HAL open science

Estimation d'une matrice de précision sous mélange de lois de Wishart

Djamila Boukehil

▶ To cite this version:

Djamila Boukehil. Estimation d'une matrice de précision sous mélange de lois de Wishart. Statistiques [math.ST]. Normandie Université, 2021. Français. NNT : 2021NORMR053 . tel-03658435

HAL Id: tel-03658435

https://theses.hal.science/tel-03658435v1

Submitted on 4 May 2022

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

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

Normandie Université

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité MATHEMATIQUES

Préparée au sein de l'Université de Rouen Normandie

Estimation d'une matrice de précision sous mélange de lois de Wishart

Présentée et soutenue par
DJAMILA BOUKEHIL

Thèse soutenue le 03/11/2021 devant le jury composé de		
MME GHISLAINE GAYRAUD	PROFESSEUR DES UNIVERSITES, UNIV TECHNOLOGIE COMPIEGNE UTC COMPIEGNE	Rapporteur du jury
M. RABAH MESSACI	PROFESSEUR DES UNIVERSITES, Université Houari BOUMEDIENE	Rapporteur du jury
M. HACÈNE BELBACHIR	PROFESSEUR DES UNIVERSITES, Université Houari BOUMEDIENE	Membre du jury
MME NADJIA EL SAADI	PROFESSEUR DES UNIVERSITES, ENSSEA	Membre du jury
MME FATIHA MEZOUED	PROFESSEUR DES UNIVERSITES, ENSSEA	Membre du jury
M. DOMINIQUE FOURDRINIER	PROFESSEUR DES UNIVERSITES, Université de Rouen Normandie	Directeur de thèse

Thèse dirigée par DOMINIQUE FOURDRINIER, Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes

UNIVERSITÉ DE ROUEN NORMANDIE

MIIΣ

litis

"Patience, persistence and perspiration make an unbeatable combination for success"

Napoleon Hill

Remerciements

Un immense merci à mes deux directeurs de thèse Fatiha Mezoued et Dominique Fourdrinier pour avoir accepté de diriger et de suivre ce projet de thèse. Je les remercie pour leur réel engagement, pour leurs disponibilités et précieux conseils qui ont fait progresser ce travail dans la bonne voie. J'ai vraiment apprécié leurs qualités humaines qui m'ont beaucoup aidé pour bien avancer pendant toutes les années de thèse. Je suis très reconnaissante pour tout ce qu'ils ont fait pour moi... grand merci.

Merci tout aussi grand à William E.Strawderman pour sa collaboration, ses suggestions très concrètes et ses idées pertinentes qui ont rendu nos publications plus instructives.

Mes remerciements les plus sincères vont également aux membres de jury de thèse : Nadjia El Saadi, Ghislaine Gayraud, Rabah Messaci et Hacène Belbachir pour avoir bien voulu lire et examiner l'intégrité de ce travail. Veuillez trouver ici ma profonde gratitude.

Je remercie le Ministère Algérien de l'Enseignement Supérieur et de la Recherche Scientifique et Campus France qui ont soutenu financièrement ce travail via le programme de coopération algéro-français Tassili.

Je remercie Stéphane Canu pour son excellent accueil au sein du laboratoire LITIS à l'INSA de Rouen, je le remercie vivement pour son aide et son encouragement.

Je tiens à remercier les membres de mon comité de suivi de thèse : Paul Raynaud de Fitte et Ghislaine Gayraud qui ont accepté d'évaluer les conditions de la formation et les avancées de ma recherche durant toutes les années de thèse. Je les remercie pour les discussions intéressantes à chaque réunion.

Je remercie les secrétaires des laboratoires LITIS et LAMOPS : Brigitte Diara, Leila Lahcen, Sandra Hague, Fabiane Bocquet, Mathieu Baum, et Mohammed Hadj Henni pour leur travail parfait, ils m'ont beaucoup aidé et ils ont résolu tous mes problèmes administratifs.

Je remercie Anis Haddouche pour son aide précieuse et surtout pour ses utiles conseils et les conversations intéressantes qu'on a eu tout au long de ma thèse.

Je remercie mes collègues de bureau à l'INSA de Rouen avec qui j'ai passé de bons moments : Rachel Blin, Ismaila Seck, Cyprien Ruffino, Imad Rida, Soufiane Belharbi, Yuan Liu, Linlin Jia et Nikolaos Adaloglou. Merci pour vos encouragements et vos conseils.

Je remercie énormément Djamila Boukhaoui, Marlène Paris, Assia Khenoussi, Belkis Taguida et Nihad Nouri pour l'écoute et le soutien pendant les moments difficiles, et pour tout les mo-

ments inoubliables que nous avons passés ensemble. J'avais de la chance d'être bien entourée de personnes aussi fidèles.

Je remercie Imen Trabelssi et son époux Sami pour tout ce qu'ils ont fait pour moi. Merci Imen pour les plats et les gâteaux délicieux, tu étais ma première amie à Rouen et j'ai vraiment passé de très bons moments avec toi.

Pour mes collègues de l'ENSSEA : Meriem Chinoune, Mohamed Benahmed, Farida Zamouche et Abderrahim Abdellaoui. Merci beaucoup pour votre soutien.

Je remercie Dalia Haider, Lounis Douadi, Sihem Semmari pour leurs aides et leurs soutiens.

À Amina ma colocataire de Kolea, merci beaucoup pour le soutien et pour les agréables moments que nous avons passé ensemble.

À mes amies de l'USTHB : Hana Bourita, Houda boughaba, Imene Boutria, Sabrina Lemzaouda, Samira Attou, Amal Benali, Bouzarra Lamia et Imene Hadjbenelezaar, je vous remercie pour vos aides et votre soutien.

Je n'oublie pas mes enseignants : de l'école primaire Taleb Abdelkrim, collège Rabeh Hocine, lycée Hillal Abdallah, université des frères Mentouri Constantine, université de l'USTHB et ceux de l'ENSSEA, je vous remercie tous pour vos efforts et votre patience.

Je remercie mes chers parents, frères, sœurs et ma belle sœur qui étaient présents au quotidien malgré la distance. Je vous remercie pour votre soutien et votre amour. Je remercie aussi mes grands parents, mes oncles, tantes, cousins et cousines pour leurs aides et leurs encouragements.

J'exprime ma profonde reconnaissance à mon époux pour le soutien continu et pour sa présence au quotidien et dans les moindres détails. Je remercie aussi ma belle famille pour les encouragements.

Enfin, Je tiens à remercier toute personne qui a contribué de près ou de loin à l'aboutissement de cette thèse.

Résumé

Dans cette thèse, nous nous intéressons à l'estimation de la matrice de précision Σ^{-1} d'un modèle de mélange de lois de Wishart, $S \mid V \sim \mathcal{W}_p(n, V\,\Sigma)$, sous des coûts de type Efron-Morris $L_k\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \mathrm{tr}[\{\hat{\Sigma}^{-1} - \Sigma^{-1}\}^2\, S^k]$, pour tout $k = 1, 2, 3 \ldots$. Ici S est une matrice d'observation de dimension $p \times p$, V est une variable de mélange qui suit une distribution $\mathcal{H}(\cdot)$ connue et définie sur $\mathbb{R}^+$ et $\mathcal{W}_p(n, V\,\Sigma)$ représente la distribution de Wishart à n degrés de liberté et de matrice de variance-covariance $V\,\Sigma$. Dans une approche unifiée des deux cas où la matrice S est inversible et où elle est singulière, nous considérons les estimateurs usuels de type $a\,S^+$ où a est une constante positive et où S^+ est l'inverse de Moore-Penrose (qui coïncide avec l'inverse S^{-1} de S lorsque S est inversible). Nous fournissons des estimateurs optimaux dans cette classe, notés $a^*\,S^+$ sous les coûts L_1, L_2 et L_3, et des estimateurs alternatifs, notés $a_0\,S^+$, sous les coûts L_0 et L_k, pour $k \geq 4$, cas où il n'existe pas d'estimateurs optimal.

Étant donné que tous les estimateurs usuels de la forme $\hat{\Sigma}_a^{-1} = a\,S^+$ sont inadmissibles, nous proposons deux types d'estimateurs alternatifs de la forme $\hat{\Sigma}_{a,c}^{-1} = a\,S^+ + c\,S\,G(S)$ et $\hat{\Sigma}_{a,c,r}^{-1} = a\,S^+ + c\,r(\mathrm{tr}\{S\})\,S\,G(S)$, où c est une constante positive, $r(\cdot)$ est une fonction réelle et $G(S)$ est une fonction matricielle d'ordre p homogène de degré α. Notons que pour les deux types d'estimateurs, la fonction $G(S)$ peut être ou pas orthogonalement invariante. Grâce à une identité de type Stein-Haff, nous établissons un estimateur sans biais de la différence de risque entre les estimateurs $\hat{\Sigma}_{a,c}^{-1}$ et $\hat{\Sigma}_a^{-1}$ et nous fournissons des conditions sur c, $G(S)$ et $r(\cdot)$ afin que les estimateurs $\hat{\Sigma}_{a,c}^{-1}$ et $\hat{\Sigma}_{a,c,r}^{-1}$ améliorent les estimateurs $\hat{\Sigma}_a^{-1}$, sous les fonctions de coût $L_k\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \mathrm{tr}[\{\hat{\Sigma}^{-1} - \Sigma^{-1}\}^2\, S^k]$, pour $k = 1, 2, 3 \ldots$. En particulier, nous donnons des exemples d'estimateurs de type Haff, Dey et Efron-Morris qui dominent les estimateurs $\hat{\Sigma}_a^{-1}$ à travers leur domination sur les estimateurs optimaux $\hat{\Sigma}_{a^*}^{-1}$ ou les estimateurs raisonnables $\hat{\Sigma}_{a_0}^{-1}$.

En dernier lieu, à la lumière des estimateurs de type Haff, Dey et Efron-Morris dont la fonction de correction $G(S)$ est homogène de degré -2, nous mettons en évidence le fait que seul ce degré d'homogénéité fait sens. Nous illustrons ce phénomène dans le cadre de l'estimation d'un vecteur moyenne d'une loi normale en montrant que, parmi les estimateurs de type James Stein d'un certain degré d'homogénéité, seule l'homogénéité d'ordre -2 est valide.

Mots clés : Lois à symétrie elliptique, mélange de lois de Wishart, identité de type Stein-Haff, matrice de précision, coûts de type Efron-Morris, fonction homogènes.

Abstract

In this thesis, we consider the problem of estimating the precision matrix Σ^{-1} of a mixture of Wishart distributions model $S \mid V \sim \mathcal{W}_p(n, V\,\Sigma)$, under various Efron-Morris type losses, $L_k\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \mathrm{tr}[\{\hat{\Sigma}^{-1} - \Sigma^{-1}\}^2\, S^k]$, for $k = 1, 2, 3 \ldots$. Here S is the $p \times p$ sample covariance matrix, V is a mixing variable which has a known distribution $\mathcal{H}(\cdot)$ on $\mathbb{R}_+$ and $\mathcal{W}_p(n, V\,\Sigma)$ denotes the Wishart distribution with n degrees of freedom and positive-definite covariance matrix $V\,\Sigma$. In a unified approach to the cases where S is invertible and S is singular, we consider the class of usuel estimator $a\,S^+$, where S^+ denotes the Moore-Penrose inverse of S (which coincides with the inverse S^{-1} of S when S is invertible) and a is a positive constant, we provide optimal estimators in this class, denoted by $a^*\,S^+$, under the losses L_1, L_2 and L_3 ; and alternative estimators, denoted by $a_0\,S^+$, under the losses L_0 and L_k, for $k \geq 4$, cases where optimal estimators do not exist.

As the usual estimators of the form $\hat{\Sigma}_a^{-1} = a\,S^+$ are inadmissibles, we propose two types of alternative estimators of the form $\hat{\Sigma}_{a,c}^{-1} = a\,S^+ + c\,S\,G(S)$ and $\hat{\Sigma}_{a,c,r}^{-1} = a\,S^+ + c\,r(\mathrm{tr}\{S\})\,S\,G(S)$, where c is a positive constant, $r(\cdot)$ is a real function and $G(S)$ is a $p \times p$ homogeneous matrix function of order α. Note that, for both types of estimators, the function $G(S)$ may be orthogonally invariant or not. We develop an unbiased estimator of the risk difference between $\hat{\Sigma}_{a,c}^{-1}$ (respectively $\hat{\Sigma}_{a,c,r}^{-1}$) and $\hat{\Sigma}_a^{-1}$, relying on a Stein-Haff type identity and we provide conditions on c, $G(S)$ and $r(\cdot)$ so that the estimators $\hat{\Sigma}_{a,c}^{-1}$ and $\hat{\Sigma}_{a,c,r}^{-1}$ improve on the estimators $\hat{\Sigma}_a^{-1}$ under the losses $L_k\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \mathrm{tr}[\{\hat{\Sigma}^{-1} - \Sigma^{-1}\}^2\, S^k]$, for $k = 1, 2, 3 \ldots$. In particular, we give examples of Haff, Dey and Efron-Morris type estimators that dominate the estimators of the form $\hat{\Sigma}_a^{-1}$, specifically for $a = a^*$ and $a = a_0$.

Finally, based on the information provided by the estimators of Haff, Dey and Efron-Morris type whose correction function $G(S)$ is homogeneous of degree -2, we highlight the fact that only this degree of homogeneity makes sense. We illustrate this phenomenon in the context of the estimation of a mean vector of a normal distribution by showing that, among the James-Stein type estimators of a certain degree of homogeneity, only the homogeneity of order -2 is valid.

Key words : Elliptically symmetric distributions, scale mixtures of Wishart distributions, Stein-Haff type identity, scatter matrix, Efron-Morris type losses, homogeneous function.

Symboles

$\mathbb{R}$	L'ensemble des nombres réels.
$\hat{A}$	L'estimateur du paramètre inconnu A.
I_m	La matrice identité de dimension $m \times m$.
A^{T}	La transposée de la matrice A.
A^{+}	L'inverse de Moore-Penrose de A.
$\mathrm{tr}(A)$	La trace de la matrice A.
$\|A\|$	La norme euclidienne du vecteur A.
$\|A\|_F$	La norme de Frobenius de la matrice A .
$\|A\|$	Le déterminant de la matrice A.
$\mathrm{rang}(A)$	Le rang de la matrice A.
$\otimes$	Le produit de Kronecker.
$\sim$	Signifie *distribué selon*.
$E_{\theta,\Sigma}$	L'espérance d'une variable distribuée selon une densite $f(\cdot)$ parametrée par θ et Σ.
δ_{ij}	Le symbole de Kronecker où $\delta_{ij} = 1$ si $i = j$ et $\delta_{ij} = 0$ si $i \neq j$.
$a \wedge b$	Le minimum entre les deux entiers a et b.
$a \vee b$	Le maximum entre les deux entiers a et b.
f'	La dérivée de la fonction f.
$\stackrel{d}{=}$	L'égalité en distribution.
$\mathcal{O}_p$	L'ensemble des matrices orthogonales $\{A \in \mathbb{R}^{p\times p} : A^{\mathrm{T}}A = AA^{\mathrm{T}} = I_n\}$.
$\mathcal{L}_{p,r}$	La sous-variété de Stiefel $\{A \in \mathbb{R}^{p\times r} : A^{\mathrm{T}}A = I_r\}$.
$\mathcal{D}_p$	L'ensemble des matrices diagonales diag $(d_1, \ldots, d_p)$ où $d_1 >, \ldots, > d_p$.
$\Gamma(\cdot)$	La fonction Gamma sur l'ensemble $\mathbb{R}_+^*$.
$\mathbb{P}$	L'ensemble des paramètres.
$\mathbb{S}_p$	L'ensemble des matrices $p \times p$ définies positives.

Liste des figures

4.1 PRIAL des estimateurs de type Haff et Dey pour $a = a_0^- = n - 2\,p - 2$. 53
4.2 PRIAL des estimateurs de type Haff et Dey pour $a = a_0^- = (n - 2\,p - 2)\mu_{-1}/\mu_{-2}$. 53
4.3 PRIAL des estimateurs de type Haff et Dey pour $a = a^* = n - p - 1$. 53
4.4 PRIAL des estimateurs de type Haff et Dey pour $a = a^* = (n - p - 1)/\mu_{-1}$. . . . 53
4.5 PRIAL des estimateurs de type Haff et Dey pour $a = a_0^- = p - 2\,n - 2$. 53
4.6 PRIAL des estimateurs de type Haff et Dey pour $a = a_0^- = (p - 2\,n - 2)\,\mu_{-1}/\mu_{-2}$. 53
4.7 PRIAL des estimateurs de type Haff et Dey pour $a = a^* = p - n - 1$. 53
4.8 PRIAL des estimateurs de type Haff et Dey pour $a = a^* = (p - n - 1)/\mu_{-1}$. . . . 53

Liste des tableaux

2.1 Exemples de densités à symétrie sphérique. 16

4.1 PRIAL de l'estimateur de type Dey quand $S \sim \mathcal{W}_p(n,\, I_p)$. 54
4.2 PRIAL de l'estimateur de type Haff quand $S \sim \mathcal{W}_p(n,\, I_p)$. 54
4.3 PRIAL de l'estimateur de type Dey quand $S \sim \mathcal{W}_p(n, V\, I_p)$ avec $V \sim \mathcal{IG}(9, 1/2)$. 54
4.4 PRIAL de l'estimateur de type Haff quand $S \sim \mathcal{W}_p(n, V\, I_p)$ avec $V \sim \mathcal{IG}(9, 1/2)$. 55

Table des matières

Remerciement ii

Résumé iv

Abstract v

Symboles vi

Liste des figures vii

Liste des tableaux vii

Table des matières viii

1 Introduction 1
- 1.1 Matrice de précision . . . 1
- 1.2 Estimation d'une matrice de précision . . . 2
- 1.3 État de l'art . . . 3
- 1.4 Objectif et structure de la thèse . . . 8

2 Rappels et préliminaires 11
- 2.1 Algèbre des matrices . . . 11
- 2.2 Théorie de la décision . . . 13
- 2.3 Lois multidimensionnelles . . . 13
- 2.4 Lois matricielles . . . 18
- 2.5 Des identités de type Stein-Haff . . . 20

3 Estimation d'une matrice de précision sous des coûts basés sur les données 23
- 3.1 Introduction . . . 23
- 3.2 Estimateurs de forme générale . . . 24
- 3.3 Estimateurs de référence . . . 27
- 3.4 Estimateurs améliorés . . . 29
- 3.5 Estimateurs orthogonalement invariant . . . 31
- 3.6 Estimateur non nécessairement orthogonalement invariant . . . 39
- 3.7 Conclusion . . . 49

4 Étude numérique 51

5 Sur l'homogénéité d'ordre -2 des estimateurs améliorés 57

5.1 Introduction 57
5.2 Résultats de domination 59
5.3 Extensions 63
5.4 Une propriété d'optimalité uniforme de l'estimateur de James-Stein 67

6 Conclusion **69**

Appendices **71**

A Lemmes importants **72**

Bibliographie **83**

CHAPITRE 1

INTRODUCTION

Sommaire

1.1 Matrice de précision **1**
1.2 Estimation d'une matrice de précision **2**
1.3 État de l'art **3**
1.3.1 Le contexte où $p \leq n$ 3
1.3.2 Le contexte où $p > n$ 6
1.3.3 Le cadre distributionnel elliptique 7
1.4 Objectif et structure de la thèse **8**

1.1 Matrice de précision

La matrice de précision, également appelée matrice de concentration, est l'inverse de la matrice d'échelle (qui se ramène à une matrice de covariance dans le cas gaussien), elle contient l'information sur la corrélation partielle entre les variables. En effet, ses éléments nous permettent de calculer la covariance entre deux variables conditionnellement à celles qui restent. Mathématiquement, pour Σ la matrice de covariance d'un ensemble de p variables X, on considère la matrice de covariance partielle $\Sigma_{a,b} = \Sigma_{a,a} - \Sigma_{a,b}\,\Sigma_{b,b}^{-1}\,\Sigma_{b,a}$ entre les deux sous-ensembles de variables $X_a = (X_1, X_2)$ constitué des deux premières variables, et X_b le sous-ensemble des variables qui restent. Supposons que $A = \Sigma^{-1}$ alors $\Sigma_{a,b} = A_{a,a}^{-1}$ (voir Hastie, Tibshirani et Friedman [HTF01]). Dans le cas particulier de la loi normale multivariée, le fait qu'un élément (i, j) de la matrice de précision Σ^{-1} vale zéro signifie que les $i^{\text{ème}}$ et $j^{\text{ème}}$ variables sont conditionnellement indépendantes. C'est d'ailleurs une des raisons pour lesquelles la matrice de précision est très importante dans beaucoup de domaines. Nous citons quelques exemples :

— En intelligence artificielle, en particulier en apprentissage automatique (Machine Learning) le modèle graphique gaussien représente les dépendances conditionnelles entre des variables aléatoires gaussiennes dans des systèmes complexes. Son estimation est équivalente à l'estimation de la matrice de précision (voir Pourahmadi [Pou13]).

— En biologie moléculaire, le problème d'estimation du réseau de régulation génique basé sur les données d'expression de gènes peut être ramené à l'estimation de la matrice de précision ou à la sélection de ses éléments non nuls (voir Li et Jackson [LJ15]).

— En finance, la matrice de précision joue un rôle important dans la détermination du portefeuille optimal, notamment au travers de l'estimation des coefficients discriminants. Elle est ainsi utilisée pour chercher la pondération des actifs qui minimise la variance du portefeuille (voir Kourtis, Dotsis et Markellos [KDM12]).

1.2 Estimation d'une matrice de précision

Soit X une matrice aléatoire de dimension $n \times p$, où n est le nombre d'observation et p est le nombre de variable, qui suit une loi normale matricielle de moyenne 0 et de covariance $I_p \otimes \Sigma$. On note $X \sim N_{n\times p}(M, \Omega \otimes \Sigma)$ et ainsi $S = X^T X$ suit une loi de Wishart à n degré de liberté et de matrice de covariance Σ. On note $S \sim \mathcal{W}_p(n, \Sigma)$. L'estimateur usuel de la matrice de précision Σ^{-1} est de la forme.

$$\hat{\Sigma}_a^{-1} = a\, S^+, \tag{1.2.1}$$

où a est une constante positive et S^+ est l'inverse de Moore Penrose (sa définition est rappelée dans le paragraphe 2.1.1), cette notation nous permet d'unifier les deux cas où la matrice S est singulière (soit quand $p > n$) et où S est inversible (soit quand $p \leq n$), auquel cas $S^+ = S^{-1}$. Les estimateurs $\hat{\Sigma}_a^{-1}$ sont des estimateurs naturels dans la mesure où S est la matrice de covariance empirique. Notons qu'alors, la matrice de précision Σ^{-1} admet l'estimateur sans biais

$$\hat{\Sigma}_{n-p-1}^{-1} = (n - p - 1)\, S^{-1}. \tag{1.2.2}$$

En théorie de la décision, les divers estimateurs $\hat{\Sigma}^{-1}$ de Σ^{-1} sont évalués en se fixant une fonction réelle L telle que

$$L\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} \geq 0,$$

qui représente le coût en Σ^{-1} de $\hat{\Sigma}^{-1}$.

À elle seule celle ci ne peut suffire à évaluer $\hat{\Sigma}^{-1}$ puisque elle dépend de chacune de ses réalisations. De manière à disposer d'un critère qui rende compte de toutes les valeurs possibles de $\hat{\Sigma}^{-1}$, on établit le risque en Σ^{-1} de $\hat{\Sigma}^{-1}$ en posant

$$R(\Sigma^{-1}, \hat{\Sigma}^{-1}) = \mathrm{E}_\Sigma[L\{\Sigma^{-1}, \hat{\Sigma}^{-1}\}], \tag{1.2.3}$$

où E_Σ est l'espérance respectivement à la loi de S. Dans la comparaison de deux estimateurs $\hat{\Sigma}_1^{-1}$ et $\hat{\Sigma}_2^{-1}$, on dit que l'estimateur $\hat{\Sigma}_1^{-1}$ améliore l'estimateur $\hat{\Sigma}_2^{-1}$ si

$$R(\Sigma^{-1}, \hat{\Sigma}_1^{-1}) \leq R(\Sigma^{-1}, \hat{\Sigma}_2^{-1}).$$

On dit que $\hat{\Sigma}_1^{-1}$ est meilleur que $\hat{\Sigma}_2^{-1}$ ou domine $\hat{\Sigma}_2^{-1}$ si, de plus, il existe au moins une valeur de Σ^{-1} telle que cette inégalité soit stricte. Un estimateur $\hat{\Sigma}_2^{-1}$ est dit inadmissible s'il existe un autre estimateur $\hat{\Sigma}_1^{-1}$ qui le domine. Inversement, si il n'existe aucun estimateur $\hat{\Sigma}_1^{-1}$ qui domine $\hat{\Sigma}_2^{-1}$ alors l'estimateur $\hat{\Sigma}_2^{-1}$ est dit admissible.

Dans la classe des estimateurs usuels $\hat{\Sigma}_a^{-1}$ dans (1.2.1), il peut exister une valeur optimale de a, soit a^*, pour laquelle le risque de l'estimateur $a^*\, S^+$ est minimal c'est à dire tel que

$$R(\Sigma^{-1}, a^*\, S^+) \leq R(\Sigma^{-1}, a\, S^+),$$

pour tout $a > 0$. Plusieurs travaux sur l'estimation de la matrice de précision ont montré que, sous les coûts de Stein et quadratique $L\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \mathrm{tr}(\hat{\Sigma}^{-1}\,\Sigma^{-1}) - \ln|\hat{\Sigma}^{-1}\,\Sigma^{-1}| - p$ et $L\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \mathrm{tr}(\hat{\Sigma}^{-1}\,\Sigma^{-1} - I_p)^2$ respectivement, les estimateurs $\hat{\Sigma}_a^{-1} = a\,S^+$ sont inadmissibles et peuvent être améliorés par d'autres estimateurs (voir, par exemple, Pal [Pal93] et Kubokawa et Inoue [KI14]).

Dans cette thèse, les fonctions de coût que nous considérons sont de la forme

$$L_k\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \mathrm{tr}[\{\hat{\Sigma}^{-1} - \Sigma^{-1}\}^2\,S^k], \tag{1.2.4}$$

pour $k = 1, 2, 3\ldots$ Notons que lorsque $k = 0$, la fonction dans (1.2.4) est le coût quadratique classique. Pour tout $k \geq 1$ ce coût est basé sur les données dans la mesure où l'observation S y intervient (sous différentes puissances). Dans ce qui suit, nous présentons les travaux réalisés sous les coûts L_0, L_1 et L_2, qui, à notre connaissance, sont les seuls rencontrés chez les auteurs traitant de l'estimation d'une matrice de précision.

1.3 État de l'art

Dans la littérature, plusieurs recherches concernant l'estimation de la matrice de précision ont été réalisées dans le but de trouver des estimateurs qui améliorent $a\,S^+$. Dans ce qui suit, nous citons les travaux effectués dans différents contextes distributionnels et sous différentes fonctions de coût. Nous en donnons une présentation selon l'ordre des dimensions : $n \geq p$ et $p > n$ où n est le nombre d'observations et p est le nombre de variables.

1.3.1 Le contexte où $p \leq n$

Soit S une matrice d'ordre p qui suit une loi de Wishart à n degré de liberté et de matrice de covariance Σ

$$S \sim \mathcal{W}_p(n,\, \Sigma).$$

En 1976, Efron et Morris [EM76] ont montré que le problème d'estimation de paramètre de position θ d'une distribution gaussienne par des estimateurs de bayes empirique se réduit à l'estimation de l'inverse de la matrice de covariance, et ils ont proposé un estimateur sans biais pour Σ^{-1} sous le coût

$$L^{EM}\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \frac{\mathrm{tr}[\{\hat{\Sigma}^{-1} - \Sigma^{-1}\}^2 S]}{n\,\mathrm{tr}\{\Sigma^{-1}\}}, \tag{1.3.1}$$

cet estimateur est de la forme $\hat{\Sigma}_a^{-1} = a\,S^{-1}$ avec $a = n - p - 1$ et $n \geq p + 1$. De plus, ils l'ont amélioré pour tout $p \geq 2$ par un estimateur minimax (qui minimise le maximum de la fonction de risque) et optimal dans la classe des estimateurs de la forme

$$\hat{\Sigma}_{EM}^{-1} = a\,S^{-1} + \frac{b}{\mathrm{tr}\{S\}}\,I_{p\times p}, \tag{1.3.2}$$

pour $a = n - p - 1$ et $b = p^2 + p - 2$.

En 1977, Haff [Haf77a] a proposé un estimateur qui généralise celui de Efron et Morris dans (1.3.2). Cet estimateur est de la forme

$$\hat{\Sigma}_H^{-1} = a\,S^{-1} + b\,r(S)\,I_{p\times p}, \tag{1.3.3}$$

où $r(S)$ est une fonction définie dans $\mathbb{R}^{p(p+2)/2}$ et où a et b sont deux constantes positives. Pour le cas particulier $r(S) = 1/\mathrm{tr}\{S\}$, Haff a montré que sous le coût

$$L_H\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \mathrm{tr}[\{\hat{\Sigma}^{-1} - \Sigma^{-1}\}^2 Q], \tag{1.3.4}$$

où Q est une matrice arbitraire définie positive, l'estimateur $\hat{\Sigma}^{-1}_{n-p-1}$ améliore l'estimateur $\hat{\Sigma}^{-1}_{EM}$ *i.e* $R_H(\Sigma^{-1}, \hat{\Sigma}^{-1}_{n-p-1}) \leq R_H(\Sigma^{-1}, \hat{\Sigma}^{-1}_{EM})$ pour tout Σ^{-1} alors que Efron et Morris [EM76] ont montré que c'est l'estimateur $\hat{\Sigma}^{-1}_{EM}$ qui améliore $\hat{\Sigma}^{-1}_{n-p-1}$ sour le coût $L^{EM}\{\Sigma^{-1}, \hat{\Sigma}^{-1}\}$ dans (1.3.1).

Pour $r(S) = t(U)/\mathrm{tr}\{S\}$ où $U = p\,|S|^{1/p}\,\mathrm{tr}\{S\}$ tel que $|S|$ est le déterminant de S et où $t(U)$ est une fonction bornée, lipschitzienne et continue sur $[0, 1]$, Haff a donné les résultats de domination de l'estimateur $\hat{\Sigma}^{-1}_H$ sur l'estimateur $\hat{\Sigma}^{-1}_{EM}$ ainsi que de l'estimateur $\hat{\Sigma}^{-1}_{EM}$ sur l'estimateur usuel $\hat{\Sigma}^{-1}_a$ sous les deux coûts L^{EM} dans (1.3.1) et L_H dans (1.3.4). Mathématiquement, il a exprimé des conditions sur la fonction $t(U)$ pour que les deux inégalités

$$R_{EM}(\Sigma^{-1}, \hat{\Sigma}^{-1}_H) \leq R_{EM}(\Sigma^{-1}, \hat{\Sigma}^{-1}_{EM}) \leq R_{EM}(\Sigma^{-1}, \hat{\Sigma}^{-1}_a)$$

et

$$R_H(\Sigma^{-1}, \hat{\Sigma}^{-1}_H) \leq R_H(\Sigma^{-1}, \hat{\Sigma}^{-1}_{EM}) \leq R_H(\Sigma^{-1}, \hat{\Sigma}^{-1}_a)$$

soient vérifiées.

Dans un autre travail, Haff [Haf79a] a montré que pour tout a tel que $n-p-5 \leq a \leq n-p-1$ l'estimateur $\hat{\Sigma}^{-1}_a$ domine l'estimateur sans biais $\hat{\Sigma}^{-1}_{n-p-1}$ sous le coût L_H dans (1.3.4), alors que sous le coût proportionnel à celui introduit par Efron et Morris, soit

$$L_{EM}\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \mathrm{tr}[\{\hat{\Sigma}^{-1} - \Sigma^{-1}\}^2 S], \tag{1.3.5}$$

c'est l'estimateur $\hat{\Sigma}^{-1}_{n-p-1}$ qui domine $\hat{\Sigma}^{-1}_a$ pour $a \neq n-p-1$. En outre, il a généralisé les résultats de son papier réalisé en 1977 [Haf77a] en proposant une large classe d'estimateurs de forme

$$\hat{\Sigma}^{-1}_{H_2} = [a + f(S)]\, S^{-1} + g(S)\, I_{p\times p}, \tag{1.3.6}$$

où $0 < a \leq n-p-1$ et où $f(S)$ et $g(S)$ sont des fonctions réelles. Il a donné des résultats de domination sur $a\,S^{-1}$ de l'estimateur $\hat{\Sigma}^{-1} = [a + f(S)]\, S^{-1}$ sous le coût L_H. Haff a remarqué aussi que tout estimateur de type (1.3.6) où la fonction $f(S)$ est nulle (par exemple celui dans (1.3.2)) peut être dominé par un estimateur de la forme (1.3.6) sous le coût L_H dans (1.3.4). Dans un autre papier [Haf79b], il a amélioré l'estimateur sans biais $\hat{\Sigma}^{-1}_{n-p-1}$ par l'estimateur usuel $a\,S^{-1}$ avec $n-p-3 \leq a \leq n-p-1$ sous le coût définit par

$$L_{H_1}\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \sum_{i\leq j} (\hat{\sigma}_{ij} - \sigma_{ij})^2\, q_{ij}, \tag{1.3.7}$$

où $\hat{\sigma}_{ij} = a\,S^{-1}_{ij}$ et les q_{ij} sont des poids arbitraires. Il a ainsi montré que sous le coût,

$$L_{H_2}\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \mathrm{tr}\{\hat{\Sigma}^{-1}\,\Sigma - I\}^2, \tag{1.3.8}$$

l'estimateur de la forme

$$\hat{\Sigma}^{-1}_{H_2} = a\,[S^{-1} + v\,t(v)\,I], \tag{1.3.9}$$

avec $a = (n-p-3)\,(n-p)/(n-1)$, $v = 1/\mathrm{tr}\{S\}$, $0 \leq t(v) \leq 2\,(p-1)/(n-p)$ et $t'(v) \leq 0$, domine l'estimateur usuel $a\,S^{-1}$ pour tout Σ. Aussi, d'une part, il a donné certaines conditions pour que cet estimateur améliore l'estimateur usuel $\hat{\Sigma}^{-1}_a$ sous le coût (1.3.7) et, d'autre part, l'estimateur optimal pour Σ^{-1} de la forme $a\,S^{-1}$ avec $a = (n-p-3)\,(n-p)/(n-1)$ sous le coût (1.3.8).

En 1987, Dey [Dey87a] a proposé un estimateur $\hat{\Sigma}^{-1}_D$ qui améliore $a\,S^{-1}$. Celui-ci est de la forme

$$\hat{\Sigma}^{-1}_D = a\,S^{-1} + \frac{\tau(\mathrm{tr}\{S^2\})}{\mathrm{tr}\{S^2\}}\,S, \tag{1.3.10}$$

où $\tau(\cdot)$ est une fonction positive, monotone croissante, absolument continue et majorée par $p-2$, sous le coût

$$L_0\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \mathrm{tr}[\{\hat{\Sigma}^{-1} - \Sigma^{-1}\}^2]. \tag{1.3.11}$$

En 1990, Dey, Ghosh et Srinivasan [DGS90] ont amélioré l'estimateur usuel $\hat{\Sigma}_a^{-1}$ par un estimateur, dit orthogonalement invariant, car de la forme

$$\hat{\Sigma}_{DGS}^{-1} = \hat{\Sigma}_a^{-1} + H_1 \Psi(L)\, H_1^\top, \tag{1.3.12}$$

où $L = \mathrm{diag}\big(l_1, ..., l_{n\wedge p}\big)$ est la matrice diagonale de dimension $(n \wedge p) \times (n \wedge p)$ des valeurs propres de S telle que $l_1 > \cdots > l_{n\wedge p} > 0$ et où H_1 est une matrice de vecteurs propres de S de dimension $p \times (n \wedge p)$ de sorte que $S = H_1 L H_1^\top$ et $\Psi(L) = \mathrm{diag}\big(\psi_1(L), \ldots, \psi_{n\wedge p}(L)\big)$ est une matrice diagonale de dimension $(n\wedge p)\times(n\wedge p)$. Il est intéressant de noter que $H_1^\top\, H_1 = I_{n\wedge p}$. Plus précisément, lorsque $p > n$, H_1 est une matrice semi-orthogonale de dimension $p\times n$ ($H_1^\top\, H_1 = I_n$) et, lorsque $p \leq n$, H_1 est une matrice orthogonale de dimension $p \times p$ ($H_1\, H_1^\top = H_1^\top\, H_1 = I_p$). Sous le coût basé sur les données (1.3.5), ces auteurs ont établi un estimateur sans biais de la différence des risques entre estimateurs $\hat{\Sigma}_{DGS}^{-1}$ et $\hat{\Sigma}_a^{-1}$. En s'appuyant sur la négativité de cet estimateur, ils ont obtenu de nouveaux estimateurs améliorés de Σ^{-1}. Parmi ceux-ci, citons $\hat{\Sigma}^{-1} = (n-p-1)\, S^{-1} + (p-1)\, p^{-1}\, |S|^{-1/p}\, S$ et $\hat{\Sigma}^{-1} = (n-p-1)\, S^{-1} + 2\,(p-1)\, S/\mathrm{tr}\{S^2\}$.

En 2006, dans le contexte gaussien, Tsukuma et Konno [TK06] se sont intéressés à l'estimation de la matrice de précision Σ^{-1} sous le coût quadratique (1.3.11). L'estimateur proposé est orthogonalement invariant de la forme générale

$$\hat{\Sigma}_{TK}^{-1} = H_1 \Psi(L)\, H_1^\top, \tag{1.3.13}$$

où H_1, L et Ψ sont des matrices comme dans l'estimateur (1.3.12), de sorte que Ψ exprimée sous la forme $\psi_i(L) = \delta_i(L)/l_i$ où $\delta_i(L)$ est une fonction de L pour tout $i = 1, 2, ..., p$. Ces auteurs ont montré que, si

$$\partial\, \delta_i(L)/\partial\, l_i \geq 0 \text{ pour tout } i = 1...p$$

et

$$n - p - 5 \leq \delta_p(L) \leq ... \leq \delta_1(L) \leq n - p - 1,$$

le risque de cet estimateur est plus petit que celui de l'estimateur sans biais $(n-p-1)\, S^{-1}$. Dans un contexte bayésien, ils ont considéré une classe d'estimateurs de la forme $\hat{\Sigma}^{-1} = a\, S^{-1} + G$ où a est une constante positive et G est une matrice symétrique de dimension $p \times p$. Puis ils ont donné des résultats de domination de ces estimateurs sur l'estimateur sans biais $\hat{\Sigma}_{n-p-1}^{-1}$. En outre, ils ont donné des conditions de domination de certains estimateurs de type Efron-Morris et de type Haff.

Lorsque p est assez grand et se rapproche de n, la matrice S peut avoir des valeurs propres proches de zéro, dans ce cas là, la matrice S^{-1} est soit instable ($p \leq n$), soit indéfinie ($p > n$). Kubokawa et Inoue dans leur article [KI14] se sont intéressés à l'estimation de Σ^{-1} dans cette situation. Ils ont proposé un estimateur ridge qui consiste à modifier l'estimateur $a\, S^{-1}$ en ajoutant à la matrice S un multiple d'une matrice qui dépend explicitement des observations afin de rendre son inverse plus stable. Ces auteurs ont ainsi considéré un estimateur de type ridge

$$\hat{\Sigma}_{KI}^{-1} = a\,(S + d\,\hat{\wedge})^{-1}, \tag{1.3.14}$$

où a et d sont des constantes dépendantes des dimensions n et p et où $\hat{\wedge}$ est une matrice définie positive de dimension $p \times p$, cette matrice étant basée sur les données (c'est-à-dire dépendant de S). Kubokawa et Inoue ont alors montré que si les deux limites $\lim\limits_{p\to\infty} \mathrm{tr}(\Sigma^i)/p$ avec $i = 1, 2$,

et $\lim_{p\to\infty} \mathrm{tr}((\wedge\,\Sigma^{-1})^j)/p$, où $j = 1, 2, 3$ et où $\wedge$ est une matrice symétrique de dimension $p \times p$ dépendant de Σ, existent et si

$\mathrm{tr}[((\hat{\wedge} - \wedge)\,\Sigma^{-1})^2]/p = O_p(n^{-1})$,
$\mathrm{E}[\mathrm{tr}[\Sigma^{-1}(\hat{\wedge} - \wedge)]]/p = O(n^{-1})$,
$\mathrm{E}[\mathrm{tr}[\Sigma^{-1} \wedge \Sigma^{-1}\,(\hat{\wedge} - \wedge)]]/p = O(n^{-1})$,
$d = o(p)$.

l'estimateur (1.3.14) améliore l'estimateur standard $a\,S^{-1}$ sous le coût (1.3.8). Ils ont donné quatre exemples pour $\hat{\wedge}$:

$$\hat{\wedge}_1 = a_1\,I_p \text{ avec } a_1 = \mathrm{tr}(S)/(np)$$
$$\hat{\wedge}_2 = (a_2/a_1)\,I \text{ avec } a_2 = \frac{1}{(n-1)(n+2)p}[\mathrm{tr}(S^2) - (\mathrm{tr}(S))^2/n]$$
$$\hat{\wedge}_3 = (b_1/b_2)\,I_p \text{ avec } b_1 = \frac{m}{p}\,\mathrm{tr}((S + a_1\,I)^{-1}) \textit{ et } b_2 = \frac{m^2}{p}\mathrm{tr}((S + a_1\,I)^{-2}) - \frac{p}{m}(b_1)^2.$$
$$\hat{\wedge}_4 = n^{-1}\mathrm{diag}(S_{11}, ..., S_{pp}) \textit{ ou } S_{ii} \text{ est le } i^{eme} \text{ element de la diagonale de la matrice } S.$$

1.3.2 Le contexte où $p > n$

Plusieurs recherches ont été développées sur les estimateurs d'une matrice de précision qui s'adaptent aux données de grande dimension où $p > n$. Nous en citons quelques-unes ci-dessous.

Rappelons que lorsque le nombre de variables est plus grand que le nombre d'observations ($p > n$), la matrice S est non inversible. Dans ce cas, en 2007, Srivastava [Sri07] a considéré $n\,S^+$, où S^+ est l'inverse de Moore-Penrose de S, comme estimateur de la matrice Σ^{-1}. En 2008, Kubokawa et Srivastava [KS08] ont étendu l'estimateur de Srivastava à ceux de la forme

$$\hat{\Sigma}_a^{-1} = a\,S^+, \tag{1.3.15}$$

avec a est une constante positive. Le travail de Kubokawa et Srivastava consiste à améliorer les estimateurs (1.3.15) sous les coûts de type Efron-Morris

$$L_k\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \mathrm{tr}[\{\hat{\Sigma}^{-1} - \Sigma^{-1}\}^2\,S^k] \tag{1.3.16}$$

pour $k = 0, 1, 2$. Ils ont considéré les estimateurs orthogonalement invariant de type (1.3.13) en utilisant la fonction $\Phi(L) = L\,\Psi(L)$ au lieu de la fonction $\Psi(L)$. Pour l'estimateur $a\,S^+$, ces auteurs ont déterminé trois valeurs particulières de a, soient $a_0 = p-n-3$, $a_1 = p-n-1$ et $a_2 = p$ qui correspondent aux estimateurs de référence sous les coûts L_0, L_1 et L_2 respectivement. Aussi, ils ont obtenu un estimateur sans biais de la différence de risques entre un estimateur alternatif et l'estimateur de référence $a_k\,S^+$, pour chacune des fonctions de coûts L_{KS}^k avec $k = 0, 1$ et 2. Ils ont établi, pour $p > n + 3$, $p > n + 1$ et $p > n$ respectivement, les conditions suivantes sur $\psi_i(l)$ pour que l'estimateur proposé $\hat{\Sigma}^{-1} = H_1\,L\,\Psi(L)\,H_1^\top$ domine l'estimateur de référence $a_k\,S^+$:

$$\text{pour } k = 0 : \sum_{i=1}^{n}\left\{\frac{\psi_i^2 - 2\,a_0\psi_i}{l_i^2} - 4\sum_{j>i}\frac{l_j\,\psi_i - l_i\,\psi_j}{l_i\,l_j(l_i - l_j)} - \frac{4}{l_i}\frac{\partial\,\psi_i}{\partial\,l_i}\right\} \leq \sum_{i=1}^{n}\left\{\frac{-a_0^2}{l_i} + \sum_{j>i}\frac{4\,a_0}{l_i\,l_j}\right\};$$

$$\text{pour } k = 1 : \sum_{i=1}^{n}\left\{\frac{\psi_i^2}{l_i} - 2\,a_1\,\frac{\psi_i}{l_i} - 4\sum_{j>i}\frac{\psi_i - \psi_j}{l_i - l_j} - 4\frac{\partial\,\psi_i}{\partial\,l_i}\right\} \leq -\sum_{i=1}^{n}\frac{a_1^2}{l_i};$$

$$\text{pour}\, k = 3 : \sum_{i=1}^{n} \left\{ \psi_i^2 - 2(p+n+1-2\,i)\,\psi_i + p^2 - 4\sum_{j>i} \frac{l_j(\psi_i - \psi_j)}{l_i - l_j} - 4 l_i \frac{\partial\,\psi_i}{\partial\, l_i} \right\} \leq 0.$$

Ainsi, ils ont fourni des exemples d'estimateurs améliorés qui sont des estimateurs de type Stein, Efron-Morris et Dey.

Kubokawa et Srivastava [Sri07] ont utilisé le cout L_2 pour construire un estimateur ridge par une méthode bayésienne empirique, cet estimateur domine l'estimateur optimal $p\,S^+$ et il est de la forme

$$\hat{\Sigma}_{KS}^{-1} = p\left(S + \frac{c}{\mathrm{tr}\{S^{-1}\}}\right)^{-1},$$

où $0 \leq c \leq 2\,(n-1)$. Kubokawa et Inoue [KI14] ont traité aussi le cas où $p > n$. Ces auteurs ont proposé un estimateur de type ridge de la forme (1.3.14) où $d = p$, soit

$$\hat{\Sigma}_{KI}^{-1} = a\,(S + p\,\hat{\wedge})^{-1}. \tag{1.3.17}$$

Ils ont fourni une approximation du risque associé au coût (1.3.8) de cet estimateur sous les conditions suivantes sur la matrice $\hat{\wedge}$:

$$\mathrm{tr}[X^t\,\hat{\wedge}^{-1}\,\Sigma\,\hat{\wedge}^{-1}\,X] = O_p(np),$$
$$\mathrm{tr}[X^t\,\hat{\wedge}^{-1}\,\Sigma\,\hat{\wedge}^{-1}\,\Sigma\,\hat{\wedge}^{-1}\,X] = O_p(np),$$

telle que $X = (X_1, ..., X_n)$ où $X_i \sim \mathcal{N}_p(0, \Sigma)$ pour tout $i = 1, ..., n$ et que $S = X\,X^t$. Comme exemple particulier de $\hat{S}$ ils ont considéré $\hat{S} = \hat{\lambda}\,I$ où $\hat{\lambda}$ est une fonction de S, soit

$$\hat{\Sigma}_{\hat{\lambda}}^{-1} = p\,(S + p\,\hat{\lambda}\,I_p)^{-1}. \tag{1.3.18}$$

Une approche basée sur la décomposition en valeur singulière de S leur a permis d'établir une approximation asymptotique de la différence de risque entre les estimateurs $\hat{\Sigma}_{\hat{\lambda}}^{-1}$ et $p\,S^+$. Ils ont montré, par des simulations, que $\hat{\Sigma}_{\hat{\lambda}}^{-1}$ dans (1.3.18) avec $\hat{\lambda} = \mathrm{tr}\{S\}/n\,p$ améliore légèrement l'estimateur $p\,S^+$.

1.3.3 Le cadre distributionnel elliptique

Dans une approche unifiée des deux cas où la matrice S est inversible ($p \leq n$) et singulière ($p > n$), Fourdrinier, Mezoued et Wells [FMW16] se sont intéressés à l'estimation de la matrice de précision Σ^{-1} pour une grande sous-classe de lois elliptiques. Le contexte distributionnel est le suivant. Soit $(X, U) = (X, U_1, ..., U_n)$ $n+1$ vecteurs aléatoires p-dimensionnels qui suivent une loi à symétrie elliptique de densité jointe définie par

$$\begin{aligned}(x, u) &\sim |\Sigma|^{-(n+1)/2}\, f((x-\theta)^T\,\Sigma^{-1}\,(x-\theta) + \sum_{i=1}^{n} U_i^T\,\Sigma^{-1}\,U_i),\\ &= |\Sigma|^{-(n+1)/2}\, f(\mathrm{tr}[\Sigma^{-1}\,(x-\theta)\,(x-\theta)^T + \Sigma^{-1}\,s]),\end{aligned} \tag{1.3.19}$$

où X et U_i avec $i = 1, ..., n$ sont des vecteurs de dimension $p \times 1$, θ est un vecteur moyenne $p \times 1$ inconnu, $S = U\,U^T$ est une matrice de dimension $p \times p$ et Σ est une matrice d'échelle inconnue.

Ces auteurs ont considéré la classe distributionnelle qui vérifie

$$c \leq \frac{F(t)}{f(t)} \leq b, \tag{1.3.20}$$

où $F(t) = \frac{1}{2}\int_t^\infty f(u)\,du$ et où b et c sont des constantes positives. Celle-ci contient le cas gaussien qui correspond à $F(t) = f(t) = 1$. Sous le coût quadratique (1.3.11), ils ont proposé comme grande classe d'estimateurs ceux de la forme

$$\hat{\Sigma}_G^{-1} = a\,S^+ + S\,G(S), \tag{1.3.21}$$

où a est une constante positive, $G(S)$ est une matrice de dimension $p \times p$ et S^+ est l'inverse de Moore-Penrose de S (lorsque $p \leq n$, S^+ est la matrice inverse S^{-1}). Étant donné qu'ils n'ont pas fourni l'estimateur sans biais de la différence de risque entre l'estimateur $\hat{\Sigma}_G^{-1}$ et l'estimateur $\hat{\Sigma}_a^{-1}$, leur condition de domination est basé sur la négativité d'un majorant de celui-ci.

Ces auteurs ont montré que l'estimateur orthogonalement invariant $\hat{\Sigma}_\phi^{-1} = H_1\,\phi(L)\,H_1^T$, où $\phi(L) = \mathrm{diag}\big(\phi_1(L), ..., \phi_{n\wedge p}(L)\big)$ tel que $\phi_k(L) = \delta_k(L)/l_k$ pour tout $k = 1, ..., n \wedge p$ et tel que $\phi_1(L) \geq \phi_2(L) \geq ... \geq \phi_{n\wedge p}(L)$, améliore $a\,S^+$ si

$$2\,b\,(|n-p| - 3) \leq \delta_k(L) \leq a.$$

Pour un estimateur de type Efron-Morris qui a la forme (1.3.21) avec $G(S) = B(t)/t\,S^+\,Q_0$ où $B(t)$ est une fonction réelle différentiable, $t = \mathrm{tr}\{S\}$ et $Q_0 = Q(Q^T\,Q)^{-1}\,Q^T$ où Q est une matrice de constantes de rang q, ils ont montré que, si $0 \leq a \leq (n-p-1)\,c - 2\,b$ et $0 \leq B(t) \leq 2\,\big((n-p-1)\,c - 2\,b\big)$, cet estimateur améliore $a\,S^+$.

1.4 Objectif et structure de la thèse

L'objectif dans cette thèse est d'améliorer, sous un coût basé sur les données (1.4.3), les estimateurs usuels dans (1.2.1) de la matrice de précision de lois qui sont un mélange de lois de Wishart. Ayant observé une matrice S semi-définie positive d'ordre p, nous proposons une approche unifiée des cas où S est inversible et S non inversible pour lesquels. Le contexte distributionnel est formalisé comme suit. Soit le modèle de mélange

$$\begin{cases} S \mid V \sim \mathcal{W}_p(n, V\,\Sigma) \\ V \sim \mathcal{H}(\cdot) \end{cases}, \tag{1.4.1}$$

où $\mathcal{W}_p(n, V\,\Sigma)$ représente la loi de Wishart à n degré de liberté et de matrice de variance-covariance $V\,\Sigma$ et où $\mathcal{H}(\cdot)$ est une loi de mélange définie sur $\mathbb{R}^+$. En conséquence, pour toute fonction g qui admet une espérance respectivement au modèle (1.4.1), on a

$$\mathrm{E}_\Sigma\,[g(S)] = \mathrm{E}_\mathcal{H}\left[\mathrm{E}_\Sigma\left[g(\tilde{S}\,V) \mid V\right]\right],$$

où E_Σ est l'espérance associée au modèle (1.4.1) et $\mathrm{E}_\mathcal{H}$ est celle associée à la loi de mélange $\mathcal{H}$. Dans le cas où la loi $\mathcal{H}$ est la loi de Dirac en 1, le modèle (1.4.1) est réduit au modèle de Wishart, c'est à dire

$$S \sim \mathcal{W}_p(n,\,\Sigma). \tag{1.4.2}$$

Il est intéressant de noter que la loi de S peut aussi être vue comme la loi de $V\,\tilde{S}$, où V est une variable aléatoire positive de loi $\mathcal{H}$ et $\tilde{S}$ est une matrice aléatoire indépendante de V et suivant une loi de Wishart $\mathcal{W}_p(n, \Sigma)$.

Les mélanges de Wishart sont apparus dans bon nombre de travaux dans différentes disciplines. Ainsi en biophysique Jian, Vemuri et Ozaeslan [JVO09], les ont utilisé, en 2009, pour modéliser un signal de résonance magnétique. En 2009 Yang, Yang, Yan, Song et Xia [YYY+16]

les ont aussi envisagé dans la détection des mouvements issus des images du radar à synthèse d'ouverture polarimétrique (PolSAR). Dans l'imagerie cérébrale (neuroimaging), Nielsen, Madsen, Schmidt et Morup [NMSM17] ont utilisé un modèle de mélange de Wishart pour modéliser la dynamique de la connectivité fonctionnelle dans un contexte bayésien. En 2011, Haff, Kim, Koo et Richards [HKKR11] ont développé une théorie sur l'estimation de la distribution de mélange pour un modèle de la forme (1.4.1) où V est une matrice aléatoire. En fournissant un estimateur non paramétrique, ils ont illustré leur résultat avec une application sur la volatilité stochastique en finance.

Dans cette thèse, nous nous intéressons à l'estimation de la matrice Σ^{-1} de dimension $p \times p$ sous des coûts de type Efron-Morris

$$L_k\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \mathrm{tr}[\{\hat{\Sigma}^{-1} - \Sigma^{-1}\}^2 \, S^k], \tag{1.4.3}$$

pour $k = 1, 2, 3 \ldots$ et sous les risques associés

$$R_k(\Sigma^{-1}, \hat{\Sigma}^{-1}) = \mathrm{E}_\Sigma[L_k\{\Sigma^{-1}, \hat{\Sigma}^{-1}\}], \tag{1.4.4}$$

où E_Σ représente l'espérance respectivement au modèle (1.4.1).

Notre objectif est d'améliorer les estimateurs usuels de la forme

$$\hat{\Sigma}_a^{-1} = a \, S^+,$$

où a est une constante positive et S^+ est l'inverse de Moore Penrose (dans le cas où S est inversible, soit le cas $p \leq n$, S^+ est la matrice inverse S^{-1}). Nous proposons une grande classe d'estimateurs de la forme

$$\hat{\Sigma}_{a,c}^{-1} = a \, S^+ + c \, S \, G(S),$$

où c est une constante positive et $G(S)$ est une fonction matricielle $p \times p$ homogène de degré α, c'est à dire, telle que

$$\exists \, \alpha \in \mathbb{R} \quad G(V \, S) = V^\alpha \, G(S) \, .$$

Dans le chapitre 2 de cette thèse, on donne quelques notions d'algèbre des matrices que nous allons utiliser dans ce travail. Nous rappelons aussi les outils de base de la théorie de décision et certaines lois matricielles, en particulier, la loi de Wishart. Dans la dernière partie de ce chapitre, nous citons les identités de type Stein-Haff qui existent dans la littérature, notamment celle établie récemment par Haddouche, Fourdrinier et Mezoued [FHM21] que nous allons utiliser pour fournir des estimateurs améliorés de la matrice de précision.

Le chapitre 3 contient les contributions théoriques de nos travaux de thèse dans lesquels, sous les coûts (1.4.3), nous proposons des estimateurs de la forme générale $\hat{\Sigma}_{a,c}^{-1} = a \, S^+ + c \, S \, G(S)$ qui améliorent l'estimateur usuel $a \, S^+$. Dans ce but, nous fournissons, dans le corollaire (3.2.1), un estimateur sans biais de la différence de risque entre ces deux estimateurs. On se basant sur la minimisation des risques $R_k(\Sigma^{-1}, \hat{\Sigma}_a^{-1})$ où $k = 1, 2, 3 \ldots$, nous présentons, dans la proposition 3.3.1, des estimateurs de référence qui correspondent à des valeurs particulières du paramètre a. Par ailleurs, nous proposons un autre type d'estimateurs de la forme $\hat{\Sigma}_{a,c,r}^{-1} = a \, S^+ + c \, r\big(\mathrm{tr}\{S\}\big) \, S \, G(S)$ et nous établissons des conditions sur c et sur la fonction r afin que ceux-ci dominent les estimateurs $a \, S^+$. Enfin, titre exemple, nous donnons des estimateurs de type Haff, Dey et Efron-Morris et leurs conditions de domination sur les estimateurs de référence $a \, S^+$. Celles-ci sont données dans les propositions 3.5.2, 3.5.1 et 3.6.1.

Le chapitre 4 se focalise sur l'évaluation des résultats de la thèse par des simulations numériques.

Dans le chapitre 5, nous montrons que les estimateurs homogènes de degré -2 sont les plus importants en considérant les estimateurs de James-Stein de vecteur multidimensionnel de moyennes où le degré d'homogénéité de terme correcteur est différent de -2.

La conclusion de ce manuscrit et des perspectives sur l'estimation de la matrice de précision sont données dans le chapitre 6.

CHAPITRE 2

RAPPELS ET PRÉLIMINAIRES

Sommaire

2.1 Algèbre des matrices **11**
2.1.1 Matrice singulière et l'inverse de Moore-Penrose 12
2.1.2 Décomposition en valeur singulière 12
2.2 Théorie de la décision **13**
2.3 Lois multidimensionnelles **13**
2.3.1 Loi normale multidimensionnelle 13
2.3.2 Lois à symétrie sphérique 14
2.3.3 Lois à symétrie elliptique 16
2.4 Lois matricielles **18**
2.4.1 Loi normale matricielle 18
2.4.2 Loi de Wishart 18
2.4.3 Mélange de lois de Wishart 19
2.5 Des identités de type Stein-Haff **20**

Nous introduisons dans ce chapitre les notions de bases que nous utilisons dans ce travail en commençant par l'algèbre des matrices, puis nous abordons les lois multidimensionnelles et matricielles et certaines de leurs propriétés. Enfin, nous fournissons les différentes identités de Stein-Haff qui ont été obtenues dans divers contextes.

2.1 Algèbre des matrices

L'algèbre des matrices est un outil important dans l'estimation par rétrécissement (shrinkage estimation) de la matrice de précision et de tout paramètre qui a la forme d'une matrice. Dans cette partie, nous rappelons les notions que nous avons utilisé pour la réalisation de ce travail en se référant au livre de Harville [Har08] et celui de Tsukuma et Konno [TK20].

2.1.1 Matrice singulière et l'inverse de Moore-Penrose

Si le déterminant d'une matrice A est nul alors l'inverse de A n'existe pas et on dit qu'elle est non inversible ou singulière. Cependant, plusieurs inverses généralisés G peuvent être définis au travers de la relation

$$A\,G\,A = A.$$

Parmi eux il existe un unique G, satisfaisant certains propriétés que nous allons citer ci-dessous, appelé l'inverse de Moore-Penrose, on le note A^+ .

Définition 2.1.1

Pour une matrice A de dimension $n \times p$, une matrice A^+ de dimension $p \times n$ est appelée l'inverse de Moore-Penrose de A si elle vérifie les égalités suivantes

(1) $A\,A^+\,A = A$,

(2) $A^+\,A\,A^+ = A^+$,

(3) $(A\,A^+)^T = A\,A^+$,

(4) $(A^+\,A)^T = A^+\,A$.

Il est clair que pour toute matrice inversible A de dimension $p \times p$,

$$A^+ = A^{-1}, \tag{2.1.1}$$

puisque $A\,A^{-1} = A^{-1}\,A = I_p$, donc A^{-1} vérifie bien les conditions de l'inverse de Moore-Penrose citées ci-dessus.

Proposition 2.1.1

Dans le cas d'une matrice symétrique S d'ordre p on a

(1) $S\,S^+\,S = S\,S\,S^+ = S$,

(2) $S^+\,S\,S^+ = S^+\,S^+\,S = S^+$,

(3) $S^+\,S = S\,S^+$.

On en déduit que

(1) $S(I_p - S\,S^+) = 0$,

(2) $S^+(I_p - S\,S^+) = 0$,

(3) $S\,S^+(I_p - S\,S^+) = 0$.

2.1.2 Décomposition en valeur singulière

Pour toute matrice A de dimension $n \times n$, il existe une matrice inversible P de même dimension que A telle que

$$A = P\,D\,P^{-1},$$

où D est une matrice $n \times n$ diagonale telle que ses éléments diagonaux, appelés valeurs propres de A, vérifient $d_1 \geq d_2, ..., \geq d_n$, la i^{me} colonne de P est le i^{me} vecteur propre associé au i^{me} élément diagonal de D.

Si A est une matrice $n \times m$ symétrique définie positive alors il existe une unique matrice D diagonale $m \times m$ d'éléments positifs et ordonnés de manière décroissante i.e $d_1 \geq d_2, ..., \geq d_n \geq 0$ et une matrice P vérifiant : $P\,P^{\mathrm{T}} = P^{\mathrm{T}}\,P = I_m$ telle que

$$A = P\,D\,P^{\mathrm{T}}.$$

Théorème 2.1.1 (Diagonalisation orthogonale)

Si A est une matrice réelle de dimension $m \times n$ ($m \leq n$) alors il existe une matrice orthogonale H de dimension $m \times m$ et une matrice orthogonale Q de dimension $n \times n$ telle que

$$H\,A\,Q^\top = \begin{bmatrix} d_1 & & 0 & . & \\ & . & & . & 0 \\ & & . & . & \\ 0 & & d_m & . & \end{bmatrix},$$

où $d_i \geq 0$ pour tout $i = 1, ..., m$.

2.2 Théorie de la décision

Soit X une variable aléatoire à valeurs dans un espace $\mathcal{X}$ dont la loi dépend d'un paramètre inconnu θ qui appartient à un ensemble (dit des paramètres) Θ ; soit P_θ cette loi. Tout estimateur de θ est une variable $\hat{\theta}$ définie sur $\mathcal{X}$ telle que $\hat{\theta} = \hat{\theta}(X)$ soit à valeurs dans Θ.

Pour évaluer un tel estimateur $\hat{\theta}$, on se fixe une fonction L de $\Theta \times \Theta$ dans $\mathcal{R}_+$ qui à toute observation x de $\mathcal{X}$ et à tout θ dans Θ associe $L(\theta, \hat{\theta}(x))$ qui est le coût encouru par l'estimation $\hat{\theta}(x)$ alors que θ est la valeur du paramètre inconnu. Comme ce coût dépend de l'observation x il ne suffit pas à évaluer l'estimateur $\hat{\theta}$. En effet $\hat{\theta}(x)$ peut être très différent de $\hat{\theta}(x')$ pour $x \neq x'$. Une telle évaluation doit reposer sur toutes les observations possibles, soit sur toute les réalisations de X. Une approche classique est de prendre l'espérance par rapport à la loi P_θ du coût, soit de considérer le risque en θ de $\hat{\theta}$ défini par

$$R(\theta, \hat{\theta}) = \mathrm{E}_\theta(L(\theta, \hat{\theta})) = \mathrm{E}_\theta(L(\theta, \hat{\theta}(X))) = \int_{\mathcal{X}} L(\theta, \hat{\theta}(x))\, d\, P_\theta(x).$$

Ici E_θ désigne l'espérance par rapport à la loi P_θ.

Le choix entre deux estimateurs $\hat{\theta}_1$ et $\hat{\theta}_2$ est guidé par les considérations suivantes. On dit que $\hat{\theta}_1$ est au moins aussi bon que $\hat{\theta}_2$ si

$$\forall \theta \in \Theta, \qquad R(\theta, \hat{\theta}_1) \leq R(\theta, \hat{\theta}_2), \tag{2.2.1}$$

et que $\hat{\theta}_1$ est meilleur que $\hat{\theta}_2$ (ou domine θ_2) si, de plus, l'inégalité (2.2.1) est stricte pour au moins une valeur θ, soit

$$\exists \theta \in \Theta, \qquad R(\theta, \hat{\theta}_1) < R(\theta, \hat{\theta}_2).$$

Un estimateur $\hat{\theta}$ est dit admissible s'il n'existe pas d'estimateur qui le domine. Ils est dit inadmissible s'il est dominé par un autre estimateur $\hat{\theta}'$.

Si Θ est un espace où une notion d'espérance peut être définie, on dit qu'un estimateur θ' est sans biais si

$$\forall \theta \in \Theta, \qquad E(\hat{\theta}) = \theta.$$

2.3 Lois multidimensionnelles

2.3.1 Loi normale multidimensionnelle

Définition 2.3.1

On dit qu'un vecteur aléatoire $X = (X_1, ..., X_p)^T$ suit une loi normale p-multidimensionnelle de moyenne $\mu \in \mathbb{R}^p$ et de matrice de covariance inversible Σ de dimension $p \times p$, si sa fonction

densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^p$ *est*

$$x \sim \frac{1}{(2\pi)^{p/2}\,|\Sigma|^{1/2}}\,\exp\left(\frac{1}{2}\,(x-\mu)^T\,\Sigma^{-1}\,(x-\mu)\right). \tag{2.3.1}$$

On note $X \sim \mathcal{N}_p(\mu, \Sigma)$.

Nous présentons maintenant quelques propriétés de la loi normale multivariée.

Propriété 2.3.1

Soit $X = (X_1, ..., X_p)^\top$ *un vecteur aléatoire qui suit une loi normale multidimensionnelle* $\mathcal{N}_p(\mu, \Sigma)$ *de moyenne* $\mu \in \mathbb{R}^p$ *et de matrice de covariance* $\Sigma \in \mathbb{S}_p$. *Alors*

(1) $\mu = (\mathrm{E}\,[X_1]\,, ..., [X_p])^\top$ *et* $\Sigma = \mathrm{E}\left[(X-\mu)\,\Sigma\,(X-\mu)^\top\right]$.

(2) *Pour tout* $t \in \mathbb{R}^p$, *la fonction caractéristique de* X *est* $\phi_X(t) = \exp\left(i\,\mu^\top\,t - \frac{1}{2}\,t^\top\,\Sigma\,t\right)$.

(3) *Pour tout* $\alpha \in \mathbb{R}^p$, *la variable* $\alpha^\top X$ *suit une loi normale unidimensionnelle.*

(4) *Pour toute matrice constante* $A_{n\times p}$ *à coefficients constants et tout vecteur* b *de dimension* $p, Y = A\,X + b \sim \mathcal{N}_p(A\,\mu + b, A\,\Sigma\,A^\top)$.

(5) *La loi marginale de* k *composantes de* X *est une loi normale* k*- multidimensionnelle.*

(6) *Si les variables* $X_1, ..., X_n$ *sont indépendantes et chacune suit une loi normale multivariée* $\mathcal{N}_p(\mu_i, \Sigma_i)$ *alors, pour tout* α_i, *où* $i = 1, ..., N$, *on a*

$$\sum_{i=1}^{N} \alpha_i\,X_i \sim \mathcal{N}_p\left(\sum_{i=1}^{N} \alpha_i\,\mu_i, \sum_{i=1}^{N} \alpha_i^2\,\Sigma_i\right).$$

Pour les démonstrations des propriétés citées ci-dessus et ou d'autres propriétés de la loi normale multivariée on peut consulter l'ouvrage de Muirhead [Mui82].

Dans ce qui suit, nous rappelons la notion de loi à symétrie sphérique qui généralise la loi normale $\mathcal{N}_p(\mu, \sigma^2\,I_p)$ en se référant au livre de Fourdrinier, Strawderman et Wells [FSW18] et celui de Fourdrinier [Fou02]. Plusieurs définitions équivalentes à celle que nous allons donner se trouvent dans d'autres ouvrages, comme celui de Frang et Zhang [FZ90].

2.3.2 Lois à symétrie sphérique

Nous commençons tout d'abord par définir la transformation orthogonale qui est une propriété très importante pour introduire les lois à symétrie sphérique.

Définition 2.3.2 (Transformation orthogonale)

On dit qu'une transformation T *de* $\mathbb{R}^p$ *est orthogonale si, pour tout* $x = (x_1, .., x_P) \in \mathbb{R}^p$ *et pour tout* $y = (y_1, .., y_p) \in \mathbb{R}^p$,

$$< Tx, Ty > = < x, y >,$$

où $< x, y > = \sum_{i=1}^{p} x_i\,y_i$ *est le produit scalaire classique de* x *et* y.

Définition 2.3.3 (Invariance orthogonale et loi radiale)

Un vecteur aléatoire $X \in \mathbb{R}^p$ *est invariant par transformation orthogonale* T *si* $T(X) \overset{d}{=} X$, *c'est-à-dire si la loi de* $T(X)$ *est la même que la loi* P *de* X. *On dit*

alors que la loi P est radiale.

Définition 2.3.4

Soit X un vecteur aléatoire de $\mathbb{R}^p$ et θ un vecteur de $\mathbb{R}^p$. On dit que X suit une loi à symétrie sphérique autour de θ, et on note $X \sim \mathcal{SS}(\theta)$, si la loi du vecteur $X - \theta$ est radiale.

Propriété 2.3.2

Un vecteur aléatoire $X \in \mathbb{R}^p$ a une loi à symétrie sphérique autour de θ si et seulement si celui-ci admet la représentation stochastique

$$X = \theta + R\,U, \tag{2.3.2}$$

où $R = ||X - R|| \geq 0$ et où $U = (X - R)/||X - R||$ est une variable aléatoire de loi uniforme $\mathcal{U}$ sur la sphère unité

$$S = \{x \in \mathbb{R}, /||x|| = 1\}, \tag{2.3.3}$$

les variables R et U étant indépendantes.

Proposition 2.3.1

Si un vecteur $X \in \mathbb{R}^p$ suit une loi à symétrie sphérique autour de $\theta \in \mathbb{R}^p$ alors, pour toute transformation orthogonale T, la variable $T\,X$ suit une loi à symétrie sphérique autour de $T\,\theta$. De plus, $X - \theta$ et $T\,X - T\,\theta$ ont la même loi.

Dans le théorème suivant, nous donnons les densités d'une variable à symétrie sphérique et du de rayon $||X - \theta||$ puis nous exprimons leur lien avec la fonction générative d'une loi radiale.

Théorème 2.3.1

Soit X un vecteur aléatoire de $\mathbb{R}^p$ qui suit une loi à symétrie sphérique autour de $\theta \in \mathbb{R}^p$. Les deux affirmations suivantes sont équivalentes :

(1) *$R = \|X - \theta\|$ a une densité $h(\cdot)$ par rapport à la mesure de Lebesgue dans $\mathbb{R}_+$.*

(2) *X a une densité $f(\cdot)$ par rapport à la mesure de Lebesgue dans $\mathbb{R}^p$.*

De plus, si l'une des deux affirmations ci-dessus est vérifiée, alors il existe une fonction $g(\cdot)$ de $\mathbb{R}_+$ dans $\mathbb{R}_+$, tel que

$$f(x) = g\left(\|x - \theta\|^2\right) \qquad \textit{presque partout,}$$

et

$$h(r) = \frac{2\,\pi^{p/2}}{\Gamma(p/2)}\, r^{p-1}\, g(r^2) \qquad \textit{presque partout,}$$

La fonction $g(\cdot)$ est appelée fonction générative et $h(\cdot)$ est la densité radiale.

Nous présentons maintenant les conditions d'existences de la moyenne et de la matrice de covariance d'une variable à symétrie sphérique (*cf.* Fourdrinier, Strawderman et Wells [FSW18]).

Théorème 2.3.2

Soit $X \in \mathbb{R}^p$ un vecteur aléatoire de loi à symétrie sphérique autour de $\theta \in \mathbb{R}^p$. Alors, la moyenne de X existe si et seulement si la moyenne de $R = \|X - \theta\|$ existe. Dans ce cas, on a

$$E(X) = \theta.$$

La matrice de covariance de X existe si et seulement si l'espérance $E(R^2)$ est finie. Dans ce cas, on a

$$Cov(X) = \frac{E(R^2)}{p} I_p\,.$$

Nous donnons ci-dessus quelques exemples de loi à symétrie sphérique de covariance $Cov(X) = \sigma^2\, I_p$ et leurs densités correspondantes. Nous allons les présenter dans le tableau suivant.

Loi	Densité de probabilité
Normale multivariée $\mathcal{N}_p(\theta, \sigma^2)$	$\frac{1}{(2\pi\sigma^2)^{\frac{p}{2}}} \exp\left(\frac{-\lVert x-\theta\rVert^2}{2\sigma^2}\right)$
Student $t_p(\theta, \sigma^2, \nu)$	$\frac{\Gamma(\frac{p+\nu}{2})}{(\pi\sigma^2\nu)^{\frac{p}{2}}\Gamma(\frac{\nu}{2})} \left(1+\frac{\lVert x-\theta\rVert^2}{\nu\sigma^2}\right)^{\frac{-(\nu+p)}{2}}$ avec $\nu \geq 1$
Kotz $\mathcal{K}_p(\theta, \sigma^2, N, r)$	$\frac{\pi^{\frac{-p}{2}}\Gamma(\frac{p}{2})\, r^{\frac{2N+p-2}{2}}}{\sigma^{2N+p-1}\Gamma\left(\frac{2N+p-2}{2}\right)} \lVert x-\theta\rVert^{2(N-1)} \exp\left(-r\frac{\lVert x-\theta\rVert^2}{\sigma^2}\right)$ avec $N>\frac{2-}{2}$
Puissance exponentielle $\mathcal{PE}_p(\theta, \sigma^2, b)$	$\frac{p\,\Gamma(\frac{p}{2})}{(\pi\sigma^2)^{\frac{p}{2}}\, 2^{1+\frac{1}{2b}}\,\Gamma(1+\frac{p}{2b})} \exp\left(-\frac{1}{2}\left(\frac{\lVert x-\theta\rVert^2}{\sigma^2}\right)^b\right)$ avec $b>0$
Logistique $\mathcal{LOG}_p(\theta, \sigma^2)$	$\frac{\left(\sum_{j=1}^{\infty}(-1)^{j-1}j^{1-\frac{p}{2}}\right)^{-1}}{(2\pi\sigma^2)^{\frac{p}{2}}} \frac{\exp\left(\frac{-\lVert x-\theta\rVert^2}{2\sigma^2}\right)}{\left(1+\exp\left(\frac{-\lVert x-\theta\rVert^2}{2\sigma^2}\right)\right)^2}$

TABLEAU 2.1 – Exemples de densités à symétrie sphérique.

2.3.3 Lois à symétrie elliptique

Les lois à symétrie elliptique peuvent être définies au travers des fonctions caractéristiques de la manière que Fang et Zang ont procédé dans leur travail [FZ90], ou bien, en utilisant une transformation linéaire des lois à symétrie sphérique. Pour une matrice d'échelle Σ^{-1} inversible, Fourdrinier, Strawderman et Wells [FSW18] les ont définies à travers de la propriété d'invariance $\mathbf{\Sigma}^{-1}-$orthogonale des vecteurs aléatoires.

Rappelons que le produit scalaire associé à toute matrice définie positive Γ est défini, pour tout $x = (x_1, ..., x_p) \in \mathbb{R}^p$ et tout $y = (y_1, ..., y_p) \in \mathbb{R}^p$ par

$$< x, y >_\Gamma = x^{\mathrm{T}}\,\Gamma\, y = \sum_{1\leq i,j\leq p}^{p} x_i\,\Gamma_{i,j}\, y_j,$$

où $\Gamma_{i,j}$ est le terme d'ordre (i, j) de Γ. Le choix de $\Gamma = \Sigma^{-1}$ est motivé par le fait que Σ^{-1} intervient naturellement dans l'expression de la densité des lois.

Définition 2.3.5 (Transformation $\mathbf{\Sigma}^{-1}-$orthogonale)
Soit Σ une matrice $p \times p$ symétrique et définie positive. Une transformation T est dite $\Sigma^{-1}-$orthogonale si, pour tout $x \in \mathbb{R}^p$ et pour tout $y \in \mathbb{R}^p$

$$< Tx, Ty >_{\Sigma^{-1}} \;=\; x^{\mathrm{T}}\, T^{\mathrm{T}}\, \Sigma^{-1}\, T\, y = < x, y >_{\Sigma^{-1}},$$

c'est à dire si $T^{\mathrm{T}}\,\Sigma^{-1}\,T = \;\Sigma^{-1}$.

Définition 2.3.6 (Invariance $\mathbf{\Sigma}^{-1}-$orthogonale)
Soit Σ une matrice $p \times p$ définie positive. Un vecteur aléatoire $Y \in \mathbb{R}^p$ est dit invariant par transformation $\Sigma^{-1}-$orthogonale T, si la loi de TY est égale à la loi de Y.

Définition 2.3.7 (Loi à symétrie elliptique)
Soit Σ une matrice $p \times p$ symétrique et définie positive. Un vecteur aléatoire $X \in \mathbb{R}^p$ a une loi à symétrie elliptique autour de $\theta \in \mathbb{R}^p$ et est de paramètre d'échelle Σ si le vecteur $X - \theta$ est invariant par toute transformation $\Sigma^{-1}-$ orthogonale T. On note alors

$$X \sim ES(\theta, \Sigma)\,.$$

Propriété 2.3.3 (Représentation stochastique)

Un vecteur aléatoire $X \in \mathbb{R}^p$ a une loi à symétrie elliptique autour de θ si et seulement si celui-ci admet la représentation stochastique

$$X = R\,U,$$

où R et où U sont des variables aléatoires indépendantes, elles sont définies respectivement par $R = \|X-\theta\|_{\Sigma^{-1}}$ et $U = X-\theta/\|X-\theta\|_{\Sigma^{-1}}$ où R est une variable aléatoire positive et U est une variable qui suit une loi uniforme sur le $\Sigma^{-1}-$ellipsoïde de rayon unitaire dans $\mathbb{R}^p$, définie par

$$\mathcal{S}_1^{\Sigma^{-1}} = \{x \in \mathbb{R}^p; \|x-\theta\|^2_{\Sigma^{-1}} = 1\}.$$

Nous exprimons dans le théorème suivant le lien entre les lois des variables R et X qui suit une loi à symétrie elliptique autour de $\theta \in \mathbb{R}^p$.

Théorème 2.3.3

Soit $X \in \mathbb{R}^p$ un vecteur aléatoire de loi à symétrie elliptique autour de $\theta \in \mathbb{R}^p$ et de paramètre d'échelle Σ. Alors les deux énoncés suivants sont équivalents :

(1) *$R = \|X-\theta\|_{\Sigma^{-1}}$ a une densité $h(\cdot)$ par rapport à la mesure de Lebesgue dans $\mathbb{R}_+$.*

(2) *X admet une densité $f(\cdot)$ par rapport à la mesure de Lebesgue dans $\mathbb{R}^p$.*

De plus, si l'une ou l'autre des conditions (1) ou (2) est vérifiée, alors

$$f(x) = |\Sigma|^{-1/2}\, g\left(\|x-\theta\|^2_{\Sigma^{-1}}\right) \qquad \textit{presque partout,}$$

et la densité $h(\cdot)$ s'exprime par

$$h(r) = \frac{2\,\pi^{p/2}}{\Gamma(p/2)}\,|\Sigma|^{-1/2}\, r^{p-1}\, g(r^2) \qquad \textit{presque partout,}$$

où $g(\cdot)$ est appelée fonction générative et $h(\cdot)$ la densité radiale.

Maintenant, nous considérons la moyenne et la matrice de covariance d'une variable à symétrie elliptique ainsi que leurs conditions d'existences (*cf.* Fourdrinier, Strawderman et Wells [FSW18]).

Théorème 2.3.4

Soit $X \in \mathbb{R}^p$ un vecteur aléatoire de loi à symétrie elliptique autour de $\theta \in \mathbb{R}^p$. Alors, la moyenne de X existe si et seulement si la moyenne de $R = \|X-\theta\|_{\Sigma^{-1}}$ existe. Dans ce cas, on note

$$E(X) = \theta.$$

La matrice de covariance de X existe si et seulement si l'espérance $E(R^2)$ est fini. Dans ce cas, elle est donnée par

$$Cov(X) = \frac{E(R^2)}{p}\Sigma\,.$$

Notons que dans le cas gaussien, R^2 suit une loi du chi-deux à p degrés de liberté dont l'espérance est égale à p et dont la matrice de covariance de X coïncide avec la matrice d'échelle Σ.

2.4 Lois matricielles

2.4.1 Loi normale matricielle

Définition 2.4.1

*Soit X une matrice aléatoire de dimension $n \times p$ telle que $\mathrm{vec}(X^{\mathrm{T}})$ suive une loi normale np-dimensionnelle de moyenne $\mathrm{vec}(M^{\mathrm{T}})$ et de matrice de covariance $\Omega \otimes \Sigma$ **i.e***

$$\mathrm{vec}(X^{\mathrm{T}}) \sim N_{np}(\mathrm{vec}(M^{\mathrm{T}}), \Omega \otimes \Sigma), \tag{2.4.1}$$

où Ω et Σ sont des matrices symétriques définies positives de dimension $n \times n$ et $p \times p$ respectivement. On di alors que X suit une loi normale matricielle de moyenne M et de covariance $\Omega \otimes \Sigma$ et on note

$$X \sim N_{n\times p}(M, \Omega \otimes \Sigma). \tag{2.4.2}$$

La densité de probabilité de cette loi est donnée comme suit.

Proposition 2.4.1

Soit $X \sim N_{n\times p}(M, \Omega \otimes \Sigma)$. Alors X admet pour densité

$$\frac{1}{(2\pi)^{np/2}\,|\Omega|^{p/2}|\Sigma|^{n/2}} \exp\left(-\frac{1}{2}\,\mathrm{tr}\,\Omega^{-1}\,(X-M)\,\Sigma^{-1}\,(X-M)^{\mathrm{T}}\right). \tag{2.4.3}$$

Dans la proposition suivante, nous donnons quelques propriétés de cette loi.

Proposition 2.4.2

Soient $X = (X_{i,j}) \sim N_{n\times p}(M, \Omega \otimes \Sigma)$ avec $M = (\mu_{ij})$, $\Omega = (w_{ij})$ et $\Sigma = (\sigma_{ij})_{1\le i\le p, 1\le j\le p}$, A et B des matrices de constantes de dimension $p \times p$ et $n \times n$ respectivement, $1 \le k \le n$ et $0 \le l \le p$ alors

(1) $X^{\mathrm{T}} = (X_{i,j}) \sim N_{n\times p}(M^{\mathrm{T}}, \Omega \otimes \Sigma)$,
(2) $\mathrm{E}(X_{ij}) = \mu_{ij}, \mathrm{E}(X_{ij}\,X_{kl}) = w_{ik}\,\sigma_{jl} + \mu_{ij}\,\mu_{kl}$,
(3) $\mathrm{E}(X) = M$,
(4) $\mathrm{E}[X\,A\,X^{\mathrm{T}}] = (\mathrm{tr}\,A\,\Sigma)\,\Omega + M\,A\,M^{\mathrm{T}}$,
(5) $\mathrm{E}[X^{\mathrm{T}}\,B\,X] = (\mathrm{tr}\,B\,\Omega)\,\Sigma + M^{\mathrm{T}}\,B\,M$

Tsukuma et Konno [TK20] ont fourni les preuves de toutes ces propositions. Gupta et Nagar [GN99] en ont donné d'autres.

2.4.2 Loi de Wishart

La loi de Wishart est la loi de la matrice de covariance empirique à entrées gaussiennes. Elle a été introduite pour la première fois dans le travail de Wishart [Wis28] en 1928. Cette loi joue un rôle important dans les analyses multivariées. Elle est apparue dans plusieurs applications de modèles graphiques [LM07], systèmes de communication [MA07] et dans d'autres domaines comme la biologie, la physique nucléaire, la finance, l'économie ainsi que l'analyse discriminante... (Janik et Nowak [JN03], Letac et Massam [LM04] ont mentionné divers travaux publiés dans les domaines cités précédemment). Dans ce qui suit, on se réfère à Tsukuma et Konno [TK20] et Muirhead [Mui82] pour définir la densité de la loi de Wishart et ses principales propriétés. Pour plus de détails, voir [Mor07] et [BB99].

Définition 2.4.2

Soit $X \sim \mathcal{N}_{n\times p}(0_{n\times p}, I_n \otimes \Sigma)$. La matrice $S = X^T X$ est appelée matrice de Wishart et est de rang $n \wedge p$ et de loi de Wishart à n degré de liberté et de matrice de covariance Σ. On note cette loi $\mathcal{W}_p(n, \Sigma)$.

Si $p \leq n$, la matrice de Wishart S est définie positive avec une probabilité égale à 1 et sa loi admet une densité par rapport à la mesure de Lebesgue donnée dans la proposition (2.4.3).

Si $p > n$, la matrice de Wishart S est singulière semi-définie positive mais pas définie positive. Pour sa loi, on parle de la loi de Wishart est singulière. Notons que cette loi ne possède pas de densité par rapport à la mesure de Lebesgue.

Définition 2.4.3 (fonction gamma multivariée)

Pour toute matrice $p\times p$ définie positive S et tout $a > (p-1)/2$, la fonction gamma multivariée $\Gamma_p(a)$ est définie par

$$\Gamma_p(a) = \int_{S>0} |S|^{a-(p+1)/2} \exp\{-\operatorname{tr} S\}(dS), \tag{2.4.4}$$

Lorsque $p = 1$, c'est la fonction gamma usuelle

$$\Gamma(a) = \Gamma_1(a) = \int_0^\infty s^{a-1} \exp\{s\}\, ds.$$

La proposition suivante donne la densité de la loi de Wishart $\mathcal{W}_p(n, \Sigma)$ dans le cas où la matrice S est inversible ($p < n$).

Proposition 2.4.3

Dans le cas où S est inversible soit $p < n$, la matrice S est définie positive et sa densité de probabilité par rapport à la mesure de Lebesgue peut être exprimée comme suit

$$\mathcal{W}(S) = \frac{1}{2^{np/2}\, |\Sigma|^{n/2}\, \Gamma_p(n/2)} |S|^{(n-p-1)/2} \exp\left(\frac{1}{2} \operatorname{tr}\{\Sigma^{-1} S\}\right),$$

où $\Gamma_p(\cdot)$ représente la fonction gamma multivariée définie dans 2.4.3.

Proposition 2.4.4

Soit $S \sim \mathcal{W}_p(n, \Sigma)$ alors

$$\mathrm{E}(S) = n\,\Sigma. \tag{2.4.5}$$

Cette proposition est importante car elle est à la base de l'estimation d'une matrice de covariance.

2.4.3 Mélange de lois de Wishart

Rappelons que l'objectif de cette thèse est d'améliorer les estimateurs de Σ^{-1} du modèle de mélange

$$\begin{cases} S \mid V \sim \mathcal{W}_p(n, V\,\Sigma) \\ V \sim \mathcal{H}(\cdot) \end{cases},$$

où S est une matrice $p \times p$ d'observations, $\mathcal{W}_p(n, V\,\Sigma)$ représente la loi de Wishart à n degré de liberté et de matrice de variance covariance $V\,\Sigma$ et $\mathcal{H}(\cdot)$ est une loi de mélange définie sur $\mathbb{R}^+$. Le mélange par rapport à la variance de Wishart est une extension naturelle de la loi de Wishart comme le mélange par rapport à la variance de loi normales étend la loi normale, cas pour lequel une large littérature existe pour l'estimation d'un paramètre de position. Haff, Kim, Koo and Richards [HKKR11] ont exprimé la densité de mélange de Wishart relativement à des lois de matrices. Chez eux V est une matrice.

2.5 Des identités de type Stein-Haff

Une identité très importante qui permet d'établir des estimateurs améliorés de certains paramètres tels que les matrices de covariance et de précision est connue sous le nom d'identité de type Stein-Haff. Celle-ci constitue le noyau de notre théorie. Son application dans notre cas, qui consiste en l'estimation de la matrice de précision dans un modèle de Wishart, permet d'obtenir un estimateur sans biais de la différence des risques entre l'estimateur usuel $\hat{\Sigma}_a^{-1}$ dans (1.2.1) et l'estimateur proposé $\hat{\Sigma}_{a,c}^{-1}$ dans (3.2.1). Dans la littérature, on trouve plusieurs versions de l'identité de type Stein-Haff pour une matrice de covariance empirique S. Certaines ont été établies dans le cas Wishart et d'autres dans un cadre distributionnel plus large, et selon l'inversibilité ($p \leq n$) ou non inversibilité ($n < p$) de la matrice S. En effet, elle a été donnée dans le cas des lois de Wishart par Stein [Ste75] et Haff [Haf79b] dans le contexte de petite dimension (S inversible), et par Kubokawa and Srivastava [KS08], Konno [Kon09], Chetelat et Wells [CW12] dans le contexte de grande dimension (S non inversible). Récemment, Fourdrinier, Haddouche et Mezoued [FHM21] ont donné une version de l'identité de type Stein-Haff sous une approche unifié des cas $p > n$ et $p \leq n$. De plus celle-ci est valide dans un cadre distributionnel assez large qui contient le cas gaussien. C'est celle-ci que nous utilisons dans notre travail et que nous rappelons dans le cas où S suit une loi de Wishart. Tout d'abord, nous rappelons la définition d'une fonction faiblement différentiable qui est nécessaire pour la définition de ces identités en se référant au livre de Fourdrinier, Starwderman et Wells [FSW18].

Définition 2.5.1 (Fonction faiblement différentiable)
Soit $\Omega \in \mathbb{R}$ un ensemble ouvert de $\mathbb{R}^n$. Une fonction f de Ω dans $\mathbb{R}$ est localement intégrable si, pour tout compact K de Ω, on a

$$\int_K |f(x)| \, \mathrm{d}x < +\infty. \tag{2.5.1}$$

Une fonction localement intégrable f de Ω dans $\mathbb{R}$ est dite faiblement différentiable s'il existe n fonctions $g_1, g_2, ..., g_n$ localement intégrable telle que, pour tout $i = 1, ..., n$ on ait

$$\int_\Omega f(x) \frac{\partial h}{\partial x_i} \mathrm{d}x = -\int_\Omega g_i(x) \, h(x) \, \mathrm{d}x, \tag{2.5.2}$$

pour toute fonction h indéfiniment différentiable à support compact de Ω dans $\mathbb{R}$ (appelée fonction de test). Les fonctions g_i $(1 \leq i \leq n)$ sont les dérivées faibles de f et sont notées $g_i = \dfrac{\partial f}{\partial x_i}$.

La faible différentiabilité est le cadre analytique naturel en estimation ponctuelle. Ceci à été mis en évidence par Stein [Ste81]. En effet, si $X \sim \mathcal{N}_p(\theta, \sigma^2 I_p)$ où σ^2 est connu, l'estimation de θ par un estimateur alternatif à X, soit par un estimateur de la forme $X + g(X)$ où $g = (g_1, ..., g_p)$, nécessite l'identité de Stein

$$\mathrm{E}_\Sigma[(X-\theta)^\top g(X)] = \sigma^2 \, \mathrm{E}_\Sigma \left[\operatorname{div} g(X)\right]$$

où $\operatorname{div} g(x) = \sum_{i=1}^p \dfrac{\partial g_i(X)}{\partial X_i}$ est la divergence faible de g. Pour les estimateurs de type James-Stein, on a $g(X) = -\dfrac{a\,\sigma^2}{||X||^2} X$ où a est une constante positive, la différentiabilité de g fait défaut en 0. En revanche, la faible différentiabilité existe dés que $p \geq 3$ et on montre que dans ce cas, $\operatorname{div} g(X) = -\dfrac{a\,\sigma^2\,(p-2)}{||X||^2}$.

Dans notre cadre, les lemmes de type Stein-Haff utilisent l'identité de Stein. Le premier lemme suivant peut être trouvé dans Haff [Haf79a].

Lemme 2.5.1 (Identité de Stein-Haff dans le cas S inversible)
Soit $S \sim \mathcal{W}_p(n, \Sigma)$ et soit $G(S)$ une matrice de dimension $p \times p$ une fonction matricielle faiblement différentiable par rapport à S alors

$$\mathrm{E}_\Sigma \left[\mathrm{tr}\{\Sigma^{-1} G(S)\}\right] = \mathrm{E}\left[\mathrm{tr}\{(n-p-1)\, S^{-1} G(S) + 2\, D_S\, G(S)\}\right], \tag{2.5.3}$$

où E_Σ désigne l'espérance par rapport à $\mathcal{W}_p(n, \Sigma)$ et où D_S désigne l'opérateur différentiel de Haff pour la matrice S qui est défini par

$$D_S = \left(\frac{1}{2}(1+\delta_{ij})\, \frac{\partial}{\partial S_{ij}}\right)_{1 \le i,j \le p}, \tag{2.5.4}$$

avec

$$\delta_{ij} = \begin{cases} 1 \ si \ i = j \\ 0 \ si \ i \neq j \end{cases}.$$

Dans plusieurs travaux, les auteurs ont défini des estimateurs dits orthogonalement invariants c'est à dire tels que $G = H_1\, \Phi(L)\, H_1^T$, où H_1 et $\Phi(L)$ sont des matrices définies dans le lemme ci-dessous. De ce fait, certaines versions de type Stein-Haff ont été établies pour ce type d'estimateurs.

Lemme 2.5.2 (Identité de type Stein-Haff dans le cas S non inversible)
Soit $S \sim \mathcal{W}_p(n, \Sigma)$ de décomposition en valeur singulière $S = H_1\, L\, H_1^T$ où $H_1\, H_1^T = I_n$ et où L est la matrice diagonale des valeurs propres $l_i > 0$ de S. Soit $\Phi(L) = \mathrm{diag}\big(\phi_1(L), \ldots, \phi_{n\wedge p}(L)\big)$, tel que les $\phi_i(L)$ soient des fonctions faiblement différentiables. Alors

$$\mathrm{E}_\Sigma \left[\mathrm{tr}\{\Sigma^{-1} H_1\, \Phi(L)\, H_1^T\}\right] = \sum_{i=1}^{p} \mathrm{E}_\Sigma \left[(p-n-1)\, \frac{\phi_i}{l_i} + 2\frac{\partial\, \phi_i}{\partial l_i} + 2 \sum_{j>i} \frac{\phi_i - \phi_j}{l_i - l_j}\right], \tag{2.5.5}$$

où E_Σ désigne l'espérance par rapport à $\mathcal{W}_p(n, \Sigma)$.

Récemment, Fourdrinier, Haddouche et Mezoued [FHM21] ont donné une version d'un lemme de type Stein-Haff pour G de forme générale, dans une approche unifiée des cas inversible et singulière de S, et pour la classe de lois à symétrie elliptique. Nous en donnons une version dans le contexte de la loi de Wishart.

Lemme 2.5.3 (Identité de type Stein-Haff dans le cas unifié)
Soit S une matrice $p \times p$ de loi de Wishart $\mathcal{W}_p(n, \Sigma)$ et soit $G(S)$ une fonction matricielle $p \times p$ faiblement différentiable par rapport à S. Nous supposons que l'espérance $\mathrm{E}_\Sigma \left[|\mathrm{tr}\{\Sigma^{-1} S\, G(S)\}|\right]$ est finie où E_Σ désigne l'espérance par rapport à $\mathcal{W}_p(n, \Sigma)$. Alors

$$\mathrm{E}_\Sigma \left[\mathrm{tr}\{\Sigma^{-1} S\, G(S)\}\right] = \mathrm{E}_\Sigma \left[\mathrm{tr}\{(n - (n \wedge p) - 1)\, S^+\, S\, G(S) + 2\, S^+\, S\, D_S\, \{G^\top(S)\, S\}\}\right],$$

où $n \wedge p = \min(n, p)$ et D_S est l'opérateur différentiel de Haff défini dans (2.5.4).

Remarquons que la matrice S apparaît devant la fonction matricielle $G(S)$. Ceci est nécessaire pour inclure le cas où S est singulière alors que ce n'est pas le cas quand S est inversible puisque S^{-1} existe (voir lemme 2.5.1).

CHAPITRE 3

ESTIMATION D'UNE MATRICE DE PRÉCISION SOUS DES COÛTS BASÉS SUR LES DONNÉES

Sommaire

3.1 Introduction **23**
3.2 Estimateurs de forme générale **24**
3.3 Estimateurs de référence **27**
3.4 Estimateurs améliorés **29**
3.5 Estimateurs orthogonalement invariant **31**
3.5.1 Estimateur de type Haff 31
3.5.2 Estimateur de type Dey 35
3.6 Estimateur non nécessairement orthogonalement invariant **39**
3.6.1 Estimateur de type Efron-Morris 39
3.6.2 Une classe plus grande d'estimateurs améliorés 45
3.7 Conclusion **49**

3.1 Introduction

Dans ce chapitre, nous nous intéressons à l'amélioration des estimateurs usuels de la matrice de précision Σ^{-1} de dimension $p \times p$ du modèle de mélange de lois de Wishart danné en (1.4.1) pour une matrice semi-définie positive S, soit

$$\begin{cases} S \mid V \sim \mathcal{W}_p(n, V\,\Sigma) \\ V \sim \mathcal{H}(\cdot) \end{cases},$$

sous les coûts basés sur les données de la forme $L_k\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \mathrm{tr}[\{\hat{\Sigma}^{-1} - \Sigma^{-1}\}^2\, S^k]$, pour $k = 0, 1, 2 \ldots$ Rappelons (voir chapitre 1) que les estimateurs usuels sont définis par

$$\hat{\Sigma}_a^{-1} = a\, S^+,$$

où $a > 0$ est une constante positive et S^+ est l'inverse de Moore-Penrose de S. Ils sont naturels dans la mesure où S est la matrice de covariance empirique et donc estime naturellement Σ.

Dans le chapitre 1 nous avons évoqué que ce problème a été étudié par Efron et Morris [EM76]. Ils ont fait remarquer que certaines estimations empiriques de Bayes pour la moyenne d'une loi normale multivariée est étroitement liée au problème de l'estimation de la matrice de précision Σ^{-1} sous un coût de type quadratique

$$L_{EM}\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \mathrm{tr}[\{\hat{\Sigma}^{-1} - \Sigma^{-1}\}^2 S], \tag{3.1.1}$$

Afin d'améliorer les estimateurs usuels, nous proposons dans la section 3.2 une grande classe d'estimateurs de la forme

$$\hat{\Sigma}_{a,c}^{-1} = a\, S^+ + c\, S\, G(S),$$

où $G(S)$ est une fonction matricielle de dimension $p \times p$ homogène de degré α et c est une constante positive. Dans une approche unifiée des deux cas où la matrice S est inversible ($p \leq n$) et où celle-ci est singulière ($n < p$) nous établissons, grâce à l'identité de Stein-Haff (2.5.3), notre résultat principal qui se base sur la détermination d'un estimateur sans biais de la différence de risque entre $\hat{\Sigma}_{a,c}^{-1}$ et $\hat{\Sigma}_a^{-1}$. Dans la section 3.3, nous fournissons les estimateurs optimaux pour Σ^{-1} de la forme $a^*\, S^+$ pour $k = 1, 2$ et 3 où a^* est la valeur qui minimise le risque de l'estimateur usuel $\hat{\Sigma}_a^{-1} = a\, S^+$. Quand $k = 0$ et $k \geq 4$, nous ne disposons pas d'estimateur optimal mais nous proposons des estimateurs de la forme $a_0\, S^+$ "dits raisonnables" dans la mesure où ils possèdent certaines propriétés intéressantes. En explicitant des conditions de négativité de l'estimateur sans biais de la différence de risque entre les estimateurs $\hat{\Sigma}_{a,c}^{-1}$ et $\hat{\Sigma}_a^{-1}$, nous établissons les conditions de domination sur les estimateurs $a\, S^+$ (où $a = a*$ ou $a = a_0$) de trois différents estimateurs correspondant à des choix fixes de la matrice $G(S)$: deux estimateurs orthogonalement invariants de type Haff et de type Dey respectivement dans la section 3.5 et un estimateur non nécessairement orthogonalement invariant de type Efron-Morris dans la section 3.6. Le développement ci-dessous suit les articles de Boukehil, Fourdrinier, Mezoued et Strawderman [BFMS21b] et [BFMS21a].

3.2 Estimateurs de forme générale

Nous considérons les estimateurs de la forme générale

$$\hat{\Sigma}_{a,c}^{-1} = a\, S^+ + c\, S\, G(S), \tag{3.2.1}$$

où a et c sont des constantes positives, S^+ est l'inverse de Moore-Penrose de S et $G(S)$ est une fonction matricielle $p \times p$ homogène de degré α, c'est-à-dire qui vérifie

$$\exists\, \alpha \in \mathbb{R} \quad G(V\, S) = V^\alpha\, G(S)\,. \tag{3.2.2}$$

Notons que lorsque S est inversible ($p \leq n$) S^+ est l'inverse S^{-1}. Dans le théorème suivant, on donne une expression de la différence de risque entre l'estimateur $\hat{\Sigma}_{a,c}^{-1}$ dans (3.2.1) et l'estimateur usuel $a\, S^+$. Cette expression nous permettra d'établir des conditions suffisantes pour que $\hat{\Sigma}_{a,c}^{-1}$ améliore $a\, S^+$ sous les coûts de type Efron-Morris

$$L_k\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \mathrm{tr}[\{\hat{\Sigma}^{-1} - \Sigma^{-1}\}^2\, S^k],$$

pour $k = 1, 2, 3 \ldots$.

Théorème 3.2.1

Supposons que sous le modèle (1.4.1) et le coût (1.4.3), les risques des estimateurs $\hat{\Sigma}_a^{-1}$ dans (1.2.1) et $\hat{\Sigma}_{a,c}^{-1}$ dans (3.2.1) soient finis. Considérons la fonction matricielle $G(S)$ dans (3.2.1)

vérifiant la propriété (3.2.2), et que le produit $S^k\,G(S)$ vérifie les conditions du lemme 2.5.3 c'est à dire $S^k\,G(S)$ une fonction matricielle $p \times p$ faiblement différentiable par rapport à S et l'espérance $\mathrm{E}_\Sigma\left[|\mathrm{tr}\{\Sigma^{-1}\,S^{k+1}\,G(S)\}|\right]$ est finie. La différence de risque entre $\hat{\Sigma}_a^{-1}$ et $\hat{\Sigma}_{a,c}^{-1}$ est donnée par

$$\begin{aligned}\Delta(G) =& \mathrm{E}_\Sigma\Big[c^2\,\mu_{k+2\alpha+2}\,\mathrm{tr}\{\tilde{S}^k\,[\tilde{S}\,G(\tilde{S})]^2\}\\ &+2\,c\left[a\,\mu_{k+\alpha}-\mu_{k+\alpha+1}(n-(n\wedge p)-1)\right]\mathrm{tr}\{\tilde{S}^+\,\tilde{S}^{k+1}\,G(\tilde{S})\}\\ &-4\,c\,\mu_{k+\alpha+1}\mathrm{tr}\{\tilde{S}^+\,\tilde{S}\,D_{\tilde{S}}\,\{G^\top(\tilde{S})\,\tilde{S}^{k+1}\}\}\Big],\end{aligned} \tag{3.2.3}$$

où $\tilde{S}=S/V$, le nombre α est l'ordre d'homogénéité de $G(\cdot)$ et où μ_β est le moment d'ordre β de la variable V, i.e. $\mu_\beta=\mathrm{E}_{\mathcal{H}}[V^\beta]$ pour tout $\beta\in\mathbb{R}$.

Preuve

Supposons que les risques $R_k\left(\Sigma^{-1}, a\,S^+\right)$ et $R_k\left(\Sigma^{-1}, a\,S^+ + c\,S\,G(S)\right)$ soient finis. Alors la différence de risque entre $\hat{\Sigma}_{a,c}^{-1}$ et $\hat{\Sigma}_a^{-1}$ est donnée par :

$$\begin{aligned}\Delta(G) &= R_k\left(\Sigma^{-1}, a\,S^+ + c\,S\,G(S)\right) - R_k\left(\Sigma^{-1}, a\,S^+\right)\\ &= \mathrm{E}_\Sigma\left[\mathrm{tr}[\{a\,S^+-\Sigma^{-1}+c\,S\,G(S)\}^2\,S^k-(a\,S^+-\Sigma^{-1})^2\,S^k]\right]\\ &= \mathrm{E}_\Sigma\left[c^2\,\mathrm{tr}\{S^k\,[S\,G(S)]^2\}+2\,c\,\mathrm{tr}\{S\,G(S)(a\,S^+-\Sigma^{-1})\,S^k\}\right]\\ &= \mathrm{E}_\Sigma\left[c^2\,\mathrm{tr}\{S^k\,[S\,G(S)]^2\}+2\,a\,c\,\mathrm{tr}\{S^+\,S^{k+1}\,G(S)\}-2\,c\,\mathrm{tr}\{\Sigma^{-1}\,S^{k+1}\,G(S)\}\right].\end{aligned} \tag{3.2.4}$$

L'expression de la différence de risque $\Delta(G)$ dans (3.2.4) peut être exprimée comme

$$\Delta(G)=\mathrm{E}_\Sigma\left[\eta(S)+\beta_\Sigma(S)\right],$$

où

$$\eta(S)=c^2\,\mathrm{tr}\{S^k\,[S\,G(S)]^2\}+2\,a\,c\,\mathrm{tr}\{S^+\,S^{k+1}\,G(S)\},$$

et

$$\beta_\Sigma(S)=-2\,c\,\mathrm{tr}\{\Sigma^{-1}\,S^{k+1}\,G(S)\}.$$

Le conditionnement de $\Delta(G)$ par la variable V donne

$$\Delta(G)=\mathrm{E}_{\mathcal{H}}\left[\gamma(V)\right], \tag{3.2.5}$$

où $\mathrm{E}_{\mathcal{H}}$ représente l'espérance relativement à la loi $\mathcal{H}$ et

$$\gamma(V)=\mathrm{E}_\Sigma\left[\eta(S)\mid V\right]+\mathrm{E}_\Sigma\left[\beta_\Sigma(S)\mid V\right].$$

Maintenant, grâce au changement de variable $S=V\,\tilde{S}$, nous obtenons

$$\gamma(V)=\mathrm{E}_\Sigma\Big[\eta(V\,\tilde{S})\mid V\Big]+\mathrm{E}_\Sigma\Big[\beta_\Sigma(V\,\tilde{S})\mid V\Big].$$

avec

$$\eta(V\,\tilde{S})=c^2\,V^{k+2\alpha+2}\,\mathrm{tr}\{\tilde{S}^k[\tilde{S}\,G(\tilde{S})]^2\}+2\,a\,c\,V^{k+\alpha}\,\mathrm{tr}\{\tilde{S}^+\,\tilde{S}^{k+1}\,G(\tilde{S})\}, \tag{3.2.6}$$

et

$$\beta_\Sigma(V\,\tilde{S})=-2\,c\,V^{k+\alpha+1}\,\mathrm{tr}\{\Sigma^{-1}\,\tilde{S}^{k+1}\,G(\tilde{S})\},$$

où l'homogénéité de G d'ordre α a été utilisée.

Remarquons que la matrice de covariance Σ qui est inconnue intervient dans l'expression de $\beta_\Sigma(V\,\tilde{S})$ ci-dessus, pour l'éliminer on utilise l'identité de Haff-Stein du lemme 2.5.3. Pour se faire, il est intéressant de noter que pour toute fonction $G(S)$ et pour tout $v > 0$

$$\mathrm{E}_\Sigma\left[G(S) \mid V = v\right] = \mathrm{E}_\Sigma\Big[G(V\,\tilde{S}) \mid V = v\Big] = \mathrm{E}_\Sigma\Big[G(v\,\tilde{S})\Big],$$

par indépendance de V et S.

Maintenant, en appliquant le lemme 2.5.3 à $\beta_\Sigma(V\,\tilde{S})$ ci-dessus, nous obtenons

$$\mathrm{E}_\Sigma\Big[\beta_\Sigma(V\,\tilde{S}) \mid V\Big] = \mathrm{E}_\Sigma\Big[\beta(V\,\tilde{S}) \mid V\Big],$$

où

$$\begin{aligned}\beta(V\,\tilde{S}) = &-2\,c\,V^{k+\alpha+1}\,(n-(n\wedge p)-1)\,\mathrm{tr}\{\tilde{S^+}\,\tilde{S}^{k+1}\,G(\tilde{S})\}\\ &-4\,c\,V^{k+\alpha+1}\,\mathrm{tr}\{\tilde{S^+}\,\tilde{S}\,D_{\tilde{S}}\,\{G^\top(\tilde{S})\,\tilde{S}^{k+1}\}\}.\end{aligned} \tag{3.2.7}$$

De (3.2.6) et (3.2.7) il vient que

$$\gamma(V) = \mathrm{E}_\Sigma\Big[\eta(V\,\tilde{S}) + \beta(V\,\tilde{S}) \mid V\Big],$$

qui ne dépend plus du paramètre inconnu Σ. Plus explicitement, la fonction $\gamma(V)$ s'écrit

$$\begin{aligned}\gamma(V) = \mathrm{E}_\Sigma[&c^2\,V^{k+2\alpha+2}\,\mathrm{tr}\{\tilde{S}^k[\tilde{S}\,G(\tilde{S})]^2\}\\ &+2\,c\,\big[a\,V^{k+\alpha} - V^{k+\alpha+1}(n-(n\wedge p)-1)\big]\,\mathrm{tr}\{\tilde{S^+}\,\tilde{S}^{k+1}\,G(\tilde{S})\}\\ &-4\,c\,V^{k+\alpha+1}\,\mathrm{tr}\{\tilde{S^+}\,\tilde{S}\,D_{\tilde{S}}\,\{G^\top(\tilde{S})\,\tilde{S}^{k+1}\}\} \mid V].\end{aligned} \tag{3.2.8}$$

Finalement, comme $\tilde{S}$ est indépendante de V, le déconditionnement aboutit au résultat (3.2.3).

□

Le corollaire suivant est découle directement du théorème (3.2.1) et fournit un estimateur sans biais $\delta(G)$ de la différence des risques $\Delta(G)$ dans (3.2.3), soit $\mathrm{E}_\Sigma[\delta(G)] = \Delta(G)$.

Corollaire 3.2.1

Sous les conditions du théorème 3.2.1, un estimateur sans biais de la différence de risque entre $\hat{\Sigma}_a^{-1}$ dans (1.2.1) et $\hat{\Sigma}_{a,c}^{-1}$ dans (3.2.1) est

$$\begin{aligned}\delta(G) = &\,c^2\,\mu_{k+2\alpha+2}\,\mathrm{tr}\{\tilde{S}^k[\tilde{S}\,G(\tilde{S})]^2\}\\ &+2\,c\,\big[a\,\mu_{k+\alpha} - \mu_{k+\alpha+1}(n-(n\wedge p)-1)\big]\,\mathrm{tr}\{\tilde{S^+}\,\tilde{S}^{k+1}\,G(\tilde{S})\}\\ &-4\,c\,\mu_{k+\alpha+1}\,\mathrm{tr}\{\tilde{S^+}\,\tilde{S}\,D_{\tilde{S}}\,\{G^\top(\tilde{S})\,\tilde{S}^{k+1}\}\},\end{aligned} \tag{3.2.9}$$

Par conséquent, l'estimateur $\hat{\Sigma}_{a,c}^{-1}$ dans (3.2.1) domine $\hat{\Sigma}_a^{-1}$ dans (1.2.1) dès que $\delta(G) \leq 0$.

Rappelons que, dans le travail de Fourdrinier, Mezoued et Wells [FMW16], sous le coût quadratique usuel (qui correspond au coût L_0) un tel estimateur sans biais de la différence des risques n'a pas pu être fourni. Afin d'obtenir des conditions de domination de leur estimateur proposé, les auteurs ont utilisé un majorant de la différence de risque. En revanche, en établissant un estimateur sans biais de la différence des risques (corolaire 3.2.1), la négativité de celui-ci garantit l'amélioration de $\hat{\Sigma}_a^{-1}$ sur $\hat{\Sigma}_{a,c}^{-1}$. Par ailleurs, notons que, dans le cas particuler où la loi de mélange est la loi de Dirac en un point fixe V, le modèle (1.4.1) correspond à une loi de Wishart $\mathcal{W}_p(n, V\,\Sigma)$. Le résultat du théorème 3.2.1 exprime que l'existence d'un estimateur sans biais de $\Delta(G)$ est préservée sous le mélange.

3.3 Estimateurs de référence

Dans cette section, on recherche pour quelles les fonctions de coûts L_k, où $k = 0, 1, 2...$, il existe un estimateur optimal $a^* S^+$ parmi la famille des estimateurs $a S^+$. En d'autre termes, on souhaite déterminer pour quelles valeurs de k il existe a^* tel que

$$\forall a > 0, \;\; R_k(\Sigma^{-1}, a^* S^+) \leq R_k(\Sigma^{-1}, a S^+). \tag{3.3.1}$$

Nous verrons plus loin, que dans le cas où la constante optimale a^* n'existe pas, nous proposons des constantes a_0 qui vérifient certaines bonnes propriétés, que nous allons préciser dans la proposition 3.3.1. Sous le modèle (1.4.1), pour tout $k = 0, 1, 2...$, le risque de l'estimateur $a S^+$ s'écrit

$$\begin{aligned} R_k(\Sigma^{-1}, a S^+) =& \mathrm{E}_\Sigma[\mathrm{tr}[\{\Sigma^{-1} - a S^+\}^2 S^k]] \\ =& \mathrm{E}_\Sigma[a^2 \,\mathrm{tr}\{S^k (S^+)^2\} - 2\, a \,\mathrm{tr}\{\Sigma^{-1} S^k S^+\}] + \mathrm{E}_\Sigma[\mathrm{tr}\{\Sigma^{-2} S^k\}]. \end{aligned} \tag{3.3.2}$$

Comme le dernier terme présent dans l'expression de la deuxième égalité de (3.3.2) ne dépend pas de a la minimisation de $R_k(a S^+, \Sigma^{-1})$ en a est équivalente à la minimisation de

$$a^2 \,\mathrm{E}_\Sigma[\mathrm{tr}\{S^k (S^+)^2\}] - 2\, a \,\mathrm{E}_\Sigma[\mathrm{tr}\{\Sigma^{-1} S^k S^+\}]\,.$$

En conditionnant par la variable V on peut écrire

$$a^2 \,\mathrm{E}_\mathcal{H}[\mathrm{E}_\Sigma[\mathrm{tr}\{S^k (S^+)^2\}] \mid V] - 2\, a \,\mathrm{E}_\mathcal{H}[\mathrm{E}_\Sigma[\mathrm{tr}\{\Sigma^{-1} S^k S^+\}] \mid V],$$

d'où, grâce au changement de variable $S = V \tilde{S}$,

$$a^2 \,\mathrm{E}_\mathcal{H}[\mathrm{E}_\Sigma[V^{k-2}\mathrm{tr}\{\tilde{S}^k (\tilde{S}^+)^2\}] \mid V] - 2\, a \,\mathrm{E}_\mathcal{H}[\mathrm{E}_\Sigma[V^{k-1}\mathrm{tr}\{\Sigma^{-1} \tilde{S}^k \tilde{S}^+\}] \mid V].$$

Comme $\tilde{S}$ et V sont indépendants, le déconditionnement donne

$$a^2 \,\mu_{k-2} \,\mathrm{E}_\Sigma[\mathrm{tr}\{\tilde{S}^k (\tilde{S}^+)^2\}] - 2\, a \,\mu_{k-1} \,\mathrm{E}_\Sigma[\mathrm{tr}\{\Sigma^{-1} \tilde{S}^k \tilde{S}^+\}]\,. \tag{3.3.3}$$

Il est alors clair que la valeur optimale de (3.3.3) est

$$a^* = \frac{\mu_{k-1} \,\mathrm{E}_\Sigma[\mathrm{tr}\{\Sigma^{-1} \tilde{S}^k \tilde{S}^+\}]}{\mu_{k-2} \,\mathrm{E}_\Sigma[\mathrm{tr}\{\tilde{S}^k (\tilde{S}^+)^2\}]}\,. \tag{3.3.4}$$

En appliquant le lemme 2.5.3 avec $G(\tilde{S}) = \tilde{S}^k (\tilde{S}^+)^2$ et en utilisant le fait que $\tilde{S}^+ = \tilde{S}^+ \tilde{S} \tilde{S}^+ = \tilde{S} (\tilde{S}^+)^2$, le numérateur de (3.3.4) devient

$$\mathrm{E}_\Sigma[\mathrm{tr}\{\Sigma^{-1} \tilde{S}^k \tilde{S}^+\}] = \big(n - (n \wedge p) - 1\big) \,\mathrm{E}_\Sigma[\mathrm{tr}\{\tilde{S}^k (\tilde{S}^+)^2\}] + 2\,\mathrm{E}_\Sigma[\mathrm{tr}\{\tilde{S}^+ \tilde{S}\, D_{\tilde{S}} \{\tilde{S}^k \tilde{S^+}\}\}]\,.$$

Donc la valeur optimale (3.3.4) s'écrit comme suit

$$a^* = \frac{\mu_{k-1}}{\mu_{k-2}} [n - (n \wedge p) - 1] + 2\, \frac{\mu_{k-1}}{\mu_{k-2}} \frac{\mathrm{E}_\Sigma[\mathrm{tr}\{\tilde{S}^+ \tilde{S}\, D_{\tilde{S}} \{\tilde{S}^k \tilde{S^+}\}\}]}{\mathrm{E}_\Sigma[\mathrm{tr}\{\tilde{S}^k (\tilde{S}^+)^2\}]}\,. \tag{3.3.5}$$

La valeur optimale a^* dans (3.3.5) peut dépendre de Σ pour certaines valeurs de k. Dans le cas contraire, elle correspond à un estimateur optimal $a^* S^+$ parmi les estimateurs de référence $a S^+$. Lorsqu'elle dépend de Σ, nous proposons des valeurs alternatives de a notées a_0 qui donnent lieu à des estimateurs ayant de bonnes propriétés de sorte que l'on puisse qualifier ces valeurs de "raisonnables". Ces valeurs de a optimales ou raisonnables sont présentées dans la proposition suivante.

Proposition 3.3.1

Supposons que S suive la loi de mélange de lois de Wishart dans (1.4.1) et que le risque de S^+ soit fini. Nous disposons des résultats suivants

Pour $k = 1, 2, 3$, la valeur optimale a^* dans (3.3.5) est constante et est égale à

- $\dfrac{1}{\mu_{-1}}\left[(n \vee p) - (n \wedge p) - 1\right]$ *si* $k = 1$;
- $\mu_1 (n \vee p)$ *si* $k = 2$;
- $\dfrac{\mu_2}{\mu_1}\left[n + p + 1\right]$ *si* $k = 3$.

Pour $k = 0$ and $k \geq 4$, la valeur optimale a^* dans (3.3.5) dépend de Σ. Dans ces cas, des bornes inférieure et supérieure, notée a_0^- et a_0^+ respectivement, existent pour a^* de sorte que, pour tout $a < a_0^-$, l'estimateur $a_0^- S^+$ améliore $a S^+$ et, pour tout $a > a_0^+$, l'estimateur $a_0^+ S^+$ améliore $a S^+$. Plus précisément,

- $a_0^- = \dfrac{\mu_{-1}}{\mu_{-2}}\left[(n \vee p) - 2\left((n \wedge p) + 1\right)\right]$ *et* $a_0^+ = \dfrac{\mu_{-1}}{\mu_{-2}}\left[(n \vee p) - (n \wedge p) - 3\right]$ *si* $k = 0$;
- $a_0^- = \dfrac{\mu_{k-1}}{\mu_{k-2}}\left[(n \vee p) + (n \wedge p) + k - 2\right]$ *et* $a_0^+ = \dfrac{\mu_{k-1}}{\mu_{k-2}}\left[(n \vee p) + (k-2)(n \wedge p) + k - 2\right]$ *si* $k \geq 4$.

Preuve

Rappelons que $\tilde{S} = S/V$. Sans perte de généralité, nous allons démontrer la proposition 3.3.1 dans le cas où $V = 1$, c'est-à-dire $\tilde{S} = S$. Clairement, on a $\mathrm{tr}\{\tilde{S}^k (\tilde{S}^+)^2\} = \mathrm{tr}\{L^{k-2}\}$. Alors, en appliquant le lemme A.3 avec $\Psi(L) = L^{k-1}$, on obtient

$$\begin{aligned} D_{\tilde{S}}\{\tilde{S}^k \tilde{S}^+\} &= D_{\tilde{S}}\{H_1 L^{k-1} H_1^\top\} \\ &= \frac{1}{2}\left[p - (n \wedge p) + 2(k-1)\right] H_1 L^{k-2} H_1^\top + \frac{1}{2} H_1 \operatorname{diag}\left(\sum_{i \neq j}^{n \wedge p} \frac{l_i^{k-1} - l_j^{k-1}}{l_i - l_j}\right)_{1 \leq i \leq n \wedge p} H_1^\top \\ &\quad + \frac{1}{2}\mathrm{tr}\{L^{k-2}\}(I_p - H_1 H_1^\top). \end{aligned} \tag{3.3.6}$$

En remplaçant (3.3.6) dans le numérateur de (3.3.5) et $\mathrm{tr}\{L^{k-2}\}$ dans son dénominateur et en utilisant le fait que $\tilde{S}\,\tilde{S}^+ (I_p - H_1 H_1^\top) = 0$, la valeur optimale a^* de (3.3.5) égale

$$\begin{aligned} a^* &= \frac{\mu_{k-1}}{\mu_{k-2}}\left[n - (n \wedge p) - 1 + \left[p - (n \wedge p) + 2(k-1)\right]\right] + \frac{\mu_{k-1}}{\mu_{k-2}} \frac{\mathrm{E}_\Sigma\left[\sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \dfrac{l_i^{k-1} - l_j^{k-1}}{l_i - l_j}\right]}{\mathrm{E}_\Sigma\left[\mathrm{tr}\{L^{k-2}\}\right]} \\ &= \frac{\mu_{k-1}}{\mu_{k-2}}\left[(n \vee p) - (n \wedge p) + 2k - 3\right] + \frac{\mu_{k-1}}{\mu_{k-2}} \frac{\mathrm{E}_\Sigma\left[\sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \dfrac{l_i^{k-1} - l_j^{k-1}}{l_i - l_j}\right]}{\mathrm{E}_\Sigma\left[\mathrm{tr}\{L^{k-2}\}\right]}. \end{aligned} \tag{3.3.7}$$

Finalement, l'application du lemme A.7 à (3.3.7) suivant les valeurs de k donne la proposition 3.3.1. □

Il est intéressant de noter que dans le cas du modèle de Wishart (1.4.2), si $p \leq n$, alors $\mathrm{E}_\Sigma[\tilde{S}^+] = \mathrm{E}_\Sigma[\tilde{S}^{-1}]$ existe si et seulement si $n - p - 1 > 0$. Si $k = 1$, cette condition coïncide avec la finitude du risque de $\tilde{S}^+$ dans (1.4.4), et donc la constante optimale $a^* = n - p - 1$ est positive. De la même manière, si $k = 0$, la finitude de risque correspond à l'existence de $\mathrm{E}_\Sigma[(\tilde{S}^+)^2] = \mathrm{E}_\Sigma[\tilde{S}^{-2}]$ qui est vérifiée si et seulement si $n - p - 3 > 0$. Ainsi a_0^+ est positif

(cependant, a_0^- peut être négatif). Notons aussi que, si $k \geq 2$, le risque est toujours fini. Dans le cas où $p > n$, $\mathrm{E}_\Sigma[\tilde{S}^+]$ existe si $p > n+1$ et $\mathrm{E}_\Sigma[(\tilde{S}^+)^2] < \infty$ si $p > n+3$ (voir Cook et Forzani [CF11]).

La finitude du risque de S^+ dans le modèle de mélange (1.4.1) nécessite l'hypothèse supplémentaire d'existence des moments $\mu_\beta = \mathrm{E}_\mathcal{H}[V^\beta]$ pour $\beta = k-2$, $k-1$ et k.

Finalement, notons que dans le cas de Wishart, si $k = 0$ et $\tilde{S}$ inversible, un estimateur sans biais pour Σ^{-1} existe ; celui ci est égal à $(n-p-1)\,\tilde{S}^{-1}$. Comme $a_0^+ = n-p-3$, nous avons $n-p-1 \geq a_0^+$, par conséquent l'estimateur "raisonnable" $a_0\,\tilde{S}^{-1}$ domine l'estimateur sans biais $(n-p-1)\,\tilde{S}^{-1}$.

3.4 Estimateurs améliorés

Dans cette section, nous proposons des estimateurs qui améliorent $a\,S^+$, pour $a > 0$ fixé. Les fonctions $G(S)$ dans (3.2.1) que nous considérons sont des fonctions homogènes d'ordre $\alpha = -2$. Dans ce contexte, nous présentons dans le résultat suivant un lien entre l'amélioration de $a\,S^+$ dans le cas d'un modèle Wishart (1.4.2) et son amélioration dans le cas de mélange de lois de Wishart (1.4.1).

Lemme 3.4.1

Supposons que la fonction $G(S)$ dans (3.2.2) soit faiblement différentiable et homogène d'ordre -2. Supposons aussi que les moments $\mu_\beta = \mathrm{E}_\Sigma[V^\beta]$ existent. Si $a\,\tilde{S}^+ + c\,\tilde{S}\,G(\tilde{S})$ améliore $a\,\tilde{S}^+$ pour certaines valeurs de c dans le cas du modèle de Wishart (1.4.2) alors l'estimateur $(\mu_{k-1}/\mu_{k-2})\left(a\,\tilde{S}^+ + c\,\tilde{S}\,G(\tilde{S})\right)$ améliore $(\mu_{k-1}/\mu_{k-2})\,a\,\tilde{S}^+$ dans le cas du modèle de mélange de lois de Wishart (1.4.1), pour tout $a \geq 0$ fixé et pour tout $k = 0, 1, 2, \ldots$.

Preuve

D'après le théorème 3.2.1, l'expression de la différence des risques pour $\alpha = -2$ est donnée par

$$\begin{aligned}\Delta(G) = \mathrm{E}_\Sigma\Big[&c^2\,\mu_{k-2}\,\mathrm{tr}\{\tilde{S}^k\,[\tilde{S}\,G(\tilde{S})]^2\}\\ &+2\,c\left[a\,\mu_{k-2} - \mu_{k-1}\,(n-(n\wedge p)-1)\right]\mathrm{tr}\{\tilde{S}^+\,\tilde{S}^{k+1}\,G(\tilde{S})\}\\ &-4\,c\,\mu_{k-1}\mathrm{tr}\{\tilde{S}^+\,\tilde{S}\,D_{\tilde{S}}\,\{G^\top(\tilde{S})\,\tilde{S}^{k+1}\}\}\Big].\end{aligned}$$

En factorisant par $c\,\mu_{k-1}$ dans $\Delta(G)$ ci-dessus, celle ci peut être écrite comme suit

$$\Delta(G) = c\,\mu_{k-1}\,\mathrm{E}_\Sigma\left[c\,\frac{\mu_{k-2}}{\mu_{k-1}}\,M(\tilde{S}) + 2\,N\left(a\,\frac{\mu_{k-2}}{\mu_{k-1}},\,\tilde{S}\right)\right], \tag{3.4.1}$$

avec

$$M(\tilde{S}) = \mathrm{tr}\{\tilde{S}^k\,[\tilde{S}\,G(\tilde{S})]^2\}$$

et

$$N\left(a\,\frac{\mu_{k-2}}{\mu_{k-1}},\,\tilde{S}\right) = a\,\frac{\mu_{k-2}}{\mu_{k-1}} - (n-(n\wedge p)-1)\,\mathrm{tr}\{\tilde{S}^+\,\tilde{S}^{k+1}\,G(\tilde{S})\} - 2\,\mathrm{tr}\{\tilde{S}^+\,\tilde{S}\,D_{\tilde{S}}\,\{G^\top(\tilde{S})\,\tilde{S}^{k+1}\}\}.$$

Si S suit une loi de Wishart $\mathcal{W}_p(n,\Sigma)$ alors la loi $\mathcal{H}(\cdot)$ est dégénérée au point $V = 1$ et les moments μ_k sont égaux 1 pour tout k. Donc, si l'expression (3.4.1) est négative pour $\mu_{k-2} = \mu_{k-1} = 1$ et pour des valeurs particulières de a et c, elle est aussi négative dans le cas du

mélange de lois de Wishart (1.4.1) où a et c sont remplacés par $a\,\mu_{k-1}/\mu_{k-2}$ et $c\,\mu_{k-1}/\mu_{k-2}$ respectivement.

□

Une conséquence immédiate du lemme (3.4.1) est que les constantes a^* et a_0 (énoncées dans la proposition 3.3.1 pour $k = 0, 1, 2...$) dans le cas Wishart correspondent aux constantes $a^*\,\mu_{k-1}/\mu_{k-2}$ et $a_0\,\mu_{k-1}/\mu_{k-2}$ dans le cas du mélange de lois de Wishart. Par conséquent l'amélioration de l'estimateur optimal (respectivement "raisonnable") dans le cas de Wishart correspond à l'amélioration de l'estimateur optimal (respectivement "raisonnable") dans le cas du mélange de lois de Wishart à travers la multiplication par μ_{k-1}/μ_{k-2}.

Maintenant, nous allons construire une plus grande classe d'estimateurs améliorés en modifiant les estimateurs (3.2.1) sous la forme

$$\hat{\Sigma}^{-1}_{a,c,r} = a\,S^+ + c\,r\big(\mathrm{tr}\{S\}\big)\,S\,G(S)\,, \tag{3.4.2}$$

où c est une constante positive, r est une fonction réelle et $G(S)$ est une fonction matricielle $p \times p$ homogènes d'ordre $\alpha = -2$. Cette dernière est supposée être orthogonalement invariante, c'est-à-dire de la forme

$$G = H_1\,\Psi(L)\,H_1^\top\,, \tag{3.4.3}$$

où L et H_1 sont des matrices comme dans (1.3.12), et où $\Psi(L) = \mathrm{diag}\big(\psi_1(L), \ldots, \psi_{n\wedge p}(L)\big)$ est une matrice diagonale de dimension $(n\wedge p)\times(n\wedge p)$. Notons que l'ordre d'homogénéité -2 signifie que $\Psi(V\,L) = V^{-2}\,\Psi(L)$. Nous fournissons dans le lemme suivant des conditions sur la fonction r dans (3.4.2) pour que l'estimateur $\hat{\Sigma}^{-1}_{a,c,r}$ améliore l'estimateur $\hat{\Sigma}^{-1}_a$.

Lemme 3.4.2

Soit $G(S)$ une fonction matricielle faiblement différentiable de forme (3.4.3) et d'ordre d'homogénéité $\alpha = -2$. Supposons aussi que, pour tout $i = 1, \ldots, n\wedge p$, $\psi_i(L) > 0$ et que les moments $\mu_\beta = \mathrm{E}_{\mathcal{H}}[V^\beta]$ existent pour $\beta = k-2, k-1$ et pour tout $k = 0, 1, 2, ...$

Sous le modèle (1.4.1) et le coût (1.4.3), nous supposons que, pour tout $k = 0, 1, 2...$ et pour certains $c > 0$, l'expression (3.2.9) est négative et donc que l'estimateur $\hat{\Sigma}^{-1}_{a,c} = a\,S^+ + c\,S\,G(S)$ dans (3.2.1) améliore l'estimateur $\hat{\Sigma}^{-1}_a = a\,S^+$ dans (1.2.1). Alors l'estimateur $\hat{\Sigma}^{-1}_{a,c,r} = a\,S^+ + c\,r\big(\mathrm{tr}\{S\}\big)\,S\,G(S)$ dans (3.4.2) améliore $\hat{\Sigma}^{-1}_a = a\,S^+$ sous réserve que, pour tout $t \geq 0$, la fonction r soit différentiable et telle que $0 \leq r(t) \leq 1$ et $r'(t) \geq 0$.

Preuve

Nous procédons de la même manière que dans la preuve de théorème 3.2.1. En remplaçant $G(\tilde{S})$ par $r\big(V\,\mathrm{tr}\{\tilde{S}\}\big)\,G(\tilde{S})$ dans (3.2.8), l'expression de $\gamma(V)$ devient

$$\begin{aligned}\gamma(V) = \mathrm{E}_\Sigma\big[&c^2\,r^2\big(V\,\mathrm{tr}\{\tilde{S}\}\big)\,V^{k-2}\,\mathrm{tr}\{\tilde{S}^k[\tilde{S}\,G(\tilde{S})]^2\}\\ &+2\,c\,r\big(V\,\mathrm{tr}\{\tilde{S}\}\big)\big[a\,V^{k-2} - V^{k-1}(n-(n\wedge p)-1)\big]\,\mathrm{tr}\{\tilde{S}^+\,\tilde{S}^{k+1}\,G(\tilde{S})\}\\ &-4\,c\,V^{k-1}\,\mathrm{tr}\big\{\tilde{S}^+\,\tilde{S}\,D_{\tilde{S}}\big\{r\big(V\,\mathrm{tr}\{\tilde{S}\}\big)\,G^\top(\tilde{S})\,\tilde{S}^{k+1}\big\}\big\}\mid V\big]\,.\end{aligned} \tag{3.4.4}$$

Utilisant le lemme A.8, on obtient

$$\begin{aligned}\gamma(V) = \mathrm{E}_\Sigma\big[&c^2\,r^2\big(V\,\mathrm{tr}\{\tilde{S}\}\big)\,V^{k-2}\,\mathrm{tr}\{\tilde{S}^k[\tilde{S}\,G(\tilde{S})]^2\}\\ &+2\,c\,r\big(V\,\mathrm{tr}\{\tilde{S}\}\big)\big[a\,V^{k-2} - V^{k-1}(n-(n\wedge p)-1)\big]\,\mathrm{tr}\{\tilde{S}^+\,\tilde{S}^{k+1}\,G(\tilde{S})\}\\ &-4\,c\,V^{k-1}\,r(V\,\mathrm{tr}\{\tilde{S}\})\,\mathrm{tr}\big\{\tilde{S}^+\,\tilde{S}\,D_{\tilde{S}}\big\{G^\top(\tilde{S})\,\tilde{S}^{k+1}\big\}\big\}\\ &-4\,c\,V^k\,r'(V\,\mathrm{tr}\{\tilde{S}\})\,\mathrm{tr}\big\{G^\top(\tilde{S})\,\tilde{S}^{k+1}\big\}\mid V\big]\,.\end{aligned} \tag{3.4.5}$$

Puisque $c > 0$, $r'(\mathrm{tr}\{\tilde{S}\}) \geq 0$, $\mathrm{tr}\{G^\top(\tilde{S})\,\tilde{S}^{k+1}\} = \mathrm{tr}\{\,\Psi\,L^{k+1}\} \geq 0$ et $0 \leq r(t) \leq 1$, l'expression $\gamma(V)$ dans (3.4.5) est majorée par

$$\gamma_1(V) = \mathrm{E}_\Sigma\big[r(V\,\mathrm{tr}\{\tilde{S}\})\,\big(A(\tilde{S})\,V^{k-2} - B(\tilde{S})\,V^{k-1}\big) \mid V\big] \tag{3.4.6}$$

où

$$A(\tilde{S}) = \big(c^2\,\mathrm{tr}\{\tilde{S}^k[\tilde{S}\,G(\tilde{S})]^2\} + 2\,c\,a\,\mathrm{tr}\{\tilde{S}^+\,\tilde{S}^{k+1}\,G(\tilde{S})\}\big) \geq 0$$

et

$$B(\tilde{S}) = 2\,c\,\big((n - (n \wedge p) - 1)\,\mathrm{tr}\{\tilde{S}^+\,\tilde{S}^{k+1}\,G(\tilde{S})\} + 2\,\mathrm{tr}\big\{\tilde{S}^+\,\tilde{S}\,D_{\tilde{S}}\big\{G^\top(\tilde{S})\,\tilde{S}^{k+1}\big\}\big\}\big).$$

De par notre hypothèse, $\delta(G)$ dans (3.2.9) est négatif. Alors, pour $\alpha = -2$, on a

$$\delta(G) = A(\tilde{S})\,\mu_{k-2} - B(\tilde{S})\,\mu_{k-1} \leq 0\,. \tag{3.4.7}$$

Il est clair que le terme $B(\tilde{S})$ dans (3.4.7) positif. Par conséquent la différence des risques dans (3.2.5) est majorée comme suit

$$\Delta(G) = \mathrm{E}_{\mathcal{H}}\,[\gamma(V)] \leq \mathrm{E}_{\mathcal{H}}\,[\gamma_1(V)] = \mathrm{E}_{\mathcal{H}}\,\Big[\mathrm{E}_\Sigma\big[r(V\,\mathrm{tr}\{\tilde{S}\})\,\big(A(\tilde{S})\,V^{k-2} - B(\tilde{S})\,V^{k-1}\big) \mid V\big]\Big]$$

En inversant le conditionnement, on obtient

$$\Delta(G) \leq \mathrm{E}\,\Big[\mathrm{E}_{\mathcal{H}}\big[r(V\,\mathrm{tr}\{\tilde{S}\})\,\big(A(\tilde{S})\,V^{k-2} - B(\tilde{S})\,V^{k-1}\big) \mid \tilde{S}\big]\Big]$$

où E désigne l'espérance par rapport à la loi $\mathcal{W}_p(n, \Sigma)$, c'est à dire la loi de $\tilde{S}$. Par conséquent, puisque $A(\tilde{S})\,V^{k-2} - B(\tilde{S})\,V^{k-1}$ change une fois de signe de $+$ à - au point $V_0(\tilde{S}) = A(\tilde{S})/B(\tilde{S})$ et puisque la fonction $r(V\,\mathrm{tr}\{\tilde{S}\})$ est croissante en V, il vient d'après (3.4.7) que

$$\begin{aligned}\Delta(G) &\leq \mathrm{E}\,\Big[r(V_0(\tilde{S})\,\mathrm{tr}\{\tilde{S}\})\mathrm{E}_{\mathcal{H}}\big[\big(A(\tilde{S})\,V^{k-2} - B(\tilde{S})\,V^{k-1}\big) \mid \tilde{S}\big]\Big] \\ &\leq 0\,.\end{aligned}$$

□

3.5 Estimateurs orthogonalement invariant

Dans ce paragraphe, nous donnons deux exemples d'estimateurs améliorés qui étendent les estimateurs proposé par Haff [Haf77b] et Dey [Dey87a].

3.5.1 Estimateur de type Haff

Nous considérons les estimateurs

$$\hat{\Sigma}_{HF}^{-1} = a\,S^+ + c\,S\,S^+/\mathrm{tr}\{S\}, \tag{3.5.1}$$

qui sont de la forme $a\,S^+ + c\,S\,G(S)$ avec $G(S) = S^+/\mathrm{tr}\{S\}$. Notons que l'ordre d'homogénéité de $G(S)$ est $\alpha = -2$. Comme mentionné dans la section 1.3, ce type d'estimateurs a été proposé par Haff [Haf77a] dans le cas gaussien lorsque S est inversible sous un coût qui est proportionnel à $L_1\{\Sigma^{-1}, \hat{\Sigma}^{-1}\}$ dans (1.4.3). Remarquons que $G(S)$ est de la forme (3.4.3) avec $\Psi = I_{n\wedge p}/\mathrm{tr}\{L\}$. Grâce au lemme 3.4.2, le développement ci-dessous s'applique aux estimateurs de la forme $a\,S^+ + c\,r\big(\mathrm{tr}\{S\}\big)\,S\,S^+/\mathrm{tr}\{S\}$ où r est une fonction réelle telle que, pour tout $t \geq 0$, $0 \leq r(t) \leq 1$ et $r'(t) \geq 0$.

En appliquant le théorème 3.2.1 avec $G(S) = S^+/\mathrm{tr}\{S\}$ et $\alpha = -2$, pour tout $k \geq 0$, le terme $\delta(G)$ dans (3.2.9) peut s'écrire comme suit :

$$\delta(G) = c^2\,\mu_{k-2}\,\frac{\mathrm{tr}\{\tilde{S}^k\,(\tilde{S}\,\tilde{S}^+)^2\}}{(\mathrm{tr}\{\tilde{S}\})^2} + 2\,c\,[a\,\mu_{k-2} - \mu_{k-1}\,(n-(n\wedge p)-1)]\,\frac{\mathrm{tr}\{\tilde{S}^{k+1}\,(\tilde{S}^+)^2\}}{\mathrm{tr}\{\tilde{S}\}}$$
$$- 4\,c\,\mu_{k-1}\,\mathrm{tr}\left\{\tilde{S}^+\,\tilde{S}\,D_{\tilde{S}}\left\{\frac{\tilde{S}^+\,\tilde{S}^{k+1}}{\mathrm{tr}\{\tilde{S}\}}\right\}\right\}. \tag{3.5.2}$$

Comme $\tilde{S}^+ = \tilde{S}^+\,\tilde{S}\,\tilde{S}^+$ est une matrice symétrique et comme $\tilde{S}^+\,\tilde{S} = \tilde{S}\,\tilde{S}^+$ l'expression de $\delta(G)$ dans (3.5.2) s'exprime par

$$\delta(G) = c^2\,\mu_{k-2}\,\frac{\mathrm{tr}\{\tilde{S}^{k+1}\,\tilde{S}^+\}}{(\mathrm{tr}\{\tilde{S}\})^2} + 2\,c\,[a\,\mu_{k-2} - \mu_{k-1}\,(n-(n\wedge p)-1)]\,\frac{\mathrm{tr}\{\tilde{S}^k\,\tilde{S}^+\}}{\mathrm{tr}\{\tilde{S}\}}$$
$$- 4\,c\,\mu_{k-1}\,\mathrm{tr}\left\{\tilde{S}^+\,\tilde{S}\,D_{\tilde{S}}\left\{\frac{\tilde{S}^+\,\tilde{S}^{k+1}}{\mathrm{tr}\{\tilde{S}\}}\right\}\right\}. \tag{3.5.3}$$

Pour développer le terme différentiel $D_{\tilde{S}}\{\tilde{S}^+\,\tilde{S}^{k+1}/\,\mathrm{tr}\{\tilde{S}\}\}$ de droite dans (3.5.3), on applique l'égalité (*A*.3) (voir appendice 6) basée sur la décomposition en valeurs singulières de $\tilde{S}$, pour

$$\psi_i = \frac{l_i^k}{\mathrm{tr}\{L\}} \quad \text{et} \quad \frac{\partial \psi_i}{\partial l_i} = k\,\frac{l_i^{k-1}}{\mathrm{tr}\{L\}} - \frac{l_i^k}{(\mathrm{tr}\{L\})^2}\,.$$

On obtient

$$D_{\tilde{S}}\left\{\frac{\tilde{S}^+\,\tilde{S}^{k+1}}{\mathrm{tr}\{\tilde{S}\}}\right\} = D_{\tilde{S}}\left\{\frac{H_1\,L^k\,H_1^\top}{\mathrm{tr}\{L\}}\right\}$$
$$= \frac{1}{2}\,[p-(n\wedge p)]\,\frac{\tilde{S}^k\,\tilde{S}^+}{\mathrm{tr}\{\tilde{S}\}} + k\,\frac{\tilde{S}^k\,\tilde{S}^+}{\mathrm{tr}\{\tilde{S}\}} - \frac{\tilde{S}^{k+1}\,\tilde{S}^+}{(\mathrm{tr}\{\tilde{S}\})^2}$$
$$+ \frac{1}{2}\,\frac{1}{\mathrm{tr}\{\tilde{S}\}}\sum_{i\neq j}^{n\wedge p}\frac{l_i^k - l_j^k}{l_i - l_j} + \frac{1}{2}\,\frac{\mathrm{tr}\{\tilde{S}^k\,\tilde{S}^+\}}{\mathrm{tr}\{\tilde{S}\}}\,\big(I_p - H_1\,H_1^\top\big). \tag{3.5.4}$$

En remplaçant (3.5.4) dans (3.5.3) et en utilisant le fait que $\tilde{S}\,\tilde{S}^+\,\big(I_p - H_1\,H_1^\top\big) = 0$, le terme $\delta(G)$ devient

$$\delta(G) = c^2\,\mu_{k-2}\,\frac{\mathrm{tr}\{\tilde{S}^{k+1}\,\tilde{S}^+\}}{(\mathrm{tr}\{\tilde{S}\})^2} + 2\,c\,[a\,\mu_{k-2} - \mu_{k-1}\,(n-(n\wedge p)-1)]\,\frac{\mathrm{tr}\{\tilde{S}^k\,\tilde{S}^+\}}{\mathrm{tr}\{\tilde{S}\}}$$
$$- 2\,c\,\mu_{k-1}\,[2\,k + p - (n\wedge p)]\frac{\mathrm{tr}\{\tilde{S}^k\,\tilde{S}^+\}}{\mathrm{tr}\{\tilde{S}\}} + 4\,c\,\mu_{k-1}\,\frac{\mathrm{tr}\{\tilde{S}^{k+1}\,\tilde{S}^+\}}{(\mathrm{tr}\{\tilde{S}\})^2}$$
$$- 2\,c\,\mu_{k-1}\,\frac{1}{\mathrm{tr}\{\tilde{S}\}}\sum_{i=1}^{n\wedge p}\sum_{j\neq i}^{n\wedge p}\frac{l_i^k - l_j^k}{l_i - l_j}$$

$$= c^2\,\mu_{k-2}\,\frac{\mathrm{tr}\{\tilde{S}^{k+1}\,\tilde{S}^+\}}{(\mathrm{tr}\{\tilde{S}\})^2} + 2\,c\,\big[a\mu_{k-2} - \mu_{k-1}\,\big(n-p+2\,(n\wedge p)+1-2\,k\big)\big]\,\frac{\mathrm{tr}\{\tilde{S}^k\,\tilde{S}^+\}}{\mathrm{tr}\{\tilde{S}\}}$$
$$+ 4\,c\,\mu_{k-1}\,\frac{\mathrm{tr}\{\tilde{S}^{k+1}\,\tilde{S}^+\}}{(\mathrm{tr}\{\tilde{S}\})^2} - 2\,c\,\mu_{k-1}\,\frac{1}{\mathrm{tr}\{\tilde{S}\}}\sum_{i=1}^{n\wedge p}\sum_{j\neq i}^{n\wedge p}\frac{l_i^k - l_j^k}{l_i - l_j}\,.$$

Exprimé en fonction de L $\delta(G)$ devient

$$\delta(G) = \frac{c}{\mathrm{tr}\{L\}} \left\{ c\,\mu_{k-2}\,\frac{\mathrm{tr}\{L^k\}}{\mathrm{tr}\{L\}} - 2\left[\mu_{k-1}\left((n\vee p)-(n\wedge p)-1+2\,k\right) - a\mu_{k-2}\right]\mathrm{tr}\{L^{k-1}\} \right.$$
$$\left. +4\,\mu_{k-1}\,\frac{\mathrm{tr}\{L^k\}}{\mathrm{tr}\{L\}} - 2\,\mu_{k-1}\,B_{k+1} \right\} \tag{3.5.5}$$

avec

$$B_{k+1} = \sum_{i=1}^{n\wedge p}\sum_{j\neq i}^{n\wedge p} \frac{l_i^k - l_j^k}{l_i - l_j}$$

(voir le lemme A.7 dans l'appendice).

À partir de (3.5.5), il est alors clair que $\delta(G) \leq 0$ pour c entre 0 et

$$C_a = 2\left[\frac{\mu_{k-1}}{\mu_{k-2}}\left((n\vee p)-(n\wedge p)-1+2\,k\right) - a\right]\frac{\mathrm{tr}\{L^{k-1}\}\,\mathrm{tr}\{L\}}{\mathrm{tr}\{L^k\}} - 4\,\frac{\mu_{k-1}}{\mu_{k-2}} + 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\,B_{k+1}\,\frac{\mathrm{tr}\{L\}}{\mathrm{tr}\{L^k\}}\,. \tag{3.5.6}$$

Cela donne lieu à la proposition suivante.

Proposition 3.5.1

Dans le contexte de la proposition 3.3.1, on dispose des conditions suivantes sur les constantes a et c sous lesquelles l'estimateur de type Haff $\hat{\Sigma}_{HF}^{-1} = a\,S^+ + c\,S\,S^+/\mathrm{tr}\{S\}$ améliore l'estimateur $a\,S^+$ selon les différentes valeurs de k.

- *Cas $k=0$: $a = a_0^-$ et $0 < c < 2\,\dfrac{\mu_{-1}}{\mu_{-2}}\,[(n\wedge p)-1]$ avec $n\wedge p \geq 2$.*
- *Cas $k=1$: $a = a^*$ et $0 < c < 2\,\dfrac{1}{\mu_{-1}}\left((n\wedge p)[(n\wedge p)+1]-2\right)$ avec $n\wedge p \geq 2$.*
- *Case $k=2$: $a = a^*$ et $0 < c < 2\,\mu_1\,[(n\wedge p)+1] - 4\,\mu_1$ avec $n\wedge p \geq 2$.*
- *Cas $k=3$: $a < a^*$ et $0 < c < 2\,\dfrac{\mu_2}{\mu_1}\,(n+p+1-a)$.*
- *Cas $k \geq 4$: $a = a_0^-$ et $0 < c < 2\,\dfrac{\mu_{k-1}}{\mu_{k-2}}\,(k-3)$.*

Remarque

Notons que dans les cas $k=3$ et $a \geq a^*$, $k=0$ et $a > a_0^+$, $k \geq 4$ et $a > a_0^+$, notre méthode ne conduit pas à un estimateur de type Haff qui améliore l'estimateur usuel $a\,S^+$. En effet dans ces cas, l'expression C_a dans (3.5.6) est bien positive mais n'admet pas de minorant strictement positif.

Preuve

En utilisant la proposition 3.3.1, pour tout $k = 1, 2, ...$, dans l'expression l'expression C_a dans (3.5.6) nous substituons à a les valeurs de a^* et a_0 respectivement.

Quand $k=1$, $a^* = \dfrac{1}{\mu_{-1}}\left[(n\vee p)-(n\wedge p)-1\right]$ et alors

$$C_{a^*} = \frac{4}{\mu_{-1}}\,(n\wedge p) - \frac{4}{\mu_{-1}} + \frac{2}{\mu_{-1}}(n\wedge p)\,[(n\wedge p)-1] = \frac{2}{\mu_{-1}}\left((n\wedge p)\,[(n\wedge p)+1]-2\right),$$

ce qui donne le premier résultat.

Quand $k = 2$, $a^* = \mu_1 (n \vee p)$ et alors,

$$\begin{aligned} C_{a^*} &= 2\,\mu_1 \left[-(n \wedge p) + 3\right] \frac{(\mathrm{tr}\{L\})^2}{\mathrm{tr}\{L^2\}} - 4\,\mu_1 + 4\,\mu_1 \left[(n \wedge p) - 1\right] \frac{(\mathrm{tr}\{L\})^2}{\mathrm{tr}\{L^2\}} \\ &= 2\,\mu_1 \left[(n \wedge p) + 1\right] \frac{(\mathrm{tr}\{L\})^2}{\mathrm{tr}\{L^2\}} - 4\,\mu_1 \\ &\geq 2\,\mu_1 \left[(n \wedge p) + 1\right] - 4\,\mu_{-1}\,, \end{aligned}$$

car $(\mathrm{tr}\{L\})^2 \geq \mathrm{tr}\{L^2\}$. D'où le deuxième résultat.

Maintenant, quand $k = 3$, on a $a^* = \dfrac{\mu_2}{\mu_1}\left[(n \vee p) + (n \wedge p) + 1\right]$, en utilisant l'expression de B_4 donnée dans le lemma A.7 (voir appendice 6), on obtient

$$\begin{aligned} C_{a^*} &= 2\,\frac{\mu_2}{\mu_1}\left[-2\,(n \wedge p) + 4\right] \frac{\mathrm{tr}\{L^2\}\,\mathrm{tr}\{L\}}{\mathrm{tr}\{L^3\}} - 4\,\frac{\mu_2}{\mu_1} \\ &\qquad + 2\,\frac{\mu_2}{\mu_1}\left((\mathrm{tr}\{L\})^2 + \left[2\,(n \wedge p) - 3\right]\mathrm{tr}\{L^2\}\right) \frac{\mathrm{tr}\{L\}}{\mathrm{tr}\{L^3\}} \\ &= 2\,\frac{\mu_2}{\mu_1}\,\frac{\mathrm{tr}\{L^2\}\,\mathrm{tr}\{L\}}{\mathrm{tr}\{L^3\}} - 4\,\frac{\mu_2}{\mu_1} + 2\,\frac{\mu_2}{\mu_1}\,\frac{(\mathrm{tr}\{L\})^3}{\mathrm{tr}\{L^3\}}\,. \end{aligned} \tag{3.5.7}$$

Notons que C_{a^*} dans (3.5.7) n'est pas constant car il dépend de L et n'admet pas de borne inférieure strictement positive. Grâce à l'inégalité (A.4) dans le lemme A.4 avec $k = 1$ et $m = 1$ et 2, on obtient

$$\frac{(\mathrm{tr}\{L\})^3}{\mathrm{tr}\{L^3\}} \geq \frac{\mathrm{tr}\{L^2\}\,\mathrm{tr}\{L\}}{\mathrm{tr}\{L^3\}} \geq 1,$$

Il est donc clair que $C_{a^*} \geq 0$ avec une probabilité égale à 1. Cependant, pour tout $a > 0$, la quantité C_a peut être écrite sous la forme

$$C_a = C_{a^*} + 2\,\frac{\mu_2}{\mu_1}\,(a^* - a)\,\frac{\mathrm{tr}\{L^2\}\,\mathrm{tr}\{L\}}{\mathrm{tr}\{L^3\}}\,,$$

Par conséquent une borne inférieure pour C_a est donnée par

$$C_a \geq 2\,\frac{\mu_2}{\mu_1}\,(a^* - a),$$

sous la condition $a < a^*$. D'où le résultat pour $k = 3$.

Quand $k = 0$, $a_0^- = \dfrac{\mu_{-1}}{\mu_{-2}}\left[(n \vee p) - 2\left((n \wedge p) + 1\right)\right]$. Il est alors facile de voir que

$$C_{a_0^-} = 2\,\frac{\mu_{-1}}{\mu_{-2}}\left[(n \wedge p) + 1\right]\frac{\mathrm{tr}\{L^{-1}\}\,\mathrm{tr}\{L\}}{n \wedge p} - 4\,\frac{\mu_{-1}}{\mu_{-2}} \geq 2\,\frac{\mu_{-1}}{\mu_{-2}}\left[(n \wedge p) - 1\right],$$

puisque $\mathrm{tr}\{L^{-1}\}\,\mathrm{tr}\{L\} \geq n \wedge p$. Ce qui donne le résultat pour $k = 0$.

Quand $k \geq 4$, on a $a_0^- = \dfrac{\mu_{k-1}}{\mu_{k-2}}\left[(n \vee p) + (n \wedge p) + k - 2\right]$. D'après l'expression de B_k dans le lemma A.7, il découle de (3.5.6) que

$$\begin{aligned} C_{a_0^-} &= 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\left[-2\,(n \wedge p) + 2k - 1\right]\frac{\mathrm{tr}\{L^{k-1}\}\,\mathrm{tr}\{L\}}{\mathrm{tr}\{L^k\}} - 4\,\frac{\mu_{k-1}}{\mu_{k-2}} \\ &\qquad + 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\left\{\sum_{m=1}^{k-2}\mathrm{tr}\{L^m\}\,\mathrm{tr}\{L^{k-1-m}\} + \left[2\,\frac{\mu_{k-1}}{\mu_{k-2}}\,(n \wedge p) - k\right]\mathrm{tr}\{L^{k-1}\}\right\}\frac{\mathrm{tr}\{L\}}{\mathrm{tr}\{L^k\}} \\ &= 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\,\frac{\mathrm{tr}\{L^{k-1}\}\,\mathrm{tr}\{L\}}{\mathrm{tr}\{L^k\}} - 4\,\frac{\mu_{k-1}}{\mu_{k-2}} + 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\,\frac{\mathrm{tr}\{L\}}{\mathrm{tr}\{L^k\}}\sum_{m=1}^{k-2}\mathrm{tr}\{L^m\}\,\mathrm{tr}\{L^{k-1-m}\}. \end{aligned}$$

D'après l'inégalité (A.4), il est facile de voir que $\sum_{m=1}^{k-2} \operatorname{tr}\{L^m\} \operatorname{tr}\{L^{k-1-m}\} \geq (k-2) \operatorname{tr}\{L^{k-1}\}$ et que $\operatorname{tr}\{L^{k-1}\} \operatorname{tr}\{L\} \geq \operatorname{tr}\{L^k\}$. Donc une majoration pour $C_{a_0^-}$ est donné par :

$$\begin{aligned} C_{a_0^-} &\geq 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\,\frac{\operatorname{tr}\{L^{k-1}\}\operatorname{tr}\{L\}}{\operatorname{tr}\{L^k\}} - 4\,\frac{\mu_{k-1}}{\mu_{k-2}} + 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\,(k-2)\,\frac{\operatorname{tr}\{L^{k-1}\}\operatorname{tr}\{L\}}{\operatorname{tr}\{L^k\}} \\ &= 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\,(k-1)\,\frac{\operatorname{tr}\{L^{k-1}\}\operatorname{tr}\{L\}}{\operatorname{tr}\{L^k\}} - 4\,\frac{\mu_{k-1}}{\mu_{k-2}} \\ &\geq 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\,(k-3), \end{aligned}$$

ce qui donne le résultat pour $k \geq 4$.

□

3.5.2 Estimateur de type Dey

Nous considérons les estimateurs

$$\hat{\Sigma}_{DY}^{-1} = a\,S^+ + c\,\frac{S}{\operatorname{tr}\{S^2\}}, \tag{3.5.8}$$

qui sont de la forme $a\,S^+ + c\,S\,G(S)$ avec $G(S) = S\,S^+/\operatorname{tr}\{S^2\}$. Notons que l'ordre d'homogénéité de $G(S)$ est $\alpha = -2$. Comme mentionné dans la section 1.3, ce type d'estimateurs a été proposé par Dey [Dey87b] dans le cas gaussien lorsque S est inversible, sous le coût $L_0\{\Sigma^{-1}, \hat{\Sigma}^{-1}\}$ dans (1.4.3). Remarquons que $G(S)$ est de la forme (3.4.3) avec $\Psi = I_{n\wedge p}/\operatorname{tr}\{L^2\}$. Grâce au lemme 3.4.2 le développement ci-dessous s'applique aux estimateurs de la forme $a\,S^+ + c\,r\big(\operatorname{tr}\{S\}\big)\,S\,/\operatorname{tr}\{S^2\}$ où r est une fonction réelle telle que, pour tout $t \geq 0$, $0 \leq r(t) \leq 1$ et $r'(t) \geq 0$.

En appliquant le théorème 3.2.1 avec $G(S) = S\,S^+/\operatorname{tr}\{S\}$ et $\alpha = -2$, pour tout $k \geq 0$, le terme $\delta(G)$ dans (3.2.9) peut s'écrire comme suit :

$$\begin{aligned} \delta(G) = c^2\,\mu_{k-2}\,\frac{\operatorname{tr}\{\tilde{S}^{k+2}\}}{(\operatorname{tr}\{\tilde{S}^2\})^2} + 2\,c\,\big[\,a\,\mu_{k-2} - \mu_{k-1}\,(n-(n\wedge p)-1)\big]\,\frac{\operatorname{tr}\{\tilde{S}^+\,\tilde{S}^{k+1}\}}{\operatorname{tr}\{\tilde{S}^2\}} \\ - 4\,c\,\mu_{k-1}\operatorname{tr}\left\{\tilde{S}^+\,\tilde{S}\,D_{\tilde{S}}\left\{\frac{\tilde{S}^{k+1}}{\operatorname{tr}\{\tilde{S}^2\}}\right\}\right\}. \end{aligned} \tag{3.5.9}$$

Pour développer le terme différentiel $D_{\tilde{S}}\big\{\,\tilde{S}^{k+1}/\operatorname{tr}\{\tilde{S}^2\}\big\}$ de droite dans (3.5.9), on applique l'égalité (A.3) (voir appendice 6) basée sur la décomposition en valeurs singulières de $\tilde{S}$, pour

$$\phi_i = \frac{l_i^{k+1}}{\operatorname{tr}\{L^2\}} \quad \text{et} \quad \frac{\partial \phi_i}{\partial l_i} = (k+1)\,\frac{l_i^k}{\operatorname{tr}\{L^2\}} - 2\,\frac{l_i^{k+2}}{(\operatorname{tr}\{L^2\})^2}\,.$$

On obtient alors

$$D_{\tilde{S}}\left\{\frac{\tilde{S}^{k+1}}{\operatorname{tr}\{\tilde{S}^2\}}\right\} = D_{\tilde{S}}\left\{\frac{H_1\,L^{k+1}\,H_1^\top}{\operatorname{tr}\{L^2\}}\right\}$$

$$
\begin{aligned}
&= H_1 \left\{ \frac{1}{2} \big(p - (n \wedge p)\big) \frac{L^k}{\mathrm{tr}\{L^2\}} + (k+1) \frac{L^k}{\mathrm{tr}\{L^2\}} - 2 \frac{L^{k+2}}{(\mathrm{tr}\{L^2\})^2} \right\} H_1^\top \\
&\quad + H_1 \mathrm{diag} \left\{ \frac{1}{2\,\mathrm{tr}\{L^2\}} \sum_{j \neq i}^{(n \wedge p)} \frac{l_i^{k+1} - l_j^{k+1}}{l_i - l_j} \right\} H_1^\top \\
&\quad + \frac{1}{2} \mathrm{tr} \left\{ \frac{L^k}{\mathrm{tr}\{L^2\}} \right\} (I_p - H_1 H_1^\top)
\end{aligned} \tag{3.5.10}
$$

Maintenant, en substituant (3.5.10) dans (3.5.9), $\delta(G)$ devient

$$
\begin{aligned}
\delta(G) = c^2 \, \mu_{k-2} \frac{\mathrm{tr}\{\tilde{S}^{k+2}\}}{(\mathrm{tr}\{\tilde{S}^2\})^2} &+ 2\,c \left[a\,\mu_{k-2} - \mu_{k-1} \left(n - (n \wedge p) - 1\right) \right] \frac{\mathrm{tr}\{\tilde{S}^+ \tilde{S}^{k+1}\}}{\mathrm{tr}\{\tilde{S}^2\}} \\
&- 4\,c\,\mu_{k-1} \mathrm{tr} \left\{ \tilde{S}^+ \tilde{S} H_1 \left\{ \frac{1}{2} \big(p - (n \wedge p)\big) \frac{L^k}{\mathrm{tr}\{L^2\}} + (k+1) \frac{L^k}{\mathrm{tr}\{L^2\}} \right.\right. \\
&\qquad \left.\left. -2 \frac{L^{k+2}}{(\mathrm{tr}\{L^2\})^2} H_1^\top \right\} \right\} \\
&- 4\,c\,\mu_{k-1} \mathrm{tr} \left\{ \tilde{S}^+ \tilde{S} H_1 \mathrm{diag} \left\{ \frac{1}{2\,\mathrm{tr}\{L^2\}} \sum_{j \neq i}^{n \wedge p} \frac{l_i^{k+1} - l_j^{k+1}}{l_i - l_j} \right\} H_1^\top \right\} \\
&- 4\,c\,\mu_{k-1} \mathrm{tr} \left\{ \frac{1}{2} \tilde{S}^+ \tilde{S} \, \mathrm{tr} \left\{ \frac{L^k}{\mathrm{tr}\{L^2\}} \right\} (I_p - H_1 H_1^\top) \right\}.
\end{aligned} \tag{3.5.11}
$$

En utilisant le fait que $\tilde{S}\,\tilde{S}^+ \left(I_p - H_1 H_1^\top\right) = 0$ et $n \vee p = n + p - (n \wedge p)$, il est clair que l'expression de $\delta(G)$ dans (3.5.11) peut être exprimée en fonction de L comme suit

$$
\begin{aligned}
\delta(G) = \frac{c}{\mathrm{tr}\{L^2\}} \Bigg\{ c\,\mu_{k-2} \frac{\mathrm{tr}\{L^{k+2}\}}{\mathrm{tr}\{L^2\}} &+ 2 \left[a\,\mu_{k-2} - \mu_{k-1} \big((n \vee p) - (n \wedge p) + 2\,k + 1\big) \right] \mathrm{tr}\{L^k\} \\
&+ 8\,\mu_{k-1} \frac{\mathrm{tr}\{L^{k+2}\}}{\mathrm{tr}\{L^2\}} - 2\,\mu_{k-1} \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \frac{l_i^{k+1} - l_j^{k+1}}{l_i - l_j} \Bigg\}.
\end{aligned} \tag{3.5.12}
$$

À partir de (3.5.12), il est clair que $\delta(G) \leq 0$ pour c entre 0 et

$$
C_a = 2 \left[\frac{\mu_{k-1}}{\mu_{k-2}} \big[(n \vee p) - (n \wedge p) + 2\,k + 1\big] - a \right] \frac{\mathrm{tr}\{L^k\}\,\mathrm{tr}\{L^2\}}{\mathrm{tr}\{L^{k+2}\}} - 8 \frac{\mu_{k-1}}{\mu_{k-2}} + 2\,B_{k+2} \frac{\mu_{k-1}}{\mu_{k-2}} \frac{\mathrm{tr}\{L^2\}}{\mathrm{tr}\{L^{k+2}\}}. \tag{3.5.13}
$$

avec

$$
B_{k+2} = \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \frac{l_i^{k+1} - l_j^{k+1}}{l_i - l_j}.
$$

Cela conduit à la proposition suivante.

Proposition 3.5.2

Dans le contexte de la proposition 3.3.1, on dispose des conditions suivantes sur les constantes a et c sous lesquelles l'estimateur de type Dey $\hat{\Sigma}_{DY}^{-1} = a\,S^+ + c\,S/\mathrm{tr}\{S^2\}$ améliore l'estimateur $a\,S^+$ selon les différentes valeurs de k.

Case $k = 0$: $a = a_0^-$ et $0 < c < 4 \dfrac{\mu_{-1}}{\mu_{-2}} \left[\{(n \wedge p) + 1\}\,(n \wedge p) - 2\right]$ avec $n \wedge p \geq 2$.

$$a = a_0^+ \ \textit{et}\ 0 < c < 2\,\frac{\mu_{-1}}{\mu_{-2}}\left[\{(n\wedge p)+3\}\,(n\wedge p)-4\right]\ \textit{avec}\ n\wedge p \geq 2.$$

Case $k=1$: $a = a^*$ *et* $0 < c < \dfrac{4}{\mu_{-1}}\left[(n\wedge p)-1\right]$ *avec* $n\wedge p \geq 2$.

Case $k=2$: $a = a^*$ *et* $0 < c < 2\,\mu_1\,[(n\wedge p)-1]$ *avec* $n\wedge p \geq 2$.

Case $k=3$: $a < a^*$ *et* $0 < c < 2\,\dfrac{\mu_2}{\mu_1}\,(n+p+1-a)$.

Case $k \geq 4$: $a = a_0^-$ *et* $0 < c < 2\,\dfrac{\mu_{k-1}}{\mu_{k-2}}\,(k-3)$.

Remarque

Dans les cas $k=3$ et $a \geq a^*$, $k \geq 4$ et $a > a_0^+$, notre méthode ne conduit pas à un estimateur de type Dey qui améliore l'estimateur usuel $a\,S^+$. Dans ce cas, l'expression de C_a dans (3.5.13) est bien positive mais elle n'admet pas de minorant strictement positif.

Preuve

En utilisant la proposition 3.3.1, pour tout $k = 1, 2, ...$, dans l'expression l'expression C_a dans (3.5.6) nous substituons à a les valeurs de a^* et a_0 respectivement.

Quand $k=1$, on a $a^* = \dfrac{1}{\mu_{-1}}\left[(n\vee p)-(n\wedge p)-1\right]$ sous réserve que $n\vee p > n\wedge p - 1$. Alors

$$C_{a^*} = \frac{8}{\mu_{-1}}\,\frac{\mathrm{tr}\{L\}\,\mathrm{tr}\{L^2\}}{\mathrm{tr}\{L^3\}} - \frac{8}{\mu_{-1}} + \frac{2}{\mu_{-1}}\,B_3\,\frac{\mathrm{tr}\{L^2\}}{\mathrm{tr}\{L^{k+2}\}}.$$

En utilisant le fait que $B_3 = 2[(n\wedge p)-1]\,\mathrm{tr}\{L\}$ (voir le lemme A.7 dans l'appendice 6), on obtient

$$\begin{aligned} C_{a^*} &= \frac{8}{\mu_{-1}}\,\frac{\mathrm{tr}\{L\}\,\mathrm{tr}\{L^2\}}{\mathrm{tr}\{L^3\}} - \frac{8}{\mu_{-1}} + \left[(n\wedge p)-1\right]\,\frac{4}{\mu_{-1}}\,\frac{\mathrm{tr}\{L\}\,\mathrm{tr}\{L^2\}}{\mathrm{tr}\{L^3\}} \\ &= \frac{4}{\mu_{-1}}\left[(n\wedge p)+1\right]\frac{\mathrm{tr}\{L\}\,\mathrm{tr}\{L^2\}}{\mathrm{tr}\{L^3\}} - \frac{8}{\mu_{-1}}\,, \end{aligned}$$

Grâce à l'inégalité (A.4) dans le lemme A.4 avec $k=1$ et $m=1$ et 2, on obtient

$$C_{a^*} \geq \frac{4}{\mu_{-1}}\left[(n\wedge p)-1\right], \tag{3.5.14}$$

ce qui donne le premier résultat.

Quand $k=2$, $a^* = \mu_1\,(n\vee p)$. En utilisant l'expression de B_k dans le lemme A.7 avec $k=4$, il vient

$$\begin{aligned} C_{a^*} &= 2\,\mu_1\,[-(n\wedge p)+5]\,\frac{(\mathrm{tr}\{L^2\})^2}{\mathrm{tr}\{L^4\}} - 8\,\mu_1 + 2\,\mu_1\,B_4\,\frac{\mathrm{tr}\{L^2\}}{\mathrm{tr}\{L^4\}} \\ &= 2\,\mu_1\,[-(n\wedge p)+5]\,\frac{(\mathrm{tr}\{L^2\})^2}{\mathrm{tr}\{L^4\}} - 8\,\mu_1 + 2\,\mu_1\left[(\mathrm{tr}\{L\})^2 + (2\,(n\wedge p)-3)\,\mathrm{tr}\{L^2\}\right]\frac{\mathrm{tr}\{L^2\}}{\mathrm{tr}\{L^4\}} \\ &= 2\,\mu_1\,[(n\wedge p)+2]\,\frac{(\mathrm{tr}\{L^2\})^2}{\mathrm{tr}\{L^4\}} - 8\,\mu_1 + 2\,\mu_1\,\frac{(\mathrm{tr}\{L\})^2\,\mathrm{tr}\{L^2\}}{\mathrm{tr}\{L^4\}} \\ &\geq 2\,\mu_1\,[(n\wedge p)+3]\,\frac{(\mathrm{tr}\{L^2\})^2}{\mathrm{tr}\{L^4\}} - 8\,\mu_1\,, \end{aligned}$$

car $(\mathrm{tr}\{L\})^2 \geq \mathrm{tr}\{L^2\}$. Maintenant, grâce à l'égalité (A.4) avec $k=2$ et $m=2$, on a

$$C_{a^*} \geq 2\,\mu_1\,[(n\wedge p)-1]\,,$$

d'où le deuxième résultat.

Quand $k = 3$, comme $a^* = \frac{\mu_2}{\mu_1}\left[(n \vee p) + (n \wedge p) + 1\right]$, en utilisant l'expression de B_5 donnée par le lemme A.7 (voir appendice 6), on obtient

$$\begin{aligned} C_{a^*} &= 2\,\frac{\mu_2}{\mu_1}\left[-2\,(n \wedge p) + 6\right]\frac{\mathrm{tr}\{L^3\}\,\mathrm{tr}\{L^2\}}{\mathrm{tr}\{L^5\}} - 8\,\frac{\mu_2}{\mu_1} + 2\,\frac{\mu_2}{\mu_1}\,B_5\,\frac{\mathrm{tr}\{L^2\}}{\mathrm{tr}\{L^5\}} \\ &= 2\,\frac{\mu_2}{\mu_1}\left[-2\,(n \wedge p) + 6\right]\frac{\mathrm{tr}\{L^3\}\,\mathrm{tr}\{L^2\}}{\mathrm{tr}\{L^5\}} - 8\,\frac{\mu_2}{\mu_1} \\ &\qquad + 2\,\frac{\mu_2}{\mu_1}\left[2\,\mathrm{tr}\{L\}\,\mathrm{tr}\{L^2\} + (2\,(n \wedge p) - 4)\,\mathrm{tr}\{L^3\}\right]\frac{\mathrm{tr}\{L^2\}}{\mathrm{tr}\{L^5\}} \\ &= 4\,\frac{\mu_2}{\mu_1}\,\frac{\mathrm{tr}\{L^3\}\,\mathrm{tr}\{L^2\}}{\mathrm{tr}\{L^5\}} + 4\,\frac{\mu_2}{\mu_1}\,\frac{\mathrm{tr}\{L\}\,(\mathrm{tr}\{L^2\})^2}{\mathrm{tr}\{L^5\}} - 8\,\frac{\mu_2}{\mu_1}\,, \end{aligned} \tag{3.5.15}$$

Notons que C_{a^*} dans (3.5.15) n'est pas constant car il dépend de L et n'admet pas de borne inférieure strictement positive. Grâce à l'inégalité (A.4) dans le lemme A.4 avec $m = 2$ et $k = 1$ et 3, on obtient

$$\frac{\mathrm{tr}\{L\}\,(\mathrm{tr}\{L^2\})^2}{\mathrm{tr}\{L^5\}} \geq \frac{\mathrm{tr}\{L^3\}\,\mathrm{tr}\{L^2\}}{\mathrm{tr}\{L^5\}} \geq 1\,.$$

Il est clair que $C_{a^*} \geq 0$. Cependant, grâce à (3.5.15), pour tout $a > 0$, C_a dans (3.5.13) peut être écrit

$$C_a = C_{a^*} + 2\,\frac{\mu_2}{\mu_1}\,(a^* - a)\,\frac{\mathrm{tr}\{L^3\}\,\mathrm{tr}\{L^2\}}{\mathrm{tr}\{L^5\}}\,,$$

Par conséquent, une borne inférieure pour C_a est donnée par

$$C_a \geq 2\,\frac{\mu_2}{\mu_1}\,(a^* - a),$$

sous la condition que $a < a^*$. D'où le résultat pour $k = 3$.

Qaund $k = 0$, en utilisant l'expression de B_2 donnée dans le lemme A.7 (voir appendice 6), C_a dans (3.5.13) devient

$$\begin{aligned} C_a &= 2\left[\frac{\mu_{-1}}{\mu_{-2}}\left[(n \vee p) - (n \wedge p) + 1\right] - a\right](n \wedge p) - 8\,\frac{\mu_{-1}}{\mu_{-2}} + 2\,\frac{\mu_{-1}}{\mu_{-2}}\,B_2 \\ &= 2\left[\frac{\mu_{-1}}{\mu_{-2}}\left[(n \vee p) - (n \wedge p) + 1\right] - a\right](n \wedge p) - 8\,\frac{\mu_{-1}}{\mu_{-2}} + 2\,\frac{\mu_{-1}}{\mu_{-2}}\,(n \wedge p)\left[(n \wedge p) - 1\right] \\ &= 2\left[\frac{\mu_{-1}}{\mu_{-2}}\,(n \vee p) - a\right](n \wedge p) - 4\,\frac{\mu_{-1}}{\mu_{-2}}\,, \end{aligned} \tag{3.5.16}$$

Pour $a = a_0^- = \frac{\mu_{-1}}{\mu_{-2}}\left[(n \vee p) - 2\,(n \wedge p) - 2\right]$, il est clair d'après (3.5.16) que

$$C_{a_0^-} = 2\,\frac{\mu_{-1}}{\mu_{-2}}\left[\{2\,(n \wedge p) + 2\}\,(n \wedge p) - 4\right] = 4\,\frac{\mu_{-1}}{\mu_{-2}}\left[\{(n \wedge p) + 1\}\,(n \wedge p) - 2\right].$$

Pour $a = a_0^+ = \frac{\mu_{-1}}{\mu_{-2}}\left[(n \vee p) - (n \wedge p) - 3\right]$, il est clair d'après (3.5.16) que

$$C_{a_0^+} = 2\,\frac{\mu_{-1}}{\mu_{-2}}\left[\{(n \wedge p) + 3\}\,(n \wedge p) - 4\right].$$

Ce qui donne le résultat pour $k = 0$.

Quand $k \geq 4$, $a_0^- = \dfrac{\mu_{k-1}}{\mu_{k-2}} [\,(n \vee p) + (n \wedge p) + k - 2\,]$. En utilisant l'expression de B_5 donnée par le lemme A.7 (voir appendice 6), on obtient

$$\begin{aligned}
C_{a_0^-} &= 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\,[-2\,(n \wedge p) + k + 3]\,\frac{\operatorname{tr}\{L^k\}\operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^{k+2}\}} - 8\,\frac{\mu_{k-1}}{\mu_{k-2}} \\
&\quad + 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\left\{\sum_{m=1}^{k-1}\operatorname{tr}\{L^m\}\operatorname{tr}\{L^{k-m}\} + [2\,(n \wedge p) - (k+1)]\operatorname{tr}\{L^k\}\right\}\frac{\operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^{k+2}\}} \\
&= 4\,\frac{\mu_{k-1}}{\mu_{k-2}}\,\frac{\operatorname{tr}\{L^k\}\operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^{k+2}\}} - 8\,\frac{\mu_{k-1}}{\mu_{k-2}} + 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\,\frac{\operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^{k+2}\}}\sum_{m=1}^{k-1}\operatorname{tr}\{L^m\}\operatorname{tr}\{L^{k-m}\},
\end{aligned}$$

d'après l'inégalité (A.4), il est facile de voir que $\sum_{m=1}^{k-1}\operatorname{tr}\{L^m\}\operatorname{tr}\{L^{k-m}\} \geq (k-1)\operatorname{tr}\{L^k\}$ et $\operatorname{tr}\{L^k\}\operatorname{tr}\{L^2\} \geq \operatorname{tr}\{L^{k+2}\}$. Donc une majoration pour $C_{a_0^-}$ est donnée par

$$\begin{aligned}
C_{a_0^-} &\geq 4\,\frac{\mu_{k-1}}{\mu_{k-2}}\,\frac{\operatorname{tr}\{L^k\}\operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^{k+2}\}} - 8\,\frac{\mu_{k-1}}{\mu_{k-2}} + 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\,(k-1)\,\frac{\operatorname{tr}\{L^k\}\operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^{k+2}\}} \\
&= 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\,(k+1)\,\frac{\operatorname{tr}\{L^k\}\operatorname{tr}\{L^2\}}{\operatorname{tr}\{L^{k+2}\}} - 8\,\frac{\mu_{k-1}}{\mu_{k-2}} \\
&\geq 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\,(k-3),
\end{aligned}$$

ce qui donne le résultat pour $k \geq 4$. □

3.6 Estimateur non nécessairement orthogonalement invariant

3.6.1 Estimateur de type Efron-Morris

Soit Q une matrice à éléments constants de dimension $(n \wedge p) \times q$ et de rang $\operatorname{rk}(Q) = q$. On considère le projecteur orthogonal de $\mathbb{R}^{n \wedge p}$ sur le sous-espace vectoriel engendré par les vecteurs colonnes de Q exprimé matriciellement par la matrice de taille $(n \wedge p) \times (n \wedge p)$ suivante

$$Q_0^* = Q\,(Q^\top Q)^{-1}\,Q^\top. \tag{3.6.1}$$

On propose comme estimateur alternatif à l'estimateur $a\,S^+$ (avec $a > 0$), qui est non nécessairement orthogonalement invariant, l'estimateur défini par

$$\hat{\Sigma}_{EM}^{-1} = a\,S^+ + c\,\frac{S\,S^+\,Q_0}{\operatorname{tr}\{S\}}, \tag{3.6.2}$$

où $Q_0 = H_1\,Q_0^*\,H_1^\top$ et $c \in \mathbb{R}$. Cet estimateur est de la forme $a\,S^+ + c\,S\,G(S)$ avec

$$G(S) = \frac{S^+\,Q_0}{\operatorname{tr}\{S\}}. \tag{3.6.3}$$

Dans le cadre du modèle Wishart (1.4.2), Tsukuma and Konno [TK06] ont considéré un estimateur de ce type dans le cas d'inversibilité de S ($n \wedge p = p$) et sous le coût quadratique L_0 dans (1.4.3). L'estimateur (3.6.2) constitue une extension au cas où la matrice S est singulière et où la loi de S est un mélange de lois de Wishart. Précisons que l'ordre d'homogénéité de

$G(S)$ est $\alpha = -2$ et notons que grâce à la forme de Q_0, l'estimateur $\hat{\Sigma}_{EM}^{-1}$ dans (3.6.2) n'est pas orthogonalement invariant, sauf dans le cas où $q = n \wedge p$ qui correspond au cas où $Q_0^* = I_{n\wedge p}$.

Dans le théorème suivant, on donne un estimateur sans biais de la différence de risque entre les estimateurs $\hat{\Sigma}_{EM}^{-1}$ dans (3.6.2) et $\hat{\Sigma}_a^{-1}$ dans (1.2.1) ainsi que les conditions de domination de $\hat{\Sigma}_{EM}^{-1}$ sur $\hat{\Sigma}_a^{-1}$.

Théorème 3.6.1

Un estimateur de la différence de risque entre l'estimateur $\hat{\Sigma}_{EM}^{-1}$ dans (3.6.2) et l'estimateur usuel a S^+ dans (1.2.1), par rapport à la famille des risques dans (3.2.3), est exprimé en fonction de la matrice L (définie dans (1.3.12)) comme suit

$$\delta(G) = c^2\,\mu_{k-2}\,\frac{\mathrm{tr}\{L^k\,Q_0^*\}}{(\mathrm{tr}\{L\})^2} + 2\,c\,\left[a\,\mu_{k-2} - \mu_{k-1}\left(n+p-2\,(n\wedge p)+2\,k-1\right)\right]\frac{\mathrm{tr}\{L^{k-1}Q_0^*\}}{\mathrm{tr}\{L\}}$$
$$+\,4\,c\,\mu_{k-1}\,\frac{\mathrm{tr}\{\,L^k\,Q_0^*\}}{(\mathrm{tr}\{L\})^2} - \frac{4\,c}{\mathrm{tr}\{L\}}\,\mu_{k-1}\sum_m\sum_{b\neq m}\frac{l_m^k}{l_m-l_b}\,(Q_0^*)_{mm}. \tag{3.6.4}$$

Une condition suffisante pour que l'estimateur $\hat{\Sigma}_{EM}^{-1}$ améliore a S^+ est que $\delta(G)$ dans (3.6.4) soit négatif, c'est à dire soit tel que

$$0 \le c \le 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\{(n\vee p)+2\,(n\wedge p)^2-3\,(n\wedge p)+2\,k-3\}-2\,a, \tag{3.6.5}$$

pour $k = 0, 1, 2, ...$

Notons que la condition (3.6.5) impose que

$$0 < a \le \frac{\mu_{k-1}}{\mu_{k-2}}\{(n\vee p)+2\,(n\wedge p)^2-3\,(n\wedge p)+2\,k-3\}\,.$$

Lorsque S est inversible et de loi de Wishart (comme dans le modèle 1.4.2) la condition ci-dessus devient

$$0 < a \le n+2\,p^2-3\,p-3,$$

Notons que, dans ce cas, nous obtenons un intervalle de variation de a plus large que celui obtenu par Tsukuma et Konno [TK06] pour lesquels

$$n-p-5 \le a \le n-p-3.$$

Dans la preuve du théorème 3.6.1 ci-dessous, nous allons voir le rôle crucial de la forme de la matrice Q_0 dans l'expression de l'estimateur $\hat{\Sigma}_{EM}^{-1}$ dans (3.6.2).

Preuve

Tout d'abord, en remplaçant $G(S)$ donné par (3.6.3) dans (3.2.3), l'estimateur sans biais de la différence en risque entre l'estimateur de type Efron-Morris dans (3.6.2) et l'estimateur usuel dans (1.2.1) a pour expression

$$\delta(G) = c^2\,\mu_{k-2}\,\frac{\mathrm{tr}\{\tilde{S}^k\,(\tilde{S}\,\tilde{S^+}\,Q_0)^2\}}{(\mathrm{tr}\{\tilde{S}\})^2} + 2\,c\,\left[a\,\mu_{k-2} - \mu_{k-1}\,(n-(n\wedge p)-1)\right]\frac{\mathrm{tr}\{\tilde{S}^+\,\tilde{S}^{k+1}\tilde{S}^+\,Q_0\}}{\mathrm{tr}\{\tilde{S}\}}$$
$$-\,4\,c\,\mu_{k-1}\,\mathrm{tr}\left\{\tilde{S}^+\,\tilde{S}\,D_{\tilde{S}}\left\{Q_0^\top\,\frac{\tilde{S}^+\,\tilde{S}^{k+1}}{\mathrm{tr}\{\tilde{S}\}}\right\}\right\}. \tag{3.6.6}$$

Maintenant, en utilisant le fait que $D_{\tilde{S}}\{A\,B\} = D_{\tilde{S}}\{A\}\,B + \{A^\top D_{\tilde{S}}\}^\top B$ avec $A = Q_0^\top$ et $B = \tilde{S}^+\,\tilde{S}^{k+1}/\mathrm{tr}\{\tilde{S}\}$, l'estimateur $\delta(G)$ dans (3.6.6) devient

$$\begin{aligned}\delta(G) = c^2\,\mu_{k-2}\,\frac{\mathrm{tr}\{\tilde{S}^k\,(\tilde{S}\,\tilde{S^+}\,Q_0)^2\}}{(\mathrm{tr}\{\tilde{S}\})^2} + 2\,c\,\big[a\,\mu_{k-2} - \mu_{k-1}\,(n-(n\wedge p)-1)\big]\,\frac{\mathrm{tr}\{\tilde{S}^+\,\tilde{S}^k Q_0\}}{\mathrm{tr}\{\tilde{S}\}}\\ - 4\,c\,\mu_{k-1}\,\mathrm{tr}\left\{\tilde{S}^+\,\tilde{S}\,D_{\tilde{S}}\,\{Q_0^\top\}\,\frac{\tilde{S}^+\,\tilde{S}^{k+1}}{\mathrm{tr}\{\tilde{S}\}}\right\} - 4\,c\,\mu_{k-1}\,\mathrm{tr}\left\{\tilde{S}^+\,\tilde{S}\,\{Q_0 D_{\tilde{S}}\}^\top\,\frac{\tilde{S}^+\,\tilde{S}^{k+1}}{\mathrm{tr}\{\tilde{S}\}}\right\}.\end{aligned} \tag{3.6.7}$$

Notons que

$$\begin{aligned}\mathrm{tr}\{\tilde{S}^k\,(\tilde{S}\,\tilde{S^+}\,Q_0)^2\} &= \mathrm{tr}\{\tilde{S}^k\,\tilde{S}\,\tilde{S^+}\,Q_0\,\tilde{S}\,\tilde{S^+}\,Q_0\}\\ &= \mathrm{tr}\{\tilde{S}^k\,\tilde{S}\,\tilde{S^+}\,H_1\,Q_0^*\,H_1^\top\,H_1\,H_1^\top\,H_1\,Q_0^*\,H_1^\top\}\\ &= \mathrm{tr}\{\tilde{S}^{k+1}\,\tilde{S^+}\,H_1\,Q_0^{*^2}\,H_1^\top\}\\ &= \mathrm{tr}\{\tilde{S}^{k+1}\,\tilde{S^+}\,H_1\,Q_0^*\,H_1^\top\}\\ &= \mathrm{tr}\{\tilde{S}^{k+1}\,\tilde{S^+}\,Q_0\},\end{aligned} \tag{3.6.8}$$

car $H_1^\top\,H_1 = I_{p\wedge n}$ et Q_0^* est un projecteur.

À partir de (3.6.8) et en appliquant l'égalité (A.17) du lemme A.11 (voir appendice 6) au troisième terme dans (3.6.7), l'expression de $\delta(G)$ peut être réécrite comme suit :

$$\begin{aligned}\delta(G) = c^2\,\mu_{k-2}\,\frac{\mathrm{tr}\{\tilde{S}^{k+1}\,\tilde{S^+}\,Q_0\}}{(\mathrm{tr}\{\tilde{S}\})^2} + 2\,c\,\big[a\,\mu_{k-2} - \mu_{k-1}\,(n-(n\wedge p)-1)\big]\,\frac{\mathrm{tr}\{\tilde{S}^+\,\tilde{S}^k Q_0\}}{\mathrm{tr}\{\tilde{S}\}}\\ - \frac{2\,c}{\mathrm{tr}\{\tilde{S}\}}\,\mu_{k-1}\,\mathrm{tr}\{\tilde{S}^+\,\tilde{S}^{k+1}\,H_1\,\Psi\,Q_0^*\,H_1^\top\} + \frac{2\,c}{\mathrm{tr}\{\tilde{S}\}}\,\mu_{k-1}\,\sum_m\sum_{b\neq m}\frac{1}{l_m-l_b}\,l_b^k\,(Q_0^*)_{mm}.\\ - 4\,c\,\mu_{k-1}\,\mathrm{tr}\left\{\tilde{S}^+\,\tilde{S}\,\{Q_0 D_{\tilde{S}}\}^\top\,\frac{\tilde{S}^+\,\tilde{S}^{k+1}}{\mathrm{tr}\{\tilde{S}\}}\right\},\end{aligned} \tag{3.6.9}$$

où $\Psi = \mathrm{diag}\left(\Psi_1, ..., \Psi_{\mathrm{p}\wedge\mathrm{n}}\right)$ et $\Psi_k = \sum_{r\neq k}\frac{1}{l_k-l_r} + \frac{p-(n\wedge p)}{l_k}$.

Maintenant, en utilisant l'égalité (A.1) (voir appendice 6) avec $A = \tilde{S}^+\,\tilde{S}$, $C = Q_0$ et $B = \tilde{S}^+\,\tilde{S}^{k+1}/\mathrm{tr}\{\tilde{S}\}$, le dernier terme dans (3.6.9) vaut

$$\mathrm{tr}\left\{\tilde{S}^+\,\tilde{S}\,\{Q_0\,D_{\tilde{S}}\}^\top\,\frac{\tilde{S}^+\,\tilde{S}^{k+1}}{\mathrm{tr}\{\tilde{S}\}}\right\} = \mathrm{tr}\left\{Q_0\,\{\tilde{S}^+\,\tilde{S}D_{\tilde{S}}\}^\top\,\frac{\tilde{S}^+\,\tilde{S}^{k+1}}{\mathrm{tr}\{\tilde{S}\}}\right\},$$

puisque $\tilde{S}^+\,\tilde{S}^{k+1}$ est symétrique. En remplaçant le terme ci-dessus dans (3.6.9), $\delta(G)$ devient

$$\begin{aligned}\delta(G) = c^2\,\mu_{k-2}\,\frac{\mathrm{tr}\{\tilde{S}^k\,(\tilde{S}\,\tilde{S^+}\,Q_0)^2\}}{(\mathrm{tr}\{\tilde{S}\})^2} + 2\,c\big[a\,\mu_{k-2} - \mu_{k-1}\,(n-(n\wedge p)-1)\big]\,\frac{\mathrm{tr}\{\tilde{S}^+\,\tilde{S}^k Q_0\}}{\mathrm{tr}\{\tilde{S}\}}\\ - \frac{2\,c}{\mathrm{tr}\{\tilde{S}\}}\,\mu_{k-1}\,\mathrm{tr}\{\tilde{S}^+\,\tilde{S}^{k+1}\,H_1\,\Psi\,Q_0^*\,H_1^\top\} + \frac{2\,c}{\mathrm{tr}\{\tilde{S}\}}\,\mu_{k-1}\,\sum_m\sum_{b\neq m}\frac{1}{l_m-l_b}\,l_b^k\,(Q_0^*)_{mm}.\\ - 4\,c\,\mu_{k-1}\,\mathrm{tr}\left\{Q_0\,\{\tilde{S}^+\,\tilde{S}D_{\tilde{S}}\}^\top\,\frac{\tilde{S}^+\,\tilde{S}^{k+1}}{\mathrm{tr}\{\tilde{S}\}}\right\}.\end{aligned} \tag{3.6.10}$$

En appliquant l'égalité (A.2) (voir appendice 6) avec $\Phi(L) = L^k/\mathrm{tr}\{L\}$, le dernier terme dans (3.6.10) est

$$\begin{aligned}
\mathrm{tr}\left\{Q_0\,\{\tilde{S}^+\,\tilde{S}D_{\tilde{S}}\}^\top\,\frac{\tilde{S}^+\,\tilde{S}^{k+1}}{\mathrm{tr}\{\tilde{S}\}}\right\} &= \mathrm{tr}\left\{Q_0\,\{H_1\,H_1^\top\,D_{\tilde{S}}\}^\top\,H_1\,\frac{L^k}{\mathrm{tr}\{L\}}\,H_1^\top\right\}\\
&= \mathrm{tr}\left\{Q_0\left[D_{\tilde{S}}\left\{H_1\,\frac{L^k}{\mathrm{tr}\{L\}}\,H_1^\top\right\} - \frac{1}{2}(p-n)\,H_1\,\frac{L^{k-1}}{\mathrm{tr}\{L\}}\,H_1^\top\right]\right\}\\
&= \mathrm{tr}\left\{Q_0\left[D_{\tilde{S}}\left\{\frac{\tilde{S}^+\,\tilde{S}^{k+1}}{\mathrm{tr}\{\tilde{S}\}}\right\} - \frac{1}{2}\big(p-(n\wedge p)\big)\,\frac{\tilde{S}^k\,\tilde{S}^+}{\mathrm{tr}\{\tilde{S}\}}\right]\right\}.
\end{aligned} \tag{3.6.11}$$

Maintenant, pour exprimer le terme $D_{\tilde{S}}\left\{\tilde{S}^+\,\tilde{S}^{k+1}/\mathrm{tr}\{\tilde{S}\}\right\}$ dans (3.6.11), on utilise l'égalité (A.3) (voir appendice 6) avec $\phi_i = \dfrac{l_i^k}{\mathrm{tr}\{L\}}$ et $\dfrac{\partial\phi_i}{\partial l_i} = k\,\dfrac{l_i^{k-1}}{\mathrm{tr}\{L\}} - \dfrac{l_i^k}{(\mathrm{tr}\{L\})^2}$. On obtient alors

$$\begin{aligned}
D_{\tilde{S}}\left\{\frac{\tilde{S}^+\,\tilde{S}^{k+1}}{\mathrm{tr}\{\tilde{S}\}}\right\} &= D_{\tilde{S}}\left\{\frac{H_1\,L^k\,H_1^\top}{\mathrm{tr}\{L\}}\right\}\\
&= \frac{1}{2}\left[p-(n\wedge p)\right]\frac{\tilde{S}^k\,\tilde{S}^+}{\mathrm{tr}\{\tilde{S}\}} + k\,\frac{\tilde{S}^k\,\tilde{S}^+}{\mathrm{tr}\{\tilde{S}\}} - \frac{\tilde{S}^{k+1}\,\tilde{S}^+}{(\mathrm{tr}\{\tilde{S}\})^2}\\
&\quad + \frac{1}{2}\,\frac{1}{\mathrm{tr}\{\tilde{S}\}}\sum_{i\neq j}^{n\wedge p}\frac{l_i^k - l_j^k}{l_i - l_j} + \frac{1}{2}\,\frac{\mathrm{tr}\{\tilde{S}^k\,\tilde{S}^+\}}{\mathrm{tr}\{\tilde{S}\}}\,\big(I_p - H_1\,H_1^\top\big).
\end{aligned} \tag{3.6.12}$$

En remplaçant (3.6.12) dans (3.6.11), on a

$$\begin{aligned}
\mathrm{tr}\left\{Q_0\,\{\tilde{S}^+\,\tilde{S}D_{\tilde{S}}\}^\top\,\frac{\tilde{S}^+\,\tilde{S}^{k+1}}{\mathrm{tr}\{\tilde{S}\}}\right\} &= k\,\frac{\mathrm{tr}\{Q_0\,\tilde{S}^k\,\tilde{S}^+\}}{\mathrm{tr}\{\tilde{S}\}} - \frac{\mathrm{tr}\{Q_0\,\tilde{S}^{k+1}\,\tilde{S}^+\}}{(\mathrm{tr}\{\tilde{S}\})^2}\\
&\quad + \frac{1}{2}\,\frac{1}{\mathrm{tr}\{\tilde{S}\}}\,\mathrm{tr}\left\{Q_0\,H_1\,\mathrm{diag}\left(\sum_{j\neq i}^{n\wedge p}\frac{l_i^k - l_j^k}{l_i - l_j}\right)H_1^\top\right\}\\
&\quad + \frac{1}{2}\left[\mathrm{tr}\{Q_0\} - \mathrm{tr}\{Q_0\;\tilde{S}\,\tilde{S}^+\}\right]\frac{\mathrm{tr}\{\,\tilde{S}^k\,\tilde{S}^+\}}{\mathrm{tr}\{\tilde{S}\}}\\
&= k\,\frac{\mathrm{tr}\{Q_0\,\tilde{S}^k\,\tilde{S}^+\}}{\mathrm{tr}\{\tilde{S}\}} - \frac{\mathrm{tr}\{Q_0\,\tilde{S}^{k+1}\,\tilde{S}^+\}}{(\mathrm{tr}\{\tilde{S}\})^2}\\
&\quad + \frac{1}{2}\,\frac{1}{\mathrm{tr}\{\tilde{S}\}}\,\mathrm{tr}\left\{Q_0\,H_1\,\mathrm{diag}\left(\sum_{j\neq i}^{n\wedge p}\frac{l_i^k - l_j^k}{l_i - l_j}\right)H_1^\top\right\},
\end{aligned} \tag{3.6.13}$$

car $\mathrm{tr}\{Q_0\;\tilde{S}\,\tilde{S}^+\} = \mathrm{tr}\{H_1\,Q_0^*\,H_1^\top\,H_1\,H_1^\top\} = \mathrm{tr}\{Q_0\}$. Finalement, en remplaçant le terme (3.6.13)

dans (3.6.10), $\delta(G)$ devient

$$\begin{aligned}
\delta(G) = {} & c^2\,\mu_{k-2}\,\frac{\mathrm{tr}\{\tilde{S}^{k+1}\,\tilde{S^+}\,Q_0\}}{(\mathrm{tr}\{\tilde{S}\})^2} + 2\,c\,\big[a\,\mu_{k-2} - \mu_{k-1}\,(n-(n\wedge p)-1)\big]\,\frac{\mathrm{tr}\{\tilde{S^+}\,\tilde{S}^k Q_0\}}{\mathrm{tr}\{\tilde{S}\}} \\
& - \frac{2\,c}{\mathrm{tr}\{\tilde{S}\}}\,\mu_{k-1}\,\mathrm{tr}\{\tilde{S^+}\,\tilde{S}^{k+1}\,H_1\,\Psi\,Q_0^*\,H_1^\top\} + \frac{2\,c}{\mathrm{tr}\{\tilde{S}\}}\,\mu_{k-1}\,\sum_m\sum_{b\neq m}\frac{l_b^k}{l_m-l_b}\,(Q_0^*)_{mm}. \\
& - 4\,c\,k\,\mu_{k-1}\,\frac{\mathrm{tr}\{Q_0\,\tilde{S}^k\,\tilde{S^+}\}}{\mathrm{tr}\{\tilde{S}\}} + 4\,c\,\mu_{k-1}\,\frac{\mathrm{tr}\{Q_0\,\tilde{S}^{k+1}\,\tilde{S^+}\}}{(\mathrm{tr}\{\tilde{S}\})^2} \\
& - \frac{2\,c}{\mathrm{tr}\{\tilde{S}\}}\,\mu_{k-1}\,\mathrm{tr}\left\{Q_0\,H_1\,\mathrm{diag}\left(\sum_{j\neq i}^{n\wedge p}\frac{l_i^k-l_j^k}{l_i-l_j}\right)H_1^\top\right\} \\
= {} & c^2\,\mu_{k-2}\,\frac{\mathrm{tr}\{\tilde{S}^{k+1}\,\tilde{S^+}\,Q_0\}}{(\mathrm{tr}\{\tilde{S}\})^2} + 2\,c\,\big[a\,\mu_{k-2} - \mu_{k-1}\,(n-(n\wedge p)+2\,k-1)\big]\,\frac{\mathrm{tr}\{\tilde{S^+}\,\tilde{S}^k Q_0\}}{\mathrm{tr}\{\tilde{S}\}} \\
& - \frac{2\,c}{\mathrm{tr}\{\tilde{S}\}}\,\mu_{k-1}\,\mathrm{tr}\{\tilde{S^+}\,\tilde{S}^{k+1}\,H_1\,\Psi\,Q_0^*\,H_1^\top\} + \frac{2\,c}{\mathrm{tr}\{\tilde{S}\}}\,\mu_{k-1}\,\sum_m\sum_{b\neq m}\frac{l_b^k}{l_m-l_b}\,(Q_0^*)_{mm}. \\
& + 4\,c\,\mu_{k-1}\,\frac{\mathrm{tr}\{Q_0\,\tilde{S}^{k+1}\,\tilde{S^+}\}}{(\mathrm{tr}\{\tilde{S}\})^2} - \frac{2\,c}{\mathrm{tr}\{\tilde{S}\}}\,\mu_{k-1}\,\mathrm{tr}\left\{Q_0\,H_1\,\mathrm{diag}\left(\sum_{j\neq i}^{n\wedge p}\frac{l_i^k-l_j^k}{l_i-l_j}\right)H_1^\top\right\}.
\end{aligned} \tag{3.6.14}$$

En exprimant $\delta(G)$ dans (3.6.14) en fonction de L, on obtient

$$\begin{aligned}
\delta(G) = {} & c^2\,\mu_{k-2}\,\frac{\mathrm{tr}\{L^k\,Q_0^*\}}{(\mathrm{tr}\{L\})^2} + 2\,c\,\big[a\,\mu_{k-2} - \mu_{k-1}\,(n-(n\wedge p)+2\,k-1)\big]\,\frac{\mathrm{tr}\{L^{k-1}Q_0^*\}}{\mathrm{tr}\{L\}} \\
& - \frac{2\,c}{\mathrm{tr}\{L\}}\,\mu_{k-1}\,\mathrm{tr}\{L^k\,\Psi\,Q_0^*\} + \frac{2\,c}{\mathrm{tr}\{L\}}\,\mu_{k-1}\,\sum_m\sum_{b\neq m}\frac{l_b^k}{l_m-l_b}\,(Q_0^*)_{mm}. \\
& + 4\,c\,\mu_{k-1}\,\frac{\mathrm{tr}\{Q_0^*\,L^k\}}{(\mathrm{tr}\{L\})^2} - \frac{2\,c}{\mathrm{tr}\{L\}}\,\mu_{k-1}\,\mathrm{tr}\left\{Q_0^*\,\mathrm{diag}\left(\sum_{j\neq i}^{n\wedge p}\frac{l_i^k-l_j^k}{l_i-l_j}\right)\right\},
\end{aligned} \tag{3.6.15}$$

où $\Psi = \mathrm{diag}\left(\Psi_1, ..., \Psi_{\mathrm{p}\wedge\mathrm{n}}\right)$ et $\Psi_k = \sum_{r\neq k}\dfrac{1}{l_k-l_r} + \dfrac{p-(n\wedge p)}{l_k}$.

Notons que

$$\mathrm{tr}\{L^k\,\Psi\,Q_0^*\} = \sum_m\sum_{b\neq m}\frac{l_b^k}{l_m-l_b}\,(Q_0^*)_{mm} + \big[p-(n\wedge p)\big]\sum_m l_m^{k-1}\,(Q_0^*)_{mm} \tag{3.6.16}$$

et

$$\begin{aligned}
\mathrm{tr}\left\{Q_0^*\,\mathrm{diag}\left(\sum_{j\neq i}^{n\wedge p}\frac{l_i^k-l_j^k}{l_i-l_j}\right)\right\} &= \sum_m (Q_0^*)_{mm}\sum_{b\neq m}\frac{l_m^k-l_b^k}{l_m-l_b} \\
&= \sum_m\sum_{b\neq m}\frac{l_m^k}{l_m-l_b}\,(Q_0^*)_{mm} - \sum_m\sum_{b\neq m}\frac{l_b^k}{l_m-l_b}\,(Q_0^*)_{mm}.
\end{aligned} \tag{3.6.17}$$

En remplaçant de (3.6.16) et (3.6.17) dans (3.6.15), on obtient

$$\begin{aligned}
\delta(G) &= c^2\,\mu_{k-2}\,\frac{\mathrm{tr}\{L^k\,Q_0^*\}}{(\mathrm{tr}\{L\})^2} + 2\,c\,\left[a\,\mu_{k-2} - \mu_{k-1}\,(n-(n\wedge p)+2\,k-1)\right]\frac{\mathrm{tr}\{L^{k-1}Q_0^*\}}{\mathrm{tr}\{L\}} \\
&\quad - \frac{2\,c}{\mathrm{tr}\{L\}}\,\mu_{k-1}\sum_m\sum_{b\neq m}\frac{l_m^k}{l_m-l_b}\,(Q_0^*)_{mm} \\
&\quad - \frac{2\,c}{\mathrm{tr}\{L\}}\,\mu_{k-1}\left[p-(n\wedge p)\right]\sum_m l_m^{k-1}\,(Q_0^*)_{mm} + 4\,c\,\mu_{k-1}\,\frac{\mathrm{tr}\{Q_0^*\,L^k\}}{(\mathrm{tr}\{L\})^2} \\
&\quad - \frac{2\,c}{\mathrm{tr}\{L\}}\,\mu_{k-1}\sum_m\sum_{b\neq m}\frac{l_m^k}{l_m-l_b}\,(Q_0^*)_{mm} \\
&= c^2\,\mu_{k-2}\,\frac{\mathrm{tr}\{L^k\,Q_0^*\}}{(\mathrm{tr}\{L\})^2} + 2\,c\,\left[a\,\mu_{k-2} - \mu_{k-1}\,(n+p-2\,(n\wedge p)+2\,k-1)\right]\frac{\mathrm{tr}\{L^{k-1}Q_0^*\}}{\mathrm{tr}\{L\}} \\
&\quad + 4\,c\,\mu_{k-1}\,\frac{\mathrm{tr}\{Q_0^*\,L^k\}}{(\mathrm{tr}\{L\})^2} - \frac{4\,c}{\mathrm{tr}\{L\}}\,\mu_{k-1}\sum_m\sum_{b\neq m}\frac{l_m^k}{l_m-l_b}\,(Q_0^*)_{mm},
\end{aligned} \tag{3.6.18}$$

puisque $\sum_m l_m^{k-1}\,(Q_0^*)_{mm} = \mathrm{tr}\{L^{k-1}Q_0^*\}$. D'où l'expression (3.6.4) du théorème.

Maintenant la détermination de la condition de domination (3.6.5) se fait à travers une borne supérieure de $\delta(G)$. Pour se faire, notons que

$$\frac{1}{l_m-l_b} \geq \frac{1}{l_m}$$

et qu'alors

$$\sum_{b\neq m}\frac{1}{l_m-l_b} \geq \frac{1}{l_m}\sum_{b\neq m} 1.$$

Donc

$$\sum_{b\neq m}\frac{1}{l_m-l_b} \geq (n\wedge p)\left[(n\wedge p)-1\right]\frac{1}{l_m}.$$

Grâce à l'inégalité précédente, il est clair que $\delta(G)$ dans (3.6.18) est majoré comme suit :

$$\begin{aligned}
\delta(G) &\leq c^2\,\mu_{k-2}\,\frac{\mathrm{tr}\{L^k\,Q_0^*\}}{(\mathrm{tr}\{L\})^2} + 2\,c\,\left[a\,\mu_{k-2} - \mu_{k-1}\,(n+p-2\,(n\wedge p)+2\,k-1)\right]\frac{\mathrm{tr}\{L^{k-1}Q_0^*\}}{\mathrm{tr}\{L\}} \\
&\quad + 4\,c\,\mu_{k-1}\,\frac{\mathrm{tr}\{Q_0^*\,L^k\}}{(\mathrm{tr}\{L\})^2} - \frac{4\,c}{\mathrm{tr}\{L\}}\,(n\wedge p)\left[(n\wedge p)-1\right]\mu_{k-1}\sum_m l_m^{k-1}\,(Q_0^*)_{mm} \\
&= c^2\,\mu_{k-2}\,\frac{\mathrm{tr}\{L^k\,Q_0^*\}}{(\mathrm{tr}\{L\})^2} + 2\,c\,\left[a\,\mu_{k-2} - \mu_{k-1}\,\{(n\vee p)+2\,(n\wedge p)^2-3\,(n\wedge p)\right. \\
&\qquad \left. +2\,k-1\}\right]\frac{\mathrm{tr}\{L^{k-1}Q_0^*\}}{\mathrm{tr}\{L\}} + 4\,c\,\mu_{k-1}\,\frac{\mathrm{tr}\{Q_0^*\,L^k\}}{(\mathrm{tr}\{L\})^2}.
\end{aligned} \tag{3.6.19}$$

En appliquant l'inégalité (A.6) du lemme A.5 (voir appendice 6) aux deux rapports des traces présents au premier et dernier termes dans (3.6.19), on obtient

$$\begin{aligned}
\delta(G) &\leq c^2\,\mu_{k-2}\,\frac{\mathrm{tr}\{L^{k-1}\,Q_0^*\}}{\mathrm{tr}\{L\}} + 2\,c\,\left[a\,\mu_{k-2} - \mu_{k-1}\,\{(n\vee p)+2\,(n\wedge p)^2-3\,(n\wedge p)\right. \\
&\qquad \left. +2\,k-1\}\right]\frac{\mathrm{tr}\{L^{k-1}Q_0^*\}}{\mathrm{tr}\{L\}} + 4\,c\,\mu_{k-1}\,\frac{\mathrm{tr}\{Q_0^*\,L^{k-1}\}}{\mathrm{tr}\{L\}}.
\end{aligned} \tag{3.6.20}$$

Par conséquent, $\delta(G)$ est négatif si

$$0 \leq c \leq 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\{(n \vee p) + 2\,(n \wedge p)^2 - 3\,(n \wedge p) + 2\,k - 3\} - 2\,a. \tag{3.6.21}$$

□

Maintenant, on se basant sur la proposition (3.3.1), nous obtenons le corollaire suivant qui est une conséquence de théorème (3.6.1). Celui ci fournit les conditions sur la constante c sous lesquelles l'estimateur de type Efron-Morris dans (3.6.2) améliore l'estimateur usuel dans (1.2.1) avec les valeurs de $a = a^*, a_0^-, a_0^+$ données dans la proposition (3.3.1), pour différentes valeurs de k.

Corollaire 3.6.1

Dans le contexte de la proposition 3.3.1, on donne des conditions sur les constantes a et c sous lesquelles l'estimateur de type Dey $\hat{\Sigma}_{EM}^{-1} = a\,S^+ + c\,S\,S^+\,Q_0/\mathrm{tr}\{S\}$ améliore l'estimateur $a\,S^+$ suivant les différentes valeurs de k.

- *Cas $k = 0$: $a < a_0^-$ et $0 \leq c \leq 2\,\dfrac{\mu_{-1}}{\mu_{-2}}\,\{(n \wedge p)[2\,(n \wedge p) - 1] - 1\}$ avec $n \wedge p \geq 1$;*
  $a > a_0^+$ et $0 \leq c \leq 2\,\dfrac{\mu_{-1}}{\mu_{-2}}\,\{(n \wedge p)\,[2\,(n \wedge p) - 2]\}$ avec $n \wedge p \geq 1$.
- *Cas $k = 1$: $a = a^*$ et $0 \leq c \leq \dfrac{2}{\mu_{-1}}\,\{(n \wedge p)[2\,(n \wedge p) - 2]\}$ avec $n \wedge p \geq 2$.*
- *Cas $k = 2$: $a = a^*$ et $0 \leq c \leq 2\,\mu_1\,\{(n \wedge p)[2\,(n \wedge p) - 3] + 1\}$ avec $n \wedge p \geq 2$.*
- *Cas $k = 3$: $a = a^*$ et $0 \leq c \leq 2\,\dfrac{\mu_2}{\mu_1}\,\{(n \wedge p)[2\,(n \wedge p) - 4] + 2\}$ avec $n \wedge p \geq 2$.*
- *Cas $k \geq 4$: $a < a_0^-$ et $0 \leq c \leq 2\,\dfrac{\mu_{k-1}}{\mu_{k-2}}\,\{(n \wedge p)[2\,(n \wedge p) - 2] + k - 1\}$ avec $n \wedge p \geq 1$;*
  $a > a_0^+$ et $0 \leq c \leq 2\,\dfrac{\mu_{k-1}}{\mu_{k-2}}\,\{(n \wedge p)\,[2\,(n \wedge p) - k - 1] + k - 1\}$ avec
  $(n \wedge p)\,[2\,(n \wedge p) - k + 1] \geq 1 - k$.

3.6.2 Une classe plus grande d'estimateurs améliorés

On considère maintenant des estimateurs de la forme

$$\hat{\Sigma}_{a,c,r}^{-1} = a\,S^+ + c\,r\big(\mathrm{tr}\{S\}\big)\,S\,G(S)\,, \tag{3.6.22}$$

où c est une constante positive, $G(S)$ est une fonction matricielle homogène d'ordre -2 et r est une fonction réelle définie et dérivable sur $\mathbb{R}_+$. Ici, G n'est pas nécessairement orthogonalement invariant. Nous allons voir dans le théorème suivant des conditions de domination de l'estimateur $\hat{\Sigma}_{a,c,r}^{-1}$ dans (3.6.22) sur l'estimateur usuel $\hat{\Sigma}_a^{-1}$ dans (1.2.1).

Théorème 3.6.2

Sous le modèle (1.4.1) et sous les coûts (1.2.4), supposons que la fonction $G(S)$ dans (3.6.22) soit homogène d'ordre -2 et faiblement différentiable. Supposons aussi que, pour certaines valeurs positives de c, l'expression dans (3.2.3) soit négative, et donc que l'estimateur $\hat{\Sigma}_{a,c}^{-1} = a\,S^+ + c\,S\,G(S)$ dans (3.2.1) améliore l'estimateur $\hat{\Sigma}_a^{-1} = a\,S^+$ dans (1.2.1).

Alors, l'estimateur $\hat{\Sigma}_{a,c,r}^{-1} = a\,S^+ + c\,r\big(\mathrm{tr}\{S\}\big)\,S\,G(S)$ dans (3.6.22) améliore l'estimateur $\hat{\Sigma}_a^{-1} = a\,S^+$ si pour tout $k = 0, 1, 2, ...$, $\mathrm{tr}\big\{G^\top(S)\,S^k\big\} = \mathrm{tr}\big\{G(S)\,S^k\big\} \geq 0$ et $\mathrm{tr}\big\{\,S^k\,[S\,G(S)]^2\big\} \geq 0$ et si pour tout $t \geq 0$, $0 \leq r(t) \leq 1$ et $r'(t) \geq 0$.

Preuve

Supposons que les risques $R_k\left(\Sigma^{-1}, \hat{\Sigma}_a^{-1}\right)$ et $R_k\left(\Sigma^{-1}, \hat{\Sigma}_{a,c,r}^{-1}\right)$ soient finis. La différence de risque entre les estimateurs $\hat{\Sigma}_{a,c,r}^{-1}$ et $\hat{\Sigma}_a^{-1}$ est donnée par

$$\begin{aligned}
\Delta(G) &= R_k\left(\Sigma^{-1}, a\,S^+ + c\;r(\mathrm{tr}\{S\})\,S\,G(S)\right) - R_k\left(\Sigma^{-1}, a\,S^+\right) \\
&= \mathrm{E}_\Sigma\left[\mathrm{tr}[\{a\,S^+ - \Sigma^{-1} + c\;r(\mathrm{tr}\{S\})\,S\,G(S)\}^2\,S^k - (a\,S^+ - \Sigma^{-1})^2\,S^k]\right] \\
&= \mathrm{E}_\Sigma\left[c^2\;r^2(\mathrm{tr}\{S\})\,\mathrm{tr}\{S^k\,[S\,G(S)]^2\} + 2\,c\;r(\mathrm{tr}\{S\})\,\mathrm{tr}\{S\,G(S)(a\,S^+ - \Sigma^{-1})\,S^k\}\right] \\
&= \mathrm{E}_\Sigma\left[c^2\;r^2(\mathrm{tr}\{S\})\,\mathrm{tr}\{S^k\,[S\,G(S)]^2\} + 2\,a\,c\,r(\mathrm{tr}\{S\})\;\mathrm{tr}\{S^+\,S^{k+1}\,G(S)\}\right. \\
&\qquad \left. -2\,c\,\mathrm{tr}\{\Sigma^{-1}\,S^{k+1}\;r(\mathrm{tr}\{S\})\,G(S)\}\right]. \qquad (3.6.23)
\end{aligned}$$

Notons que la différence $\Delta(G)$ dans (3.6.23) peut être réécrite sous la forme

$$\Delta(G) = \mathrm{E}_\Sigma\left[\eta(S) + \beta_\Sigma(S)\right],$$

où

$$\eta(S) = c^2\;r^2(\mathrm{tr}\{S\})\,\mathrm{tr}\{S^k\,[S\,G(S)]^2\} + 2\,a\,c\,r(\mathrm{tr}\{S\})\;\mathrm{tr}\{S^+\,S^{k+1}\,G(S)\},$$

et

$$\beta_\Sigma(S) = -2\,c\,\mathrm{tr}\{\Sigma^{-1}\,S^{k+1}\;r(\mathrm{tr}\{S\})\,G(S)\}.$$

de sorte que, conditionnellement par rapport la variable V on ait

$$\Delta(G) = \mathrm{E}_{\mathcal{H}}\left[\gamma(V)\right], \qquad (3.6.24)$$

où $\mathrm{E}_{\mathcal{H}}$ est l'espérance associée à la loi $\mathcal{H}$ et où

$$\gamma(V) = \mathrm{E}_\Sigma\left[\eta(S) \mid V\right] + \mathrm{E}_\Sigma\left[\beta_\Sigma(S) \mid V\right].$$

Maintenant, grâce au changement de variable $S = V\,\tilde{S}$, il vient que

$$\gamma(V) = \mathrm{E}_\Sigma\left[\eta(V\,\tilde{S}) \mid V\right] + \mathrm{E}_\Sigma\left[\beta_\Sigma(V\,\tilde{S}) \mid V\right], \qquad (3.6.25)$$

où, utilisant l'ordre d'homogénéité $\alpha = 2$ de $G(S)$,

$$\eta(V\,\tilde{S}) = c^2\;r^2(V\,\mathrm{tr}(\tilde{S}))\,V^{k-2}\,\mathrm{tr}\{\tilde{S}^k[\tilde{S}\,G(\tilde{S})]^2\} + 2\,a\,c\;r(V\,\mathrm{tr}\{\tilde{S}\})\,V^{k-2}\,\mathrm{tr}\{\tilde{S}^+\,\tilde{S}^{k+1}\,G(\tilde{S})\}, \qquad (3.6.26)$$

et

$$\beta_\Sigma(V\,\tilde{S}) = -2\,c\,V^{k-1}\,\mathrm{tr}\{\Sigma^{-1}\,\tilde{S}^{k+1}\,r(V\,\mathrm{tr}\{\tilde{S}\})\,G(\tilde{S})\}.$$

Alors, en appliquant le lemme (2.5.3) au deuxieme terme de $\gamma(V)$ dans (3.6.25) où $G(\tilde{S})$ est remplacé par $\tilde{S}^k\,r(\mathrm{tr}\{V\,\tilde{S}\})\,G(\tilde{S})$, on obtient

$$\mathrm{E}_\Sigma\left[\beta_\Sigma(V\,\tilde{S}) \mid V\right] = \mathrm{E}_\Sigma\left[\beta(V\,\tilde{S}) \mid V\right],$$

avec

$$\begin{aligned}
\beta(V\,\tilde{S}) = &-2\,c\,r(V\,\mathrm{tr}\{\tilde{S}\})\,V^{k-1}\,(n - (n \wedge p) - 1)\,\mathrm{tr}\{\tilde{S^+}\,\tilde{S}^{k+1}\,G(\tilde{S})\} \\
&- 4\,c\,V^{k-1}\,\mathrm{tr}\{\tilde{S^+}\,\tilde{S}\,D_{\tilde{S}}\,\{r(V\,\mathrm{tr}\{\tilde{S}\})\,G^\top(\tilde{S})\,\tilde{S}^{k+1}\}\}. \qquad (3.6.27)
\end{aligned}$$

Grâce aux égalités (3.6.26) et (3.6.27), on reformule l'expression de $\gamma(V)$ dans (3.6.25), soit

$$\begin{aligned}\gamma(V) = \mathrm{E}_\Sigma \Big[& c^2\, r^2(V\,\mathrm{tr}(\tilde{S}))\, V^{k-2}\, \mathrm{tr}\{\tilde{S}^k[\tilde{S}\, G(\tilde{S})]^2\} \\ & +2\, a\, c\, r(V\,\mathrm{tr}\{\tilde{S}\})\, V^{k-2}\, \mathrm{tr}\{\tilde{S}^+\, \tilde{S}^{k+1}\, G(\tilde{S})\} \\ & -2\, c\, r(V\,\mathrm{tr}\{\tilde{S}\})\, V^{k-1}\, (n-(n\wedge p)-1)\, \mathrm{tr}\{\tilde{S^+}\, \tilde{S}^{k+1}\, G(\tilde{S})\} \\ & -4\, c\, V^{k-1}\, \mathrm{tr}\{\tilde{S^+}\, \tilde{S}\, D_{\tilde{S}}\, \{r(V\,\mathrm{tr}\{\tilde{S}\})\, G^\top(\tilde{S})\, \tilde{S}^{k+1}\}\} \mid V \Big]. \end{aligned} \tag{3.6.28}$$

En appliquant le lemme A.12 sur le dernier terme de (3.6.28) (voir appendice 6), on obtient

$$\begin{aligned}\gamma(V) = \mathrm{E}_\Sigma \Big[& c^2\, r^2(V\,\mathrm{tr}(\tilde{S}))\, V^{k-2}\, \mathrm{tr}\{\tilde{S}^k[\tilde{S}\, G(\tilde{S})]^2\} \\ & +2\, a\, c\, r(V\,\mathrm{tr}\{\tilde{S}\})\, V^{k-2}\, \mathrm{tr}\{\tilde{S}^+\, \tilde{S}^{k+1}\, G(\tilde{S})\} \\ & -2\, c\, r(V\,\mathrm{tr}\{\tilde{S}\})\, V^{k-1}\, (n-(n\wedge p)-1)\, \mathrm{tr}\{\tilde{S^+}\, \tilde{S}^{k+1}\, G(\tilde{S})\} \\ & -4\, c\, V^{k-1}\, r(V\,\mathrm{tr}\{\tilde{S}\})\, \mathrm{tr}\big\{\tilde{S}^+\, \tilde{S}\, D_{\tilde{S}}\{\, G^\top(\tilde{S})\, \tilde{S}^{k+1}\}\, \big\} \\ & -4\, c\, V^{k-1}\, G\, r'(V\,\mathrm{tr}\{\tilde{S}\})\, \mathrm{tr}\big\{G(\tilde{S})\, \tilde{S}^{k+1}\big\} \mid V \Big]. \end{aligned} \tag{3.6.29}$$

Comme $0 \le r(t) \le 1$, $c > 0$, $r'(\mathrm{tr}\{\tilde{S}\}) \ge 0$ et $\mathrm{tr}\big\{G^\top(\tilde{S})\, \tilde{S}^{k+1}\big\} \ge 0$, un majorant pour $\gamma(V)$ dans (3.6.29) est donné par

$$\gamma_1(V) = \mathrm{E}_\Sigma\big[r(V\,\mathrm{tr}\{\tilde{S}\})\, \big(A(\tilde{S})\, V^{k-2} - B(\tilde{S})\, V^{k-1}\big) \mid V\big] \tag{3.6.30}$$

avec

$$A(\tilde{S}) = \big(c^2\, \mathrm{tr}\{\tilde{S}^k[\tilde{S}\, G(\tilde{S})]^2\} + 2\, c\, a\, \mathrm{tr}\{\tilde{S}^+\, \tilde{S}^{k+1}\, G(\tilde{S})\}\big)$$

et

$$B(\tilde{S}) = 2\, c\, \big((n-(n\wedge p)-1)\, \mathrm{tr}\{\tilde{S}^+\, \tilde{S}^{k+1}\, G(\tilde{S})\} + 2\, \mathrm{tr}\big\{\tilde{S}^+\, \tilde{S}\, D_{\tilde{S}}\big\{G^\top(\tilde{S})\, \tilde{S}^{k+1}\big\}\big\}\big).$$

De par l'hypothèse sur les traces de $S^k\, G(S)$, on a $\mathrm{tr}\{\tilde{S}^+\, \tilde{S}^{k+1}\, G(\tilde{S})\} = \mathrm{tr}\{\tilde{S}^k\, G(\tilde{S})\} \ge 0$ et $\mathrm{tr}\big\{\, \tilde{S}^k\, [\tilde{S}\, G(\tilde{S})]^2\big\} \ge 0$, et donc $A(\tilde{S}) \ge 0$. En outre, par hypothèse, l'estimateur $\delta(G)$ dans (3.2.3) est négatif, soit

$$\delta(G) = A(\tilde{S})\, \mu_{k-2} - B(\tilde{S})\, \mu_{k-1} \le 0\,, \tag{3.6.31}$$

et donc le terme $B(\tilde{S})$ est aussi positif.

Il résulte de ce qui précède que la différence en risque dans (3.6.24) vérifie

$$\begin{aligned}\Delta(G) \le & \mathrm{E}_\mathcal{H}\Big[\mathrm{E}_\Sigma\big[r(V\,\mathrm{tr}\{\tilde{S}\})\, \big(A(\tilde{S})\, V^{k-2} - B(\tilde{S})\, V^{k-1}\big) \mid V\big]\Big] \\ \le & \mathrm{E}_\mathcal{H}\Big[\mathrm{E}_\Sigma\big[r(V_0(\tilde{S})\,\mathrm{tr}\{\tilde{S}\})\big]\, \mathrm{E}_\mathcal{H}\big[\big(A(\tilde{S})\, V^{k-2} - B(\tilde{S})\, V^{k-1}\big) \mid V\big]\Big], \end{aligned}$$

l'inégalité ci-dessus provenant du fait que $A(\tilde{S})\, V^{k-2} - B(\tilde{S})\, V^{k-1}$ change de signe du $+$ au - au point $V_0(\tilde{S}) = A(\tilde{S})/B(\tilde{S})$ et du fait que la fonction $r(V\,\mathrm{tr}\{\tilde{S}\})$ est croissante en V. Inversant le conditionnement et utilisant (3.6.31) et le fait que $r(t) \ge 0$ pour tout $t \ge 0$, il vient que

$$\begin{aligned}\Delta(G) & \le \mathrm{E}_\Sigma\Big[r(V_0(\tilde{S})\,\mathrm{tr}\{\tilde{S}\})\, \mathrm{E}_\mathcal{H}\big[\big(A(\tilde{S})\, V^{k-2} - B(\tilde{S})\, V^{k-1}\big) \mid \tilde{S}\big]\Big], \\ & \le 0. \end{aligned}$$

□

Pour l'estimateur de type Efron-Morris introduit dans (3.6.2),

$$\hat{\Sigma}_{EM}^{-1} = a\,S^{+} + c\,\frac{S\,S^{+}\,Q_0}{\mathrm{tr}\{S\}},$$

avec $Q_0 = H_1\,Q_0^{*}\,H_1^{\top}$ et $c \in \mathbb{R}$, le corollaire suivant est immédiat.

Corollaire 3.6.2

Dans le cadre du théorème 3.6.2, l'estimateur $\hat{\Sigma}_{a,c,r}^{-1} = a\,S^{+} + c\,r\big(\mathrm{tr}\{S\}\big)\,S\,S^{+}\,Q_0/\mathrm{tr}\{S\}$ *améliore l'estimateur* $\hat{\Sigma}_{a}^{-1} = a\,S^{+}$ *si la condition (3.6.5) est vérifiée, c'est à dire si*

$$0 \le c \le 2\,\frac{\mu_{k-1}}{\mu_{k-2}}\{(n \vee p) + 2\,(n \wedge p)^2 - 3\,(n \wedge p) + 2\,k - 3\} - 2\,a,$$

et si, pour tout $t \ge 0$, $0 \le r(t) \le 1$ *et* $r'(t) \ge 0$.

Exemples

Nous allons donner deux exemples de modèle de mélange de lois de Wishart dans (1.4.1) en considérant les lois gamma et gamma inverse comme loi de mélange $\mathcal{H}$. Rappelons que, si V suit une loi gamma de paramètre de forme $\lambda > 2$ et de paramètre d'échelle θ, sa densité a pour forme

$$v \mapsto \frac{\theta^{-\lambda}}{\Gamma(\lambda)} e^{-v/\theta} v^{\lambda-1}, \tag{3.6.32}$$

et, pour $k = 0, 1, ...$, son moment d'ordre k est

$$\mu_k = \mathrm{E}[V^k] = \frac{\Gamma(\lambda + k)}{\Gamma(\lambda)}\,\theta^k.$$

Alors la condition de domination de l'estimateur $\hat{\Sigma}_{a,c,r}^{-1} = a\,S^{+} + c\,r\big(\mathrm{tr}\{S\}\big)\,S\,S^{+}\,Q_0/\mathrm{tr}\{S\}$ sur l'estimateur $\hat{\Sigma}_{a}^{-1} = a\,S^{+}$ donnée dans le corollaire (3.6.2) est

$$0 \le c \le 2\,\theta\,(\lambda + k - 2)\,\{(n \vee p) + 2\,(n \wedge p)^2 - 3\,(n \wedge p) + 2\,k - 3\} - 2\,a.$$

Remarquons que, pour que la borne supérieure de c reste positive, il est nécessaire que

$$0 < a \le \theta\,(\lambda + k - 2)\,\{(n \vee p) + 2\,(n \wedge p)^2 - 3\,(n \wedge p) + 2\,k - 3\}$$

et donc qu'il n'y a amélioration que pour un intervalle de valeurs de a.

Maintenant, si V suit une loi gamma inverse de paramètre de forme $\lambda > k-1$ et de paramètre d'échelle θ sa densité a pour forme

$$v \mapsto \frac{\theta^{-\lambda}}{\Gamma(\lambda)} e^{-\gamma/\theta\,v} \left(\frac{1}{v}\right)^{\lambda+1}, \tag{3.6.33}$$

et, pour $k = 0, 1, ...$, le moment d'ordre k est

$$\mu_k = \mathrm{E}[V^k] = \frac{\Gamma(\lambda - k)}{\Gamma(\lambda)}\,\theta^{-k}.$$

Alors la condition de domination de l'estimateur $\hat{\Sigma}_{a,c,r}^{-1} = a\,S^{+} + c\,r\big(\mathrm{tr}\{S\}\big)\,S\,S^{+}\,Q_0/\mathrm{tr}\{S\}$ sur l'estimateur $\hat{\Sigma}_{a}^{-1} = a\,S^{+}$ donnée dans le corollaire (3.6.2) est la suivante

$$0 \le c \le \frac{2}{\theta\,(\lambda - k + 1)}\,\{(n \vee p) + 2\,(n \wedge p)^2 - 3\,(n \wedge p) + 2\,k - 3\} - 2\,a,$$

Remarquons que, pour que la borne supérieure de c reste positive, il est nécessaire que

$$0 < a \le \frac{1}{\theta\,(\lambda - k + 1)}\,\{(n \vee p) + 2\,(n \wedge p)^2 - 3\,(n \wedge p) + 2\,k - 3\}.$$

et donc qu'il n'y a amélioration que pour un intervalle de valeur de a.

3.7 Conclusion

Dans ce chapitre, nous avons considéré des estimateurs alternatifs à $a\,S^+$ de la forme $a\,S^+ + c\,S\,G(S)$ pour la matrice de précision Σ^{-1} pour le modèle de mélange de lois de Wishart (1.4.1). Sous une large classe de coûts (1.4.3) et grâce à une identité de type Stein-Haff (2.5.3), nous avons développé des conditions sur c et G pour que les estimateurs de type Haff et Dey, qui sont des estimateurs orthogonalement invariants, améliorent l'estimateurs usuels $a\,S^+$. Nous avons aussi donné des conditions de domination de l'estimateur de type Efron-Morris qui n'appartient pas nécessairement à la classe des estimateurs orthogonalement invariants. Notons que nos résultats sont valides dans les deux cas où la matrice de covariance empirique S est inversible ou singulière.

CHAPITRE 4

ÉTUDE NUMÉRIQUE

Dans ce chapitre, nous poursuivons notre travail par une application où nous discutons les résultats de domination basés sur des scénarios de simulations, ces derniers reposent sur certaines lois fixées pour deux types d'estimateurs.

Pour évaluer la performance des estimateurs de type Haff et Dey que nous avons considéré dans la section 3.4, nous effectuons des simulations de Monte-Carlo en générant 10000 répétitions indépendantes à partir du modèle de mélange (1.4.1) où la matrice de covariance Σ est sous la forme $\Sigma = \mathrm{diag}\big(b^{p-i+1}, 1 \leq i \leq p\big)$ avec $0 < b < 3$. Nous considérons deux lois différentes de mélange : $\mathcal{H}$ qui est la loi de Dirac au point 1 et donc S suit la loi de Wishart $\mathcal{W}_p(n, \Sigma)$; ensuite, $\mathcal{H}$ est la loi gamma inverse $\mathcal{GI}(9, 1/2)$. Dans ce qui suit, nous développons un rappel des définitions des lois gamma et gamma-inverse, suivant le livre de Barra [Bar81].

Loi gamma

Une variable aléatoire positive V suit une loi gamma de paramètres $\alpha > 0$ et $\beta > 0$, et on note $V \sim \Gamma(\alpha, \beta)$, si sa densité de probabilité par rapport à la mesure de Lebesgue sur $\mathbb{R}_+^*$ a pour expression

$$v \mapsto \frac{\beta^{-\alpha}}{\Gamma(\alpha)} e^{-v/\beta} v^{\alpha-1}, \tag{4.0.1}$$

où $\Gamma(\alpha) = \int_0^\infty v^{\alpha-1}\, e^{-v}\, dv$ est la fonction de gamma de Euler. Ici α est le paramètre de forme et β est le paramètre d'échelle. La fonction caractéristique de cette densité est

$$t \mapsto (1 - it\beta)^{-\alpha} \tag{4.0.2}$$

et son moment d'ordre k est égal à :

$$\mu_k = \beta^k \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)} = \beta^k (\alpha + k - 1)...\alpha. \tag{4.0.3}$$

Si V suit une loi Gamma $\Gamma(\alpha, \beta)$ alors la variable $1/V$ suit une loi dite Gamma Inverse $\mathcal{GI}(\alpha, 1/\beta)$ dont on précise la définition dans ce qui suit.

Loi de gamma inverse

Une variable aléatoire V suit une loi gamma inverse de paramètres $\alpha > 0$ et $\beta > 0$, et on note $V \sim \mathcal{GI}(\alpha, \beta)$, si sa densité de probabilité de V par rapport à la mesure de Lebesgue sur $\mathbb{R}_*^+$ a pour expression :

$$v \mapsto \frac{\beta^{-\alpha}}{\Gamma(\alpha)} e^{-\gamma/\beta\, v} \left(\frac{1}{v}\right)^{\alpha+1}, \tag{4.0.4}$$

où $\Gamma(\alpha) = \int_0^\infty v^{\alpha-1}\, e^{-v}\, dv$ est la fonction de gamma de Euler. Ici α est le paramètre de forme et β est l'inverse du paramètre d'échelle. Le moment d'ordre k de cette loi est donné par la formule suivante :

$$\mu_k = \gamma^k \frac{\Gamma(\alpha - k)}{\Gamma(\alpha)} = \frac{\gamma^k}{(\alpha-1)...(\alpha-k)}. \tag{4.0.5}$$

La deuxième loi de mélange $\mathcal{H}$ que nous considérons dans les simulations est la loi gamma inverse de paramètre de forme $\alpha = 9$ et de paramètre d'échelle $\beta = 2$.

$$v \mapsto \frac{1}{2^9\, \Gamma(9)} \frac{1}{v^{10}} \exp\left(-\frac{1}{2\, v}\right),$$

pour laquelle nous avons besoin des moments $\mu_{-1} = 18$ et $\mu_{-2} = 360$. Pour les deux cas $p \leq n$ et $p > n$, nous présentons dans les figures 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8 et dans les tableaux 4.1, 4.4 le pourcentage d'amélioration (PRIAL [1]) de l'estimateur alternatif $\hat{\Sigma}_{a,c}^{-1}$ (de type Dey ou Haff) par rapport à l'estimateur de référence $\hat{\Sigma}_a^{-1}$ pour $a = a^*, a_0^-, a_0^+$ (valeurs données dans la proposition 3.3.1). Le PRIAL est défini par

$$PRIAL = 100 \times \frac{R_k(\Sigma^{-1}, \hat{\Sigma}_a^{-1}) - R_k(\Sigma^{-1}, \hat{\Sigma}_{a,c}^{-1})}{R_k(\Sigma^{-1}, \hat{\Sigma}_a^{-1})}.$$

Quant aux figures, nous allons tracer les PRIAL uniquement dans les deux cas où $k = 0$ et $k = 1$, car nous n'avons pas eu d'amélioration importante pour les autres valeurs de k ; les tableaux ci-dessous l'illustrent.

1. Percentage Reduction In Average Loss.

Effet de la constante b sur les PRIAL pour $\Sigma = \mathrm{diag}\big(b^{p-i+1}, 1 \leq i \leq p\big)$.

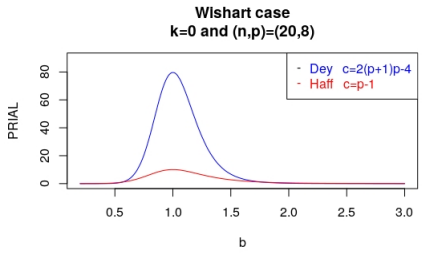

FIGURE 4.1 – PRIAL des estimateurs de type Haff et Dey pour $a = a_0^- = n - 2\,p - 2$.

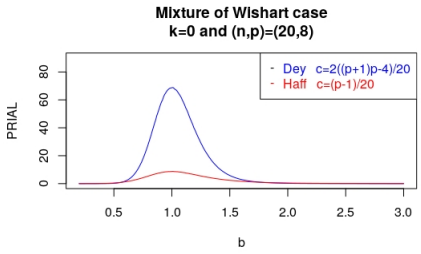

FIGURE 4.2 – PRIAL des estimateurs de type Haff et Dey pour $a = a_0^- = (n - 2\,p - 2)\mu_{-1}/\mu_{-2}$.

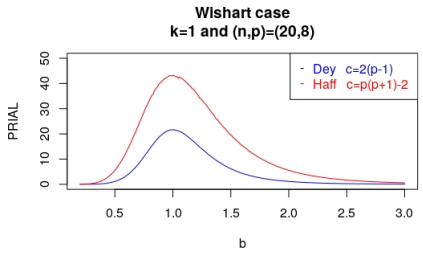

FIGURE 4.3 – PRIAL des estimateurs de type Haff et Dey pour $a = a^* = n - p - 1$.

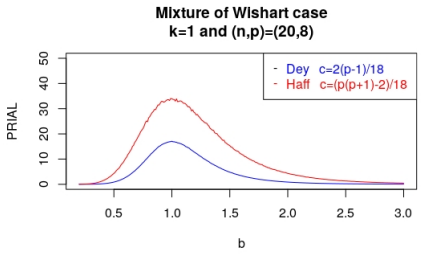

FIGURE 4.4 – PRIAL des estimateurs de type Haff et Dey pour $a = a^* = (n - p - 1)/\mu_{-1}$.

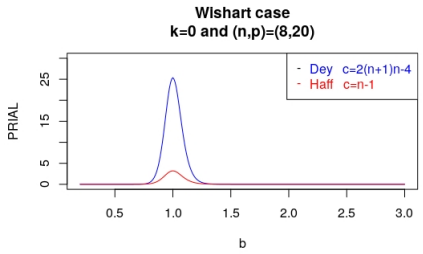

FIGURE 4.5 – PRIAL des estimateurs de type Haff et Dey pour $a = a_0^- = p - 2\,n - 2$.

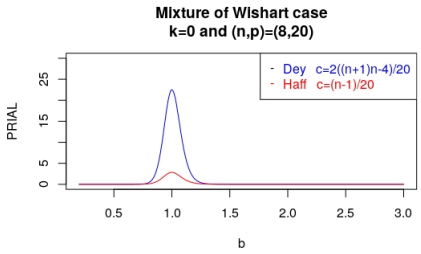

FIGURE 4.6 – PRIAL des estimateurs de type Haff et Dey pour $a = a_0^- = (p - 2\,n - 2)\,\mu_{-1}/\mu_{-2}$.

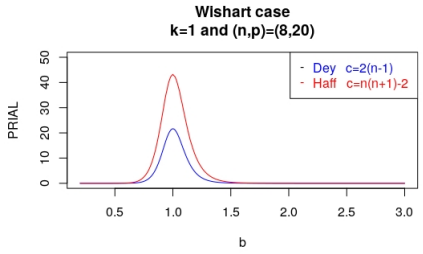

FIGURE 4.7 – PRIAL des estimateurs de type Haff et Dey pour $a = a^* = p - \ n - 1$.

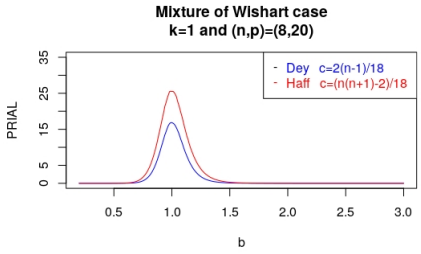

FIGURE 4.8 – PRIAL des estimateurs de type Haff et Dey pour $a = a^* = (p - \ n - 1)/\mu_{-1}$.

Dans tous les cas, les PRIAL sont d'autant plus élevés que b est proche de 1 et ils sont faibles lorsque les éléments diagonaux de Σ sont dispersés. Ceci est vérifié dans les deux cas loi

de Wishart et mélange de lois de Wishart. Pour cette raison, nous considérons dans les tableaux suivants le cas où $b = 1$, c'est à dire où $\Sigma = I_p$. Dans le tableau 4.1 (respectivement le tableau 4.2), nous fournissons les PRIAL de l'estimateur de type Dey (respectivement de type Haff) quand $S \sim \mathcal{W}_p(n,\, I_p)$, tandis que les tableaux 4.3 et 4.4 correspondent respectivement aux estimateurs de types Dey et Haff, lorsque S suit le modèle (1.4.1) avec $\mathcal{H} = \mathcal{IG}(9, 1/2)$. Outre les cas $k = 0$ et $k = 1$, nous considérons d'autres valeurs de k pour lesquels nous allons faire des commentaires ci-dessous qui justifieront le fait que nous n'avons pas présenté les graphiques correspondants ci-dessus.

		$n = 20$ $p = 45$	$n = 45$ $p = 20$	$n = 50$ $p = 120$	$n = 120$ $p = 50$
k=0	a_0^-	21.97	57.16	19.02	63.60
	a_0^+	8.71	22.98	6.46	19.22
k=1	a^*	11.07	11.02	4.50	4.50
k=2	a^*	6.85	6.83	2.74	2.74
k=3	a^*	0.45	0.45	0.06	0.06
k=6	a_0^-	2.75	2.75	0.42	0.42
k=7	a_0^-	3.74	3.74	0.57	0.57

Tableau 4.1 – PRIAL de l'estimateur de type Dey quand $S \sim \mathcal{W}_p(n,\, I_p)$.

		$n = 20$ $p = 45$	$n = 45$ $p = 20$	$n = 50$ $p = 120$	$n = 120$ $p = 50$
k=0	a_0^-	1.80	4.69	0.60	2.00
k=1	a^*	46.47	46.66	42.39	42.45
k=2	a^*	8.95	4.05	3.82	1.60
k=3	a^*	0.23	0.23	0.03	0.03
k=6	a_0^-	1.63	1.63	0.25	0.25
k=7	a_0^-	2.21	2.22	0.33	0.33

Tableau 4.2 – PRIAL de l'estimateur de type Haff quand $S \sim \mathcal{W}_p(n,\, I_p)$.

		$n = 20$ $p = 45$	$n = 45$ $p = 20$	$n = 50$ $p = 120$	$n = 120$ $p = 50$
k=0	a_0^-	19.57	50.43	16.87	71.57
	a_0^+	7.67	19.76	5.76	16.36
k=1	a^*	8.78	8.56	3.47	3.47
k=2	a^*	4.86	4.79	1.87	1.94
k=3	a^*	0.25	0.25	0.03	0.03
k=6	a_0^-	1.38	1.25	0.20	0.16
k=7	a_0^-	0.20	0.28	0.27	0.28

Tableau 4.3 – PRIAL de l'estimateur de type Dey quand $S \sim \mathcal{W}_p(n, V\, I_p)$ avec $V \sim \mathcal{IG}(9, 1/2)$.

		$n = 20$ $p = 45$	$n = 45$ $p = 20$	$n = 50$ $p = 120$	$n = 120$ $p = 50$
k=0	a_0^-	1.62	4.18	0.53	1.66
k=1	a^*	33.79	36.50	33.32	31.09
k=2	a^*	2.61	1.04	2.61	1.04
k=3	a^*	0.01	0.01	0.01	0.01
k=6	a_0^-	0.41	0.59	0.11	0.11
k=7	a_0^-	0.78	0.88	0.17	0.16

Tableau 4.4 – PRIAL de l'estimateur de type Haff quand $S \sim \mathcal{W}_p(n, V\, I_p)$ avec $V \sim \mathcal{IG}(9, 1/2)$.

D'après ces tableaux, nous pouvons voir que les PRIAL les plus grands sont obtenus lorsque $k = 0$ pour l'estimateur de type Dey et lorsque $k = 1$ pour l'estimateur de type Haff. Pour les autres valeurs de k (c'est-à-dire $k = 3, 6$ et 7), les PRIAL sont très faibles, tandis que pour $k = 2$ ils ne sont que légèrement meilleurs.

CHAPITRE 5

SUR L'HOMOGÉNÉITÉ D'ORDRE -2 DES ESTIMATEURS AMÉLIORÉS

5.1 Introduction

Remarquons que les estimateurs améliorés dans le contexte de la loi de Wishart dans la littérature, qui sont homogènes d'un certains degré, tels les estimateurs de type Haff et Dey, sont, de fait, homogènes de degré -2. Ceci semble ainsi être le cas le plus important. Il est imaginable que des estimateurs d'autre degré d'homogénéité soient des estimateurs dominants, mais, bien que nous n'ayons pas de démonstration formelle de ce fait, de tels estimateurs sont probablement moins intéressants et moins utiles. Ceci, dans un certain sens, est analogue au cas de l'estimation d'un vecteur multidimensionnel de moyenne où pratiquement tous les estimateurs améliorés dans la littérature, qui présentent une homogénéité, ont le même degré d'homogénéité que l'estimateur de James-Stein. C'est ce dernier phénomène que nous étudions, dans ce chapitre, en considérant des estimateurs de type James-Stein où le degré d'homogénéité du terme correcteur est différent de -2. Plus précisément, soit X un vecteur aléatoire qui suit la loi normale $p-$dimensionnelle $\mathcal{N}_p(\theta, I)$, de vecteur moyenne inconnu θ et de matrice de covariance l'identité I. Nous estimons θ sous le coût quadratique classique

$$L(\theta, d) = \|d - \theta\|^2 \tag{5.1.1}$$

et son risque associé

$$R(\theta, \delta(X)) = E_\theta(\|\delta(X) - \theta\|^2)\,, \tag{5.1.2}$$

où E_θ désigne l'espérance par rapport à la loi $\mathcal{N}_p(\theta, I)$.

Nous considérons les estimateurs de la forme

$$\delta_{a,b}(X) = \big(1 - a/\|X\|^{2b}\big)X, \tag{5.1.3}$$

où $a > 0$ et $b > 0$. Nous allons monter dans la section 5.2 que, si $b \geq 1$ et si p est suffisamment grand dépendant de b ($p \geq p(b)$), de tels estimateurs sont minimax et améliorent l'estimateur sans biais $\delta_0(X) = X$ pour certains valeurs de a ($0 < a < a(p, b)$). Par exemple, lorsque $b = 2$, si $p > 6$ et $0 < a < 2\,(p-4)\,(p-6)$, l'estimateur (5.1.3) domine X. De même, la domination est valable pour $b = 3$, lorsque $p > 10$ et $0 < a < 2\,(p-6)\,(p-8)\,(p-10)$. En revanche, nous montrerons également qu'une telle amélioration ne tient pas pour $b < 1$.

James et Stein [JS61] ont montré que pour $b = 1$, $p \geq 3$ et $0 < a < 2\,(p-2)$, l'estimateur (5.1.3) domine X, et donc est minimax. De plus, $a = p - 2$ est le choix optimal dans la mesure où il minimise le risque de l'estimateur de la forme (5.1.3). La méthode typique pour prouver ce résultat est de construire un estimateur sans biais de la différence des risques entre $\delta_{a,b}(X)$ et X (voir Stein [Ste81]), on montre ensuite que cet estimateur est toujours négatif et donc que la différence des risques est négative.

Plus précisément, et d'une manière générale, si l'on considère des estimateurs de la forme $\delta(X) = X + g(X)$, le résultat de Stein donne (à condition que $g(\cdot)$ soit faiblement différentiable et que les risques soient finis) :

$$R(\theta, X + g(X)) - R(\theta, X) = E_\theta[\|g(X)\|^2 + 2\,\mathrm{div}g(X)], \tag{5.1.4}$$

où $\mathrm{div}g(X)$ est la divergence faible de $g(X)$.

Ainsi, puisque son espérance donne la différence des risques, un estimateur sans biais de cette différence est $\|g(X)\|^2 + 2\,\mathrm{div}g(X)$. Il s'avère que cette estimation de la différence des risques n'est pas toujours négative (elle change de signe) pour l'estimateur (5.1.3) lorsque $b \neq 1$ et donc que la preuve classique doit être modifiée. Un exemple où la domination du risque peut être établie

lorsque l'estimateur sans biais de la différence des risques change de signe est donné dans Shao et Strawderman [SS94] où une classe d'estimateurs dominant la partie positive de l'estimateur de James-Stein est proposée. À notre connaissance, les estimateurs de la forme (5.1.3) n'ont pas été étudiés dans la littérature pour $b \neq 1$. Ils font l'objet de l'article Boukehil, Fourdrinier, Mezoued et Srawderman[BFMS21c].

5.2 Résultats de domination

Nous donnons dans le lemme suivant, l'estimateur sans biais de la différence des risques entre $\delta_{a,b}(X)$ un estimateur de la forme (5.1.3) et X.

Lemme 5.2.1

Supposons que $\max(2\,b-1, b) < p/2$. *Un estimateur sans biais de la différence des risques entre les estimateurs* $\delta_{a,b}(X)$ *dans (5.1.3) et* X *est*

$$a^2/Y^{2b-1} - 2\,a\,(p-2\,b)/Y^b \tag{5.2.1}$$

où $Y = \|X\|^2$, *au sens où cette différence des risques s'écrit*

$$E_\theta[a^2/Y^{2b-1} - 2\,a\,(p-2\,b)/Y^b] \tag{5.2.2}$$

La preuve découle d'une application directe de l'estimateur sans biais de la différence des risques donnée dans (5.1.4) et du fait que $Y = \|X\|^2$ suive une loi du chi-deux non centrée à p degrés de liberté et de paramètre de décentralisation $\|\theta\|^2/2$. L'inégalité impliquant b soit $\max(2\,b-1, b) < p/2$, garantit que la différence des risques existe et qu'elle est finie.

Notons que si $b = 1$ alors $2\,b - 1 = b$. Dans ce cas, le terme intégrant dans (5.2.2) est égal à $\big(a^2 - 2\,a\,(p-2)\big)/Y$ et, puisque $Y > 0$, l'estimateur sans biais de la différence des risques, et donc la différence des risques elle-même, est négative si $0 < a < 2\,(p-2)$ et $p > 2$. Il s'agit du célèbre résultat de James et Stein [JS61].

En revanche, dans tous les autres cas ($b \neq 1$), l'estimateur sans biais de la différence des risques, change de signe et la preuve habituelle échoue. Néanmoins, (5.2.2) permet une discussion sur la domination telle que donnée dans le résultat principal ci-dessous.

À cette fin, les deux lemmes techniques suivants sont utiles. Une version du résultat suivant est utilisée pour prouver l'admissibilité des tests pour les familles exponentielles avec des régions d'acceptation convexes.

Lemme 5.2.2

Soit Y une variable aléatoire réelle de densité p_θ de la forme $p_\theta(y) = c(\theta)\,\exp(\theta y)$ par rapport à une mesure σ-finie μ, d'espace de paramètres naturel $]c_0, \infty[$ et telle que $\mu(]A, \infty[) > 0$ pour tout $A > 0$. Alors

a. $\lim_{\theta\to\infty} P_\theta[Y > A] = 1$ *pour tout* $A > 0$.

et

b. $\lim_{\theta\to\infty} E_\theta[g(Y)] = \infty$ *pour toute fonction positive croissante* $g(\cdot)$ *telle que* $\lim_{y\to\infty} g(y) = \infty$.

Preuve

Soient δ un point intérieur fixe de l'espace naturel des paramètres $]c_0, \infty[$, $\epsilon > 0$ et $\theta > \delta$. Alors

$$\begin{aligned}
\frac{P_\theta[Y \le A]}{P_\theta[Y > A]} &= \frac{\int_{-\infty}^{A} \exp(\theta y)\,\mu(dy)}{\int_{A}^{\infty} \exp(\theta y)\,\mu(dy)} \\
&= \frac{\int_{-\infty}^{A} \exp\big((\theta-\delta)\,y\big)\exp(\delta y)\,\mu(dy)}{\int_{A}^{\infty} \exp((\theta-\delta)\,y) exp(\delta y)\,\mu(dy)} \\
&\le \frac{\int_{-\infty}^{A} \exp\big((\theta-\delta)\,y\big)\exp(\delta y)\,\mu(dy)}{\int_{A+\epsilon}^{\infty} \exp((\theta-\delta)\,y) exp(\delta y)\,\mu(dy)} \\
&\le \frac{\exp\big((\theta-\delta)\,A)\big)}{\exp((\theta-\delta)(A+\epsilon))}\,\frac{\int_{-\infty}^{A} \exp(\delta y)\,\mu(dy)}{\int_{A+\epsilon}^{\infty} exp(\delta y)\,\mu(dy)} \\
&= \exp\big((\delta-\theta)\,\epsilon)\big)\,\frac{P_\delta[Y \le A]}{P_\delta[Y > A+\epsilon]}\,.
\end{aligned}$$

Cette dernière expression tend vers 0 lorsque θ tend vers l'infini pour chaque ϵ et A fixés. Cela donne lieu à la partie a du lemme puisque $P_\theta[Y \le A] + P_\theta[Y > A] = 1$.

Par propriétés de positivité et de monotonie de la fonction $g(\cdot)$, nous avons $E_\theta[g(Y)] \ge g(A)P_\theta[Y > A]$. Cela donne la partie b puisque, grâce à $\lim_{y\to\infty} g(y) = \infty$, ce dernier terme peut être rendu arbitrairement grand, soit supérieur à M, en choisissant A suffisamment grand, soit $g(A) > 2\,M$, et en choisissant ensuite θ suffisamment grand pour que $P_\theta[Y > A] > 1/2$ (grâce à la partie a). □

Lemme 5.2.3

Soit Y une variable aléatoire de loi du chi-deux non centrée à p degrés de liberté et de paramètre de décentralisation λ. Alors la fonction $h(\lambda) = E_\lambda[1/Y^b]/E_\lambda[1/Y^c]$ est finie, strictement croissante (resp. décroissante) et tend vers ∞(respectivement 0) quand λ tend vers ∞ lorsque $p/2 > c > b$ (resp. $p/2 > b > c$).

Preuve

Supposons que $p/2 > c > b$. Notons que, grâce à la représentation de Poisson d'une loi non centrée du chi-deux, à p degrés de liberté, en termes de lois du chi-deux centrée, on a

$$h(\lambda) = \frac{E_\lambda^K[E^X[1/X_{p+2K}^b|K]]}{E_\lambda^K[E^X[1/1/X_{p+2K}^c|K]]},$$

où $X_{p+2K}|K \sim \chi^2_{p+2K}$ et $K \sim Poisson(\lambda)$. Posant $A(K,a) = E^X\left[1/X_{p+2K}^a|K\right]$, la fonction $h(\lambda)$ peut être exprimée comme suit

$$h(\lambda) = \frac{E_\lambda[A(K,b)]}{E_\lambda[A(K,c)]} = E_\lambda^*\left[\frac{A(K,b)}{A(K,c)}\right],$$

où E_λ^* représente l'espérance par rapport à la famille exponentielle discrète des densités

$$p_\lambda^*(k) \propto (\lambda^k/k!)\,\frac{A(k,c)}{E_\lambda^K[A(K,c)]}\,.$$

Par hypothèse sur c et b, la fonction $A(K,b)/A(K,c)$ est croissante en K de sorte que, par propriété de vraisemblance monotone croissante en λ de la loi non centrée du chi-deux de Y, $h(\lambda)$ est strictement croissante en λ. De plus, puisque $A(K,b)/A(K,c)$ a pour limite ∞ lorsque K tend vers ∞ quand $c > b$, la première partie du lemme découle du lemme 5.2.2. Les inégalités $p/2 > c > b$ assurent la finitude de toutes les espérances. La preuve pour le cas $p/2 > b > c$ est similaire. □

Théorème 5.2.1

Soit $Y = \|X\|^2$. On dispose les deux résultats suivants.

a. Pour $b < 1$ aucun estimateur de la forme (5.1.3) ne domine X pour toute dimension p.

b. Pour $b > 1$ et $p > 2\,(2\,b - 1)$, l'estimateur (5.1.3) domine X et donc est minimax pour

$$0 < a < 2\,(p - 2b)\,\frac{E_0[1/Y^b]}{E_0[1/Y^{2b-1}]}\,.$$

Preuve

Il découle du lemme 5.2.1 que la différence des risques est négative si toutes les espérances sont finies et si $a^2 - 2\,a\,(p - 2\,b)\,E_\lambda[1/Y^b]/E_\lambda[1/Y^{2b-1}] < 0$ ou, de manière équivalente, si $a^2 - 2\,a\,(p - 2\,b)\,E^*_\lambda[Y^{b-1}] < 0$ pour tout $\lambda \geq 0$ $\big($en utilisant la notation de la preuve du lemme (5.2.3)$\big)$.

Pour la partie a, notons que, puisque X est admissible quand $p = 1$ ou 2, nous avons $p - 2\,b > 0$ pour tout $p > 2$. Ceci implique que $a > 0$ pour tout estimateur amélioré. Cependant, d'après le lemme 5.2.3, puisque $E^*_\lambda[Y^{b-1}]$ se rapproche de 0 lorsque λ se rapproche de ∞ (puisque $b - 1 < 0$) aucune valeur positive de a ne satisfait à la condition ci-dessus. Ceci prouve la partie a.

Pour prouver la partie b, l'hypothèse que $b > 1$ implique que $2\,b - 1 > b$, de sorte que toutes les espérances impliquées dans la différence des risques sont finies lorsque $p > 2\,(2\,b - 1)$. En outre, la différence des risques est négative pour tous les λ à condition que $0 < a < 2\,(p - 2\,b)\,E_0[1/Y^b]/E_0[1/Y^{2b-1}]$ puisque la fonction $E_\lambda[1/Y^b]/E_\lambda[1/Y^{2b-1}]$ est croissante en λ d'après le lemme 5.2.3. Ceci prouve la partie b. □

Notons que, puisque les moments d'ordre négatif d'une loi du chi-deux centrée ont une forme simple, soit $E\left[\,1/\big(\chi^2_p\big)^b\right] = 1/[(p-2)(p-4)\ldots(p-2\,b)]$, les conditions de domination des estimateurs minimax données par le théorème 5.2.1 peuvent être exprimées sous la forme des inéquations suivantes :

— pour $b = 2$, les conditions sont $p > 6$ et

$$0 < a < 2\,(p-4)\,\frac{E[1/Y^2]}{E[1/Y^3]} = 2\,(p-4)\,\frac{(p-2)\,(p-4)\,(p-6)}{(p-2)\,(p-4)} = 2\,(p-4)\,(p-6)\,;$$

— pour $b = 3$, elles sont $p > 10$ et

$$0 < a < 2\,(p-4)\,\frac{E[1/Y^3]}{E[1/Y^5]} = 2\,(p-6)\,(p-8)\,(p-10)\,;$$

— de même, pour $b = 4$, elles sont $p > 14$ et

$$0 < a < 2\,(p-8)\,(p-10)\,(p-12)\,(p-14)\,.$$

5.3 Extensions

Dans cette section, nous considérons quelques extensions du résultat principal. L'extension à $X \sim \mathcal{N}_p(\theta, \sigma^2 I)$ est relativement directe. L'estimation sans biais de Stein de la différence de risque pour un estimateur de θ de la forme $X + \sigma^2\, g(X)$ est $\sigma^4\big(\|g(X)\|^2 + 2\mathrm{div} g(X)\big)$. Une version du résultat principal pour ce cas lorsque σ est connu est donnée dans le théorème suivant.

Théorème 5.3.1

Soient $X \sim N_p(\theta, \sigma^2 I)$ avec σ connu et $\delta_{a,b}(X) = \big(\,1 - a\,\sigma^2/\|X\|^{2b}\big)X$. Sous la fonction de coût donnée par (5.1.1), nous avons les deux résultats suivants.

a. Pour $b < 1$ aucun estimateur de la forme ci-dessus ne domine X pour toute dimension p.

b. Pour $b > 1$ et $p > 2\,(2\,b - 1)$, l'estimateur ci-dessus domine X et est minimax pour

$$0 < a < 2\,(p - 2\,b)\,(\sigma^2)^{b-1}\,\frac{E_0[1/Y^b]}{E_0[1/Y^{2b-1}]}\,,$$

où E_θ désigne l'espérance par rapport à $X \sim \mathcal{N}_p(\theta, \sigma^2 I)$ et où Y suit une loi du chi-deux centrée à p degrés de liberté.

Notons qu'une forme équivalente, et un peu plus naturelle, de l'estimateur amélioré donné par la partie b est $(1 - a\,(\sigma^2)^b/\|X\|^{2b})\,X$ pour $0 < a < 2\,(p - 2\,b)\,E_0[1/Y^b]/E_0[1/Y^{2b-1}]$.

Preuve

La preuve suit de près celle du théorème 5.2.1. La preuve de la partie a est essentiellement inchangée.

La preuve de la partie b utilise la forme ci-dessus de l'estimation sans biais de la différence de risque, ce qui donne un estimateur amélioré si toutes les espérances sont finies et si, pour tous les θ,

$$a^2 - 2\,a\,(p - 2\,b)\,E_\theta[1/(\|X\|^2)^b]/E_\theta[1/(\|X\|^2)^{2b-1}] < 0\,.$$

Puisque $Y = \|X\|^2/\sigma^2$ suit une loi du chi-deux non centrée à p degrés de liberté et de paramètre de

décentralisation $\lambda = \|\theta\|^2/2\sigma^2$, cette condition devient (avec un changement évident de notation)

$$0 < a < 2\,(p-2b)\,(\sigma^2)^{b-1} E_\lambda[1/Y^b]/E_\lambda[1/Y^{2b-1}]$$

pour tout λ. D'après le lemme 5.2.3, le rapport des espérances est croissant en λ, et la condition est donc valide si

$$0 < a < 2\,(p-2b)\,(\sigma^2)^{b-1} E_0[1/Y^b]/E_0[1/Y^{2b-1}]\,,$$

ce qui est la condition donnée pour a. La finitude du risque découle des conditions sur b. □

Comme extension des estimateurs de type James-Stein, on considère des estimateurs la forme de Baranchik (voir [Bar70]), soit

$$\delta_{a,b,r}(X) = \big(1 - a\,\sigma^2\, r\big(\|X\|^2\big)/\big(\|X\|^2\big)^b\big)X.$$

où r est une fonction de $\mathbb{R}_+$ dans $\mathbb{R}_+$.

Théorème 5.3.2

Soit $X \sim \mathcal{N}_p(\theta, \sigma^2 I)$ avec σ connu. Soit $\delta_{a,b,r}(X) = \big(1 - a\,\sigma^2\, r\big(\|X\|^2\big)/\big(\|X\|^2\big)^b\big)X$ où $r(\cdot)$ est une fonction absolument continue de $\mathbb{R}_+$ dans $\mathbb{R}_+$. Pour $b > 1$, $p > 2\,(2\,b-1)$ et

$$0 < a < 2\,(p-2\,b)\,(\sigma^2)^{b-1}\,\frac{E_0[1/Y^b]}{E_0[1/Y^{2b-1}]}\,,$$

où $Y = \|X\|^2$ suit une loi du chi-deux centrée à p degrés de liberté, l'estimateur $\delta_{a,b,r}(X)$ domine X et est minimax sous la fonction de coût donnée par (5.1.1) dès que $0 \le r(t) \le 1$ et $r'(t) \ge 0$.

Preuve

Une application directe de l'estimation sans biais de la différence des risques de Stein donne

$$\begin{aligned}
\Delta(\theta) &= R\big(\theta, \delta_{a,b,r}(X)\big) - R(\theta, X) \\
&= a\,\sigma^4\, E_\theta\left[a\,\frac{r^2(\|X\|^2)}{\|X\|^{2(2b-1)}} - 2\,(p-2\,b)\,\frac{r(\|X\|^2)}{\|X\|^{2b}} - 4\,\frac{r'(\|X\|^2)}{\|X\|^{2(b-1)}}\right] \\
&\le a\,\sigma^4\, E_\theta\left[r(\|X\|^2)\left\{\frac{a}{\|X\|^{2(2b-1)}} - \frac{2\,(p-2\,b)}{\|X\|^{2b}}\right\}\right],
\end{aligned}$$

où la dernière inégalité vient du au fait que $r'(t) \ge 0$ et $r^2(t) \le r(t)$ par les conditions sur $r(\cdot)$.

Puisque la fonction $r(\cdot)$ est croissante et que le terme entre crochets change de signe une seule fois en passant du positif au négatif en le point $\|X\|_0^2 = \big(a/2\,(p-2b)\big)^{1/(b-1)}$, il s'ensuit que

$$\Delta(\theta) \leq a\,\sigma^4\,r\big(\|X\|_0^2\big)\,E_\theta\left[\frac{a}{\|X\|^{2(2b-1)}} - \frac{2\,(p-2\,b)}{\|X\|^{2b}}\right] \leq 0\,,$$

par la partie b du théorème 5.3.1. □

Le cas $X \sim \mathcal{N}_p(\theta, \sigma^2 I)$ où le paramètre d'échelle σ est inconnu peut également être considéré. Ci-dessous $E_{\theta,\sigma}$ désignera l'espérance par rapport à la loi $\mathcal{N}_p(\theta, \sigma^2 I)$.

Corollaire 5.3.1

Soit $X \sim \mathcal{N}_p(\theta, \sigma^2 I)$ avec σ inconnu et soit $S/\sigma^2 \sim \chi_k^2$ indépendant de X. Soit $\delta_{a,b}(X,S) = \Big(1 - a\,\big(S/\|X\|^2\big)^b\Big)\,X$ et soit la fonction de coût soit donnée par (5.1.1). Alors, pour $b > 1$ et $p > 2\,(2\,b-1)$, l'estimateur ci-dessus domine X pour

$$0 < a < 2\,(p-2\,b)\,C_{b,k}\,\frac{E_0[1/Y^b]}{E_0[1/Y^{2b-1}]}\,,$$

où Y suit une loi du chi-deux centrée à p degrés de liberté et où $C_{b,k} = E\Big[\big(\chi_k^2\big)^b\Big]\Big/E\Big[\big(\chi_k^2\big)^{2b}\Big]$.

Preuve

En écrivant l'estimateur sous la forme $\delta_{a,b}(X,S) = \Big(1 - a\,\sigma^2\,(S/\sigma^2)\,S^{b-1}\big(1/\|X\|^2\big)^b\Big)\,X$, et en utilisant le résultat de Stein, la différence des risques peut être exprimée comme suit :

$$\begin{aligned}
\Delta(\theta,\sigma) &= R\big(\theta, \delta_{a,b}(X)\big) - R(\theta, X) \\
&= \sigma^4\,E_{\theta,\sigma}\left[a^2\,\frac{S^{2b}/\sigma^4}{\|X\|^{2(2b-1)}} - 2\,(p-2\,b)\,a\,\frac{S^b/\sigma^2}{\|X\|^{2b}}\right] \\
&= a\,\sigma^2\,E_{\theta,\sigma}\left[a\,\frac{\big(S/\sigma^2\big)^{2b}}{\big(\|X\|^2/\sigma^2\big)^{2b-1}} - 2\,(p-2\,b)\,\frac{\big(S/\sigma^2\big)^b}{\big(\|X\|^2/\sigma^2\big)^b}\right] \\
&= a\,\sigma^2\,E\Big[\big(\chi_k^2\big)^{2b}\Big]\,E_{\theta,\sigma}\left[\frac{a}{\big(\|X\|^2/\sigma^2\big)^{2b-1}} - C_{b,k}\,\frac{2\,(p-2\,b)}{\big(\|X\|^2/\sigma^2\big)^b}\right],
\end{aligned}$$

puisque $S/\sigma^2 \sim \chi_k^2$ est indépendant de $\|X\|^2/\sigma^2 \sim \chi_p^2(\|\theta\|^2/2\sigma^2)$. Il s'ensuit, en utilisant l'argument du théorème 5.2.1, que cette différence des risques est négative si les conditions du théorème sont satisfaites. □

Ce résultat peut également être étendu aux estimateurs de type Baranchik.

Corollaire 5.3.2

Soit $X \sim \mathcal{N}_p(\theta, \sigma^2 I)$ avec σ inconnu et soit $S/\sigma^2 \sim \chi^2_k$ indépendant de X. Soit la fonction de coût soit donnée par (5.1.1). Nous avons les deux résultats suivants.

a. Soit $\delta_{a,b,r}(X,S) = (1 - ar(\|X\|^2)(S/\|X\|^2)^b)X$, où $r(\cdot)$ satisfait les conditions du théorème 5.3.2 et où p, b, et a satisfont les conditions de la partie b du théorème 5.3.1. Alors $\delta_{a,b,r}(X,S)$ domine X.

b. De même, sous les mêmes conditions, $\delta^*_{a,b,r}(X,S) = (1 - ar(\|X\|^2/S)(S/\|X\|^2)^b)X$ domine X.

Preuve

En suivant la preuve du Corollaire 5.3.1, et en utilisant les méthodes du Théorème 5.3.2 pour éliminer le terme impliquant $r'(\|X\|^2)$, un majorant de la différence des risques est donné par

$$\begin{aligned}\Delta(\theta,\sigma) &= R\big(\theta, \delta_{a,b,r}(X)\big) - R(\theta, X) \\ &\leq a\,\sigma^2\, E\Big[\big(\chi^2_k\big)^{2b}\Big]\, E_{\theta,\sigma}\left[r(\|X\|^2)\left\{\frac{a}{\big(\|X\|^2/\sigma^2\big)^{2b-1}} - C_{b,k}\,\frac{2\,(p-2\,b)}{\big(\|X\|^2/\sigma^2\big)^b}\right\}\right],\end{aligned}$$

avec $C_{b,k}$ donné dans le Corollaire 5.3.1. Par conséquent, cette différence des risques est négative si

$$0 < a < 2\,(p-2\,b)\,C_{b,k}\,E_{\theta,\sigma}\left[\frac{r(\|X\|^2)}{\big(\|X\|^2/\sigma^2\big)^b}\right] \Big/ E_{\theta,\sigma}\left[\frac{r(\|X\|^2)}{\big(\|X\|^2/\sigma^2\big)^{2b-1}}\right].$$

Mais, posant $Y = \|X\|^2/\sigma^2$, nous pouvons exprimer et minorer le dernier rapport des espérances comme suit :

$$E_{\theta,\sigma}[r(\sigma^2 Y)/Y^b]/E_{\theta,\sigma}[r(\sigma^2 Y)/Y^{2b-1}] \geq E_{\theta,\sigma}[1/Y^b)]/E_{\theta,\sigma}[1/Y^{2b-1}] \geq E_{0,1}[1/Y^b)]/E_{0,1}[1/Y^{2b-1}]\,,$$

en utilisant un argument de rapport de vraisemblance monotone. Ceci prouve la partie a.

De même, pour prouver la partie b, on peut écrire

$$\begin{aligned}\Delta(\theta,\sigma) &= R\big(\theta, \delta^*_{a,b,r}(X)\big) - R(\theta, X) \\ &\leq a\,\sigma^2\, E_{\theta,\sigma}\left[r(Y/S)\left\{\frac{a}{(Y/S)^{2b-1}} - \frac{2\,(p-2\,b)}{(Y/S)^b}\right\}\right].\end{aligned}$$

Par conséquent, la différence de risque est négative si

$$0 < a < E_{\theta,\sigma}[r(Y/S)/(Y/S)^b]/E_{\theta,\sigma}[r(Y/S)/(Y/S)^{2b-1}]\,.$$

Encore une fois, un argument de rapport de vraisemblance monotone montre que

$$\begin{aligned} E_{\theta,\sigma}[r(Y/S)/(Y/S)^b]/E_{\theta,\sigma}[r(Y/S)/(Y/S)^{2b-1}] &\geq E_{\theta,\sigma}[1/(Y/S)^b]/E_{\theta,\sigma}[1/(Y/S)^{2b-1}] \\ &\geq C_{b,k}E_0[1/Y^b]/E_0[1/Y^{2b-1}]\,, \end{aligned}$$

d'après le lemme 5.2.3 puisque $b < 2\,b - 1 < p/2$. Cela donne la partie b. □

5.4 Une propriété d'optimalité uniforme de l'estimateur de James-Stein

Le résultat suivant étend une propriété d'optimalité bien connue de l'estimateur de James-Stein à la grande classe d'estimateurs de la forme (5.1.3) avec $b \geq 1$.

Théorème 5.4.1

Soit $X \sim \mathcal{N}_p(\theta, I)$ et soit la fonction de coût donnée par (5.1.1). Alors, pour tout $p \geq 3$, l'estimateur usuel de James-Stein $\delta_{p-2,1}(X) = \big(\,1 - (p-2)/\|X\|^2\big)X$ est le meilleur estimateur uniformément dans la classe de tous les estimateurs de la forme (5.1.3) avec $b \geq 1$.

Preuve

Nous supposons dans les calculs ci-dessous que, pour chaque $p \geq 3$ pour lequel nous comparons les risques, le risque de $\delta_{a,b}(X)$ est fini (c'est-à-dire que $\max(2\,b-1, b) < p/2$). Sinon, l'estimateur de James-Stein est clairement meilleur. D'après les calculs du lemme 5.2.1, la différence de risque (5.1.4) entre les estimateurs $\delta_{a,b}(X)$ et $\delta_{p-2,1}(X)$ est égale à

$$R(\theta, \delta_{a,b}(X)) - R(\theta, \delta_{p-2,1}(X)) = E_\theta\ \left[a^2/Y^{2b-1} - 2\,a\,(p-2\,b)/Y^b\right] + (p-2)^2\,E_\theta\ [1/Y]\ ,$$

où $Y = \|X\|^2$ et où E_θ est l'espérance par rapport à $\chi^2_p\big(\|\theta\|^2/2\big)$. Nous montrons que cette expression est positive pour tout $b > 1$ et pour tout a. Pour $b = 1$, le résultat classique donne

$a = p - 2$ comme valeur optimale. Pour tout $1 < b < 2\,b - 1 < p/2$ l'expression ci-dessus est finie et la valeur minimale de l'expression quadratique en a pour tout θ fixé est réalisée pour $a = (p - 2\,b)\,E[1/Y^b]/E[1/Y^{2b-1}]$ et prend la valeur

$$A = -(p - 2\,b)^2\,\frac{\big(E_\theta[1/Y^b]\big)^2}{E_\theta[1/Y^{2b-1}]} + (p-2)^2\,E_\theta[1/Y]\,.$$

Pour compléter la preuve, il suffit de montrer que $A > 0$, ce qui est équivalent à

$$\frac{E_\theta[1/Y^{2b-1}]\,E_\theta[1/Y]}{\big(E_\theta[1/Y^b]\big)^2} > \frac{(p-2\,b)^2}{(p-2)^2}.$$

Or le membre gauche de l'inégalité ci-dessus est supérieur ou égal à 1 puisque, par inégalité de covariance,

$$\frac{E_\theta[1/Y^{2b-1}]}{E_\theta[1/Y^b]}\,\frac{E_\theta[1/Y]}{E_\theta[1/Y^b]} = E^*_\theta[1/Y^{b-1}]\,E^*_\theta[Y^{b-1}] \geq 1\,,$$

où, $f_\theta(y)$ étant la densité de $\chi^2_p\big(\|\theta\|^2/2\big)$, l'espérance E^*_θ est prise par rapport à la densité proportionnelle à $f_\theta(y)\,y^{-b}$. Par conséquent, comme $(p-2\,b)^2/(p-2)^2/ < 1$ si $b > 1$, la preuve est achevée. □

Il est facile d'étendre ce résultat au cas $X \sim \mathcal{N}(\theta, \sigma^2 I)$. Il est intéressant de noter que le résultat n'étend pas l'optimalité à toutes les valeurs de b, et en fait un tel résultat n'est pas vrai. En particulier si $b = 0$, les estimateurs sont de la forme $(1-a)\,X$ qui sont des estimateurs de Bayes propre et ils sont admissibles pour tout p si $0 \leq a < 1$, et donc, ne peuvent pas être dominés par l'estimateur inadmissible de James-Stein. Cependant, comme montré ci-dessus, de tels estimateurs ne peuvent pas être minimax lorsque $b < 1$ (à moins, bien sûr, que $a = 0$).

CHAPITRE 6

CONCLUSION

Dans cette thèse nous nous sommes intéressés au problème d'estimation de la matrice de précision Σ^{-1} de dimension $p \times p$ du modèle de mélange de lois de Wishart

$$\begin{cases} S \mid V \sim \mathcal{W}_p(n, V\,\Sigma) \\ V \sim \mathcal{H}(\cdot) \end{cases}$$

où $\mathcal{W}_p(n, V\,\Sigma)$ représente la loi de Wishart à n degrés de liberté et de matrice de variance-covariance $V\,\Sigma$ et où $\mathcal{H}(\cdot)$ est une loi de mélange définie sur $\mathbb{R}^+$. Pour les estimateurs usuels de la forme

$$\Sigma_a^{-1} = a\,S^+,$$

où a est une constante positive et S^+ est l'inverse de Moore-Penrose de la matrice de covariance empirique S, nous avons donné sous le coût

$$L_k\{\Sigma^{-1}, \hat{\Sigma}^{-1}\} = \mathrm{tr}[\{\hat{\Sigma}^{-1} - \Sigma^{-1}\}^2\,S^k],$$

des valeurs optimales a^* de a pour $k = 1, 2$ et 3 et des valeurs "raisonnables" a_0 pour lesquelles les estimateurs $\Sigma_{a_0}^{-1}$ vérifient certaines propriétés importantes pour $k = 0$ et $k \geq 4$. Nous avons montré que ce type d'estimateurs n'est pas admissible, en proposant des estimateurs alternatifs dominant $a\,S^+$ de la forme générale

$$\hat{\Sigma}_{a,c}^{-1} = a\,S^+ + c\,S\,G(S),$$

où a et c sont des constantes positives, S^+ est l'inverse de Moore-Penrose de S et $G(S)$ est une fonction matricielle de dimension $p \times p$ homogène de degré α. Grâce à l'identité de type Stein-Haff qui a été établie par Fourdrinier, Haddouche et Mezoued [FHM21], nous avons fourni, dans une approche unifiée des deux cas où la matrice S est inversible ($p \leq n$) et singulière ($n < p$), un estimateur sans biais de la différence de risque entre $\hat{\Sigma}_a^{-1}$ et $\hat{\Sigma}_{a,c}^{-1}$ à travers les moments de la loi de mélange $\mathcal{H}$, soit

$$\begin{aligned} \delta(G) = c^2\,\mu_{k+2\alpha+2}\,\mathrm{tr}\{\tilde{S}^k[\tilde{S}\,G(\tilde{S})]^2\} \\ + 2\,c\left[a\,\mu_{k+\alpha} - \mu_{k+\alpha+1}(n - (n \wedge p) - 1)\right]\mathrm{tr}\{\tilde{S}^+\,\tilde{S}^{k+1}\,G(\tilde{S})\} \\ - 4\,c\,\mu_{k+\alpha+1}\,\mathrm{tr}\{\tilde{S}^+\,\tilde{S}\,D_{\tilde{S}}\,\{G^\top(\tilde{S})\,\tilde{S}^{k+1}\}\}, \end{aligned}$$

où $\tilde{S} = S/V$ et où μ_β est le moment d'ordre β de la loi $\mathcal{H}$. En explicitant des conditions de négativité de cet estimateur, nous avons donné des résultats de domination sur l'estimateur $a\,S^+$, où a correspond à une constante optimale ou raisonnable, des estimateurs orthogonalement invariants de type Haff et Dey et d'un estimateur de type Efron-Morris qui n'appartient pas nécessairement à la classe des estimateurs orthogonalement invariant.

Nous avons construit une classe d'estimateurs plus grande en modifiant le terme correcteur de l'estimateur $\hat{\Sigma}^{-1}_{a,c}$, ces estimateurs ont la forme

$$\hat{\Sigma}^{-1}_{a,c,r} = a\,S^+ + c\;r\big(\mathrm{tr}\{S\}\big)\,S\,G(S)\,,$$

où c est une constante positive, r est une fonction réelle et $G(S)$ est une fonction homogène d'ordre $\alpha = -2$ qui peut être ou non orthogonalement invariante. Nous avons donné des conditions sur c, r et G pour que cet estimateur améliore l'estimateur $a\,S^+$. Finalement, pour illustrer les résultats théoriques obtenus dans ce travail, nous avons effectué des simulations numériques pour deux choix de loi de mélange $\mathcal{H}$: une loi de Dirac et une loi gamma inverse.

Dans tous nos exemples d'estimateurs amélioré, c'est l'homogénéité de degré -2 qui intervient. Aucun estimateur où ce degré d'homogénéité diffère de -2 n'a pu être exhibé. Le fait que -2 soit le degré d'homogénéité adéquat a été établi pour le problème de l'estimation de la moyenne d'une loi normale multidimensionnelle.

Dans notre travail la matrice de covariance Σ est considérée inversible. Récemment, Chételat et Wells [CW16] se sont intéressés à l'estimation de la matrice de précision d'un modèle normal multivarié dont la matrice de covariance Σ est singulière. Ils ont utilisé la fonction de coût qui coïncide avec L_0, soit

$$L\{\Sigma^+, \hat{\Sigma}^+\} = \mathrm{tr}[\{\hat{\Sigma}^+ - \Sigma^+\}^2],$$

où Σ^+ est l'inverse de Morre-Penrose de la matrice Σ et où $\hat{\Sigma}^+$ estime Σ^+. Nous envisageons d'élargir le travail de Chételat et Wells en considérant notre modèle de mélange de lois de Wishart au travers des autres fonctions de coût $(L_1, L_2, ...)$.

Rappelons aussi que, dans notre travail, la variable de mélange V est une variable aléatoire réelle positive. Haff, Kim, Koo et Richards [HKKR11] ont présenté un modèle de mélange de lois de Wishart où la variable V de mélange est une matrice définie positive. Il serait intéressant de prolonger directement cette thèse dans un tel contexte.

Appendices

ANNEXE A

LEMMES IMPORTANTS

Dans cette partie, nous rappelons les lemmes A.1, A.2 et A.3 qui ont été énoncés et/ou établis par Fourdrinier, Haddouche et Mezoued [FHM21] et nous donnons d'autres lemmes que nous avons utilisés dans ce travail.

Lemme A.1

Soient A, B et C des fonctions matricielles d'une matrice S symétrique de dimension $p \times p$. Supposant que toutes les dérivées partielles existent, on a

$$\mathrm{tr}\{A\,(C\,D_S)^\top\,B^\top\} = \mathrm{tr}\{C\,(A\,D_S)^\top\,B\}. \tag{A.1}$$

où D_S est l'opérateur différentiel de Haff défini dans (2.5.4).

Lemme A.2

Pour toute fonction matricielle diagonale de L, $\Phi(L) = \mathrm{diag}(\phi_i(L))_{1\le i\le n\wedge p}$, nous avons

$$(H_1\,H_1^\top\,D_S)^\top\{H_1\,\Phi(L)\,H_1^\top\} = D_S\,\{H_1\,\Phi(L)\,H_1^\top\} - \frac{1}{2}\big(p-(n\wedge p)\big)\,H_1\,L^{-1}\,\Phi(L)\,H_1^\top. \tag{A.2}$$

Lemme A.3

Soit $G = H_1\Psi H_1^{\mathrm{T}}$ où $\Psi = \Psi(L) = \mathrm{diag}\big(\psi_1(L),\dots,\psi_{n\wedge p}(L)\big)$ est une matrice diagonale de dimension $(n\wedge p)\times(n\wedge p)$. Alors $SS^+\,G = G$ et

$$\mathcal{D}_s\{SS^+\,G\} = \mathcal{D}_s\{H_1\Psi H_1^{\mathrm{T}}\} = H_1\Psi^{(1)}H_1^{\mathrm{T}} + \frac{1}{2}\,\mathrm{tr}\big(L^{-1}\Psi\big)\,\big(I_p - H_1H_1^{\mathrm{T}}\big) \tag{A.3}$$

où $\Psi^{(1)}$ est une matrice $(n\wedge p)\times(n\wedge p)$ diagonale telle que

$$\psi_i^{(1)} = \frac{1}{2}(p-(p\wedge n))\frac{\psi_i}{l_i} + \frac{\partial\psi_i}{\partial l_i} + \frac{1}{2}\sum_{j\neq i}^{p\wedge n}\frac{\psi_i-\psi_j}{l_i-l_j}\,.$$

Nous avons établi les lemmes suivants.

Lemme A.4

Pour tout $k \in \mathbb{R}$ et $m \in \mathbb{R}$, on a

$$\mathrm{tr}\{L^{k+m}\} \leq \mathrm{tr}\{L^k\}\,\mathrm{tr}\{L^m\}. \tag{A.4}$$

si $k\,m \geq 0$, nous avons

$$\mathrm{tr}\{L^k\}\,\mathrm{tr}\{L^m\} \leq (n \wedge p)\,\mathrm{tr}\{L^{k+m}\}. \tag{A.5}$$

Preuve

on obtient le premier résultat immédiatement grâce à l'inégalité suivante

$$\sum_{i=1}^{p\wedge n} l_i^{k+m} = \sum_{i=1}^{p\wedge n} l_i^k\, l_i^m \leq \sum_{i=1}^{p\wedge n} l_i^k \sum_{j=1}^{p\wedge n} l_j^m\,.$$

□

Le deuxième résultat vient de l'inégalité de covariance.

Lemme A.5

Soit A une matrice de taille $p \times p$ admettant une racine carré c'est-à-dire telle que l'on peut écrire $A = A^{1/2}\,A^{1/2}$ et L soit une matrice diagonale comme dans (1.3.12). Alors

$$\mathrm{tr}\{L^{k-1}\,A\} \geq \frac{\mathrm{tr}\{L^k\,A\}}{\mathrm{tr}\{L\}}. \tag{A.6}$$

Preuve

Pour tout vecteur x de dimension $(n \wedge p) \times 1$, on a

$$\begin{aligned} x^\top\,A^{\frac{1}{2}}\,L^{k-1}\,A^{\frac{1}{2}}\,x = (A^{\frac{1}{2}}\,x)^\top L^{k-1}(\,A^{\frac{1}{2}}\,x) = \sum_{j=1}^{n\wedge p} \frac{l_i^k}{l_i}\,(\,A^{\frac{1}{2}}\,x)_i^2 &\geq \sum_{j=1}^{n\wedge p} \frac{l_i^k}{\sum_{j=1}^{n\wedge p} l_i}\,(\,A^{\frac{1}{2}}\,x)_i^2 \\ &= x^\top \frac{A^{\frac{1}{2}}\,L^k\,A^{\frac{1}{2}}}{\mathrm{tr}\{L\}} x. \end{aligned}$$

Ceci donne l'inégalité matricielle suivante

$$A^{\frac{1}{2}}\,L^{k-1}\,A^{\frac{1}{2}} \geq \frac{A^{\frac{1}{2}}\,L^k\,A^{\frac{1}{2}}}{\mathrm{tr}\{L\}}.$$

En prenant la trace des deux membres de l'inégalité précédente, on obtient

$$\mathrm{tr}\{L^{k-1}\,A\} \geq \frac{\mathrm{tr}\{L^{k+}\,A\}}{\mathrm{tr}\{L\}}.$$

□

Lemme A.6

Soit A est une matrice semi-définie positive de dimension $p \times p$ et soit Q_0^* une matrice de projection de dimension $p \times p$ comme dans (3.6.1). Alors

$$\mathrm{tr}\{A\} \geq \mathrm{tr}\{Q_0^*\,A\} \geq 0.$$

Preuve

Nous avons

$$\mathrm{tr}\{A\} = \mathrm{tr}\{Q_0^* \, A\} + \mathrm{tr}\{[I_p - Q_0^*] \, A\}. \tag{A.7}$$

Soit $A = R \, D \, R^\top$ la décomposition en valeur singulière de A telle que $D = \mathrm{diag}(d_1, d_2, ..., d_p)$ où $d_i \geq 0$ sont les valeurs singulières de A pour $1 \leq i \leq p$ et où R est la matrice des vecteurs propres associés. Soit $Q_0^* = Q_1 \, Q_1^\top$ où Q_1 est une matrice de dimension $p \times q$ où q est le nombre des valeurs propres qui valent 1 ; cette matrice est de la forme $Q_1 = Q \, (Q^\top \, Q)^{-1/2}$.

La démonstration de lemme A.6 consiste à prouver que les termes de la partie droite de (A.7) sont tous les deux positifs. D'abord, on a

$$\begin{aligned} \mathrm{tr}\{Q_0^* \, A\} &= \mathrm{tr}\{Q_1 \, Q_1^\top \, R \, D \, R^\top\} = \mathrm{tr}\{Q_1^\top \, R \, D^{1/2} \, D^{1/2} \, R^\top \, Q_1\} \\ &= \mathrm{tr}\{(D^{1/2} \, R^\top \, Q_1)^\top \, D^{1/2} \, R^\top \, Q_1\} = \| \, D^{1/2} \, R^\top \, Q_1 \, \|_F^2, \end{aligned} \tag{A.8}$$

qui est est positif car il s'agit du carré de la norme de Frobenius. Aussi, comme $I_p - Q_0^* = I_p - Q_1 \, Q_1^\top = Q_2 \, Q_2^\top$ est un projecteur, alors

$$\mathrm{tr}\{[I_p - Q_0^*] \, A\} = \mathrm{tr}\{Q_2 \, Q_2^\top \, R \, D \, R^\top\} = \| \, D^{1/2} \, R^\top \, Q_2 \, \|_F^2, \tag{A.9}$$

donc est positif. Alors, grâce à (A.8) et (A.9), nous avons

$$\mathrm{tr}\{A\} \geq \mathrm{tr}\{Q_0^* \, A\} \geq 0.$$

□

Lemme A.7

Pour tout $k \in \mathbb{N}$, soit

$$B_k = \sum_{i=1}^{n \wedge p} \sum_{j \neq i}^{n \wedge p} \frac{l_i^{k-1} - l_j^{k-1}}{l_i - l_j} \, .$$

Alors, pour $k = 0$,

$$B_0 = \mathrm{tr}\{L^{-2}\} - \left(\mathrm{tr}\{L^{-1}\}\right)^2$$

et

$$[1 - (n \wedge p)] \, \mathrm{tr}\{L^{-2}\} \leq B_0 \leq 0 \, ;$$

pour $k = 1$,

$$B_1 = 0 \, ;$$

pour $k = 2$,

$$B_2 = (n \wedge p) \, [(n \wedge p) - 1] \, ;$$

pour $k = 3$,

$$B_3 = 2 \, [(n \wedge p) - 1] \, \mathrm{tr}\{L\} \, ;$$

pour $k \geq 4$,

$$B_k = \sum_{m=1}^{k-3} \mathrm{tr}\{L^m\} \, \mathrm{tr}\{L^{k-2-m}\} + [2 \, (n \wedge p) - (k-1)] \, \mathrm{tr}\{L^{k-2}\}$$

et

$$[2 \, (n \wedge p) - (k-1)] \, \mathrm{tr}\{L^{k-2}\} \leq B_k \leq (k-1) \, [(n \wedge p) - 1] \, \mathrm{tr}\{L^{k-2}\} \, .$$

Preuve

Pour les égalités, un simple calcul donne

$$B_0 = \sum_{i=1}^{n\wedge p}\sum_{j\neq i}^{n\wedge p} \frac{l_i^{-1} - l_j^{-1}}{l_i - l_j} = -\sum_{i=1}^{n\wedge p}\sum_{j\neq i}^{n\wedge p} \frac{1}{l_i\, l_j} = -\sum_{i=1}^{n\wedge p}\left(\sum_{j=1}^{n\wedge p}\frac{1}{l_i\, l_j} - \frac{1}{l_i^2}\right)$$
$$= -\sum_{i=1}^{n\wedge p}\left(\frac{1}{l_i}\right)\sum_{j=1}^{n\wedge p}\left(\frac{1}{l_j}\right) + \sum_{i=1}^{n\wedge p}\left(\frac{1}{l_i^2}\right) = -\big(\mathrm{tr}\{L^{-1}\}\big)^2 + \mathrm{tr}\{L^{-2}\}.$$

De la même manière, il vient que

$$B_1 = \sum_{i=1}^{n\wedge p}\sum_{j\neq i}^{n\wedge p} \frac{1-1}{l_i - l_j} = 0\,,$$

$$B_2 = \sum_{i=1}^{n\wedge p}\sum_{j\neq i}^{n\wedge p} \frac{l_i - l_j}{l_i - l_j} = \sum_{i=1}^{n\wedge p}\sum_{j\neq i}^{n\wedge p} 1 = (n\wedge p)\,[(n\wedge p) - 1],$$

et

$$B_3 = \sum_{i=1}^{n\wedge p}\sum_{j\neq i}^{n\wedge p} \frac{l_i^2 - l_j^2}{l_i - l_j} = \sum_{i=1}^{n\wedge p}\sum_{j\neq i}^{n\wedge p}(l_i + l_j) = [(n\wedge p) - 1]\sum_{i=1}^{n\wedge p} l_i + \sum_{i=1}^{n\wedge p}\left(\sum_{j=1}^{n\wedge p} l_j - l_i\right)$$
$$= [(n\wedge p) - 1]\,\mathrm{tr}\{L\} + (n\wedge p)\,\mathrm{tr}\{L\} - \mathrm{tr}\{L\} = 2\,[(n\wedge p) - 1]\,\mathrm{tr}\{L\}\,.$$

Finalement, pour $k \geq 4$, nous avons

$$\begin{aligned}
B_k &= \sum_{i=1}^{n\wedge p}\sum_{j\neq i}^{n\wedge p} \frac{l_i^{k-1} - l_j^{k-1}}{l_i - l_j} = \sum_{i=1}^{n\wedge p}\sum_{j\neq i}^{n\wedge p}\sum_{m=0}^{k-2} l_i^m\, l_j^{k-2-m} \\
&= \sum_{m=1}^{k-3}\sum_{i=1}^{n\wedge p}\sum_{j\neq i}^{n\wedge p} l_i^m\, l_j^{k-2-m} + \sum_{i=1}^{n\wedge p}\sum_{j\neq i}^{n\wedge p}\left(l_j^{k-2} + l_i^{k-2}\right) \\
&= \sum_{m=1}^{k-3}\sum_{i=1}^{n\wedge p} l_i^m\left(\sum_{j=1}^{n\wedge p} l_j^{k-2-m} - l_i^{k-2-m}\right) + \sum_{i=1}^{n\wedge p}\left(\sum_{j=1}^{n\wedge p} l_j^{k-2} - l_i^{k-2}\right) + [(n\wedge p) - 1]\sum_{i=1}^{n\wedge p} l_i^{k-2} \\
&= \sum_{m=1}^{k-3}\Big(\mathrm{tr}\{L^m\}\,\mathrm{tr}\{L^{k-2-m}\} - \mathrm{tr}\{L^{k-2}\}\Big) + 2\,[(n\wedge p) - 1]\,\mathrm{tr}\{L^{k-2}\} \\
&= \sum_{m=1}^{k-3}\mathrm{tr}\{L^m\}\,\mathrm{tr}\{L^{k-2-m}\} + [2\,(n\wedge p) - (k-1)]\,\mathrm{tr}\{L^{k-2}\}\,.
\end{aligned}$$

Quant aux inégalités, appliquant l'inégalité (A.5) du lemme A.4 avec $m = k = -1$, on obtient

$$\big(\mathrm{tr}\{L^{-1}\}\big)^2 \leq (n\wedge p)\,\mathrm{tr}\{L^{-2}\}\,,$$

ce qui donne la première inégalité vérifiée par B_0. La deuxième inégalité est obtenue en utilisant le lemme A.4 avec $k = m = -1$.

Maintenant, pour m et $k = k - 2 - m$, on obtient

$$\mathrm{tr}\{L^m\}\,\mathrm{tr}\{L^{k-2-m}\} \leq (n\wedge p)\,\mathrm{tr}\{L^{k-2}\}\,,$$

d'où la deuxième inégalité vérifiée par B_k pour $k \geq 4$ (la première inégalité est évidente). □

Lemme A.8

Pour toute fonction matricielle $G(\tilde{S})$ de la forme $G(\tilde{S}) = H_1\,\Psi(L)\,H_1^\top$, on a

$$\begin{aligned}\operatorname{tr}\{S^+\,S\,D_S\{r(\operatorname{tr}\{S\})\,G^\top(S)\,S^{k+1}\}\} &= r(\operatorname{tr}\{S\})\operatorname{tr}\{S^+\,S\,D_S\{G^\top(S)\,S^{k+1}\}\}\\ &\quad + r'(\operatorname{tr}\{S\})\operatorname{tr}\{G^\top(S)\,S^{k+1}\}. \qquad \text{(A.10)}\end{aligned}$$

Preuve

Comme $S\,S^+\,G^\top(S) = G^\top(S)$, il est clair que

$$\begin{aligned}D_S\{r(\operatorname{tr}\{S\})\,G^\top(S)\,S^{k+1}\} &= D_S\{S^+\,S\,r(\operatorname{tr}\{S\})\,G^\top(S)\,S^{k+1}\}\\ &= D_S\{H_1\,r(\operatorname{tr}\{S\})\,\Psi\,L^{k+1}\,H_1^{\mathrm{T}}\}\end{aligned}$$

En utilisant le lemme A.3 avec $r(\operatorname{tr}\{\tilde{S}\})\,\Psi\,L^{k+1}$ à la place de Ψ et

$$\begin{aligned}\psi_i^{(1)} &= \frac{1}{2}\big(p-(p\wedge n)\big)\,r(\operatorname{tr}\{L\})\,\frac{(\psi\,L^{k+1})_i}{l_i} + \frac{\partial\left(r(\operatorname{tr}\{L\})(\psi\,L^{k+1})_i\right)}{\partial l_i}\\ &\qquad + \frac{1}{2}\,r(\operatorname{tr}\{L\})\sum_{j\neq i}^{p\wedge n}\frac{(\psi\,L^{k+1})_i-(\psi\,L^{k+1})_j}{l_i-l_j}\\ &= r(\operatorname{tr}\{L\})\left(\frac{1}{2}\big(p-(p\wedge n)\big)\,\frac{(\psi\,L^{k+1})_i}{l_i} + \frac{\partial(\psi\,L^{k+1})_i}{\partial l_i} + \frac{1}{2}\sum_{j\neq i}^{p\wedge n}\frac{(\psi\,L^{k+1})_i-(\psi\,L^{k+1})_j}{l_i-l_j}\right)\\ &\qquad + r'(\operatorname{tr}\{L\})\,(\psi\,L^{k+1})_i\,,\end{aligned}$$

on obtient

$$D_S\{r(\operatorname{tr}\{S\})\,G^\top(S)\,S^{k+1}\} = H_1\Psi^{(1)}H_1^{\mathrm{T}} + \frac{1}{2}\operatorname{tr}\Big(L^{-1}r(\operatorname{tr}\{S\})\,\Psi\,L^{k+1}\Big)\,(I_p - H_1H_1^{\mathrm{T}}).$$

Or on sait que $S\,S^+\,(I_p - H_1\,H_1^\top) = 0$. Donc

$$S\,S^+\,D_S\{r(\operatorname{tr}\{\tilde{S}\})\,G^\top(S)\,S^{k+1}\} = r(\operatorname{tr}\{\tilde{S}\})\,S\,S^+\,D_S\{G^\top(S)\,S^{k+1}\} + r'(\operatorname{tr}\{\tilde{S}\})\,S\,S^+\,\Psi\,L^{k+1}\,.$$

En prenant la trace des deux membres de cette égalité il vient

$$\begin{aligned}\operatorname{tr}\{S^+\,S\,D_S\{r(\operatorname{tr}\{\tilde{S}\})\,G^\top(S)\,S^{k+1}\}\} &= r(\operatorname{tr}\{\tilde{S}\})\operatorname{tr}\{S^+\,S\,D_S\{G^\top(S)\,S^{k+1}\}\}\\ &\quad + r'(\operatorname{tr}\{\tilde{S}\})\operatorname{tr}\{G^\top(S)\,S^{k+1}\}.\end{aligned}$$

D'où le résultat. □

Lemme A.9

Pour la matrice semi-orthogonale H_1 de dimension $p\times(p\wedge n)$ en (1.3.12) , on a

$$D_S\,\{H_1\} = \frac{1}{2}\,H_1\,\Psi, \qquad \text{(A.11)}$$

où $\Psi = \operatorname{diag}\big(\Psi_1,\ldots,\Psi_{\mathrm{p\wedge n}}\big)$ *et* $\Psi_k = \sum_{r\neq k}\dfrac{1}{l_k-l_r} + \dfrac{p-(n\wedge p)}{l_k}$.

Preuve

En utilisant l'égalité (A.4.6) dans Fourdrinier, Haddouche et Mezoued [FHM21] (qui se trouve dans la première ligne ci-dessous) avec $a = j$, on trouve

$$\begin{aligned}
(D_S\,\{H_1\})_{ik} &= \sum_j D_{S_{ij}}\, H_{1_{jk}} \\
&= \frac{1}{2}\sum_j \sum_{r\neq k} \frac{H_{1_{jr}}}{l_k - l_r}\,[(H_1)_{ir}\,(H_1)_{jk} + (H_1)_{jr}\,(H_1)_{ik}] \\
&\qquad + \frac{1}{l_k}\,[(I_p - H_1\,H_1^\top)_{ji}\,(H_1)_{jk} + (I_p - H_1\,H_1^\top)_{jj}\,(H_1)_{ik}] \\
&= \sum_{r\neq k} \frac{H_{1_{jr}}}{l_k - l_r}\,[\,(H_1)_{ir}\,(H_1^\top\,H_1)_{rk} + (H_1^\top\,H_1)_{rr}\,(H_1)_{ik}] \\
&\qquad + \frac{1}{l_k}\,[(H_1 - H_1\,H_1^\top\,H_1)_{ik}\,(H_1)_{jk} + \mathrm{tr}\{I_p - H_1\,H_1^\top\}\,(H_1)_{ik}] \\
&= (H_1)_{ik}\left(\sum_{r\neq k}\frac{1}{l_k - l_r} + \frac{(p - p\wedge n)}{l_k}\right), \qquad \text{(A.12)}
\end{aligned}$$

car $H_1^\top\,H_1 = I_{n\wedge p}$. Cette égalité est équivalente à celle donnée dans (A.11). □

Rappelons que Q_0^* est une matrice de taille $(n\wedge p)\times(n\wedge p)$ définie par

$$Q_0^* = Q\,(Q^\top\,Q)^{-1}\,Q^\top, \qquad \text{(A.13)}$$

où Q une matrice à éléments constants de dimension $(n\wedge p)\times q$ et de rang $\mathrm{rk}(Q) = q$.

Lemme A.10

Pour toute matrice de projection Q_0^* comme dans (3.6.1), on a

$$D_S\,\{\,H_1\,Q_0^*\,H_1^\top\} = \frac{1}{2}\,H_1\,\Psi\,Q_0^*\,H_1^\top + \{(H_1\,Q_0^*)^\top\,D_S\}^\top\,(H_1\,Q_0^*)^\top. \qquad \text{(A.14)}$$

où $\Psi = \mathrm{diag}\left(\Psi_1, ..., \Psi_{\mathrm{p}\wedge\mathrm{n}}\right)$ *avec* $\Psi_k = \sum_{r\neq k}\dfrac{1}{l_k - l_r} + \dfrac{p-(n\wedge p)}{l_k}$.

Preuve

En utilisant la propriété des projecteurs, soit $Q_0^* = {Q_0^*}^2$, et en appliquant l'identité (i) de la proposition $A.4$ dans Fourdrinier, Haddouche and Mezoued [FHM21], soit $D_S(A\,B) = D_S(A)B + (A^\top\,D_S)^\top\,B$ avec $A = H_1\,Q_0^*$ et $B = (H_1\,Q_0^*)^\top$ et on obtient

$$D_S\,\{\,H_1\,Q_0^*\,H_1^\top\} = D_S\,\{H_1\}\,Q_0^*\,(H_1\,Q_0^*)^\top + \{H_1^\top\,D_S\}^\top(\,Q_0^*\,H_1^\top). \qquad \text{(A.15)}$$

Le deuxième terme de droite dans (A.15) est nul car Q^* ne dépend pas de S. Finalement, pour exprimer $D_S\,\{\,H_1\,Q_0^*\,H_1^\top\}$ on utilise (A.11), on obtient

$$\begin{aligned}
D_S\,\{\,H_1\,Q_0^*\,H_1^\top\} &= \frac{1}{2}\,H_1\,\Psi\,Q_0^*\,(H_1\,Q_0^*)^\top + \{H_1^\top\,D_S\}^\top(\,Q_0^*\,H_1^\top) \\
&= \frac{1}{2}\,H_1\,\Psi\,Q_0^*\,H_1^\top + \{H_1^\top\,D_S\}^\top(\,Q_0^*\,H_1^\top). \qquad \text{(A.16)}
\end{aligned}$$

□

Lemme A.11

Pour tout projecteur Q_0^* comme dans (A.13) et pour $Q_0 = H_1\, Q_0^*\, H_1^\top$, on a

$$\operatorname{tr}\left\{S^+\, S\, D_S\,\{Q_0^\top\}\,\frac{S^+\, S^{k+1}}{\operatorname{tr}\{S\}}\right\} = \frac{1}{2\operatorname{tr}\{S\}}\operatorname{tr}\{S^+\, S^{k+1}\, H_1\, \Psi\, Q_0^*\, H_1^\top\} + \frac{1}{2\operatorname{tr}\{S\}}\sum_m\sum_{o\neq m}\frac{1}{l_m - l_0}\, l_0^k\,(Q_0^*)_{mm}. \quad \text{(A.17)}$$

où $\Psi = \operatorname{diag}\left(\Psi_1, ..., \Psi_{\mathrm{p}\wedge\mathrm{n}}\right)$ *avec* $\Psi_k = \sum_{r\neq k}\dfrac{1}{l_k - l_r} + \dfrac{p-(n\wedge p)}{l_k}$.

Preuve

Pour $Q_0 = H_1\, Q_0^*\, H_1^\top$, on a

$$\begin{aligned}\operatorname{tr}\left\{S^+\, S\, D_S\,\{Q_0^\top\}\,\frac{S^+\, S^{k+1}}{\operatorname{tr}\{S\}}\right\} &= \frac{1}{\operatorname{tr}\{S\}}\operatorname{tr}\left\{S^+\, S\, S^+\, S^{k+1}\, D_S\,\{\, H_1\, Q_0^*\, H_1^\top\}\right\}\\ &= \frac{1}{\operatorname{tr}\{S\}}\operatorname{tr}\left\{S^+\, S^{k+1}\, D_S\,\{\, H_1\, Q_0^*\, H_1^\top\}\right\}. \end{aligned} \quad \text{(A.18)}$$

En remplaçant (A.16) dans (A.18) il est clair que

$$\begin{aligned}\operatorname{tr}\left\{S^+\, S\, D_S\,\{Q_0^\top\}\,\frac{S^+\, S^{k+1}}{\operatorname{tr}\{S\}}\right\} &= \frac{1}{\operatorname{tr}\{S\}}\operatorname{tr}\left\{S^+\, S^{k+1}\left[\frac{1}{2}\, H_1\, \Psi\, Q_0^*\, H_1^\top + \{H_1^\top\, D_S\}^\top(\, Q_0^*\, H_1^\top)\right]\right\}\\ &= \frac{1}{2\operatorname{tr}\{S\}}\operatorname{tr}\{S^+\, S^{k+1}\, H_1\, \Psi\, Q_0^*\, H_1^\top\}\\ &\quad + \frac{1}{\operatorname{tr}\{S\}}\operatorname{tr}\{S^+\, S^{k+1}\,\{H_1^\top\, D_S\}^\top(\, Q_0^*\, H_1^\top)\}. \end{aligned} \quad \text{(A.19)}$$

Maintenant, le deuxième terme de droite dans (A.19) peut être écrit comme suit :

$$\begin{aligned}\operatorname{tr}\{S^+\, S^{k+1}\,\{H_1^\top\, D_S\}^\top\,(Q_0^*\, H_1^\top)\} &= \sum_{j,i}(S^+\, S^{k+1})_{ji}\,[\{H_1^\top\, D_S\}^\top(\, Q_0^*\, H_1^\top)]_{ij}\\ &= \sum_{j,i}(S^+\, S^{k+1})_{ji}\,\sum_k\{H_1^\top\, D_S\}_{ki}\,(Q_0^*\, H_1^\top)_{kj}\\ &= \sum_{j,i}(S^+\, S^{k+1})_{ji}\,\sum_k\sum_r(H_1)_{rk}\,(D_S)_{ri}\,(Q_0^*\, H_1^\top)_{kj}\\ &= \sum_{j,i}(S^+\, S^{k+1})_{ji}\,\sum_k\sum_r(H_1)_{rk}\,(D_S)_{ri}\sum_m(Q_0^*)_{mk}\,(H_1)_{jm}\\ &= \sum_{j,i}(S^+\, S^{k+1})_{ji}\,\sum_k\sum_r(H_1)_{rk}\,\sum_m\{(D_S)_{ri}\;(H_1)_{jm}\}\,(Q_0^*)_{km}, \end{aligned} \quad \text{(A.20)}$$

car Q^* ne dépend pas de S. De plus, grâce à l'égalité (A.4.5) de Fourdrinier, Haddouche et

Mezoued [FHM21] avec $i = r$, $a = j$, $j = i$, $k = m$ et $r = b$, le terme dans (A.20) devient

$$\begin{aligned}
\operatorname{tr}\{S^+ S^{k+1} \{H_1^\top D_S\}^\top (Q_0^* H_1^\top)\} &= \frac{1}{2}\sum_{j,i}(S^+ S^{k+1})_{ji} \sum_{k,r,m} (H_1)_{rk} \\
&\sum_{b\neq m} \frac{(H_1)_{jb}}{l_m - l_b} [(H_1)_{rb} (H_1)_{im} + (H_1)_{ib} (H_1)_{rm}] (Q_0^*)_{km} \\
&+ \frac{1}{2}\sum_{j,i}(S^+ S^{k+1})_{ji} \sum_{k,r,m} \frac{1}{l_m} (H_1)_{rk} \\
&\times [(I_p - H_1 H_1^\top)_{jr} (H_1)_{im} + (I_p - H_1 H_1^\top)_{ji} (H_1)_{rm}] (Q_0^*)_{km} \\
&= \frac{1}{2}\sum_{j,i}(S^+ S^{k+1})_{ji} \sum_{r,m} \sum_{b\neq m} \frac{(H_1)_{jb}}{l_m - l_b} \\
&\times [(H_1)_{rb} (H_1)_{im} + (H_1)_{ib} (H_1)_{rm}] (H_1 Q_0^*)_{rm} \\
&+ \frac{1}{2}\sum_{j,i}(S^+ S^{k+1})_{ji} \sum_{r,m} \frac{1}{l_m} \Big[(I_p - H_1 H_1^\top)_{jr} (H_1)_{im} \\
&(I_p - H_1 H_1^\top)_{ji} (H_1)_{rm}\Big] (H_1 Q_0^*)_{rm} \\
&= \frac{1}{2}\sum_{j,i}(S^+ S^{k+1})_{ji} \sum_{m} \sum_{b\neq m} \frac{(H_1)_{jb}}{l_m - l_b} \\
&\times [(H_1^\top H_1 Q_0^* Q^{*^\top})_{bm} (H_1)_{im} + (H_1)_{ib} (H_1^\top H_1 Q_0^*)_{mm}] \\
&+ \frac{1}{2}\sum_{j,i}(S^+ S^{k+1})_{ji} \sum_{m} \frac{1}{l_m} \Big[(I_p - H_1 H_1^\top)(H_1 Q_0^*)_{jm} (H_1)_{im} \\
&(I_p - H_1 H_1^\top)_{ji} (H_1^\top H_1 Q_0^*)_{mm}\Big] \\
&= \frac{1}{2}\sum_{j,i}(S^+ S^{k+1})_{ji} \sum_{m} \sum_{b\neq m} \frac{(H_1)_{jb}}{l_m - l_b} \\
&\times [(Q_0^*)_{bm} (H_1)_{im} + (H_1)_{ib} (Q_0^*)_{mm}] \\
&+ \frac{1}{2}\sum_{j,i}(S^+ S^{k+1})_{ji} \sum_{m} \frac{1}{l_m} (I_p - H_1 H_1^\top)_{ji} (Q_0^*)_{mm}, \qquad \text{(A.21)}
\end{aligned}$$

car Q_0^* est un projecteur et $H_1^\top H_1 = I_{n\wedge p}$. Par conséquent

$$\begin{aligned}
\operatorname{tr}\{S^+ S^{k+1} \{H_1^\top D_S\}^\top (Q_0^* H_1^\top)\} &= \frac{1}{2}\sum_{j}\sum_{m}\sum_{b\neq m} \frac{(H_1)_{jb}}{l_m - l_b} \\
&\times [(Q_0^*)_{bm} (S^+ S^{k+1} H_1)_{jm} + (S^+ S^{k+1} H_1)_{bj} (Q_0^*)_{mm}] \\
&+ \frac{1}{2}\sum_{j}\sum_{m} \frac{1}{l_m} [(S^+ S^{k+1}) (I_p - H_1 H_1^\top)]_{jj} (Q_0^*)_{mm} \qquad \text{(A.22)}
\end{aligned}$$

Notons que le dernier terme de l'égalité précédente est nul puisque $(S^+ S^{k+1}) (I_p - H_1 H_1^\top) = 0$,

et donc

$$\begin{aligned}
\mathrm{tr}\{S^+\, S^{k+1}\, \{H_1^\top\, D_S\}^\top\, (Q_0^*\, H_1^\top)\} &= \frac{1}{2} \sum_m \sum_{b\neq m} \frac{1}{l_m - l_b} \\
&\quad \times [Q_{bm}^*\, (H_1^\top\, S^+\, S^{k+1}\, H_1)_{bm} + (H_1^\top\, S^+\, S^{k+1}\, H_1)_{bb}\, (Q_0^*)_{mm}] \\
&= \frac{1}{2} \sum_m \sum_{b\neq m} \frac{1}{l_m - l_b} [(Q_0^*)_{bm}\, L_{bm}^k + L_{bb}^k\, (Q_0^*)_{mm}] \\
&= \frac{1}{2} \sum_m \sum_{b\neq m} \frac{1}{l_m - l_b}\, l_b^k\, (Q_0^*)_{mm}, \qquad \text{(A.23)}
\end{aligned}$$

puisque L^k est une matrice diagonale. Finalement, en remplaçant (A.23) dans (A.19) nous trouvons le résultat dans (A.17). □

Lemme A.12

Pour toute fonction matricielle $G(S)$ de dimension $p\times p$ et pour toute fonction réelle $r(\mathrm{tr}\{S\})$, on a

$$\begin{aligned}
\mathrm{tr}\big\{S^+\, S\, D_S\{\, r(V\, \mathrm{tr}\{S\})\, G^\top(\tilde{S})\, S^{K+1}\}\big\} &= r(V\, \mathrm{tr}\{S\})\, \mathrm{tr}\big\{S^+\, S\, D_S\{\, G^\top(\tilde{S})\, S^{K+1}\}\, \big\} \qquad \text{(A.24)} \\
&\quad + G\, r^{'}(V\, \mathrm{tr}\{S\})\, \mathrm{tr}\big\{G(\tilde{S})\, S^{K+1}\big\}.
\end{aligned}$$

Preuve

Comme $r(\mathrm{tr}\{S\})$ est une fonction réelle, on a

$$\mathrm{tr}\big\{S^+\, S\, D_S\{r(V\, \mathrm{tr}\{S\})\, G^\top(\tilde{S})\, S^{K+1}\}\big\} = \mathrm{tr}\big\{S^+\, S\, D_S\{G^\top(\tilde{S})\, S^{K+1}\, r(V\, \mathrm{tr}\{S\})\, \}\big\}. \qquad \text{(A.25)}$$

En utilisant le fait que $D_S\{A\, B\} = D_S\{A\}\, B + \{A^\top D_S\}^\top\, B$ avec $A = G^\top(\tilde{S})\, S^{K+1}$ et $B = r(\mathrm{tr}\{V\, S\})$, le terme dans (A.25) devient

$$\begin{aligned}
\mathrm{tr}\big\{S^+\, S\, D_S\{r(V\, \mathrm{tr}\{S\})\, G^\top(\tilde{S})\, S^{K+1}\}\big\} &= r(V\, \mathrm{tr}\{S\})\, \mathrm{tr}\big\{S^+\, S\, D_S\{\, G^\top(\tilde{S})\, S^{K+1}\}\, \big\} \\
&\quad + \mathrm{tr}\big\{S^+\, S\, \{S^{K+1}\, G(\tilde{S})\, D_S\}^\top\, r(V\, \mathrm{tr}\{S\})\, \big\}
\end{aligned} \qquad \text{(A.26)}$$

En appliquant (A.1) avec $A = S^+\, S$, $C = S^{K+1}\, G(\tilde{S})$ et $B = r(V\, \mathrm{tr}\{S\})$, le dernier terme dans (A.26) peut être réécrit comme suit

$$\begin{aligned}
\mathrm{tr}\big\{S^+\, S\, \{S^{K+1}\, G(\tilde{S})\, D_S\}^\top\, r(V\, \mathrm{tr}\{S\})\, \big\} &= \mathrm{tr}\big\{S^+\, S\, \{S^{K+1}\, G(\tilde{S})\, D_S\}^\top\, r(V\, \mathrm{tr}\{S\})\, \big\} \\
&= \sum_i [S^+\, S\, \{S^{K+1}\, G(\tilde{S})\, D_S\}^\top\, r(V\, \mathrm{tr}\{S\})]_{ii} \\
&= \sum_{i,j} (S^+\, S)_{ij}\, (S^{K+1}\, G(\tilde{S})\, D_S)_{ij}\, r(V\, \mathrm{tr}\{S\}) \\
&= \sum_{i,j,k} (S^+\, S)_{ij}\, (S^{K+1}\, G(\tilde{S}))_{ik}\, (D_S)_{kj}\, r(V\, \mathrm{tr}\{S\}).
\end{aligned} \qquad \text{(A.27)}$$

En remplaçant (2.5.4) dans (A.27), on trouve

$$\begin{aligned}
\mathrm{tr}\big\{S^+\,S\,\{S^{K+1}\;G(\tilde{S})\,D_S\}^\top\,r(V\,\mathrm{tr}\{S\})\,\big\} &= \frac{1}{2}\,\sum_{i,j,k}(S^+\,S)_{ij}\;(S^{K+1}\;G(\tilde{S}))_{ik}\,(1+\delta_{kj})\,\frac{\partial\,r(V\,\mathrm{tr}\{S\})}{\partial S_{kj}} \\
&= \sum_{i,j}(S^+\,S)_{ij}\;(S^{K+1}\;G(\tilde{S}))_{ij}\;\frac{\partial\,r(V\,\mathrm{tr}\{S\})}{\partial S_{jj}} \\
&= V\,\sum_{i,j}(S^+\,S)_{ji}\;(S^{K+1}\,G(\tilde{S}))_{ij}\,r^{'}(V\,\mathrm{tr}\{S\}) \\
&= V\,\mathrm{tr}\{S^{K+1}\,G(\tilde{S})\}\,r^{'}(V\,\mathrm{tr}\{S\}), \qquad \text{(A.28)}
\end{aligned}$$

car $\partial\,r(V\,\mathrm{tr}\{S\})/\partial S_{ij} = 0$ si $i \neq j$, $S^+\,S$ est symétrique et $\partial\,r(V\,\mathrm{tr}\{S\})/\partial S_{jj} = V\,r^{'}(V\,\mathrm{tr}\{S\})$. En remplaçant (A.28) dans (A.26), nous trouvons le résultat donné dans (A.12). □

Bibliographie

[Bar70] A. J. Baranchik. A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.*, 41 :642–645, 1970. 64

[Bar81] J. R. Barra. *Mathematical Basis of Statistics*. Academic Press, 1981. 51

[BB99] M. Bilodeau and D. Brenner. *Theory of Multivariate Statistics*. Institute of mathematical statistics, 1999. 18

[BFMS21a] D. Boukehil, D. Fourdrinier, F. Mezoued, and W. E. Strawderman. Efron-morris type estimators of an inverse scatter matrix. *Technical report, Université de Rouen Normandie and ENSSEA de Tipaza*, 2021. 24

[BFMS21b] D. Boukehil, D. Fourdrinier, F. Mezoued, and W. E. Strawderman. Estimation of the inverse scatter matrix for a scale mixture of wishart matrices under efron-morris type losses. *Journal of Statistical Planning and Inference*, 215 :368–387, 2021. 24

[BFMS21c] D. Boukehil, D. Fourdrinier, F. Mezoued, and W. E. Strawderman. On homogeneous James-Stein type estimators. *Technical report, Université de Rouen Normandie and ENSSEA de Tipaza*, 2021. 59

[CF11] R. D. Cook and L. Forzani. On the mean and variance of the generalized inverse of a singular Wishart matrix. *Electronic Journal of Statistics*, 5 :146–158, 2011. 29

[CW12] D. Chételat and M. T. Wells. Improved multivariate normal mean estimation with unknown covariance when p is greater than n. *Annals of Statistics*, 40 :3137–3160, 2012. 20

[CW16] D. Chételat and M. T. Wells. Improved second order estimation in the singular multivariate normal model. *Journal of Multivariate Analysis*, 147(C), 2016. 70

[Dey87a] D. K. Dey. Improved estimation of a multinormal precision matrix. *Statistics and probability letters*, (6) :125–128, 1987. 4, 31

[Dey87b] D. K. Dey. Improved estimation of a multinormal precision matrix. *Statistics & Probability Letters*, 6 :125–128, 1987. 35

[DGS90] D. K. Dey, M. Ghosh, and C. Srinivasan. A new class of improved estimators of a multinormal precision matrix. *Statistics and decisons*, (8) :141–151, 1990. 5

[EM76] B. Efron and C. Morris. Multivariate empirical Bayes and estimation of covariance matrices. *Annals of Statistics*, 4(1) :22–32, 1976. 3, 4, 24

[FHM21] D. Fourdrinier, A. M. Haddouche, and F. Mezoued. Scale matrix estimation of an elliptically symmetric distribution in high and low dimensions. *Journal of Multivariate Analysis*, 181, 2021. 9, 20, 21, 69, 72, 77, 79

[FMW16] D. Fourdrinier, F. Mezoued, and M. T. Wells. Estimation of the inverse scatter matrix of an elliptically symmetric distribution. *Journal of Multivariate Analysis*, 143 :32–55, 2016. 7, 26

[Fou02] D. Fourdrinier. *Statistique inférentielle*. Paris : Dunod, 2002. 14

[FSW18] D. Fourdrinier, W. E. Strawderman, and M. T. Wells. *Shrinkage estimation*. Springer, 2018. 14, 15, 16, 17, 20

[FZ90] K. Fang and Y. T. Zhang. *Generalized multivariate analysis*. Berlin/Heidelberg/New York/London/Paris/Tokyo/Hong Kong, 1990. 14, 16

[GN99] A. K. Gupta and D. K. Nagar. *Matrix Variate Distributions*. Chapman and Hall/CRC, 1999. 18

[Haf77a] L. R. Haff. Minimax estimators for a multinormal precision matrix. *Journal of Multivariate Analysis*, (7) :374–385, 1977. 3, 4, 31

[Haf77b] L. R. Haff. Minimax estimators for a multinormal precision matrix. *Journal of Multivariate Analysis*, 7 :374–385, 1977. 31

[Haf79a] L. R. Haff. Estimation of the inverse covariance matrix : Random mixtures of the inverse Wishart matrix and the identity. *JThe Annals of Statistics*, 7(6) :1264–1276, 1979. 4, 20

[Haf79b] L. R. Haff. An identity for the Wishart distribution with applications. *Journal of Multivariate Analysis*, 9(4) :531–544, 1979. 4, 20

[Har08] D. A. Harville. *Matrix Algebra from statistician's perspective*. Springer, 2008. 11

[HKKR11] L. R. Haff, P. T. Kim, J. Y. Koo, and D. St. P. Richards. Minimax estimation for mixtures of Wishart distributions. *The Annals of Statistics*, 39(6) :3417–3440, 2011. 9, 19, 70

[HTF01] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics, 2001. 1

[JN03] R. K . Janik and M. A . Nowak. Wishart and anti-wishart random matrices. *J. Phys. A, Special Issue on Random Matrix Theory*, 26(12) :3629–3637, 2003. 18

[JS61] W. James and C. Stein. Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1 : Contributions to the Theory of Statistics*, pages 361–379, Berkeley, Calif., 1961. University of California Press. 58, 59

[JVO09] B. Jian, B. C. Vemuri, and E. Ozaeslan. A mixture of wisharts (mow) model for multifiber reconstruction. In *Visualisation and Processinf of Tensor Fields*, pages 39–56, 2009. 8

[KDM12] A. Kourtis, G. Dotsis, and R. N Markellos. Parameter uncertainty in portfolio selection : Shrinking the inverse covariance matrix. *Journal of Banking and Finance*, 36(9) :2522–2531, 2012. 2

[KI14] T. Kubokawa and A. Inoue. Estimation of covariance and precision matrices under scale-invariant quadratic loss in high dimension. *Electronic Journal of Statistics*, 8 :130–158, 2014. 3, 5, 7

[Kon09] Y. Konno. Shrinkage estimators for large covariance matrices in multivariate real and complex normal distributions under an invariant quadratic loss. *Journal of Multivariate Analysis*, 100 :2237–2253, 2009. 20

[KS08] T. Kubokawa and M. S. Srivastava. Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data. *Journal of Multivariate Analysis*, 99(9) :1906–1928, 2008. 6, 20

[LJ15] Y. Li and S. A. Jackson. Gene network reconstruction by integration of prior biological knowledge. *G3 Genes Genomes Genitecs*, 5, 2015. 2

[LM04] G. Letac and H. Massam. All invariant moments of the Wishart distribution. *Scandinavian Journal of Statistics*, 31(2) :295–318, 2004. 18

[LM07] G. Letac and H. Massam. Wishart distribution for decomposable graphs. *The Annals of Statistics*, 35(3) :1278–1323, 2007. 18

[MA07] A. Maaref and S. Aissa. Eigenvalue distributions of wishart-type random matrices with application to the performance analysis of mimo mrc systems. *G3 Genes Genomes Genitecs*, 6 :2678 – 2689, 2007. 18

[Mor07] E. Morris, L. *Multivariate Statistics : A Vector Space Approach.* Institute of mathematical statistics, 2007. 18

[Mui82] R. J. Muirhead. *Aspects of Multivariate Statistical Theory.* Wiley, New York, 1982. 14, 18

[NMSM17] S. F. V. Nielsen, K. H. Madsen, M. N. Schmidt, and M. Mørup. Modeling dynamic functional connectivity using a wishart mixture model. In *In 2017 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, 2017. 9

[Pal93] N. Pal. Estimating of the normal dispersion matrix and th precision matrix from a decision-theoretic point of view : a review. *Statistical papers*, 34, 1993. 3

[Pou13] M. Pourahmadi. High-dimensional covariance estimation : With high-dimensional data. *Wiley*, 2013. 1

[Sri07] M. S. Srivastava. Multivariate theory for analyzing high dimensional data. *JOURNAL OF THE JAPAN STATISTICAL SOCIETY*, 37(1) :53–86, 2007. 6, 7

[SS94] P. Y. S. Shao and W. E. Strawderman. Improving on the james-stein positive-part estimator. *annals of statistics*, 22 :1517–1538, 1994. 59

[Ste75] C. Stein. Estimation of a covariance matrix. In *Rietz Lecture, 39th Annual Meeting of the IMS*. Atlanta, Georgia, 1975. 20

[Ste81] C. Stein. Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, 9 :1135–1151, 1981. 20, 58

[TK06] H. Tsukuma and Y. Konno. On improved estimation of normal precision matrix and discriminant coeffecients. *Journal of multivariate analysis*, 97 :1477–1500, 2006. 5, 39, 40

[TK20] T. Tsukuma and T. Kubokawa. *shrinkage estimation for mean and covariance matrices.* Springer Singapore, 2020. 11, 18

[Wis28] J. Wishart. The generalised product moment distribution in samples from a normal multivariate. *Biometrika*, 20A :32–52, 1928. 18

[YYY+16] W. Yang, X. Yang, T. Yan, H. Song, and G. S Xia. Region-based change detection for polarimetric sar images using Wishart mixture models. In *IEEE Transactions on Geoscience and Remote Sensing*, 2016. 8