{i=1}^n \mathbb{1}_{\{U_i=m\}} X_i \otimes X_i - p_m \mathbb{E}(X \otimes X | \{U=m\}) \right) \right)_{1 \leq m \leq M} \\
&= \left( \pi_5^m \left( \widehat{W}^{(n)} \right) \right)_{1 \leq m \leq M}.
\end{aligned}$$

Or

$$\widehat{V}_m^{(n)} = \frac{1}{\widehat{N}_m^{(n)}} \sum_{i=1}^n \mathbb{1}_{\{U_i=m\}} X_i \otimes X_i - \widehat{\mu}_m^{(n)} \otimes \widehat{\mu}_m^{(n)},$$

d'où

$$\frac{1}{\widehat{N}_m^{(n)}} \sum_{i=1}^n \mathbb{1}_{\{U_i=m\}} X_i \otimes X_i = \widehat{V}_m^{(n)} + \widehat{\mu}_m^{(n)} \otimes \widehat{\mu}_m^{(n)}.$$

et

$$V_m = \mathbb{E}(X \otimes X | U = m) - \mu_m \otimes \mu_m \Rightarrow \mathbb{E}(X \otimes X | U = m) = V_m + \mu_m \otimes \mu_m.$$

Finalement pour tout  $m \in \{1, \dots, M\}$

$$\begin{aligned}\pi_5^m(\widehat{W}^{(n)}) &= \sqrt{n}(\widehat{p}_m^{(n)}\widehat{V}_m^{(n)} + \widehat{p}_m^{(n)}\widehat{\mu}_m^{(n)} \otimes \widehat{\mu}_m^{(n)} - p_m V_m - p_m \mu_m \otimes \mu_m) \\ &= \sqrt{n}(\widehat{p}_m^{(n)} - p_m)\widehat{V}_m^{(n)} + p_m \sqrt{n}(\widehat{V}_m^{(n)} - V_m) \\ &\quad + \sqrt{n}(\widehat{p}_m^{(n)} - p_m)\widehat{\mu}_m^{(n)} \otimes \widehat{\mu}_m^{(n)} + p_m \sqrt{n}(\widehat{\mu}_m^{(n)} \otimes \widehat{\mu}_m^{(n)} - \mu_m \otimes \mu_m)\end{aligned}$$

$$\begin{aligned}\pi_5^m(\widehat{W}^{(n)}) &= \sqrt{n}(\widehat{p}_m^{(n)} - p_m)\widehat{V}_m^{(n)} + p_m \sqrt{n}(\widehat{V}_m^{(n)} - V_m) \\ &\quad + \sqrt{n}(\widehat{p}_m^{(n)} - p_m)\widehat{\mu}_m^{(n)} \otimes \widehat{\mu}_m^{(n)} + p_m \sqrt{n}(\widehat{\mu}_m^{(n)} - \mu_m) \otimes \widehat{\mu}_m^{(n)} \\ &\quad + \mu_m p_m \otimes \sqrt{n}(\widehat{\mu}_m^{(n)} - \mu_m) \\ &= p_m \sqrt{n}(\widehat{V}_m^{(n)} - V_m) + \sqrt{n}(\widehat{p}_m^{(n)} - p_m) (\widehat{V}_m^{(n)} + \widehat{\mu}_m^{(n)} \otimes \widehat{\mu}_m^{(n)}) \\ &\quad + p_m (\sqrt{n}(\widehat{\mu}_m^{(n)} - \mu_m) \otimes \widehat{\mu}_m^{(n)} + \mu_m \otimes \sqrt{n}(\widehat{\mu}_m^{(n)} - \mu_m)).\end{aligned}$$

Alors

$$\begin{aligned}p_m \sqrt{n}(\widehat{V}_m^{(n)} - V_m) &= \pi_5^m(\widehat{W}^{(n)}) - \sqrt{n}(\widehat{p}_m^{(n)} - p_m) (\widehat{V}_m^{(n)} + \widehat{\mu}_m^{(n)} \otimes \widehat{\mu}_m^{(n)}) \\ &\quad - p_m (\sqrt{n}(\widehat{\mu}_m^{(n)} - \mu_m) \otimes \widehat{\mu}_m^{(n)} + \mu_m \otimes \sqrt{n}(\widehat{\mu}_m^{(n)} - \mu_m)) \\ \sqrt{n}(\widehat{V}_m^{(n)} - V_m) &= p_m^{-1} (\pi_5^m(\widehat{W}^{(n)}) - \sqrt{n}(\widehat{p}_m^{(n)} - p_m) (\widehat{V}_m^{(n)} + \widehat{\mu}_m^{(n)} \otimes \widehat{\mu}_m^{(n)})) \\ &\quad - (\sqrt{n}(\widehat{\mu}_m^{(n)} - \mu_m) \otimes \widehat{\mu}_m^{(n)} + \mu_m \otimes \sqrt{n}(\widehat{\mu}_m^{(n)} - \mu_m))\end{aligned}$$

$$\begin{aligned}\sqrt{n}(\widehat{V}_m^{(n)} - V_m) &= p_m^{-1} (\pi_5^m(\widehat{W}^{(n)}) - \pi_4^m(\widehat{W}^{(n)}) (\widehat{V}_m^{(n)} + \widehat{\mu}_m^{(n)} \otimes \widehat{\mu}_m^{(n)})) \\ &\quad - p_m^{-1} ((\pi_2^m(\widehat{W}^{(n)}) - \pi_4^m(\widehat{W}^{(n)})\widehat{\mu}_m^{(n)}) \otimes \widehat{\mu}_m^{(n)} \\ &\quad + \mu_m \otimes (\pi_2^m(\widehat{W}^{(n)}) - \pi_4^m(\widehat{W}^{(n)})\widehat{\mu}_m^{(n)}))\end{aligned}$$

□

Dans le Théorème 9, nous établissons la convergence du critère  $n\widehat{\xi}_{K|m}^{(n)}$ , pour  $m \in \{1, \dots, M\}$  fixé.

**Théorème 9**  $\forall K \subset I$  et  $\forall m \in \{1, \dots, M\}$ , on a :

1.  $\widehat{\xi}_{K|m}^{(n)}$  converge presque sûrement vers  $\xi_{K|m}$  quand  $n \rightarrow +\infty$ .

## 4.2 Estimation et convergence du critère

$$2. n\widehat{\xi}_{K|m}^{(n)} = \sum_{\ell=1}^q \left( \widehat{\Phi}_{\ell,K|m}^{(n)}(\widehat{W}^{(n)}) + p_{\ell|m} \|\widehat{\Psi}_{\ell,K|m}^{(n)}(\widehat{W}^{(n)}) + \sqrt{n}\Delta_{\ell,K|m}\|_{\mathbb{R}^p} \right)^2$$

où  $(\widehat{\Phi}_{\ell,K|m}^{(n)})_{n \in \mathbb{N}}$  et  $(\widehat{\Psi}_{\ell,K|m}^{(n)})_{n \in \mathbb{N}}$  sont des suites aléatoires d'opérateurs linéaires continus définis dans  $\mathcal{L}(F, \mathbb{R})$ , et qui convergent presque sûrement vers  $\Phi_{\ell,K|m}$  (resp.  $\Psi_{\ell,K|m}$ ) quand  $n \rightarrow +\infty$ , tels que pour tout  $\mathbf{T} \in F$ ,

$$\widehat{\Phi}_{\ell,K|m}^{(n)}(\mathbf{T}) = p_m^{-1} \left( \pi_3^{\ell m}(\mathbf{T}) - \pi_4^m(\mathbf{T}) \widehat{p}_{\ell|m}^{(n)} \right) \|\widehat{\Delta}_{\ell,K|m}^{(n)}\|_{\mathbb{R}^p},$$

et

$$\begin{aligned} \widehat{\Psi}_{\ell,K|m}^{(n)}(\mathbf{T}) &= \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left\{ p_{\ell,m}^{-1} \left( \pi_1^{\ell m}(\mathbf{T}) - \pi_3^{\ell m}(\mathbf{T}) \widehat{\mu}_{\ell,m}^{(n)} \right) \right. \\ &\quad \left. - p_m^{-1} \left( \pi_2^m(\mathbf{T}) - \pi_4^m(\mathbf{T}) \widehat{\mu}_m^{(n)} \right) \right\} - p_m^{-1} \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \\ &\quad \times \left\{ \left( \pi_5^m(\mathbf{T}) - \pi_4^m(\mathbf{T}) \left( \widehat{V}_m^{(n)} + \widehat{\mu}_m^{(n)} \otimes \widehat{\mu}_m^{(n)} \right) \right) \right. \\ &\quad \times \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \left. \right\} \\ &\quad + p_m^{-1} \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left\{ \left( \left( \pi_2^m(\mathbf{T}) - \pi_4^m(\mathbf{T}) \widehat{\mu}_m^{(n)} \right) \otimes \widehat{\mu}_m^{(n)} \right) \right. \\ &\quad \times \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \left. \right\} \\ &\quad + p_m^{-1} \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left\{ \left( \mu_m \otimes \left( \pi_2^m(\mathbf{T}) - \pi_4^m(\mathbf{T}) \widehat{\mu}_m^{(n)} \right) \right) \right. \\ &\quad \times \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \left. \right\}. \end{aligned}$$

**Preuve:** On a :

1.  $\forall K \subset I, \forall m \in \{1, \dots, M\}$  et  $\forall \ell \in \{1, \dots, q\}$ , on a

$$\widehat{\mu}_{\ell,m}^{(n)} = \frac{1}{\widehat{N}_{\ell m}^{(n)}} \sum_{i=1}^n \mathbb{1}_{\{Z_i=\ell, U_i=m\}} X_i = \widehat{p}_{\ell,m}^{(n)-1} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i=\ell, U_i=m\}} X_i.$$

D'après la loi forte des grands nombres,  $\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i=\ell, U_i=m\}} X_i$  converge presque sûrement vers  $\mathbb{E} \left( \mathbb{1}_{\{Z=\ell, U=m\}} X \right)$  quand  $n \rightarrow +\infty$ , et  $\widehat{p}_{\ell,m}^{(n)} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i=\ell, U_i=m\}}$  converge presque sûrement vers  $\mathbb{E} \left( \mathbb{1}_{\{Z=\ell, U=m\}} \right) = p_{\ell,m}$  quand  $n \rightarrow +\infty$ . Comme l'application  $x \mapsto x^{-1}$  est continue, on en déduit que  $\widehat{p}_{\ell,m}^{(n)-1}$  converge presque sûrement vers  $p_{\ell,m}^{-1}$  quand  $n \rightarrow +\infty$ . Alors  $\widehat{\mu}_{\ell,m}^{(n)} = \widehat{p}_{\ell,m}^{(n)-1} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i=\ell, U_i=m\}} X_i$  converge presque sûrement vers  $p_{\ell,m}^{-1} \mathbb{E} \left( \mathbb{1}_{\{Z=\ell, U=m\}} X \right)$  quand  $n \rightarrow +\infty$ . Or,

$$p_{\ell,m}^{-1} \mathbb{E} \left( \mathbb{1}_{\{Z=\ell, U=m\}} X \right) = \mathbb{E} (X | Z = \ell, U = m) = \mu_{\ell,m}$$

## 4.2 Estimation et convergence du critère

On en déduit que  $\hat{\mu}_{\ell,m}^{(n)}$  converge presque sûrement vers  $\mu_{\ell,m}$  quand  $n \rightarrow +\infty$ . De façon similaire, on montre que  $\hat{\mu}_m^{(n)}$  et  $\hat{\mathbb{V}}_m^{(n)}$  convergent presque sûrement respectivement vers  $\mu_m$  et  $\mathbb{V}_m$  quand  $n \rightarrow +\infty$ , d'où  $\hat{Q}_{K|m}^{(n)}$  converge presque sûrement vers  $Q_{K|m}$  quand  $n \rightarrow +\infty$ .

$\hat{p}_m^{(n)} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{U_i=m\}}$  converge presque sûrement vers  $\mathbb{E}(\mathbb{1}_{\{U=m\}}) = p_m$  quand  $n \rightarrow +\infty$ . Ainsi  $\hat{p}_{\ell|m}^{(n)} = \frac{\hat{p}_{\ell,m}^{(n)}}{\hat{p}_m^{(n)}}$  converge presque sûrement vers  $p_{\ell|m}$  quand  $n \rightarrow +\infty$  comme quotient d'estimateurs qui convergent presque sûrement. Ainsi, on en déduit que  $\hat{\xi}_{K|m}^{(n)}$  converge presque sûrement vers  $\xi_{K|m}$  quand  $n \rightarrow +\infty$ .

2. On a :

$$\begin{aligned} n\hat{\xi}_{K|m}^{(n)} &= \sum_{\ell=1}^q \left(\hat{p}_{\ell|m}^{(n)}\right)^2 n \left\| \left( I_{\mathbb{R}^p} - \hat{\mathbb{V}}_m^{(n)} \hat{Q}_{K|m}^{(n)} \right) \left( \hat{\mu}_{\ell,m}^{(n)} - \hat{\mu}_m^{(n)} \right) \right\|_{\mathbb{R}^p}^2 \\ &= \sum_{\ell=1}^q \left(\hat{p}_{\ell|m}^{(n)}\right)^2 \left\| \sqrt{n} \left( I_{\mathbb{R}^p} - \hat{\mathbb{V}}_m^{(n)} \hat{Q}_{K|m}^{(n)} \right) \left( \hat{\mu}_{\ell,m}^{(n)} - \hat{\mu}_m^{(n)} \right) \right\|_{\mathbb{R}^p}^2 \end{aligned}$$

$\forall m \in \{1, \dots, M\}$  et  $\forall \ell \in \{1, \dots, q\}$ , posons :

$$\hat{\Delta}_{\ell,K|m}^{(n)} = \left( I_{\mathbb{R}^p} - \hat{\mathbb{V}}_m^{(n)} \hat{Q}_{K|m}^{(n)} \right) \left( \hat{\mu}_{\ell,m}^{(n)} - \hat{\mu}_m^{(n)} \right)$$

on a :

$$\begin{aligned} \hat{\Delta}_{\ell,K|m}^{(n)} &= \left( \hat{\mu}_{\ell,m}^{(n)} - \hat{\mu}_m^{(n)} \right) - \hat{\mathbb{V}}_m^{(n)} \hat{Q}_{K|m}^{(n)} \left( \hat{\mu}_{\ell,m}^{(n)} - \hat{\mu}_m^{(n)} \right) \\ &= \left[ \left( \hat{\mu}_{\ell,m}^{(n)} - \hat{\mu}_m^{(n)} \right) - \left( \mu_{\ell,m} - \mu_m \right) + \left( \mu_{\ell,m} - \mu_m \right) \right] \\ &\quad - \left( \hat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \hat{Q}_{K|m}^{(n)} \left( \hat{\mu}_{\ell,m}^{(n)} - \hat{\mu}_m^{(n)} \right) \\ &\quad - \mathbb{V}_m \left( \hat{Q}_{K|m}^{(n)} - Q_{K|m} \right) \left( \hat{\mu}_{\ell,m}^{(n)} - \hat{\mu}_m^{(n)} \right) \\ &\quad - \mathbb{V}_m Q_{K|m} \left[ \left( \hat{\mu}_{\ell,m}^{(n)} - \hat{\mu}_m^{(n)} \right) - \left( \mu_{\ell,m} - \mu_m \right) \right] \\ &\quad - \mathbb{V}_m Q_{K|m} \left( \mu_{\ell,m} - \mu_m \right) \\ &= \left[ \left( \hat{\mu}_{\ell,m}^{(n)} - \hat{\mu}_m^{(n)} \right) - \left( \mu_{\ell,m} - \mu_m \right) \right] \\ &\quad - \left( \hat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \hat{Q}_{K|m}^{(n)} \left( \hat{\mu}_{\ell,m}^{(n)} - \hat{\mu}_m^{(n)} \right) \\ &\quad - \mathbb{V}_m \left( \hat{Q}_{K|m}^{(n)} - Q_{K|m} \right) \left( \hat{\mu}_{\ell,m}^{(n)} - \hat{\mu}_m^{(n)} \right) \\ &\quad - \mathbb{V}_m Q_{K|m} \left[ \left( \hat{\mu}_{\ell,m}^{(n)} - \hat{\mu}_m^{(n)} \right) - \left( \mu_{\ell,m} - \mu_m \right) \right] + \Delta_{\ell,K|m}. \end{aligned}$$

Ainsi

## 4.2 Estimation et convergence du critère

$$\begin{aligned}
\widehat{\Delta}_{\ell,K|m}^{(n)} &= \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left[ \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) - (\mu_{\ell,m} - \mu_m) \right] \\
&- \left( \widehat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \\
&- \mathbb{V}_m \left( \widehat{Q}_{K|m}^{(n)} - Q_{K|m} \right) \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) + \Delta_{\ell,K|m}.
\end{aligned}$$

En ajoutant puis en retranchant le terme  $\left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left( \widehat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right)$  à l'équation précédente, on obtient :

$$\begin{aligned}
\widehat{\Delta}_{\ell,K|m}^{(n)} &= \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left[ \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) - (\mu_{\ell,m} - \mu_m) \right] \\
&- \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left( \widehat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \\
&+ \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left( \widehat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \\
&- \left( \widehat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \\
&- \mathbb{V}_m \left( \widehat{Q}_{K|m}^{(n)} - Q_{K|m} \right) \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) + \Delta_{\ell,K|m}.
\end{aligned}$$

D'où

$$\begin{aligned}
\widehat{\Delta}_{\ell,K|m}^{(n)} &= \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left[ \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \right. \\
&- \left. (\mu_{\ell,m} - \mu_m) - \left( \widehat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \right] \\
&- \mathbb{V}_m Q_{K|m} \left( \widehat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \\
&- \mathbb{V}_m \left( \widehat{Q}_{K|m}^{(n)} - Q_{K|m} \right) \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) + \Delta_{\ell,K|m},
\end{aligned}$$

$$\begin{aligned}
\widehat{\Delta}_{\ell,K|m}^{(n)} &= \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left[ \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) - (\mu_{\ell,m} - \mu_m) \right. \\
&- \left. \left( \widehat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \right] \\
&- \mathbb{V}_m Q_{K|m} \left( \widehat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \\
&+ \mathbb{V}_m \left( Q_{K|m} - \widehat{Q}_{K|m}^{(n)} \right) \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) + \Delta_{\ell,K|m} \\
\widehat{\Delta}_{\ell,K,m}^{(n)} &= \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K,m} \right) \left[ \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) - (\mu_{\ell,m} - \mu_m) \right. \\
&- \left. \left( \widehat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \right] \\
&- \mathbb{V}_m Q_{K|m} \left( \widehat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \\
&+ \mathbb{V}_m \left[ -Q_{K|m} \left( \mathbb{V}_m - \widehat{\mathbb{V}}_m^{(n)} \right) \widehat{Q}_{K|m}^{(n)} \right] \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) + \Delta_{\ell,K|m},
\end{aligned}$$

## 4.2 Estimation et convergence du critère

$$\begin{aligned}
\widehat{\Delta}_{\ell,K|m}^{(n)} &= \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left[ \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) - (\mu_{\ell,m} - \mu_m) \right. \\
&\quad - \left. \left( \widehat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \right] \\
&\quad - \mathbb{V}_m Q_{K|m} \left( \widehat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \\
&\quad + \mathbb{V}_m Q_{K|m} \left( \widehat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) + \Delta_{\ell,K|m}.
\end{aligned}$$

D'où

$$\begin{aligned}
\widehat{\Delta}_{\ell,K|m}^{(n)} &= \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left[ \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) - (\mu_{\ell,m} - \mu_m) \right. \\
&\quad - \left. \left( \widehat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \right] + \Delta_{\ell,K|m} \\
&= \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left[ \left( \widehat{\mu}_{\ell,m}^{(n)} - \mu_{\ell,m} \right) - \left( \widehat{\mu}_m^{(n)} - \mu_m \right) \right. \\
&\quad - \left. \left( \widehat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \right] + \Delta_{\ell,K|m} \\
\sqrt{n} \widehat{\Delta}_{\ell,K|m}^{(n)} &= \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left[ \sqrt{n} \left( \widehat{\mu}_{\ell,m}^{(n)} - \mu_{\ell,m} \right) - \sqrt{n} \left( \widehat{\mu}_m^{(n)} - \mu_m \right) \right. \\
&\quad - \left. \left( \sqrt{n} \left( \widehat{\mathbb{V}}_m^{(n)} - \mathbb{V}_m \right) \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \right] + \sqrt{n} \Delta_{\ell,K|m}.
\end{aligned}$$

En utilisant les Lemmes 7 et 8, on obtient :

$$\begin{aligned}
\sqrt{n} \widehat{\Delta}_{\ell,K|m}^{(n)} &= \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left[ p_{\ell,m}^{-1} \left( \pi_1^{\ell m} (\widehat{W}^{(n)}) - \pi_3^{\ell m} (\widehat{W}^{(n)}) \widehat{\mu}_{\ell,m}^{(n)} \right) \right. \\
&\quad - \left. p_m^{-1} \left( \pi_2^m (\widehat{W}^{(n)}) - \pi_4^m (\widehat{W}^{(n)}) \widehat{\mu}_m^{(n)} \right) \right] \\
&\quad - p_m^{-1} \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left[ \left( \pi_5^m (\widehat{W}^{(n)}) - \pi_4^m (\widehat{W}^{(n)}) \right) \right. \\
&\quad \times \left. \left( \widehat{\mathbb{V}}_m^{(n)} + \widehat{\mu}_m^{(n)} \otimes \widehat{\mu}_m^{(n)} \right) \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \right] \\
&\quad + p_m^{-1} \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left[ \left( \left( \pi_2^m (\widehat{W}^{(n)}) - \pi_4^m (\widehat{W}^{(n)}) \widehat{\mu}_m^{(n)} \right) \otimes \widehat{\mu}_m^{(n)} \right) \right. \\
&\quad \times \left. \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \right] \\
&\quad + p_m^{-1} \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left[ \left( \mu_m \otimes \left( \pi_2^m (\widehat{W}^{(n)}) - \pi_4^m (\widehat{W}^{(n)}) \widehat{\mu}_m^{(n)} \right) \right) \right. \\
&\quad \times \left. \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \right] + \sqrt{n} \Delta_{\ell,K|m} \\
&= \widehat{\Psi}_{\ell,K|m}^{(n)} \left( \widehat{W}^{(n)} \right) + \sqrt{n} \Delta_{\ell,K|m}
\end{aligned}$$

où



$$\begin{aligned}
\widehat{\Psi}_{\ell,K|m}^{(n)}(\mathbf{T}) &= \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left[ p_{\ell,m}^{-1} \left( \pi_1^{\ell m}(\mathbf{T}) - \pi_3^{\ell m}(\mathbf{T}) \widehat{\mu}_{\ell,m}^{(n)} \right) \right. \\
&\quad \left. - p_m^{-1} \left( \pi_2^m(\mathbf{T}) - \pi_4^m(\mathbf{T}) \widehat{\mu}_m^{(n)} \right) \right] \\
&\quad - p_m^{-1} \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left[ \left( \pi_5^m(\mathbf{T}) - \pi_4^m(\mathbf{T}) \left( \widehat{V}_m^{(n)} + \widehat{\mu}_m^{(n)} \otimes \widehat{\mu}_m^{(n)} \right) \right) \right. \\
&\quad \times \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \left. \right] \\
&\quad + p_m^{-1} \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left[ \left( \left( \pi_2^m(\mathbf{T}) - \pi_4^m(\mathbf{T}) \widehat{\mu}_m^{(n)} \right) \otimes \widehat{\mu}_m^{(n)} \right) \right. \\
&\quad \times \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \left. \right] \\
&\quad + p_m^{-1} \left( I_{\mathbb{R}^p} - \mathbb{V}_m Q_{K|m} \right) \left[ \left( \mu_m \otimes \left( \pi_2^m(\mathbf{T}) - \pi_4^m(\mathbf{T}) \widehat{\mu}_m^{(n)} \right) \right) \right. \\
&\quad \times \widehat{Q}_{K|m}^{(n)} \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \left. \right].
\end{aligned}$$

Ainsi

$$\begin{aligned}
n \widehat{\xi}_{K|m}^{(n)} &= \sum_{\ell=1}^q \left( \widehat{p}_{\ell|m}^{(n)} \right)^2 \left\| \sqrt{n} \left( I_{\mathbb{R}^p} - \widehat{V}_m^{(n)} \widehat{Q}_{K|m}^{(n)} \right) \left( \widehat{\mu}_{\ell,m}^{(n)} - \widehat{\mu}_m^{(n)} \right) \right\|_{\mathbb{R}^p}^2 \\
&= \sum_{\ell=1}^q \left( \left( \widehat{p}_{\ell|m}^{(n)} \right) \left\| \sqrt{n} \widehat{\Delta}_{\ell,K|m}^{(n)} \right\|_{\mathbb{R}^p} \right)^2 \\
&= \sum_{\ell=1}^q \left( \left( \widehat{p}_{\ell|m}^{(n)} \right) \left\| \sqrt{n} \widehat{\Delta}_{\ell,K|m}^{(n)} \right\|_{\mathbb{R}^p} - p_{\ell|m} \left\| \sqrt{n} \widehat{\Delta}_{\ell,K|m}^{(n)} \right\|_{\mathbb{R}^p} \right. \\
&\quad \left. + p_{\ell|m} \left\| \sqrt{n} \widehat{\Delta}_{\ell,K|m}^{(n)} \right\|_{\mathbb{R}^p} \right)^2 \\
&= \sum_{\ell=1}^q \left( \sqrt{n} \left( \widehat{p}_{\ell|m}^{(n)} - p_{\ell|m} \right) \left\| \widehat{\Delta}_{\ell,K|m}^{(n)} \right\|_{\mathbb{R}^p} + p_{\ell|m} \left\| \sqrt{n} \widehat{\Delta}_{\ell,K|m}^{(n)} \right\|_{\mathbb{R}^p} \right)^2.
\end{aligned}$$

A partir du Lemme 6 on obtient :

$$\begin{aligned}
n \widehat{\xi}_{K|m}^{(n)} &= \sum_{\ell=1}^q \left( p_m^{-1} \left( \pi_3^{\ell m}(\widehat{W}^{(n)}) - \pi_4^m(\widehat{W}^{(n)}) \widehat{p}_{\ell|m}^{(n)} \right) \left\| \widehat{\Delta}_{\ell,K|m}^{(n)} \right\|_{\mathbb{R}^p} \right. \\
&\quad \left. + p_{\ell|m} \left\| \sqrt{n} \widehat{\Delta}_{\ell,K|m}^{(n)} \right\|_{\mathbb{R}^p} \right)^2 \\
&= \sum_{\ell=1}^q \left( \widehat{\Phi}_{\ell,K|m}^{(n)}(\widehat{W}^{(n)}) + p_{\ell|m} \left\| \widehat{\Psi}_{\ell,K|m}^{(n)}(\widehat{W}^{(n)}) + \sqrt{n} \widehat{\Delta}_{\ell,K|m}^{(n)} \right\|_{\mathbb{R}^p} \right)^2,
\end{aligned}$$

où

$$\widehat{\Phi}_{\ell,K|m}^{(n)}(\mathbf{T}) = p_m^{-1} \left( \pi_3^{\ell m}(\mathbf{T}) - \pi_4^m(\mathbf{T}) \widehat{p}_{\ell|m}^{(n)} \right) \left\| \widehat{\Delta}_{\ell,K|m}^{(n)} \right\|_{\mathbb{R}^p}$$

## 4.2 Estimation et convergence du critère

Comme  $\hat{\mu}_{\ell,m}^{(n)}, \hat{\mu}_m^{(n)}, \hat{V}_m^{(n)}$  et  $\hat{p}_{\ell|m}^{(n)}$  convergent presque sûrement respectivement vers  $\mu_{\ell,m}, \mu_m, \mathbb{V}_m$  et  $p_{\ell|m}$  quand  $n \rightarrow +\infty$ , alors  $\hat{Q}_{K|m}^{(n)}$  et  $\hat{\Delta}_{\ell,K|m}^{(n)}$  convergent presque sûrement vers  $Q_{K|m}$  et  $\Delta_{\ell,K|m}$  quand  $n \rightarrow +\infty$ . Cela implique la convergence presque sûre des opérateurs aléatoires  $\hat{\Phi}_{\ell,K|m}^{(n)}$  (resp.  $\hat{\Psi}_{\ell,K|m}^{(n)}$ ) vers  $\Phi_{\ell,K|m}$  (resp. vers  $\Psi_{\ell,K|m}$ ) quand  $n \rightarrow +\infty$ .

□

Dans le Théorème 10, nous établissons la convergence du critère  $\hat{\xi}_K^{(n)'}$ .

**Théorème 10**  $\forall K \subset I$ , on a :

1.  $\hat{\xi}_K^{(n)'}$  converge presque sûrement vers  $\xi_K$  quand  $n \rightarrow +\infty$ .
- 2.

$$\begin{aligned} n\hat{\xi}_K^{(n)'} &= \sum_{m=1}^M \sqrt{n}\hat{\Lambda}_{K|m}^{(n)}(\widehat{W}^{(n)}) + \sum_{m=1}^M \sum_{\ell=1}^q (p_m \hat{\Phi}_{\ell,K|m}^{(n)}(\widehat{W}^{(n)} \\ &+ p_m p_{\ell|m} \|\hat{\Psi}_{\ell,K|m}^{(n)}(\widehat{W}^{(n)}) + \sqrt{n}\Delta_{\ell,K|m}\|_{\mathbb{R}^p})^2 \end{aligned}$$

avec  $\hat{\Lambda}_{K|m}^{(n)}(\widehat{W}^{(n)}) = \pi_4^m(\widehat{W}^{(n)}) (\hat{p}_m^{(n)} + p_m) \hat{\xi}_{K|m}^{(n)}$ , où  $(\hat{\Lambda}_{K|m}^{(n)})_{n \in \mathbb{N}}$  est une suite aléatoire d'opérateurs linéaires continus définis dans  $\mathcal{L}(F, \mathbb{R})$ , et qui converge presque sûrement vers  $\Lambda_{K|m} \in \mathcal{L}(F, \mathbb{R})$  telle que pour tout  $\mathbf{T} \in F$ ,  $\Lambda_{K|m}(\mathbf{T}) = 2p_m \pi_4^m(\mathbf{T}) \xi_{K|m}$ .

**Preuve:** On a :

1. On a  $\hat{\xi}_K^{(n)'} = \sum_{m=1}^M \hat{p}_m^{(n)2} \hat{\xi}_{K|m}^{(n)}$ ,  $\forall m \in \{1, \dots, M\}$ ,  $\hat{p}_m^{(n)}$  (resp.  $\hat{\xi}_{K|m}^{(n)}$ ) converge presque sûrement vers  $p_m$  (resp. vers  $\xi_{K|m}$ ) quand  $n \rightarrow +\infty$  (Théorème 9). Comme la fonction  $x \mapsto x^2$  est continue, il vient que  $\hat{p}_m^{(n)2}$  converge presque sûrement vers  $p_m^2$  quand  $n \rightarrow +\infty$ . Ainsi  $\hat{p}_m^{(n)2} \hat{\xi}_{K|m}^{(n)}$  converge presque sûrement vers  $p_m^2 \xi_{K|m}$  quand  $n \rightarrow +\infty$ . D'où  $\hat{\xi}_K^{(n)'}$  converge presque sûrement vers  $\xi_K'$  quand  $n \rightarrow +\infty$  comme somme finie d'estimateurs qui convergent presque sûrement.
2. on a  $n\hat{\xi}_K^{(n)'} = \sum_{m=1}^M n\hat{p}_m^{(n)2} \hat{\xi}_{K|m}^{(n)}$ ,  $\forall m \in \{1, \dots, M\}$ ,

$$\begin{aligned} \hat{p}_m^{(n)2} \hat{\xi}_{K|m}^{(n)} &= (\hat{p}_m^{(n)2} - p_m^2) \hat{\xi}_{K|m}^{(n)} + p_m^2 \hat{\xi}_{K|m}^{(n)} \\ &= (\hat{p}_m^{(n)} - p_m) (\hat{p}_m^{(n)} + p_m) \hat{\xi}_{K|m}^{(n)} + p_m^2 \hat{\xi}_{K|m}^{(n)}. \end{aligned}$$

Ainsi

$$\begin{aligned}
n\widehat{p}_m^{(n)2} \widehat{\xi}_{K|m}^{(n)} &= n \left( \widehat{p}_m^{(n)} - p_m \right) \left( \widehat{p}_m^{(n)} + p_m \right) \widehat{\xi}_{K|m}^{(n)} + p_m^2 n \widehat{\xi}_{K|m}^{(n)} \\
&= \sqrt{n} \pi_4^m(\widehat{W}^{(n)}) \left( \widehat{p}_m^{(n)} + p_m \right) \widehat{\xi}_{K|m}^{(n)} + p_m^2 n \widehat{\xi}_{K|m}^{(n)} \\
&= \sqrt{n} \widehat{\Lambda}_{K|m}^{(n)}(\widehat{W}^{(n)}) + \sum_{\ell=1}^q \left( p_m \widehat{\Phi}_{\ell,K|m}^{(n)}(\widehat{W}^{(n)}) \right. \\
&\quad \left. + p_m p_{\ell|m} \|\widehat{\Psi}_{\ell,K|m}^{(n)}(\widehat{W}^{(n)}) + \sqrt{n} \Delta_{\ell,K|m}\|_{\mathbb{R}^p} \right)^2.
\end{aligned}$$

D'où

$$\begin{aligned}
n\widehat{\xi}_K^{(n)'} &= \sum_{m=1}^M \sqrt{n} \widehat{\Lambda}_{K|m}^{(n)}(\widehat{W}^{(n)}) + \sum_{m=1}^M \sum_{\ell=1}^q \left( p_m \widehat{\Phi}_{\ell,K|m}^{(n)}(\widehat{W}^{(n)}) \right. \\
&\quad \left. + p_m p_{\ell|m} \|\widehat{\Psi}_{\ell,K|m}^{(n)}(\widehat{W}^{(n)}) + \sqrt{n} \Delta_{\ell,K|m}\|_{\mathbb{R}^p} \right)^2,
\end{aligned}$$

où  $\widehat{\Lambda}_{K|m}^{(n)}(\widehat{W}^{(n)}) = \pi_4^m(\widehat{W}^{(n)}) \left( \widehat{p}_m^{(n)} + p_m \right) \widehat{\xi}_{K|m}^{(n)}$ . On sait que  $\widehat{p}_m^{(n)}$  converge presque sûrement vers  $p_m$  quand  $n \rightarrow +\infty$ , alors  $\left( \widehat{p}_m^{(n)} + p_m \right)$  converge presque sûrement vers  $2p_m$  quand  $n \rightarrow +\infty$ . Comme  $\widehat{\xi}_{K|m}^{(n)}$  converge presque sûrement vers  $\xi_{K|m}$  quand  $n \rightarrow +\infty$ , alors  $\forall \mathbf{T} \in F$ ,  $\pi_4^m(\mathbf{T}) \left( \widehat{p}_m^{(n)} + p_m \right) \widehat{\xi}_{K|m}^{(n)}$  converge presque sûrement vers  $\Lambda_{K|m}(\mathbf{T}) = 2p_m \pi_4^m(\mathbf{T}) \xi_{K|m}$  quand  $n \rightarrow +\infty$ .

□

## 4.3 Sélection des variables

Dans la Section 4.1.2, nous avons montré que le sous-ensemble  $I_1$  est entièrement caractérisé par la permutation aléatoire  $\sigma$  et l'entier  $s$ . Dans cette section, nous estimons ces deux paramètres pour déduire l'estimation du sous-ensemble  $I_1$ , puis nous établirons leur propriété de convergence.

### 4.3.1 Estimation et convergence des paramètres $\sigma$ et $s$

La procédure d'estimation des paramètres  $\sigma$  et  $s$  reste identique à celle établie dans les Chapitre 2, Section 2.2. Ainsi, lorsque les paramètres  $\sigma$  et  $s$  ont été estimés comme indiqué, la sélection des variables est obtenue en prenant le sous-ensemble aléatoire  $\hat{I}_1^{(n)}$  comme estimateur de  $I_1$ , avec :

$$\hat{I}_1^{(n)} = \{\hat{\sigma}^{(n)}(i); 1 \leq i \leq \hat{s}^{(n)}\}.$$

Dans le théorème 11, nous établissons la convergence des estimateurs  $\hat{\sigma}^{(n)}$  et  $\hat{s}^{(n)}$ .

**Théorème 11** On a :

1.  $\lim_{n \rightarrow +\infty} \mathbb{P}(\hat{\sigma}^{(n)} = \sigma) = 1$ ;
2.  $\hat{s}^{(n)}$  converge en probabilité vers  $s$  quand  $n \rightarrow +\infty$ .

On déduit de ce Théorème que  $\lim_{n \rightarrow +\infty} \mathbb{P}\{\hat{I}_1^{(n)} = I_1\} = 1$ , ce qui établit la convergence en probabilité de notre méthode de sélection de variables en analyse discriminante mixte.

Remarquons de plus que le Théorème 11 est identique au Théorème 3.1 dans [Nkiet \(2012\)](#). Ainsi, pour la démonstration, il nous suffit de montrer le Lemme 9 :

**Lemme 9** Considérons les ensembles  $E$  et  $F_\ell$  tels que :

$$E = \{\ell \in \mathbb{N}^*; 1 \leq \ell \leq t, m_\ell \geq 2\}, \text{ ou } t \in I,$$

et

$$F_\ell := \left\{ \sum_{k=0}^{\ell-1} m_k + 1, \dots, \sum_{k=0}^{\ell} m_k - 1 \right\},$$

on suppose que  $m_0 = 0$ . Supposons que  $E \neq \emptyset$ , alors  $\forall \ell \in E$  et  $\forall i \in F_\ell$ , la suite  $n^\beta \left( \hat{\xi}_{K_{\sigma(i)}}^{(n)'} - \hat{\xi}_{K_{\sigma(i+1)}}^{(n)'} \right)$  converge en probabilité vers 0 quand  $n \rightarrow +\infty$ .

**Preuve:** En utilisant le point 2 du Théorème 10, on a :

## 4.3 Sélection des variables

$$\begin{aligned}
n^\beta \left( \widehat{\xi}_{K_{\sigma(i)}}^{(n)'} - \widehat{\xi}_{K_{\sigma(i+1)}}^{(n)'} \right) &= n^\beta \left[ \sum_{m=1}^M n^{-1/2} (\widehat{\Lambda}_{K_{\sigma(i)}|m}^{(n)}) (\widehat{W}^{(n)}) \right. \\
&+ \sum_{m=1}^M \sum_{\ell=1}^q n^{-1} \left( p_m \widehat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)} (\widehat{W}^{(n)}) \right. \\
&+ p_m p_{\ell|m} \|\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)} (\widehat{W}^{(n)}) + \sqrt{n} \Delta_{\ell, K_{\sigma(i)}|m}\|_{\mathbb{R}^p} \|^2 \\
&- \sum_{m=1}^M n^{-1/2} (\widehat{\Lambda}_{K_{\sigma(i+1)}|m}^{(n)}) (\widehat{W}^{(n)}) \\
&- \sum_{m=1}^M \sum_{\ell=1}^q n^{-1} \left( p_m \widehat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)} (\widehat{W}^{(n)}) \right. \\
&\left. \left. + p_m p_{\ell|m} \|\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)} (\widehat{W}^{(n)}) + \sqrt{n} \Delta_{\ell, K_{\sigma(i+1)}|m}\|_{\mathbb{R}^p} \|^2 \right) \right].
\end{aligned}$$

$$\begin{aligned}
n^\beta \left( \widehat{\xi}_{K_{\sigma(i)}}^{(n)'} - \widehat{\xi}_{K_{\sigma(i+1)}}^{(n)'} \right) &= \sum_{m=1}^M n^{\beta-1/2} \left( \widehat{\Lambda}_{K_{\sigma(i)}|m}^{(n)} - \widehat{\Lambda}_{K_{\sigma(i+1)}|m}^{(n)} \right) (\widehat{W}^{(n)}) \\
&+ n^{\beta-1} \sum_{m=1}^M \sum_{\ell=1}^q p_m^2 \left( \widehat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)} (\widehat{W}^{(n)})^2 - \widehat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)} (\widehat{W}^{(n)})^2 \right) \\
&+ \sum_{m=1}^M \sum_{\ell=1}^q p_m^2 p_{\ell|m}^2 n^{\beta-1} \left( \|\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)} (\widehat{W}^{(n)}) + \sqrt{n} \Delta_{\ell, K_{\sigma(i)}|m}\|_{\mathbb{R}^p} \|^2 \right. \\
&- \left. \|\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)} (\widehat{W}^{(n)}) + \sqrt{n} \Delta_{\ell, K_{\sigma(i+1)}|m}\|_{\mathbb{R}^p} \|^2 \right) \\
&+ \sum_{m=1}^M \sum_{\ell=1}^q 2p_m^2 p_{\ell|m} n^{\beta-1} \left( \widehat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)} (\widehat{W}^{(n)}) \|\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)} (\widehat{W}^{(n)}) \right. \\
&+ \sqrt{n} \Delta_{\ell, K_{\sigma(i)}|m}\|_{\mathbb{R}^p} \\
&- \left. \widehat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)} (\widehat{W}^{(n)}) \|\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)} (\widehat{W}^{(n)}) \right. \\
&\left. + \sqrt{n} \Delta_{\ell, K_{\sigma(i+1)}|m}\|_{\mathbb{R}^p} \right).
\end{aligned}$$

Posons :

$$\begin{aligned}
\widehat{A}_i^{(n)} &= \sum_{m=1}^M n^{\beta-1/2} \left( \widehat{\Lambda}_{K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)}) - \widehat{\Lambda}_{K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)}) \right) \\
\widehat{B}_i^{(n)} &= \sum_{m=1}^M \sum_{\ell=1}^q p_m^2 n^{\beta-1} \left( \widehat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)})^2 - \widehat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)})^2 \right) \\
\widehat{C}_i^{(n)} &= \sum_{m=1}^M \sum_{\ell=1}^q p_m^2 p_{\ell|m}^2 n^{\beta-1} \left( \|\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)}) + \sqrt{n} \Delta_{\ell, K_{\sigma(i)}|m}\|_{\mathbb{R}^p}^2 \right. \\
&\quad \left. - \|\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)}) + \sqrt{n} \Delta_{\ell, K_{\sigma(i+1)}|m}\|_{\mathbb{R}^p}^2 \right), \text{ et} \\
\widehat{D}_i^{(n)} &= \sum_{m=1}^M \sum_{\ell=1}^q 2p_m^2 p_{\ell|m} n^{\beta-1} \left( \widehat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)}) \|\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)}) + \sqrt{n} \Delta_{\ell, K_{\sigma(i)}|m}\|_{\mathbb{R}^p} \right. \\
&\quad \left. - \widehat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)}) \|\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)}) + \sqrt{n} \Delta_{\ell, K_{\sigma(i+1)}|m}\|_{\mathbb{R}^p} \right).
\end{aligned}$$

Soit  $\|\cdot\|_{\infty}$  la norme de la convergence uniforme usuelle, telle que pour tout opérateur  $T$ ,

$$\|T\|_{\infty} = \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|}, \text{ et } \|\cdot\|_F \text{ la norme de } F. \text{ On a } \forall m \in \{1, \dots, M\}$$

$$\begin{aligned}
\widetilde{A}_i^{(n)} &= n^{\beta-1/2} \left( \widehat{\Lambda}_{K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)}) - \widehat{\Lambda}_{K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)}) \right) \\
|\widetilde{A}_i^{(n)}| &\leq n^{\beta-1/2} \left( \|\widehat{\Lambda}_{K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)}) + \widehat{\Lambda}_{K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)})\| \right) \\
&\leq n^{\beta-1/2} \left( \|\widehat{\Lambda}_{K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)})\| + \|\widehat{\Lambda}_{K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)})\| \right) \\
&\leq n^{\beta-1/2} \left( \|\widehat{\Lambda}_{K_{\sigma(i)}|m}^{(n)}\|_{\infty} \|\widehat{W}^{(n)}\|_F + \|\widehat{\Lambda}_{K_{\sigma(i+1)}|m}^{(n)}\|_{\infty} \|\widehat{W}^{(n)}\|_F \right) \\
&\leq n^{\beta-1/2} \left( \|\widehat{\Lambda}_{K_{\sigma(i)}|m}^{(n)}\|_{\infty} + \|\widehat{\Lambda}_{K_{\sigma(i+1)}|m}^{(n)}\|_{\infty} \right) \|\widehat{W}^{(n)}\|_F
\end{aligned}$$

$\forall m \in \{1, \dots, M\}$ ,  $\widehat{\Lambda}_{K_{\sigma(i)}|m}^{(n)}$  (resp.  $\widehat{\Lambda}_{K_{\sigma(i+1)}|m}^{(n)}$ ) converge presque sûrement vers  $\Lambda_{K_{\sigma(i)}|m}$  (resp. vers  $\Lambda_{K_{\sigma(i+1)}|m}$ ) quand  $n \rightarrow +\infty$  (Théorème 9),  $\widehat{W}^{(n)}$  converge en loi vers une variable aléatoire normale centrée  $W \in F$  et de matrice de covariance  $\Gamma$  quand  $n \rightarrow +\infty$  (application du Théorème Limite centrale). Comme la fonction qui a  $x \mapsto \|x\|_{\infty}$  est continue et que  $\beta - 1 < 0$ , en utilisant la dernière inégalité (relation  $|\widetilde{A}_i^{(n)}|$ ), on en déduit que la suite de variables aléatoires  $\widetilde{A}_i^{(n)}$  converge en probabilité vers 0 quand  $n \rightarrow +\infty$ . D'où la suite  $\widehat{A}_i^{(n)}$  converge en probabilité vers 0 quand  $n \rightarrow +\infty$  comme somme finie de suites de variables aléatoires qui convergent en probabilité vers 0 quand  $n \rightarrow +\infty$ .

## 4.3 Sélection des variables

$\forall m \in \{1, \dots, M\}, \forall \ell \in \{1, \dots, q\}$  on a :

$$\begin{aligned} \tilde{B}_i^{(n)} &= n^{\beta-1} \left( \hat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)} (\widehat{W}^{(n)})^2 - \hat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)} (\widehat{W}^{(n)})^2 \right) \\ |\tilde{B}_i^{(n)}| &\leq n^{\beta-1} \left( \|\hat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)} (\widehat{W}^{(n)})\|^2 + \|\hat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)} (\widehat{W}^{(n)})\|^2 \right) \\ &\leq n^{\beta-1} \left( \|\hat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)}\|_\infty^2 \|\widehat{W}^{(n)}\|_F^2 + \|\hat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}\|_\infty^2 \|\widehat{W}^{(n)}\|_F^2 \right) \\ &\leq n^{\beta-1} \left( \|\hat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)}\|_\infty^2 + \|\hat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}\|_\infty^2 \right) \|\widehat{W}^{(n)}\|_F^2 \end{aligned}$$

$\forall m \in \{1, \dots, M\}$  et  $\forall \ell \in \{1, \dots, q\}$ ,  $\hat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)}$  (resp.  $\hat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}$ ) converge presque sûrement vers  $\Phi_{\ell, K_{\sigma(i)}|m}$  (resp. vers  $\Phi_{\ell, K_{\sigma(i+1)}|m}$ ) quand  $n \rightarrow +\infty$  (Théorème 8).  $\widehat{W}^{(n)}$  converge en loi vers la variable aléatoire  $W$  quand  $n \rightarrow +\infty$  (Théorème 10). Comme la fonction qui a  $x \mapsto \|x\|_\infty^2$  est continue et comme  $\beta - 1 < 0$ . En utilisant la dernière inégalité (relation  $|\tilde{B}_i^{(n)}|$ ), on en déduit que la suite de variables aléatoires  $\tilde{B}_i^{(n)}$  converge en probabilité vers 0 quand  $n \rightarrow +\infty$ . D'où la suite  $\tilde{B}_i^{(n)}$  converge en probabilité vers 0 quand  $n \rightarrow +\infty$  comme somme finie de suites de variables aléatoires qui convergent en probabilité vers 0 quand  $n \rightarrow +\infty$ .

Pour tout  $i \in F_\ell$ , on a

$$\|\Delta_{\ell, K_{\sigma(i)}|m}\| = \xi'_{K_{\sigma(i)}} = \xi'_{K_{\sigma(i+1)}} = \|\Delta_{\ell, K_{\sigma(i+1)}|m}\|.$$

$\forall m \in \{1, \dots, M\}, \forall \ell \in \{1, \dots, q\}$ , on a

$$\begin{aligned} &\|\hat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)} (\widehat{W}^{(n)}) + \sqrt{n} \Delta_{\ell, K_{\sigma(i)}|m}\|^2 - \|\hat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)} (\widehat{W}^{(n)}) + \sqrt{n} \Delta_{\ell, K_{\sigma(i+1)}|m}\|^2 \\ &= \|\hat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)} (\widehat{W}^{(n)})\|^2 - \|\hat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)} (\widehat{W}^{(n)})\|^2 + 2\sqrt{n} \langle \hat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)} (\widehat{W}^{(n)}), \Delta_{\ell, K_{\sigma(i)}|m} \rangle \\ &- 2\sqrt{n} \langle \hat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)} (\widehat{W}^{(n)}), \Delta_{\ell, K_{\sigma(i+1)}|m} \rangle. \end{aligned}$$

Ainsi,

$$\begin{aligned} \tilde{C}_i^{(n)} &= n^{\beta-1} \left( \|\hat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)} (\widehat{W}^{(n)})\|^2 - \|\hat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)} (\widehat{W}^{(n)})\|^2 \right) \\ &+ 2\sqrt{n} \langle \hat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)} (\widehat{W}^{(n)}), \Delta_{\ell, K_{\sigma(i)}|m} \rangle \\ &- 2\sqrt{n} \langle \hat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)} (\widehat{W}^{(n)}), \Delta_{\ell, K_{\sigma(i+1)}|m} \rangle \end{aligned}$$

## 4.3 Sélection des variables

$$\begin{aligned}
\tilde{C}_i^{(n)} &\leq n^{\beta-1/2} \left( n^{-1/2} (\|\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)})\|^2 + \|\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)})\|^2) \right) \\
&+ 2 \left\langle \widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)}), \Delta_{\ell, K_{\sigma(i)}|m} \right\rangle \\
&- 2 \left\langle \widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)}), \Delta_{\ell, K_{\sigma(i+1)}|m} \right\rangle.
\end{aligned}$$

Ainsi

$$\begin{aligned}
|\widehat{C}_i^{(n)}| &\leq n^{\alpha-1/2} \|\widehat{W}^{(n)}\|_F \left( n^{-1/2} (\|\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)}\|_\infty^2 + \|\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}\|_\infty^2) \|\widehat{W}^{(n)}\|_F \right) \\
&+ 2 \|\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)}\|_\infty \|\Delta_{\ell, K_{\sigma(i)}|m}\| + 2 \|\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}\|_\infty \|\Delta_{\ell, K_{\sigma(i+1)}|m}\|.
\end{aligned}$$

$\forall m \in \{1, \dots, M\}$  et  $\forall \ell \in \{1, \dots, q\}$ ,  $\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)}$  (resp.  $\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}$ ) converge presque sûrement vers  $\Psi_{\ell, K_{\sigma(i)}|m}$  (resp. vers  $\Psi_{\ell, K_{\sigma(i+1)}|m}$ ) quand  $n \rightarrow +\infty$ .  $\widehat{W}^{(n)}$  converge en loi vers la variable aléatoire  $W$  quand  $n \rightarrow +\infty$ . De plus, les normes  $\|\cdot\|_\infty$  et  $\|\cdot\|_F$  sont continues, il vient que  $n^{-1/2} \|\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)}\|_\infty \|\widehat{W}^{(n)}\|_F$  (resp.  $n^{-1/2} \|\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}\|_\infty \|\widehat{W}^{(n)}\|_F$ ) converge en probabilité vers 0 quand  $n \rightarrow +\infty$ . Comme  $\beta - 1/2 < 0$ , on en déduit que la suite de variables aléatoires  $\tilde{C}_i^{(n)}$  converge en probabilité vers 0 quand  $n \rightarrow +\infty$ . D'où la suite  $\widehat{C}_i^{(n)}$  converge en probabilité vers 0 comme somme finie de suites de variables aléatoires qui convergent en probabilité vers 0 quand  $n \rightarrow +\infty$ .

$\forall m \in \{1, \dots, M\}, \forall \ell \in \{1, \dots, q\}$  on a :

$$\begin{aligned}
|\widetilde{D}_i^{(n)}| &= n^{\beta-1} \left| \left( \widehat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)}) \|\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)}) + \sqrt{n} \Delta_{\ell, K_{\sigma(i)}|m}\|_{\mathbb{R}^p} \right) \right. \\
&- \left. \widehat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)}) \|\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)}) + \sqrt{n} \Delta_{\ell, K_{\sigma(i+1)}|m}\|_{\mathbb{R}^p} \right|
\end{aligned}$$

$$\begin{aligned}
|\widetilde{D}_i^{(n)}| &\leq n^{\beta-1} \left( \left| \widehat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)}) \right| \|\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)}) + \sqrt{n} \Delta_{\ell, K_{\sigma(i)}|m}\|_{\mathbb{R}^p} \right. \\
&+ \left. \left| \widehat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)}) \right| \|\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)}) + \sqrt{n} \Delta_{\ell, K_{\sigma(i+1)}|m}\|_{\mathbb{R}^p} \right)
\end{aligned}$$



## 4.3 Sélection des variables

$$|\widetilde{D}_i^{(n)}| \leq n^{\beta-1} \left( \left| \widehat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)}) \right| \left( \|\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)})\|_{\mathbb{R}^p} + \sqrt{n} \|\Delta_{\ell, K_{\sigma(i)}|m}\|_{\mathbb{R}^p} \right) \right. \\ \left. + \left| \widehat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)}) \right| \left( \|\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)})\|_{\mathbb{R}^p} + \sqrt{n} \|\Delta_{\ell, K_{\sigma(i+1)}|m}\|_{\mathbb{R}^p} \right) \right)$$

$$|\widetilde{D}_i^{(n)}| \leq n^{\beta-1} \left( \left| \widehat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)}) \right| \left( \|\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)}\|_{\infty} \|\widehat{W}^{(n)}\|_F + \sqrt{n} \|\Delta_{\ell, K_{\sigma(i)}|m}\|_{\mathbb{R}^p} \right) \right. \\ \left. + \left| \widehat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)}) \right| \left( \|\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}\|_{\infty} \|\widehat{W}^{(n)}\|_F + \sqrt{n} \|\Delta_{\ell, K_{\sigma(i+1)}|m}\|_{\mathbb{R}^p} \right) \right)$$

$$|\widetilde{D}_i^{(n)}| \leq n^{\beta-1/2} \left( \left| \widehat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)}(\widehat{W}^{(n)}) \right| \left( n^{-1/2} \|\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)}\|_{\infty} \|\widehat{W}^{(n)}\|_F + \|\Delta_{\ell, K_{\sigma(i)}|m}\|_{\mathbb{R}^p} \right) \right. \\ \left. + \left| \widehat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}(\widehat{W}^{(n)}) \right| \left( n^{-1/2} \|\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}\|_{\infty} \|\widehat{W}^{(n)}\|_F + \|\Delta_{\ell, K_{\sigma(i+1)}|m}\|_{\mathbb{R}^p} \right) \right)$$

$$|\widetilde{D}_i^{(n)}| \leq n^{\beta-1/2} \left( \left\| \widehat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)} \right\|_{\infty} \|\widehat{W}^{(n)}\|_F \left( n^{-1/2} \|\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)}\|_{\infty} \|\widehat{W}^{(n)}\|_F + \|\Delta_{\ell, K_{\sigma(i)}|m}\|_{\mathbb{R}^p} \right) \right. \\ \left. + \left\| \widehat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)} \right\|_{\infty} \|\widehat{W}^{(n)}\|_F \left( n^{-1/2} \|\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}\|_{\infty} \|\widehat{W}^{(n)}\|_F + \|\Delta_{\ell, K_{\sigma(i+1)}|m}\|_{\mathbb{R}^p} \right) \right)$$

$$|\widetilde{D}_i^{(n)}| \leq n^{\beta-1/2} \left( \left\| \widehat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)} \right\|_{\infty} \left( n^{-1/2} \|\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)}\|_{\infty} \|\widehat{W}^{(n)}\|_F + \|\Delta_{\ell, K_{\sigma(i)}|m}\|_{\mathbb{R}^p} \right) \right. \\ \left. + \left\| \widehat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)} \right\|_{\infty} \left( n^{-1/2} \|\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}\|_{\infty} \|\widehat{W}^{(n)}\|_F \right. \right. \\ \left. \left. + \|\Delta_{\ell, K_{\sigma(i+1)}|m}\|_{\mathbb{R}^p} \right) \right) \|\widehat{W}^{(n)}\|_F$$

$\forall m \in \{1, \dots, M\}$  et  $\forall \ell \in \{1, \dots, q\}$ ,  $\widehat{\Phi}_{\ell, K_{\sigma(i)}|m}^{(n)}$ ,  $\widehat{\Phi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}$  et  $\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)}$ ,  $\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}$  convergent presque sûrement vers  $\Phi_{\ell, K_{\sigma(i)}|m}$ ,  $\Phi_{\ell, K_{\sigma(i+1)}|m}$  et  $\Psi_{\ell, K_{\sigma(i)}|m}$ ,  $\Psi_{\ell, K_{\sigma(i+1)}|m}$  quand  $n \rightarrow +\infty$ .  $\widehat{W}^{(n)}$  converge en loi vers une variable aléatoire normale centrée dans  $F$  quand  $n \rightarrow +\infty$ . Comme les opérateurs  $\Phi_{\ell, K_{\sigma(i)}|m}$  et  $\Phi_{\ell, K_{\sigma(i+1)}|m}$  sont continus, il vient que  $\Phi_{\ell, K_{\sigma(i)}|m}(\widehat{W}^{(n)})$  et  $\Phi_{\ell, K_{\sigma(i+1)}|m}(\widehat{W}^{(n)})$  convergent en loi vers  $\Phi_{\ell, K_{\sigma(i)}|m}(W)$  et  $\Phi_{\ell, K_{\sigma(i+1)}|m}(W)$  quand  $n \rightarrow +\infty$ . De plus les normes  $\|\cdot\|_{\infty}$  et  $\|\cdot\|_F$  sont continues, il vient que  $n^{-1/2} \|\widehat{\Psi}_{\ell, K_{\sigma(i)}|m}^{(n)}\|_{\infty} \|\widehat{W}^{(n)}\|_F$  et  $n^{-1/2} \|\widehat{\Psi}_{\ell, K_{\sigma(i+1)}|m}^{(n)}\|_{\infty} \|\widehat{W}^{(n)}\|_F$  convergent en probabilité vers 0 quand  $n \rightarrow +\infty$ . Comme  $\beta - 1/2 < 0$ , on en déduit que la suite de variables aléatoires  $\widetilde{D}_i^{(n)}$  converge en probabilité vers 0 quand  $n \rightarrow +\infty$ . D'où la suite  $\widehat{D}_i^{(n)}$  converge en probabilité vers 0 comme

somme finie de suites de variables aléatoires qui convergent en probabilité vers 0 quand  $n \rightarrow +\infty$ .

Ainsi, on en déduit que  $\widehat{A}_i^{(n)}$ ,  $\widehat{B}_i^{(n)}$ ,  $\widehat{C}_i^{(n)}$  et  $\widehat{D}_i^{(n)}$  convergent en probabilité vers 0 quand  $n \rightarrow +\infty$ .  $\square$

## 4.4 Etude numérique

Dans cette partie, nous proposons une étude numérique dans le but d'étudier et de prouver l'efficacité de notre méthode. Nous étudions principalement l'influence des fonctions de pénalité  $f_n, g_n$  et des paramètres  $\alpha$  et  $\beta$ . Nous effectuons une comparaison avec la méthode de *Mahat et al.* (2007) et nous donnons un exemple avec des données réelles.

### 4.4.1 Simulations des données

Soient  $X$  un vecteur aléatoire de  $\mathbb{R}^5$  ( $p = 5$ ) et  $U$  une variable aléatoire contenant 8 cellules ( $M = 2^3$ ), c'est-à-dire que  $Y$  est un vecteur aléatoire contenant trois variables binaires ( $d = 3$ ).

**Simulation 1 :** Nous nous plaçons dans le cas de deux groupes. La vecteur aléatoire  $X$  est générée comme une variable normale multivariée dans  $\mathbb{R}^5$ , de moyenne  $\mu$  et de matrice de covariance  $\Gamma$ , tels que les éléments du premier et second groupes suivent respectivement une distribution normale  $\mathbf{N}(\mu_1, \Gamma)$  et  $\mathbf{N}(\mu_2, \Gamma)$ , avec  $\Gamma = \frac{1}{2}(I_5 + J)$ , où  $I_5$  est la matrice identité de taille  $5 \times 5$  et  $J$  la matrice de taille  $5 \times 5$  dont tous les éléments valent 1. Nous prenons  $\mu_1 = (0, 0, 0, 0, 0)^T$  et  $\mu_2 = \frac{1}{4}(1, 0, 2, 0, 3)^T$ . La variable aléatoire  $U$  est générée à partir d'une distribution notée  $\Pi = (q_1, q_2, \dots, q_8) \in \{1, \dots, 8\}$  avec  $\sum_{k=1}^8 q_k = 1$ . Cela est équivalent à générer la variable aléatoire  $Y$  comme un vecteur aléatoire contenant trois variables binaires. Pour cela, deux modèles sont considérés :

**Modèle 1 :**  $q_1 = q_2 = \dots = q_8 = 1/8$  : c'est la distribution uniforme ;

**Modèle 2 :**  $q_1 = q_3 = q_5 = 0.001$ ,  $q_2 = q_4 = q_6 = q_7 = q_8 = 0.1994$ . Le Modèle 2 correspond au cas où le nombre d'individu dans les cellules 1, 3 et 5 est très faible.

**Simulation 2 :** Nous nous plaçons dans le cas de trois groupes. Les trois groupes suivent respectivement les lois  $\mathbf{N}(\mu_1, \Gamma)$ ,  $\mathbf{N}(\mu_2, \Gamma)$  et  $(U[0, 1])^5$ , avec  $\mu_1 = (0, 0, 0, 0, 0)^T$ ,  $\mu_2 = (\frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1, \frac{5}{4})^T$  et  $\mu_3 = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})^T$ . Le vecteur  $X$  et la variable  $U$  sont simulés comme dans le Modèle 1.

Nous simulons ensuite deux échantillons (validation croisée Chapitre 1, Section 2.3), à savoir un échantillon d'apprentissage et un échantillon test.

**Echantillon 1 :** L'échantillon 1 est considéré comme échantillon d'apprentissage. Il est utilisé pour sélectionner les variables. Pour la simulation 1, nous prenons  $n = 100, 200, 300, 400, 500$  et  $n_1 = n_2 = n/2$ . Pour la simulation 2, nous prenons  $n = 6000, 7500, 9000, 10500$  et  $n_1 = n_2 = n_3 = n/3$ . Nous étudions aussi l'influence des fonctions de pénalité  $f_n, g_n$  et des paramètres  $\alpha, \beta$  et  $\lambda$  ( $\lambda$  est le paramètre de lissage de la méthode de [Mahat et al. \(2007\)](#)) puis, nous regardons leur impact sur notre méthode de sélection.

**Echantillon 2 :** L'échantillon 2 est considéré comme échantillon test. Il est utilisé pour calculer le taux de bon classement. C'est le rapport entre le nombre de variables bien classées sur la taille de l'échantillon. Le taux de bon classement est obtenu grâce à la fonction de Fisher sur la base des variables sélectionnées. Dans le cas de deux groupes, la règle de décision est donnée grâce à la fonction Fisher [Mahat et al. \(2007\)](#).  
 $\forall (x, y) \in (X, Y) :$

$$\mathbf{D}_{\ell, m}(x, y) = (\mu_{1, m} - \mu_{2, m})^T V^{-1} \left\{ x - \frac{1}{2}(\mu_{1, m} - \mu_{2, m}) \right\} \geq \log \left( \frac{p_{2, m}}{p_{1, m}} \right) + \log(a). \quad (4.10)$$

Dans le cas de plus de deux groupes, on utilise la fonction de Fisher généralisée pour plus de deux groupes ([Leon et al. \(2011\)](#)) :

$$\mathbf{D}_{\ell, m}(x, y) = (\mu_{\ell, m})^T V^{-1} x - \frac{1}{2}(\mu_{\ell, m})^T V^{-1} \mu_{\ell, m} + \log p_{\ell, m} + \log p_{\ell}. \quad (4.11)$$

Ainsi, on est dans le groupe  $\ell$  si  $\forall (\ell, k) \in \{1, \dots, p\}^2$ , et  $\forall m \in \{1, \dots, M\}$  on a :

$$\mathbf{D}_{\ell, m} - \mathbf{D}_{k, m} \geq 0, \quad \ell \neq k.$$

### 4.4.2 Influence des fonctions de pénalité

On pose :

$$f_n(i) = n^{-0.45}f(i), \quad g_n(i) = n^{-0.6}g(i) \text{ avec } f(i) = 1/g(i).$$

Nous considérons les fonctions  $g_n$  ( $n = 1, \dots, 13$ ) comme au Chapitre 2 Section 2.3.2 :  $g_1(x) = x$ ,  $g_2(x) = x^{0.1}$ ,  $g_3(x) = x^{0.5}$ ,  $g_4(x) = x^{0.9}$ ,  $g_5(x) = x^{10}$ ,  $g_6(x) = \ln(x)$ ,  $g_7(x) = (\ln(x))^{0.1}$ ,  $g_8(x) = (\ln(x))^{0.5}$ ,  $g_9(x) = (\ln(x))^{0.9}$ ,  $g_{10}(x) = x \ln(x)$ ,  $g_{11}(x) = (x \ln(x))^{0.1}$ ,  $g_{12}(x) = (x \ln(x))^{0.5}$ ,  $g_{13}(x) = (x \ln(x))^{0.9}$  (**Kundu et Murali (1996)**). Pour chaque choix de la fonction de pénalité  $g_n$ , nous calculons le taux de bon classement lié à notre méthode en utilisant les estimateurs empiriques et les estimateurs non paramétriques de **Mahat et al. (2007)**.

Les résultats sont reportés dans les Tableaux 4.1 - 4.3. Nous voyons que quelles que soit les fonctions de pénalité  $f_n$ ,  $g_n$  et quelque soit la taille de l'échantillon  $n = 100, 300, 500$ , le taux de bon classement des deux estimateurs reste quasiment identique. Alors ni les fonctions de pénalité, ni les estimateurs utilisés n'influencent le taux de bon classement de notre méthode de sélection de variables.

Par contre, les résultats des Tableaux 4.1 - 4.3 ne montrent pas réellement la différence entre les estimateurs empiriques et les estimateurs non paramétriques lorsque le nombre d'éléments dans les cellules sont nuls ou négligeables. Pour le vérifier, nous considérons le Modèle 2. Nous supposons que le nombre d'éléments dans les cellules 1, 3 et 5 est très faible ( $p_1 = p_3 = p_5 = 0.001$ ). Puis, la probabilité dans les cellules 2, 4, 6, 7 et 8 est de 0.1994. Nous voyons dans le Tableau 4.6 que les taux de bon classement basé sur les estimateurs empiriques sont non définis (NA= Not Available) pour tout  $n = 100, 300, 500$ . Cela est certainement dû à la présence d'éléments en très faible quantité (ou pas) dans les cellules 1, 3 et 5. Les résultats sont différents avec estimateurs non paramétriques. On observe des taux de bon classement de 0.54000, 0.56670, 0.52400 respectivement pour  $n = 100, 300, 500$ . Ce qui montre que cette estimateur est plus robuste que l'estimateur empirique. Pour cette raison, nous utilisons dans la suite les estimateurs non paramétriques.

### 4.4.3 Influence des paramètres $\alpha$ et $\beta$

Nous étudions ici l'influence des paramètres  $\alpha$  et  $\beta$  sur notre méthode de sélection. Nous prenons :

$$f_n(i) = 1/n^\alpha (\ln(i))^{0.1}, \quad g_n(i) = (\ln(i))^{0.1} / n^\beta, \quad \text{où } f(i) = 1/(\ln(i))^{0.1},$$

avec  $\alpha \in ]0, 1/2[$ , et  $\beta \in ]0, 1[$ . Les résultats sont reportés sur les graphes 4.1 - 4.3. On observe une variation significative du taux de bon classement en fonction des paramètres  $\alpha$  et  $\beta$ , pour  $n = 100, 300, 500$ . Pour différentes valeurs de  $\alpha$  fixées, le taux de bon classement n'est pas constant quand  $\beta$  varie dans  $]0, 1[$ . Ainsi un choix approprié de  $\alpha$  et  $\beta$  devient nécessaire, c'est-à-dire déterminer le couple  $(\alpha, \beta) \in ]0, 1/2[ \times ]0, 1[$  qui maximise le taux de bon classement. Une approche via la méthode de validation croisée est proposée dans la section suivante.

### 4.4.4 Choix optimal des paramètres $\alpha$ et $\beta$ : Algorithme

Nous savons maintenant que les paramètres  $\alpha$  et  $\beta$  influencent notre méthode de sélection. Nous adoptons une méthode afin de déterminer de façon automatique les paramètres  $(\alpha, \beta)$  qui maximise le taux de bon classement. Nous utilisons pour cela la méthode de validation croisée. Pour tout  $k \in \{1, \dots, n\}$ , on retire la ligne  $k$  dans les observations de  $X$  et  $Y$  de l'échantillon d'apprentissage. Nous appliquons ensuite notre méthode de sélection de variable sur l'échantillon restant pour une valeur de  $\alpha$  fixée, on fait varier  $\beta$  dans  $]0, 1[$ . L'observation qui a été retirée est attribuée à un groupe noté  $\tilde{g}_{\alpha, \beta}(k)$  dans  $\{1, \dots, q\}$ , en utilisant les règles de classification de la Section 4.4.1 avec les variables sélectionnées à l'étapes précédente. Ensuite, nous considérons :

$$CV(\alpha, \beta) = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{Z_k = \tilde{g}_{\alpha, \beta}(k)\}},$$

et nous prenons pour valeurs optimal de  $(\alpha, \beta)$  le couple  $(\alpha_{opt}, \beta_{opt})$  définie par :

$$(\alpha_{opt}, \beta_{opt}) = \underset{(\alpha, \beta) \in ]0, 1/2[ \times ]0, 1[}{\operatorname{argmax}} CV(\alpha, \beta).$$

**Algorithm 1** Calcul du taux de bon classement avec les données simulées

Simuler l'échantillon d'apprentissage  $\mathcal{S}_1$  comme indiqué dans la Section 4.4.1

Discrétiser minutieusement le rectangle  $R = ]0, 1/2[ \times ]0, 1[$  de manière à obtenir une grille de points  $G$

**for all**  $(\alpha, \beta) \in G$  **do**

**for**  $k = 1, \dots, n$  **do**

Retirer la  $k^{\text{ème}}$  observation de l'échantillon d'apprentissage, l'ensemble de données restantes est noté  $\mathcal{S}_1^{(-k)}$

Appliquer la méthode de sélection des variables sur  $\mathcal{S}_1^{(-k)}$  afin d'obtenir le sous-ensemble  $\hat{I}_1^{(-k)}$

Attribuer l'observation retirée à l'un des groupes en utilisant la règle (4.10) ou (4.11), les variables sélectionnées sont rangées dans  $\hat{I}_1^{(-k)}$ , tout ceci en utilisant l'échantillon  $\mathcal{S}_1^{(-k)}$

Notons par  $\tilde{g}_{\alpha, \beta}(k)$   $k \in \{1, \dots, q\}$  le groupe correspondant à l'observation retirée à l'étape précédente. On note  $\epsilon_k = 1$  if  $\tilde{g}_{\alpha, \beta}(k)$  est le véritable groupe de cette observation, et  $\epsilon_k = 0$  sinon

**end for**

$$CV(\alpha, \beta) = n^{-1} \sum_{k=1}^n \epsilon_k$$

**end for**

On choisit  $(\hat{\alpha}, \hat{\beta})$  qui maximisent  $CV(\alpha, \beta)$  dans  $G$

Appliquer la méthode de sélections de variable sur l'ensemble de l'échantillon d'apprentissage  $\mathcal{S}_1$  avec  $(\hat{\alpha}, \hat{\beta})$  afin d'obtenir le sous-ensemble  $\hat{I}_1$  des variables adéquates à la sélection

Simuler un échantillon test  $\mathcal{S}_2$  comme indiqué dans la Section 4.4.1

Calculer le taux de bon classement sur  $\mathcal{S}_2$  en utilisant les variables sélectionnées dans  $\hat{I}_1$

#### 4.4.5 Comparaison avec la méthode Mahat et al. (2007)

La comparaison se fait dans le cas de deux groupes. Les résultats sont reportés dans les Tableaux 4.4 et 4.5. Dans le cas de deux groupes, c'est-à-dire le modèle 1, notre méthode de sélection et celle de *Mahat et al.* (2007) donnent quasiment les mêmes résultats, c'est-à-dire même taux de bon classement pour  $n = 100, 200, 300, 400, 500$  (Tableau 4.4). Dans le cas de trois groupes, c'est-à-dire le modèle 2, nous donnons uniquement le taux de bon classement de notre méthode. Nous ne la comparons avec aucune autre méthode car notre approche est la seule pour le moment (sur la base de la littérature effectuée) à s'étendre à plus de deux groupes. Les résultats sont reportés dans le Tableau 4.5. Le taux de bon classement atteint 50% à partir d'une taille d'échantillon  $n_1 = n_2 = n_3 = 2000$ , puis elle augmente avec la taille de  $n$ .

Ainsi, dans le cas de deux groupes, notre méthode de sélection des variables et celle de *Mahat et al.* (2007) sont équivalentes pour  $n = 100, 200, 300, 400, 500$ . Dans le cas de trois groupes, le taux de bon classement de notre méthode donne de bonnes performances pour des grandes tailles d'échantillon. On atteint 50% de bon classement lorsque la taille de l'échantillon dans chaque groupe est de  $n_1 = n_2 = n_3 = 2000$ , puis doucement, elle augmente avec la taille de l'échantillon  $n$ .

#### 4.4.6 Exemple sur des données réelles

Pour appliquer notre méthode aux données réelles, nous utilisons des données de demande de carte de crédit australienne contenues dans les bases de données de l'Université de Californie à Irvine (voir *Lichman M* (2013)). Tous les noms ont été changés en symboles sans signification pour protéger la confidentialité des demandeurs. Cet ensemble de données est intéressant car il contient un bon mélange de variables continues et binaires.

Les données sont composées de 690 demandeurs MasterCard (Carte de Crédit) subdivisés en deux groupes : le premier groupe contient 307 candidats qui sont classés comme positifs et le deuxième groupe contient 383 candidats classés comme négatifs. Nous considérons dix variables : quatre variables binaires et six variables continues et une onzième variable pour caractériser les groupes (deux groupes). Sur ces données, nous avons appliqué notre méthode et celle de *Mahat et al.* (2007) afin de faire des comparaisons. Nous

calculons le taux de bon classement moyen et le nombre de variables sélectionnées (NV) à partir de la méthode 10-fold cross-validation. Les résultats sont présentés dans le Tableau 4.7. Nous voyons que les deux méthodes donnent des résultats semblables en ce qui concerne le taux de bon classement respectivement 0.87800 pour la méthode de *Mahat et al. (2007)* et 0.87900 pour notre méthode, tandis que notre méthode sélectionne moins de variables (4 variables sélectionnées) que celle de *Mahat et al. (2007)* (6 variables sélectionnées).

### Informations sur les variables

---

$X_1$	: {0, 1}. Variables binaire
$X_2$	: Variable continue
$X_3$	: Variable continue
$X_4$	: Variable continue
$X_5$	: {0, 1}. Variables binaire
$X_6$	: {0, 1}. Variables binaire
$X_7$	: Variable continue
$X_8$	: {0, 1}. Variables binaire
$X_9$	: Variable continue
$X_{10}$	: Variable continue
$X_{11}$	: Variable qui caractérise les groupes (+,-) (deux groupes)

---

## 4.5 Conclusion

Dans ce chapitre, nous avons généralisé la méthode de sélection des variables de *Nkiet (2012)*, dans le cas des variables aléatoires mixtes. Plus précisément, nous avons prolongé dans le cadre de l'analyse discriminante des variables aléatoires mixtes, le critère de sélection développé dans le Chapitre 2. Notre approche repose sur un critère (Equation (4.7)). Il est utilisé pour ranger les variables dans l'ordre décroissant, de sorte à ramener le problème de sélection à un problème d'estimation d'une permutation aléatoire et d'un paramètre appelé dimension. Cette nouvelle approche a deux avantages : Elle est utilisée lorsque le nombre de groupes est supérieur à deux et deuxièmement, il suffit que le vec-



teur aléatoire constitué des variables continues ait un moment d'ordre quatre fini. Cela suppose que nous ne formulons aucune hypothèse sur la distribution du vecteur aléatoire contenant les variables continues. Nous avons considéré deux types d'estimateurs, à savoir les estimateurs empiriques et des estimateurs non paramétriques (*Mahat et al. (2007)*). Le second estimateur est essentiellement utilisé pour résoudre le problème de cellules vides, problème couramment rencontré en analyse discriminante mixte. Pour chaque type d'estimateur, nous avons établi un estimateur du critère de sélection. Avec les estimateurs empiriques, nous avons établi les propriétés de convergence et la convergence en probabilité de la méthode. Nous avons dans ensuite proposé une étude numérique, dans le but d'étudier principalement l'influence des paramètres de contrôle  $f_n$ ,  $g_n$ ,  $\alpha$ ,  $\beta$ , puis la comparer à celle de *Mahat et al. (2007)* dans le cas de deux groupes. Cette étude relève qu'il n'y a pas une influence des fonctions de pénalité  $f_n$  et  $g_n$  sur notre méthode de sélection. En effet, pour chaque fonction  $f_n$  et  $g_n$  choisies (*Kundu et Murali (1996)*), il n'y a pas de variation du taux de bon classement lié à notre méthode et cela, quel que soit le type d'estimateur considéré (Tableaux 4.1 - 4.3). Notons que cette observation peut être différente si le nombre de cellules  $M > 8$ . La différence entre les estimateurs empiriques et les estimateurs non paramétriques ce remarque lorsque dans certaines cellules, le nombre d'éléments est négligeables ou nul (voir Modèle 2, Section 4.4.1). En effet, le Tableau 4.6 montre que les taux de bon classement basés sur les estimateurs empiriques sont non définis (NA= Not Available) pour tout  $n = 100, 300, 500$ . Cela est dû à une faible quantité d'éléments dans les cellules 1, 3 et 5. Par contre, avec les estimateurs non paramétriques, le taux de bon classement varie entre 0.52400 et 0.56670 (Tableau 4.6). Ce qui justifie l'utilisation d'estimateurs non paramétriques dans la suite du travail. L'étude de simulation révèle aussi qu'il y a une influence des paramètres  $\alpha$  et  $\beta$  sur la méthode de sélection (Figures 4.1 - 4.3). Cela nous a conduit à utiliser la méthode de validation croisée dans le but de calculer automatiquement les paramètres optimaux  $\alpha$  et  $\beta$  qui maximisent le taux de bon classement de notre méthode. L'étude comparative avec la méthode de *Mahat et al. (2007)* à montrer que les deux méthodes de sélection sont équivalentes sur la base du taux de bon classement (Tableau 4.4).

L'application aux données réelles vient confirmée la comparaison faite par simulations. En effet, le Tableau 4.7 montre que le taux de bon classement des deux méthodes reste équivalent, par contre notre méthode sélectionne moins de variables que celle de *Mahat et al. (2007)*, ce qui est un avantage considérable.

TABLE 4.1 – Taux de bon classement moyen sur 1000 réplifications avec différentes fonctions de pénalité ( $n = 100, n_1=n_2=50$ ),  $\alpha = 0.25, \beta = 0.5, \lambda = 0.5$ ).

Fonctions	Taux de bon classement	
	Estimateur empirique	Estimateur non-paramétrique
$g_1$	0.608	0.608
$g_2$	0.609	0.609
$g_3$	0.610	0.610
$g_4$	0.609	0.609
$g_5$	0.608	0.608
$g_6$	0.61028	0.61028
$g_7$	0.60903	0.60903
$g_8$	0.61066	0.61066
$g_9$	0.60898	0.60898
$g_{10}$	0.60842	0.60842
$g_{11}$	0.60948	0.60948
$g_{12}$	0.60915	0.60915
$g_{13}$	0.60908	0.60908

TABLE 4.2 – Taux de bon classement moyen sur 1000 réplifications avec différentes fonctions de pénalité ( $n = 300, n_1=n_2=150, \alpha = 0.25, \beta = 0.5, \lambda = 0.5$ ).

Fonctions	Taux de bon classement	
	Estimateur empirique	Estimateur non-paramétrique
$g_1$	0.56421	0.5642433
$g_2$	0.5632833	0.5633233
$g_3$	0.56417	0.56417
$g_4$	0.56346	0.56346
$g_5$	0.56382	0.56382
$g_6$	0.56343	0.56343
$g_7$	0.5637433	0.5637433
$g_8$	0.5635433	0.5635433
$g_9$	0.5631267	0.5631267
$g_{10}$	0.5637933	0.5637933
$g_{11}$	0.5640433	0.5640433
$g_{12}$	0.5634567	0.5634567
$g_{13}$	0.5637533	0.5637533

TABLE 4.3 – Taux de bon classement moyen sur 1000 réplifications avec différentes fonctions de pénalité ( $n = 500, n_1=n_2=250, \alpha = 0.25, \beta = 0.5, \lambda = 0.5$ ).

Fonctions	Taux de bon classement	
	Estimateur empirique	Estimateur non-paramétrique
$g_1$	0.54998	0.54998
$g_2$	0.550478	0.550676
$g_3$	0.55032	0.550396
$g_4$	0.550494	0.550584
$g_5$	0.549824	0.559824
$g_6$	0.550584	0.550622
$g_7$	0.550504	0.550546
$g_8$	0.550088	0.550132
$g_9$	0.549724	0.54934
$g_{10}$	0.550132	0.550186
$g_{11}$	0.550362	0.55037
$g_{12}$	0.550464	0.550464
$g_{13}$	0.550284	0.550284

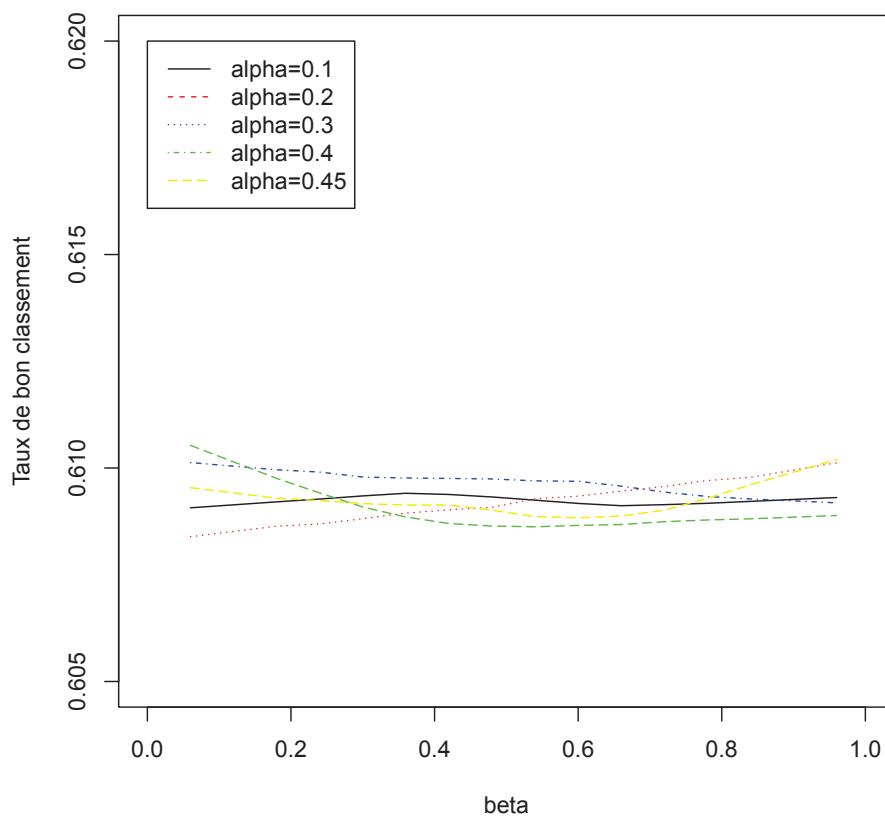


FIGURE 4.1 – Taux de bon classement moyen sur 1000 réplifications pour différentes valeurs de  $\alpha$  en fonction de  $\beta$ .  $n = 100$

## 4.5 Conclusion

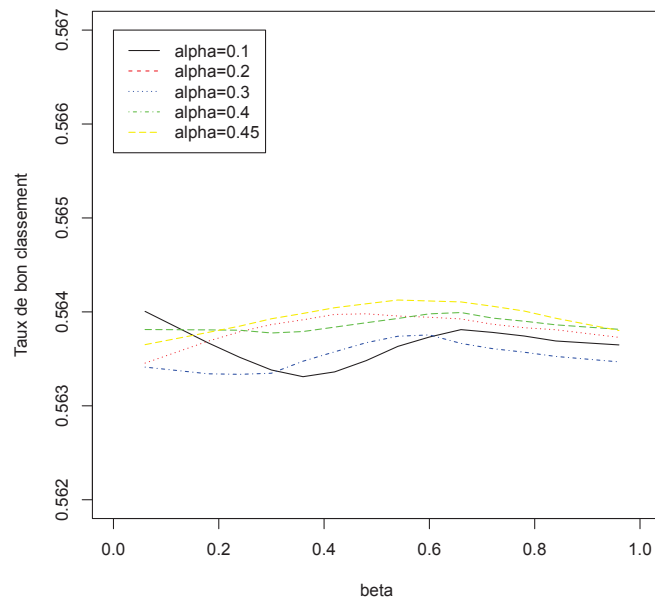


FIGURE 4.2 – Taux de bon classement moyen sur 1000 réplifications pour différentes valeurs de  $\alpha$  en fonction de  $\beta$ .  $n=300$

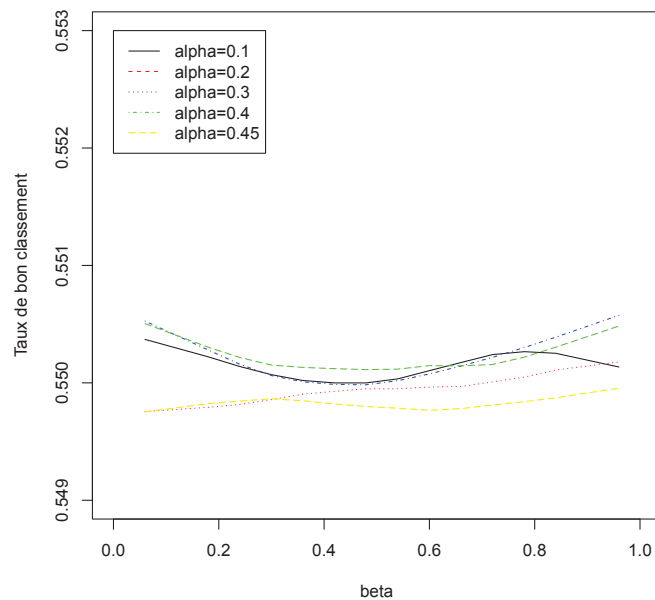


FIGURE 4.3 – Taux de bon classement moyen sur 1000 réplifications pour différentes valeurs de  $\alpha$  en fonction de  $\beta$ .  $n=500$

TABLE 4.4 – *Taux de bon classement moyen sur 1000 réplifications de notre méthode contre celle de Mahat.*

		Taux de bon classement	
n	$n_1=n_2$	Notre méthode	Méthode de Mahat
100	50	0.6086	0.6082667
200	100	0.5787	0.5783667
300	150	0.5624444	0.5622111
400	200	0.5553417	0.5554667
500	250	0.5498933	0.5495067

TABLE 4.5 – *Taux de bon classement moyen de notre méthode sur 500 réplifications.*

Taux de bon classement		
n	$n_1=n_2=n_3$	Notre Méthode
6000	2000	0.4998
7500	2500	0.5015
9000	3000	0.5159
10500	3500	0.5301

TABLE 4.6 – Taux de bon classement moyen sur 1000 réplifications avec  $f_7 = n^{-1/4}/h_7$ ,  $g_7 = n^{-1/4}h_7$ ,  $\alpha = 0.25$  et  $\beta = 0.5$ . Cas du Modèle 2.

$n$	$n_1 = n_2$	Taux de bon classement	
		Estimateur empirique	Estimateur non paramétrique
100	50	NA	0.54000
300	150	NA	0.56670
500	250	NA	0.52400

TABLE 4.7 – Taux de bon classement moyen et nombre de variables sélectionnées (NV) par la méthode 10-fold cross-validation.

	Notre méthode	Méthode de Mahat
CC	0.87900	0.87800
NV	4.00	6.00



## **Troisième partie**

# **Sélection des variables sur un modèle linéaire fonctionnel additif**

---

# ÉTAT DE L'ART SUR LES MODÈLES LINÉAIRES FONCTIONNELS : ESTIMATION ET TESTS

---

## Sommaire

---

<b>5.1</b>	<b>Présentation du modèle et notations . . . . .</b>	<b>124</b>
<b>5.2</b>	<b>Estimation . . . . .</b>	<b>126</b>
5.2.1	Analyse en composantes principales . . . . .	126
5.2.2	Modèle linéaire fonctionnel additif . . . . .	128
<b>5.3</b>	<b>Tests . . . . .</b>	<b>129</b>
5.3.1	Tests dans les modèles linéaires fonctionnels . . . . .	129
5.3.2	Tests dans les modèles linéaires additifs . . . . .	130

---

---

## Introduction

---

Les récentes innovations technologiques permettent aujourd'hui de recueillir des données discrétisées de plus en plus fines, les rendant ainsi fonctionnelles. Les courbes de croissances, les images observées par satellites, les séries chronologiques etc, sont là quelques exemples qui montrent le champ d'application lié aux données fonctionnelles. Les premiers travaux dans ce domaine sont obtenus sous l'impulsion de [Grenander \(1981\)](#), [Dauxois et al. \(1982\)](#) et [Ramsay \(1982\)](#). L'objectif de ce chapitre est celui de rappeler quelques méthodes d'estimation et de test dans les modèles fonctionnels et les modèles fonctionnels additifs. La littérature n'est pas très fournie dans les techniques d'estimation et de test dans le cas fonctionnel comparé au cas multidimensionnel. Ainsi, il n'est bien entendu pas question de dresser un panorama complet de tous les résultats obtenus dans le domaine de la statistique fonctionnelle, on se limitera à présenter les résultats issues des travaux de [Cardot et al. \(1999, 2003a, 2004\)](#). D'autres méthodes d'analyse de données fonctionnelles peuvent être aussi consultées dans la thèse de [Laurent Delsol \(2008\)](#).

## 5.1 Présentation du modèle et notations

Un modèle linéaire fonctionnel est un modèle qui caractérise la relation entre une variable aléatoire réelle  $Y$  et une covariable aléatoire fonctionnelle  $X$ . Il s'écrit comme le produit scalaire entre un coefficient fonctionnel aléatoire  $\phi$  et une covariable fonctionnelle aléatoire  $X$  ([Cardot et al. \(1999\)](#)) :

$$Y = \langle \phi, X \rangle + \varepsilon. \quad (5.1)$$

Or,

$$\langle \phi, X \rangle = \int_{\mathcal{E}} \phi(t)X(t)dt,$$

alors on définit le modèle linéaire fonctionnel par la relation ([Cardot et al. \(1999\)](#)) :

$$Y = \int_{\mathcal{E}} \phi(t)X(t)dt + \varepsilon, \quad (5.2)$$

---

## 5.1 Présentation du modèle et notations

---

où  $\varepsilon$  est une variable aléatoire indépendante de  $X$ . En général, la variable aléatoire fonctionnelle  $X$  est notée  $X = \{X(t), t \in \mathcal{C}\}$  et on suppose que le couple  $(X, Y)$  est défini dans le même espace de probabilité  $(\Omega, \mathcal{A}, P)$ . Alternativement, l'équation (5.2) peut être réécrite dans le cas où la variable  $X$  est une variable aléatoire à valeurs dans un espace de Hilbert  $\mathcal{H}$ . Elle s'écrit de la façon suivante ([Cardot et al. \(1999\)](#)) :

$$Y = \Phi \{X(t); t \in \mathcal{C}\} + \varepsilon, \quad (5.3)$$

où  $\Phi$  est un opérateur linéaire continu et défini dans  $\mathcal{L}(\mathcal{H}, \mathbb{R})$ , qui est l'ensemble des opérateurs linéaires continus définis sur  $\mathcal{H}$  et à valeurs dans  $\mathbb{R}$ . Dans ce modèle, on suppose que les variables aléatoires  $X$  et  $\varepsilon$  sont centrées, c'est-à-dire  $\forall t \in \mathcal{C}, \mathbb{E}(X(t)) = 0$  et  $\mathbb{E}(\varepsilon) = 0$ . De plus, on suppose que  $\mathbb{E}(\|X\|^2) < +\infty$ .

On définit l'opérateur de covariance associé à la variable aléatoire  $X$  par ([Cardot et al. \(1999\)](#)) :

$$\Gamma(x)(t) = \int_{\mathcal{C}} \mathbb{E}[X(t)X(s)](x)(s)ds, \quad \forall x \in \mathcal{H}, \quad \text{et } \forall t \in \mathcal{C}, \quad (5.4)$$

et l'opérateur de covariance croisé associé aux variables aléatoires  $X$  et  $Y$  par :

$$\Delta x = \int_{\mathcal{C}} \mathbb{E}[X(t)Y]x(t)dt, \quad x \in \mathcal{H}. \quad (5.5)$$

De l'équation (5.3), [Cardot et al. \(1999\)](#) montrent que les opérateurs  $\Gamma$  et  $\Delta$  sont liés par la relation  $\Delta = \Phi\Gamma$ .

La problématique de l'estimation du coefficient fonctionnel  $\phi$  (ou de l'opérateur  $\Phi$ ) a été largement abordée dans la littérature. On distingue principalement deux méthodes : la première méthode consiste à exprimer le coefficient  $\phi$  sur une base d'éléments de  $L^2_{\mathcal{C}}$  qui minimisent un critère des moindres carré pénalisé. On peut retrouver cette méthode dans [Hastie et Mallows \(1993\)](#), [Marx et al. \(1996, 1999\)](#), [?, Cardot et al. \(2003a\)](#). La seconde méthode est basée sur l'analyse en composantes principales fonctionnelle. Elle consiste à réduire la dimension de l'espace fonctionnel initial en un sous-espace de dimension fini [Cardot et al. \(1999\)](#). D'autres méthodes d'estimation peuvent être trouver dans la thèse de [Crambes \(2006\)](#).

## 5.2 Estimation

### 5.2.1 Analyse en composantes principales

L'idée principale de cette approche est de réduire la dimension des variables fonctionnelles en les décomposant dans une base orthonormale en utilisant les  $k$  premières valeurs propres de l'opérateur  $\Gamma$ .

#### Existence, unicité et estimation de $\phi$

L'existence et l'unicité de la constante aléatoire  $\phi$  sont obtenues en supposant que les valeurs propres de l'opérateur  $\Gamma$  sont non nulles et que la condition suivante est vérifiée :

$$\sum_{j=1}^{+\infty} \frac{\langle \mathbb{E}(X_i(Y_i - \mu)), V_j \rangle^2}{\lambda_j^2} < +\infty,$$

où  $\lambda_j$  et  $V_j$  sont respectivement les valeurs propres et vecteurs propres de l'opérateur  $\Gamma$ . Une condition nécessaire et suffisante est obtenue si  $\phi$  est solution du problème d'optimisation suivant (*Cardot et al. (2003b)*) :

$$\phi = \arg \min_{(\theta, \beta) \in \mathbb{R} \times \mathcal{H}} \mathbb{E} \left( (Y_i - \theta - \langle \beta, X_i \rangle)^2 \right).$$

Sous ses conditions,  $\phi$  est explicitement donné par l'expression

$$\phi = \sum_{j=1}^{+\infty} \frac{\Delta V_j}{\lambda_j} V_j.$$

Soit  $(X_i, Y_i)_{i \leq n}$  un échantillon i.i.d de  $(X, Y)$ . Les estimateurs empiriques des opérateurs  $\Gamma$  et  $\Delta$  sont donnés par :

$$\Gamma_n x(t) = \frac{1}{n} \sum_{i=1}^n \langle X_i, x \rangle_{\mathcal{H}} X_i(t), \quad \forall x \in \mathcal{H}, \quad \text{et } t \in \mathcal{C},$$

et

$$\Delta_n x = \frac{1}{n} \sum_{i=1}^n \langle X_i, x \rangle_{\mathcal{H}} Y_i, \quad \forall x \in \mathcal{H}.$$

En utilisant les estimateurs empiriques de  $\Gamma$  et  $\Delta$ , on obtient un estimateur de  $\phi$  par :

$$\hat{\phi} = \sum_{j=1}^K \frac{\Delta_n \hat{V}_j}{\hat{\lambda}_j} \hat{V}_j,$$

où  $(\hat{\lambda}_j, \hat{V}_j)_{j \in \mathbb{N}^*}$  sont les composantes principales de  $\Gamma_n$ .

En général, l'opérateur  $\Gamma$  n'est pas inversible, et s'il arrive qu'il soit inversible, son inverse quant à lui n'est pas nécessairement borné, car  $\mathcal{H}$  est un espace de dimension infinie. Alors, le calcul direct de l'inverse de  $\Gamma$  devient très difficile voir impossible.

Pour régler ce problème, [Cardot et al. \(1999\)](#) réduisent la dimension de la base formée des vecteurs propres de l'opérateur  $\Gamma$  de l'espace infini  $\mathcal{H}$ . On obtient un sous-espace de dimension finie noté  $\widehat{\mathcal{H}}_{k_n}$ , engendré par  $\{\hat{V}_j, j = 1, \dots, k_n\}$ , où  $(k_n)_{n \in \mathbb{N}}$  est une suite d'entiers positifs telle que  $\lim_{n \rightarrow +\infty} k_n = +\infty$ , avec  $n \leq k_n$ .

On considère  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_n \geq 0 = \hat{\lambda}_{n+1} = \dots$ , les valeurs propres de  $\Gamma_n$ , et  $\hat{V}_1, \hat{V}_2, \dots$  ses vecteurs propres orthogonaux associés, on définit dans  $\mathcal{H}$  le projecteur orthogonal  $\hat{\Pi}_{k_n}$  tel que :

$$\hat{\Pi}_{k_n} = \sum_{j=1}^{k_n} \hat{V}_j \otimes \hat{V}_j.$$

En supposant que  $\hat{\lambda}_{k_n} > 0$ , on définit l'estimateur  $\hat{\Phi}_{k_n}$  de  $\Phi$  par ([Cardot et al. \(1999\)](#))

$$\hat{\Phi}_{k_n} = \Delta_n \hat{\Pi}_{k_n} (\hat{\Pi}_{k_n} \Gamma_n \hat{\Pi}_{k_n})^{-1} = \sum_{j=1}^K \frac{\Delta_n \hat{V}_j}{\hat{\lambda}_j} \langle \hat{V}_j, \cdot \rangle. \quad (5.6)$$

Soit l'hypothèse suivante :

( $H_0$ )  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_{k_n} \geq 0$  presque sûrement. On obtient le théorème suivant :

**Théorème 12** *Considérons l'hypothèse ( $H_0$ ), et supposons que les hypothèses suivantes sont sa-*

atisfaites :

- (H<sub>1</sub>)  $\lambda_1 \geq \lambda_2 > \dots > 0$ ,
- (H<sub>2</sub>)  $\mathbb{E}(\|X\|_{\mathcal{H}}^4) < +\infty$ ,
- (H<sub>3</sub>)  $\lim_{n \rightarrow +\infty} n\lambda_{k_n}^4 = +\infty$ ,
- $\lim_{n \rightarrow +\infty} \frac{n\lambda_{k_n}^2}{\left(\sum_{j=1}^{k_n} a_j\right)^2} = +\infty$ .

Alors  $\|\widehat{\Phi}_{k_n} - \Phi\|_{\mathcal{L}(\mathcal{H})} \rightarrow 0$  en probabilité quand  $n \rightarrow +\infty$ .

Le coefficient  $a_j$  est défini par :

- $a_j = \frac{2\sqrt{2}}{\lambda_1 - \lambda_2}$  si  $j = 1$ ,
- $a_j = \frac{2\sqrt{2}}{\min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})}$  si  $j \neq 1$ .

**Théorème 13** *Considérons les hypothèses (H<sub>0</sub>), (H<sub>1</sub>) et supposons que les hypothèses suivantes sont satisfaites :*

- (H<sub>4</sub>)  $\|X\|_{\mathcal{H}} \leq c_1$ ,
- (H<sub>5</sub>)  $|\varepsilon| \leq c_2$ ,
- (H<sub>6</sub>)  $\lim_{n \rightarrow +\infty} \frac{n\lambda_{k_n}^2}{\left(\sum_{j=1}^{k_n} a_j\right)^2} \log n = +\infty$ .

Alors  $\|\widehat{\Phi}_{k_n} - \Phi\|_{\mathcal{L}(\mathcal{H})} \rightarrow 0$  presque sûrement quand  $n \rightarrow +\infty$ .

### 5.2.2 Modèle linéaire fonctionnel additif

*Cardot et al.* (2004) utilisent une généralisation du modèle (5.2) en considérant plusieurs prédicteurs fonctionnels  $X_1, X_2, \dots, X_p$ , où les  $X_k = \{X_k(t), t \in \mathcal{C}_k\}$  ( $k = 1, \dots, p$ ) sont des prédicteurs fonctionnels centrés et de carrés intégrables, définis sur un compact  $\mathcal{C}_k$  de  $\mathbb{R}$ . Ce modèle est défini par l'équation suivante :

$$Y = \sum_{k=1}^p \int_{\mathcal{C}_k} \phi_k(t) X_k(t) dt + \varepsilon = \sum_{k=1}^p \Phi_k(X_k) + \varepsilon. \quad (5.7)$$

Un estimateur de chaque coefficient fonctionnel  $\phi_k$  ( $k = 1, \dots, p$ ) est obtenu en effectuant

une approximation de l'équation (5.7) de la façon suivante :

$$\sum_{k=1}^p \int_{\mathcal{C}_k} \phi_k(t) X_k(t) dt \approx \sum_{k=1}^p \sum_{j=1}^{k_{n,k}} \beta_{jk} \langle X_k, \widehat{V}_{jk} \rangle, \quad (5.8)$$

où  $\forall k = 1, \dots, p$ ,  $\widehat{V}_{jk}$  sont les vecteurs propres de l'opérateur de covariance empirique  $\Gamma_{n,k}$  définis sur l'échantillon  $(X_{k,i})_{1 \leq i \leq n}$ . Ainsi, un estimateur des coefficients fonctionnels  $\phi_k$  est donné par :

$$\widehat{\phi}_k(t) = \sum_{j=1}^{k_{n,k}} \widehat{\beta}_{jk} \widehat{V}_{jk}(t), \quad k = 1, \dots, p, \quad (5.9)$$

$\widehat{\beta}_{jk}$  est obtenu en utilisant la méthode des moindres carrés sur le modèle approché (Equation (5.8)).

L'analyse en composantes principales fonctionnelles est aussi utilisée dans le but de classer les couples  $(X, Y)$ , en utilisant le coefficient  $\phi$  comme fonction discriminante. Pour plus de détails, on peut consulter [Preda, C.; Saporta, G. \(2004\)](#); [Preda, C. et Saporta, G. \(2005a, 2007\)](#).

## 5.3 Tests

### 5.3.1 Tests dans les modèles linéaires fonctionnels

On trouve un nombre assez restreint de tests dans la littérature consacrés à la statistique fonctionnelle. On peut consulter la thèse de [Laurent Delsol \(2008\)](#) pour avoir un aperçu de tests déjà proposés. Dans cette section, nous nous limiterons à présenter les tests proposés dans [Cardot et al. \(2003a, 2004\)](#).

La notion de test dans le modèle (5.2) permet de vérifier le lien entre la variable aléatoire réelle  $Y$  et la variable aléatoire fonctionnelle  $X$ . Cela revient à tester la nullité du coefficient



fonctionnel  $\phi$ . [Cardot et al. \(2003a\)](#) proposent le test suivant :

$$H_0 : \phi = 0 \text{ contre } H_1 : \phi \neq 0.$$

Ce test est équivalent sous l'hypothèse  $H_0$ , à tester la nullité de l'opérateur de covariance croisé  $\Delta$ . Deux statistiques de test basées sur la norme de l'opérateur empirique  $\Delta_n$  sont établies :

$$D_n = \frac{1}{\hat{\sigma}_n^2} \|\sqrt{n}\Delta_n \hat{A}_n\|^2, \text{ et } T_n = \frac{1}{\sqrt{p_n}} \left( \frac{1}{\hat{\sigma}_n^2} \|\sqrt{n}\Delta_n \hat{A}_n\|^2 - k_n \right).$$

Les Théorèmes 1 et 2 de [Cardot et al. \(2003a\)](#) montrent que  $D_n$  suit approximativement une loi de  $\chi^2$  (quand  $n \rightarrow +\infty$ ), et sous des conditions appropriées sur la suite  $(k_n)_{n \in \mathbb{N}^*}$ ,  $T_n$  converge en probabilité sous l'hypothèse  $H_0$  vers une variable gaussienne centrée et de variance 2. Sous la même hypothèse  $H_0$  et de l'indépendance des variables aléatoire  $X$  et  $\varepsilon$ , ils montrent que  $\sqrt{n}\Delta_n$  converge en distribution vers une variable gaussienne centrée  $G_\Delta$  avec un opérateur de covariance  $\sigma^2\Gamma$ , où  $\sigma^2$  est la variance de  $Y$ .

### 5.3.2 Tests dans les modèles linéaires additifs

Dans le cas d'un modèle linéaire fonctionnel additif (équation (5.7)), la nullité de la fonctionnelle  $\phi_k$  ne signifie pas que l'opérateur de covariance croisé entre  $Y$  et  $X_k$  est nul. Cela est dû au fait que  $X_k$  et les autres covariables fonctionnelles  $X_\ell$ , ( $\ell \neq k$ ) ne sont pas nécessairement corrélées ([Cardot et al. \(2004\)](#)). Ainsi, l'approche proposée par [Cardot et al. \(2003b\)](#) dans ce contexte n'est plus valable.

Pour y remédier, plusieurs méthodes ont été envisagées. On peut utiliser un test non paramétrique lorsque  $Y$  et  $X$  dans le modèle (5.7) sont approchées par un lien non paramétrique. C'est ce qu'ont proposé [Ferraty et Vieu \(2002\)](#) en introduisant un estimateur à noyau de  $\Phi$ . Il y a le test du rapport de vraisemblance de [Azzalini et Bowman \(1993\)](#). On peut approcher l'intégrale de l'équation (5.7) en utilisant une approximation de Riemann, ou l'écrire comme combinaison linéaire des vecteurs propres de l'opérateur de covariance croisé entre  $Y$  et  $X_k$  ( voir équation (5.8)), et utiliser le test du rapport de vraisemblance sous les hypothèses suivantes ([Cardot et al. \(2004\)](#)) :

$$\tilde{H}_0 : \hat{\beta}_{jk_0} = 0, j = 1, \dots, k_0.$$

La statistique de test est donnée par :

$$F_n = \frac{RSS_0 - RSS_1}{RSS_1} \times \frac{(n - 1 - \sum_{k=1}^p k_{n,k})}{\sum_{k \neq k_0} k_{n,k}},$$

où

$$RSS_1 = \sum_{i=1}^n \left( Y_i - \sum_{k=1}^p \langle \hat{\phi}_k, X_i \rangle \right)^2,$$

et

$$RSS_0 = \sum_{i=1}^n \left( Y_i - \sum_{k \neq k_0} \langle \tilde{\phi}_k, \tilde{X}_i \rangle \right)^2,$$

pour  $k \neq k_0$ ,  $\tilde{\phi}_k$  est l'estimateur de  $\phi$  en considérant le modèle linéaire fonctionnel additif sous l'hypothèse nulle.

---

# SÉLECTION DES VARIABLE SUR UN MODÈLE LINÉAIRE FONCTIONNEL ADDITIF

---

## Sommaire

---

<b>6.1</b>	<b>Modèle et critère de sélection . . . . .</b>	<b>134</b>
6.1.1	Modèle, définition et motivation . . . . .	134
6.1.2	Critère de sélection et propriétés . . . . .	135
6.1.3	Estimation et convergence du critère . . . . .	140
6.1.4	Sélection des variables . . . . .	144
<b>6.2</b>	<b>Sélection des variables dans un modèle linéaire fonctionnel additif . . .</b>	<b>147</b>
6.2.1	Approximation du modèle linéaire fonctionnel additif : critère de sélection pour les variables fonctionnelles . . . . .	147
<b>6.3</b>	<b>Conclusion . . . . .</b>	<b>151</b>

---

---

## Introduction

---

Parmi les domaines d'études où la statistique fonctionnelle a été abordée y figure l'analyse discriminante. Son objectif est d'attribuer une classe à une nouvelle donnée fonctionnelle à partir d'un échantillon de données dont on connaît la classe d'appartenance : on parle d'analyse discriminante fonctionnelle. Il semble que peu de travaux aient été consacrés à cette partie de la statistique fonctionnelle. Ce constat, est également observé dans le domaine de la sélection des variables fonctionnelles.

Ainsi, dans ce chapitre, on s'intéresse au problème de sélection des variables dans un modèle fonctionnel additif. On se propose d'étendre au cas fonctionnel, le critère de sélection développé dans les Chapitres 2 et 4. Plus précisément, on propose un critère de sélection des variables fonctionnelles en ce plaçant dans le cas de la régression d'une variable aléatoire réelle sur une somme de variables aléatoires fonctionnelles.

Une première approche consiste à considérer le modèle linéaire de la forme :

$$Y = \sum_{k=1}^p b_k^T X_k + \varepsilon, \quad (6.1)$$

où,  $Y$  et  $\varepsilon$  sont des variables aléatoires réelles,  $b_k$  est un vecteur de  $\mathbb{R}^q$  et  $X_k$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^q$ , où les variables  $\varepsilon$  et  $X_k$  sont indépendantes. On suppose que le vecteur  $X_k$  et la variable  $Y$  sont centrées. Ce modèle est en fait une généralisation du modèle (1) de [Nkiet \(2001\)](#).

Notre démarche de sélection de variables fonctionnelles s'effectue de la façon suivante :

1. Nous généralisons la méthode de sélection des variables de [Nkiet \(2001\)](#) à partir du modèle linéaire (6.1) ;
2. En utilisant une approximation semblable à celle de [Cardot et al. \(2004\)](#) (Section 5.2.2, Equation (5.8)), nous proposons un critère de sélection aux modèles linéaires fonctionnels additifs, critère qui sera à la base de toute la procédure de sélection comme établi dans les Chapitres 2 et 4.

## 6.1 Modèle et critère de sélection

### 6.1.1 Modèle, définition et motivation

Considérons à nouveau l'équation (6.1) :

$$Y = \sum_{k=1}^p b_k^T X_k + \varepsilon,$$

où

$$X_k = \begin{pmatrix} X_{k1} \\ X_{k2} \\ \vdots \\ X_{kq} \end{pmatrix} \in \mathbb{R}^q, b_k = \begin{pmatrix} b_{k1} \\ b_{k2} \\ \vdots \\ b_{kq} \end{pmatrix} \in \mathbb{R}^q. \text{ Posons } X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} = (X_1^T | X_2^T | \cdots | X_p^T)^T \in \mathbb{R}^{pq}.$$

$Y$  et  $\varepsilon$  sont des variables aléatoires réelles. On suppose que  $X, Y$  et  $\varepsilon$  sont centrées. On note par  $\langle \cdot, \cdot \rangle_{\mathbb{R}^{pq}}$  le produit scalaire dans  $\mathbb{R}^{pq}$  tel que pour tout  $U, V \in \mathbb{R}^{pq}$  :

$$\langle U, V \rangle_{\mathbb{R}^{pq}} = \sum_{k=1}^p \langle U_k, V_k \rangle_{\mathbb{R}^q}.$$

Remarquons que si  $q = 1$ , le modèle (6.1) se ramène au modèle (1) de [Nkiet \(2001\)](#). Ainsi, nous proposons une généralisation de la méthode de sélection des variables de [Nkiet \(2001\)](#) à travers le modèle linéaire (6.1).

Dans l'équation (6.1), on s'intéresse à la sélection des vecteurs  $X_k$  ( $k = 1, \dots, p$ ), c'est-à-dire identifier les vecteurs  $X_k$  qui ne contribuent pas à expliquer la variable aléatoire réelle  $Y$ . Nous savons qu'un vecteur  $X_k$  n'explique pas la variable aléatoire  $Y$  dans le modèle (6.1) si le vecteur de coefficients  $b_k$  est nul. Ainsi, on note par  $\mathcal{I}$ , l'ensemble des indices des vecteurs candidats à la sélection, tel que  $\mathcal{I} = \{1, \dots, p\}$ . Posons :

$$\mathcal{I}_0 = \{k \in \mathcal{I} | b_k = 0\},$$

le sous-ensemble des indices des vecteurs qui n'expliquent pas la variable  $Y$ , et

$$\mathcal{I}_1 = \mathcal{I} - \mathcal{I}_0 = \{k \in \mathcal{I} | b_k \neq 0\},$$

le sous-ensemble des indices des vecteurs qui expliquent la variable  $Y$ , et on suppose que le sous-ensemble  $\mathcal{I}_1 \neq \emptyset$ .

### 6.1.2 Critère de sélection et propriétés

Tout comme au Chapitre 2, nous supposons que :

1.  $\mathbb{E}(Y^4) < +\infty$  et  $\mathbb{E}(\|X\|_{\mathbb{R}^{pq}}^4) < +\infty$ ,
2.  $X_1, \dots, X_p$  sont linéairement indépendants.

On note par  $V_X$  l'opérateur de covariance associé au vecteur aléatoire  $X$ . Il peut se mettre sous la forme :

$$V_X = \mathbb{E}(X \otimes X) = \begin{pmatrix} V_{11} & V_{12} & \cdots & V_{1q} \\ V_{21} & V_{22} & \cdots & V_{2q} \\ \vdots & \vdots & \cdots & \vdots \\ V_{p1} & V_{p2} & \cdots & V_{pq} \end{pmatrix},$$

et on le suppose inversible. On note par  $V_{XY} = \mathbb{E}(YX)$  l'opérateur de covariance croisé associé au vecteur aléatoire  $X$  et à la variable aléatoire réelle  $Y$ , puis on considère le critère suivant :

$$C_{\mathcal{K}} = \|V_{XY} - V_X \Pi_{\mathcal{K}} V_{XY}\|, \quad (6.2)$$

où  $\|\cdot\|$  désigne la norme d'opérateur définie par  $\|A\| = \sqrt{\text{tr}(A^*A)}$ ,  $\Pi_{\mathcal{K}}$  est un opérateur défini par :

$$\Pi_{\mathcal{K}} = A_{\mathcal{K}}^* (A_{\mathcal{K}} V_X A_{\mathcal{K}}^*)^{-1} A_{\mathcal{K}},$$

où  $A_{\mathcal{K}}$  est un opérateur linéaire, défini de la façon suivante :

$$A_{\mathcal{K}} : x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \in \mathbb{R}^{pq} \mapsto (x_i)_{i \in \mathcal{K}} \in \mathbb{R}^{\text{card}(\mathcal{K})q},$$

où  $\text{card}(\mathcal{K})$  désigne le cardinal du sous-ensemble  $\mathcal{K}$ . Le critère (6.2) permet de mesurer la perte d'information lorsqu'un sous-ensemble de vecteurs  $\mathcal{K}$  est sélectionné. Dans le Lemme 10, nous donnons une caractérisation du sous-ensemble  $\mathcal{I}_1$ .

**Lemme 10**  $\forall \mathcal{H} \subset \mathcal{I}, \mathcal{I}_1 \subset \mathcal{H} \Leftrightarrow C_{\mathcal{H}} = 0.$

**Preuve:** Considérons l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ , et considérons les opérateurs linéaires suivants :

$$L_1 : x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \in \mathbb{R}^{pq} \mapsto \sum_{k=1}^p \sum_{\ell=1}^q x_{k\ell} X_{k\ell} \in L^2(\Omega, \mathcal{A}, P)$$

et

$$L_2 : y \in \mathbb{R} \mapsto \langle y, Y \rangle_{\mathbb{R}} \in L^2(\Omega, \mathcal{A}, P).$$

On note par  $L_1^*$  et  $L_2^*$  les adjoints respectifs des opérateurs  $L_1$  et  $L_2$  et définis tels que :

$$L_1^* : Z \in L^2(\Omega, \mathcal{A}, P) \mapsto \mathbb{E}(ZX) \in \mathbb{R}^{pq}, \text{ et } L_2^* : Z \in L^2(\Omega, \mathcal{A}, P) \mapsto \mathbb{E}(ZY) \in \mathbb{R}.$$

En plus, on a :

$$\begin{cases} V_X = L_1^* L_1 \\ V_{XY} = L_1^* L_2. \end{cases}$$

On a ainsi :

$$V_{XY} - V_X \Pi_{\mathcal{H}} V_{XY} = L_1^* L_2 - L_1^* L_1 A_{\mathcal{H}}^* (A_{\mathcal{H}}^* L_1^* L_1 A_{\mathcal{H}}^*)^{-1} A_{\mathcal{H}} L_1^*.$$

Notons par  $Im(T)$ , l'image de l'opérateur  $T$  et  $\Pi_{Im(T)} = T(T^*T)^{-1}T^*$  le projecteur sur  $Im(T)$ . Alors :

$$\begin{aligned} \Pi_{Im(L_1 A_{\mathcal{H}}^*)} &= L_1 A_{\mathcal{H}}^* ((L_1 A_{\mathcal{H}}^*)^* (L_1 A_{\mathcal{H}}^*))^{-1} (L_1 A_{\mathcal{H}}^*)^* \\ &= L_1 A_{\mathcal{H}}^* (A_{\mathcal{H}} L_1^* L_1 A_{\mathcal{H}}^*)^{-1} A_{\mathcal{H}} L_1^*. \end{aligned}$$

D'où

$$\begin{aligned} V_{XY} - V_X \Pi_{\mathcal{H}} V_{XY} &= L_1^* L_2 - L_1^* L_1 A_{\mathcal{H}}^* (A_{\mathcal{H}}^* L_1^* L_1 A_{\mathcal{H}}^*)^{-1} A_{\mathcal{H}} L_1^* L_2 \\ &= L_1^* L_2 - L_1^* \Pi_{Im(L_1 A_{\mathcal{H}}^*)} L_2 \\ &= L_1^* (I_{\mathcal{L}(\mathbb{R}^{pq})} - \Pi_{Im(L_1 A_{\mathcal{H}}^*)}) L_2 \\ &= L_1^* \Pi_{Im(L_1 A_{\mathcal{H}}^*)^\perp} L_2 \end{aligned}$$

## 6.1 Modèle et critère de sélection

où  $Im(L_1 A_{\mathcal{X}}^*)^\perp$  désigne l'orthogonal de  $Im(L_1 A_{\mathcal{X}}^*)$ . Ainsi

$$C_{\mathcal{X}} = \|L_1^* \Pi_{Im(L_1 A_{\mathcal{X}}^*)^\perp} L_2\| \quad (6.3)$$

$\forall \alpha \in \mathbb{R}$ , on a

$$\begin{aligned} L_2(\alpha) &= \alpha Y \\ &= \alpha \left( \sum_{k=1}^p b_k^T X_k + \varepsilon \right) \\ &= \alpha \sum_{k=1}^p b_k^T X_k + \alpha \varepsilon. \end{aligned}$$

Alors

$$\begin{aligned} L_1^* \Pi_{Im(L_1 A_{\mathcal{X}}^*)^\perp} L_2(\alpha) &= L_1^* \Pi_{Im(L_1 A_{\mathcal{X}}^*)^\perp} \left( \alpha \sum_{k=1}^p b_k^T X_k + \alpha \varepsilon \right) \\ &= \alpha \sum_{k=1}^p L_1^* \Pi_{Im(L_1 A_{\mathcal{X}}^*)^\perp} b_k^T X_k + L_1^* \Pi_{Im(L_1 A_{\mathcal{X}}^*)^\perp} \alpha \varepsilon. \end{aligned}$$

**Montrons que**  $\alpha \varepsilon \in Im(L_1 A_{\mathcal{X}}^*)^\perp$

En effet,

$$\begin{aligned} \langle L_1 x, \alpha \varepsilon \rangle_{L^2(\Omega, \mathcal{A}, P)} &= \mathbb{E}(\alpha \varepsilon L_1 x) \\ &= \mathbb{E}(\alpha \varepsilon \langle x, X \rangle_{\mathbb{R}^{pq}}) \\ &= \alpha \mathbb{E}(\langle x, \varepsilon X \rangle_{\mathbb{R}^{pq}}) \\ &= \alpha \langle x, \mathbb{E}(\varepsilon X) \rangle_{\mathbb{R}^{pq}} \\ &= \alpha \langle x, \mathbb{E}(\varepsilon) \mathbb{E}(X) \rangle_{\mathbb{R}^{pq}} \\ &= 0. \end{aligned}$$



## 6.1 Modèle et critère de sélection

Alors  $\alpha\varepsilon \in \text{Im}(L_1)^\perp$ . Or  $\text{Im}(L_1 A_{\mathcal{K}}^*) \subset \text{Im}(L_1)$ , ce qui implique que  $\text{Im}(L_1)^\perp \subset \text{Im}(L_1 A_{\mathcal{K}}^*)^\perp$ , d'où  $\alpha\varepsilon \in \text{Im}(L_1 A_{\mathcal{K}}^*)^\perp$ . Ainsi

$$\begin{aligned} L_1^* \Pi_{\text{Im}(L_1 A_{\mathcal{K}}^*)^\perp} \alpha\varepsilon &= L_1^* \alpha\varepsilon \\ &= \mathbb{E}(\alpha\varepsilon X) \\ &= \alpha \mathbb{E}(\varepsilon X) \\ &= \alpha \mathbb{E}(\varepsilon) \mathbb{E}(X) \\ &= 0. \end{aligned}$$

Alors

$$\begin{aligned} L_1^* \Pi_{\text{Im}(L_1 A_{\mathcal{K}}^*)^\perp} L_2(\alpha) &= \alpha \sum_{k=1}^p L_1^* \Pi_{\text{Im}(L_1 A_{\mathcal{K}}^*)^\perp} b_k^T X_k \\ &= \alpha L_1^* \Pi_{\text{Im}(L_1 A_{\mathcal{K}}^*)^\perp} L_1(b) \end{aligned}$$

d'où

$$L_1^* \Pi_{\text{Im}(L_1 A_{\mathcal{K}}^*)^\perp} L_2(\alpha) = \alpha L_1^* \Pi_{\text{Im}(L_1 A_{\mathcal{K}}^*)^\perp} L_1(b), \quad (6.4)$$

avec

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix} \in \mathbb{R}^{pq}.$$

Alors si  $C_{\mathcal{K}} = 0$ , en prenant  $\alpha = 1$ , on a :

$$L_1^* \Pi_{\text{Im}(L_1 A_{\mathcal{K}}^*)^\perp} L_1(b) = 0.$$

Puisque pour tout opérateur  $T$ ,  $\text{Ker}(T^*T) = \text{Ker}(T)$ , alors

$$\Pi_{\text{Im}(L_1 A_{\mathcal{K}}^*)^\perp} L_1(b) = 0, \quad (6.5)$$

## 6.1 Modèle et critère de sélection

d'où  $L_1(b) \in \text{Im}(L_1 A_{\mathcal{H}}^*)$ . Soit  $\text{card}(\mathcal{H})$  le cardinal du sous-ensemble  $\mathcal{H}$  tel que  $\mathcal{H} = \{k_1, k_2, \dots, k_{\text{card}(\mathcal{H})}\}$ . Ainsi, il existe  $\beta \in \mathbb{R}^{\text{card}(\mathcal{H})q}$  où

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{\text{card}(\mathcal{H})} \end{pmatrix}$$

tel que :

$$\begin{cases} L_1(b) = L_1 A_{\mathcal{H}}^*(\beta) \\ \sum_{k=1}^p b_k^T X_k = \sum_{\ell=1}^{\text{card}(\mathcal{H})} \beta_\ell^T X_{k_\ell}, \end{cases}$$

alors,

$$\sum_{\ell=1}^{\text{card}(\mathcal{H})} (b_{k_\ell}^T - \beta_\ell^T) X_{k_\ell} + \sum_{\ell \in \mathcal{I} - \mathcal{H}} b_\ell^T X_\ell = 0$$

$$\begin{cases} \sum_{\ell=1}^{\text{card}(\mathcal{H})} (b_{k_\ell}^T - \beta_\ell^T) X_{k_\ell} = 0 \\ \sum_{\ell \in \mathcal{I} - \mathcal{H}} b_\ell^T X_\ell = 0. \end{cases}$$

Ainsi, pour tout  $\ell \in \mathcal{I} - \mathcal{H}$ , on a  $b_\ell = 0$ . On déduit que  $\mathcal{I} - \mathcal{H} \subset \mathcal{I}_0$  d'où  $\mathcal{I}_1 \subset \mathcal{H}$ . De même,

$$\begin{aligned} L_1^* \Pi_{\text{Im}(L_1 A_{\mathcal{H}}^*)^\perp} L_1(b) &= L_1^* \Pi_{\text{Im}(L_1 A_{\mathcal{H}}^*)^\perp} \sum_{k=1}^p b_k^T X_k \\ &= L_1^* \Pi_{\text{Im}(L_1 A_{\mathcal{H}}^*)^\perp} \left( \sum_{\ell=1}^{\text{card}(\mathcal{H})} b_{k_\ell}^T X_{k_\ell} + \sum_{\ell \in \mathcal{I} - \mathcal{H}} b_\ell^T X_\ell \right). \end{aligned}$$

Si  $\mathcal{I}_1 \subset \mathcal{H}$ , alors  $\mathcal{I} - \mathcal{H} \subset \mathcal{I}_0$ , alors  $\forall \ell \in \mathcal{I} - \mathcal{H}$ ,  $\ell \in \mathcal{I}_0$ , c'est-à-dire  $b_\ell = 0$ . Donc

$$\begin{aligned} L_1^* \Pi_{\text{Im}(L_1 A_{\mathcal{H}}^*)^\perp} L_1(b) &= L_1^* \Pi_{\text{Im}(L_1 A_{\mathcal{H}}^*)^\perp} \sum_{\ell=1}^{\text{card}(\mathcal{H})} b_{k_\ell}^T X_{k_\ell} \\ &= L_1^* \Pi_{\text{Im}(L_1 A_{\mathcal{H}}^*)^\perp} L_1 A_{\mathcal{H}}^*(b) \\ &= 0 \end{aligned}$$

car  $L_1 A_{\mathcal{H}}^*(b) \in \text{Im}(L_1 A_{\mathcal{H}}^*)$ . Des équations (6.3) et (6.4) on déduit que  $C_{\mathcal{H}} = 0$ .  $\square$

### 6.1.3 Estimation et convergence du critère

Soit  $(X^{(i)}, Y^{(i)})_{1 \leq i \leq n}$  un échantillon i.i.d de  $(X, Y)$ , avec  $p \leq n$  et  $q \leq n$ . On a :

$$X^{(i)} = \begin{pmatrix} \frac{X_1^{(i)}}{X_2^{(i)}} \\ \vdots \\ X_p^{(i)} \end{pmatrix},$$

d'où :

$$\bar{X}^{(n)} = \frac{1}{n} \sum_{i=1}^n X^{(i)} = \begin{pmatrix} \frac{\frac{1}{n} \sum_{i=1}^n X_1^{(i)}}{\frac{1}{n} \sum_{i=1}^n X_2^{(i)}} \\ \vdots \\ \frac{\frac{1}{n} \sum_{i=1}^n X_p^{(i)}}{\frac{1}{n} \sum_{i=1}^n X_p^{(i)}} \end{pmatrix} \quad \text{et} \quad \bar{Y}^{(n)} = \frac{1}{n} \sum_{i=1}^n Y^{(i)},$$

$$\hat{V}_{XY}^{(n)} = \frac{1}{n} \sum_{i=1}^n \left( Y^{(i)} - \bar{Y}^{(n)} \right) \left( X^{(i)} - \bar{X}^{(n)} \right)$$

et

$$\hat{V}_X^{(n)} = \frac{1}{n} \sum_{i=1}^n \left( X^{(i)} - \bar{X}^{(n)} \right) \otimes \left( X^{(i)} - \bar{X}^{(n)} \right) = \begin{pmatrix} \hat{V}_{11}^{(n)} & \hat{V}_{12}^{(n)} & \cdots & \hat{V}_{1q}^{(n)} \\ \hat{V}_{21}^{(n)} & \hat{V}_{22}^{(n)} & \cdots & \hat{V}_{2q}^{(n)} \\ \vdots & \vdots & \cdots & \vdots \\ \hat{V}_{p1}^{(n)} & \hat{V}_{p2}^{(n)} & \cdots & \hat{V}_{pq}^{(n)} \end{pmatrix}.$$

On estime alors le critère  $C_{\mathcal{H}}$  par :

$$\hat{C}_{\mathcal{H}}^{(n)} = \|\hat{V}_{XY}^{(n)} - \hat{V}_X^{(n)} \hat{\Pi}_{\mathcal{H}}^{(n)} \hat{V}_{XY}^{(n)}\|, \quad (6.6)$$

où  $\hat{\Pi}_{\mathcal{H}}^{(n)} = A_{\mathcal{H}}^* \left( A_{\mathcal{H}} \hat{V}_X^{(n)} A_{\mathcal{H}}^* \right)^{-1} A_{\mathcal{H}}$ .

Posons  $\xi_{\mathcal{H}} = V_{XY} - V_X \Pi_{\mathcal{H}} V_{XY}$  et considérons les projecteurs  $\pi_1$  (resp.  $\pi_2$ ) définis de  $\mathcal{L}(\mathbb{R}^{pq+1})$  vers  $\mathcal{L}(\mathbb{R}^{pq})$  (resp.  $\mathcal{L}(\mathbb{R}^{pq+1})$  vers  $\mathcal{L}(\mathbb{R}, \mathbb{R}^{pq})$ ).

**Théorème 14** On a :

1.  $\hat{C}_{\mathcal{H}}^{(n)}$  converge presque sûrement vers  $C_{\mathcal{H}}$  lorsque  $n \rightarrow +\infty$ .
2.  $\forall \mathcal{H} \subset \mathcal{I}$ , on a

$$\sqrt{n} \hat{C}_{\mathcal{H}}^{(n)} = \|\hat{\Psi}_{\mathcal{H}}^{(n)} \left( \hat{H}^{(n)} \right) + \sqrt{n} \xi_{\mathcal{H}}\|,$$

## 6.1 Modèle et critère de sélection

où  $(\widehat{\Psi}_{\mathcal{H}}^{(n)})_{n \in \mathbb{N}^*}$  est une suite de variables aléatoires à valeurs dans  $\mathcal{L}(\mathcal{L}(\mathbb{R}^{pq+1}), \mathbb{R}^{pq})$  qui converge presque sûrement vers l'opérateur  $\Psi_{\mathcal{H}}$  à valeurs dans  $\mathcal{L}(\mathcal{L}(\mathbb{R}^{pq+1}), \mathbb{R}^{pq})$  défini par

$$\Psi_{\mathcal{H}}(T) = \pi_2(T) - \pi_1(T)\Pi_{\mathcal{H}}V_{XY} + V_X\Psi_{\mathcal{H}}\pi_1(T)\Psi_{\mathcal{H}}V_{XY} + V_{XY} - V_X\Psi_{\mathcal{H}}\pi_2(T),$$

et  $(\widehat{H}^{(n)})_{n \in \mathbb{N}^*}$  est une suite de variables aléatoires à valeurs dans  $\mathcal{L}(\mathbb{R}^{pq+1})$  qui converge en loi vers une variable aléatoire  $H$  de loi normale centrée et d'opérateur de covariance

$$\Gamma = \mathbb{E} \left( \left( Z^{\otimes 2} - \mathbb{E}(Z^{\otimes 2}) \right)^{\otimes 2} \right),$$

où

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \frac{X_1}{X_2} \\ \vdots \\ \frac{X_p}{Y} \end{pmatrix} \in \mathbb{R}^{pq+1},$$

et  $\otimes$  est le produit tensoriel dans  $\mathcal{L}(\mathbb{R}^{pq+1})$ .

**Preuve:**

1. Application directe de la loi forte des grands nombres (voir Chapitre 2).
2. On a :

$$\begin{aligned} \sqrt{n}\widehat{C}_{\mathcal{H}}^{(n)} &= \left\| \sqrt{n}(\widehat{V}_{XY}^{(n)} - V_{XY}) - \sqrt{n}(\widehat{V}_X^{(n)} - V_X)\widehat{\Pi}_{\mathcal{H}}^{(n)}\widehat{V}_{XY}^{(n)} - V_X \left( \sqrt{n}(\widehat{\Pi}_{\mathcal{H}}^{(n)} - \Pi_{\mathcal{H}}) \right) \widehat{V}_{XY}^{(n)} \right. \\ &\quad \left. - V_X\Pi_{\mathcal{H}} \left( \sqrt{n}(\widehat{V}_{XY}^{(n)} - V_{XY}) \right) + \sqrt{n}(V_{XY} - V_X\Pi_{\mathcal{H}}V_{XY}) \right\|, \end{aligned}$$

or

$$\begin{aligned} \widehat{\Pi}_{\mathcal{H}}^{(n)} - \Pi_{\mathcal{H}} &= A_{\mathcal{H}}^* \left( (A_{\mathcal{H}}\widehat{V}_X^{(n)}A_{\mathcal{H}}^*)^{-1} - (A_{\mathcal{H}}V_XA_{\mathcal{H}}^*)^{-1} \right) A_{\mathcal{H}} \\ &= A_{\mathcal{H}}^* \left( -(A_{\mathcal{H}}\widehat{V}_X^{(n)}A_{\mathcal{H}}^*)^{-1} (A_{\mathcal{H}}\widehat{V}_X^{(n)}A_{\mathcal{H}}^* - A_{\mathcal{H}}V_XA_{\mathcal{H}}^*) (A_{\mathcal{H}}V_XA_{\mathcal{H}}^*)^{-1} \right) A_{\mathcal{H}} \\ &= -\widehat{\Pi}_{\mathcal{H}}^{(n)} \left( \widehat{V}_X^{(n)} - V_X \right) \Pi_{\mathcal{H}}. \end{aligned}$$

D'où

$$\begin{aligned} \sqrt{n}\widehat{C}_{\mathcal{H}}^{(n)} &= \|\sqrt{n}(\widehat{V}_{XY}^{(n)} - V_{XY}) - \sqrt{n}(\widehat{V}_X^{(n)} - V_X)\widehat{\Pi}_{\mathcal{H}}^{(n)}\widehat{V}_{XY}^{(n)} \\ &\quad + V_X\widehat{\Pi}_{\mathcal{H}}^{(n)}\left(\sqrt{n}(\widehat{V}_X^{(n)} - V_X)\right)\Pi_{\mathcal{H}}\widehat{V}_{XY}^{(n)} \\ &\quad - V_X\Pi_{\mathcal{H}}\left(\sqrt{n}(\widehat{V}_{XY}^{(n)} - V_{XY})\right) + \sqrt{n}\xi_{\mathcal{H}}\|. \end{aligned} \quad (6.7)$$

où  $\xi_{\mathcal{H}} = V_{XY} - V_X\Pi_{\mathcal{H}}V_{XY}$ .

Considérons la variable aléatoire  $Z$  à valeurs dans  $\mathbb{R}^{pq+1}$  telles que :

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \frac{X_1}{X_2} \\ \vdots \\ \frac{X_p}{Y} \end{pmatrix} \text{ et } Z^{(k)} = \begin{pmatrix} X^{(k)} \\ Y^{(k)} \end{pmatrix}_{1 \leq k \leq n}.$$

Notons par  $V = \mathbb{E}(Z \otimes Z)$  l'opérateur de covariance de la variable  $Z$  tel que

$$V = \begin{pmatrix} V_X & V_{XY} \\ V_{YX} & V_Y \end{pmatrix}, \quad (6.8)$$

où  $V_Y = \mathbb{E}(YY)$  et  $V_{XY} = V_{YX}^*$ . On a

$$\bar{Z}^{(n)} = \frac{1}{n} \sum_{k=1}^n Z^{(k)} \text{ et } \widehat{V}^{(n)} = \frac{1}{n} \sum_{k=1}^n \left( Z^{(k)} - \bar{Z}^{(n)} \right) \otimes \left( Z^{(k)} - \bar{Z}^{(n)} \right)$$

ou encore

$$\widehat{V}^{(n)} = \begin{pmatrix} \widehat{V}_X^{(n)} & \widehat{V}_{XY}^{(n)} \\ \widehat{V}_{YX}^{(n)} & \widehat{V}_Y^{(n)} \end{pmatrix}, \quad (6.9)$$

avec  $\widehat{V}_Y^{(n)} = \frac{1}{n} \sum_{k=1}^n \left( Y^{(k)} - \bar{Y}^{(n)} \right) \left( Y^{(k)} - \bar{Y}^{(n)} \right)$  et  $\widehat{V}_{XY}^{(n)} = \widehat{V}_{YX}^{(n)*}$ .

En considérant les projecteurs  $\pi_1$  (resp.  $\pi_2$ ) définis de  $\mathcal{L}(\mathbb{R}^{pq+1})$  vers  $\mathcal{L}(\mathbb{R}^{pq})$  (resp.  $\mathcal{L}(\mathbb{R}^{pq+1})$  vers  $\mathcal{L}(\mathbb{R}, \mathbb{R}^{pq})$ ) tels que  $\pi_1(V) = V_X$  et  $\pi_2(V) = V_{XY}$ , puis des équations (6.7) - (6.9) on déduit que

$$\sqrt{n}\widehat{C}_{\mathcal{H}}^{(n)} = \|\widehat{\Psi}_{\mathcal{H}}^{(n)}\left(\widehat{H}^{(n)}\right) + \sqrt{n}\xi_{\mathcal{H}}\|,$$

## 6.1 Modèle et critère de sélection

où  $\widehat{H}^{(n)} = \sqrt{n} (\widehat{V}^{(n)} - V)$  et  $\widehat{\Psi}_{\mathcal{X}}^{(n)}$  est un opérateur défini dans  $\mathcal{L}(\mathcal{L}(\mathbb{R}^{pq+1}), \mathbb{R}^{pq})$  tel que pour tout  $T \in \mathcal{L}(\mathcal{L}(\mathbb{R}^{pq+1}), \mathbb{R}^{pq})$ ,

$$\widehat{\Psi}_{\mathcal{X}}^{(n)}(T) = \pi_2(T) - \pi_1(T)\widehat{\Pi}_{\mathcal{X}}^{(n)}\widehat{V}_{XY}^{(n)} + \widehat{V}_X^{(n)}\widehat{\Psi}_{\mathcal{X}}^{(n)}\pi_1(T)\widehat{\Psi}_{\mathcal{X}}^{(n)}\widehat{V}_{XY}^{(n)} + V_{XY} - V_X\Psi_{\mathcal{X}}\pi_2(T).$$

Soit  $\|\cdot\|_{\infty}$ , la norme usuelle d'opérateurs définie dans  $\mathcal{L}(E, F)$  par

$$\|T\|_{\infty} = \sup_{x \in E - \{0\}} \|Tx\|_F / \|x\|_E,$$

et pour tout opérateur  $A$  et  $B$ ,  $\|AB\|_{\infty} \leq \|A\|_{\infty}\|B\|_{\infty}$ , on a ainsi

$$\begin{aligned} \|\widehat{\Psi}_{\mathcal{X}}^{(n)}(T) - \Psi_{\mathcal{X}}(T)\|_{\infty} &= \left\| -\pi_1(T) \left( \widehat{\Pi}_{\mathcal{X}}^{(n)} - \Pi_{\mathcal{X}} \right) \widehat{V}_{XY}^{(n)} - \pi_1(T)\Pi_{\mathcal{X}} \left( \widehat{V}_{XY}^{(n)} - V_{XY} \right) \right. \\ &\quad \left. + V_X \left( \widehat{\Pi}_{\mathcal{X}}^{(n)} - \Pi_{\mathcal{X}} \right) \pi_1(T)\Pi_{\mathcal{X}}\widehat{V}_{XY}^{(n)} + V_X\Pi_{\mathcal{X}}\pi_1(T)\Pi_{\mathcal{X}} \left( \widehat{V}_{XY}^{(n)} - V_{XY} \right) \right\|_{\infty}. \end{aligned}$$

Or,  $\|B\|_{\infty, \infty} := \sup_{T \in \mathcal{L}(\mathbb{R}^{p+q}) - \{0\}} \|B(T)\|_{\infty} / \|T\|_{\infty}$ . D'où

$$\begin{aligned} \|\widehat{\Psi}_{\mathcal{X}}^{(n)}(T) - \Psi_{\mathcal{X}}(T)\|_{\infty} &\leq \|\pi_1(T)\|_{\infty} \left[ \|\widehat{\Pi}_{\mathcal{X}}^{(n)} - \Pi_{\mathcal{X}}\|_{\infty} \|\widehat{V}_{XY}^{(n)}\|_{\infty} + \|\Pi_{\mathcal{X}}\|_{\infty} \|\widehat{V}_{XY}^{(n)} - V_{XY}\|_{\infty} \right. \\ &\quad \left. + \|V_X\|_{\infty} \|\Pi_{\mathcal{X}}\|_{\infty} \|\widehat{\Pi}_{\mathcal{X}}^{(n)} - \Pi_{\mathcal{X}}\|_{\infty} \|\widehat{V}_{XY}^{(n)}\|_{\infty} \right. \\ &\quad \left. + \|V_X\|_{\infty} \|\Pi_{\mathcal{X}}\|_{\infty}^2 \|\widehat{V}_{XY}^{(n)} - V_{XY}\|_{\infty} \right]. \end{aligned}$$

Alors

$$\begin{aligned} \|\widehat{\Psi}_{\mathcal{X}}^{(n)} - \Psi_{\mathcal{X}}\|_{\infty} &\leq \left[ \|\widehat{\Pi}_{\mathcal{X}}^{(n)} - \Pi_{\mathcal{X}}\|_{\infty} \|\widehat{V}_{XY}^{(n)}\|_{\infty} + \|\Pi_{\mathcal{X}}\|_{\infty} \|\widehat{V}_{XY}^{(n)} - V_{XY}\|_{\infty} \right. \\ &\quad \left. + \|V_X\|_{\infty} \|\Pi_{\mathcal{X}}\|_{\infty} \|\widehat{\Pi}_{\mathcal{X}}^{(n)} - \Pi_{\mathcal{X}}\|_{\infty} \|\widehat{V}_{XY}^{(n)}\|_{\infty} \right. \\ &\quad \left. + \|V_X\|_{\infty} \|\Pi_{\mathcal{X}}\|_{\infty}^2 \|\widehat{V}_{XY}^{(n)} - V_{XY}\|_{\infty} \right] \|\pi_1\|_{\infty, \infty} \|T\|_{\infty}. \end{aligned} \quad (6.10)$$

$\widehat{V}_X^{(n)}$  (resp.  $\widehat{V}_{XY}^{(n)}$ ) converge presque sûrement vers  $V_X$  (resp. vers  $V_{XY}$ ) quand  $n \rightarrow +\infty$ . Alors  $\widehat{\Pi}_{\mathcal{X}}^{(n)}$  converge presque sûrement vers  $\Pi_{\mathcal{X}}$  quand  $n \rightarrow +\infty$ . De l'équation (6.10), on déduit que  $\widehat{\Psi}_{\mathcal{X}}^{(n)}$  converge presque sûrement vers  $\Psi_{\mathcal{X}}$  quand  $n \rightarrow +\infty$ .

De l'expression de  $\widehat{H}^{(n)}$ , on a :

$$\widehat{H}_1^{(n)} = \sqrt{n} \left( \frac{1}{n} \sum_{k=1}^n Z^{(k)} \otimes Z^{(k)} - V \right) \quad \text{et} \quad \widehat{H}_2^{(n)} = \frac{1}{\sqrt{n}} \left( (\sqrt{n} \overline{Z}^{(n)}) \otimes (\sqrt{n} \overline{Z}^{(n)}) \right).$$

Le théorème limite central montre que  $\widehat{H}_1^{(n)}$  (resp.  $\sqrt{n} \overline{Z}^{(n)}$ ) converge en loi vers la variable

aléatoire  $H$  (resp.  $U$ ), quand  $n \rightarrow +\infty$ , de loi normale centrée et de d'opérateur de covariance  $\Gamma$  (resp.  $\Gamma'$ ) tel que

$$\Gamma = \mathbb{E} \left( (Z \otimes Z - V) \tilde{\otimes} (Z \otimes Z - V) \right) \quad (\text{resp. } \Gamma' = \mathbb{E} (Z \otimes Z) ).$$

Par conséquent,  $\widehat{H}_2^{(n)}$  converge en probabilité vers 0 quand  $n \rightarrow +\infty$ , et le théorème de Slutsky permet de conclure que  $\widehat{H}^{(n)}$  converge en loi vers  $H$  quand  $n \rightarrow +\infty$   $\square$

### 6.1.4 Sélection des variables

On se propose maintenant d'estimer le sous-ensemble  $\mathcal{I}_1$ . Le Lemme 10 montre que pour  $i \in \mathcal{I}$ ,  $C_{\mathcal{K}_i} = 0 \Leftrightarrow i \notin \mathcal{I}_1$ , où  $\mathcal{K}_i = \mathcal{I} - \{i\}$ . En ordonnant les vecteurs  $X_k$  à travers le critère  $C_{\mathcal{K}_i}$  dans l'ordre décroissant, cela nous amène à considérer la permutation aléatoire  $\sigma$  de  $\mathcal{I}$  telle que

$$C_{\mathcal{K}_{\sigma(1)}} \geq C_{\mathcal{K}_{\sigma(2)}} \geq \dots \geq C_{\mathcal{K}_{\sigma(p)}}.$$

Si

$$C_{\mathcal{K}_{\sigma(i)}} = C_{\mathcal{K}_{\sigma(j)}}, \quad \text{et si } i < j, \quad \text{alors } \sigma(i) < \sigma(j).$$

Comme le sous-ensemble  $\mathcal{I}_1$  est supposé non vide, il va exister un entier  $q_0 \in \mathcal{I}$  égal à  $p$  lorsque  $\mathcal{I}_1 = \mathcal{I}$  (ce qui implique que  $C_{\mathcal{K}_i} > 0, \forall i \in \mathcal{I}$ ) et

$$C_{\mathcal{K}_{\sigma(1)}} \geq C_{\mathcal{K}_{\sigma(2)}} \geq \dots \geq C_{\mathcal{K}_{\sigma(q_0)}} > C_{\mathcal{K}_{\sigma(q_0+1)}} = \dots = C_{\mathcal{K}_{\sigma(p)}} = 0, \quad \text{si } \mathcal{I}_1 \neq \mathcal{I}$$

On déduit alors que

$$\mathcal{I}_1 = \{\sigma(k); 1 \leq k \leq q_0\}.$$

**Théorème 15** On a :

(i)  $\lim_{n \rightarrow +\infty} P(\widehat{\sigma}^{(n)} = \sigma) = 1;$

(ii)  $\widehat{q}^{(n)}$  converge en probabilité vers  $q$ , quand  $n \rightarrow +\infty$ .

Partant du Lemme 10 et des Théorèmes 14 et 15, la sélection des vecteurs (l'estimation et les propriétés de convergence des paramètres  $\sigma$ ,  $q_0$  et du sous-ensemble  $\mathcal{I}_1$ ) reste identique

## 6.1 Modèle et critère de sélection

à celle établie aux Chapitres 2 et 4. La preuve du Théorème 15 repose sur le Théorème 16 qui suit :

Soit  $r \in \mathbb{N}^*$  et  $(m_1, \dots, m_r) \in (\mathbb{N}^*)^r$  tels que  $\sum_{\ell=1}^r m_\ell = p$  et

$$C_{\mathcal{H}_{\sigma(1)}} = \dots = C_{\mathcal{H}_{\sigma(m_1)}} > C_{\mathcal{H}_{\sigma(m_1+1)}} = \dots = C_{\mathcal{H}_{\sigma(m_1+m_2)}} > \dots > \\ C_{\mathcal{H}_{\sigma(m_1+m_2+\dots+m_{r-1}+1)}} = \dots = C_{\mathcal{H}_{\sigma(m_1+m_2+\dots+m_r)}}.$$

Posons :

$$E = \{\ell \in \mathbb{N}^* \mid 1 \leq \ell \leq r, m_\ell \geq 2\}, \text{ et } F_\ell = \left\{ \left( \sum_{k=0}^{\ell-1} m_k \right) + 1, \dots, \left( \sum_{k=0}^{\ell} m_k \right) - 1 \right\},$$

avec  $m_0 = 0$ . On a :

**Théorème 16** Si  $E \neq \emptyset$ , alors  $\forall \ell \in E$  et  $i \in F_\ell$ ,  $n^\alpha \left( \widehat{C}_{\mathcal{H}_{\sigma(i)}}^{(n)} - \widehat{C}_{\mathcal{H}_{\sigma(i+1)}}^{(n)} \right)$  converge en probabilité vers 0 quand  $n \rightarrow +\infty$ .

**Preuve:** Posons  $\lambda_\ell = \|\xi_{\mathcal{H}_{\sigma(i)}}\| = \|\xi_{\mathcal{H}_{\sigma(i+1)}}\|$ . Si  $\lambda_\ell = 0$ , alors

$$\begin{aligned} \left| n^\alpha \left( \widehat{C}_{\mathcal{H}_{\sigma(i)}}^{(n)} - \widehat{C}_{\mathcal{H}_{\sigma(i+1)}}^{(n)} \right) \right| &= n^\alpha \left| n^{-\frac{1}{2}} \left( \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)})\| - \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)})\| \right) \right| \\ &= n^{\alpha-\frac{1}{2}} \left| \left( \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)})\| - \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)})\| \right) \right| \\ &\leq n^{\alpha-\frac{1}{2}} \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) - \widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)})\| \\ &\leq n^{\alpha-\frac{1}{2}} \left\| \left( \widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)} - \widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)} \right) (\widehat{H}^{(n)}) \right\| \\ &\leq n^{\alpha-\frac{1}{2}} \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)} - \widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}\|_\infty \|\widehat{H}^{(n)}\|. \end{aligned}$$

Puisque  $\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}$  et  $\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}$  convergent presque sûrement respectivement vers  $\Psi_{\mathcal{H}_{\sigma(i)}}$  et  $\Psi_{\mathcal{H}_{\sigma(i+1)}}$  quand  $n \rightarrow +\infty$ ,  $\widehat{H}^{(n)}$  converge en loi vers  $H$  quand  $n \rightarrow +\infty$  (Théorème 14), et que  $\|\cdot\|_\infty$  et  $\|\cdot\|$  sont continues, il vient que  $\|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)} - \widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}\|_\infty \|\widehat{H}^{(n)}\|$  converge en loi vers  $\|\Psi_{\mathcal{H}_{\sigma(i)}} - \Psi_{\mathcal{H}_{\sigma(i+1)}}\|_\infty \|H\|$  quand  $n \rightarrow +\infty$ . Comme  $\alpha - \frac{1}{2} < 0$ , alors on déduit que  $n^\alpha \left( \widehat{C}_{\mathcal{H}_{\sigma(i)}}^{(n)} - \widehat{C}_{\mathcal{H}_{\sigma(i+1)}}^{(n)} \right)$  converge en probabilité vers 0 quand  $n \rightarrow +\infty$ .

Si  $\lambda_\ell \neq 0$ , on a :

$$n^\alpha \left( \widehat{C}_{\mathcal{H}_{\sigma(i)}}^{(n)} - \widehat{C}_{\mathcal{H}_{\sigma(i+1)}}^{(n)} \right)$$



## 6.1 Modèle et critère de sélection

$$\begin{aligned}
&= n^{\alpha-\frac{1}{2}} \left( \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\xi_{\mathcal{H}_{\sigma(i)}}\| - \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\xi_{\mathcal{H}_{\sigma(i+1)}}\| \right) \\
&= \frac{n^{\alpha-\frac{1}{2}} \left( \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\xi_{\mathcal{H}_{\sigma(i)}}\|^2 - \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\xi_{\mathcal{H}_{\sigma(i+1)}}\|^2 \right)}{\|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\xi_{\mathcal{H}_{\sigma(i)}}\| + \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\xi_{\mathcal{H}_{\sigma(i+1)}}\|} \\
&= \frac{n^{\alpha-\frac{1}{2}} \left( \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)})\|^2 + n\|\xi_{\mathcal{H}_{\sigma(i)}}\|^2 - \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)})\|^2 - n\|\xi_{\mathcal{H}_{\sigma(i+1)}}\|^2 \right)}{\|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\xi_{\mathcal{H}_{\sigma(i)}}\| + \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\xi_{\mathcal{H}_{\sigma(i+1)}}\|} \\
&+ \frac{2n^\alpha \left( \langle \xi_{\mathcal{H}_{\sigma(i)}}, \widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) \rangle - \langle \xi_{\mathcal{H}_{\sigma(i+1)}}, \widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) \rangle \right)}{\|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\xi_{\mathcal{H}_{\sigma(i)}}\| + \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\xi_{\mathcal{H}_{\sigma(i+1)}}\|} \\
&= \frac{n^{\alpha-\frac{1}{2}} \left( \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)})\|^2 - \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)})\|^2 \right)}{\|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\xi_{\mathcal{H}_{\sigma(i)}}\| + \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\xi_{\mathcal{H}_{\sigma(i+1)}}\|} \\
&+ \frac{2n^\alpha \left( \langle \xi_{\mathcal{H}_{\sigma(i)}}, \widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) \rangle - \langle \xi_{\mathcal{H}_{\sigma(i+1)}}, \widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) \rangle \right)}{\|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\xi_{\mathcal{H}_{\sigma(i)}}\| + \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\xi_{\mathcal{H}_{\sigma(i+1)}}\|} \\
&= \frac{n^{\alpha-1} \left( \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)})\|^2 - \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)})\|^2 \right)}{\|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\xi_{\mathcal{H}_{\sigma(i)}}\| + \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) + \sqrt{n}\xi_{\mathcal{H}_{\sigma(i+1)}}\|} \\
&+ \frac{2n^{\alpha-\frac{1}{2}} \left( \langle \xi_{\mathcal{H}_{\sigma(i)}}, \widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) \rangle - \langle \xi_{\mathcal{H}_{\sigma(i+1)}}, \widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) \rangle \right)}{\|n^{-\frac{1}{2}}\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) + \xi_{\mathcal{H}_{\sigma(i)}}\| + \|n^{-\frac{1}{2}}\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) + \xi_{\mathcal{H}_{\sigma(i+1)}}\|},
\end{aligned}$$

où  $\langle \cdot, \cdot \rangle$  désigne le produit scalaire d'opérateurs tel que  $\langle A, B \rangle = \text{tr}(A^*B)$ . Ainsi, on a d'une part :

$$\begin{aligned}
&\left| n^{\alpha-1} \left( \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)})\|^2 - \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)})\|^2 \right) \right| \\
&\leq n^{\alpha-1} \left( \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)})\|^2 + \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)})\|^2 \right) \\
&\leq n^{\alpha-1} \left( \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}\|_\infty^2 \|\widehat{H}^{(n)}\|^2 + \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}\|_\infty^2 \|\widehat{H}^{(n)}\|^2 \right) \\
&\leq n^{\alpha-1} \left( \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}\|_\infty^2 + \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}\|_\infty^2 \right) \|\widehat{H}^{(n)}\|^2,
\end{aligned} \tag{6.11}$$

et d'autre part :

$$\begin{aligned}
&\left| 2n^{\alpha-\frac{1}{2}} \left( \langle \xi_{\mathcal{H}_{\sigma(i)}}, \widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) \rangle - \langle \xi_{\mathcal{H}_{\sigma(i+1)}}, \widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) \rangle \right) \right| \\
&\leq 2n^{\alpha-\frac{1}{2}} \left( \left| \langle \xi_{\mathcal{H}_{\sigma(i)}}, \widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)}) \rangle \right| + \left| \langle \xi_{\mathcal{H}_{\sigma(i+1)}}, \widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)}) \rangle \right| \right) \\
&\leq 2n^{\alpha-\frac{1}{2}} \left( \|\xi_{\mathcal{H}_{\sigma(i)}}\| \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}(\widehat{H}^{(n)})\| + \|\xi_{\mathcal{H}_{\sigma(i+1)}}\| \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}(\widehat{H}^{(n)})\| \right) \\
&\leq 2n^{\alpha-\frac{1}{2}} \left( \|\xi_{\mathcal{H}_{\sigma(i)}}\| \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}\|_\infty \|\widehat{H}^{(n)}\| + \|\xi_{\mathcal{H}_{\sigma(i+1)}}\| \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}\|_\infty \|\widehat{H}^{(n)}\| \right) \\
&\leq 2n^{\alpha-\frac{1}{2}} \|\xi_{\mathcal{H}_{\sigma(i)}}\| \left( \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i)}}^{(n)}\|_\infty + \|\widehat{\Psi}_{\mathcal{H}_{\sigma(i+1)}}^{(n)}\|_\infty \right) \|\widehat{H}^{(n)}\|.
\end{aligned} \tag{6.12}$$

---

## 6.2 Sélection des variables dans un modèle linéaire fonctionnel additif

---

Des équations (6.11) et (6.12) et des propriétés de convergence des estimateurs  $\widehat{\Psi}_{\mathcal{H}_{\sigma^{(i)}}}^{(n)}$ ,  $\widehat{\Psi}_{\mathcal{H}_{\sigma^{(i+1)}}}^{(n)}$  et  $\widehat{H}^{(n)}$ , on déduit que  $n^\alpha \left( \widehat{C}_{\mathcal{H}_{\sigma^{(i)}}}^{(n)} - \widehat{C}_{\mathcal{H}_{\sigma^{(i+1)}}}^{(n)} \right)$  converge en probabilité vers 0 quand  $n \rightarrow +\infty$ .  $\square$

## 6.2 Sélection des variables dans un modèle linéaire fonctionnel additif

Dans cette section, nous proposons un critère de sélection pour les variables fonctionnelles. Considérons le modèle linéaire fonctionnel additif de la forme :

$$Y = \sum_{k=1}^p \int_{\mathcal{C}_k} b_k(t) X_k(t) dt + \varepsilon = \sum_{k=1}^p \langle b_k, X_k \rangle_{L^2(\mathcal{C}_k)} + \varepsilon, \quad (6.13)$$

où  $\forall (k = 1, \dots, p)$ ,  $b_k$  est une constante fonctionnelle de carré intégrable et  $X_k$  un processus continu défini sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$  et indépendant de  $\varepsilon$ , ou  $Y$  et  $\varepsilon$  sont des variables aléatoires réelles.

On s'intéresse ainsi dans ce modèle (équation (6.13)) au problème de sélection des variables fonctionnelles  $X_k(t)$  ( $t \in \mathcal{C}_k$ ). Notons que notre méthode de sélection n'est pour le moment valable que dans le cas des modèles linéaires, à savoir modèles linéaires unidimensionnels (*Nkiet* (2001)), modèles linéaires multidimensionnels Chapitre 2 et dans les cas de l'analyse discriminante mixte Chapitres 4. Ainsi, nous proposons dans la section qui suit, un critère de sélection des variables dans les modèles linéaires fonctionnels additifs, sur la base des critères développés dans les Chapitre 2 et 4.

### 6.2.1 Approximation du modèle linéaire fonctionnel additif : critère de sélection pour les variables fonctionnelles

Considérons le modèle de l'équation (6.1). Soit  $(\phi_{k,\ell})_\ell$  une base de l'espace  $L^2(\mathcal{C}_\ell)$ ,  $\forall k = 1, \dots, p$ , où  $\mathcal{C}_\ell$  est un compact de  $\mathbb{R}$ . Dans cette base, les fonctions  $b_k$  et  $X_k$  peuvent

---

## 6.2 Sélection des variables dans un modèle linéaire fonctionnel additif

---

s'écrire :

$$b_k(t) = \sum_{\ell=1}^{+\infty} \eta_{k\ell} \phi_{k,\ell}(t) \quad \text{et} \quad X_k(t) = \sum_{\ell=1}^{+\infty} X_{k\ell} \phi_{k,\ell}(t), \forall k = 1, \dots, p.$$

Considérons maintenant une suite croissante d'entiers positifs  $(q_n)_{n \in \mathbb{N}^*}$  telle que  $\lim_{n \rightarrow +\infty} q_n = +\infty$ , avec  $q_n \geq n$ . On réécrit les variables fonctionnelles  $b_k$  et  $X_k$  en tronquant la base  $(\phi_{k,\ell})_\ell$  aux  $q_n$  premières composantes :

$$\tilde{b}_k(t) = \sum_{\ell=1}^{q_n} \eta_{k\ell} \phi_{k,\ell}(t) \quad \text{et} \quad \tilde{X}_k(t) = \sum_{\ell=1}^{q_n} X_{k\ell} \phi_{k,\ell}(t), \forall k = 1, \dots, p.$$

d'où il vient que

$$\langle \tilde{b}_k, \tilde{X}_k \rangle_{L^2(\mathcal{C}_\ell)} = \sum_{\ell=1}^{q_n} \eta_{k\ell} X_{k\ell}.$$

On approche ainsi le modèle (6.13) par :

$$Y \approx \sum_{k=1}^p \sum_{\ell=1}^{q_n} \eta_{k\ell} X_{k\ell} + \varepsilon = \sum_{k=1}^p \eta_k^T X_k + \varepsilon, \quad (6.14)$$

où

$$\eta_k = \begin{pmatrix} \eta_{k1} \\ \eta_{k2} \\ \vdots \\ \eta_{kq_n} \end{pmatrix} \quad \text{et} \quad X_k = \begin{pmatrix} X_{k1} \\ X_{k2} \\ \vdots \\ X_{kq_n} \end{pmatrix}.$$

Le modèle (6.14) ainsi obtenu est semblable au modèle (6.1). Nous donnons maintenant dans la suite, une expression du critère (Equation 6.2) dans le cas fonctionnel.

Pour tout  $k = 1, \dots, p$  et  $\ell = 1, \dots, q_n$ , on exprime les coefficients  $X_{k\ell}$  et  $\eta_{k\ell}$  de la façon suivante :

$$X_{k\ell} = \langle \tilde{X}_k, \phi_{k,\ell} \rangle_{L^2(\mathcal{C}_\ell)} \quad \text{et} \quad \eta_{k\ell} = \langle \tilde{b}_k, \phi_{k,\ell} \rangle_{L^2(\mathcal{C}_\ell)},$$

où  $\phi_{k,\ell}$  est une base quelconque de  $L^2(\mathcal{C}_\ell)$ ,  $\forall k=1, \dots, p$ . Ainsi,

$$X_{k\ell} = \langle \tilde{X}_k, \phi_{k,\ell} \rangle_{L^2(\mathcal{C}_\ell)} = \int_{\mathcal{C}_\ell} \tilde{X}_k(t) \phi_{k,\ell}(t) dt \quad \text{et} \quad \eta_{k\ell} = \langle \tilde{b}_k, \phi_{k,\ell} \rangle_{L^2(\mathcal{C}_\ell)} = \int_{\mathcal{C}_\ell} \tilde{b}_k(t) \phi_{k,\ell}(t) dt. \quad (6.15)$$

On calcule les intégrales (6.15) par la méthode des trapèzes. Pour cela, on discrétise l'intervalle  $\mathcal{C}_\ell$  en  $N$  points de pas constant  $(t_1, \dots, t_N)$ . Ainsi, pour tout  $k = 1, \dots, p$  et  $\ell = 1, \dots, q_n$ , on a :

## 6.2 Sélection des variables dans un modèle linéaire fonctionnel additif

$$X_{k\ell} = \frac{(t_N - t_1)}{2N} \left( \tilde{X}_k(t_1)\phi_{k,\ell}(t_1) + \tilde{X}_k(t_N)\phi_{k,\ell}(t_N) + 2 \sum_{j=2}^{N-1} \tilde{X}_k(t_j)\phi_{k,\ell}(t_j) \right).$$

Il vient que

$$X_k = \frac{(t_N - t_1)}{2N} \begin{pmatrix} \tilde{X}_k(t_1)\phi_{k,1}(t_1) + \tilde{X}_k(t_N)\phi_{k,1}(t_N) + 2 \sum_{j=2}^{N-1} \tilde{X}_k(t_j)\phi_{k,1}(t_j) \\ \tilde{X}_k(t_1)\phi_{k,2}(t_1) + \tilde{X}_k(t_N)\phi_{k,2}(t_N) + 2 \sum_{j=2}^{N-1} \tilde{X}_k(t_j)\phi_{k,2}(t_j) \\ \vdots \\ \tilde{X}_k(t_1)\phi_{k,q_n}(t_1) + \tilde{X}_k(t_N)\phi_{k,q_n}(t_N) + 2 \sum_{j=2}^{N-1} \tilde{X}_k(t_j)\phi_{k,q_n}(t_j) \end{pmatrix}.$$

De même

$$\eta_{k\ell} = \frac{(t_N - t_1)}{2N} \left( \tilde{b}_k(t_1)\phi_{k,\ell}(t_1) + \tilde{b}_k(t_N)\phi_{k,\ell}(t_N) + 2 \sum_{j=2}^{N-1} \tilde{b}_k(t_j)\phi_{k,\ell}(t_j) \right),$$

alors

$$\eta_k = \frac{(t_N - t_1)}{2N} \begin{pmatrix} \tilde{b}_k(t_1)\phi_{k,1}(t_1) + \tilde{b}_k(t_N)\phi_{k,1}(t_N) + 2 \sum_{j=2}^{N-1} \tilde{b}_k(t_j)\phi_{k,1}(t_j) \\ \tilde{b}_k(t_1)\phi_{k,2}(t_1) + \tilde{b}_k(t_N)\phi_{k,2}(t_N) + 2 \sum_{j=2}^{N-1} \tilde{b}_k(t_j)\phi_{k,2}(t_j) \\ \vdots \\ \tilde{b}_k(t_1)\phi_{k,q_n}(t_1) + \tilde{b}_k(t_N)\phi_{k,q_n}(t_N) + 2 \sum_{j=2}^{N-1} \tilde{b}_k(t_j)\phi_{k,q_n}(t_j) \end{pmatrix}.$$

Ainsi, pour tout  $i = 1, \dots, n$  :

$$X_k^{(i)} = \frac{(t_N - t_1)}{2N} \begin{pmatrix} \tilde{X}_k^{(i)}(t_1)\phi_{k,1}(t_1) + \tilde{X}_k^{(i)}(t_N)\phi_{k,1}(t_N) + 2 \sum_{j=2}^{N-1} \tilde{X}_k^{(i)}(t_j)\phi_{k,1}(t_j) \\ \tilde{X}_k^{(i)}(t_1)\phi_{k,2}(t_1) + \tilde{X}_k^{(i)}(t_N)\phi_{k,2}(t_N) + 2 \sum_{j=2}^{N-1} \tilde{X}_k^{(i)}(t_j)\phi_{k,2}(t_j) \\ \vdots \\ \tilde{X}_k^{(i)}(t_1)\phi_{k,q_n}(t_1) + \tilde{X}_k^{(i)}(t_N)\phi_{k,q_n}(t_N) + 2 \sum_{j=2}^{N-1} \tilde{X}_k^{(i)}(t_j)\phi_{k,q_n}(t_j) \end{pmatrix}.$$

Ainsi

$$\bar{X}_k^{(n)} = \frac{(t_N - t_1)}{2N} \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} \tilde{X}_k^{(i)}(t_1)\phi_{k,1}(t_1) + \tilde{X}_k^{(i)}(t_N)\phi_{k,1}(t_N) + 2 \sum_{j=2}^{N-1} \tilde{X}_k^{(i)}(t_j)\phi_{k,1}(t_j) \\ \tilde{X}_k^{(i)}(t_1)\phi_{k,2}(t_1) + \tilde{X}_k^{(i)}(t_N)\phi_{k,2}(t_N) + 2 \sum_{j=2}^{N-1} \tilde{X}_k^{(i)}(t_j)\phi_{k,2}(t_j) \\ \vdots \\ \tilde{X}_k^{(i)}(t_1)\phi_{k,q_n}(t_1) + \tilde{X}_k^{(i)}(t_N)\phi_{k,q_n}(t_N) + 2 \sum_{j=2}^{N-1} \tilde{X}_k^{(i)}(t_j)\phi_{k,q_n}(t_j) \end{pmatrix}.$$

---

## 6.2 Sélection des variables dans un modèle linéaire fonctionnel additif

---

La composante  $k$  de la matrice  $X^{(i)} - \bar{X}^{(n)}$  est donnée par

$$\begin{aligned} (X^{(i)} - \bar{X}^{(n)})_k &= \frac{(t_N - t_1)}{2N} \times \\ &\begin{pmatrix} \left( \tilde{X}_k^{(i)}(t_1) - \tilde{\bar{X}}_k^{(n)}(t_1) \right) \phi_{k,1}(t_1) + \left( \tilde{X}_k^{(i)}(t_N) - \tilde{\bar{X}}_k^{(n)}(t_N) \right) \phi_{k,1}(t_N) + 2 \sum_{j=2}^{N-1} \left( \tilde{X}_k^{(i)}(t_j) - \tilde{\bar{X}}_k^{(n)}(t_j) \right) \phi_{k,1}(t_j) \\ \left( \tilde{X}_k^{(i)}(t_1) - \tilde{\bar{X}}_k^{(n)}(t_1) \right) \phi_{k,2}(t_1) + \left( \tilde{X}_k^{(i)}(t_N) - \tilde{\bar{X}}_k^{(n)}(t_N) \right) \phi_{k,2}(t_N) + 2 \sum_{j=2}^{N-1} \left( \tilde{X}_k^{(i)}(t_j) - \tilde{\bar{X}}_k^{(n)}(t_j) \right) \phi_{k,2}(t_j) \\ \vdots \\ \left( \tilde{X}_k^{(i)}(t_1) - \tilde{\bar{X}}_k^{(n)}(t_1) \right) \phi_{k,q_n}(t_1) + \left( \tilde{X}_k^{(i)}(t_N) - \tilde{\bar{X}}_k^{(n)}(t_N) \right) \phi_{k,q_n}(t_N) + 2 \sum_{j=2}^{N-1} \left( \tilde{X}_k^{(i)}(t_j) - \tilde{\bar{X}}_k^{(n)}(t_j) \right) \phi_{k,q_n}(t_j) \end{pmatrix}, \end{aligned}$$

où  $\forall t \in \mathcal{C}_k, \tilde{\bar{X}}_k^{(n)}(t) = \frac{1}{n} \sum_{i=1}^n \tilde{X}_k^{(i)}(t)$ . Ainsi, la composante  $k$  de l'opérateur de covariance croisé entre les variables  $X$  et  $Y$  est donnée par :

$$\begin{aligned} (\tilde{V}_{XY}^{(n)})_k &= \frac{(t_N - t_1)}{2N} \frac{1}{n} \sum_{i=1}^n (Y^{(i)} - \bar{Y}^{(n)}) \times \\ &\begin{pmatrix} \left( \tilde{X}_k^{(i)}(t_1) - \tilde{\bar{X}}_k^{(n)}(t_1) \right) \phi_{k,1}(t_1) + \left( \tilde{X}_k^{(i)}(t_N) - \tilde{\bar{X}}_k^{(n)}(t_N) \right) \phi_{k,1}(t_N) + 2 \sum_{j=2}^{N-1} \left( \tilde{X}_k^{(i)}(t_j) - \tilde{\bar{X}}_k^{(n)}(t_j) \right) \phi_{k,1}(t_j) \\ \left( \tilde{X}_k^{(i)}(t_1) - \tilde{\bar{X}}_k^{(n)}(t_1) \right) \phi_{k,2}(t_1) + \left( \tilde{X}_k^{(i)}(t_N) - \tilde{\bar{X}}_k^{(n)}(t_N) \right) \phi_{k,2}(t_N) + 2 \sum_{j=2}^{N-1} \left( \tilde{X}_k^{(i)}(t_j) - \tilde{\bar{X}}_k^{(n)}(t_j) \right) \phi_{k,2}(t_j) \\ \vdots \\ \left( \tilde{X}_k^{(i)}(t_1) - \tilde{\bar{X}}_k^{(n)}(t_1) \right) \phi_{k,q_n}(t_1) + \left( \tilde{X}_k^{(i)}(t_N) - \tilde{\bar{X}}_k^{(n)}(t_N) \right) \phi_{k,q_n}(t_N) + 2 \sum_{j=2}^{N-1} \left( \tilde{X}_k^{(i)}(t_j) - \tilde{\bar{X}}_k^{(n)}(t_j) \right) \phi_{k,q_n}(t_j) \end{pmatrix}, \end{aligned}$$

et celle de la matrice de covariance de  $\tilde{V}_X^{(n)}$  est donnée par la relation suivante :

$\forall (k, \ell') \in 1, \dots, p^2 :$

$$\tilde{V}_{k\ell'}^{(n)} = \frac{1}{n} \sum_{i=1}^n \left( X_{\ell'}^{(i)} - \bar{X}_{\ell'}^{(n)} \right) \left( X_k^{(i)} - \bar{X}_k^{(n)} \right)^T = \frac{1}{n} \sum_{i=1}^n X_{\ell'}^{(i)} X_k^{(i)T} - \bar{X}_{\ell'}^{(n)} \bar{X}_k^{(n)}.$$

Ainsi, nous remplaçons dans le critère  $\hat{C}_{\mathcal{X}}^{(n)}$  (équation (6.6)),  $\hat{V}_{XY}^{(n)}$  et  $\hat{V}_X^{(n)}$  par  $\tilde{V}_{XY}^{(n)}$  et  $\tilde{V}_X^{(n)}$ , respectivement, on obtient :

$$\tilde{C}_{\mathcal{X}}^{(n)} = \|\tilde{V}_{XY}^{(n)} - \tilde{V}_X^{(n)} \tilde{\Pi}_{\mathcal{X}}^{(n)} \tilde{V}_{XY}^{(n)}\|, \quad \text{où } \tilde{\Pi}_{\mathcal{X}}^{(n)} = A_{\mathcal{X}}^* \left( A_{\mathcal{X}} \tilde{V}_X^{(n)} A_{\mathcal{X}}^* \right)^{-1} A_{\mathcal{X}} \quad (6.16)$$

Reste maintenant dans cette partie à étudier la convergence du critère  $\tilde{C}_{\mathcal{X}}^{(n)}$ , un choix du paramètre  $q_n$  et une étude de simulations semblable à celle des Chapitres 2 et 4.

## 6.3 Conclusion

Dans ce chapitre, nous avons proposé un critère de sélection des variables dans le cas d'un modèle linéaire fonctionnel additif. Nous avons premièrement généralisé la méthode de sélection de variables de *Nkiet* (2001) en considérant le modèle (6.1). Ce modèle est une généralisation du modèle (1) de *Nkiet* (2001). La généralisation repose sur le critère (6.2) qui est une extension du critère de *Nkiet* (2001) dans le cadre du modèle (6.1). Ce critère joue le même rôle que celui des Chapitres 2 et 4, c'est-à-dire qu'il permet de ramener le problème de sélection des variables à un problème d'estimation. Nous avons ensuite approché le modèle linéaire fonctionnel additif (6.13) au modèle linéaire multivarié (6.1), en troquant la base  $(L^2(\mathcal{C}_\ell))$  aux  $q_n$  premières composantes, où  $(q_n)_{n \in \mathbb{N}}$  est une suite croissante d'entiers positifs telle que  $\lim_{n \rightarrow +\infty} q_n = +\infty$ . Reste maintenant dans cette partie à proposer une étude de convergence du critère (6.16), un choix du paramètre  $q_n$  et une étude de simulations.

# Conclusion générale et perspectives

---

# Conclusion et perspectives

---

Les résultats obtenus dans cette thèse ont porté sur la généralisation d'un critère et d'une méthode de sélection des variables. Considérant le critère de sélection introduit dans [Nkiet \(2001\)](#), nous avons proposé de nouvelles méthodes de sélection des variables en régression linéaire multidimensionnelle, en analyse discriminante mixte et dans un modèle linéaire fonctionnel additif, respectivement pour des variables aléatoires continues, des variables aléatoires mixtes et des variables aléatoires fonctionnelles. Nous avons ainsi proposé des méthodes de sélection de variables à trois modèles statistiques. Cette étude a été motivée sur un plan purement théorique par les travaux de [Nkiet \(2001\)](#) et [Nkiet \(2012\)](#). Une autre motivation, d'ordre pratique cette fois-ci, a consisté à montrer à partir d'une étude de simulation, l'influence et le choix des paramètres de contrôles utilisés dans notre méthode de sélection des variables.

## Sélection des variables en régression linéaire multidimensionnelle

Dans la première partie de cette thèse que nous avons intitulée "sélection des variables en régression linéaire pour des variables aléatoires continues", la première contribution a été donnée au Chapitre 2. Nous avons proposé une méthode de sélection des variables, pour des variables aléatoires continues sur un modèle linéaire multidimensionnel (Equation (2.1)). C'est le chapitre qui établit la première contribution (théorique et pratique) de cette thèse.



---

---

---

On se place dans le cas d'un modèle linéaire multidimensionnel. Nous avons établi un critère de sélection (équation (2.5)), adapté au modèle linéaire multidimensionnel (équation (2.1)). Il est vu comme une mesure de perte d'information lorsqu'un sous-ensemble de variables est sélectionné. On l'utilise pour ranger toutes les variables dans l'ordre décroissant de façon à ce que le problème de sélection de variable se ramène à un problème d'estimation d'une permutation aléatoire et d'un paramètre appelé dimension. À partir du critère (2.1), nous avons donné une caractérisation du sous-ensemble  $I_1$  des variables pertinentes (Lemme 1), puis la convergence en loi de notre méthode (Théorèmes 1 et 2). Le principal avantage de notre approche sur le plan théorique, vient du fait qu'elle est valable quelle que soit la loi de probabilité suivie par les variables  $X$  et  $Y$ , contrairement aux méthodes classiques qui imposent une loi de probabilité (une loi gaussienne en générale) aux variables étudiées. Nous clôturons ce chapitre avec une étude de simulations et d'une application aux données réelles avec le logiciel *R Core Team* (2013), nous avons montré grâce au calcul de la perte de prédiction (Section 2.3), que notre méthode est meilleure que celle de *An et al.* (2013) sur la base de la perte de prédiction du modèle.

## Sélection des variables en analyse discriminante mixte

Dans la seconde partie intitulée "sélection des variables en analyse discriminante mixte", nous avons établi au Chapitre 4, la deuxième contribution de nature plus théorique que pratique. Cette approche va suivre les mêmes étapes que celles du Chapitre 2. Nous proposons une méthode de sélection des variables en analyse discriminante pour des variables aléatoires mixtes pour plus de deux groupes. Dans cette étude, nous avons considéré un mélange de variables aléatoires continues et discrètes.

---

---

---

Deux principaux avantages sont à relever dans cette approche : le premier avantage vient du fait que notre méthode est valable même lorsque le nombre de groupe est supérieur à deux, et le second avantage suppose seulement que le moment d'ordre quatre du vecteur aléatoire  $X$  soit fini, c'est-à-dire que la méthode reste valable comme au Chapitre 2, quelle que soit la loi de probabilité suivie par la variable aléatoire continue  $X$ . Un nouveau critère de sélection pour les variables mixtes a été établi (Equation (4.7)). Il est regardé comme une mesure de perte d'information lorsqu'un sous-ensemble de variables est sélectionné. Le critère est utilisé pour ranger toutes les variables dans l'ordre décroissant du pouvoir discriminant afin que le problème de sélection se ramène à un problème d'estimation. Une étude théorique a été proposée à travers les Lemmes 6 - 9 et les Théorèmes 9 - 11 dans le but d'établir la convergence en probabilité de notre méthode de sélection. Cette étude généralise celle de *Nkiet* (2012) dans le cadre mixte. Pour achever notre étude, nous avons montré par simulation et sur un exemple de données réelles, à partir du taux de bon classement que notre méthode est équivalente à celle de *Mahat et al.* (2007) dans le cas de deux groupes, mais sélectionne moins de variables que celle de *Mahat et al.* (2007), ce qui lui donne un certain avantage.

## Sélection des variables en statistique fonctionnelle

La troisième et dernière contribution de cette thèse intitulée "sélection des variables en régression linéaire fonctionnelle", est une contribution purement théorique contrairement aux Chapitres 2 et 4. Le problème est relativement nouveau dans la statistique fonctionnelle. Notre objectif était de proposer une méthode de sélection pour des fonctions aléatoires. Pour cela, nous avons commencé par établir un critère de sélection

---

---

---

des variables aléatoires fonctionnelles. Pour introduire notre approche dans les modèles fonctionnels additifs, nous avons considéré dans une première étape un modèle linéaire multivarié (Equation (6.1)). Ce modèle est une généralisation du modèle (1) de *Nkiet* (2001). Il nous a permis d'étendre la méthode de sélection de *Nkiet* (2001) dans la cas du modèle (6.1). En utilisant l'équation approchée (6.14), nous avons proposé un critère de sélection pour les variables mixtes. Reste maintenant dans cette partie à étudier la convergence du critère  $\tilde{C}_K^{(n)}$  et proposer une étude de simulations semblable à celle des Chapitres 2 et 4.

## Perspectives

Immédiatement après cette thèse, dans le but d'approfondir les résultats obtenus et de les étendre dans d'autres types de modèles statistiques, nous nous pencherons sur les axes suivants :

1. Proposer une étude convergence du critère  $\tilde{C}_K^{(n)}$  (Chapitre 6) et un choix judicieux du paramètre  $q_n$ .
2. Proposer un package sur la base des programmes réalisés, dans le but de vulgariser notre méthode de sélection des variables et de proposer des solutions pratiques liées à la problématique de la sélection des variables.
3. Etendre notre méthode de sélection des variables aux modèles linéaires partiellement linéaire et aux modèles de régression logistiques.

### 6.3 Conclusion

---

Il est à noter que les limites de notre méthode de sélection pourront faire l'objet de futur travaux. Il sera ainsi question de regarder notre méthode de sélection des variables dans le cas où les variables  $X_1, \dots, X_p$  sont fortement corrélées. Puis, réduire l'influence des paramètres de contrôle  $\alpha, \beta, f_n$  et  $g_n$  sur notre méthode de sélection.

# Bibliographie

---

# Bibliographie

---

- Asparoukhov, O. and Krzanowski, W.J. Non-parametric Smoothing of the Location Model in Mixed Variables Discrimination, *Statistics and Computing.*, 10 :289–297, 2000.
- Azzalini, A. and Bowman, A. On the use of nonparametric regression for checking linear relationships. *J. Roy. Statist. Soc., Ser.B* 55 :549–557, 1993.
- Baiguo An and Jianhua Guo and Hansheng Wang, Multivariate regression shrinkage and selection by canonical correlation analysis, *Computational Statistics & Data Analysis*, 62 :93–107, 2013.
- Bar-Hen. A and Daudin. J.J Generalisation of the Mahanobis distance in the mixed case. *J Multivar Anal.*, 53, 332–342, 1995.
- Billingsley. *Convergence of probability measures*, Wiley., 1968.
- Breiman L, Spector P. Submodel selection and evaluation in regression. *Linear Algebra Appl.*, 210 :59–88, 1992.
- Cardot, H., Ferraty, F., Mas, A. and Sarda, P. Testing Hypotheses in the Functional Linear Model. *Fondation of the Scandinavian Journal of Statistics*, 30 :241–255, 2003a.
- Cardot, H., Ferraty, F., Sarda, P. Spline estimators for the functional linear model. *Statist. Sinica*, 13 :571–591, 2003b.
- Cardot, H., Goia, A. and Sarda, P. Testing for no effect in functional linear regression models, some computational approaches. *Comm.Statist.Simul.Comput.*, 33 :179–199, 2004.
- Cardot, H., Ferraty, F. and Sarda, P. Functional linear model, *Statistics & Probability Letters*, 45 :11–22, 1999.

- 
- Christophe Crambes. Modèles de régression linéaire pour des variables explicatives fonctionnelles. *Thèse de doctorat, Université Paul Sabatier, Toulouse II, 2006.*
- Daudin, J.J. Selection of variables in mixed-variable discriminant analysis. *Biometrics*, 42, 473–481, 1986.
- Daudin, J.J. and Bar-Hen, A. Selection in discriminant analysis with continuous and discrete variables. *Comput.Statist.Anal.*, 32, 161–175, 1999.
- Dauxois, J. et Pousse, A. *Les analyses factorielles en calcul des probabilités et en statistique : essai d'étude synthétique.* Thèse de doctorat, Université Paul Sabatier, Toulouse, 1976
- Dauxois, J., Pousse, A. and Romain, Y. Asymptotic theory for the principal component analysis of a random vector function : some applications to statistical inference. *Journal of Multivariate Analysis*, 12 :136–154, 1982.
- Dauxois, J., Romain, Y., Viguier, S. Tensor products and statistics., *Linear Algebra Appl.*, 210 :59–89, 1994.
- Dauxois, J. and Nkiet, G.M. Canonical analysis of two Euclidean subspaces and its applications., *Linear Algebra Appl.*, 264 :355–388, 1997.
- Dauxois, J. Nkiet, G.M and Romain Y. Linear relative canonical analysis of euclidean random variables, asymptotic study and some applications., *The Institute of Statistical Mathematics.*, 56(2) :279–304, 2004.
- Debasis kundu, Murali, G. Model selection in linear regression, *Computational Statistics and Data Analysis*, 22 :461–469, 1996.
- De Leon A.R, A. Soo, and T. Williamson, Classification with discrete and continuous variables via general mixed-data models, *Journal of Applied Statistics*, 38 :1021–1032, 2011.
- Deville, J.-C. Méthodes statistiques et numériques de l'analyse harmonique. *Annales de l'I.N.S.E.E.*, 15 :3–101, 1974.
- Droesbeke J.J., Fichet B., Tassi Ph., Modèles pour l'analyse des données multidimensionnelles, *Economica.*, 1992.
- Ferraty, F. and Vieu, P. Functional nonparametric model and application to spectrometric data. *Comput. Statist.*, to appear, 2002.
-

- 
- Frank, I.E. and J. H. Friedman. A statistical view of some chemometrics regression tools (with discussion), *Technometrics* 35 :109-148, 1993.
- Fujikoshi, Y., Kan, T., Takahashi S., Sakurai, T. Prediction error criterion for selection variables in linear regression model. *Ann. Inst. Stat. Math.*, 63 :387-403, 2011.
- Fujikoshi, Y., Sato, K. Modified AIC and  $C_p$  multivariate linear regression. *Biometrika*, 84 :707-716, 1997.
- Fujikoshi, Y. Selection of variables in two-group discriminant analysis by error rate and Akaike's information criterion. *J. Multivariate Anal.*, 17 :27-37, 1985.
- Fujikoshi, Yasunori, A test for additional information in canonical correlation analysis., *Annals of the Institute of Statistical Mathematics*, 34 :523-530, 1982.
- Gareth M. James, Jing Wang and Ji Zhu. Functional linear regression that's interpretable. *The Annals of Statistics*, 37(5A) :2083-2108, 2009.
- Gerhard Tutz & Jan Gertheiss. Feature Extraction in Signal Regression : A Boosting Technique for Functional Data Regression Technical Report Number 011, Department of Statistics University of Munich, <http://www.stat.uni-muenchen.de>, 2007
- Grenander, U., *Abstract inference*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York. ix+526 pp., 1981.
- Hastie, T., Mallows, C. A discussion of "A Statistical View of some Chemometrics Regression Tools". I.E. Frank and J.H. Friedman. *Technometrics*, 35 :140-143, 1993.
- Helland, I.S. Partial least squares regression and statistical models. *Scandinavian Journal of Statistics*, 17 :97-114, 1990.
- Hocking RR. The analysis and selection in linear regression. *Biometrics*, 32 :1-49, 1976.
- Hoerl, A.E. and Kennard, R.W. Ridge regression : advances, algorithms and applications. *American Journal of Mathematical Management Sciences*, 1 :5-83, 1980.
- Johnson, R.A, Wichern, Y.Y., *Applied Multivariate Statistical Analysis*. fifth ed. Pearson Education, 2003.
- Krishnaiah, P.R. *Selection of variables in discriminant analysis*. In :P.R Krishnaiah, L.N. Kanal (eds), *Handbook of Statistic*, 2, North-Holland, Amsterdam :883-892, 1982.
-



- 
- Krzanowski, W.J. Stepwise location model choice in mixed-variable discrimination. *Appl. Stat.*, 32, 160–166, 1983.
- Krzanowski, W.J. Mixtures of continuous and categorical variables in discriminant analysis. *Biometrics*, 36, 493–499, 1980.
- Krzanowski, W.J. Discrimination and Classification Using Both Binary and Continuous Variables. *Journal of the American Statistical Association*, 352, 782–790, 1975.
- Laurent Delsol. Régression sur variable fonctionnelle : Estimation, Tests de structure et Applications. *Thèse de doctorat, Université Paul Sabatier, Toulouse III, 2008*
- Lichman M, Classification with discrete and continuous variables via general mixed-data models, *UCI Machine Learning Repository* [<http://archive.ics.uci.edu/ml>]. Irvine, CA :University of California, School of Information and Computer Science.
- Linhart, H. and Zucchini, W. *Model selection*. Wiley, New York, 1986.
- Liu, Y. C., & Yeh, I. C. Using mixture design and neural networks to build stock selection decision support systems. *Neural Computing and Applications*, 1-15. (Print ISSN 0941-0643, Online ISSN 1433-3058, First online : 16 November 2015, DOI 10.1007/s00521-015-2090-x)
- Mahat, N.I., Krzanowski, W.J., and Hernandez, A. Variable selection in discriminant analysis based on the location model for mixed variables., *Adv. Data Anal.*, 1 :105–122, 2007.
- Marx, B. D., and Eilers, P. H. *Generalized linear regression on sampled signals with penalized likelihood*. In : Forcina, A., Marchetti, G.M., Hatzinger, R., Galmacci, G. (Eds), *Statistical Modelling, Proceedings of the Eleventh International Workshop on Statistical Modelling*, Orvieto, 1996.
- Marx, B. D., and Eilers, P. H. Generalized linear regression on sampled signals and curves : a P-spline approach., *Technometrics*, 41 :1–13, 1999.
- McKay, R.J. Simultaneous procedures for variable selection in multiple discriminant analysis., *Biometrika*, 64 :283–290, 1977.
- McKay, R.J. Simultaneous procedures in discriminant analysis involving two groups. *Technometrics*, 18, 47–53, 1976.
- McLachlan, G.J. On the relationship between the F test and the overall error rate for variable selection in two-group discriminant analysis. *Biometrics*, 36, 501–510, 1980.
-

- 
- Miller AJ. *Subset selection in regression*. Chapman and Hall, London, 1990.
- Nkiet, G. M. Sélection des variables dans un modèle structurel de régression linéaire. *C. R. Acad. sci. paris I*, 333 :1105–1110, 2001.
- Nkiet, G. M. Inference for the invariance of canonical analysis under linéaire transformations. *Journal of Multivariate Analysis*, 84 :1–18, 2003.
- Nkiet, G. M. On estimation of the dimensionality in linéaire canonical analysis. *Statistics & Probability Letters*, 75 :103–112, 2005.
- Nkiet, G. M. Direct variable selection for discrimination among several groups. *J. Multivariate Anal.*, 105 :151–163, 2012.
- Olkin, I. and Tate, R.F. Multivariate correlation models with mixed discrete and continuous variables. *Ann. Math. Stat.*, 32, 448–465, 1961.
- Preda, C. ; Saporta, G. PLS approach for clusterwise linear regression on functional data. *Classification, clustering, and data mining applications 167-176*. Stud. Classification Data Anal. Knowledge Organ., Springer, Berlin., 2004.
- Preda, C. ; Saporta, G. PLS regression on a stochastic process. *Comput. Statist. Data Anal.*, 48, 149–158., 2005.
- Preda, C. ; Saporta, G. PCR and PLS for Clusterwise Regression on Functional Data. *Selected Contributions in Data Analysis and Classification 589-598*. Springer, Berlin, Heidelberg, 2007.
- Ramsay, J.O. *When the data are functions*. *Psychometrika* 47, 379–396, 1982.
- Ramsay J.O., Silverman B.W. *Functional data analysis*. Springer-Verlag, Berlin, 1997.
- Ramsay J.O., Silverman B.W., *Applied functional data analysis : methods and case studies*. Springer-Verlag, 2002.
- Ramsay, J.O., and Silverman, B. W. *Functional data analysis (2nd ed.)*. New York : Springer-Verlag, 2005.
- Rao CR. *Linear statistical inference and its applications*. 2nd edn. Wiley, New York, 1973.
- Le Roux, N.J., Steel, S.J., Louw, N. Variable selection and error rate estimation in discriminant analysis. *J.Stat.Comput.Simul.*, 59,195–219, 1997.
-

- R Core Team, R : A Language and Environment for Statistical Computing, R Foundation for Statistical Computing. *Vienna, Austria*, 3-900051-07-0, <http://www.R-project.org/> 2013.
- Shao J. Linear model selection by cross-validation. *J. Amer. Statist. Assoc.*, 88 :486–494, 1993.
- Shibata R. Approximate efficiency of a selection procedure for the number of regression variables. *Biometrika*, 71 :43–49, 1994.
- Tamoudi, A. Analyse Canonique des données aléatoires. *Thèse de troisième cycle*, Université Paul Sabatier, Toulouse, 1987.
- Thompson M.L., Selection of variables in multiple regression, Part I, *Internat., Statist.Rev.*, 46 :1–19, 1978.
- Volle M., L'analyse des données, *Economica.*, 1985.
- A.F. Yao. L.Ferré. Functional sliced inverse regression analysis. *Statistics*, 37 :475–488, 2003.
- Zhang P. On the distributional properties of model selection criteria. *J. Amer. Statist. Assoc.*, 87 :732–737, 1992.
- Zhang P. Model selection via multifold cross validation. *Ann. Statist.*, 21 :299–313, 1993.
- Zheng X., Loh W.Y., A consistent variable selection criterion for linear models with high-dimensional covariates., *Statist. Sinica.*, 7 :311–325, 1997.