Université de Montréal

ESTIMATION ÉQUIVARIANTE DE PARAMÈTRES MULTIVARIÉS AVEC CONTRAINTES

par

N'deye Rokhaya Gueye

Département de mathématiques et de statistique
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Statistique

janvier 2000

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé

ESTIMATION ÉQUIVARIANTE DE PARAMÈTRES MULTIVARIÉS AVEC CONTRAINTES

présenté par

N'deye Rokhaya Gueye

a été évalué par un jury composé des personnes suivantes :

Martin Bilodeau
(président-rapporteur)

François Perron
(directeur de recherche)

Eric Marchand
(co-directeur)

Jean François Angers
(membre du jury)

Mémoire accepté le :
24 janvier 2000

SOMMAIRE

Ce mémoire porte sur l'étude et la recherche d'estimations de paramètres multivariés. On vise à obtenir des estimateurs équivariants qui minimisent le risque fréquentiste.

Dans un premier temps, on introduit des notations, définitions, propriétés et résultats nécessaires pour capitaliser sur la notion d'invariance. Ces notations, définitions, propriétés et résultats sont utilisés pour dégager la forme générale de la meilleure règle équivariante ainsi que sa fonction de risque. Dans un second temps, on évalue le meilleur estimateur équivariant et sa fonction de risque à partir de leur formes générales et dans le cas d'un modèle restreint. Ces résultats sont appliqués à plusieurs problèmes d'estimation :

– estimation de la moyenne μ d'un mélange de lois normales lorsque l'information paramétrique est de la forme $\mu^t \Sigma^{-1} \mu = \lambda^2$ où λ^2 est connu et Σ est la matrice de variance covariance qui est inconnue,

– estimation de la moyenne μ d'un mélange de lois normales sous la restriction $\Sigma = \frac{\mu^t \mu}{C^2} I$, où C est connu et désigne le coefficient de variation en dimension un.

Dans un troisième temps, on propose une étude numérique et comparative entre le meilleur estimateur équivariant et d'autres types estimateurs, lorsque le meilleur estimateur équivariant associé à une certaine contrainte est utilisé sous une autre contrainte. C'est-à-dire, on trouve δ_{λ_0} le meilleur estimateur équivariant lorsque la restriction dit que $\lambda = \lambda_0$ et on évalue le risque de δ_{λ_0} lorsque $\lambda \neq \lambda_0$.

REMERCIEMENTS

Je me permettrais d'exprimer ma reconnaissance la plus profonde à mon directeur François Perron et mon codirecteur Éric Marchand pour leur encadrement, leur très grande disponibilité, leurs conseils pédagogiques forts utiles. J'ai découvert en eux des maîtres courtois, ouverts et efficaces dotés d'un sens aigu dans la méthode, la rigueur et l'organisation du travail. Très sincères reconnaissances.

Je remercie tout le personnel enseignant du Département de mathématiques et de statistique de l'Université de Montréal pour la qualité des cours et les services dont j'ai pu bénéficier durant mes années de scolarité.

Je tiens également à exprimer ma reconnaissance au Programme Canadien de Bourses de la Francophonie pour l'appui financier qu'il m'a offert.

Mes remerciements vont à Claude Beauchamp pour sa disponibilité et pour m'avoir initié aux logiciels Fortran 77 et Matlab.

Finalement, je veux remercier de façon spéciale mon mari Saer pour tout le soutien moral et la patience dont il a fait preuve.

TABLE DES MATIÈRES

CHAPITRE 2

LISTE DES GRAPHIQUES

INTRODUCTION

Un des buts de la statistique est de fournir une inférence sur un paramètre θ disposant d'un échantillon $(X_1, X_2, ..., X_n)$ associé à une loi de probabilité P_θ. Les problèmes inférentiels peuvent se ramener à des problèmes d'estimations lorsqu'ils sont considérés sous un angle décisionnel, par le choix convenable d'une fonction de perte. Il s'agit alors de trouver la règle de décision qui minimise la fonction de risque fréquentiste parmi toutes les règles possibles. Si cela existe.

La recherche d'une solution générale s'avère parfois très difficile. Pour pallier à ce problème, nous allons nous restreindre à une sous-classe des règles qui est la classe des règles aléatoires équivariantes.

Notre étude portera alors à l'estimation équivariante de paramètres multivariés avec contraintes.

Ce mémoire est divisé en trois chapitres. Le premier chapitre traitera de l'invariance et de l'estimation équivariante. Ce sera une synthèse de plusieurs résultats existants et diffusés de manière éparpillée qu'on trouve dans Brown (1980), Andersson (1982), Strasser (1985), Lehmann (1986), Le Cam (1986), Eaton (1989), Wijsman (1990), Robert (1992) et Giri (1995).

Nous utiliserons d'abord l'approche de Bourbaki (1963) pour l'étude de l'intégration sur les groupes topologiques et les groupes de transformations. Ensuite nous nous servirons des propriétés de ces derniers pour introduire et étudier les problèmes invariants en caractérisant les règles équivariantes. Nous étudierons les liens existants entre les règles équivariantes et les règles aléatoires en s'inspirant de Giri (1975). Ensuite ceux existants entre les règles équivariantes et les règles de Bayes par l'intermédiaire de la mesure de Haar en se reférant de Nachbin (1965), Eaton (1989) et Robert (1992). Et enfin ceux existants entre les règles équivariantes par rapport à l'admissibilité et la minimaxité par le théorème de Hunt-Stein et sa généralisation en rassemblant les différents résultats publiés par Brown (1980), Lehmann (1986), Strasser (1985) et Giri (1995). Ce chapitre introduit les notations, les définitions, propriétés et les résultats, dont on a besoin pour l'application du deuxième chapitre.

Dans le deuxième chapitre, on abordera l'estimation équivariante en modèle admettant une statistique libre. La caractérisation des règles équivariantes faite dans le premier chapitre servira à trouver de façon explicite le meilleur estimateur équivariant pour le cas du modèle restreint. Notre façon de faire s'inspire de Kariya (1989). L'approche de Kariya (1989) est basée sur une généralisation de Pitman (1938), Girshich et Savage (1951) et Kiefer (1957). Le modèle restreint forme une orbite sous l'action du groupe induit agissant sur l'espace des paramètres, la statistique libre se comporte comme un invariant maximal. On utilisera les résultats obtenus pour l'application de quelques problèmes d'estimations équivariants et on évaluera le risque minimal sur un sous-espace de l'espace paramétrique comme dans les problèmes ci-dessous :

- L'estimation de la moyenne μ d'une distribution de mélange de lois normales $N_p(\mu, \Sigma)$ sous la contrainte $\mu'\Sigma^{-1}\mu = \lambda^2$ avec λ^2 connu.

- L'estimation de la moyenne μ d'une distribution de mélange de lois normales $\left(N_p(\mu, \frac{\mu^t \mu}{C^2} I)\right)$, où le coefficient de variation C^2 est connu.

Dans le premier problème, Kariya, Perron and Giri (1988) ont étudié le cas normal et ont évalué le meilleur estimateur équivariant pour $p = 2$. Nous nous sommes servis de l'approche de ces derniers pour évaluer le meilleur estimateur équivariant pour p quelconque. C'est par la suite que nous avons retrouvé les résultats de Marchand (1994) pour le même problème. Nous avons utilisé l'approche de Marchand (1990, 1993) et les résultats du deuxième chapitre pour étudier le comportement du meilleur estimateur équivariant. Le cas général des mélanges de lois normales $N_p(\mu, \Sigma)$ lorsque $\mu^t \Sigma^{-1} \mu = \lambda^2$ avec λ^2 connu est également proposé.

Dans le second problème, Nous étudierons d'abord le cas normal et ensuite le cas d'un mélange de lois normales. Perron (1987) a étudié le cas normal.

Le troisième chapitre portera sur une étude numérique et comparative du meilleur estimateur équivariant par rapport à d'autres types d'estimateurs équivariants lorsque l'information paramétrique est erronée. C'est-à-dire qu'on part de l'hypothèse $\mu^t \Sigma^{-1} \mu = \lambda_0^2$, $\lambda^2 \neq \lambda_0^2$ alors qu'en réalité $\mu^t \Sigma^{-1} \mu = \lambda^2$. Marchand (1990, 1993) étudie et compare le meilleur estimateur équivariant avec la classe des estimateurs linéaires et l'estimateur du maximum de vraisemblance dans le cas d'une loi normale lorsque $\mu^t \Sigma^{-1} \mu = \lambda_0^2$, $\lambda^2 \neq \lambda_0^2$. Nous reprenons cette approche de Marchand (1994) dans le cas le plus général d'un mélange de lois normales. La classe des estimateurs linéaires, la classe des estimateurs de James-Stein et la classe des estimateurs du maximum de vraisemblance sont les sous-classes d'estimateurs étudiées. Ce chapitre nous permettra d'étudier la robustesse du meilleur estimateur équivariant de la moyenne d'un mélange de lois normales vu dans le deuxième chapitre. Pour chaque sous-classe, on étudiera le meilleur estimateur équivariant et son comportement par rapport à ses autres estimateurs. Le risque minimal pour chaque sous-classe sera également évalué.

CHAPITRE 1

INVARIANCE ET ESTIMATION ÉQUIVARIANTE

La notion d'invariance est l'une des restrictions intuitives imposées par les fréquentistes aux estimateurs. En statistique, elle formalise la symétrie et elle permet de résoudre beaucoup de problèmes reliés aux tests d'hypothèses et à l'estimation. Le but de ce chapitre est de considérer le principe de l'invariance en statistique, pour chercher des estimateurs optimaux et de voir dans quelles conditions ces estimateurs sont minimax, admissibles ou coïncident avec des estimateurs de Bayes.

1.1. INTÉGRATION SUR LES GROUPES TOPOLOGIQUES.

Définition 1.1.1. Sur un ensemble G, on dit qu'une loi de composition notée $(\circ)$ partout définie détermine une structure de groupe si :

i) elle est interne, c'est-à-dire $\forall\, g_1, g_2 \in G,\ g_1 \circ g_2 \in G$,

ii) elle est associative, c'est-à-dire

$g_1 \circ (g_2 \circ g_3) = (g_1 \circ g_2) \circ g_3, \forall g_1, g_2, g_3 \in G,$

iii) elle possède un élément neutre, c'est-à-dire il existe un élément e de G tel que $e \circ g = g \circ e = g, \forall g \in G,$

iv) tout élément de G admet un élément inverse c'est-à-dire $\forall g \in G,$ il existe un élément noté $g^{-1} \in G,$ tel que $g^{-1} \circ g = g \circ g^{-1} = e, \forall g \in G.$

Un ensemble muni d'une structure de groupe prend le nom de groupe.

Dans ce mémoire, nous noterons multiplicativement la loi de composition interne et le groupe $(G,\circ)$ par G.

Définition 1.1.2. On appelle groupe topologique tout espace topologique G muni d'une structure de groupe telle que la fonction de $G \times G \to G$ définie par $(g_1,g_2) \to g_1 g_2^{-1}$ est continue.

Exemple 1.1.3. Soit $G = GL(n)$ où n est un entier naturel différent de zéro, l'espace des matrices réelles carrées $(n \times n)$ régulières muni de l'opération multiplicative des matrices. Cet espace forme un groupe.

En effet :

i) $\forall A, B \in GL(n)$ on a $AB \in GL(n)$,

ii) $(AB)\,C = A\,(BC), \forall A, B, C \in GL(n)$,

iii) La matrice identité I est l'élément neutre de $GL(n)$,

iv) $\forall A \in GL(n)$, il existe l'élément A^{-1}(inverse de la matrice A) tel que : $AA^{-1} = A^{-1}A = I$.

Il est clair que $GL(n) \subset \mathscr{L}_{n,n}$, l'espace des matrices réelles carrées $(n \times n)$.

Exemple 1.1.4. Soit $G = O(n)$ l'espace des matrices réelles carrées $(n \times n)$ orthogonales. On vérifie facilement que $O(n)$ est un sous-groupe de $GL(n)$, qu'il est un ensemble fermé et compact de $\mathscr{L}_{n,n}$ et qu'il possède les propriétés topologiques de $\mathscr{L}_{n,n}$. Les opérations de groupe sur $O(n)$ sont continues. D'où le groupe des matrices réelles carrées $(n \times n)$ orthogonales est un groupe topologique.

Par la suite nous allons toujours supposer les autres conditions suivantes sur G :

i) G est un espace séparable (Hausdorff),

ii) G est localement compact,

iii) la topologie de G dispose d'une base dénombrable.

Nous allons à présent donner les définitions d'une intégrale et d'une mesure de Radon pour ensuite définir la mesure de Haar sur un groupe.

Soit $\mathcal{X}$ un espace topologique quelconque. On note $K(\mathcal{X})$, l'ensemble de toutes les fonctions continues de $\mathcal{X}$ vers $\mathcal{X}$ à support compact.

Définition 1.1.5. On appelle intégrale toute fonctionnelle réelle J définie sur $K(\mathcal{X})$ qui satisfait les relations suivantes :

i) $J(a_1 f_1 + a_2 f_2) = a_1 J(f_1) + a_2 J(f_2)$, (linéarité)

ii) $J(f) \geq 0, \forall f \in K(\mathcal{X})$ et f non négative, (positivité)

iii) Il existe $f \in K(\mathcal{X})$ tel que $J(f) > 0$.

Définition 1.1.6. Soit $\mathcal{X}$ un espace topologique donné et μ une mesure sur $\sigma(\mathcal{X})$ les boréliens de $\mathcal{X}$. La mesure μ est une mesure de Radon si $\mu \neq 0$ et $\mu(C) < \infty$ pour tout ensemble compact $C \subset \mathcal{X}$.

Théorème 1.1.7. (théorème de représentation de Riez)

Si J est une intégrale définie sur $K(\mathcal{X})$ alors il existe une unique mesure de Radon μ telle que :

$$J(f) = \int f(x)\mu(dx), f \in K(\mathcal{X}).$$

Inversement, une mesure de Radon $\mu \neq 0$ définit une intégrale.

Définition 1.1.8. Soit G un groupe topologique. Pour tout $g \in G$, on définit la transformation L_g sur $K(G)$ par $(L_g f)(x) = f(gx), \quad \forall f \in K(G), \forall x \in \mathcal{X}, \; g \in G$.

Remarque 1.1.9. On a $L_g \circ L_h = L_{goh'}, \quad \forall g, h \in K(G)$.

En effet $\forall f \in K(G)$,

$$(L_g \circ L_h)(f)(x) = L_g(L_h(f))(x) = L_g(f)(hx) = f(g(h(x))) = f(g \circ h(x)) = L_{goh}(f)(x).$$

Définition 1.1.10. Une intégrale J sur $K(G)$ est invariante à gauche si pour tout $f \in K(G)$ et $g \in G$, on a :

$$J(L_g f) = J(f).$$

Si μ est la mesure de Radon associée à une intégrale J invariante à gauche alors elle satisfait la relation suivante :

$$\int f(gx)\mu(dx) = \int f(x)\mu(x),\ g \in G,\ f \in K(G),\ f \text{ intégrable.} \qquad (1.1.1)$$

La relation (1.1.3) entraîne qu'une mesure μ est invariante à gauche si :

$$\mu(gA) = \mu(A),\ g \in G \text{ et } A \subset \sigma(G).$$

Définition 1.1.11. Si une intégrale J définie sur $K(G)$ est invariante à gauche alors la mesure de Radon associée à J est dite mesure de Haar à gauche et elle est notée ν_ℓ.

Cette mesure ν_ℓ vérifie alors la relation (1.1.3). On montre qu'elle est unique à un coefficient multiplicatif près.

De même, on définit sur $K(G)$ la transformation R_g par :

$$fR_g(x) = f(xg), f \in K(G), x \in G,\ g \in G.$$

De manière similaire, on définit par cette transformation R_g, l'intégrale invariante à droite J_r et la mesure de Haar à droite ν_r (définie à un coefficient près).

Théorème 1.1.12. Si ν_ℓ (respectivement ν_r) désigne la mesure de Haar à gauche sur G (respectivement la mesure de Haar à droite sur G) alors il existe une fonction $\Delta : G \to IR^+$ continue qui relie ν_ℓ et ν_r par la relation suivante :

$$\nu_r(dx) = \Delta(x^{-1})\nu_\ell(dx).$$

Cette fonction est appelée le module de G.

De plus, elle vérifie la relation suivante :

$$\Delta(g_1 g_2) = \Delta(g_1)\Delta(g_2),\ \forall g_1, g_2 \in G.$$

Démonstration. Voir Eaton (1989, page 7)

Propriétés 1.2.13 Soit Δ un module de G. Alors Δ vérifie les propriétés suivantes :

- $\Delta(e) = 1$,
- $\Delta(g_1 g_2) = \Delta(g_2 g_1)$,
- $\Delta(g^{-1}) = \dfrac{1}{\Delta(g)}$,
- $\Delta(G) = \{\Delta(g); g \in G\}$ est un sous-groupe de IR^+.

Si $\Delta(g) = 1, \forall g \in G$, alors on dit que G est unimodulaire.

Théorème 1.1.14. Si le groupe G est localement compact, une condition nécessaire et suffisante pour qu'il existe une mesure de Haar finie sur G est que G soit compact.

Démonstration. Voir Nachbin (1965, page 75).

Exemple 1.1.15. Soit le groupe IR^n muni de l'addition des vecteurs. Alors la mesure de Lebesgue sur IR^n est invariante à gauche et à droite. De plus, le module Δ est égal à un.

Exemple 1.1.16. Prenons $G = GL(n)$ comme dans l'exemple 1.1.3. Soit dx la mesure de Lebesgue restreinte à $GL(n)$. On sait que $GL(n)$ est un ouvert de $\mathscr{L}_{n,n}$ (espace des matrices carrées $(n \times n)$) et qu'il a une mesure de Lebesgue infinie. On définit l'intégrale J par :

$$J(f) = \int f(x) \frac{dx}{|\det(x)|^n}.$$

Montrons que J est invariante à gauche et à droite.

$$J(L_g f) = \int f(gx) \frac{dx}{|\det(x)|^n}.$$

Faisons le changement de variable $y = gx$. Le jacobien de la transformation est $|\det(g)|^{-n}$ (le calcul du jacobien est donné par Eaton (1989, exemple 1.9).

Donc on obtient l'égalité suivante :

$$J(L_g f) = \int f(y) \left| \frac{\det(g)^{-n}}{\det(g^{-1}y)^n} \right| dy$$

$$= \int f(y) \frac{dy}{|\det(y)|^n}$$

$$= J(f).$$

D'où J est invariante à gauche.

La démonstration J invariante à droite est similaire à la précédente. Il suffit seulement de remplacer gx par xg.

Le module d'un groupe topologique G est un cas particulier du multiplicateur de G. Définissons la notion de multiplicateur d'un groupe topologique G.

Définition 1.1.17. On appelle multiplicateur de G, toute fonction χ définie de $G \to (0,\infty)$ qui satisfait les conditions suivantes :

i) χ est continue,

ii) $\chi(g_1g_2) = \chi(g_1)\chi(g_2), \forall g_1, g_2 \in G.$

Propriétés 1.1.18. Soit χ un multiplicateur de G. Alors χ vérifie les propriétés suivantes :

- $\chi(e) = 1$,
- $\chi(g_1g_2) = \chi(g_2g_1)$,
- $\chi(g^{-1}) = \dfrac{1}{\chi(g)}$,
- $\chi(G) = \{\chi(g) : g \in G\}$ est un sous-groupe de IR^+.

Définissons l'intégrale relativement invariante.

Définition 1.1.19. Une intégrale J sur $K(G)$ est relativement invariante avec multiplicateur χ, si elle vérifie la relation suivante :

$$J(L_g f) = \chi(g) J(f), \forall g \in G.$$

Remarque 1.1.20. Si G est compact alors $\chi(G)$ est aussi compacte sur IR^+.

Dans ce cas, on a $\chi(G) = \{1\}$.

Le multiplicateur $\chi \equiv 1$, dit multiplicateur trivial, est le seul multiplicateur définit sur un compact. Il faut voir dans ce cas que si J est relativement invariante alors J est invariante à gauche c'est-à-dire $J(L_g f) = J(f)$.

Soit χ_ℓ (respectivement χ_r) un multiplicateur à gauche sur G (respectivement à droite sur G).

Définition 1.1.21. Soit G un groupe localement compact. On dit qu'une mesure μ est relativement invariante à gauche (respectivement à droite) sur G avec multiplicateur à gauche χ_ℓ (respectivement à droite χ_r), si :

$$\mu(gf)=\chi_\ell(g)\mu(f),\text{ pour } f\in K(G),\ g\in G$$

(respectivement $\mu(fg)=\chi_r(g)\mu(f)$, pour $f\in K(G)$ et $g\in G$).

La mesure de Haar à gauche est relativement invariante avec multiplicateur à gauche $\chi_g\equiv 1$, et le multiplicateur à droite devient $\chi_d=\Delta_d$.

La mesure de Haar à droite est relativement invariante avec multiplicateur à droite $\chi_d\equiv 1$ et le multiplicateur à droite $\chi_\ell=\Delta_\ell$.

Théorème 1.1.22. Soit χ est un multiplicateur sur G, on a les assertions suivantes :

i) Si $J(f)=\int f(x)\mu(dx)$ est relativement invariante avec multiplicateur χ, alors $J(f)=\int f(x)\chi(x)\mu(dx)$ est invariante.

ii) Si $J(f)=\int f(\mathrm{x})v_g(d\mathrm{x})$ est invariante à gauche avec multiplicateur χ, alors $J(f)=\int f(\mathrm{x})\chi(\mathrm{x})\mu(d\mathrm{x})$ est relativement invariante à gauche avec multiplicateur χ.

Démonstration.

i)
$$\begin{aligned}J(L_g f)&=\int f(g\mathrm{x})\chi(\mathrm{x})\mu(d\mathrm{x})\\&=\int f(g\mathrm{x})\chi\left(g^{-1}g\mathrm{x}\right)\mu(d\mathrm{x})\end{aligned}$$

d'où la relation suivante :

$$J\left(L_g f\right)=\chi\left(g^{-1}\right)\chi(g)\int f(x)\chi(x)\mu(dx)=J(f).$$

De façon similaire, on montre ii).

La définition suivante servira à énoncer le théorème de Hunt-Stein présenté dans la dernière section. Ce théorème permet d'établir la relation entre la minimaxité et l'invariance.

L'existence d'une suite de probabilités est la condition nécessaire sur laquelle repose le théorème de Hunt-Stein.

Définition 1.1.23. Le groupe G est moyennable s'il existe une suite de probabilités $\{\nu_n\}_{n\in IN}$ sur G telles que :

$$\lim_{n\to\infty}\left|\nu_n(Bg)-\nu_n(B)\right|=0,\ \forall\, g\in G,\ B\in\sigma(G).$$

Remarques 1.1.24.

i) Tout groupe compact est moyennable.

ii) Tout groupe commutatif est moyennable.

iii) Tout sous-groupe fermé d'un groupe moyennable est moyennable.

iv) Le groupe $GL(p)$ des matrices carrées régulières et le groupe $SL(p)$ des matrices carrées $(n \times n)$ de déterminant égal à 1 ne sont pas moyennables.

Remarque 1.1.25. Les groupes additifs (groupes commutatifs), multiplicatifs sur IR^+, de changements d'échelle, des matrices orthogonales (car groupe compact) sont moyennables. Le groupe $T(p)$ des matrices triangulaires supérieures inversibles est aussi moyennable.

1.2. L'ACTION D'UN GROUPE SUR UN ENSEMBLE.

Pour un ensemble non vide $\mathcal{X}$. On note $\sigma_{\mathcal{X}}$ l'ensemble des applications biunivoques (des transformations mesurables) de $\mathcal{X}$ sur lui-même. Cet ensemble muni de la loi de composition $(\circ)$ est un groupe. On l'appelle groupe symétrique ou groupe des permutations de l'ensemble $\mathcal{X}$. (Pour plus de détail sur cet groupe consulter Bourbaki (1963)).

1.2.1. Opération d'un groupe sur un ensemble $\mathcal{X}$.

Définition 1.2.1 On appelle groupe de transformations G (ou groupe de permutations) de l'ensemble $\mathcal{X}$, tout sous-groupe du groupe symétrique $\sigma_{\mathcal{X}}$.

Donc un groupe de transformations G sur $\mathcal{X}$ doit satisfaire les assertions suivantes :

i) le groupe $G \subset \sigma_{\mathcal{X}}$,

ii) pour tout $g_1, g_2 \in G,\ g_1 \circ g_2 \in G$,

iii) pour tout $g_1, g_2, g_3 \in G$, $(g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3)$,

iv) la transformation identité e, définie par $e(x) = x,\ \forall x \in G$ est dans G.

v) pour tout $g \in G$, il existe la transformation inverse $g^{-1} \in G$ telle que

$$g^{-1} \circ g = g \circ g^{-1} = e.$$

Dans beaucoup d'exemples le groupe G est un groupe de transformations sur $\mathcal{X}$. Donc les éléments de G sont des fonctions bijectives définies de $\mathcal{X}$ vers $\mathcal{X}$.

Définition 1.2.2 On dit qu'un groupe G opère à gauche sur un ensemble $\mathcal{X}$ s'il existe une fonction $H : G \times \mathcal{X} \to \mathcal{X}$ qui satisfait les conditions suivantes :

i) $\forall g \in G$, la fonction $\mathrm{H}(\mathrm{g}, \cdot) : \mathcal{X} \to \mathcal{X}$ est bijective,

ii) $H(e,x) = x, \forall x \in \mathcal{X}$,

iii) $H(g_1, H(g_2, x)) = H(g_1 g_2, x), \forall g_1, g_2 \in G, x \in \mathcal{X}$.

La définition pour l'opération à droite est la même sauf au iii) où on a :

$H(g_2, H(g_1, x)) = H(g_1 g_2, x), \ \forall g_1, g_2 \in G$ et $x \in X$.

Pour simplifier les notations, on supprime le symbole de la fonction H et on écrit simplement gx au lieu de $H(g,x)$ si l'action est à gauche et xg au lieu de $H(g,x)$ si l'action est à droite. Avec cette notation les trois propriétés définies pour l'action à gauche s'écrivent :

i) $\forall g \in G$, la fonction $\mathrm{H}(\mathrm{g}, \cdot) : \mathcal{X} \to \mathcal{X}$ est bijective,

ii) $ex = x, \forall x \in \mathcal{X}$,

iii) $\mathrm{g}_2(\mathrm{g}_1, \mathrm{x}) = (g_2 g_1)(x), \ \forall \mathrm{g}_1, \mathrm{g}_2 \in \mathrm{G}$ et $\mathrm{x} \in \mathcal{X}$.

Remarques 1.2.3. Supposons que G opère à gauche sur $\mathcal{X}$ alors on a les assertions suivantes :

- Si $gx = x, \forall x \in \mathcal{X}$ alors l'action est dite triviale.

- Si $gx \neq x$, $\forall g \in G - \{e\}$ et $\forall x \in \mathcal{X}$, alors on dit que l'action est libre.

Exemple 1.2.4 Supposons que $G = GL(n)$ et $\mathcal{X} = \mathrm{IR}^n$. On définit la fonction $F : GL(n) \mathrm{x}\ \mathrm{IR}^n \to \mathrm{IR}^n$ par $F(C,x) = Cx$ où Cx est le produit d'une matrice C et d'un vecteur x. On vérifie que :

i) pour tout $C \in GL(n)$, la fonction $x \to Cx$ est bijective.

En effet nous avons les résultats suivants :

- si $Cx = Cy,\ \forall x,\ y \in \mathrm{IR}^n$ alors $C^{-1}Cx = y$, d'où l'injection,

- soit $x, y \in \mathrm{IR}^n$ tel que $x = C^{-1}y$ alors $Cx = y$. On a alors la surjection.

ii) $Ix = x,\ \forall x \in \mathrm{IR}^n$,

iii) $B(Cx) = (\mathrm{B}C)x, \forall B, \mathrm{C} \in GL(n), \forall x \in \mathrm{IR}^n$.

Donc le groupe $GL(n)$ opère à gauche sur IR^n.

Tout sous-groupe de $GL(n)$ opère à gauche sur IR^n en particulier $O(n)$.

L'opération du groupe G sur $\mathcal{X}$ induit une relation d'équivalence à gauche sur les éléments de $\mathcal{X}$. Notons cette relation $\Re_g$. Elle est définie par : $x\Re_g y$ si et seulement s'il existe $g \in G$ tel que $y = gx$.

Les orbites de $\mathcal{X}$ sont les classes d'équivalences induites par la relation $\Re_g$. On note $O_x = \{gx : g \in G\}$ l'orbite de x. Ce qui veut dire que deux points sont équivalents s'ils sont sur la même orbite.

Définition 1.2.5. On dit que G opère transitivement sur $\mathcal{X}$ si l'ensemble $\mathcal{X}$ est formé d'une seule orbite.

Exemple 1.2.6. Considérons l'exemple 1.2.4 où $GL(n)$ opère à gauche sur IR^{n}. On ne peut pas dire que l'action de $GL(n)$ sur IR^{n} est transitive.

Mais l'ensemble $\mathrm{IR}^{\mathrm{n}} - \{0\}$ est invariant sous le groupe $GL(n)$ et l'action de G sur cet ensemble est donc transitive.

En effet soit $x, y \in \mathrm{IR}^{\mathrm{n}} - \{0\}$, il existe $C_1, C_2 \in GL(n)$ où leur première colonne est respectivement x, y. Si on pose $C = C_1 C_2^{-1}$, on a $Cx = y$ d'où x et y sont dans la même orbite. L'action de $GL(n)$ sur $\mathrm{IR}^{\mathrm{n}} - \{0\}$ est transitive. Cette action est donc transitive.

Exemple 1.2.7. (exemple d'une action non transitive).

On avait vu dans l'exemple 1.2.4 que $O(n)$ opère à gauche sur IR^n et même sur $\mathrm{IR}^{\mathrm{n}} - \{0\}$. Soit $x \in \mathcal{X}$, l'orbite de x est $O_x = \{y : \|y\| = \|x\|\}$. C'est donc la sphère de centre 0 et de rayon $\|x\|$. Mais l'espace $\mathcal{X}$ est la réunion de toutes les sphères de centre 0 qui constituent donc les orbites de $\mathcal{X}$. D'où l'action de $O(n)$ sur $\mathcal{X}$ n'est pas transitive.

Définition 1.2.8. Un groupe opère topologiquement à gauche sur $\mathcal{X}$ si G opère à gauche sur $\mathcal{X}$ et si l'action de G $(H : G \times \mathcal{X} \to \mathcal{X})$ est continue.

Nous allons à présent introduire la notion d'opération propre. Pour cela nous avons besoin de la définition suivante.

Définition 1.2.9. Soient $\mathcal{X}$ et $\mathcal{Y}$ deux espaces topologiques. Une fonction f de $\mathcal{X} \to \mathcal{Y}$ est fermée si l'image de tout fermé de $\mathcal{X}$ est un fermé de $\mathcal{Y}$. (Pour plus de détails voir Bourbaki (1963)).

Définition 1.2.10. Soient $\mathcal{X}$ et $\mathcal{Y}$ deux espaces topologiques et f une fonction de $\mathcal{X} \to \mathcal{Y}$, on dit que f est une fonction propre si elle vérifie les conditions suivantes :

- elle est continue,

- pour tout espace topologique Z, la fonction $f \times i_z : \mathcal{X} \times Z \to \mathcal{Y} \times Z$ est fermée, où i_z désigne la fonction identité sur Z.

Théorème 1.2.11. Considérons $\mathcal{X}$ et $\mathcal{Y}$ deux espaces localement compacts. Soit $f : \mathcal{X} \to \mathcal{Y}$ une fonction continue. Alors f est propre si et seulement si l'image réciproque de tout compact inclus dans $\mathcal{Y}$ est un compact de $\mathcal{X}$, c'est-à-dire $f^{-1}(K)$ est compact pour tout compact $K \subset \mathcal{Y}$.

Démonstration. Voir Wijsman (1990, page 27).

Soit la fonction $\theta : G \times \mathcal{X} \to \mathcal{X} \times \mathcal{X}$ définie par :

$$\theta(g, x) = (gx, x).$$

Définition 1.2.12. On dit que G opère proprement sur un espace topologique $\mathcal{X}$ si la fonction θ est propre.

Proposition 1.2.13 Soit G un groupe localement compact opérant topologiquement sur un espace localement compact $\mathcal{X}$. Alors l'action est propre si et seulement si pour toute paire de sous espaces compacts A, B de $\mathcal{X}$,

$((A, B)) = \{g \in G : gA \cap B \neq \varnothing\}$ est un compact fermé.

Démonstration. Voir Wijsman (1990, page 32).

Corollaire 1.2.14. Si G est un groupe compact opérant topologiquement sur un espace localement compact alors l'action est propre.

Démonstration. En utilisant la proposition précédente et le fait que G soit compact, $((A,B)) = \{g \in G : gA \cap B \neq \emptyset\} \subset G$ est compact.

Ce qui achève la démonstration.

Théorème 1.2.15. Considérons G un groupe topologique localement compact opérant proprement sur un espace localement compact $\mathcal{X}$. Soit $x \in \mathcal{X}$ alors on a les résultats suivants :

i) l'espace quotient $\mathcal{X} / G$ est localement compact,

ii) la fonction $\alpha_x : G \to \mathcal{X}$ par $\alpha_x(g) = gx$, est propre,

iii) le sous-groupe $G_x = \{g \in G : gx = x\}$ de G est compact,

iv) l'espace $Gx = \{gx \in \mathcal{X} : g \in G\}$ est un fermé de $\mathcal{X}$,

v) La bijection $\Psi_x : G / G_x \to Gx$ définie par $\Psi_x(gG_x) = gx$ est un homéomorphisme.

Démonstration. Voir Wijsman (1990, page 34).

Exemple 1.2.16. L'action de $O(n)$ sur IR^n est propre car $O(n)$ est compact.

Exemple 1.2.17. Par contre l'action de $GL(n)$ sur IR^n n'est pas propre.

En effet pour $x = (1,0,\ldots,0)'$, le groupe G_x égal à l'ensemble de toutes les matrices régulières qui ont pour première colonne x n'est pas compact. Puisque G_x n'est pas compact, d'après le théorème précédent, l'action n'est pas propre.

1.2.2. Intégrale relativement invariante.

On suppose que G opère transitivement sur $\mathcal{X}$. On définit la fonction L_g par :

$$L_g : K(\mathcal{X}) \to K(\mathcal{X})$$
$$f \to f(gx), \forall x \in X.$$

On vérifie facilement que :

$$L_g \circ L_h = L_{gh}.$$

Définissons les intégrales invariantes, relativement invariantes avec multiplicateur χ sur $K(G)$.

Définition 1.2.18. Soit χ un multiplicateur de G, une intégrale J sur $K(\mathcal{X})$ est relativement invariante (à gauche) avec multiplicateur χ si :

$$J(L_g f) = \chi(g) J(f), \forall f \in K(\mathcal{X}) \text{et } g \in G. \tag{1.2.1}$$

Si μ est la mesure de Radon associée à J, la relation (1.2.1) est équivalente à la relation suivante :

$$\int_X f(g^{-1}x)\mu(dx) = \chi(g)\int_{\mathcal{X}} f(x)\mu(dx), \forall f \in K(G) \text{ et } g \in G.$$

Exemple 1.2.19. Si on pose $G = GL(n)$ et $\mathcal{X} = \mathrm{IR}^n$, comme dans l'exemple 1.2.4, alors la mesure de Lebesgue est relativement invariante avec multiplicateur

$$\chi(g) = (\det(g))^{-1}.$$

En effet nous avons l'égalité suivante :

$$J(L_g f) = \int_{\mathcal{X}} f(gx) dx.$$

En faisant un changement de variable $y = gx$, on a dy=det$(g)\,dx$.

On obtient alors la relation suivante :

$$J(L_g f) = (\det(g))^{-1} \int f(y)\,dy.$$

D'où la relation précédente devient :

$$J(L_g f) = \chi(g) J(f).$$

Exemple 1.2.20. Posons $G = O(n)$ et $\mathcal{X} = \mathrm{IR}^n$. Considérons la mesure de probabilité P sur $\mathcal{X}$. Soit $X \in \mathcal{X}$, un vecteur aléatoire qui a pour distribution P, on pose $P = \mathcal{L}(X)$.

On dit que le vecteur aléatoire X à une distribution sphérique si et seulement si il vérifie :

$$\mathcal{L}(X) = \mathcal{L}(gX), \ \forall g \in O(n).$$

Ce qui est équivalent à X a une distribution sphérique si et seulement si P est invariante sous $O(n)$. Donc les distributions sphériques sur IR^n sont relativement invariantes (avec multiplicateur 1) sous l'action de $O(n)$.

Cherchons les conditions d'existences d'une mesure relativement invariante avec multiplicateur.

On suppose que G opère transitivement sur $\mathcal{X}$. Soit x_0 un point fixé. Définissons la fonction π_{x_o} par :

$$\pi_{x_o} : G \to \mathcal{X}$$

$$g \to gx_0$$

Il est facile de montrer que cette fonction est ouverte c'est-à-dire que l'image de tout ouvert de G est un ouvert de $\mathcal{X}$. Elle est aussi surjective (car G est transitif).

On pose $G_{x_o} = \{g : gx_0 = x_0\}$ alors G_{x_o} est un sous-groupe fermé de G.

Notons par Δ_{x_o} le module de G_{x_o} et par Δ le module de G.

Théorème 1.2.21 (Weil). Sous les hypothèses précédentes, il existe une intégrale J sur $K(\mathcal{X})$ relativement invariante avec multiplicateur χ sur $\mathcal{X}$ si et seulement si χ vérifie la relation suivante :

$$\Delta_{x_o}(g) = \chi(g)\Delta(g), \forall g \in G_{x_o}. \tag{1.2.2}$$

De plus, lorsque J existe, elle est unique à une constante près.

Démonstration. Voir Nachbin (1965, page 138).

On en déduit de ce théorème l'existence de la mesure de Radon μ associée à J si et seulement si χ vérifie la relation (1.2.2).

Trois corollaires très importants découlent de ce théorème. Pour ces trois corollaires, nous travaillerons sous les conditions du théorème 1.2.21.

Corollaire 1.2.22. Soit χ un multiplicateur sur G. Si le groupe G_{x_o} est "unimodulaire" alors il existe une mesure de probabilité $\mu \neq 0$ sur $\mathcal{X}$ relativement invariante avec multiplicateur χ.

Démonstration. Voir Nachbin (1965, page 140).

Corollaire 1.2.23. Si G et G_{x_o} sont unimodulaires alors il existe une mesure de probabilité $\mu \neq 0$ sur $\mathcal{X}$ G-invariante.

Démonstration. Voir Nachbin (1965, page 140).

Corollaire 1.2.24. Si G est compact alors il existe une mesure de probabilité sur G invariante.

Démonstration. Voir Nachbin (1965, page 140).

1.2.3. Invariant maximal.

On suppose que G opère sur $\mathcal{X}$. Donc les classes d'équivalences sont les orbites de $x \in \mathcal{X}$. Introduisons la notion de fonction invariante maximale.

Définition 1.2.25. Soit une fonction $t : \mathcal{X} \to \mathcal{Y}$. Alors on a les définitions suivantes :

- t est dite invariante sous l'action du groupe G si $t(X) = t(gX), \forall X \in \mathcal{X}$, $\forall g \in G$

- t est dite invariante maximale sous l'action du groupe G si elle est invariante sous G et si elle vérifie la condition suivante :

si $t(x) = t(y)$ alors il existe $g \in G$ tel que $y = gx$.

Exemple 1.2.26. Considérons le groupe $G = O(n)$. Ce groupe opère sur IR^n.

Soit $x \in \mathrm{IR}^n$ alors il existe $g \in G$ tel que :

$gX = \|X\| X_0$ *où* $\|X\|$ désigne la norme de X et $X_0 = (1,0...,0)'$.

La fonction $f(X) = \|X\|$ est invariante maximale.

En effet

i) $f(gX) = \|gX\| = \|X\| = f(X), \forall g \in O(n)$.

ii) Soient X, Y tels que $f(X) = f(Y)$ alors $\|X\| = \|Y\|$.

Soient $g_1, g_2 \in O(n)$ tels que $g_1 X = \|X\| X_0$ et $g_2 Y = \|Y\| X_0$. Comme $\|X\| = \|Y\|$ alors on a $g_1 X = g_2 Y$. Donc en prenant $g = {g_2}^{-1} g_1$, on a $Y = gX$. Ce qui veut dire que X et Y sont sur la même orbite.

Il faut remarquer que :

- toute fonction bijective basée sur $\|X\|$ est invariante maximale.

- une fonction f sur $\mathcal{X}$ est invariante maximale sur $O(n)$ si et seulement si elle s'écrit comme fonction de $\|X\|$.

Remarques 1.2.27.

- La fonction t est invariante si et seulement si t est constante sur chaque orbite de $\mathcal{X}$ (donc t est une fonction de toute statistique invariante maximale).

- La fonction t est invariante maximale si et seulement si elle est constante sur chaque orbite de $\mathcal{X}$ et prend des valeurs différentes sur différentes orbites.

- Une statistique invariante maximale indexe les orbites de G. Et si G est transitif alors les seules statistiques invariantes maximales sont les fonctions constantes.

Théorème 1.2.28. Si t est une fonction invariante maximale sur G alors toute fonction h est invariante sous G si et seulement si h est une fonction de t.

Démonstration .

- Soit h une fonction satisfaisant la factorisation suivante :

$$h(X) = s(t(X)).$$

On a alors $h(gX) = s(t(gX))$, comme t est invariante donc $t(X) = t(gX)$, on obtient alors $h(gX) = s(t(X)) = h(X)$, d'où h est invariante.

- Supposons que h est invariante donc $h(X) = h(gX), \forall X \in \mathcal{X}$ et $g \in G$, comme t est invariante maximale donc si $t(X) = t(Y)$ alors $Y = gX$.

On obtient alors $h(Y) = h(gX) = h(X)$, d'où h est une fonction de t.

Ce théorème est très utile dans le sens où si on connait une fonction invariante maximale alors toutes les fonctions invariantes seront connues.

1.3 PROBLÈME INVARIANT.

1.3.1. Problème statistique décisionnel.

On considère $\mathcal{X}$ l'espace échantillonnal, $\mathcal{A}$ une tribu sur $\mathcal{X}$. Soit une famille de lois de probabilités $\mathcal{P}$ définie sur $\mathcal{A}$. Si le vecteur aléatoire X associé aux observations prend ses valeurs dans $\mathcal{X}$ et a pour loi de probabilité $\mathcal{L}(X) \in \mathcal{P}$, alors on dit que le triplet $\{\mathcal{X}, \mathcal{A}, \mathcal{P}\}$ est un modèle statistique.

Lorsque $\mathscr{P}$ s'écrit $\{P_\theta, \theta \in \Theta\}$, Θ est dit l'espace paramétrique. On dira dans ce cas que le modèle statistique est paramétrisé et est noté $\{\mathscr{X}, \mathscr{A}, \Theta\}$.

Une fois le modèle statistique construit, on cherche à établir une inférence sur le paramètre θ. On utilise l'observation x connue pour estimer la valeur du paramètre θ inconnue du statisticien. Le rôle de ce dernier consiste à tirer une conclusion concernant ce paramètre θ en utilisant toujours l'observation x. Cette conclusion est appelée décision en statistique. Elle est choisie parmi un ensemble de décisions noté $\mathscr{D}$. Cet ensemble est appelé l'espace des décisions. On munit cette espace d'une tribu $\mathscr{B}$ tel que $(\mathscr{D}, \mathscr{B})$ est un espace mesurable et que la tribu $\mathscr{B}$ soit assez grande pour contenir tous les singletons de $\mathscr{D}$.

Définissons à présent une probabilité de transition δ^* sur $(\mathscr{X} \times \mathscr{D}, \mathscr{A} \times \mathscr{B})$. Elle vérifie :

i) pour tout $x \in \mathscr{X}$, $\delta^*(x, \cdot)$ est une loi de probabilité sur $\mathscr{B}$,

ii) pour tout $B \in \mathscr{B}$, la fonction $\delta^*(\cdot, B) : \mathscr{X} \to [0,1]$ est mesurable.

Cette probabilité de transition est appelée règle de décision aléatoire. On note Λ l'ensemble de toutes les règles de décision aléatoires.

Il existe un sous-ensemble très important de Λ, celui des règles pures.

On dit que la règle de décision δ^* est une règle pure, si pour chaque $x \in \mathscr{X}$, $\delta^*(x, \cdot)$ donne probabilité un à l'atome $\{\delta(x)\}$ dans $\mathscr{B}$, où $\delta : \mathscr{X} \to \mathscr{D}$. Dans ce cas on la note $\delta(x)$. On note Δ, l'ensemble de toutes les règles pures.

Théorème 1.3. 1. Si δ^* est une règle de décision aléatoire telle que :

$\delta^*(x,\{\delta(x)\}) = 1,\ \forall x \in \mathcal{X}$. Alors δ est une fonction mesurable de $\mathcal{X}$ dans $\mathcal{D}$.

Démonstration. Voir Perron (1987, page 8).

En pratique, le statisticien réalise au moyen de l'observation $x \in \mathcal{X}$ connue, en se basant sur la loi de probabilité de $\delta^*(x,\cdot)$, ce qu'on appelle une expérience aléatoire.

C'est donc le résultat de cette expérience aléatoire qui est la décision d prise par le statisticien.

Pour mesurer la précision de cette décision, on introduit une fonction appelée fonction de perte.

Définition 1.3.2. On appelle fonction de perte (ou de coût espéré), la fonction L définie sur $\Theta \times \mathcal{D}$ à valeurs dans IR^+ telle que $\forall \theta \in \Theta$ la fonction suivante :

$$L(\theta,\cdot): \mathcal{D} \rightarrow \mathrm{IR}^+,$$

est mesurable.

Elle prend une petite valeur si la décision $d \in D$ est bonne et une grande valeur si le contraire se produit.

Définition 1.3.3. On appelle modèle paramétrique décisionnel l'ensemble $\{\mathcal{X},\mathcal{A},\Theta,\mathcal{D},\mathcal{B},L\}$ où chaque élément est défini plus haut.

Dans tout ce qui suit, on désigne par $\{\mathcal{X},\mathcal{A},\Theta,\mathcal{D},\mathcal{B},L\}$ un modèle décisionnel.

L'existence de la fonction de perte, sous l'hypothèse de rationalité du statisticien n'est pas établie dans ce mémoire. Pour plus d'informations voir Berger (1985, chapitre 2), Robert (1992, chapitre 2).

On cherche à minimiser cette fonction de coût par l'intermédiaire le risque fréquentiste donné par la formule suivante :

$$R(\theta,\delta) = \int_{\mathcal{X}} \int_{\mathcal{D}} L(\theta, y)\,\delta^*(x, dy) P_\theta(dx).$$

Définition 1.3.4. On appelle fonction de risque la fonction

$$R : \Lambda \text{ x } \Theta \to IR^+ \cup \{\infty\}$$

$$(\delta,\theta) \to \int_{\mathcal{X}} \int_{\mathcal{D}} L(\theta, y)\delta^*(x, dy) P_\theta(dx).$$

Si δ est une règle pure alors $R(\theta,\delta) = \int_{\mathcal{X}} L(\theta,\delta(x)) P_\theta(dx)$.

Cette fonction est bien définie car δ^* est une probabilité de transition et L une fonction positive mesurable. Elle est utilisée pour comparer deux règles de décisions. La meilleure règle est celle qui a le plus petit risque possible.

1.3.2. Problème décisionnel invariant.

Pour introduire l'invariance dans les problèmes décisionnels, il faut nécessairement utiliser les groupes de transformations.

Soit le modèle $\{\mathcal{X}, \mathcal{A}, \Theta, \mathcal{D}, \mathcal{B}, L\}$ et soit G un groupe topologique quelconque.

On associe au groupe G des transformations bijectives de $\mathcal{X} \to \mathcal{X}$. Ainsi, on crée le groupe de transformations (noté $\widetilde{G}$) induit par G. On a par définition

$$\widetilde{G} = \{\widetilde{g} : \mathcal{X} \to \mathcal{X} \text{ bijective, } g \in G\}.$$

On suppose toujours l'identifiabilité c'est-à-dire $\forall \theta_1, \theta_2 \in \Theta$, $\theta_1 \neq \theta_2$ il existe $A \in \mathcal{A}$ tel que $P_{\theta_1}(A) > P_{\theta_2}(A)$.

Définition 1.3.5. Le modèle statistique $\{\mathcal{X}, \mathcal{A}, \Theta\}$ est dit invariant sous l'action du groupe G, si $\forall g \in G$ et $\theta \in \Theta$, il existe $\theta' \in \Theta$ tel que la distribution de $\tilde{g}(X)$ est $P_{\theta'}$ si P_θ est la distribution de X.

Par identifiabilité, θ' est unique. Ce qui définit $\bar{g}$ par la relation $\bar{g}\theta = \theta'$. On vérifie facilement que $\overline{G}$ est un groupe de transformations sur Θ.

Remarque 1.3.6. Cette définition implique la relation suivante :

$$P_\theta(\tilde{g}(X) \in A) = P_{\bar{g}\theta}(X \in A), \text{ pour } A \in \mathcal{A}. \tag{1.3.1}$$

La relation (1.3.1) est équivalente à la relation suivante :

$$E_\theta(\Phi(\tilde{g}(X))) = E_{\bar{g}\theta}(\Phi(X)), \ \forall \Phi \text{ mesurable et bornée.}$$

On suppose que L est identifiable c'est-à-dire $\forall d_1, d_2 \in D$, $d_1 \neq d_2$, on a la relation suivante :

$$L(\cdot, d_1) \neq L(\cdot, d_2).$$

Définition 1.3.7. Soit un modèle statistique $\{\mathcal{X}, \mathcal{A}, \Theta, \mathcal{D}, \mathcal{B}, L\}$ invariant sous G. On dit que la fonction de coût est invariante sous G si $\forall d \in \mathcal{D}$, il existe $d' \in \mathcal{D}$, tel que :

$$L(\theta, d) = L(\bar{g}\theta, d'), \forall \theta \in \Theta.$$

Par identifiabilité de L, d' est unique. Ce qui définit g^* par la relation $g^*d = d'$. On vérifie facilement que G^* est un groupe de transformations sur $\mathcal{D}$. L'unicité des décisions d entraîne que la fonction $L(\cdot, d)$ est identifiable.

Dans ce qui suit, on suppose que $\{P_\theta, \theta \in \Theta\}$ et $L(\cdot, d)$ sont identifiables.

Définition 1.3.8. On dit que le modèle décisionnel est invariant sous l'action du groupe G si :

i) le modèle statistique est invariant,

ii) la fonction de perte est invariante.

Par extension G^* opère sur l'espace des règles aléatoires de la façon suivante :

i) si δ est une règle pure $(g^*\delta)(x) = g^*(\delta(x))$,

ii) si δ^* est une règle aléatoire alors $(g^*\delta^*)(x, B) = \delta^*(x, g^*B), \forall B \in \mathcal{B}, x \in \mathcal{D}$.

Cette définition est équivalente à : Si une variable aléatoire $Y \in \mathcal{D}$ a pour loi $\delta^*(x, \cdot)$ alors $\tilde{g}Y$ a pour loi $(g^*\delta^*)(x, \cdot)$.

Théorème 1.3.9. Soit un problème décisionnel invariant $\{\mathcal{X}, \mathcal{A}, \Theta, \mathcal{D}, \mathcal{B}, L\}$.

Soit $\delta \in \Lambda$ alors $R(\theta, \delta) = R(\bar{g}\theta, g^*\delta^*\tilde{g}^{-1}), \forall \theta \in \Theta, g \in G$.

Démonstration. On considère deux cas : le cas d'une règle pure et le cas d'une règle aléatoire.

i) Si δ est une règle pure, on a

$$R(\theta, \delta) = \int_{\mathcal{X}} L(\theta, \delta(x)) P_\theta(dx)$$

$$= \int_{\mathcal{X}} L\left(\bar{g}\theta, g^{*}\delta(x)\right)P_{\theta}(dx) \text{ car } L \text{ est invariant}$$

$$= \int_{\mathcal{X}} L\left(\bar{g}\theta, g^{*}\delta\left(\tilde{g}^{-1}\tilde{g}x\right)\right)P_{\theta}(dx).$$

En faisant un changement de variable x à $\tilde{g}x$, on obtient :

$$R(\theta,\delta) = \int_{\mathcal{X}} L\left(\bar{g}\theta, g^{*}\delta\left(\tilde{g}^{-1}x\right)\right)P_{\theta}\left(d\tilde{g}^{-1}x\right).$$

En vertu de la relation (1.3.1), nous avons :

$$R(\theta,\delta) = \int_{\mathcal{X}} L\left(\bar{g}\theta, g^{*}\delta\tilde{g}^{-1}(x)\right)P_{\bar{g}\theta}(dx).$$

De la relation précédente, il en découle le résultat.

ii) si δ^{*} est une règle aléatoire, on a :

$$R\left(\theta,\delta^{*}\right) = \int_{\mathcal{X}}\int_{\mathcal{D}} L(\theta,y)\,\delta^{*}(x,dy)P_{\theta}(dx)$$

$$= \int_{\mathcal{X}}\int_{\mathcal{D}} L\left(\theta, g^{*^{-1}}g^{*}y\right)\delta^{*}(x,dy)P_{\theta}(dx).$$

En faisant un changement de variable y à $g^{*}y$, on obtient :

$$R\left(\theta,\delta^{*}\right) = \int_{\mathcal{X}}\int_{\mathcal{D}} L\left(\theta, g^{*-1}y\right)\left(g*\delta^{*}\right)(x,dy)P_{\theta}(dx)$$

$$= \int_{\mathcal{X}}\int_{\mathcal{D}} L(\bar{g}\theta,y)\left(g*\delta^{*}\right)(x,dy)P_{\theta}(dx) \text{ car L est invariant,}$$

$$= \int_{\mathcal{X}}\int_{\mathcal{D}} L(\bar{g}\theta,y)\left(g*\delta^{*}\right)\left(\tilde{g}^{-1}\tilde{g}x,dy\right)P_{\theta}(dx).$$

En faisant un autre changement de variable x à $\tilde{g}x$ et en utilisant la relation (1.3.1), on obtient :

$$R(\theta,\delta) = \int_{\mathcal{X}} \int_{\mathcal{D}} L(\overline{g}\theta, y)\left(g * \delta^*\right)\left(\widetilde{g}^{-1}x, dy\right) P_{\overline{g}\theta}(dx)$$

$$= R\left(\overline{g}\theta, g * \delta^* \widetilde{g}^{-1}\right),$$

Ce qui achève la démonstration.

Définition 1.3.10. Une règle aléatoire δ^* est équivariante si :

$$\delta = g * \delta^* \widetilde{g}^{-1}, \forall g \in G. \tag{1.3.2}$$

On note par Λ_E l'ensemble des règles aléatoires équivariantes et Δ_E l'ensemble des règles pures équivariantes.

Corollaire 1.3.11. Pour un problème décisionnel, si une règle aléatoire δ^* est équivariante alors

$$R\left(\theta, \delta^*\right) = R\left(\overline{g}\theta, \delta^*\right), \forall \theta \in \Theta, g \in G.$$

Démonstration. En utilisant la relation (1.3.2) et le théorème 1.3.9, on a le résultat.

Introduisons à présent $\Re_g$ la relation d'équivalence à gauche induite par le groupe de transformations $\overline{G}$ sur les éléments de Θ.

Rappelons que pour $\theta_0 \in \Theta$, l'orbite de θ_0 est égale à $O_{\theta_0} = \{\overline{g}\theta_0 : g \in G\}$.

Remarque 1.3.12. Le risque d'une règle équivariante est constant sur chaque orbite de Θ.

En effet si $\theta \in O_{\theta_0}$ alors $\theta = \overline{g}\theta_0$.

Donc la fonction de risque d'une règle équivariante évaluée en $\theta \in O_{\theta_0}$ est :

$$R(\theta,\delta)=R(\overline{g}\theta_0,\delta)=R(\theta_0,\delta)$$ (d'après le corollaire 1.3.11).

Théorème 1.3.13. Si l'action de $\overline{G}$ sur Θ est transitive alors la fonction de risque d'une règle équivariante est constante sur Θ.

Démonstration. Si $\overline{G}$ est transitif alors $\Theta = O_{\theta_0}$ pour un θ_0 fixé.

En utilisant la remarque précédente, on a $R(\theta_0,\delta)=R(\theta,\delta)$.

Si on est sous les conditions du théorème 1.3.13, la meilleure règle équivariante est celle qui minimise le risque constant.

L'objet des invariants maximaux dans ce chapitre est de déterminer la meilleure règle équivariante en calculant la perte moyenne conditionnelle par rapport à la valeur prise par l'invariant maximal. Le risque sera l'espérance de la moyenne conditionnelle, où l'espérance est prise sur l'invariant maximal.

On suppose que l'action du groupe $\overline{G}$ sur Θ est transitive. Et soit $T(X)$ une statistique invariante maximale sur $\mathcal{X}$.

En utilisant le théorème 1.3.13, on a :

$$R(\theta,\delta)=R(\theta_0,\delta)=E(L(\theta_0,\delta(X)))=E(E(L(\theta_0,\delta(X))\mid T(X))).$$

Lemme 1.3.14. Soit $T(X)$ une statistique invariante maximale sur X. Si l'action de $\overline{G}$ sur Θ est transitive alors la meilleure règle équivariante δ_0 est celle qui minimise pour tout $T(X)$ l'expression :

$$E(L(\theta_0,\delta(X))\mid T(X)=t).$$

Si on se restreint à Δ_E, d'après Perron (1987), une décomposition des règles $\delta \in \Delta_E$ est donnée par la relation suivante :

$$\delta(x) = q(x, \mathrm{v}(T(x))).$$

où T est une statistique invariante maximale, q et T sont des fonctions connues indépendantes de δ et v appartenant à un certain ensemble de fonctions V. La fonction de risque devient alors

$$R(\theta,\delta) = E_{\theta_0}(E(L(\theta_0, q(X, \mathrm{v}(u))) \mid T(X) = u)).$$

Ainsi la recherche de la meilleure règle équivariante revient à déterminer la fonction v_0 appartenant à V telle que :

$$E\{L(\theta_0, q(x, \mathrm{v}_0(u))) \mid T(X) = u\} \leq \inf_{\mathrm{v} \in V} E\{L(\theta_0, q(X, \mathrm{v}(u))) \mid T(X) = u\}.$$

Ceci revient à chercher la fonction $\mathrm{v}_0 \in V$ qui minimise, en tout point $u \in T(x)$, l'expression suivante :

$$E\{L(\theta_0, q(X, \mathrm{v}_0(u))) \mid T(X) = u\}.$$

(Pour plus de détails voir Perron, 1987)

1.4. RÉSULTATS

1.4.1. Liens avec la règle de Bayes.

Dans cette sous section, nous allons établir les résultats concernant les relations entre la meilleure règle équivariante et la règle de Bayes (généralisée) par rapport à une certaine loi à priori λ que nous déterminerons.

Considérons le modèle $\{\mathcal{X},\mathcal{A},\Theta,\mathcal{D},\mathcal{B},L\}$ invariant sous l'action du groupe G. Supposons que $\overline{G}$ opère transitivement et proprement sur Θ. Supposons aussi l'existence d'une mesure de Radon μ relativement invariante avec multiplicateur Δ^{-1} telle que P_θ soit absolument continue par rapport à μ. Dans ces conditions, on cherche la meilleure règle équivariante et ces liens avec la règle de Bayes.

Comme P_θ est absolument continue par rapport à μ, on a :

$$f(x \mid \theta) = \frac{d}{d\mu}(P_\theta(x)).$$

Lemme 1.4.1. Dans les conditions du problème, on a la relation suivante :

$$f(x \mid \theta) = \Delta^{-1} f(\tilde{g}x \mid \overline{g}\theta).$$

Démonstration. Nous avons les égalités suivantes :

$$\begin{aligned}
P_{\overline{g}\theta}(\tilde{g}B) &= \int_{\mathcal{X}} 1_{\tilde{g}B}(x) f(x \mid \overline{g}\theta)\,\mu(dx) \\
&= \int_{\mathcal{X}} 1_B\left(\tilde{g}^{-1}x\right) f(x \mid \overline{g}\theta)\,\mu(dx) \\
&= \int_{\mathcal{X}} 1_B\left(\tilde{g}^{-1}x\right) f\left(\tilde{g}\tilde{g}^{-1}x \mid \overline{g}\theta\right)\mu(dx) \\
&= \int_{\mathcal{X}} 1_B\left(\tilde{g}^{-1}x\right) f(\tilde{g}x \mid \overline{g}\theta)\,\mu(d\tilde{g}x) \text{ en remplaçant } \tilde{g}^{-1}x \text{ par } x \\
&= \int_{\mathcal{X}} 1_B(x) f(\tilde{g}x \mid \overline{g}\theta)\,\Delta^{-1}(g)\,\mu(dx).
\end{aligned} \tag{1.4.1}$$

Nous avons aussi la relation suivante :

$$P_\theta(B) = \int_{\mathcal{X}} 1_B(x) f(x \mid \theta)\, \mu(dx). \tag{1.4.2}$$

Comme le problème est invariant, on a la relation :

$$P_{\bar{g}\theta}(\tilde{g}B) = P_\theta(B), \forall g \in G, \theta \in \Theta \text{ et } \mathrm{B} \in A. \tag{1.4.3}$$

En réunissant les relations (1.4.1), (1.4.2) et (1.4.3), on obtient l'égalité suivante :

$$\int_{\mathcal{X}} 1_B(x) \Delta^{-1}(g) f(\tilde{g}x \mid \bar{g}\theta)\, \mu(dx) = \int_{\mathcal{X}} 1_B(x) f(x \mid \theta)\, \mu(dx).$$

Par une propriété de l'intégrale, nous avons le résultat suivant :

$$\forall g \in G, \mathrm{x} \in B,\ f(x \mid \theta) = \Delta^{-1}(g) f(\tilde{g}x \mid \bar{g}\theta)\ \mu\text{-presque partout.}$$

On définit une fonction T de $K(\mathcal{X}) \to K(\mathcal{X} \mid G)$ par l'égalité suivante :

$$T(f)(\pi(x)) = \int_G f(\tilde{g}x)\, \nu_R(dg),$$

où f est une fonction μ-intégrable, ν_R la mesure de Haar à droite sur G et π la projection de $\mathcal{X}$ sur l'espace des orbites $\mathcal{X} \mid G$.

Théorème 1.4.2. Soit G opérant proprement sur $\mathcal{X}$. Et soit J une intégrale relativement invariante sur $K(\mathcal{X})$ avec multiplicateur Δ^{-1}. Alors il existe une intégrale J_1 sur $K(\mathcal{X} \mid G)$ tel que :

$$J(f) = J_1(T(f)), f \in K(\mathcal{X}).$$

Démonstration. Voir Eaton (1989, page 77).

Théorème 1.4.3. Dans les conditions du problème, la fonction de risque est donnée par la relation suivante :

$$R(\theta,\delta^*) = J_1(T(f_0)),\text{ où } f_0 = \int_{\mathcal{D}} L(y,\theta)\delta^*(dy,x)f(x\mid\theta).$$

Démonstration. La fonction de risque d'une règle δ évaluée en θ est donnée par :

$$R(\theta,\delta) = \int_{\mathcal{X}}\int_{\mathcal{D}} L(y,\theta)\delta^*(x,dy)f(x\mid\theta)\mu(dx) = \int_{\mathcal{X}} f_0(x)\mu(dx) = J(f_0). \quad (1.4.4)$$

Montrons d'abord que J est relativement invariante avec multiplicateur Δ^{-1}.

Comme la mesure μ étant relativement invariante avec multiplicateur Δ^{-1} alors on a le résultat suivant :

$$J(L_g f) = \int_{\mathcal{X}} f(\tilde{g}x)\,\mu(dx) = \int_{\mathcal{X}} f(x)\,\mu(d\tilde{g}^{-1}x) = \Delta^{-1}(g)\int_{\mathcal{X}} f(x)\,\mu(dx) = \Delta^{-1}(g)J(f).$$

Mais en utilisant le Théorème 1.4.2 et la relation (1.4.4), nous avons le résultat.

Pour trouver la meilleure règle équivariante, on considère la fonction H définie sur $\mathcal{D} \times \mathcal{X}$ par

$$H(d,x) = \int_G L(\bar{g}\theta,d)f(x\mid\bar{g}\theta)\nu_R(dg).$$

Cette fonction ne dépend pas du paramètre $\theta \in \Theta$ car $\overline{G}$ opère transitivement sur l'espace Θ.

Théorème 1.4.4. S'il existe une fonction mesurable $a_0 : \mathcal{X} \to \mathcal{D}$ telle que :

i) $H(d,x) \geq H(a_0(x),x), \forall d \in \mathcal{D} \text{ et } x \in \mathcal{X},$

ii) $a_0(\tilde{g}x) = g^* a_0(x), \forall g \in G \text{ et } x \in \mathcal{X},$

alors $\delta_0 = a_0$ définit la meilleure règle de décision équivariante (δ_0 est une règle pure).

Démonstration. Soit $\delta^* \in \Lambda$ et $\theta \in \Theta$, d'après le i), on a :

$$\int_G L(\bar{g}\theta, \delta^*(x)) f(x \mid \bar{g}\theta) \nu_R(dg) \geq \int_G L(\bar{g}\theta, a_0(x)) f(x \mid \bar{g}\theta) \nu_R(dg).$$

Donc on a l'inégalité suivante :

$$\int_G \int_{\mathcal{D}} L(\bar{g}\theta, y) \delta^*(x, dy) f(x \mid \bar{g}\theta) \nu_R(dg) \geq \int_G L(\bar{g}\theta, a_0(x)) f(x \mid \bar{g}\theta) \nu_R(dg).$$

Mais on sait que :

$$\int_G L(\bar{g}\theta, a_0(x)) f(x \mid \bar{g}\theta) \nu_R(dg) = \int_G \int_{\mathcal{D}} L(\bar{g}\theta, y) \delta_0^*(x, dy) f(x \mid \bar{g}\theta) \nu_R(dg).$$

En appliquant l'intégrale sur les deux membres par rapport à J_1 de la relation précédente et en utilisant le théorème 1.4.3, on a :

$$R(\theta, \delta^*) \geq R(\theta, \delta_0^*), \forall \theta \in \Theta.$$

La meilleure règle équivariante est celle qui, pour chaque $x \in \mathcal{X}$, minimise la fonction $H(\cdot, x)$.

Étudions à présent la relation qui existe entre cette meilleure règle équivariante et la règle de Bayes. Pour cela nous considérons, pour $\theta_0 \in \Theta$ fixé, la fonction suivante :

$$\tau_{\theta_0} : G \to \Theta$$

$$g \to g\theta_0.$$

La fonction τ_{θ_0} définit une surjection de G sur Θ du fait de la transitivité de $\overline{G}$ sur Θ. Elle induit une mesure λ sur Θ, définie par :

$$\lambda(B) = \nu_R\left(\tau_{\theta_0}^{-1}(B)\right), \forall\, B \in \sigma(\Theta).$$

Cette relation est équivalente à :

$$\int_\Theta f(\theta)\lambda(d\theta) = \int_G f(g\theta_0)\nu_R(dg), \forall f \in K(\Theta). \tag{1.4.5}$$

Si on suppose que λ est une mesure de Radon. Nous pouvons affirmer que λ est relativement invariante avec multiplicateur Δ^{-1} sous l'action de $\overline{G}$. En utilisant la relation (1.4.5), on a :

$$H(a,x) = \int_\Theta L(a,\theta)f(x \mid \theta)\lambda(d\theta).$$

Mais l'expression $\int_\Theta L(a,\theta)f(x \mid \theta)\lambda(d\theta)$ est proportionnelle au risque *a posteriori* associé à λ. D'où $\delta_0 = a_0$ est une règle de Bayes (généralisée) associée à la mesure λ.

Si G est compact, en utilisant le théorème de Weil (1.2.21), la mesure λ est une mesure de probabilité finie. Donc l'expression $\int_\Theta L(a,\theta)f(x \mid \theta)\lambda(d\theta)$ est proportionnelle au risque *a posteriori*. Ce qui nous conduit à une règle de Bayes.

1.4.2. Liens avec la minimaxité.

On introduit les notations et les définitions pour pouvoir énoncer le théorème de Hunt-stein généralisé. Pour plus de détails sur les définitions, voir Strasser (1985). On dispose d'un modèle décisionnel $\{\mathcal{X}, \mathcal{A}, \Theta, \mathcal{D}, \mathcal{B}, L\}$ tel que, pour tout $\theta \in \Theta$, $L(\theta,\cdot)$ est une fonction continue sur $\mathcal{D}$.

Soit $ca(\mathcal{X},\mathcal{A})$ l'ensemble des mesures signées et bornées sur $(\mathcal{X},\mathcal{A})$ et $C_b(\mathcal{D})$ l'ensemble des fonctions continues et bornées et soit un sous espace de $ca(\mathcal{X},\mathcal{A})$ donné par l'expression suivante :

$$L(\chi)=\{\mu \in ca(\mathcal{X},\mathcal{A}) : \sigma \perp \mu \text{ si } \sigma \perp P_\theta, \forall \theta \in \Theta\}.$$

Lemme 1.4.5. Si P est dominé par une mesure σ-finie v et si P équivalent à v, alors $L(\chi)=\{\mu \in ca(\mathcal{X},\mathcal{A}) : \mu << v\}$.

Démonstration. Voir Strasser (1985, page 227).

Définition 1.4.6. On appelle fonction de décision généralisée de $\mathcal{X}$ et $\mathcal{D}$, la fonction bilinéaire β sur $C_b(\mathcal{D}) \text{x } L(\mathcal{X})$ à valeurs dans IR satisfaisant les conditions suivantes :

1) $|\beta(f,\mu)| \leq \|f\|\|\mu\|$, pour $f \in C_b(\mathcal{D}), \mu \in L(\chi)$,

2) $\beta(f,\mu) \geq 0$, pour $f \geq 0$ et $\mu \geq 0$,

3) $\beta(1,\mu)=\mu(\mathcal{X}), \mu \in L(\mathcal{X})$.

On note $\beta(\mathcal{X},\mathcal{D})$, l'ensemble des fonctions de décisions généralisées.

Soit $\delta \in \Lambda$. La fonction bilinéaire définie par :

$$\beta_\delta : C_b(\mathcal{D}) \text{x } L(\mathcal{X}) \to \text{IR}$$

$$(f,\mu) \to \iint f(a)\delta(x,da)\mu(dx),$$

est une fonction de décisions généralisée.

Outre la fonction de risque fréquentiste définie dans la sous section 1.3.1 sur $\Delta \times \Theta$, on peut donner une définition plus générale d'une fonction de risque de $\beta \in \beta(\mathcal{X},\mathcal{D})$.

Définition 1.4.7. Le risque de $\beta \in \beta(\mathcal{X},\mathcal{D})$ est $\beta(L(\theta,\cdot),P_\theta)$ et le risque pour $\delta \in \Lambda$ est $\beta_\delta(L(\theta,\cdot),P_\theta)$ égal encore à $R(\theta,\delta)$. On pose

$$R(\mathcal{X},\mathcal{D}) = \{(\beta(L(\theta,\cdot),P_\theta))_{\theta\in\Theta}, \beta \in \beta(\mathcal{X},\mathcal{D})\}$$

et

$$R_0(\mathcal{X},\mathcal{D}) = \{(\beta_\delta(L(\theta,\cdot),P_\theta))_{\theta\in\Theta}, \beta \in \beta(\mathcal{X},\mathcal{D})\}.$$

Définition 1.4.8. Soit $(G,\mathcal{B}(G))$ un groupe localement compact et à base d'ouverts dénombrable où $\mathcal{B}(G)$ est le σ-algèbre sur G.

- Une fonction linéaire positive $M : L(\mathcal{X}) \to L(\mathcal{X})$ est un opérateur stochastique si $\|M\sigma\| = \sigma, \forall \sigma \in L(\mathcal{X}), \sigma \geq 0$.

- Une famille $(M_g)_{g\in G}$ d'opérateurs stochastiques est un groupe si on a : $M_{g_1 \circ g_2} = M_{g_1} \circ M_{g_2}, \forall g_1, g_2 \in G$ et $M_e = id_{L(\chi)}$ où e désigne l'élément neutre du groupe G.

Définition 1.4.9. Soit $(M_g)_{g\in G}$ un groupe d'opérateurs stochastiques sur $L(\mathcal{X})$, on dit que $\mathcal{X}$ est invariant sous $(M_g)_{g\in G}$ si $M_g P_\theta \in \mathcal{P}, \forall g \in G$ et $\theta \in \Theta$.

Exemple 1.4.10. Soit le modèle décisionnel $\{\mathcal{X},\mathcal{A},\Theta,\mathcal{D},\mathcal{B},L\}$ tel que P soit dominé par une mesure σ-finie. Supposons que le modèle soit invariant sous le groupe de transformations G. Pour tout $g \in G$, on pose

$$M_g \mu = \mu \circ g^{-1}, \forall \mu \in L(\mathcal{X}).$$

Comme le modèle est invariant sous G, on a $P_\theta \circ g^{-1} \in \mathcal{P}, \forall \theta \in \Theta$ et $g \in G$. Donc en utilisant ce résultat et le lemme 1.4.5, on a :

- $M_g \mu \in L(\mathcal{X}), \forall \mu \in L(\mathcal{X})$,

- $M_{g_1} \circ M_{g_2} = M_{g_1 \circ g_2}$,

en effet $M_{g_1} \circ M_{g_2} \mu = M_{g_1}\left(\mu \circ g_2^{-1}\right) = \left(\mu \circ g_2^{-1}\right) \circ g_1^{-1} = M_{g_1 \circ g_2} \mu$.

- $M_e = id_{L(\mathcal{X})}$. En effet $M_e \mu = \mu = id_{L(\mathcal{X})} \mu$

D'où $(M_g)_{g \in G}$ est un groupe d'opérateurs stochastiques et il est évident que $\mathcal{X}$ est invariant sous $(M_g)_{g \in G}$.

Théorème 1.4.11. Supposons qu'il existe une mesure μ σ-finie telle que $\forall \theta \in \Theta, P_\theta << \mu$. Supposons que $\mathcal{D}$ soit localement compact avec base d'ouverts dénombrable. Soit $\beta \in \beta(\mathcal{X}, \mathcal{D})$ telle que la fonction $f \to \beta(\mathcal{X}, \mathcal{D})$ est intégrable selon l'approche de Daniell, sur $C_b(\mathcal{D}), \forall \theta \in \Theta$. Alors il existe $\delta^* \in \Lambda$ tel que :

$$\beta(f, \mu) = \iint f(a) \delta^*(x, da) \mu(dx), \forall f \in C_b(\mathcal{D}), \mu \in L(\mathcal{X}).$$

Démonstration. Voir Strasser (1985, page 233).

Théorème 1.4.12. (Farell, 1967) Supposons qu'il existe une mesure μ σ-finie telle que : $\forall \theta \in \Theta, P_\theta << \mu$. Supposons que D soit localement compact avec base d'ouverts dénombrable. Alors pour toute fonction $\beta \in \beta(\mathcal{X}, \mathcal{D})$, il existe $\delta \in \Lambda$ telle que :

$$\beta(f,\mu) \geq \iint f(a)\delta^*(x,da)\mu(dx), \mu \in L^+(\mathcal{X}),$$

pour toute fonction $f : D \to \mathrm{IR}$ compacte.

Démonstration. Voir Strasser (1985, page 233).

Définition 1.4.13. Une fonction de décision généralisée est strictement équivariante si on a l'égalité suivante :

$$\beta(f \circ g, \mu) = \beta(f, M_g \mu), \forall f \in C_b(\mathcal{D}), \mu \in L(\mathcal{X}), g \in G.$$

Théorème 1.4.14. (Hunt-Stein généralisé)

Soit le modèle $\{\mathcal{X}, \mathcal{A}, \Theta, \mathcal{D}, \mathcal{B}, L\}$ invariant sous le groupe $(M_g)_{g\in G}$. Supposons que G opère à gauche continûment sur D. Si le groupe $(M_g)_{g\in G}$ a la propriété du point fixe alors, $\forall \beta \in \beta(\mathcal{X}, \mathcal{D})$, il existe une fonction $\beta_0 \in \beta(\mathcal{X}, \mathcal{D})$ strictement équivariante telle que :

$$\sup_{g\in G} \beta(f \circ g^{-1}, M_g \mu) \geq \beta_0(f,\mu), \forall \mu \in L^+(\mathcal{X}), f \in C_b(\mathcal{D}).$$

Démonstration. Voir Strasser (1985, page 251).

Corollaire 1.4.15. Si on est dans les conditions du théorème alors pour toute fonction $\beta \in \beta(\mathcal{X}, \mathcal{D})$, il existe $\beta_0 \in \beta(\mathcal{X}, \mathcal{D})$ telle que :

$$\sup_{\theta\in\Theta_x} \beta(f_\theta, P_\theta) \geq \sup_{\theta\in\Theta_x} \beta_0(f_\theta, P_\theta), \forall x \in \mathcal{X},$$

et $\forall \theta \in \Theta$, $f_\theta : \mathcal{D} \to IR$ est une fonction continue inférieurement.

Démonstration. Voir Strasser (1985, page 253).

Bondar and Miles (1981) montrent que si le groupe G est localement compact avec base d'ouverts dénombrable alors la moyennabilité implique la propriété du point fixe. Et de plus, si G est connexe alors on a équivalence entre ces deux hypothèses.

Dans ce qui suit, on considère que le modèle décisionnel est invariant sous le groupe topologique G et que $\mathcal{P}$ est dominé par une mesure σ-finie sur $\mathcal{X}$.

On avait vu dans la sous-section 1.3.2 que G induit des groupes de transformations sur $\mathcal{X}$, $\Theta, \mathcal{D}$.

Théorème 1.4.16. Supposons que G soit connexe et localement compact avec base d'ouverts dénombrable. Supposons aussi $\mathcal{D}$ localement compact avec base dénombrable. Si G est moyennable alors, $\forall \beta \in \beta(\mathcal{X},\mathcal{D})$, il existe une fonction de décision $\beta_0 \in \beta(\mathcal{X},\mathcal{D})$ strictement équivariante telle que :

$$\sup_{\theta \in \Theta_x} \beta(f_\theta, P_\theta) \geq \sup_{\theta \in \Theta_x} \beta_0(f_\theta, P_\theta), \forall x \in \mathcal{X},$$

et pour toute fonction $f : \mathcal{D} \to IR$ continue inférieurement.

Démonstration. On pose $M_g \mu = \mu \circ g^{-1}$, $\forall \mu \in L(\mathcal{X})$. On a vu dans l'exemple 1.4.10 que $(M_g)_{g \in G}$ est un groupe d'opérateurs stochastiques, que $\mathcal{X}$ est invariant sous $(M_g)_{g \in G}$ et G opère continûment à gauche sur D.

Il en découle, en utilisant la remarque précédente, que $(M_g)_{g \in G}$ possède la propriété du point fixe.

On est donc dans les conditions du théorème de Hunt-Stein généralisé, ce qui donne le résultat.

Théorème 1.4.17. Si en plus des hypothèses du théorème précédent, la fonction de perte $L(\theta,d)$ vérifie pour tout $\theta \in \Theta$, $\{d \in \mathcal{D} : L(\theta,d) \geq \tau\}$ est compact pour tout $\tau \geq 0$. Alors pour toute règle aléatoire δ^*, il existe $\delta_0^* \in \Lambda_E$ telle que :

$$\sup_{\theta \in \Theta_x} R(\theta,\delta^*) \geq \sup_{\theta \in \Theta_x} R(\theta,\delta_0^*), \ \forall x \in \chi .$$

Démonstration. C'est une conséquence immédiate des théorèmes 1.4.11, 1.4.12 et 1.4.16.

Remarque 1.4.18. Notons Λ_{SE} la sous classe de Λ des règles strictement équivariantes. Sous les conditions du théorème précédent, Λ_{SE} est une classe complète.

Considérons à présent le problème d'un test d'hypothèses. On veut préciser les conditions sous lesquelles le meilleur test invariant est aussi optimal par rapport à tous les tests. La réponse à ce problème est encore basée sur le théorème de Hunt-Stein généralisé appliqué aux tests d'hypothèses. Initialement le théorème de Hunt-Stein a été établi pour les tests d'hypothèses. Les propriétés et définitions concernant les tests d'hypothèses ne sont pas exposées dans ce travail. Pour plus de détails, on peut consulter Giri (1975) et Lehmann (1986).

Soit H_0 l'hypothèse nulle et H_1 l'hypothèse alternative.

On note Θ_{H_i}, l'ensemble des $\theta \in \Theta$ vérifiant $H_i, i = 0,1$.

Soit le modèle statistique $(\mathcal{X},\mathcal{A},\Theta)$ invariant sous l'action d'un groupe de transformations $\overline{G}$ induit par un groupe topologique G. Supposons qu'il existe une mesure μ σ-finie telle que : $P_\theta << \mu, \forall\theta \in \Theta$.

Lorsqu'on applique le théorème de Hunt-Stein généralisé aux tests d'hypothèses, ce théorème porte simplement le nom de théorème de Hunt-Stein. Cela a été le point de départ de la généralisation.

Théorème 1.4.19. (Hunt-Stein). Soit $\sigma(G)$ la σ-algèbre associée à G,

– pour tout $B \in \mathcal{A}$, l'ensemble des couples (x, g) avec $\tilde{g}x \in B$ est dans $\mathcal{A}$ x$\sigma(G)$,

– pour tout $H \in \sigma(G)$, $g \in G$, l'ensemble des Hg est dans $\sigma(G)$.

Alors si G est moyennable, pour tout test φ il existe un test ψ presque invariant tel que :

$$\sup_{\theta \in \Theta_{H_0}} E_\theta \varphi \geq \sup_{\theta \in \Theta_{H_0}} E_\theta \psi,$$

$$\inf_{\theta \in \Theta_{H_1}} E_\theta \varphi \leq \inf_{\theta \in \Theta_{H_1}} E_\theta \psi.$$

Démonstration. La démonstration de ce théorème est donnée par plusieurs auteurs dont Wesler (1959), Giri (1975) et Lehmann (1986).

Voici deux exemples traités par Lehmann (1986), dont le premier ne vérifie pas l'une des hypothèses de Hunt-Stein et le second vérifie toutes les hypothèses de Hunt-Stein, que nous allons reprendre dans ce mémoire.

Le premier exemple est celui d'un problème invariant sous l'action d'un groupe G ne vérifiant pas l'hypothèse de moyennabilité du théorème de Hunt-Stein. Dans ce cas, on montre que le test invariant uniformément le plus puissant est inadmissible.

Exemple 1.4.20. Soit $X = (X_1, X_2, \ldots, X_n)^t$ et $Y = (Y_1, Y_2, \ldots, Y_n)^t$ deux réalisations indépendantes de lois multinormales de moyennes zéro et de matrices de variances covariances $E(X^t X) = \Sigma$, $E(Y^t Y) = \Delta\,\Sigma$, où Δ est un scalaire inconnu et Σ est une matrice définie positive inconnue.

On veut tester l'hypothèse $(H_0): \Delta \leq \Delta_0$ contre l'alternative $(H_1): \Delta \geq \Delta_1$ avec $\Delta_1 > \Delta_0$.

Ce problème devient invariant sous le groupe de transformations $GL(p)$ opérant comme suit :

$$(X, Y, \Sigma) \to (gX, gY, g\Sigma g^t), \forall g \in GL(p).$$

Puisque le groupe $GL(p)$ est transitif avec probabilité un sur l'espace échantillonnal, le test uniformément le plus puissant de niveau α est $\varphi(x, y) \equiv \alpha$. Ainsi, le maximum de la puissance minimale sur l'hypothèse (H_1) parmi les tests invariant est α.

Par contre le niveau du test α, avec la région de rejet $\frac{X_i^2}{Y_i^2} > C, \forall i$, a une fonction de puissance $\beta(\Delta)$ strictement croissante en Δ et son maximum de puissance minimal sur l'ensemble $\Delta \geq \Delta_1$ est $\beta(\Delta_1) > \beta(\Delta_0) = \alpha_0$.

Exemple 1.4.21. Soit Y_i, $i = 1,\ldots,n$ des variables aléatoires indépendants et identiquement distribuées normale $N(\eta_i, \sigma^2)$ avec $\eta_{s+1} = \eta_{s+2} = \ldots = \eta_n = 0$ pour $(s < n)$.

On veut tester l'hypothèse $(H_0): \eta_1 = \eta_2 = \ldots = \eta_r = 0$ pour $(r \leq s < n)$.

Le problème est invariant sous les groupes :

- G_1 des translations transformant $Y_i \to \begin{cases} Y_i + c_i,\ i = r+1,\dots,s \\ Y_i,\ i = 1,\dots,r; s+1,\dots,n \end{cases}$,

- G_2 des transformations orthogonales,

- G_3 des scalaires transformant $Y_i \to cY_i, i = 1,\dots,n$ et $c \neq 0$.

Chacun des trois groupes vérifie les hypothèses de Hunt-Stein. (Lehmann, 1986) montre que le maximal invariant de l'espace échantillonnal sous G_1 ou G_2 ou G_3 est :

$$W = \frac{\sum_{i=1}^{r} Y_i^2}{\sum_{i=s+1}^{n} Y_i^2}.$$

Et le maximal invariant de l'espace des paramètres sous $\overline{G}_1$ ou $\overline{G}_2$ ou $\overline{G}_3$, est :

$$\psi^2 = \frac{\sum_{i=1}^{r} \eta_i^2}{\sigma^2}.$$

L'hypothèse $(H_0): \eta_1 = \eta_2 = \dots = \eta_r = 0$ est équivalent à l'hypothèse $(H_0^{'}): \psi = 0$.

En utilisant le lemme de Neyman-Pearson, le test uniformément le plus puissant sous $(H_0^{'}): \psi = 0$ contre $(H_1^{'}): \psi \neq 0$ est de rejeter $(H_0^{'})$ lorsque $\frac{n-s}{r} W$ est grand.

C'est-à-dire qu'on rejette $(H_0^{'})$ au niveau α lorsqu'on a :

$$\frac{n-s}{r}W = \frac{\frac{\sum_{i=1}^{r} Y_i^2}{r}}{\frac{\sum_{i=s+1}^{n} Y_i^2}{n-s}} > K \text{ où } P\left(\frac{n-s}{r}W > K\right) = \alpha .$$

Puisque la variable aléatoire $\frac{n-s}{r}W$ admet une distribution de Fisher $F(r,n-s)$, la valeur critique $K = F_{r,n-s}(\alpha)$ est le quantile d'ordre α de $F(r,n-s)$.

Comme les groupes vérifient les hypothèses de Hunt-Stein, ils demeurent invariants pour le problème de maximiser la puissance minimum sur l'hypothèse alternative. Plus généralement, le test qui est uniformément le plus puissant sous le groupe G_1 ou G_2 ou G_3, pour tester $(H_0^{'})$, maximise la puissance minimale sous l'alternative.

CHAPITRE 2

ESTIMATION ÉQUIVARIANTE EN MODÈLE RESTREINT

Dans ce chapitre, nous reprenons l'approche de Kariya (1989) pour caractériser les estimateurs équivariants dans le modèle admettant une statistique libre. Ensuite nous utiliserons les résultats obtenus pour l'application de quelques problèmes intéressants d'estimations tels :

- l'estimation de la moyenne μ d'un mélange de lois multinormales de paramètres (μ, Ω, Z), où Z est une variable aléatoire positive connue, sous la contrainte :

$$\mu^t \Sigma^{-1} \mu = \lambda^2, \text{ avec } \Sigma = E(Z)\Omega \text{ et } \lambda^2 \text{ connu.}$$

La définition d'un mélange de lois normales est donnée dans ce chapitre.

- l'estimation de la moyenne μ d'un mélange de lois normales de paramètres $\left(\mu, \dfrac{\mu^t \mu}{E(Z)C^2} I, Z\right)$, où Z est une variable aléatoire positive connue et C^2 est connu.

Dans chacune des situations, on étudiera le cas particulier d'une loi normale multivariée, ensuite le cas général d'un mélange de lois normales multivariées et enfin on appliquera les résultats au cas d'une loi de "Student" multivariée. Et pour chaque étude, on dégagera le meilleur estimateur équivariant et le risque minimal. Le comportement du meilleur estimateur équivariant pour le cas normal, comme pour le cas de "Student" multivariée, sera également étudié.

2.1. ESTIMATION DANS UN MODÈLE AVEC RESTRICTION

Soit $\{\mathcal{X}, \mathrm{A}, \mathcal{P}, \Theta^*, \mathcal{D}, \mathcal{B}, L\}$ un modèle décisionnel où Θ^* est un ensemble ouvert sur IR^p et G est un groupe de transformations topologique opérant mesurablement à gauche sur $\mathcal{X}$.

On suppose que $\mathcal{X}$ est l'espace des statistiques minimales exhaustives, que le modèle décisionnel $\{\mathcal{X}, \mathrm{A}, \mathcal{P}, \Theta^*, \mathcal{D}, \mathcal{B}, L\}$ est invariant et que la famille $\mathcal{P}$ est identifiable.

Soit $\Theta = \{\theta \in \Theta^* : \theta = \Psi(\eta), \eta \in \Upsilon\}$ où Υ est un sous-ensemble ouvert de IR^q $(q < p)$ et Ψ est une fonction mesurable bijective de Υ dans l'espace paramétrique $\Psi(\Upsilon) \subset \Theta^*$.

On note par $\overline{G}$ le groupe de transformations induit par G opérant à gauche sur Θ^*. Introduisons $\lambda(\theta)$ un invariant maximal des paramètres sous $\overline{G}$.

Posons l'espace paramétrique $\Theta_{\lambda_0} = \{\theta \in \Theta^* : \lambda(\theta) = \lambda_0\}$ où λ_0 est connu.

Remarque 2.1.1. On définit un groupe opérant sur Υ par :

$G^{**} = \{g^{**} : g^{**} = \Psi^{-1} g \Psi, \forall g \in G\}$. L'action de G^{**} sur Υ est transitive.

Pour plus de détails voir Kariya (1989).

Définition 2.1.2. La sous famille $\{P_\theta\}_{\theta\in\Theta}$ est appelée famille restreinte.

De plus, elle est invariante sous le groupe de transformation G^{**} sur Υ.

Un invariant maximal devient une statistique libre dans le modèle restreint.

On cherche à décomposer $\mathcal{X}$ comme produit cartésien de deux espaces afin de caractériser les estimateurs équivariants. Supposons l'existence de la décomposition suivante :

$$\pi : \mathcal{X} \to G \times V$$

$$x \to (h(x), u(x)),$$

où V est un espace mesurable.

Cette décomposition satisfait les trois conditions suivantes :

- la fonction $x \to (h(x), u(x))$ est une bijection bimesurable,

- la fonction $h : \mathcal{X} \to G$ est une fonction mesurable et équivariante c'est-à-dire :

$$h(gx) = gh(x), \forall g \in G,$$

- la fonction $u : \mathcal{X} \to V$ est invariante maximale.

Problème : Sous ces conditions, on veut estimer θ avec la fonction de perte $L(\theta,d)$.

On utilise le fait que π est bijective. On a, pour $\delta \in \Delta_I$, la relation suivante :

$$\delta(x) = \delta\left(\pi^{-1}(h(x), u(x))\right), \forall\ x \in \mathcal{X}\ . \tag{2.1.1}$$

Théorème 2.1.3. (caractérisation des estimateurs équivariants). Un estimateur δ est équivariant si et seulement s'il satisfait la relation :

$$\delta(x) = h^*(x)k(u(x)), \ \forall x \in \mathcal{X}$$

pour une certaine fonction k mesurable sur V.

De plus, on vérifie l'expression suivante :

$$k(u(x)) = \delta(\pi^{-1}(e, u(x))), \ \forall x \in \mathcal{X},$$

où e désigne l'élément neutre du groupe G.

Démonstration. Soit un estimateur δ défini par $\delta(x) = h^*(x)k(u(x)), \ \forall x \in \mathcal{X}$ et k une fonction mesurable sur V.

Puisque la fonction h est équivariante et la fonction u est invariante maximale, nous avons :

$$\delta(gx) = h^*(gx)k(u(gx)) = g^* h(x)k(u(x)) = \delta(x), \ \forall x \in \mathcal{X}.$$

Donc δ est équivariant.

Soit un estimateur équivariant $\delta \in \Delta_E$ vérifiant la relation suivante :

$$k(u(x)) = \delta(\pi^{-1}(e, u(x))), \ \forall x \in \mathcal{X}.$$

Comme $\delta \in \Delta_E$ alors nous avons la relation suivante :

$$\delta(gx) = g^* \delta(x). \tag{2.1.2}$$

Mais en utilisant (2.1.1), on obtient aussi la relation suivante :

$$\delta(gx) = \delta o \pi^{-1}(g^* h(x), u(x)). \tag{2.1.3}$$

En réunissant (2.1.2) et (2.1.3), on obtient alors :

$$\delta \circ \pi^{-1}(g^* h(x), u(x)) = g^* \delta \circ \pi^{-1}(h(x), u(x)).$$

Comme $h(x) \in G$ alors on a la relation suivante :

$$\delta(x) = \delta \circ \pi^{-1}(h(x), u(x)) = h^*(x)\delta \circ \pi^{-1}(e, u(x)).$$

En posant $k(u(x)) = \delta(\pi^{-1}(e, u(x)))$, on a le résultat.

Une conséquence de ce théorème est l'équivalence entre la recherche d'estimateurs équivariants et la recherche de fonctions mesurables sur U.

Théorème 2.1.4. Sous les conditions du problème, le meilleur estimateur équivariant (s'il existe) est donné par :

$$\hat{\delta}(x) = h^*(x)\hat{k}(u(x)),$$

où $\hat{k}(u(x))$ minimise la fonction $E((e, I); L(h(x)k(u(x))) \mid \mathrm{u}(\mathrm{x}))$.

Démonstration. En vertu du lemme 1.3.14. du chapitre 1, la meilleure règle équivariante (si elle existe) minimise la fonction $E[L\{(e, I); h(x)k(u(x))\} \mid u(x)]$ où $e^t = (1, 0, \ldots, 0)$ et I est la matrice identité. Ce qui revient à chercher une fonction k qui minimise l'expression $E[L\{(e, I); \delta(x)\} \mid u(\mathrm{x})]$.

Dans tout ce qui suit, on adoptera les notations suivantes :

- $N_p(\mu, \Sigma)$, la loi normale de dimension p de paramètres μ et Σ;

- $\mathscr{B}e(a, b)$, la loi bêta de paramètres a et b;

- $\mathscr{G}(a, b)$, la loi gamma de paramètres a et b dont la densité est :

$$f(x \mid a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} 1_{[0,\infty]}(x);$$

- $W_p(n, \Sigma)$, la loi Wishart de dimension p de paramètres n et Σ;

– χ_n^2, la loi de khi-deux à n degrés de liberté ;

– $F_{n,m}(\alpha)$, la loi de Fisher décentrée à n, m degrés de liberté et α est le paramètre de décentrage.

2.2. ESTIMATION DE LA MOYENNE μ D'UNE DISTRIBUTION DE MÉLANGE DE LOIS MULTINORMALES SOUS LA CONTRAINTE $\mu^t\Sigma^{-1}\mu = \lambda^2$, AVEC λ^2 CONNU.

Soient $X_1, X_2, ..., X_n$ des variables aléatoires de loi multinormale avec moyenne μ et matrice variance-covariance Σ. Considérons la fonction de perte donnée par :

$$L((\mu, \Sigma), d) = (\mu - d)^t\Sigma^{-1}(\mu - d). \qquad (2.2.1)$$

On pose, $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{\text{n}}$ et $S = \sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^t$.

On veut d'abord estimer la moyenne μ d'une distribution multinormale de paramètres μ et Σ, avec la fonction de perte (2.2.1) lorsque :

$$\mu^t\Sigma^{-1}\mu = \lambda^2 \text{ avec } \lambda^2 \text{ connu.} \qquad (2.2.2)$$

Ensuite on traite le cas le plus général d'une estimation de la moyenne μ d'un mélange de lois normales avec la fonction de perte (2.2.1) sous la contrainte (2.2.2).

Considérons l'espace des paramètres :

$$\Theta_\lambda = \{(\mu, \Sigma) \in IR^p x S(p) : \mu^t\Sigma^{-1}\mu = \lambda^2\},$$

où $S(p)$ désigne l'ensemble des matrices carrées définies positives.

Considérons le groupe de transformations $GL(p)$ des matrices carrées $(p \text{ x } p)$ régulières transformant :

$$(\overline{X}, S) \rightarrow (g\overline{X}, gSg'), \;\; g \in GL(p).$$

Ce problème devient invariant sous la fonction de perte (2.2.1). Le groupe $\overline{GL}(p)$ induit par $GL(p)$ dans l'espace des paramètres transforme :

$$\theta = (\mu, \Sigma) \rightarrow (g\mu, g\Sigma g').$$

Le groupe $\overline{GL}(p)$ opère transitivement sur Θ_λ (car Θ_λ est constitué de tous les

$\theta = (\mu, \Sigma) \in IR^p \text{x} \, S(p)$ tels que : $\mu'\Sigma^{-1}\mu = \lambda^2$).

La fonction de risque d'un estimateur $\hat{\mu}$ de μ est constante sur Θ_λ. Elle est égale à l'expression suivante :

$$R(\theta, \hat{\mu}) = E\{(\mu - \hat{\mu}(X))' \Sigma^{-1} (\mu - \hat{\mu}(X))\}, \; \forall \, \theta = (\mu, \Sigma) \in \Theta_\lambda.$$

La statistique $U = Y'S^{-1}Y$ est invariante maximale. Elle est basée sur (Y, S) où $Y = \sqrt{n}\,\overline{X}$.

Donc U est une statistique libre dans le modèle restreint (par définition) et l'invariant maximal dans l'espace des paramètres Θ est $\mu'\Sigma^{-1}\mu = \lambda^2$.

Comme U est une statistique invariante maximale, on sait *a priori* que sa distribution ne dépend de (μ, Σ) qu'à travers $\lambda^2 = \mu'\Sigma^{-1}\mu$. De plus, pour le cas d'une loi multinormale $N_p(\sqrt{n}\mu, \Sigma)$, U est distribuée selon une loi de Fisher décentrée $F_{p,n-p}(n\lambda^2)$ voir Muirhead (1982) et Giri (1995) parmi d'autres.

Maintenant, nous présenterons une caractérisation des règles équivariantes.

Proposition 2.2.1 Un estimateur $\hat{\mu}_e \in \Delta$ basé sur le couple (Y,S) est équivariant si et seulement s'il existe une fonction mesurable k_1 telle que :

$$\hat{\mu}_e(Y,S) = k_1\left(Y'S^{-1}Y\right)Y .$$

Démonstration. En utilisant le théorème 2.1.3, nous avons :

$$\hat{\mu}_e(Y,S) = h(Y,S)k\left(Y'S^{-1}Y\right).$$

Mais nous avons aussi :

$$h(Y,S) = S^{\frac{1}{2}}h\left(S^{-\frac{1}{2}}Y,I\right).$$

Définissons une matrice $H \in O(p)$ ayant pour première colonne $Q = S^{-\frac{1}{2}}Y$.

On obtient alors :

$$h\left(S^{-\frac{1}{2}}Y,I\right) = h(HH'S^{-\frac{1}{2}}Y,I) = Hh((Y'S^{-1}Y)^{\frac{1}{2}}e,I).$$

Mais les composantes de $h((Y'S^{-1}Y)^{\frac{1}{2}}e,I)$, excepté la première composante $h_1((Y'S^{-1}Y)^{\frac{1}{2}}e,I)$, sont nulles.

Nous avons alors la relation suivante :

$$h(Y,S) = S^{\frac{1}{2}}Qh_1((Y'S^{-1}Y)^{\frac{1}{2}}e,I) = Yh_1\left((Y'S^{-1}Y)^{\frac{1}{2}}e,I\right).$$

On obtient :

$$\mu_e(Y,S) = Y\,h_1\left((Y'S^{-1}Y)^{\frac{1}{2}}e,I\right)k\left(Y'S^{-1}Y\right).$$

En posant $k_1\left(Y'S^{-1}Y\right) = k\left(Y'S^{-1}Y\right)h_1\left((Y'S^{-1}Y)^{\frac{1}{2}}e,I\right)$, on a le résultat.

Nous recherchons le meilleur estimateur équivariant basé sur le couple (Y,S).

Théorème 2.2.2 Le meilleur estimateur équivariant de la moyenne μ, d'un mélange de lois normales avec la fonction de perte (2.2.1) et la contrainte (2.2.2), est donné par :

$$\hat{\mu}(Y,S) = \frac{E_\theta(Y^t\Sigma^{-1}\mu \mid U)}{E_\theta(Y^t\Sigma^{-1}Y \mid U)} Y,\ \theta = (\mu,\Sigma) \in \Theta_\lambda . \tag{2.2.3}$$

Démonstration. En utilisant la proposition précédente, la fonction de risque d'un estimateur équivariant $\delta(Y,S) = k(U)Y$ est donnée par :

$$R((\mu,\Sigma),\delta) = E_\theta\left\{E_\theta\left\{(\mu - k(U)Y)^t\Sigma^{-1}(\mu - k(U)Y) \mid U\right\}\right\}, \forall\, \theta = (\mu,\Sigma) \in \Theta_\lambda .$$

Donc le meilleur estimateur équivariant devient $\hat{\mu}(Y,S) = \hat{k}(U)Y$ où $\hat{k}(U)$ est la fonction qui minimise l'expression $E_\theta\left\{(\mu - k(U)Y)^t\Sigma^{-1}(\mu - k(U)Y) \mid U\right\}$.

Mais cette fonction atteint son minimum en $\hat{k}(U) = \dfrac{E_\theta(Y^t\Sigma^{-1}\mu \mid U)}{E_\theta(Y^t\Sigma^{-1}Y \mid U)}$.

Il est important de préciser que la formule (2.2.3) est générale et s'applique aux membres de la famille elliptique qui ne sont pas des mélanges de lois multinormales. Cependant une expression explicite pour le meilleur estimateur équivariant semble difficilement abordable pour le cas général. Mais nous obtenons ci-dessous une expression explicite pour la sous-famille qui sont des mélanges de lois multinormales.

Traitons le cas particulier d'une famille de lois $N_p(\mu,\Sigma)$ normales avec la contrainte $\mu^t\Sigma^{-1}\mu = \lambda^2$ où λ^2 connu.

2.2.1. Estimation de la moyenne μ d'une loi multinormale $N_p(\mu,\Sigma)$ sous la contrainte $\mu'\Sigma^{-1}\mu=\lambda^2$ avec λ^2 connu.

Cette étude a été faite par Kariya, Perron et Giri (1988) qui ont évalué le meilleur estimateur équivariant pour $p=2$. Marchand (1994) la reprend en traitant le cas le plus général où $p\geq 2$. Nous reprenons ces approches pour dégager le meilleur estimateur équivariant, le risque minimal et leurs propriétés.

Notons par **P.1** ce problème.

La statistique $(\overline{X},S)$ est exhaustive minimale pour (μ,Σ). Et il est bien connu que Y et S sont des variables aléatoires indépendantes dont leurs distributions respectives sont normales $N_p(\sqrt{n}\mu,\Sigma)$ et Wishart $W_p(n-1,\Sigma)$.

Pour évaluer le meilleur estimateur équivariant, nous avons besoin des lemmes suivants. Le premier lemme est utilisé dans l'évaluation de la densité conjointe de (Y,U), proposée au second lemme. Cette densité conjointe de (Y,U) nous servira à évaluer le meilleur estimateur équivariant en utilisant le troisième lemme.

Lemme 2.2.3. Soit A une variable aléatoire admettant une distribution de Wishart $W_p(n,\Sigma)$ avec n un entier positif et $n>p-1$. Alors pour tout vecteur aléatoire M de dimension p indépendant de A, la variable $W=\dfrac{M'\Sigma^{-1}M}{M'A^{-1}M}$ a une distribution de khi-deux χ^2_{n-p+1}.

De plus, elle est indépendante de M.

Démonstration. Voir Muirhead (1982, page 96).

Lemme 2.2.4. Pour $\theta = (\mu, \Sigma) \in \Theta_\lambda$, la densité conjointe de (Y, U) est donnée par :

$$f_\theta(y,u) = c(\lambda)(\det \Sigma)^{-1/2} u^{-\left(\frac{m}{2}+1\right)} (y'\Sigma^{-1}y)^{\frac{m}{2}} e^{\sqrt{n}\mu'\Sigma^{-1}y} e^{-\frac{u+1}{2u}y'\Sigma^{-1}y}, \qquad (2.2.4)$$

où $c^{-1}(\lambda) = (2)^{\frac{n}{2}} (\pi)^{\frac{p}{2}} (\det(\Sigma))^{1/2} \Gamma\left(\frac{m}{2}\right) e^{\frac{n\lambda^2}{2}}$, et $m = n - p$.

Démonstration. En prenant $M = Y$ et $A = S$ dans le théorème précédent, la variable aléatoire $W = \dfrac{Y'\Sigma^{-1}Y}{Y^t S^{-1} Y}$ suit une loi χ^2_{n-p} et elle est indépendante de la variable Y. Donc la densité conjointe de (Y, W) est le produit d'une densité multinormale $N_p(\sqrt{n}\mu, \Sigma)$ et d'une densité d'un Khi-deux χ^2_{n-p}.

Faisons la transformation (Y, W) à (Y, U). Le jacobien de cette transformation est $\dfrac{Y'\Sigma^{-1}Y}{U^2}$.

Alors la densité conjointe de Y et U est donnée par la formule suivante :

$$f_\theta(y,u) = f_Y(y) f_W\left(\frac{y'\Sigma^{-1}y}{u}\right) \frac{y'\Sigma^{-1}y}{u^2}, \qquad (2.2.5)$$

où f_Y et f_W désignent respectivement les fonctions de densités de Y et W.

Ces fonctions f_Y et f_W sont données respectivement par les expressions suivantes :

$$f_Y(y) = (2\pi)^{-\frac{p}{2}} (\det \Sigma)^{-1/2} \exp\left(-\frac{1}{2}(y - \sqrt{n}\mu)' \Sigma^{-1} (y - \sqrt{n}\mu)\right)$$

$$= (2\pi)^{-\frac{p}{2}} (\det \Sigma)^{-1/2} \exp\left(-\frac{1}{2} y'\Sigma^{-1}y\right) \exp\left(-\frac{1}{2} n\lambda^2\right) \exp\left(\sqrt{n}\mu'\Sigma^{-1}y\right) (2.2.6)$$

et

$$f_W\left(\frac{y^t\Sigma^{-1}y}{u}\right)=(2)^{-\frac{m}{2}}u^{-\frac{m}{2}+1}\left(y^t\Sigma^{-1}y\right)^{\frac{m}{2}-1}\left(\Gamma\left(\frac{m}{2}\right)\right)^{-1}\exp\left(-\frac{1}{2}\frac{y^t\Sigma^{-1}y}{u}\right),\quad (2.2.7)$$

où $m=n-p$.

En combinant les relations (2.2.5), (2.2.6) et (2.2.7), on a :

$$f_\theta(y,u)=\frac{(2)^{-\frac{m+p}{2}}(\pi)^{-\frac{p}{2}}}{\Gamma\left(\frac{m}{2}\right)}e^{-\frac{n\lambda^2}{2}}(\det\Sigma)^{-1/2}u^{-\left(\frac{m}{2}+1\right)}(y^t\Sigma^{-1}y)^{\frac{m}{2}}e^{\sqrt{n}\mu^t\Sigma^{-1}y}e^{-\frac{u+1}{2u}y^t\Sigma^{-1}y}.$$

Ce qui achève la démonstration.

Lemme 2.2.5. Soit $k\in IN$ et $b,c\in IR$. Nous avons le résultat suivant :

$$\int_{\mathbb{R}^n}y_1^k(y^ty)^b\exp\left(-cy^ty\right)dy=\begin{cases}c^{-\left(\frac{p+k}{2}+b\right)}\dfrac{\left(\Gamma\left(\frac{1}{2}\right)\right)^{p-1}\Gamma\left(\frac{p+k}{2}+b\right)\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k+p}{2}\right)}, & \text{si } k \text{ est pair,}\\[2ex] 0,\ \text{sinon.}\end{cases}$$

Démonstration. Faisons le changement de variable pour passer en coordonnées polaires, plus précisément :

$$y_1=\sqrt{r}\sin\theta$$

$$y_2=\sqrt{r}\cos\theta_1\sin\theta_2$$

.

.

.

$$y_{p-1}=\sqrt{r}\cos\theta_1\cos\theta_2\ldots\cos\theta_{p-2}\sin\theta_{p-1}$$

$$y_p=\sqrt{r}\cos\theta_1\cos\theta_2\ldots\cos\theta_{p-2}\cos\theta_{p-1},$$

où $(r,\theta_1,\theta_2...\theta_{p-2},\theta_{p-1}) \in IR^{+} \text{x} \left[-\frac{\pi}{2},\frac{\pi}{2}\right]^{p-2} \text{x}\,[0,2\pi]$.

Le jacobien associé à cette transformation est :

$$J(r,\theta_1,\theta_2,...,\theta_{p-2},\theta_{p-1}) = \frac{1}{2} r^{\frac{p}{2}-1} \prod_{j=1}^{p-2} \left(\cos\theta_j\right)^{p-1-j}$$.(Voir Perron 1987, page 59).

L'intégrale devient alors :

$$\int_{\mathrm{IR}^n} y_1^k (y^t y)^b \exp\left(-c y^t y\right) dy = \int_0^{+\infty} \frac{1}{2} r^{\frac{p+k}{2}+b-1} \exp\left(-cr\right) dr$$

$$\text{x} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cdots \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^{2\pi} (\sin\theta_1)^k \prod_{j=1}^{p-2} \left(\cos\theta_j\right)^{p-1-j} d\theta_{p-1} ... d\theta_1 .$$

Mais nous savons que :

$$\int_0^{+\infty} \frac{1}{2} r^{\frac{p+k}{2}+b-1} \exp\left(-cr\right) dr = \frac{1}{2} c^{-\left(\frac{p+k}{2}+b\right)} \Gamma\left(\frac{p+k}{2}+b\right) dr$$

et

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} .. \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^{2\pi} (\sin\theta_1)^k \prod_{j=1}^{p-2} \left(\cos\theta_j\right)^{p-1-j} d\theta_{p-1} ... d\theta_1 = \begin{cases} 2\dfrac{\left(\Gamma\left(\frac{1}{2}\right)\right)^{p-1} \Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k+p}{2}\right)}, \text{ si } k \text{ est pair} \\ 0, \text{ sinon.} \end{cases}$$

D'où on obtient bien :

$$\int_{\mathrm{IR}^n} y_1^k (y^t y)^b \exp\left(-c y^t y\right) dy = \begin{cases} c^{-\left(\frac{p+k}{2}+b\right)} \dfrac{\left(\Gamma\left(\frac{1}{2}\right)\right)^{p-1} \Gamma\left(\frac{p+k}{2}+b\right)\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k+p}{2}\right)}, & \text{si } k \text{ est pair} \\ 0, \text{ sinon.} \end{cases}$$

Théorème 2.2.6. Le meilleur estimateur équivariant de la moyenne μ, d'une loi $N_p(\mu,\Sigma)$ sous la fonction de perte (2.2.1) et la contrainte (2.2.2), est donné par l'expression suivante :

$$\hat{\mu}(Y,S) = \hat{k}(U)Y$$

avec

$$\hat{k}(U) = \frac{\sqrt{n}\lambda^2}{p} \frac{{}_1F_1\left(\frac{n}{2}+1; \frac{p}{2}+1; \frac{n\lambda^2}{2}\frac{U}{U+1}\right)}{{}_1F_1\left(\frac{n}{2}+1; \frac{p}{2}; \frac{n\lambda^2}{2}\frac{U}{U+1}\right)}, \tag{2.2.8}$$

où ${}_kF_\ell$ représente la fonction hypergéométrique généralisée dont l'expression est donnée par

$${}_kF_\ell(a_1,\ldots,a_k;b_1,\ldots,b_\ell;x) = \sum_{i=0}^{+\infty} \frac{(a_1)_i,\ldots,(a_k)_i}{(b_1)_i,\ldots,(b_\ell)_i} \frac{x^i}{i!} \quad \text{et}$$

$$(a)_i = \begin{cases} 1 & \text{si } i = 0 \\ a(a+1)(a+2)\ldots(a+i-1) & \text{sinon.} \end{cases}$$

Démonstration. Sans perte de généralité, on pose $\theta = (\mu_0, \Sigma_0)$ où $\Sigma_0 = I_p$ et $\mu_0 = (\lambda, 0, \ldots, 0)$. On obtient, en utilisant la relation (2.2.3), l'expression suivante :

$$\hat{k}(u) = \frac{E(Y^t \mu_0 \mid \mathrm{U} = \mathrm{u})}{E(Y^t Y \mid \mathrm{U} = \mathrm{u})} \qquad (2.2.9).$$

Il reste à calculer explicitement $E(Y^t \mu_0 \mid \mathrm{U} = \mathrm{u})$ et $E(Y^t Y \mid \mathrm{U} = \mathrm{u})$.

Nous avons :

$$E(Y^t \mu_0 \mid \mathrm{U} = \mathrm{u}) = \frac{\lambda}{f_U(u)} \int_{\mathrm{IR}^p} y_1 f_\theta(y, u) dy .$$

Comme la densité de U est donnée par l'expression :

$$f_U(u) = \int_{\mathrm{IR}^p} f_\theta(y, u) dy ,$$

alors

$$E(Y^t \mu_0 \mid \mathrm{U} = \mathrm{u}) = \lambda \frac{\int_{\mathrm{IR}^p} y_1 f_\theta(y, u) dy}{\int_{\mathrm{IR}^p} f_\theta(y, u) dy} .$$

De même on montre que :

$$E(Y^t Y \mid \mathrm{U} = \mathrm{u}) = \frac{\int_{\mathrm{IR}^p} y^t y f_\theta(y, u) dy}{\int_{\mathrm{IR}^p} f_\theta(y, u) dy} .$$

En utilisant la densité conjointe de Y et U, pour $\theta_0 = (\mu_0, \Sigma_0)$ où $\Sigma_0 = I$ et $\mu_0 = (\lambda, 0, \ldots, 0)$, on a :

$$E(Y^t \mu_0 \mid \mathrm{U} = \mathrm{u}) = \frac{\int_{\mathrm{IR}^p} y_1 (y^t y)^{\frac{m}{2}} e^{\sqrt{n\lambda}\, y_1} e^{-\frac{u+1}{2u} y^t y} dy}{\int_{\mathrm{IR}^p} (y^t y)^{\frac{m}{2}} e^{\sqrt{n\lambda}\, y_1} e^{-\frac{u+1}{2u} y^t y} dy} \qquad (2.2.10)$$

et

$$E(Y^tY \mid \mathrm{U}=\mathrm{u}) = \frac{\int_{\mathrm{IR}^p} (y^t y)^{\frac{m}{2}+1} e^{\sqrt{n}\lambda y_1} e^{-\frac{u+1}{2u} y^t y} dy}{\int_{\mathrm{IR}^p} (y^t y)^{\frac{m}{2}} e^{\sqrt{n\lambda} e^t y_1} e^{-\frac{u+1}{2u} y^t y} dy}. \tag{2.2.11}$$

Mais on sait que :

$$e^{\sqrt{n}\lambda y_1} = \sum_{k=0}^{\infty} \frac{\left(\sqrt{n}\lambda y_1\right)^k}{k!}. \tag{2.2.12}$$

Alors en utilisant la relation (2.2.12), les relations (2.2.10) et (2.2.11) deviennent respectivement :

$$E\left(Y^t \mu_0 \mid U=u\right) = \frac{\sum_{k=0}^{\infty} \frac{\left(\sqrt{n}\lambda\right)^k}{k!} \int_{\mathrm{IR}^p} y_1^{k+1} (y^t y)^{\frac{m}{2}} e^{-\frac{u+1}{2u} y^t y} dy}{\sum_{k=0}^{\infty} \frac{\left(\sqrt{n}\lambda\right)^k}{k!} \int_{\mathrm{IR}^p} y_1^k \left(y^t y\right)^{\frac{m}{2}} e^{-\frac{u+1}{2u} y^t y} dy} \tag{2.2.13}$$

et

$$E\left(Y^t Y \mid U=u\right) = \frac{\sum_{k=0}^{\infty} \frac{\left(\sqrt{n}\lambda\right)^k}{k!} \int_{\mathrm{IR}^p} y_1^k (y^t y)^{\frac{m}{2}+1} e^{-\frac{u+1}{2u} y^t y} dy}{\sum_{k=0}^{\infty} \frac{\left(\sqrt{n}\lambda\right)^k}{k!} \int_{\mathrm{IR}^p} y_1^k \left(y^t y\right)^{\frac{m}{2}} e^{-\frac{u+1}{2u} y^t y} dy}. \tag{2.2.14}$$

En réunissant les expressions (2.2.9), (2.2.13) et (2.2.14), on obtient :

$$\hat{k}(U) = \frac{\sum_{k=0}^{\infty} \frac{\left(\sqrt{n}\lambda\right)^k}{k!} \int_{\mathrm{IR}^p} y_1^{k+1} (y^t y)^{\frac{m}{2}} e^{-\frac{U+1}{2U} y^t y} dy}{\sum_{k=0}^{\infty} \frac{\left(\sqrt{n}\lambda\right)^k}{k!} \int_{\mathrm{IR}^p} y_1^k \left(y^t y\right)^{\frac{m}{2}+1} e^{-\frac{U+1}{2U} y^t y} dy}. \tag{2.2.15}$$

D'après le lemme 2.2.5, l'expression (2.2.15) devient :

$$\hat{k}(U) = \sqrt{n}\lambda^2 \frac{\displaystyle\sum_{k=0}^{\infty} \frac{\lambda^{2k} n^k}{(2k+1)!} 2^k \frac{\Gamma\left(k+1+\frac{m-p}{2}\right)\Gamma\left(k+\frac{3}{2}\right)}{\Gamma\left(k+\frac{p}{2}+1\right)} \left(\frac{U}{1+U}\right)^k}{\displaystyle\sum_{k=0}^{\infty} \frac{\lambda^{2k} n^k}{(2k)!} 2^k \frac{\Gamma\left(k+1+\frac{m-p}{2}\right)\Gamma\left(k+\frac{1}{2}\right)}{\Gamma\left(k+\frac{p}{2}\right)} \left(\frac{U}{1+U}\right)^k}.$$

Mais nous avons aussi la relation suivante :

$\Gamma\left(\frac{1}{2}\right)(2k+t)! = 2^{2k+t} k!\Gamma\left(k+t+\frac{1}{2}\right)$, $t = 0,\ 1$ (Perron 1987, page 61).

En utilisant les deux relations précédentes, on obtient la formule suivante :

$$\hat{k}(U) = \frac{\sqrt{n}\lambda^2}{2} \frac{\displaystyle\sum_{k=0}^{\infty} \frac{\lambda^{2k} n^k}{2^k k!} \frac{\Gamma\left(k+1+\frac{n}{2}\right)}{\Gamma\left(k+\frac{p}{2}+1\right)} \left(\frac{U}{1+U}\right)^k}{\displaystyle\sum_{k=0}^{\infty} \frac{\lambda^{2k} n^k}{2^k k!} \frac{\Gamma\left(k+1+\frac{n}{2}\right)}{\Gamma\left(k+\frac{p}{2}\right)} \left(\frac{U}{1+U}\right)^k}.$$

D'où par définition d'une fonction hypergéométrique, nous avons :

$$\hat{k}(U) = \frac{\sqrt{n}\lambda^2}{p} \frac{{}_1F_1\left(\frac{n}{2}+1; \frac{p}{2}+1; \frac{n\lambda^2}{2}\frac{U}{U+1}\right)}{{}_1F_1\left(\frac{n}{2}+1; \frac{p}{2}; \frac{n\lambda^2}{2}\frac{U}{U+1}\right)}.$$

Calculons à présent le risque minimal.

Théorème 2.2.7. Sous les conditions du problème, le risque minimal dans la classe des estimateurs équivariants est égal à :

$$R((\mu,\Sigma),\hat{\mu}) = \lambda^2 - \frac{n\lambda^4}{p} e^{-\frac{n\lambda^2}{2}} E\left(\frac{{}_1F_1{}^2\left(\frac{n}{2}+1, \frac{p}{2}+1, \frac{n\lambda^2}{2}V\right)}{{}_1F_1\left(\frac{n}{2}+1; \frac{p}{2}; \frac{n\lambda^2}{2}V\right)} \right), \forall(\mu,\Sigma) \in \Theta_\lambda \qquad (2.2.16)$$

où V est une variable aléatoire de densité bêta $\mathscr{Be}\left(\frac{p}{2}+1, \frac{m}{2}\right)$.

Démonstration. Sous la fonction de perte (2.2.1), le risque du meilleur estimateur équivariant évalué en $\theta_0 = (\mu_0, \Sigma_0)$ où $\mu_0^t = (\lambda, 0, \ldots, 0)$ et $\Sigma_0 = I_p$, est donné par la relation suivante :

$$R((\mu_0, I), \hat{\mu}) = E\left\{ E\left[\left(\hat{k}(U)Y - \mu_0\right)^t \left(\hat{k}(U)Y - \mu_0\right) | U \right] \right\}.$$

On peut également écrire :

$$R((\mu_0, I), \hat{\mu}) = E\left\{\hat{k}^2(U) E\left(Y'Y \mid U\right) - 2\hat{k}(U) E(Y_1 \mid U) + \lambda^2\right\}$$

$$= \lambda^2 - E\left\{ \frac{\left(E\left(Y'\mu \mid U\right)\right)^2}{E\left(Y'Y \mid U\right)} \right\}. \qquad (2.2.17)$$

En vertu du lemme 2.2.5, les relations (2.2.13) et (2.2.14) deviennent respectivement :

$$E\left(Y'\mu \mid U = u\right) = \frac{n^{1/2}\lambda^2}{p} \frac{u}{1+u} \frac{{}_1F_1\left(\frac{n}{2}+1; \frac{p}{2}+1; \frac{n\lambda^2}{2}\frac{u}{u+1}\right)}{{}_1F_1\left(\frac{n}{2}; \frac{p}{2}; \frac{n\lambda^2}{2}\frac{u}{u+1}\right)} \qquad (2.2.18)$$

et

$$E\left(Y'Y \mid U=u\right)=n\frac{u}{1+u}\frac{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}\frac{u}{u+1}\right)}{{}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\frac{u}{u+1}\right)}. \qquad (2.2.19)$$

Donc en réunissant les relations (2.2.17), (2.2.18) et (2.2.19), le risque devient :

$$R\left(\left(\mu_0,I\right),\hat{\mu}\right)=\lambda^2-\frac{n^2\lambda^4}{p^2}E^U\left(\frac{u}{1+u}\frac{{}_1F_1^{\,2}\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}\frac{U}{U+1}\right)}{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}\frac{U}{U+1}\right){}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\frac{U}{U+1}\right)}\right). \qquad (2.2.20)$$

Comme la loi de probabilité de U est Fisher décentrée $F_{p,n-p}\left(n\lambda^2\right)$ alors la densité de U est donnée par l'expression suivante :

$$f_U(u)=\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{p}{2}\right)}e^{-\frac{n\lambda^2}{2}}u^{\frac{p}{2}-1}\left(1+u\right)^{-\frac{n}{2}}{}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\frac{u}{1+u}\right). \qquad (2.2.21)$$

D'où en utilisant les relations (2.2.20) et (2.2.21), on a :

$$R\left(\left(\mu_0,I\right),\hat{\mu}\right)=\lambda^2-\frac{n^2\lambda^4}{p^2}e^{-\frac{n\lambda^2}{2}}\int_0^{+\infty}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{p}{2}\right)}\frac{u^{\frac{p}{2}}}{\left(1+u\right)^{\frac{n}{2}+1}}\frac{{}_1F_1^{\,2}\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}\frac{u}{u+1}\right)}{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}\frac{u}{u+1}\right)}du.$$

En faisant le changement de variable $v=\frac{u}{1+u}$, on a :

$$R\left(\left(\mu_0, I\right), \hat{\mu}\right) = \lambda^2 - \frac{n^2\lambda^4}{p^2} e^{-\frac{n\lambda^2}{2}} \int_0^{+\infty} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{p}{2}\right)} v^{\frac{p}{2}} (1-v)^{\frac{m}{2}-1} \frac{{}_1F_1^{\,2}\left(\frac{n}{2}+1; \frac{p}{2}+1; \frac{n\lambda^2}{2}v\right)}{{}_1F_1\left(\frac{n}{2}+1; \frac{p}{2}; \frac{n\lambda^2}{2}v\right)} dv$$

$$= \lambda^2 - \frac{n\lambda^4}{p} e^{-\frac{n\lambda^2}{2}} \int_0^{+\infty} \frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{p}{2}+1\right)} v^{\frac{p}{2}} (1-v)^{\frac{m}{2}-1} \frac{{}_1F_1^{\,2}\left(\frac{n}{2}+1; \frac{p}{2}+1; \frac{n\lambda^2}{2}v\right)}{{}_1F_1\left(\frac{n}{2}+1; \frac{p}{2}; \frac{n\lambda^2}{2}v\right)} dv.$$

Comme la fonction de risque est constante sur Θ_λ alors nous avons le résultat.

Lemme 2.2.8. Si $\alpha - \beta$ et 2β sont deux entiers positifs alors

$$ {}_1F_1(\alpha; \beta; x) = e^x \Gamma(\beta) \sum_{i=0}^{\alpha-\beta} \binom{\alpha-\beta}{i} \frac{x^i}{\Gamma(\beta+i)}. \tag{2.2.22}$$

Démonstration. On utilise le fait que, si une variable aléatoire Y_k soit distribuée selon une loi χ_k^2 alors le moment d'ordre α de Y_k est :

$$E(\chi_k^\alpha) = 2^\alpha \frac{\Gamma\left(\frac{k}{2}+\alpha\right)}{\Gamma\left(\frac{k}{2}\right)}, \quad \forall\, \alpha > -\frac{k}{2}.$$ (voir Johnson, Kotz & Balakrifhman, 1995)

Donc nous avons :

$$ {}_1F_1(\alpha; \beta; x) = e^x \frac{\Gamma(\beta)}{\Gamma(\alpha)} \sum_{k=0}^{+\infty} e^{-x} \frac{\Gamma(\alpha+k)}{\Gamma(\beta+k)} \frac{x^k}{k!}$$

$$= 2^{\beta-\alpha} e^x \frac{\Gamma(\beta)}{\Gamma(\alpha)} \sum_{k=0}^{+\infty} e^{-x} E\left(\chi_{2\beta+2k}^{\alpha-\beta}\right) \frac{x^k}{k!}$$

$$= 2^{\beta-\alpha} e^x \frac{\Gamma(\beta)}{\Gamma(\alpha)} E\left(V^{\alpha-\beta}\right),$$

où V est de loi de khi-deux décentrée $\chi^2_{2\beta}(2x)$.(voir Johnson, Kotz & Balakrifhman, 1995)

Mais le moment d'ordre r d'une variable aléatoire de densité $\chi^2_\nu(\delta^2)$ est donné par la relation suivante :

$$E(V^r) = 2^r \sum_{k=0}^{r} \binom{r}{k} \left(\frac{\delta^2}{2}\right)^k \frac{\Gamma\left(\frac{\nu}{2}+r\right)}{\Gamma\left(\frac{\nu}{2}+k\right)},$$

On obtient alors le résultat suivant :

$${}_1F_1(\alpha;\beta;x) = 2^{\beta-\alpha}\Gamma(\beta)e^x \sum_{k=0}^{\alpha-\beta} \binom{\alpha-\beta}{k} \frac{x^k}{\Gamma(\beta+k)}.$$

Ce qui nous donne le résultat.

Appliquons ce résultat pour notre problème **P.1** lorsque $\frac{n-p}{2}$ est un entier positif.

Théorème 2.2.9. Si on est sous les conditions du problème. Si on suppose de plus que $\frac{n-p}{2}$ est un entier positif. Alors le meilleur estimateur équivariant de la moyenne μ est donné par :

$$\hat{\mu}(Y,S) = \hat{k}(U)Y = g(\mathrm{V})Y$$

où $g(V) = \frac{\sqrt{n}\lambda^2}{2}\frac{a(V)}{b(\mathrm{V})}$, $a(V) = \sum_{k=0}^{\ell} \binom{\ell}{k} \frac{\left(\frac{n\lambda^2\mathrm{V}}{2}\right)^k}{\Gamma\left(\frac{p}{2}+k+1\right)}$, $b(V) = \sum_{k=0}^{\ell+1} \binom{\ell+1}{k} \frac{\left(\frac{n\lambda^2\mathrm{V}}{2}\right)^k}{\Gamma\left(\frac{p}{2}+k\right)}$,

$\ell = \frac{m}{2}$ et $V = \frac{U}{U+1}$.

Démonstration. C'est une conséquence du théorème 2.2.6 et du lemme 2.2.8.

On reprend à présent l'approche du problème **P.1**, mais dans le cas plus général où la famille des distributions de X est formée de mélanges de lois normales.

2.2.2. Estimation de la moyenne μ d'une distribution de mélanges de lois normales sous $\mu^t\Sigma^{-1}\mu = \lambda^2, \lambda^2$ connu.

Nous reprenons l'estimation de μ sous la contrainte $\mu^t\Sigma^{-1}\mu = \lambda^2$ pour le cas d'un mélange de lois normales, où Σ représente la matrice de variance-covariance. Et on note par **P.2**, ce problème.

Définition 2.2.10. Une variable aléatoire X est issue d'une distribution dite mélange de lois normales si elle admet la représentation suivante :

la loi conditionnelle de X, étant donné $Z = z$, est normale $N_p(\mu, z\Omega)$, $\forall z \geq 0$.

Nous supposons que Ω est définie positive, $P(Z > 0) = 1$ et $E(Z) < \infty$. Les paramètres de cette distribution sont μ, Ω, Z.

L'ensemble des mélanges de lois normales, pour $\Sigma = E(Z)\Omega$, est une sous-classe des lois elliptiques $E_p(\mu, \Sigma)$. La classe des lois elliptiques, ainsi que la classe des mélanges de lois normales ont été étudiées par beaucoup de statisticiens dont Kelker (1970), Strawderman (1974), Berger (1975), Muirhead (1982) et Fang, Kotz and Ng (1990).

Lemme 2.2.11. Soit $X_1, X_2, \ldots, X_n$ des variables aléatoires conditionnellement indépendantes étant donné l'événement $\{Z = z\}$ et de lois identiques multinormales $N_p(\mu, z\Omega)$ étant donné $\{Z = z\}$. Alors les variables aléatoires Y et S

étant donné $Z = z$ sont indépendantes et leurs distributions conditionnelles étant donné $Z = z$ sont respectivement :

$$N_p\left(\sqrt{n}\mu, z\Omega\right) \text{ et } W_p(n-1, z\Omega).$$

Lemme 2.2.12. Soit X un vecteur aléatoire de mélanges de lois normales de paramètres μ, Ω, Z. Notons par Σ la matrice de variance et covariance de X. On a la relation suivante :

$$\Sigma = E(Z)\Omega.$$

Démonstration. On sait que

$$\Sigma = \operatorname{cov}(X) = E\left(XX^t\right) - \mu\mu^t.$$

Mais nous avons aussi la relation suivante :

$$E\left(XX^t\right) = E\left(E\left(XX^t \mid Z\right)\right) = E\left(Z\Omega + \mu\mu^t\right) = E(Z)\Omega + \mu\mu^t.$$

D'où on obtient :

$$\Sigma = E(Z)\Omega + \mu\mu^t - \mu\mu^t = E(Z)\Omega.$$

On veut trouver le meilleur estimateur équivariant de la moyenne d'un mélange de lois normales de paramètres μ, Ω, Z lorsque l'information paramétrique est de la forme suivante :

$$\Theta_\lambda = \left\{(\mu, \Sigma) \in IR^p \text{x } S(p) : \mu^t \Sigma^{-1} \mu = \lambda^2\right\},$$

où λ^2 est positif et connu.

Nous évaluerons le risque du meilleur estimateur équivariant. À partir de maintenant on supposera toujours que les lois elliptiques considérées sont des mélanges de lois normales. On note par H la fonction de répartition de la variable Z.

Supposons qu'on soit sous les conditions du lemme 2.2.11 et qu'on veuille estimer la moyenne μ avec la fonction de perte (2.2.1) avec la contrainte $\mu' \Sigma^{-1} \mu = \lambda^2$ (connu).

Théorème 2.2.13. Sous les conditions du problème, le meilleur estimateur équivariant de la moyenne μ, est donné par :

$$\hat{\mu}(Y,S) = \hat{k}(U)Y$$

avec

$$\hat{k}(U) = \frac{\sqrt{n}\lambda^2}{p} \frac{\displaystyle\int_0^{+\infty} \frac{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\frac{U}{U+1}\right)}{{}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\frac{U}{U+1}\right)} H^{Z|U}(dz)}{\displaystyle\int_0^{+\infty} \frac{Z}{E(Z)} \frac{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\frac{U}{U+1}\right)}{{}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\frac{U}{U+1}\right)} H^{Z|U}(dz)}. \tag{2.2.23}$$

Démonstration. Pour évaluer le multiplicateur $\hat{k}(u)$, il suffit, d'après le théorème 2.2.2, de calculer $E(Y'\Sigma^{-1}\mu \mid \mathrm{U}=\mathrm{u})$ et $E(Y'\Sigma^{-1}Y \mid \mathrm{U}=\mathrm{u})$.

Mais nous avons :

$$E(Y'\Sigma^{-1}\mu \mid \mathrm{U}=\mathrm{u}) = E\left\{E\left(Y'\Sigma^{-1}\mu \mid \mathrm{U}=\mathrm{u},\mathrm{Z}\right) \mid \mathrm{U}=\mathrm{u}\right\}$$

$$= E\left\{\frac{Z}{E(Z)} E\left(Y'(\mathrm{Z}\Omega)^{-1}\mu \mid \mathrm{U}=\mathrm{u},\mathrm{Z}\right) \mid \mathrm{U}=\mathrm{u}\right\} \tag{2.2.24}$$

et

$$E(Y^t\Sigma^{-1}Y \mid \mathrm{U}=\mathrm{u}) = E\left\{E\left(Y^t\Sigma^{-1}\mathrm{Y} \mid \mathrm{U}=\mathrm{u},\mathrm{Z}\right) \mid \mathrm{U}=\mathrm{u}\right\}$$

$$= E\left\{\frac{Z}{E(Z)}E\left(Y^t(Z\Omega)^{-1}Y \mid \mathrm{U}=\mathrm{u},\mathrm{Z}\right) \mid \mathrm{U}=\mathrm{u}\right\}. \qquad (2.2.25)$$

Pour $Z=z$, on trouve $\mu^t(z\Omega)^{-1}\mu = \frac{1}{z}\mu^t\Omega^{-1}\mu = \frac{E(Z)}{z}\mu^t\Sigma^{-1}\mu = \frac{E(Z)}{z}\lambda^2$.

Donc d'après des résultats antérieurs, les expressions $E\left(Y^t(Z\Omega)^{-1}\mu \mid \mathrm{U}=\mathrm{u},\mathrm{Z}\right)$ et $E\left(Y^t(Z\Omega)^{-1}Y \mid \mathrm{U}=\mathrm{u},\mathrm{Z}\right)$ sont obtenues en remplaçant respectivement dans les expressions (2.2.18) et (2.2.19), λ^2 par $\frac{E(Z)}{z}\lambda^2$. On obtient alors :

$$E\left(Y^t(Z\Omega)^{-1}\mu \mid \mathrm{U}=\mathrm{u},\mathrm{Z}\right) = \frac{n^{3/2}\lambda^2}{p}\frac{E(Z)}{Z}\frac{u}{1+u}\frac{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\frac{u}{u+1}\right)}{{}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\frac{u}{u+1}\right)}$$

et

$$E\left(Y^t(Z\Omega)^{-1}Y \mid \mathrm{U}=\mathrm{u},\mathrm{Z}\right) = n\frac{u}{1+u}\frac{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\frac{u}{u+1}\right)}{{}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\frac{u}{u+1}\right)}.$$

D'où les relations (2.2.24) et (2.2.25) deviennent :

$$E(Y^t\Sigma^{-1}\mu \mid \mathrm{U}=\mathrm{u}) = \frac{n^{3/2}\lambda^2}{p}\frac{u}{1+u}\int_0^{+\infty}\frac{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\frac{u}{u+1}\right)}{{}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\frac{u}{u+1}\right)}H^{Z|U=u}(dz)$$

(2.2.26)

et

$$E(Y'\Sigma^{-1}Y \mid \mathrm{U}=\mathrm{u}) = n\frac{u}{1+u}\int_0^{+\infty}\frac{Z}{E(Z)}\frac{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\frac{u}{u+1}\right)}{{}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\frac{u}{u+1}\right)}H^{Z|U=u}(dz).$$

(2.2.27)

En combinant les relations (2.2.3), (2.2.26) et (2.2.27), on obtient le résultat.

On peut écrire la relation (2.2.23) d'une autre manière.

Corollaire 2.2.14. Sous les conditions du problème, le meilleur estimateur équivariant est donné par :

$$\hat{\mu}(Y,S) = \hat{k}(U)Y$$

avec

$$\hat{k}(U) = \frac{\sqrt{n\lambda^2}\mathrm{E}(Z)}{p}\frac{\int_0^{+\infty} e^{-\frac{n\lambda^2}{2}\frac{E(Z)}{Z}}\,{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}\frac{\mathrm{E}(Z)}{Z}\frac{U}{1+U}\right)H(dz)}{\int_0^{+\infty} ze^{-\frac{n\lambda^2}{2}\frac{E(Z)}{Z}}\,{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}\frac{\mathrm{E}(Z)}{Z}\frac{U}{1+U}\right)H(dz)}. \quad (2.2.28)$$

Démonstration. Utilisons la relation suivante :

$$dH^{Z|U=u}(z) = \frac{f^{U|Z}(u)\;dH(z)}{f^U(u)}. \quad (2.2.29)$$

Mais $f^{U|Z}(u)$ est obtenue en remplaçant dans (2.2.21), λ^2 par $\frac{E(Z)}{z}\lambda^2$.

Dans ce cas on a :

$$f^{U|Z}(u)=\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{p}{2}\right)}e^{-\frac{n\lambda^2}{2}\frac{E(Z)}{Z}}u^{\frac{p}{2}-1}(1+u)^{-\frac{n}{2}}{}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\frac{u}{1+u}\right). \quad (2.2.30)$$

En utilisant (2.2.29) et (2.2.30) dans la relation (2.2.39.), on obtient :

$$\hat{k}(U)=\frac{\sqrt{n}\lambda^2E(Z)}{p}\frac{\int_0^{+\infty}e^{-\frac{n\lambda^2}{2}\frac{E(Z)}{Z}}{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\frac{U}{1+U}\right)H(dz)}{\int_0^{+\infty}ze^{-\frac{n\lambda^2}{2}\frac{E(Z)}{Z}}{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\frac{U}{1+U}\right)H(dz)}.$$

Remarque 2.2.15. Sous les conditions du problème **P.2**, le multiplicateur du meilleur estimateur équivariant de la moyenne μ ne change pas lorsque $Z\to cZ$, $c>0$ et que matrice Σ reste constante.

Théorème 2.2.16. Sous les conditions du problème, le risque minimal dans la classe des estimateurs équivariants évalué en $(\mu,\Sigma)\in\Theta_\lambda$ est égal à :

$$R((\mu,\Sigma),\hat{\mu})=\lambda^2-\frac{n\lambda^4E(Z)}{p}E\left(\frac{\left(\int_0^{\infty}e^{-\frac{n\lambda^2}{2}\frac{E(Z)}{Z}}{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\mathrm{V}\right)H(dz)\right)^2}{\int_0^{\infty}ze^{-\frac{n\lambda^2}{2}\frac{E(Z)}{Z}}{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\mathrm{V}\right)H(dz)}\right) \quad (2.2.49)$$

où V est une variable aléatoire de loi bêta $\mathscr{B}e\left(\frac{p}{2}+1,\frac{m}{2}\right)$.

Démonstration. On sait que :

$$R((\mu,\Sigma),\hat{\mu})=\lambda^2-E\left(\frac{E^2\left(Y'\Sigma^{-1}\mu\,|\,U\right)}{E\left(Y'\Sigma^{-1}Y\,|\,U\right)}\right). \quad (2.2.32)$$

Mais en réunissant les relations (2.2.26), (2.2.27) et (2.2.32), le risque minimal devient :

$$R((\mu,\Sigma),\hat{\mu})=\lambda^{2}-\frac{n^{2}\lambda^{4}E(Z)}{p^{2}}E\left\{\frac{U}{1+U}\frac{\left(\int_{0}^{\infty}\frac{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^{2}}{2}\frac{E(Z)}{Z}\frac{U}{1+U}\right)}{{}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{n\lambda^{2}}{2}\frac{E(Z)}{Z}\frac{U}{1+U}\right)}H^{Z|U}(dz)\right)^{2}}{\int_{0}^{\infty}z\frac{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^{2}}{2}\frac{E(Z)}{Z}\frac{U}{1+U}\right)}{{}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{n\lambda^{2}}{2}\frac{E(Z)}{Z}\frac{U}{1+U}\right)}H^{Z|U}(dz)}\right\}$$

Utilisons la relation (2.2.29) et la densité de U conditionnellement à Z. on obtient alors la relation suivante :

$$R((\mu,\Sigma),\hat{\mu})=\lambda^{2}-\frac{n\lambda^{4}E(Z)}{p}\int_{0}^{\infty}\left\{\frac{\left(\int_{0}^{\infty}e^{-\frac{n\lambda^{2}}{2}\frac{E(Z)}{Z}}\,{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^{2}}{2}\frac{E(Z)}{Z}\frac{u}{u+1}\right)H(dz)\right)^{2}}{\int_{0}^{\infty}ze^{-\frac{n\lambda^{2}}{2}\frac{E(Z)}{Z}}\,{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^{2}}{2}\frac{E(Z)}{Z}\frac{u}{u+1}\right)H(dz)}\right.$$

$$\left.\mathrm{x}\quad\frac{u^{\frac{p}{2}}(1+u)^{-\frac{n}{2}-1}}{B\left(\frac{m}{2},\frac{p}{2}\right)}\right\}du.$$

En faisant un changement de variable $v=\dfrac{u}{1+u}$, nous avons le résultat.

Proposition 2.2.17. Sous les conditions du problème, le meilleur estimateur équivariant et le risque minimal dans la classe des estimateurs équivariants sont donnés par les expressions suivantes :

$$\hat{k}(U)=\frac{2t}{\sqrt{np}}\frac{E\left(e^{-tW}\,{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;tW\frac{U}{1+U}\right)\right)}{E\left(W^{-1}e^{-tW}\,{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};tW\frac{U}{1+U}\right)\right)},\qquad(2.2.33)$$

$$R((\mu,\Sigma),\hat{\mu})=\lambda^{2}-\frac{4t^{2}}{np}E_{V}\left(\frac{E^{2}\left(e^{-tW}\,{}_{1}F_{1}\left(\frac{n}{2}+1;\frac{p}{2}+1;tVW\right)\right)}{E\left(W^{-1}e^{-tW}\,{}_{1}F_{1}\left(\frac{n}{2}+1;\frac{p}{2};tVW\right)\right)}\right),\ (\mu,\Sigma)\in\Theta_{\lambda} \qquad (2.2.34)$$

où $t=\frac{n\lambda^{2}}{2}$, W a la même distribution que la variable $\frac{E(Z)}{Z}$ et V admet une distribution bêta $\mathcal{B}e\left(\frac{p}{2}+1;\frac{m}{2}\right)$.

De plus, on a $E(W)\geq 1$. Le cas normal correspond à $P(W=1)=1$.

Démonstration. En posant $W=\frac{E(Z)}{Z}$ dans les relations (2.2.28) et (2.2.31), on a les expressions (2.2.33) et (2.2.34).

Si $E(Z^{2})$ et $E(Z^{-2})$ existent, la covariance entre Z et Z^{-1} est négative. Donc on peut écrire la relation suivante :

$\operatorname{cov}(Z,Z^{-1})=1-E(Z)E(Z^{-1})\leq 0$ donc $E(Z)E(Z^{-1})\geq 1$.

Mais on sait par définition de W que $E(W)=E(Z)E(Z^{-1})$.

Ce qui nous donne $E(W)\geq 1$.

Le théorème suivant nous montre que le multiplicateur $\hat{k}(U)$ est un rapport de deux polynômes en V quand $\frac{n-p}{2}$ est un entier.

Théorème 2.2.18. Supposons qu'on soit sous les conditions du problème **P.2** et que $\frac{n-p}{2}$ soit un entier positif alors le meilleur estimateur équivariant est :

$$\hat{\mu}(Y,S)=\hat{k}(U)Y$$

avec

$$\hat{k}(\mathrm{U})=\frac{t}{\sqrt{n}}\frac{a(U)}{b(\mathrm{U})},\quad a(U)=\sum_{k=0}^{\ell}\binom{\ell}{k}\frac{t^k}{\Gamma\left(k+\frac{p}{2}+1\right)}\left(\frac{U}{1+U}\right)^k E\left(W^k e^{-\frac{tW}{1+U}}\right),$$

$$b(U)=\sum_{k=0}^{\ell+1}\binom{\ell+1}{k}\frac{t^k}{\Gamma\left(k+\frac{p}{2}\right)}\left(\frac{U}{1+U}\right)^k E\left(W^{k-1} e^{-\frac{tW}{1+U}}\right),\ \ell=\frac{m}{2}=\frac{n-p}{2}\ \text{et}\ t=\frac{n\lambda^2}{2}.$$

Démonstration. En utilisant les relations (2.2.22) et (2.2.28), nous avons :

$$\hat{k}(U)=\frac{\sqrt{n}\lambda^2}{2}\frac{\displaystyle\sum_{k=0}^{\ell}\binom{\ell}{k}\frac{t^k}{\Gamma\left(k+\frac{p}{2}+1\right)}\left(\frac{U}{1+U}\right)^k E\left(W^k e^{-\frac{tW}{1+U}}\right)}{\displaystyle\sum_{k=0}^{\ell+1}\binom{\ell+1}{k}\frac{t^k}{\Gamma\left(k+\frac{p}{2}\right)}\left(\frac{U}{1+U}\right)^k E\left(W^{k-1} e^{-\frac{tW}{1+U}}\right)}.$$

2.2.3 Étude du comportement du meilleur estimateur équivariant et le risque minimal pour P.1 et P.2.

Il faut remarquer que, le groupe de transformations $GL(p)$ des matrices carrées $(p \times p)$ régulières utilisé dans ce problème n'est ni moyennable, ni compact. Et en plus, l'opération de ce groupe sur IR^p n'est pas propre. Donc il nous est impossible d'appliquer les résultats du chapitre 1 pour démontrer la minimaxité et l'admissibilité de la meilleure règle équivariante. Il nous est aussi impossible de dégager les liens existants entre la meilleure règle équivariante et la règle de Bayes pour une loi *a priori*.

Perron (1987) et Marchand (1990) ont étudié le comportement du coefficient multiplicatif dans le cas où les observations seraient indépendantes et ont même distribution multinormale $N_p(\mu, I)$ sous la contrainte $\mu'\mu=\lambda^2$.

Nous allons reprendre ces approches pour étudier le comportement du multiplicateur et du risque minimal dans le cas où les observations seraient indépendantes et issues d'une distribution multinormale $N_p(\mu,\Sigma)$ avec la contrainte $\mu^t\Sigma^{-1}\mu = \lambda^2$.

Nous allons tout d'abord énoncer le lemme suivant qui servira à dégager les propriétés du multiplicateur et du risque minimal.

Lemme 2.2.19. Soient $S_a(x)=\sum_{i=0}^{\infty} a_i x^i$ et $S_b(x)=\sum_{i=0}^{\infty} b_i x^i$ deux séries de puissances en x où les suites de constantes $\{a_i\}_{i=0}^{\infty}$ et $\{b_i\}_{i=0}^{\infty}$ satisfont les trois conditions suivantes :

(i) $a_i > 0$, $b_i \geq 0$ et $a_i b_{i+1} > a_{i+1} b_i$ pour $i = 0,1,\ldots,$

(ii) les séries $S_a(x)$ et $S_b(x)$ convergent pour tout $x \geq 0$,

(iii) $\lim_{i\to\infty} \frac{b_i}{a_i} = c$, où c représente : soit une constante positive ou nulle, soit "$+\infty$".

Alors sous ces conditions

(a) le rapport $R(x)=\frac{S_b(x)}{S_a(x)}$ est une fonction strictement croissante $\forall\, x \geq 0$,

(b) $\lim_{x\to 0} R(x)=\frac{b_0}{a_0}$ et $\lim_{x\to\infty} R(x)=c$.

Démonstration. Voir Marchand (1990, page 35).

De ce lemme, on en déduit les résultats suivants.

Le prochain corollaire présente quelques propriétés de la fonction R définie par, pour $\alpha_1 > 0$, $\alpha_2 \geq 0$, $\beta_1 > 0$, $\beta_2 \geq 0$,

$$R(\alpha_1;\beta_1;\alpha_2;\beta_2;x) = \frac{{}_1F_1(\alpha_1;\beta_1;x)}{{}_1F_1(\alpha_2;\beta_2;x)}, \quad x \geq 0.$$

Corollaire 2.2.20. La fonction R vérifie les propriétés suivantes :

(a) les fonctions $R(\alpha+1;\beta;\alpha;\beta;\cdot)$ et $R(\alpha;\beta;\alpha;\beta+1;\cdot)$ sont strictement croissantes sur $(0,\infty)$,

(b) la fonction $R(\alpha+1;\beta+1;\alpha;\beta;\cdot)$ est strictement croissante sur $(0,\infty)$ si $\alpha<\beta$ et strictement décroissante si $\alpha>\beta$,

(c) les expressions $xR(\alpha+1;\beta+1;\alpha;\beta;x)$ et $xR(\alpha+1;\beta;\alpha;\beta;x)$ sont strictement croissantes en x pour $x\in(0,\infty)$.

Démonstration. Voir Marchand (1994).

(a) En prenant $a_i = \frac{1}{i!}\frac{(\alpha)_i}{(\beta)_i}$ et $b_i = \frac{1}{i!}\frac{(\alpha+1)_i}{(\beta)_i}$, on vérifie facilement les relations (i), (ii) et (iii) du lemme précédent.

Donc la fonction $R(\alpha+1;\beta;\alpha;\beta;\cdot)$ est strictement croissante sur $(0,\infty)$. Il en est de même pour la fonction $R(\alpha;\beta;\alpha;\beta+1;\cdot)$.

(b) Si on pose $a_i = \frac{1}{i!}\frac{(\alpha)_i}{(\beta)_i}$ et $b_i = \frac{1}{i!}\frac{(\alpha+1)_i}{(\beta+1)_i}$, $i>0$. Alors $a_i>0$, $b_i\geq 0$ et

$$\frac{b_{i+1}a_i}{b_i a_{i+1}} = \frac{(\alpha+i+1)}{(\beta+i+1)}\frac{(\beta+i)}{(\alpha+i)} \geq 1 \text{ si } \alpha\leq\beta \text{ et } \frac{b_{i+1}a_i}{b_i a_{i+1}} < 1 \text{ si } \alpha>\beta.$$

De plus, on vérifie facilement les conditions (ii) et (iii) du lemme précédent, ce qui implique le (b).

(c) On pose $a_i = \frac{1}{i!}\frac{(\alpha)_i}{(\beta)_i}$ et $b_i = \frac{1}{(i-1)!}\frac{(\alpha+1)_{i-1}}{(\beta+1)_{i-1}}$, $i>0$.

Alors on a

$$\frac{b_{i+1}a_i}{b_i a_{i+1}} = \frac{i+1}{i} > 1.$$

Nous vérifions facilement les conditions (ii) et (iii) du lemme précédent. Ce qui nous donne la première expression du (c). La démonstration de la deuxième expression du (c) est similaire, il suffit de poser :

$$a_i = \frac{1}{i!}\frac{(\alpha)_i}{(\beta)_i} \text{ et } b_i = \frac{1}{(i-1)!}\frac{(\alpha+1)_{i-1}}{(\beta)_{i-1}},\ i > 0.$$

Rappelons la représentation $\hat{k}(u) = \dfrac{k_1(u)}{k_2(u)}$ (voir les relations (2.2.9), (2.2.18) et (2.2.19)) où les fonctions k_1 et k_2 sont données respectivement par les relations à la page suivante :

$$k_1(u) = E_\theta\left(Y^t\mu \mid U=u\right) = \frac{n^{3/2}\lambda^2}{p}\frac{u}{1+u}\frac{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}\frac{u}{u+1}\right)}{{}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\frac{u}{u+1}\right)}, \quad (2.2.35)$$

et

$$k_2(u) = E_\theta\left(Y^tY \mid U=u\right) = n\frac{u}{1+u}\frac{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}\frac{u}{u+1}\right)}{{}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\frac{u}{u+1}\right)}. \quad (2.2.36)$$

Corollaire 2.2.21. Les fonctions $k_1(u)$ et $k_2(u)$ sont strictement croissantes sur $(0,\infty)$. Et la fonction $\hat{k}(u)$ est strictement décroissante sur $(0,\infty)$.

De plus, nous avons $\hat{k}(0) = \dfrac{\sqrt{n}\lambda^2}{p}$ et $\lim\limits_{u\to\infty}\hat{k}(u) < 1$.

Démonstration. Soient les fonctions h_1 et h_2 définies respectivement par les relations suivantes :

$$h_1(\mathrm{v}) = \frac{n^{3/2}\lambda^2}{p}\mathrm{v}\frac{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}\mathrm{v}\right)}{{}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\mathrm{v}\right)},\ \forall\, \mathrm{v} \in (0,1)$$

et

$$h_2(\mathrm{v}) = n\mathrm{v}\frac{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}\mathrm{v}\right)}{{}_1F_1\left(\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\mathrm{v}\right)},\ \forall\, \mathrm{v} \in (0,1).$$

Alors nous avons $k_1(u) = h_1\left(\frac{u}{1+u}\right)$ et $k_2(u) = h_2\left(\frac{u}{1+u}\right)$.

En utilisant le (c) du corollaire précédent, les fonctions h_1 et h_2 sont strictement croissantes sur $(0,1)$. Mais aussi la fonction $u \to \frac{u}{1+u}$ est aussi strictement croissante. Alors k_1 et k_2 sont composées de fonctions strictement croissantes. Donc elles sont strictement croissantes.

On écrit $\hat{k}(u)$ de la manière suivante :

$$\hat{k}(u) = \frac{\sqrt{n}\lambda^2}{p}\frac{R\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\frac{u}{u+1}\right)}{R\left(\frac{n}{2}+1;\frac{p}{2};\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\frac{u}{u+1}\right)}.$$

Le multiplicateur $\hat{k}(u)$ est donc le produit de deux fonctions décroissantes et positives (d'après le (a) et (b) du corollaire précédent). Elle est donc décroissante.

On a $\hat{k}(0)=\dfrac{\sqrt{n}\lambda^2}{p}\dfrac{R\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n}{2};\frac{p}{2};0\right)}{R\left(\frac{n}{2}+1;\frac{p}{2};\frac{n}{2};\frac{p}{2};0\right)}=\dfrac{\sqrt{n}\lambda^2}{p}$.

Posons

$$h_\lambda(\mathrm{v})=\frac{\sqrt{n}\lambda^2}{p}\frac{R\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\mathrm{v}\right)}{R\left(\frac{n}{2}+1;\frac{p}{2};\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\mathrm{v}\right)},\ \text{pour } \mathrm{v}\in(0,1).$$

On a alors $\lim_{u\to\infty}\hat{k}(u)=h_\lambda(1)$.

Comme $h_\lambda(1)$ est une fonction strictement croissante en λ, on a la relation suivante :

$$\lim_{u\to\infty}\hat{k}(u)<\lim_{\lambda\to\infty}h_\lambda(1).$$

Mais d'après le (b) du lemme 2.2.19 nous avons :

$$\lim_{\lambda\to\infty}\lambda^2\,R\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\right)=\frac{p}{n}$$

et

$$\lim_{\lambda\to\infty}R\left(\frac{n}{2}+1;\frac{p}{2};\frac{n}{2};\frac{p}{2};\frac{n\lambda^2}{2}\right)=1.$$

D'où nous avons les inégalités suivantes :

$$\lim_{u\to\infty}\hat{k}(u)\leq\frac{1}{\sqrt{n}}\leq 1.$$

La conjecture suivante nous permet de démontrer le théorème 2.2.23. Le théorème 2.2.23 est très important car il montre que le risque minimal est borné.

Conjecture 2.2.22. La fonction $\frac{k_1^{\,2}(u)}{k_2(u)}$, pour $u \in (0,\infty)$, est strictement croissante en u.

Théorème 2.2.23. Sous les conditions du problème, le risque minimal de la moyenne μ est compris entre $\frac{\lambda(p-1)}{n\lambda+p}$ et $\frac{\lambda p}{n\lambda+p}$.

De plus, $\lim\limits_{\lambda\to\infty} R((\mu,\Sigma),\hat{\mu}_\lambda) = \frac{p}{n}$ et $\lim\limits_{\lambda\to 0} R((\mu,\Sigma),\hat{\mu}_\lambda) = 0$.

Démonstration : le résultat est lié à la conjecture.

2.2.4 Estimation équivariante de la moyenne μ d'une distribution de Student $St_p(\mu,\Sigma,\alpha,\beta)$ sous la contrainte $\mu^t\Sigma^{-1}\mu=\lambda^2$.

Marchand (1990) étend la définition usuelle de familles de "Student" multivariées à n degrés de liberté à des familles qu'il appelle "Student" multivariées de paramètre de position μ et de paramètres de dispersion α et β. Il montre que ces dernières sont des mélanges de lois normales de paramètres $(\mu,\ (E(Z))^{-1}I,\ Z)$ où la variable aléatoire Z^{-1} est gamma $\mathcal{G}(\alpha,\beta)$. Et lorsque $\alpha=\beta=\frac{n}{2}$, on retrouve la loi usuelle de "Student" à n degrés de liberté.

On considère les familles des mélanges de lois normales de paramètres μ, Ω, Z. En utilisant l'approche de Marchand (1994), il est facile de montrer que si la variable aléatoire Z^{-1} est de loi gamma $\mathcal{G}(\alpha,\beta)$ alors nous obtenons les familles de "Student" multivariées de paramètres de position μ, de paramètre d'échelle Σ et de paramètres de dispersion α et β. Dans le reste du mémoire, la loi "Student" multivariée de paramètres μ, Σ, α et β est notée $St_p(\mu, \Sigma, \alpha, \beta)$.

Étudions le problème d'estimation de la moyenne μ d'une telle famille lorsque la contrainte est $\mu^t\Sigma^{-1}\mu = \lambda^2$. Pour ce faire, nous allons d'abord évaluer le multiplicateur du meilleur estimateur équivariant de la moyenne μ (noté $\hat{k}_{St}(u)$), ensuite le risque minimal dans la classe des estimateurs équivariants et enfin nous allons obtenir quelques propriétés du multiplicateur et du risque minimal.

Corollaire 2.2.24. Si la variable aléatoire Z^{-1} admet une distribution gamma $\mathcal{G}(\alpha, \beta)$, alors le multiplicateur du meilleur estimateur équivariant $\hat{k}_{St}(u)$ ne dépend pas du paramètre β.

Démonstration. On sait que si une variable aléatoire A admet une loi gamma $\mathcal{G}(\alpha, \beta)$ alors pour $c > 0$, cA est de loi gamma $\mathcal{G}\left(\alpha, \frac{\beta}{c}\right)$.

En utilisant la remarque 2.2.15 pour $c = \beta$, on en conclut que le multiplicateur $\hat{k}_{St}(u)$ ne dépend pas de β.

Lemme 2.2.25. Si la variable aléatoire Z^{-1} a pour loi gamma $\mathcal{G}(\alpha, \beta)$ (avec $\alpha > 1$) alors nous avons :

i) $$\hat{k}_{St}(U) = \frac{2t}{p\sqrt{n}} \frac{E^W\left(W\,{}_1F_1\left(\frac{n}{2}+1; \frac{p}{2}+1; tVW\right)\right)}{E^W\left({}_1F_1\left(\frac{n}{2}+1; \frac{p}{2}+1; tVW\right)\right)}, \tag{2.2.37}$$

ii) $$R((\mu,\Sigma),\hat{\mu}) = \lambda^2 - \frac{4t^2}{np}\left(\frac{\alpha-1}{t+\alpha-1}\right)^{\alpha-1} E\left(\frac{E^2\left(W\,{}_1F_1\left(\frac{n}{2}+1; \frac{p}{2}+1; tVW\right)\right)}{E\left({}_1F_1\left(\frac{n}{2}+1; \frac{p}{2}; tVW\right)\right)}\right)$$

où W est de loi gamma $\mathcal{G}(\alpha-1, t+\alpha-1)$, $t=\frac{n\lambda^2}{2}$ et $V=\frac{U}{1+U}$.

Démonstration. En utilisant la proposition 2.2.17, on a :

i) $$\hat{k}_{St}(u)=\frac{2t}{p\sqrt{n}}\frac{E\left(e^{-tW}{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;tvW\right)\right)}{E\left(W^{-1}e^{-tW}{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;tvW\right)\right)},$$

où W a la même loi que $\frac{E(Z)}{Z}$.

Mais si Z^{-1} est de loi gamma $\mathcal{G}(\alpha,\beta)$ alors $W=E(Z)Z^{-1}$ est de loi gamma $\mathcal{G}\left(\alpha,\frac{\beta}{\mathrm{E(Z)}}\right)$.

Donc en développant, nous obtenons la relation suivante :

$$\hat{k}_{St}(U)=\frac{2t}{p\sqrt{n}}\frac{\int_0^{+\infty}\frac{\left(\frac{\beta}{E(Z)}\right)^{\alpha}}{\Gamma(\alpha)}ww^{\alpha-2}e^{-\left(t+\frac{\beta}{E(Z)}\right)w}{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;tVw\right)dw}{\int_0^{+\infty}\frac{\left(\frac{\beta}{E(Z)}\right)^{\alpha}}{\Gamma(\alpha)}w^{\alpha-2}e^{-\left(t+\frac{\beta}{E(Z)}\right)w}{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};tVw\right)dw}.$$

Comme Z^{-1} est $\mathcal{G}(\alpha,\beta)$ alors $\mathrm{E}(Z)=\frac{\beta}{\alpha-1}$, pour $\alpha>1$.

On obtient alors le résultat suivant :

$$\hat{k}_{st}(U)=\frac{2t}{p\sqrt{n}}\frac{\int_0^{+\infty}\frac{(t+\alpha-1)^{\alpha-1}}{\Gamma(\alpha-1)}ww^{\alpha-2}e^{-(t+\alpha-1)w}{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;tVw\right)dw}{\int_0^{+\infty}\frac{(t+\alpha-1)^{\alpha-1}}{\Gamma(\alpha-1)}w^{\alpha-2}e^{-(t+\alpha-1)w}{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};tVw\right)dw}.$$

Ce qui conduit à la relation (2.2.37).

ii) En utilisant la proposition 2.2.17, on peut écrire

$$R((\mu,\Sigma),\hat{\mu}) = \lambda^2 - \frac{2t}{\sqrt{n}} E\left(\hat{k}\left(\frac{V}{1-V}\right) E\left(e^{-tW}{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;tVW\right)\right)\right).$$

Mais on a aussi la relation suivante :

$$E\left(e^{-tW}{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;tVW\right)\right) = \frac{(\alpha-1)^{\alpha-1}}{\Gamma(\alpha-1)} \int_0^{+\infty} w w^{\alpha-2} e^{-(t+\alpha-1)w}{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;tVw\right) dw.$$

$$= \frac{(\alpha-1)^{\alpha-1}}{(t+\alpha-1)^{\alpha-1}} E\left(W{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;tVw\right)\right).$$

D'où le risque est donné par la relation suivante :

$$R((\mu,\Sigma),\hat{\mu}) = \lambda^2 - \frac{4t^2}{np}\left(\frac{\alpha-1}{t+\alpha-1}\right)^{\alpha-1} E\left(\frac{E^2\left(W{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;tVW\right)\right)}{E\left({}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};tVW\right)\right)}\right).$$

Lemme 2.2.26. Soit une variable W admet une loi $\mathcal{G}(a,b)$ et soit γ_1, γ_2, γ_3 et ℓ des réels positifs avec $\gamma_3 < b$, alors

$$E\left(W^{\ell}{}_1F_1(\gamma_1;\gamma_2;\gamma_3 W)\right) = (a)_{\ell} b^{-\ell}{}_2F_1\left(\gamma_1, a+\ell;\gamma_2;\frac{\gamma_3}{b}\right). \qquad (2.2.38)$$

Démonstration. On a :

$$\mathrm{E}^{\mathrm{W}}\left(w^{\ell}{}_1F_1(\gamma_1;\gamma_2;\gamma_3 W)\right) = \sum_{k=0}^{\infty} \frac{(\gamma_1)_k}{(\gamma_2)_k} \frac{(\gamma_3)^k}{k!} \mathrm{E}^{\mathrm{W}}\left(w^{k+\ell}\right).$$

Mais on sait que :

$$E\left(W^{k+\ell}\right)=b^{-(k+\ell)}\frac{\Gamma(a+k+\ell)}{\Gamma(a)},\quad \text{pour } k+\ell>-a$$

$$=b^{-(k+\ell)}\frac{\Gamma(a+k+\ell)}{\Gamma(a+\ell)}\frac{\Gamma(a+\ell)}{\Gamma(a)}$$

$$=b^{-k-\ell}(a+\ell)_k(a)_\ell.$$

D'où nous avons :

$$\mathrm{E}\left(W^{\ell}\,{}_1F_1(\gamma_1;\gamma_2;\gamma_3W)\right)=(a)_\ell b^{-\ell}\sum_{k=0}^{\infty}\frac{(\gamma_1)_k(a+\ell)_k}{(\gamma_2)_k}\frac{\left(\frac{\gamma_3}{b}\right)^k}{k!}=(a)_\ell b^{-\ell}\,{}_2F_1\left(\gamma_1,a+\ell;\gamma_2;\frac{\gamma_3}{b}\right).$$

Théorème 2.2.27. Sous les conditions du problème et pour $\alpha>1$, le meilleur estimateur équivariant de la moyenne μ est donné par :

$$\hat{\mu}_{St}(Y,S)=\hat{k}_{St}(U)Y$$

avec

$$\hat{k}(U)=\frac{2(\alpha\text{-}1)h}{p\sqrt{n}}\frac{{}_2F_1\left(\frac{n}{2}+1,\alpha;\frac{p}{2}+1;hV\right)}{{}_2F_1\left(\frac{n}{2}+1,\alpha-1;\frac{p}{2};hV\right)},\tag{2.2.39}$$

$$h=\frac{t}{t+\alpha-1},\quad t=\frac{n\lambda^2}{2}\quad et\quad V=\frac{U}{1+U}.$$

Démonstration. Le résultat découle directement des expressions (2.2.37) et (2.2.38) (en posant $\ell=0$ et $\ell=1$ en (2.2.38)).

Théorème 2.2.28. Sous les conditions du problème et pour $\alpha>1$, le risque minimal de la moyenne μ est égal à l'expression suivante :

$$R((\mu,\Sigma),\hat{\mu})=\lambda^{2}-\frac{4t^{2}}{np}\left(\frac{\alpha-1}{t+\alpha-1}\right)^{\alpha+1}E\left(\frac{\left({}_{2}F_{1}\left(\frac{n}{2}+1,\alpha;\frac{p}{2}+1;hV\right)\right)^{2}}{{}_{2}F_{1}\left(\frac{n}{2}+1,\alpha-1;\frac{p}{2};hV\right)}\right), \quad (2.2.40)$$

où V est une variable aléatoire de densité bêta $\mathscr{B}e\left(\frac{p}{2}+1;\frac{m}{2}\right)$, $t=\frac{n\lambda^{2}}{2}$ et

$h=\dfrac{t}{t+\alpha-1}$.

Démonstration. C'est une conséquence directe du lemme 2.2.25 et du lemme 2.2.26.

Lemme 2.2.29. Si la famille des distributions est $St_{p}(\mu,\Sigma,\alpha,\beta)$ alors le meilleur estimateur équivariant pour cette classe de distribution converge vers le meilleur estimateur équivariant pour la famille des distributions normales $N_{p}(\mu,\Sigma)$ lorsque $\alpha\to\infty$.

Démonstration. Il faut remarquer que :

$$\lim_{\alpha\to\infty}(\alpha h)^{k}=\lim_{\alpha\to\infty}\left(\frac{(\alpha)n\lambda^{2}}{n\lambda^{2}+2(\alpha-1)}\right)^{k}=\left(\frac{n\lambda^{2}}{2}\right)^{k} \quad (2.2.41)$$

$$\lim_{\alpha\to\infty}\frac{(\alpha)_{k}}{\alpha^{k}}=1,\quad \ell=0,1. \quad (2.2.42)$$

En utilisant le théorème de la convergence dominée et les relations (2.2.41) et (2.2.42), nous avons l'égalité suivante :

$$\lim_{\alpha\to\infty} {}_2F_1\left(\frac{n}{2}+1,\alpha;\frac{p}{2}+1;hV\right)=\sum_{k=0}^{\infty}\lim_{\alpha\to\infty}\frac{\left(\frac{n}{2}+1\right)_k}{\left(\frac{p}{2}+1\right)_k}\frac{(\alpha)_k}{\alpha^k}(\alpha h)^k\frac{V^k}{k!}={}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}V\right).$$

(2.2.43)

De même on montre la relation suivante :

$$\lim_{\alpha\to\infty} {}_2F_1\left(\frac{n}{2}+1,\alpha-1;\frac{p}{2};hV\right)={}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}V\right). \quad (2.2.44)$$

En utilisant les relations (2.2.39), (2.2.43) et (2.2.44), on a le résultat.

Lemme 2.2.30. Si la famille des distributions est $St_p(\mu,\Sigma,\alpha,\beta)$ alors le risque minimal dans la classe des estimateurs équivariants est asymptotiquement égal au risque minimal pour la famille de distributions normales (Problème **P.1**).

Démonstration. La démonstration est similaire à celle du lemme précédent.

Propriétés 2.2.31. Si la famille des distributions est $St_p(\mu,\Sigma,\alpha,\beta)$ alors, pour $\alpha>1$, nous avons les résultats suivants :

(a) le multiplicateur $\hat{k}_{St}(u)$ est décroissant en u si $\alpha\geq\frac{p}{2}+1$ et croissante en u sinon,

(b) $\hat{k}_{st}(0)=\frac{2(\alpha\text{-}1)h}{p\sqrt{n}}$.

Démonstration.

(a) En utilisant une propriété des fonctions hypergéométriques voir Abramowitz et Stegun (1964, relation (15.2.17), page 558), on a la relation suivante :

$$\left(\frac{p}{2}+1-\alpha\right)_2F_1\left(\frac{n}{2}+1,\alpha-1;\frac{p}{2}+1,hV\right)+(\alpha-1)_2F_1\left(\frac{n}{2}+1,\alpha;\frac{p}{2}+1,hV\right)$$
$$-\frac{p}{2}\,{}_2F_1\left(\frac{n}{2}+1,\alpha-1;\frac{p}{2},hV\right)=0.$$

Cette relation est équivalente à :

$$\frac{{}_2F_1\left(\frac{n}{2}+1,\alpha;\frac{p}{2}+1,hV\right)}{{}_2F_1\left(\frac{n}{2}+1,\alpha-1;\frac{p}{2},hV\right)}=\frac{\frac{p}{2}}{\alpha-1}+\frac{\alpha-1-\frac{p}{2}}{\alpha-1}\frac{{}_2F_1\left(\frac{n}{2}+1,\alpha-1;\frac{p}{2}+1,hV\right)}{{}_2F_1\left(\frac{n}{2}+1,\alpha-1;\frac{p}{2},hV\right)}.$$

Mais en vertu de la relation (2.2.39), nous avons l'expression suivante :

$$\hat{k}_{St}(u)=\frac{2(\alpha\text{-}1)h}{p\sqrt{n}}\left(\frac{\frac{p}{2}}{\alpha-1}+\frac{\alpha-1-\frac{p}{2}}{\alpha-1}\frac{{}_2F_1\left(\frac{n}{2}+1,\alpha-1;\frac{p}{2}+1,h\text{v}\right)}{{}_2F_1\left(\frac{n}{2}+1,\alpha-1;\frac{p}{2},h\text{v}\right)}\right). \quad (2.2.45)$$

On avait montré par le corollaire 2.2.21 que le multiplicateur du meilleur estimateur équivariant de la moyenne μ pour le cas d'une loi multinormale, donné par $\hat{k}(u)=\frac{\sqrt{n}\lambda^2}{2}\frac{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1,\frac{n\lambda^2}{2}\text{v}\right)}{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2},\frac{n\lambda^2}{2}\text{v}\right)}$, est strictement décroissante en u.

De la même manière on montre que $\frac{{}_2F_1\left(\frac{n}{2}+1,\alpha;\frac{p}{2}+1,h\text{v}\right)}{{}_2F_1\left(\frac{n}{2}+1,\alpha-1;\frac{p}{2},h\text{v}\right)}$ est strictement décroissant en u.

D'où en utilisant la relation (2.2.45), le multiplicateur est décroissant en u si $\alpha\geq\frac{p}{2}+1$ et croissante en u sinon.

(b) La démonstration est similaire à celle du corollaire 2.2.21.

Proposition 2.2.32. Sous les conditions du problème, si $\alpha = \frac{p}{2}+1$ alors on a les relations suivantes :

$$\hat{\mu}_{St}(Y,S) = \frac{n\lambda^2}{n\lambda^2 + p}\overline{X}$$

et

$$R((\mu,\Sigma),\hat{\mu}_{St}) = \frac{p\lambda^2}{n\lambda^2 + p}.$$

C'est-à-dire le meilleur estimateur équivariant est linéaire en $\overline{X}$.

Démonstration. En remplaçant α par $\frac{p}{2}+1$ dans (2.2.40), on obtient la relation suivante :

$$R((\mu,\Sigma),\hat{\mu}) = \lambda^2 - \frac{n\lambda^4}{p}\left(\frac{p}{n\lambda^2 + p}\right)^{\alpha+1} E\left({}_2F_1\left(\frac{n}{2}+1,\frac{p}{2}+1;\frac{p}{2}+1;hV\right)\right).$$

D'après une propriété des fonctions hypergéomètriques Abramowitz et Stegun (1964, relation (15.1.8), page 556), on a :

$${}_2F_1\left(\frac{n}{2}+1,\frac{p}{2}+1;\frac{p}{2}+1;hV\right) = {}_1F_0\left(\frac{n}{2}+1;hV\right) = (1\text{-}hV)^{-\left(\frac{n}{2}+1\right)}.$$

Donc nous avons l'égalité suivante :

$$E\left({}_2F_1\left(\frac{n}{2}+1,\frac{p}{2}+1;\frac{p}{2}+1;hV\right)\right) = \frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{p}{2}+1\right)\Gamma\left(\frac{n-p}{2}\right)}\int_0^1 \mathrm{v}^{\frac{p}{2}}(1-\mathrm{v})^{\frac{n-p}{2}}(1-h\mathrm{v})^{-\left(\frac{n}{2}+1\right)}d\mathrm{v}.$$

Mais en utilisant l'égalité précédente et une relation sur les représentations de l'intégrale des fonctions hypergéométriques voir Abramawitz et Stegun (1964, relation (15.3.1), page 558), on a :

$$E\left({}_2F_1\left(\frac{n}{2}+1,\frac{p}{2}+1;\frac{p}{2}+1;h\mathrm{V}\right)\right)={}_2F_1\left(\frac{n}{2}+1,\frac{p}{2}+1;\frac{n}{2}+1;h\right)=(1-h)^{-\left(\frac{p}{2}+1\right)}=\left(\frac{p}{n\lambda^2+p}\right)^{-\left(\frac{p}{2}+1\right)}.$$

D'où on en déduit la relation suivante :

$$R((\mu,\Sigma),\hat{\mu})=\lambda^2-\frac{n\lambda^4}{n\lambda^2+p}.$$

Ce qui complète la démonstration.

Remarque 2.2.33. Sous les conditions du théorème précédent, on verra dans le chapitre 3 que le meilleur estimateur équivariant coïncide avec le meilleur estimateur linéaire (c'est l'estimateur linéaire qui rend le risque le plus petit possible dans la classe des estimateurs linéaires).

2.3. ESTIMATION ÉQUIVARIANTE DE LA MOYENNE μ D'UNE DISTRIBUTION DE MÉLANGES DE LOIS MULTINORMALES $N_p\left(\mu,\left(\frac{\mu^t\mu}{C^2}\right)I\right)$, OÙ C, LE COEFFICIENT DE VARIATION, EST CONNU.

Soient $X_1,X_2,\ldots,X_n$ des observations issues d'une loi de probabilité commune de moyenne μ et de matrice variance-covariance $\frac{\mu^t\mu}{C^2}I$. Considérons la fonction de perte suivante :

$$L((\mu,\Sigma),d)=C^2\frac{(d-\mu)^t(d-\mu)}{\mu^t\mu} \tag{2.3.1}$$

On veut estimer la moyenne μ avec la fonction de perte (2.3.1).

On pose $Y=\sqrt{n}\ \overline{X}$ et $W=tr(S)$.

Considérons le groupe de transformations $G=\mathrm{IR}^{+}\mathrm{x}\,O(p)$, où IR^{+} est le groupe des réels positifs et $O(p)$ est le groupe de transformations des matrices carrées orthogonales, transformant :

$$X_i \to b\Gamma X_i\text{, pour } i=1,\ldots,n,\ b\in \mathrm{IR}^{+} \text{ et } \Gamma\in O(\mathrm{p}).$$

La transformation induite par le groupe G sur l'espace (Y,W) donne :

$$(Y,W)\to(b\Gamma Y,b^2W),\ b\in \mathrm{IR}^{+} \text{ et } \Gamma\in O(p).$$

Le groupe induit $\mathrm{IR}^{+}\mathrm{x}\,O(p)$, par $\mathrm{IR}^{+}\mathrm{x}\,O(p)$, dans l'espace des paramètres transforme :

$$\left(\mu,\left(\frac{\mu'\mu}{C^2}\right)I\right)\to\left(b\Gamma\mu,b^2\left(\frac{\mu'\mu}{C^2}\right)I\right),\ b\in \mathrm{IR}^{+} \text{ et } \Gamma\in O(\mathrm{p}).$$

Sous la fonction de perte (2.3.1), ce problème devient invariant. Le groupe $\mathrm{IR}^{+}\mathrm{x}\,O(p)$, opère transitivement sur l'espace paramétrique.

La statistique $U_1=\dfrac{Y'Y}{W}$ est invariante maximale sur l'espace basé sur (Y,W) et la correspondance de U dans l'espace des paramètres est donnée par :

$$\mu'\left(\left(\frac{\mu'\mu}{C^2}\right)I\right)^{-1}\mu=C^2.$$

Comme G opère transitivement sur l'espace paramétrique, la fonction de risque pour un estimateur équivariant δ est constante :

$$R\left(\left(\mu,\frac{\mu'\mu}{C^2}I\right),\delta\right)=R((\mu_0,I),\delta)\text{, pour } \mu_0=(C,0,0,\ldots,0)' \text{ et } \forall\mu\in \mathrm{IR}.$$

Le meilleur estimateur équivariant basé sur (Y,W) est obtenu en remplaçant dans l'expression (2.2.3), U par U_1. Et on a :

$$\hat{\mu}(Y,W)=\hat{k}_1(u)Y \text{ avec } \hat{k}_1(u)=\frac{E_{\mu_0}\left(Y^t\mu_0 \mid U_1=u\right)}{E_{\mu_0}\left(Y^tY \mid U_1=u\right)}.$$

Nous avons donc besoin de la densité conjointe de Y et U_1 pour calculer le multiplicateur $\hat{k}_1(u)$.

Traitons d'abord le cas particulier où la famille des lois de X est celle des lois de probabilité multinormales $N_p\left(\mu,\frac{\mu^t\mu}{C^2}I\right)$.

2.3.1 Estimation de la moyenne μ d'une loi multinomale de paramètre $\left(\mu,\frac{\mu_t\mu}{C_2}I\right)$.

Le problème de l'estimation de la moyenne μ (positive) d'une distribution normale $N\left(\mu,\sigma^2\right)$ $(p=1)$, lorsque le coefficient de variation $\frac{\sigma}{\mu}$ est connu, a été abordé par Fisher (1956). Il a été repris par plusieurs auteurs dont Cox et Hinkley (1974), Efron (1975) et Kariya (1989). La généralisation à $p\geq 1$ de ce problème a été traitée par Perron (1987) et reprise par Kariya, Perron et Giri (1988).

Nous reprenons l'approche de Perron (1987) pour trouver le meilleur estimateur équivariant de la moyenne μ d'une loi $N_p\left(\mu,\frac{\mu^t\mu}{C^2}I\right)$, C où est le coefficient de variation, avec la fonction de perte (2.3.1). Ensuite nous déterminerons le risque minimal dans la classe des estimateurs équivariants. Notons **P.3** ce problème.

La statistique(Y, W) est minimale et exhaustive pour ce problème.

Pour évaluer le meilleur estimateur équivariant, nous avons besoin de la distribution conjointe de Y et U_1. Mais avant tout, nous allons énoncer un résultat sur les distributions de Wishart.

Lemme 2.3.1 Si une variable aléatoire A admet une loi Wishart $W_p(m)$, alors la trace de A (notée $tr(A)$) admet une loi de probabilité khi-deux χ^2_{mp}.

Démonstration :

Par définition d'une distribution de Wishart $W_p(m)$, on a :

$A = ZZ^t = \sum_{i=1}^{m} Z_i Z_i^t$, où les Z_i sont indépendantes de lois $N_p(0, I)$.

Donc la trace de A est $tr(A) = \sum_{i=1}^{m} Z_i Z_i^t = \sum_{i=1}^{m} Z_i^t Z_i$.

Comme Z_i est de loi normale $N_p(0, I)$, par les propriétés des formes quadratiques d'une loi normale, $Z_i^t Z_i$ est χ^2_p.

D'où la loi de $tr(A)$ est χ^2_{mp} (par propriétés des distributions de khi-deux).

Lemme 2.3.2. La densité de conjointe de Y et U_1 est donnée par :

$$f_{Y,U_1}(y,u) = q(C)u^{-\left(\frac{m_1}{2}+1\right)}(y^t y)^{\frac{m_1}{2}} e^{\sqrt{n}\frac{C^2}{\mu^t \mu}\mu^t y} e^{-\frac{1}{2}\frac{u+1}{u}\frac{C^2}{\mu^t \mu}y^t y} \tag{2.3.2}$$

où $q(C) = (2)^{\frac{np}{2}}(\pi)^{\frac{p}{2}}\Gamma\left(\frac{m_1}{2}\right)e^{\frac{nC^2}{2}}\left(\frac{C^2}{\mu^t \mu}\right)^{\frac{np}{2}}$.

Démonstration. Comme la variable aléatoire $\left(\frac{\mu^t\mu}{C^2}\right)^{-1} S$ admet la loi de probabilité $W_p(n-1)$ alors d'après le lemme précédent $W_1 = \left(\frac{\mu^t\mu}{C^2}\right)^{-1} tr(S)$ admet une distribution de khi-deux $\chi^2_{(n-1)p}$.

De plus, elle est indépendante de Y (car Y et S sont des variables aléatoires indépendantes).

Donc la densité conjointe de Y et U_1 est obtenue en remplaçant dans la relation (2.2.4) $m = n - p$ par $m_1 = np - p$, λ^2 par C^2 et Σ par $\frac{\mu^t\mu}{C^2} I$.

On obtient alors le résultat suivant :

$$f_{Y,U_1}(y,u) = q(C) u^{-\left(\frac{m_1}{2}+1\right)} (y^t y)^{\frac{m_1}{2}} e^{\sqrt{n}\frac{C^2}{\mu^t\mu}\mu^t y} e^{-\frac{1}{2}\frac{u+1}{u}\frac{C^2}{\mu^t\mu} y^t y}.$$

D'où on a le résultat.

Théorème 2.3.3. Sous les conditions du problème **P.3**, le meilleur estimateur équivariant de la moyenne μ est donné par :

$$\hat{\mu}(Y,S) = \hat{k}(U_1)Y$$

avec

$$\hat{k}(U_1) = \frac{\sqrt{n}C^2}{p} \frac{{}_1F_1\left(\frac{np}{2}+1; \frac{p}{2}+1; \frac{nC^2}{2}\frac{U_1}{U_1+1}\right)}{{}_1F_1\left(\frac{np}{2}+1; \frac{p}{2}; \frac{nC^2}{2}\frac{U_1}{U_1+1}\right)}. \tag{2.3.3}$$

Démonstration. On remplace dans la relation (2.2.8), m par m_1, λ par C et U par U_1.

On fait de même pour trouver le risque minimal. On remplace dans l'expression (2.2.16), m par m_1 et λ par C. On obtient alors :

$$R(\mu,\hat{\mu})=C^2-\frac{nC^4}{p}e^{-\frac{nC^2}{2}}E\left(\frac{{}_1F_1{}^2\left(\frac{np}{2}+1;\frac{p}{2}+1;\frac{nC^2}{2}V\right)}{{}_1F_1\left(\frac{np}{2}+1;\frac{p}{2};\frac{nC^2}{2}V\right)}\right)$$

où V est une variable aléatoire de densité $\mathcal{B}e\left(\frac{p}{2}+1;\frac{m_1}{2}\right)$.

Théorème 2.3.4. Si on suppose que $\ell_1=\frac{m_1}{2}=\frac{(n-1)p}{2}$ est un entier positif alors sous les conditions du problème, le meilleur estimateur équivariant de la moyenne μ est donné par :

$$\hat{\mu}(Y,S)=\hat{k}(U_1)Y$$

avec

$$\hat{k}(U_1)=\frac{\sqrt{n}C^2}{2}\frac{a(U_1)}{b(U_1)},$$

$$a(U_1)=\sum_{k=0}^{\ell_1}\binom{\ell_1}{k}\frac{\left(\frac{nC^2}{2}\frac{U_1}{1+U_1}\right)^k}{\Gamma\left(\frac{p}{2}+k+1\right)}\text{ et }b(U_1)=\sum_{k=0}^{\ell_1+1}\binom{\ell_1+1}{k}\frac{\left(\frac{nC^2}{2}\frac{U_1}{1+U_1}\right)^k}{\Gamma\left(\frac{p}{2}+k\right)}.$$

Démonstration. En utilisant la relation (2.3.3) et le théorème 2.2.8, on obtient le résultat.

Reprenons à présent l'approche du problème **P.1** mais dans le cas plus général où la famille des distributions de X est formée d'un mélange de lois normales.

2.3.2 Estimation de la moyenne μ d'un mélange de lois multinormales de paramètres $\left(\mu, \frac{\mu^t \mu}{C^2 E(Z)} I, Z\right)$.

On veut à présent généraliser le problème précédent (problème **P.3**) en considérant les mélanges de lois normales. On note par **P.4**, ce problème.

On considère $X_1, X_2, \dots, X_n$ des observations présentées comme ci-dessus. On adopte les mêmes notations que le problème **P.3**.

On veut estimer la moyenne μ avec la fonction de perte (2.3.1)

En utilisant le lemme 2.2.12, on a $\frac{\mu^t \mu}{C^2} I = E(Z)\Omega$ donc $\Omega = (E(Z))^{-1} \frac{\mu^t \mu}{C^2} I$.

D'où les distributions de Y et S conditionnellement à $Z = z$ deviennent respectivement $N_p\left(\mu, \frac{z}{E(Z)} \frac{\mu^t \mu}{C^2} I\right)$ et $W_p\left(n-1, \frac{z}{E(Z)} \frac{\mu^t \mu}{C^2} I\right)$.

D'après des résultats antérieurs, la densité de $(Y, U_1 \mid Z = z)$ est obtenue en remplaçant dans (2.3.2), C^2 par $\frac{E(Z)}{z} C^2$. Et on a :

$$f^{Y,U_1|Z=z}(y,u) = q e^{\frac{nC^2}{2}\frac{E(Z)}{z}} \left(\frac{E(Z)}{Z}\frac{C^2}{\mu^t \mu}\right)^{\frac{np}{2}} u^{-\left(\frac{m_1}{2}+1\right)} (y^t y)^{\frac{m_1}{2}}$$

$$\text{x}\, e^{\sqrt{n}\frac{E(Z)}{z}\frac{C^2}{\mu^t \mu}\mu^t y}\, e^{-\frac{1}{2}\frac{E(Z)}{z}\frac{u+1}{u}\frac{C^2}{\mu^t \mu} y^t y},$$

avec

$$q = \frac{(2)^{-\frac{np}{2}}(\pi)^{-\frac{p}{2}}}{\Gamma\left(\frac{m_1}{2}\right)}.$$

Théorème 2.3.5. Sous les conditions du problème **P.4**, le meilleur estimateur équivariant est donné par :

$$\hat{k}(U_1) = \frac{\sqrt{n}C^2}{p} \frac{\int_0^{+\infty} e^{-\frac{nC^2}{2}\frac{E(Z)}{Z}} {}_1F_1\left(\frac{np}{2}+1; \frac{p}{2}+1; \frac{nC^2}{2}\frac{E(Z)}{Z}\frac{U_1}{1+U_1}\right) H(dz)}{\int_0^{+\infty} z e^{-\frac{nC^2}{2}\frac{E(Z)}{Z}} {}_1F_1\left(\frac{np}{2}+1; \frac{p}{2}; \frac{nC^2}{2}\frac{E(Z)}{Z}\frac{U_1}{1+U_1}\right) H(dz)}. \quad (2.3.4)$$

Démonstration. Ce résultat est obtenu en remplaçant dans (2.2.28), m par m_1 et λ par C.

On fait de même pour trouver le risque minimal.

Théorème 2.3.6. Le risque minimal du problème **P.4** dans la classe des estimateurs équivariants est égal à :

$$R(\mu_0, \hat{\mu}) = C^2 - \frac{nC^4 E(Z)}{p} E\left(\frac{\left(\int_0^{+\infty} e^{-\frac{nC^2}{2}\frac{E(Z)}{Z}} {}_1F_1\left(\frac{np}{2}+1; \frac{p}{2}+1; \frac{nC^2}{2}\frac{E(Z)}{Z} \mathrm{V}\right) H(dz) \right)^2}{\int_0^{+\infty} z e^{-\frac{nC^2}{2}\frac{E(Z)}{Z}} {}_1F_1\left(\frac{np}{2}+1; \frac{p}{2}+1; \frac{nC^2}{2}\frac{E(Z)}{Z} \mathrm{V}\right) H(dz)} \right) \text{où} \quad (2.3.5)$$

V est une variable aléatoire de densité $\mathcal{B}e\left(\frac{p}{2}+1; \frac{m}{2}\right)$.

Démonstration On remplace dans l'expression (2.2.31), m par m_1 et λ par C.

Remarque 2.3.7. De façon similaire à la proposition 2.2.17, les relations (2.3.4) et (2.3.5) peuvent s'écrire respectivement :

$$\hat{k}(u) = \frac{\sqrt{n}C^2}{p} \frac{E\left(e^{-\frac{nC^2}{2}W} {}_1F_1\left(\frac{n}{2}+1; \frac{p}{2}+1; tVW\right)\right)}{E\left(W^{-1}e^{-\frac{nC^2}{2}W} {}_1F_1\left(\frac{n}{2}+1; \frac{p}{2}; tVW\right)\right)},$$

et

$$R(\mu_0, \hat{\mu}) = \lambda^2 - \frac{nC^4}{p} E\left(\frac{E^2\left(e^{-\frac{nC^2}{2}W} {}_1F_1\left(\frac{n}{2}+1; \frac{p}{2}+1; tVW\right)\right)}{E\left(W^{-1}e^{-\frac{nC^2}{2}W} {}_1F_1\left(\frac{n}{2}+1; \frac{p}{2}; tVW\right)\right)}\right)$$

où $t = \frac{nC^2}{2}$, W admet la même distribution que $\frac{E(Z)}{Z}$ et V admet la densité Bêta $\mathscr{B}\mathrm{e}\left(\frac{\mathrm{p}}{2}+1; \frac{m}{2}\right)$.

De plus, on a $E(W) \geq 1$. Le cas normal correspond à $P(W=1)=1$.

Théorème 2.3.7. Si on est sous les conditions du problème et si $\frac{np}{2}$ est un entier positif alors le meilleur estimateur équivariant est :

$$\hat{\mu}(Y, S) = \hat{k}(U_1)Y$$

où

$$\hat{k}(U) = \frac{t}{\sqrt{n}} \frac{a(U_1, C)}{b(U_1, C)},$$

$$a(U_1, C) = \sum_{k=0}^{\ell_1+1} \binom{\ell_1}{k} \frac{t^k}{\Gamma\left(k+\frac{p}{2}+1\right)} \left(\frac{U_1}{1+U_1}\right)^k E\left(W^k e^{-\frac{tW}{1+U_1}}\right),$$

$$b(U_1,C)=\sum_{k=0}^{\ell_1+1}\binom{\ell_1+1}{k}\frac{t^k}{\Gamma\left(k+\frac{p}{2}\right)}\left(\frac{U_1}{1+U_1}\right)^k E\left(W^{k-1}e^{-\frac{tW}{1+U_1}}\right),\ \ell_1=\frac{m_1}{2}\text{ et }t=\frac{nC^2}{2}.$$

CHAPITRE 3

ÉTUDE D'AUTRES ESTIMATEURS ÉQUIVARIANTS.

Dans cette partie, nous reprenons l'approche de Marchand (1990, 1994) pour étudier et comparer quelques sous-classes de la classe des estimateurs équivariants dans le cas plus général où la famille des mesures de probabilité de X est un mélange de lois normales de paramètres (μ, Σ, Z), avec Z une variable aléatoire positive dont l'espérance existe. Les sous-classes étudiées sont la classe des estimateurs linéaires, la classe des estimateurs du type James-Stein et l'estimateurs du maximum de vraisemblance. On note par H la fonction de répartition de la variable Z.

Nous étudierons d'abord deux cas.

– Le premier cas est celui où l'information paramétrique est $\mu^t \Sigma^{-1} \mu = \lambda_0^2$.

– Le second cas est celui où l'information paramétrique est erronée c'est-à-dire $\mu^t \Sigma^{-1} \mu \neq \lambda_0^2$.

Dans chaque cas, nous évaluerons pour chacune des classes (citées ci-dessus), le meilleur estimateur équivariant, le risque d'un estimateur ainsi le risque minimal dans la classe considérée.

Ensuite une étude comparative et numérique sera présentée dans la dernière section.

3.1. ÉTUDE DE QUELQUES CLASSES D'ESTIMATEURS DE LA MOYENNE μ D'UNE DISTRIBUTION DE MÉLANGES DE LOIS NORMALES SOUS $\mu' \Sigma^{-1} \mu = \lambda^2$.

Sous la contrainte $\mu' \Sigma^{-1} \mu = \lambda^2$, nous étudierons d'abord la sous-classe des estimateurs linéaires, ensuite celle des estimateurs du maximum de vraisemblance et enfin celle des estimateurs de James-Stein.

3.1.1. Étude de la sous-classe des estimateurs linéaires de la moyenne μ d'une distribution de mélange de lois normale de paramètres (μ, Ω, Z) sous $\mu' \Sigma^{-1} \mu = \lambda^2$ avec $\Sigma = E(Z)\Omega$.

Soit $X_1, X_2, \ldots, X_n$ des variables aléatoires indépendantes de moyenne μ et de matrice de variance-covariance Σ. On veut étudier la classe des estimateurs linéaires lorsque le paramètre $\theta = (\mu, \Sigma)$ appartient à l'espace donné par l'expression suivante :

$$\Theta_\lambda = \{\theta = (\mu, \Sigma) \in \Theta : \mu' \Sigma^{-1} \mu = \lambda^2\}. \tag{3.1.1}$$

Ces estimateurs sont de la forme $\delta(X) = k\overline{X}$, colinéaire au vecteur observé $\overline{X}$ et dont la constante multiplicative ne dépend que de l'information paramétrique $\lambda^2 = \mu' \Sigma^{-1} \mu$.

Il est certain que ces estimateurs sont équivariants (en considérant la fonction $k_1(x)=k$ dans la proposition 2.2.11). On note par Δ_L cette sous-classe. Elle est donnée par :

$$\Delta_L = \{\delta \in \Delta_E : \delta_k(x) = kx, k \in IR\}.$$

La fonction de risque d'un estimateur δ_k est constante pour tout $(\mu, \Sigma) \in \Theta_\lambda$.

Théorème 3.1.1. Sous les conditions du problème, la fonction de risque d'un estimateur linéaire δ_k est donnée par :

$$R(\theta, \delta_k) = \lambda^2 - 2k\lambda^2 + k^2\left(\frac{p}{n} + \lambda^2\right) \qquad (3.1.2).$$

Démonstration. La fonction de risque d'un estimateur linéaire δ_k, $k \in IR$, évaluée en $\theta = (\mu, \Sigma)$ est :

$$R(\theta, \delta_k) = E\left\{\left(\mu - k\overline{X}\right)^t \Sigma^{-1}\left(\mu - k\overline{X}\right)\right\}.$$

En développant la formule précédente, on a :

$$R(\theta, \delta_k) = \lambda^2 - 2k\lambda^2 + k^2 E\left(\overline{X}^t \Sigma^{-1} \overline{X}\right). \qquad (3.1.3)$$

Mais nous avons les égalités suivantes

$$E\left(\overline{X}^t \Sigma^{-1} \overline{X}\right) = E\left(tr\left(\overline{X}^t \Sigma^{-1} \overline{X}\right)\right) = E\left(tr\left(\Sigma^{-1} \overline{X}\,\overline{X}^t\right)\right) = tr\left(\Sigma^{-1} E\left(\overline{X}\,\overline{X}^t\right)\right).$$

On sait que :

$$\operatorname{cov}(\overline{X}) = E\left(\overline{X}\,\overline{X}^t\right) - \mu\mu^t = \frac{\Sigma}{n},$$

alors la relation précédente devient :

$$E\left(\overline{X}^t\Sigma^{-1}\overline{X}\right)=tr\left(\Sigma^{-1}\left(\frac{\Sigma}{n}+\mu^t\mu\right)\right)=\frac{p}{n}+\lambda^2. \qquad (3.1.5)$$

En utilisant les relations (3.1.3) et (3.1.5), la fonction de risque devient :

$$R(\theta,\delta_k)=\lambda^2-2k\lambda^2+k^2\left(\frac{p}{n}+\lambda^2\right).$$

On veut à présent rechercher le meilleur estimateur linéaire (optimal) c'est-à-dire l'estimateur qui rend la fonction de risque minimale sur Δ_L.

Corollaire 3.1.2. Sous les conditions du problème, le meilleur estimateur linéaire (MEL) dans Δ_L est donné par :

$$\delta_{k^*(\lambda)}(\overline{X})=k^*(\lambda)\overline{X} \text{ où } k^*(\lambda)=\frac{\lambda^2}{\lambda^2+\frac{p}{n}}.$$

Démonstration. En utilisant la relation (3.1.2), on a :

$$\inf_{\delta_k\in\Delta_L} R(\theta_0,\delta_k)=\min_{k\in IR}\left(\lambda^2-2k\lambda^2+k^2\left(\frac{p}{n}+\lambda^2\right)\right).$$

Mais le minimum de cette fonction est atteint en $k^*(\lambda)=\dfrac{\lambda^2}{\lambda^2+\frac{p}{n}}$.

Corollaire 3.1.3. Sous les conditions du problème, le risque minimal linéaire dans Δ_L est donné par :

$$R(\theta_0,\delta_{k^*(\lambda)})=\frac{p\lambda^2}{n\lambda^2+p}.$$

Démonstration. Le risque minimal est obtenu en remplaçant la valeur de $k^*(\lambda)$ dans la relation (3.1.2).

On en déduit des résultats de cette sous-section que le meilleur estimateur de la moyenne μ d'un mélange de lois normales de paramètres (μ,Ω,Z), dans la classe des estimateurs linéaires, ne dépend pas de la variable du mélange Z.

De plus, les résultats obtenus dans cette sous-section sont valables si on avait une distribution elliptique au lieu d'un mélange de lois normales.

3.1.2. Étude de la classe des estimateurs de type James-Stein.

Soit $X_1, X_2,..., X_n$ des observations indépendantes, admettant la même distribution de mélanges de lois normales de paramètres μ,Ω,Z. On suppose que les observations conditionnellement à $Z=z$ soient indépendantes et de loi multinormale $N_p(\mu, z\Omega)$.

On note par Σ, la matrice de variance et covariance de X.

On veut étudier la classe des estimateurs de type James-Stein, lorsque la contrainte $\mu'\Sigma^{-1}\mu = \lambda^2$, λ^2 connu. On considère la classe Δ_{JS} donnée par :

$$\Delta_{JS} = \left\{ \delta : IR^p \text{x} S^+(p) \to IR^p \text{ avec } \delta(\overline{X},S) = \left(1 - \frac{c}{\overline{X}^t S^{-1}\overline{X}}\right)\overline{X}, c \in IR \right\}$$

Dans cette classe, un élément de Δ_{JS} est noté par δ_c avec,

$$\delta_c(\overline{X},S) = \left(1 - \frac{c}{\overline{X}^t S^{-1}\overline{X}}\right)\overline{X},\ c \in IR.$$

Posons $U = n\overline{X}^t S^{-1}\overline{X}$. Il est clair que Δ_{JS} est une sous-classe de la classe des estimateurs équivariants, car $\delta_c(\overline{X},S) = k(U)\overline{X},\ c \in IR\ et\ k(U) = 1 - \frac{c}{U}$.

Donc la fonction de risque d'un estimateur linéaire δ_c, est constante sur l'espace $\Theta_\lambda = \{(\mu, \Sigma) \in IR^p \text{x}\, S(p) : \mu^t \Sigma^{-1} \mu = \lambda^2\}$.

On a besoin des lemmes suivants pour évaluer la fonction de risque d'un estimateur de type James-Stein.

Lemme 3.1.4. Si X est une variable aléatoire de loi $N_p(\theta, I)$, alors on a :

(a) $$E_\theta\left(\frac{1}{X^t X}\right) = E_\theta\left(\frac{1}{p + 2K - 2}\right),$$

(b) $$E_\theta\left(\frac{(X-\theta)^t X}{X^t X}\right) = (p-2)E_\theta\left(\frac{1}{p + 2K - 2}\right),$$

où K est une variable aléatoire de Poisson avec paramètre de moyenne égal à $\dfrac{\theta^t \theta}{2}$.

Démonstration. Voir Muirhead (1982, page 124) parmi d'autres.

Remarque 3.1.5. Si X une variable aléatoire admettant la loi $N_p(\mu, \Sigma)$, avec $(\mu, \Sigma) \in \Theta_\lambda$, on a :

(a) $$E_{(\mu,\Sigma)}\left(\frac{1}{X^t \Sigma^{-1} X}\right) = E\left(\frac{1}{p + 2K - 2}\right),$$

(b) $$E_{(\mu,\Sigma)}\left(\frac{(X-\mu)^t \Sigma^{-1} X}{X^t \Sigma^{-1} X}\right) = (p-2)E\left(\frac{1}{p + 2K - 2}\right),$$

où K est une variable aléatoire de Poisson de moyenne égale à $\dfrac{\lambda^2}{2}$.

Lemme 3.1.6. Dans les conditions du problème, pour $\theta = (\mu, \Sigma) \in \Theta_\lambda$ et $Y = \sqrt{n}\, \bar{X}$, on a les résultats suivants :

i. $$E_\theta\left(\frac{\left(Y-\sqrt{n}\mu\right)^t\Sigma^{-1}Y}{Y^tS^{-1}Y}\right)=(n-p)(p-2)E\left(W^{-1}E\left(\frac{1}{p+2M-2}\mid W\right)\right),$$

ii. $$E_\theta\left(\frac{Y^t\Sigma^{-1}Y}{\left(Y^tS^{-1}Y\right)^2}\right)=(n-p)(n-p+2)E\left(W^{-1}E\left(\frac{1}{p+2M-2}\mid W\right)\right),$$

où W est une variable aléatoire de même loi que $\frac{E(Z)}{Z}$ et M est une variable aléatoire discrète dont la loi conditionnelle étant donné l'événement $W=w$ est une loi de Poisson de moyenne égale à $\frac{n\lambda^2}{2}w$.

Démonstration.

i. En utilisant le lemme 2.2.12, on a la relation suivante :

$$E_\theta\left(\frac{\left(Y-\sqrt{n}\mu\right)^t\Sigma^{-1}Y}{Y^tS^{-1}Y}\right)=E_\theta\left(\frac{Z}{E(Z)}E\left(\frac{\left(Y-\sqrt{n}\mu\right)^t(Z\Omega)^{-1}Y}{Y^tS^{-1}Y}\mid Z\right)\right)$$

On peut également écrire

$$E_\theta\left(\frac{\left(Y-\sqrt{n}\mu\right)^t\Sigma^{-1}Y}{Y^tS^{-1}Y}\right)=E_\theta\left(\frac{Z}{E(Z)}E\left(\frac{\left(Y-\sqrt{n}\mu\right)^t(Z\Omega)^{-1}Y}{Y^t(Z\Omega)^{-1}Y}\frac{Y^t(Z\Omega)^{-1}Y}{Y^tS^{-1}Y}\mid Z\right)\right).$$

En utilisant le lemme 2.2.11, les variables aléatoires Y et S étant donné $Z=z$ sont indépendantes et ont pour lois respectives $N_p\left(\sqrt{n}\mu, z\Omega\right)$ et $W_p(n-1,z\Omega)$. Alors les variables aléatoires $\frac{\left(Y-\sqrt{n}\mu\right)^t(Z\Omega)^{-1}Y}{Y^t(Z\Omega)^{-1}Y}$ et $\frac{Y^t(Z\Omega)^{-1}Y}{Y^tS^{-1}Y}$ sont conditionnellement indépendantes étant donnée $Z=z$.

Mais en vertu du lemme 2.2.3, la loi conditionnelle de $\frac{Y^t(z\Omega)^{-1}Y}{Y^tS^{-1}Y}$ étant donnée $Z=z$ est χ^2_{n-p}. En se servant du lemme précédent, on a :

$$E_\theta\left(\frac{\left(Y-\sqrt{n}\mu\right)^t\Sigma^{-1}Y}{Y^tS^{-1}Y}\right)=(n-p)(p-2)E^Z\left(\frac{Z}{E(Z)}E_\theta\left(\frac{1}{p+2K-2}\mid Z\right)\right)$$

$$=(n-p)(p-2)E\left(\frac{1}{W}E\left(\frac{1}{p+2K-2}\mid W\right)\right).$$

ii. De la même manière qu'en (i), on a :

$$E_\theta\left(\frac{Y^t\Sigma^{-1}Y}{\left(Y^tS^{-1}Y\right)^2}\right)=E_\theta\left(\frac{Z}{E(Z)}E\left(\left(\frac{Y^t(Z\Omega)^{-1}Y}{Y^tS^{-1}Y}\right)^2\mid Z\right)E\left(\frac{1}{Y^t(Z\Omega)^{-1}Y}\mid Z\right)\right). \quad (3.1.4)$$

Mais comme la loi conditionnelle de $\frac{Y^t(z\Omega)^{-1}Y}{Y^tS^{-1}Y}$ étant donné $Z=z$ est χ^2_{n-p}, on donc la relation suivante :

$$E\left(\left(\frac{Y^t(Z\Omega)^{-1}Y}{Y^tS^{-1}Y}\right)^2\mid Z\right)=2(n-p)+(n-p)^2=(n-p)(n-p+2).$$

Et d'après le lemme précédent et la relation précédente , la relation (3.1.4) devient :

$$E_\theta\left(\frac{Y^t\Sigma^{-1}Y}{\left(Y^tS^{-1}Y\right)^2}\right)=(n-p)(n-p+2)E_\theta\left(\frac{Z}{E(Z)}E\left(\frac{1}{p+2K-2}\mid Z\right)\right).$$

En posant $W=\frac{E(Z)}{Z}$, on obtient le résultat.

Théorème 3.1.7 La fonction de risque d'un estimateur de James-Stein δ_c sur Θ_λ est donnée par :

$$R(\theta, \delta_c) = ac^2 - 2bc + p, \tag{3.1.5}$$

où $a = (n-p)(n-p+2)E\left(W^{-1}E\left(\frac{1}{p+2M-2}|W\right)\right)$,

$b = (n-p)(p-2)E\left(W^{-1}E\left(\frac{1}{p+2M-2}|W\right)\right)$,

W est une variable aléatoire de même distribution que $\frac{E(Z)}{Z}$ et M est une variable aléatoire discrète dont la loi conditionnelle étant donné $W = w$ est Poisson de moyenne égale à $\frac{n\lambda^2}{2}w$.

Démonstration. La fonction de risque de $\delta_c \in \Delta_{JS}$, pour $\theta = (\mu, \Sigma) \in \Theta_\lambda$ est :

$$R(\theta, \delta_c) = E_\theta\left(\left(\sqrt{n}\mu - \left(1 - \frac{c}{Y^t S^{-1} Y}\right)Y\right)^t \Sigma^{-1}\left(\sqrt{n}\mu - \left(1 - \frac{c}{Y^t S^{-1} Y}\right)Y\right)\right).$$

En développant cette formule, on a :

$$R(\theta, \delta_c) = E\left(\left(Y - \sqrt{n}\mu\right)^t \Sigma^{-1}\left(Y - \sqrt{n}\mu\right)\right) - 2cE\left(\frac{\left(Y - \sqrt{n}\mu\right)^t \Sigma^{-1} Y}{Y^t S^{-1} Y}\right) + c^2 E\left(\frac{Y^t \Sigma^{-1} Y}{\left(Y^t S^{-1} Y\right)^2}\right).$$

Mais nous avons :

$$E\left(\left(Y - \sqrt{n}\mu\right)^t \Sigma^{-1}\left(Y - \sqrt{n}\mu\right)\right) = E\left\{\frac{Z}{E(Z)}E\left[\left(Y - \sqrt{n}\mu\right)^t (Z\Omega)^{-1}\left(Y - \sqrt{n}\mu\right) | Z\right]\right\} = p.$$

Alors en utilisant le lemme précédent et les deux dernières relations, on obtient le résultat.

Corollaire 3.1.8. Dans les conditions du problème, le meilleur estimateur (MESJ) dans Δ_{JS} est donné par :

$$\delta_{c^*}(Y,S)=\left(1-\frac{c^*}{Y'S^{-1}Y}\right)Y, \text{ avec } c^*=\frac{p-2}{n-p+2}.$$

Démonstration. C'est une conséquence directe de la relation (3.1.5).

On en déduit de ce corollaire que le meilleur estimateur de la moyenne μ d'un mélange de lois normales de paramètres (μ, Z, Ω), dans la classe des estimateurs de type James-Stein, ne dépend pas de la variable du mélange Z.

Corollaire 3.1.9. Dans les conditions du problème, le risque minimal dans Δ_{JS} est donné par :

$$R\left(\theta,\delta_{c^*}\right)=p-\frac{(n-p)(n-p+2)^2}{(p-2)}E\left(W^{-1}E\left(\frac{1}{p+2M-2}\,|\,W\right)\right),$$

W est une variable aléatoire de même distribution que $\frac{E(Z)}{Z}$ et M est une variable aléatoire discrète dont la loi conditionnelle étant donné $W=w$ est Poisson de moyenne égale à $\frac{n\lambda^2}{2}w$.

Démonstration. C'est une conséquence direct du théorème 3.1.7 et du corollaire 3.1.8.

3.1.3. Étude de l'estimateur du maximum de vraisemblance.

Dans cette sous-section, nous allons nous restreindre au cas où les observations sont normales. Soit $X_1, X_2, ..., X_n$ des variables aléatoires indépendantes et ayant la même loi normale $N_p(\mu,\Sigma)$. On veut évaluer l'estimateur du maximum de vraisemblance (EMV) de (μ,Σ) et son risque lorsque (μ,Σ) appartient à Θ_λ. Ce problème a été traité par Kariya, Giri et Perron (1988) pour $\lambda^2=1$. Nous allons reprendre l'approche de ces derniers, dans le but de comparer cet estimateur avec d'autres types d'estimateurs cités au début de ce chapitre.

Lemme 3.1.10. Soient A et B deux matrices carrées régulières de tailles respectives $(p \text{ x } p)$ et $(q \text{ x } q)$, C une matrice $(p \text{ x } q)$ et D une matrice $(q \text{ x } p)$. Si $P = A + CBD$ alors on a l'égalité suivante :

$$P^{-1} = A^{-1} - A^{-1}CB\left(B + BDA^{-1}CB\right)^{-1} BDA^{-1}.$$

Démonstration. Voir Muirhead (1982, page 580).

Théorème 3.1.11. L'estimateur du maximum de vraisemblance de (μ, Σ) sous la relation (3.1.1) est :

$$\delta_{EVM} = \left(\hat{\mu}_{EVM}, \hat{\Sigma}_{EVM}\right),$$

avec

$$\hat{\mu}_{EVM} = \frac{\sqrt{\lambda^4 + 4\lambda^2 V^{-1}} - \lambda^2}{2}\overline{X}$$

et

$$\hat{\Sigma}_{EVM} = \frac{1}{n}S + \frac{\sqrt{\lambda^4 + 4\lambda^2 V^{-1}} + \lambda^2}{2}\overline{X}\,\overline{X}'$$

où $V = \dfrac{U}{1+U}$ et $U = n\overline{X}S^{-1}\overline{X}$.

Démonstration. Nous utilisons la méthode des multiplicateurs de Lagrange. Ainsi nous cherchons à maximiser la fonction suivante :

$$\frac{n}{2}\log|\Sigma| - \frac{1}{2}trS\Sigma^{-1} - \frac{n}{2}(\bar{x} - \mu)^t \Sigma^{-1}(\bar{x} - \mu)^t - \frac{\gamma}{2}\left(\mu^t \Sigma^{-1}\mu - \lambda^2\right),$$

où γ est le multiplicateur de Lagrange.

Mais cette fonction atteint son maximum en

$$\hat{\mu}_{EVM} = \frac{n}{n+\gamma}\overline{X}$$

et

$$\hat{\Sigma}_{EVM} = \frac{1}{n}S + \frac{\gamma}{n+\gamma}\overline{X}\,\overline{X}^t .$$

En utilisant la contrainte (3.1.1), on obtient :

$$a^2\overline{X}\left(\frac{1}{n}S + (1-a)\overline{X}\,\overline{X}^t\right)^{-1}\overline{X} = \lambda_0^2 \text{ avec } a = \frac{n}{n+\gamma}.$$

En vertu du lemme précèdent, on a :

$$a^2\left(U - \frac{1-a}{1+(1-a)U}U^2\right) = \lambda_0^2.$$

Alors en résolvant cette équation, on obtient :

$$a = \frac{-\lambda_0^2 + \sqrt{\lambda_0^4 + 4\lambda_0^2 V^{-1}}}{2}.$$

D'où nous avons :

$$\hat{\mu}_{EVM} = \frac{\sqrt{\lambda_0^4 + 4\lambda_0^2 V^{-1}} - \lambda_0^2}{2}\overline{X}$$

et

$$\hat{\Sigma}_{EVM} = \frac{1}{n}S + \frac{\sqrt{\lambda_0^4 + 4\lambda_0^2 V^{-1}} + \lambda_0^2}{2}\overline{X}\,\overline{X}^t .$$

Théorème 3.1.12. Le risque de δ_{EVM} est donné par l'expression suivante :

$$R(\theta,\delta_{EVM}) = \lambda^2 + pe^{-\frac{n\lambda^2}{2}} E\left\{k_{EVM}^2(V)_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}V\right)\right.$$

$$\left.-2\frac{\sqrt{n\lambda^2}}{p}k_{EVM}(V)_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}V\right)\right\}$$

avec $k_{EVM}(V) = \dfrac{\sqrt{\lambda_0^4 + 4\lambda_0^2 V^{-1}} - \lambda_0^2}{2}$ et V est une variable aléatoire de densité Bêta $\mathscr{B}e\left(\frac{p}{2}+1,\frac{n-p}{2}\right)$.

Démonstration. La fonction de risque de δ_{EVM} évaluée en $\theta = (\mu,\Sigma) \in \Theta_\lambda$ est :

$$R(\theta,\delta_{EVM}) = \lambda^2 + E_\theta\left(k_{EVM}^2(U)E\left(Y^t\Sigma^{-1}Y \mid U\right) - 2k_{EVM}(U)E\left(\mu^t\Sigma^{-1}Y \mid U\right)\right).$$

Mais en utilisant les relations (2.2.18), (2.2.19) et (2.2.20), on obtient :

$$R(\theta,\delta_{EVM}) = \lambda^2 + pe^{-\frac{n\lambda^2}{2}} \int_0^{+\infty} \frac{u^{\frac{p}{2}}(1+u)^{-\frac{n}{2}-1}}{B\left(\frac{p}{2}+1,\frac{n-p}{2}\right)}\left(k_{EVM}^2(u)_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}\frac{u}{u+1}\right)\right.$$

$$\left.-2k_{EVM}(u)\frac{\sqrt{n\lambda^2}}{p}{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}\frac{u}{u+1}\right)\right)du.$$

En posant $v = \dfrac{u}{1+u}$, on obtient le résultat.

3.2. ÉTUDE COMPARATIVE DE QUELQUES CLASSES D'ESTIMATEURS DE LA MOYENNE μ D'UNE DISTRIBUTION DE MÉLANGES DE LOIS NORMALES LORSQUE L'INFORMATION PARAMÈTRIQUE EST ÉRONNÉE.

Dans cette section, nous allons comparer des estimateurs équivariants lorsque l'information paramétrique supposée égale à λ_0^2 est fausse.

Dans un premier temps, on proposera les comparaisons entre le meilleur estimateur équivariant et quelques types d'estimateurs équivariants tels le meilleur estimateur dans la classe des estimateurs de type James-Stein, le meilleur estimateur dans la classe des estimateurs linéaires et l'estimateur du maximum de vraisemblance. L'étude se fera seulement dans le cas d'une loi normale.

Dans un second temps on évaluera le risque minimal pour une loi de Student. Ceci dans le but de comparer le risque minimal pour différentes valeurs du degré de liberté.

3.2.1 Étude de quelques classes d'estimateurs de la moyenne μ d'une loi multinormale lorsque $\mu'\Sigma^{-1}\mu \neq \lambda_0^2$.

Dans cette sous-section, nous allons évaluer, pour chacun des estimateurs ci-dessus, le risque lorsque l'information paramétrique est erronée.

Donc il s'agit d'étudier le comportement d'un estimateur $\hat{\mu}_{EVM}$ construit comme l'estimateur du maximum de vraisemblance pour le cas normal et dans un contexte de robustesse.

Théorème 3.2.1. La fonction de risque d'un estimateur $\delta(Y,S) = k(U)Y$ lorsque $\lambda^2 \neq \lambda_0^2$ est :

$$R(\theta,\delta(Y,S))=\lambda^2+p\int_0^{+\infty}\frac{\mathrm{v}^{\frac{p}{2}}(1-\mathrm{v})^{\frac{n-p}{2}-1}}{e^{\frac{n\lambda^2}{2}}B\left(\frac{p}{2}+1,\frac{n-p}{2}\right)}h(\mathrm{v})\left(h(\mathrm{v})\,{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}\mathrm{v}\right)\right.$$

$$\left.-2\frac{\sqrt{n}\lambda^2}{p}\,{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}\mathrm{v}\right)d\mathrm{v}\right)$$

avec $h(\mathrm{v})=k\left(\frac{\mathrm{v}}{1-\mathrm{v}}\right)$.

Démonstration. Pour $\theta=(\mu,\Sigma)$, la fonction de risque d'un estimateur $\delta(Y,S)=k(U)Y$ est donnée par :

$$R(\theta,\delta(Y,S))=\lambda^2+E\{k^2(U)E(Y'\Sigma^{-1}Y\mid U)-2k(U)E(\mu'\Sigma^{-1}Y\mid U)\}.$$

En utilisant les relations (2.2.35) et (2.2.36), on a :

$$R(\theta,\delta(Y,S))=\lambda^2+E\{k^2(U)k_2(U)-2k(U)k_1(U)\}.$$

Et en vertu de la relation (2.2.21), on a :

$$R(\theta,\delta(Y,S))=\lambda^2+\frac{\Gamma\left(\frac{n}{2}\right)e^{-\frac{n\lambda^2}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{p}{2}\right)}n\int_0^{\infty}u^{\frac{p}{2}}(1+u)^{-\frac{n}{2}-1}\left(k^2(u){}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}\frac{u}{u+1}\right)\right.$$

$$\left.-2k(u)\frac{\sqrt{n}\lambda^2}{p}{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}\frac{u}{u+1}\right)\right)du.$$

En posant $\mathrm{v}=\frac{u}{1+u}$, et $h(\mathrm{v})=k\left(\frac{\mathrm{v}}{1-\mathrm{v}}\right)$, on obtient la relation suivante :

$$R(\theta,\delta(Y,S))=\lambda^2+p\frac{e^{-\frac{n\lambda^2}{2}}}{B\left(\frac{p}{2}+1,\frac{n-p}{2}\right)}\int_0^\infty \mathrm{v}^{\frac{p}{2}}(1-\mathrm{v})^{\frac{n-p}{2}-1}\left(h^2(\mathrm{v})\,{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}\mathrm{v}\right)\right.$$

ce

$$\left.-2h(\mathrm{v})\frac{\sqrt{n}\lambda^2}{p}\,{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}\mathrm{v}\right)\right)d\mathrm{v},$$

qui achève la démonstration.

Corollaire 3.2.2. La fonction de risque du meilleur estimateur $\hat{\delta}(Y,S,\lambda_0)=\hat{k}(U,\lambda_0)Y$ lorsque $\lambda^2=\mu'\Sigma^{-1}\mu\neq\lambda_0^2$ est :

$$R\left(\theta,\hat{\delta}_E(Y,S)\right)=\lambda^2+pe^{-\frac{n\lambda^2}{2}}E\left\{\hat{h}(V,\lambda_0^2)\left[\hat{h}(V,\lambda_0^2)\,{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}V\right)\right.\right.$$

$$\left.\left.-2\frac{\sqrt{n}\lambda^2}{p}\,{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}V\right)\right]\right\},$$

où $$\hat{h}(V,\lambda_0)=\frac{\sqrt{n}\lambda_0^2}{p}\,\frac{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda_0^2}{2}V\right)}{{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda_0^2}{2}V\right)}.$$

Démonstration. Il s'agit d'une conséquence directe du théorème 3.2.1.

Théorème 3.2.3. La fonction de risque du meilleur estimateur linéaire dans la classe Δ_L $\delta_{k^*}(Y,S,\lambda_0)=k^*(\lambda_0)Y$ lorsque $\lambda^2=\mu'\Sigma^{-1}\mu\neq\lambda_0^2$ est :

$$R(\theta,\delta_k)=\frac{n^{-2}p^2}{\left(\lambda_0^2+n^{-1}p\right)^2}\lambda^2+\frac{n^{-1}p\lambda_0^4}{\left(\lambda_0^2+n^{-1}p\right)^2}.$$

Démonstration. En utilisant la relation (3.1.2) pour $\delta_{k^*}(Y,S,\lambda_0)=k^*(\lambda_0)Y$, la fonction de risque devient :

$$R(\theta,\delta_k)=\lambda^2-2\lambda^2\frac{\lambda_0^2}{\lambda_0^2+n^{-1}p}+\frac{\lambda_0^4}{\left(\lambda_0^2+n^{-1}p\right)^2}\left(\frac{p}{n}+\lambda^2\right)=\frac{n^{-1}p\left(\lambda_0^4+n^{-1}p\lambda^2\right)}{\left(\lambda_0^2+n^{-1}p\right)^2}.$$

Théorème 3.2.4. La fonction de risque du meilleur estimateur de type James-Stein $\delta_{c^*}(Y,S)=\left(1-\frac{c^*}{Y'S^{-1}Y}\right)Y$ dans la classe Δ_{JS} lorsque $\lambda^2=\mu'\Sigma^{-1}\mu\neq\lambda_0^2$ est donnée par :

$$R\left(\theta,\delta_c^{\ *}\right)=p-\frac{(n-p)(n-p+2)^2}{(p-2)}E\left(\frac{1}{p+2M-2}\right),$$

où M est une variable aléatoire de Poisson de paramètre de moyenne égal à $\frac{n\lambda_0^2}{2}$.

Démonstration. On a le résultat en utilisant la relation (3.1.5) pour

$\delta(Y,S)=\delta_{c^*}(Y,S,\lambda_0)$.

Théorème 3.2.5. La fonction de risque de l'estimateur du maximum de vraisemblance $\delta_{EVM}\left(Y,S,\lambda_0^2\right)=\frac{\sqrt{\lambda_0^4+4\lambda_0^2V^{-1}}-\lambda_0^2}{2}\overline{X}$ lorsque $\lambda^2\neq\lambda_0^2$ est donnée par :

$$R(\theta,\delta(Y,S))=\lambda^2+p\int_0^\infty\frac{v^{\frac{p}{2}}(1-v)^{\frac{n-p}{2}-1}}{e^{\frac{n\lambda^2}{2}}B\left(\frac{p}{2}+1,\frac{n-p}{2}\right)}k_{MEV}\left(v,\lambda_0^2\right)\left\{k_{MEV}\left(v,\lambda_0^2\right){}_1F_1\left(\frac{n}{2}+1;\frac{p}{2};\frac{n\lambda^2}{2}v\right)\right.$$

$$\left.-2\frac{\sqrt{n}\lambda^2}{p}\,{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}v\right)\right\}dv.$$

Démonstration. On a le résultat en utilisant le théorème 3.2.1 pour

$\delta(Y,S)=\delta_{MEV}\left(Y,S,\lambda_0^2\right)$.

3.2.2 Étude du meilleur estimateur équivariant de la moyenne μ d'une loi Student multivariée $St_p(m,\mu,\Sigma)$ lorsque $\mu'\Sigma^{-1}\mu \neq \lambda_0^2$, pour différentes valeurs de *m*.

On veut évaluer le risque minimal du meilleur estimateur de la moyenne μ d'une loi Student multivariée $St_p(m,\mu,\Sigma)$ lorsque $\mu'\Sigma^{-1}\mu = \lambda^2 \neq \lambda_0^2$. Ceci dans le but d'étudier le comportement du meilleur estimateur de la moyenne μ d'une loi Student multivariée $St_p(m,\mu,\Sigma)$ lorsque $\mu'\Sigma^{-1}\mu \neq \lambda_0^2$, pour différentes valeurs du degré de liberté m.

Théorème 3.2.6. Dans les conditions du problème **P.2**, la fonction de risque d'un estimateur $\delta(Y,S)=k(U)Y$ lorsque $\lambda^2 = \mu'\Sigma^{-1}\mu \neq \lambda_0^2$ pour un mélange de lois normales est égale à :

$$R(\theta,\delta(Y,S)) = \lambda^2 + p\int_0^\infty \frac{v^{\frac{p}{2}}(1-v)^{\frac{n-p}{2}-1}}{B\left(\frac{p}{2}+1,\frac{n-p}{2}\right)}\left(h^2(v)g_2(v)-2h(v)g_1(v)\right)dv,$$

avec

$$g_1(v) = \frac{\sqrt{n}\lambda^2}{p}E\left(e^{-\frac{n\lambda^2}{2}W}{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}vW\right)\right),$$

$$g_2(v) = E\left(W^{-1}e^{-\frac{n\lambda^2}{2}W}{}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}vW\right)\right),$$

$h(v) = k\left(\frac{v}{1-v}\right), \forall\, v \in [0,1]$ et W admet la même loi que $\frac{E(Z)}{Z}$.

Démonstration. Pour $\theta=(\mu,\Sigma)$, la fonction de risque d'un estimateur $\delta(Y,S)=k(U)Y$ est donnée par :

$$R(\theta,\delta(Y,S)) = \lambda^2 + E\{k^2(U)E(Y'\Sigma^{-1}Y \mid U) - 2k(U)E(\mu'\Sigma^{-1}Y \mid U)\}$$

Mais en utilisant les relations (2.2.29) et (2.2.30), dans (2.2.26) et (2.2.27) on a :

$$E(Y'\Sigma^{-1}\mu \mid \mathrm{U} = \mathrm{u}) = \frac{n^{3/2}\lambda^2}{pB\left(\frac{p}{2},\frac{n-p}{2}\right)} \frac{u^{\frac{p}{2}}(1+u)^{-\frac{n}{2}-1}}{f^U(u)}$$

$$\mathrm{x}\int_0^{+\infty} e^{-\frac{n\lambda^2}{2}\frac{E(Z)}{Z}} {}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\frac{u}{u+1}\right)dH^Z(z)$$

et

$$E(Y'\Sigma^{-1}Y \mid \mathrm{U} = \mathrm{u}) = \frac{n}{B\left(\frac{p}{2},\frac{n-p}{2}\right)} \frac{u^{\frac{p}{2}}(1+u)^{-\frac{n}{2}-1}}{f^{\mathrm{U}}(\mathrm{u})}$$

$$\mathrm{x}\int_0^{+\infty} \frac{E(Z)}{Z} e^{-\frac{n\lambda^2}{2}\frac{E(Z)}{Z}} {}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}\frac{E(Z)}{Z}\frac{u}{u+1}\right)dH^Z(z).$$

D'où en posant $W = \frac{Z}{E(Z)}$, et $\mathrm{v} = \frac{u}{1+u}$, on a le résultat.

On a vu dans le chapitre 2 que la loi Student multivariée de paramètres $(\alpha,\beta,\mu,\Sigma)$ (notée $St_p(\alpha,\beta,\mu,\Sigma)$), s'obtient par un mélange de lois normales. La loi du mélange Z^{-1} est gamma $\mathcal{G}(\alpha,\beta)$. Les lois normales sont conditionnellement à Z $N_p(\mu, Z\Omega)$. Le cas où $\alpha = \frac{m}{2}$ correspond à la loi de Student multivariée standard $St_p(m,\mu,\Sigma)$.

Corollaire 3.2.7. La fonction de risque en $\mu'\Sigma^{-1}\mu = \lambda^2$ du meilleur estimateur équivariant d'une loi Student $St_p(\alpha,\beta,\mu,\Sigma)$ pour $\alpha > 1$ et $\beta > 0$ lorsque l'information paramétrique $\lambda^2 \neq \lambda_0^2$, est donnée par :

$$R(\theta,\delta(Y,S)) = \lambda^2 + p\left(\frac{\alpha-1}{t+\alpha-1}\right)^{\alpha-1}\int_0^\infty \frac{\mathrm{v}^{\frac{p}{2}}(1-\mathrm{v})^{\frac{n-p}{2}-1}}{B\left(\frac{p}{2}+1,\frac{n-p}{2}\right)}\hat{h}(\mathrm{v})\left(\hat{h}(\mathrm{v})\ {}_2F_1\left(\frac{n}{2}+1,\alpha-1;\frac{p}{2};r\mathrm{v}\right)\right.$$

$$\left.-4\frac{(\alpha-1)r}{p\sqrt{n}}\ {}_2F_1\left(\frac{n}{2}+1,\alpha;\frac{p}{2}+1;r\mathrm{v}\right)\right)d\mathrm{v},$$

avec

$$\hat{h}(\mathrm{v}) = \frac{2(\alpha-1)r_0}{p\sqrt{n}}\frac{{}_2F_1\left(\frac{n}{2}+1,\alpha;\frac{p}{2}+1;\mathrm{r}_0\mathrm{v}\right)}{{}_2F_1\left(\frac{n}{2}+1,\alpha-1;\frac{p}{2};\mathrm{r}_0\mathrm{v}\right)},\ \mathrm{t}=\frac{n\lambda^2}{2},\ \mathrm{r}=\frac{\mathrm{t}}{\mathrm{t}+\alpha-1},\ \mathrm{t}_0=\frac{n\lambda_0^2}{2},$$

$\mathrm{r}_0 = \dfrac{\mathrm{t}_0}{\mathrm{t}_0+\alpha-1}$ et $\mathrm{v} = \dfrac{u}{1+u}$.

Démonstration. Il faut remarquer que si Z admet la distribution de gamma $\mathcal{G}(\alpha,\beta)$, $\alpha>1$ et $\beta>0$ alors

$$g_1(\mathrm{v}) = \frac{\sqrt{n}\lambda^2}{p}\left(\frac{\alpha-1}{t+\alpha-1}\right)^{\alpha}\int_0^\infty \frac{(\mathrm{t}+\alpha-1)^\alpha}{\Gamma(\alpha)}\mathrm{w}^{\alpha-1}e^{-(\mathrm{t}+\alpha-1)\mathrm{w}}\ {}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;\frac{n\lambda^2}{2}\mathrm{vw}\right)d\mathrm{w}$$

$$= \frac{\sqrt{n}\lambda^2}{p}\left(\frac{\alpha-1}{t+\alpha-1}\right)^{\alpha-1}E^W\left(\mathrm{W}\ {}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;t\mathrm{vW}\right)\right)$$

et

$$g_2(\mathrm{v}) = \left(\frac{\alpha-1}{t+\alpha-1}\right)^{\alpha-1}E^W\left({}_1F_1\left(\frac{n}{2}+1;\frac{p}{2}+1;t\mathrm{vW}\right)\right),$$

où W est de loi $G(\alpha-1,\mathrm{t}+\alpha-1)$.

Mais en utilisant le lemme 2.2.26, on obtient :

$$g_1(\mathrm{v}) = \frac{\sqrt{n}\lambda^2}{p}\left(\frac{\alpha-1}{t+\alpha-1}\right)^{\alpha}{}_2F_1\left(\frac{n}{2}+1,\alpha;\frac{p}{2}+1;\frac{t}{t+\alpha-1}\mathrm{v}\right) \qquad (3.2.1)$$

et

$$g_2(\mathrm{v}) = \left(\frac{\alpha-1}{t+\alpha-1}\right)^{\alpha\text{-}1} {}_2F_1\left(\frac{n}{2}+1, \alpha-1; \frac{p}{2}; \frac{t}{t+\alpha-1}\mathrm{v}\right). \tag{3.2.2}$$

D'où en utilisant le théorème 2.2.27, le théorème 3.2.6, les relations (3.2.1) et (3.2.2), on obtient le résultat.

Corollaire 3.2.8. La fonction de risque en $\mu{}'\Sigma^{-1}\mu = \lambda^2$ du meilleur estimateur équivariant d'une loi Student $St_p(m, \mu, \Sigma)$ lorsque l'information paramétrique $\lambda^2 \neq \lambda_0^2$, est donnée par :

$$R(\theta, \delta(Y,S)) = \lambda^2 + p\left(\frac{m-2}{2t+m-2}\right)^{\frac{m-2}{2}} \int_0^\infty \frac{\mathrm{v}^{\frac{p}{2}}(1-\mathrm{v})^{\frac{n-p}{2}-1}}{B\left(\frac{p}{2}+1, \frac{n-p}{2}\right)} \hat{h}(\mathrm{v}) \left(\hat{h}(\mathrm{v})\, {}_2F_1\left(\frac{n+2}{2}, \frac{m-2}{2}; \frac{p}{2}; r\mathrm{v}\right) \right.$$

$$\left. - 2\frac{(m-2)r}{p\sqrt{n}}\, {}_2F_1\left(\frac{n+2}{2}, \frac{m}{2}; \frac{p+2}{2}; r\mathrm{v}\right)\right) d\mathrm{v},$$

avec $\hat{h}(\mathrm{v}) = \dfrac{(\mathrm{m}-2)r}{p\sqrt{n}} \dfrac{{}_2F_1\left(\frac{n}{2}+1, \frac{\mathrm{m}}{2}; \frac{p}{2}+1; \mathrm{rv}\right)}{{}_2F_1\left(\frac{n}{2}+1, \frac{\mathrm{m}}{2}-1; \frac{p}{2}; \mathrm{rv}\right)}$, $\mathrm{t} = \dfrac{n\lambda^2}{2}$, $\mathrm{r} = \dfrac{2\mathrm{t}}{2\mathrm{t}+m-2}$ et $\mathrm{v} = \dfrac{u}{1+u}$.

Démonstration. Il suffit d'appliquer le lemme précèdent.

3.3. COMMENTAIRES ET INTERPRÉTATIONS

Dans cette sous section, on évalue numériquement les risques pour des cas particuliers et on trace des courbes. Les calculs sont effectués en utilisant des programmes faits en *FORTRAN 77*. Les graphiques sont tracés à l'aide de *MATLAB*.

Dans un premier temps, on considère le problème où la loi est $N_p(\mu,\Sigma)$. Pour chacun des estimateurs considérés précédemment, on représente graphiquement les cas suivants :

– Risque minimal lorsque λ^2 est connu.

– Fonction de risque de l'estimateur localement meilleur en λ_0^2 c'est-à-dire que dans la classe considéré, les estimateur sont optimaux lorsque $\lambda^2 = \lambda_0^2$.

– Evaluation du multiplicateur optimal.

Ceci lorsque $n = 1, 20$, $p = 1, 2, 3, 4$ et $\lambda_0^2 = 0.5, 2, 4$.

Dans un second temps, on considère le problème où la loi est Student $St_p(m, \mu, \Sigma)$. On refait la même procédure avec $m = 3, 4, 5$ et 6.

Dans le premier cas, où $\mu^t\Sigma^{-1}\mu = \lambda^2$ connu, il est certain que le Meilleur Estimateur Linéaire (MEL), le Meilleur Estimateur de James Stein(MEJS), l'Estimateur du Maximum de Vraisemblance (EMV) sont tous dominés par le Meilleur Estimateur Équivariant (MEE). Et ceci est illustré par les graphiques 7 et 8, qui présentent les risques minimaux en λ^2 atteints par MEE, MEL, MEJS, EMV. Ces graphiques montrent en outre que ces estimateurs ont la même allure, donc presque le même comportement pour $p > 1$. La prédominance du MEE sur MEL, MEJS et EMV, est très faible pour $p > 1$. Le cas $p = 1$ est un peu particulier du point de vue comportement et écart entre les courbes MEE et MEL sauf pour les valeurs de λ^2 très petites. Ces graphiques vérifient les propriétés de la section 2.2.3.

Les multiplicateurs de MEE, MEL, MEJS, EMV sont aussi tracés et présentés aux figures 9 à 14. Ces graphiques vérifient aussi les propriétés vues dans la section 2.2.3.

Pour le cas où l'information paramétrique $\mu'\Sigma^{-1}\mu = \lambda^2 \neq \lambda_0^2$, les courbes représentant les fonctions de risque du MEE, MEL, EMV, en λ^2 pour différentes valeurs de n, p et λ_0^2 sont présentées aux figures 1 à 6.

Ces courbes montrent que le MEE et EMV ont des comportements assez semblables.

Pour chaque n, p et λ_0^2, le MEE semble dominer légèrement le MEL sur un intervalle contenant λ_0^2, assez restreint. Mais en dehors de cet intervalle, le MEL semble dominer assez largement le MEE. On peut alors dire, d'après ces graphiques, que le MEL semble plus stable que le MEE.

Les comparaisons entre le MEL et EMV décèlent des résultats similaires à ceux trouvés entre le MEL et le MEE. Mais l'intervalle où EMV domine le MEL ne contient pas forcement λ_0^2.

Il est intéressant de comparer graphiquement les courbes des fonctions de risque minimal en λ^2, pour les familles de Student multivariées $St_p(\mu, \Sigma, m)$, pour différentes valeurs du degrés de liberté m. Mais on a présenté seulement les courbes pour $(m = 3, 4, 5, 6)$ dans ce mémoire. Ces courbes sont proposées aux figures 21 à 22. Pour n, p fixé, les courbes des fonctions de risque minimal ont le même comportement. L'écart entre ces courbes croit en λ^2. Mais néanmoins l'écart n'est pas considérable et tend à disparaître lorsque λ^2 est très petit. Les propriétés vues dans la sous section 2.2.4. sont confirmées par les graphiques.

La comparaison des fonctions de risque en λ^2, lorsque : $\mu'\Sigma^{-1}\mu = \lambda^2 \neq \lambda_0^2$, sont reproduites aux figures 15 à 20. Pour $p = 1$ et pour chaque n et λ_0^2 fixés, il existe un intervalle contenant λ_0^2 sur lequel le risque diminue en fonction de m (plus m est grand plus le risque devient intéressant). Mais en dehors de cet intervalle on observe le contraire. Pour $p \neq 1$ et pour tout n et λ_0^2, la fonction de risque en $m = 3$ est plus petite que les autres fonctions de risque. Si $m = p + 2$, alors la fonction de risque en m coïncide avec celle du MEL.

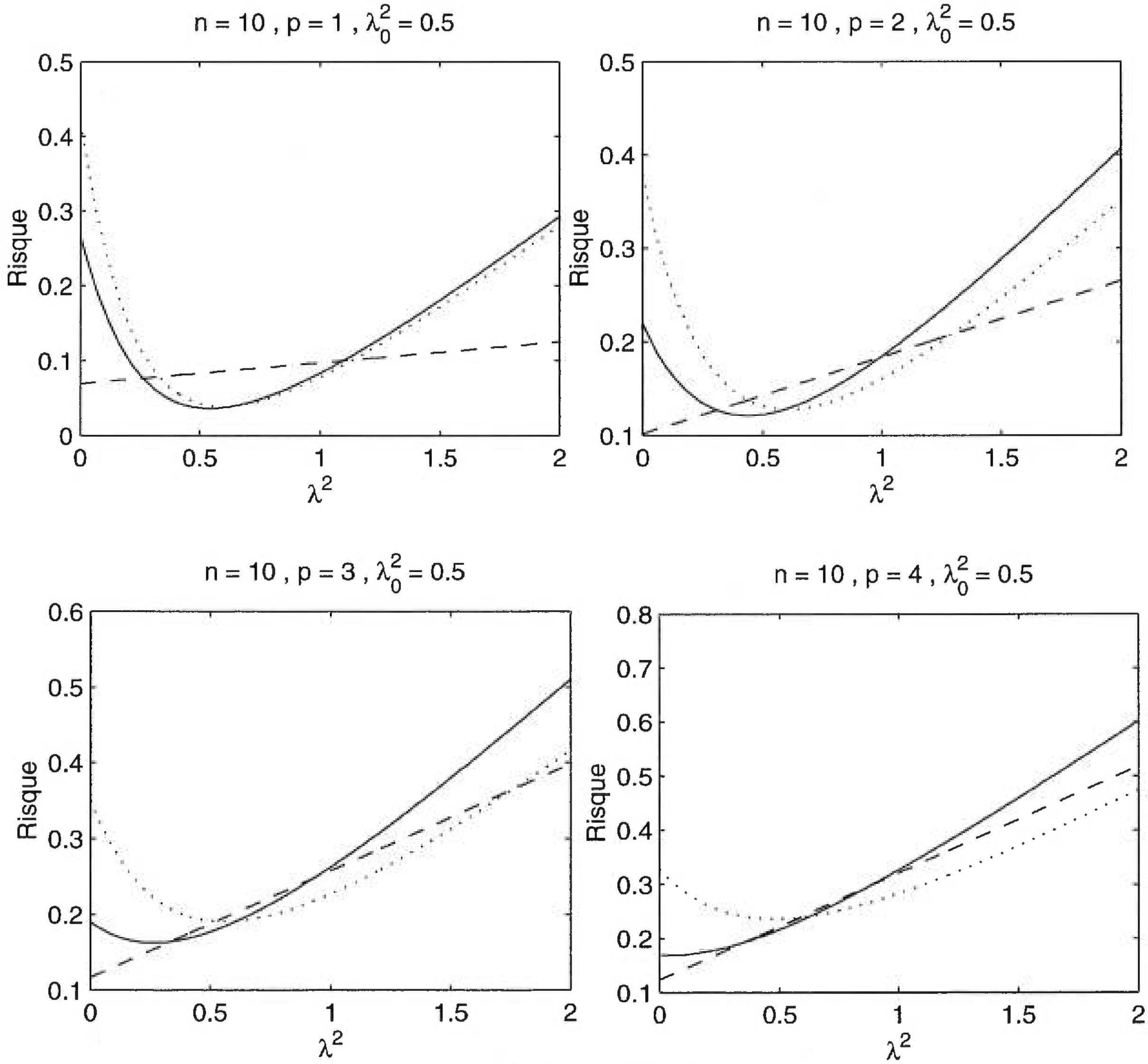
GRAPHIQUES

FONCTION DE RISQUE D'UN ESTIMATEUR ÉQUIVARIANT LOCALEMENT MEILLEUR EN $\lambda_0^2 = 0.5$ (n =10).

———: *Risque de l'estimateur équivariant localement meilleur en λ_0^2 (MEE).*

$- - -$: *Risque de l'estimateur linéaire localement meilleur en λ_0^2 (MEL).*

$\cdots\cdots$: *Risque de l'estimateur du maximum de vraisemblance localement meilleur en λ_0^2 (MEV).*



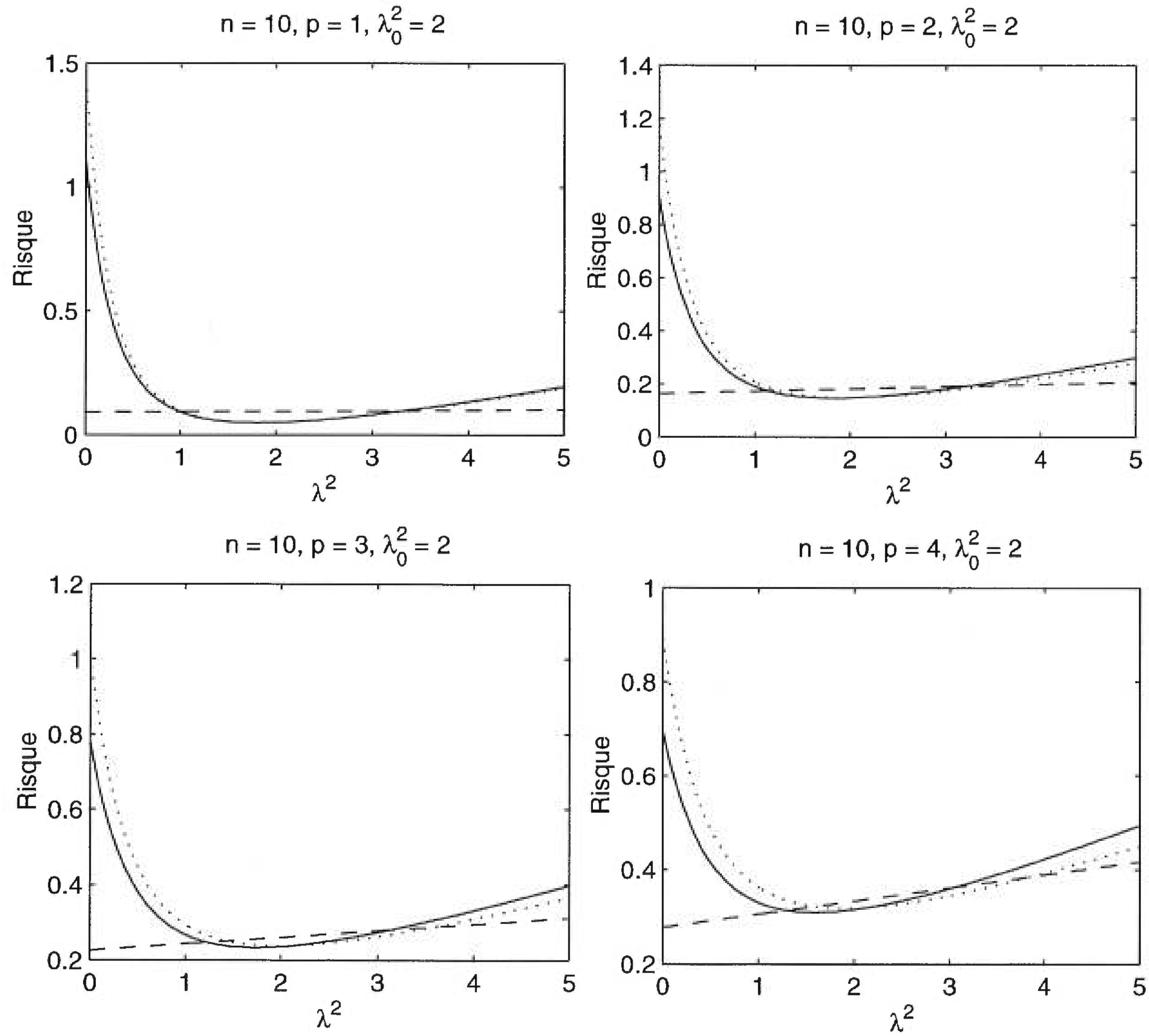
Graphique 1

FONCTION DE RISQUE D'UN ESTIMATEUR ÉQUIVARIANT LOCALEMENT MEILLEUR EN $\lambda_0^2 = 2$ (n =10).

———: *Risque de l'estimateur équivariant localement meilleur en* λ_0^2 *(MEE).*

— — —: *Risque de l'estimateur linéaire localement meilleur en* λ_0^2 *(MEL).*

······: *Risque de l'estimateur du maximum de vraisemblance localement meilleur en* λ_0^2 *(MEV).*



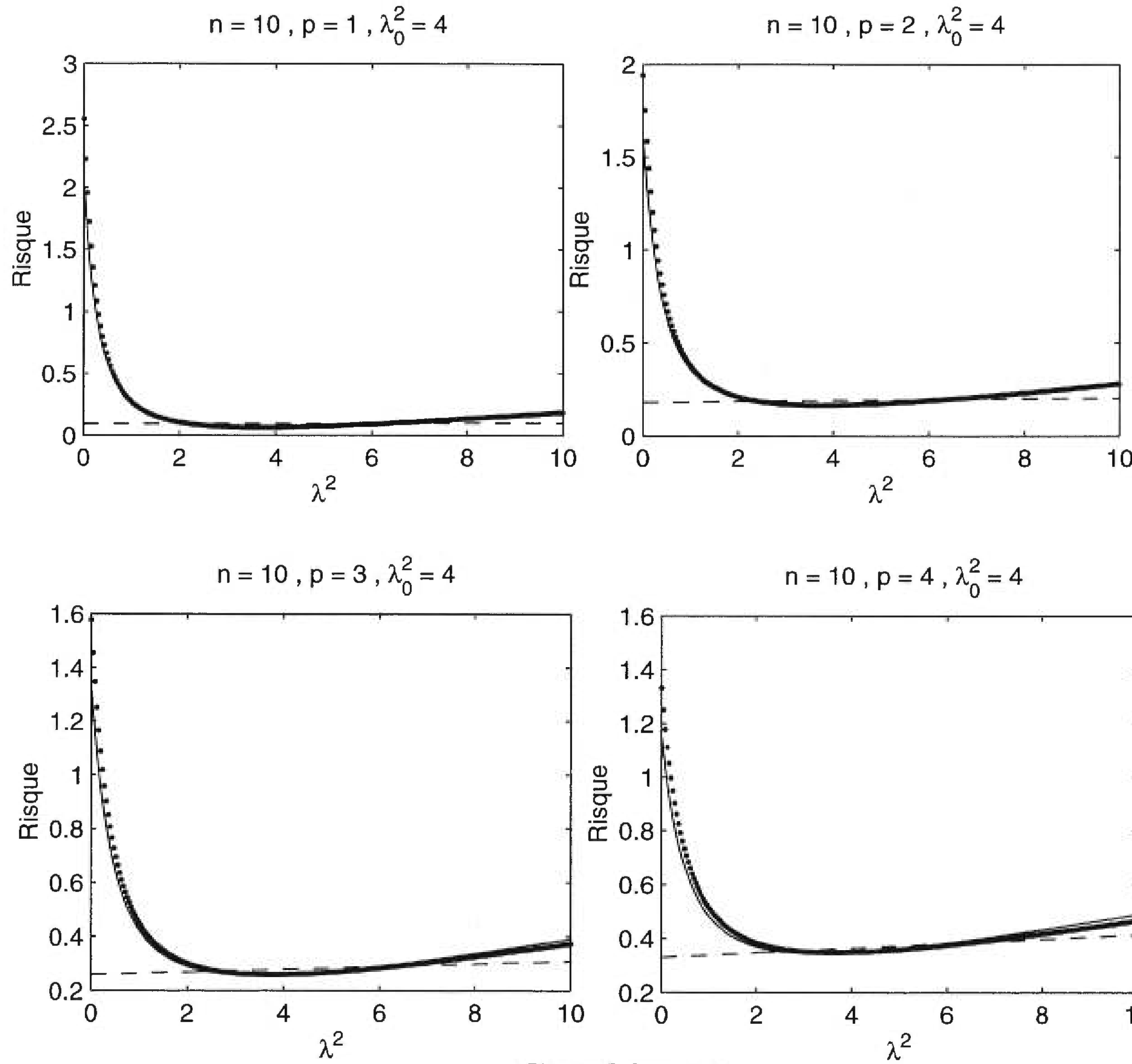
Graphique 2

FONCTION DE RISQUE D'UN ESTIMATEUR ÉQUIVARIANT LOCALEMENT MEILLEUR EN $\lambda_0^2 = 4$ (n =10).

———: *Risque de l'estimateur équivariant localement meilleur en λ_0^2 (MEE).*

– – – : *Risque de l'estimateur linéaire localement meilleur en λ_0^2 (MEL).*

······ : *Risque de l'estimateur du maximum de vraisemblance localement meilleur en λ_0^2 (MEV).*



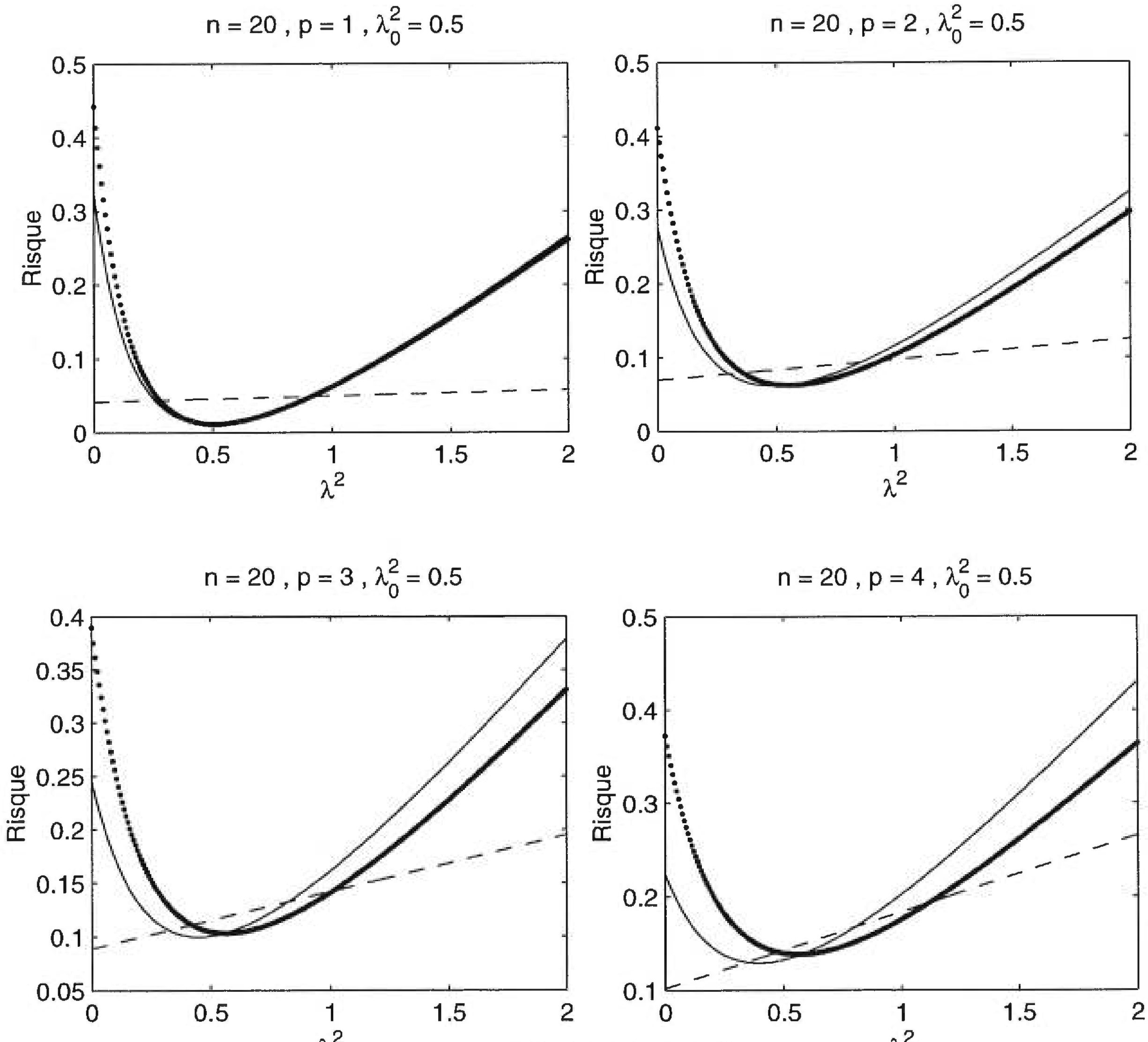
Graphique 3

FONCTION DE RISQUE D'UN ESTIMATEUR ÉQUIVARIANT LOCALEMENT MEILLEUR EN $\lambda_0^2 = 0.5$ (n =20).

———: *Risque de l'estimateur équivariant localement meilleur en λ_0^2 (MEE).*

— — —: *Risque de l'estimateur linéaire localement meilleur en λ_0^2 (MEL).*

······ : *Risque de l'estimateur du maximum de vraisemblance localement meilleur en λ_0^2 (MEV).*



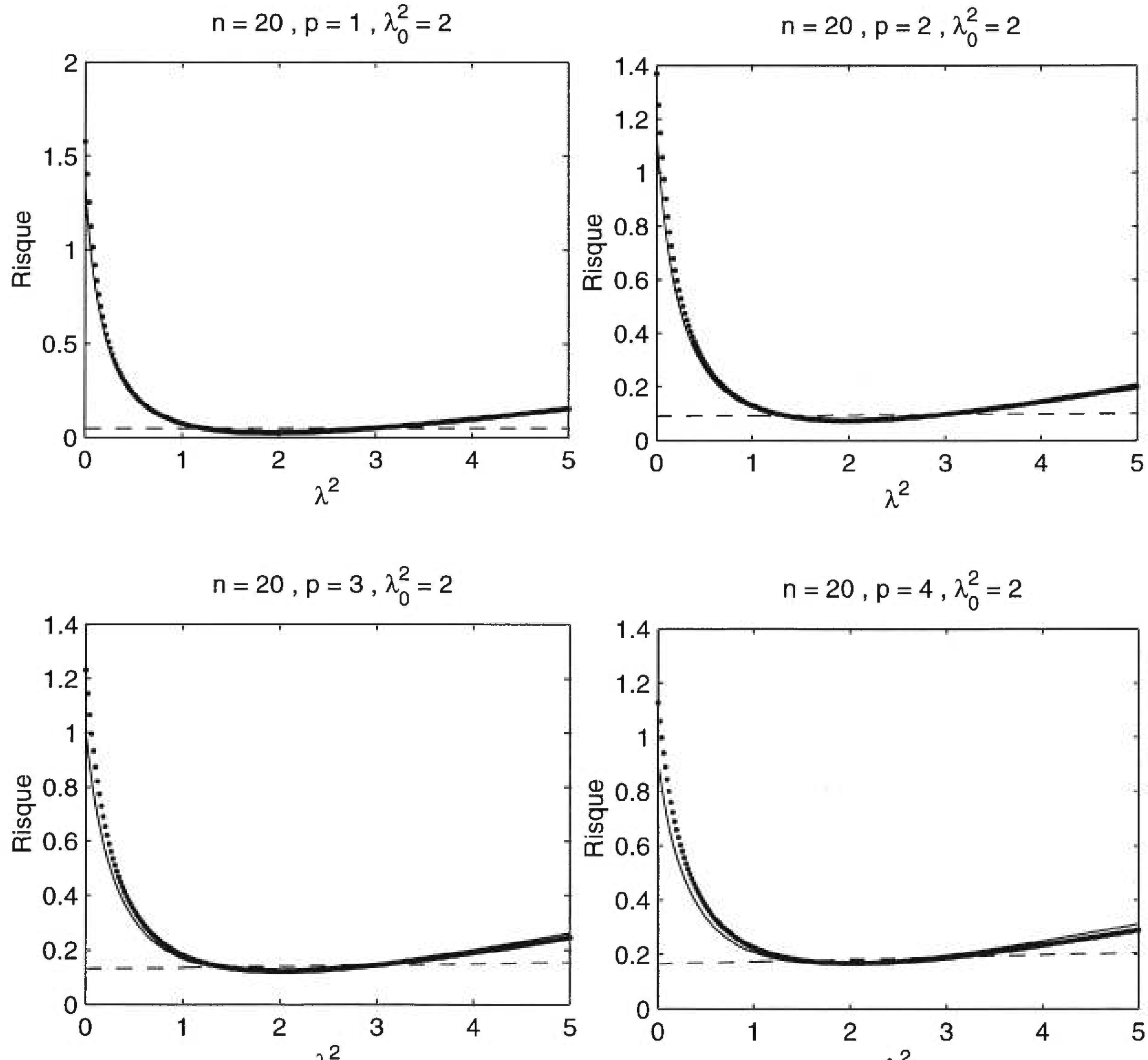
Graphique 4

FONCTION DE RISQUE D'UN ESTIMATEUR ÉQUIVARIANT LOCALEMENT MEILLEUR EN $\lambda_0^2 = 2$ (n =20).

———: *Risque de l'estimateur équivariant localement meilleur en λ_0^2 (MEE).*

— — —: *Risque de l'estimateur linéaire localement meilleur en λ_0^2 (MEL).*

······ : *Risque de l'estimateur du maximum de vraisemblance localement meilleur en λ_0^2 (MEV).*



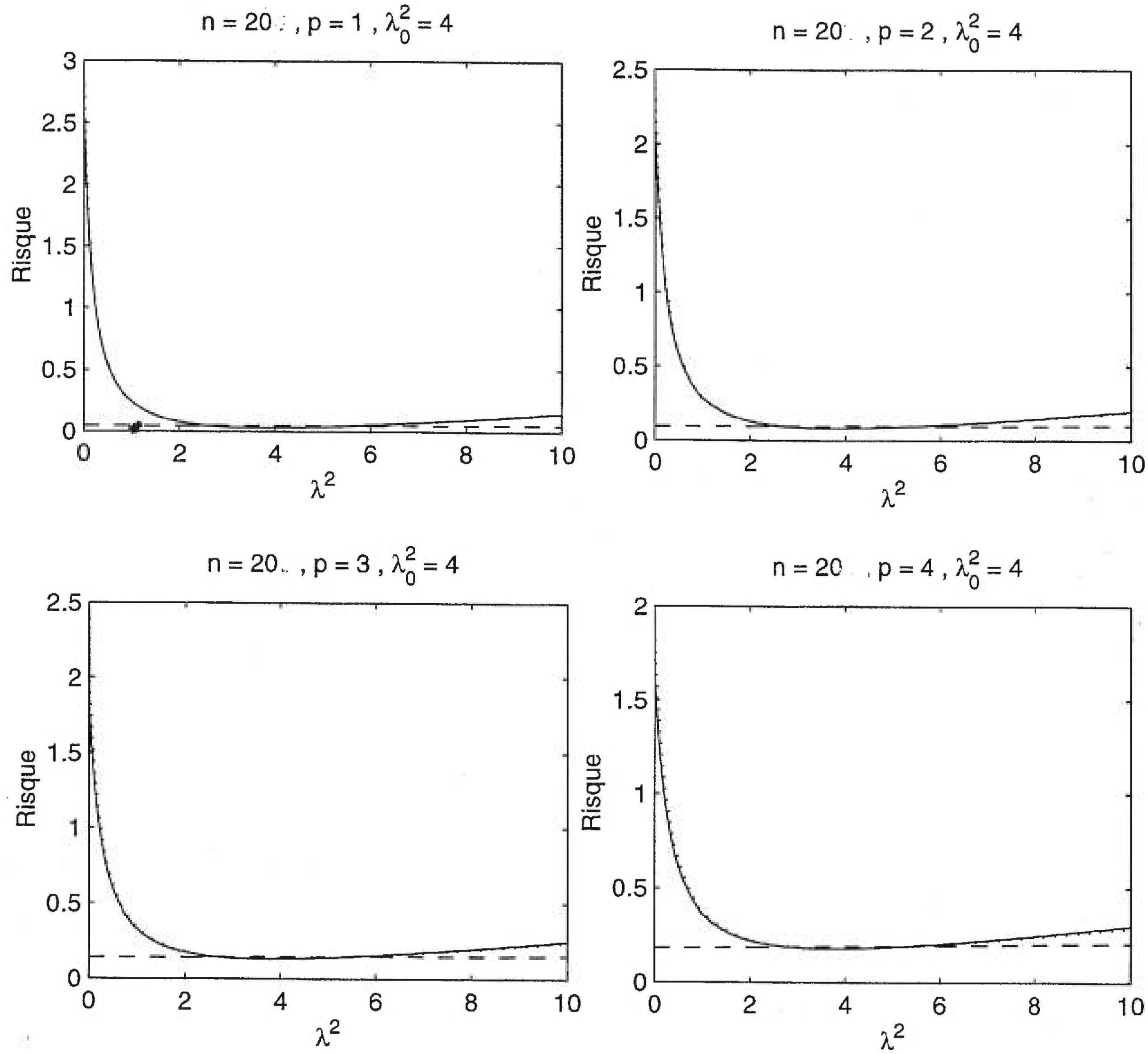
Graphique 5

FONCTION DE RISQUE D'UN ESTIMATEUR ÉQUIVARIANT LOCALEMENT MEILLEUR EN $\lambda_0^2 = 4$ (n =20).

——: *Risque de l'estimateur équivariant localement meilleur en λ_0^2 (MEE).*

— — —: *Risque de l'estimateur linéaire localement meilleur en λ_0^2 (MEL).*

······ : *Risque de l'estimateur du maximum de vraisemblance localement meilleur en λ_0^2 (MEV).*



Graphique 6

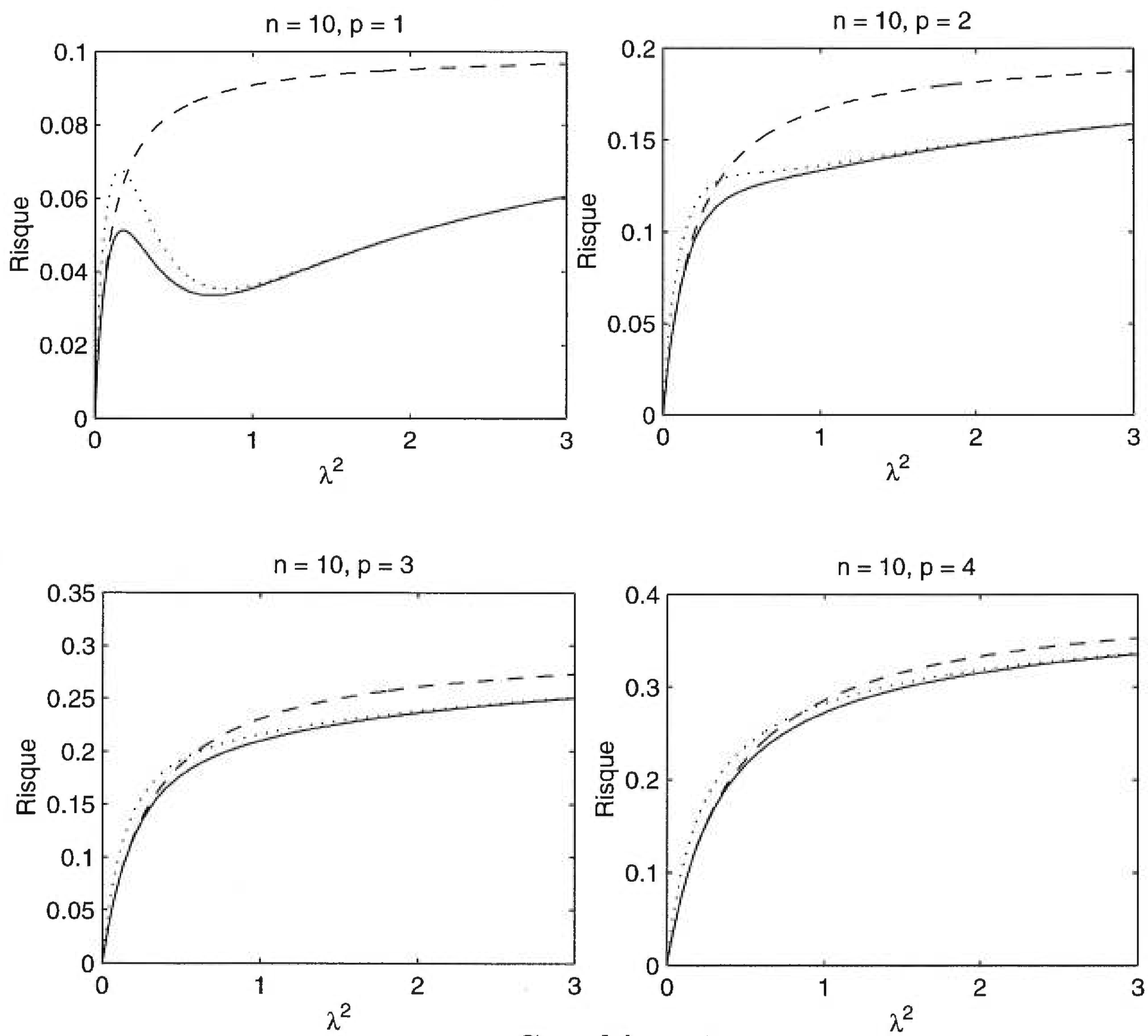
RISQUE MINIMAL D'UN ESTIMATEUR ÉQUIVARIANT

$(n = 10)$.

———: *Risque minimal de l'estimateur équivariant.*

— — —: *Risque minimal de l'estimateur linéaire.*

······ : *Risque minimal de l'estimateur du maximum de vraisemblance.*



Graphique 7

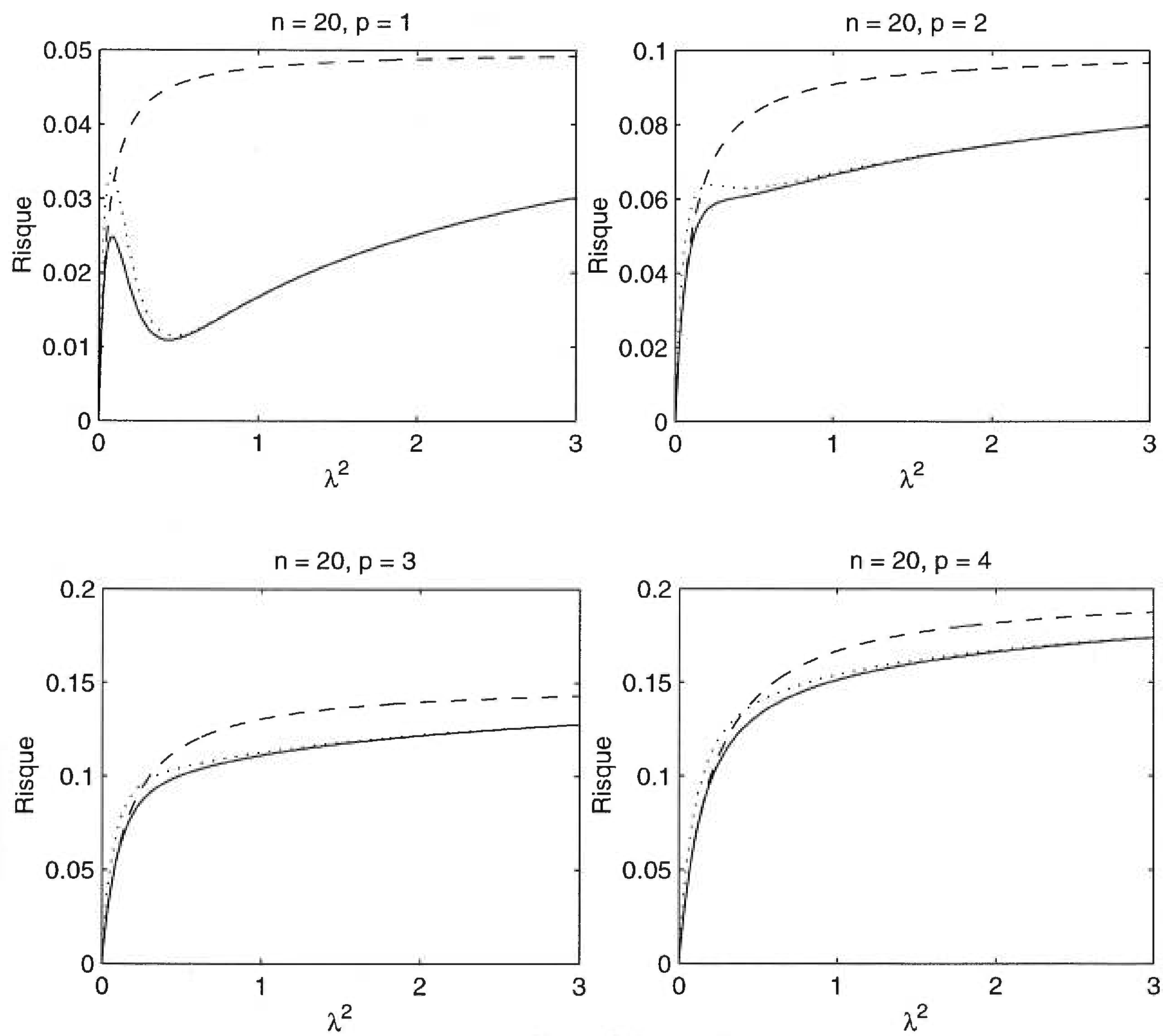
RISQUE MINIMAL D'UN ESTIMATEUR ÉQUIVARIANT

$(n = 20)$.

———: *Risque minimal de l'estimateur équivariant.*

— — —: *Risque minimal de l'estimateur linéaire.*

······ : *Risque minimal de l'estimateur du maximum de vraisemblance.*



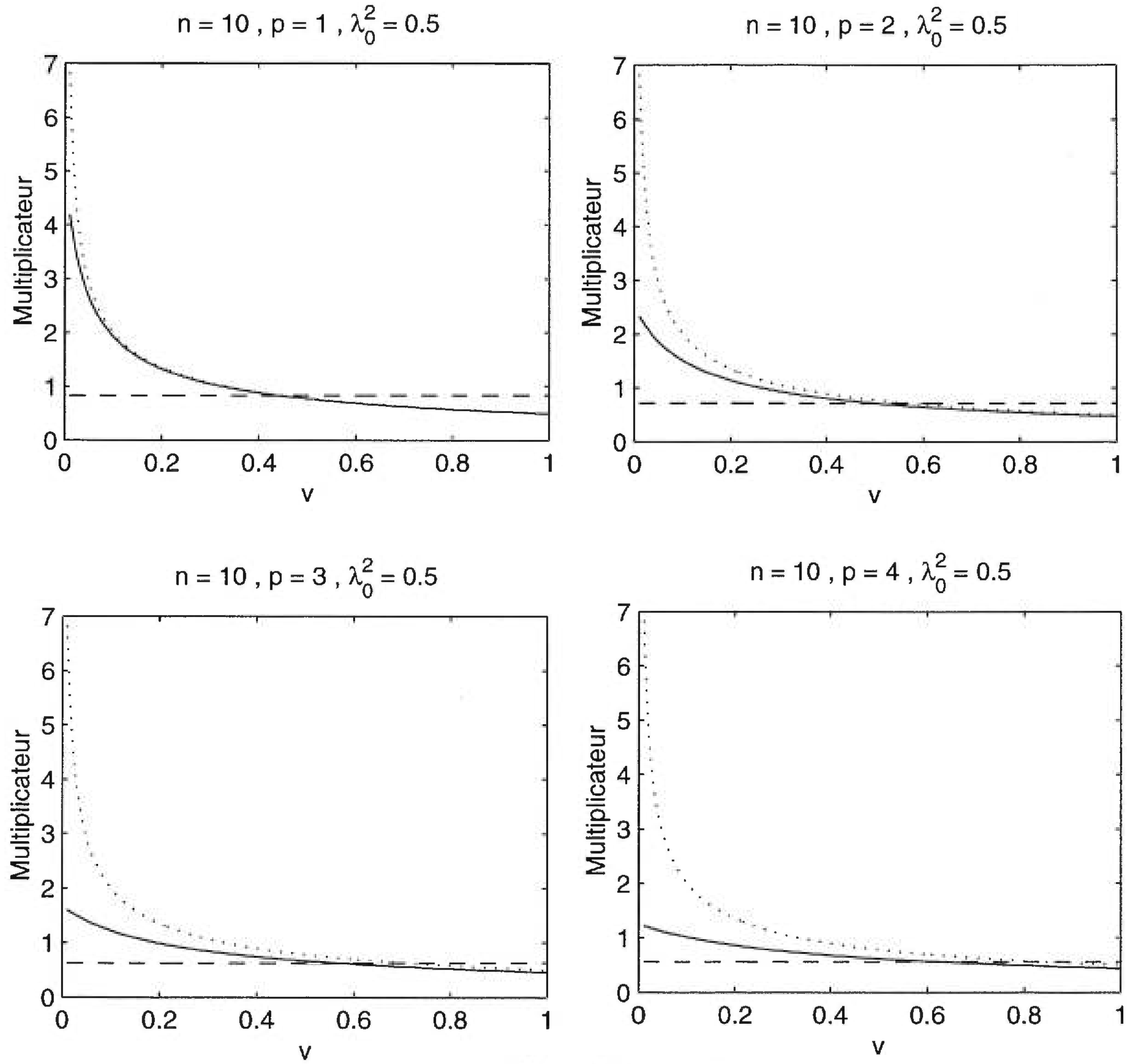
Graphique 8

ÉVALUATION DU MULTIPLICATEUR OPTIMAL LORSQUE $\lambda^2 \neq \lambda_0^2$ $(n = 10)$.

———: *Multiplicateur équivariant optimal.*

— — — : *Multiplicateur linéaire optimal.*

$\cdots\cdots$: *Multiplicateur du maximum de vraisemblance optimal.*



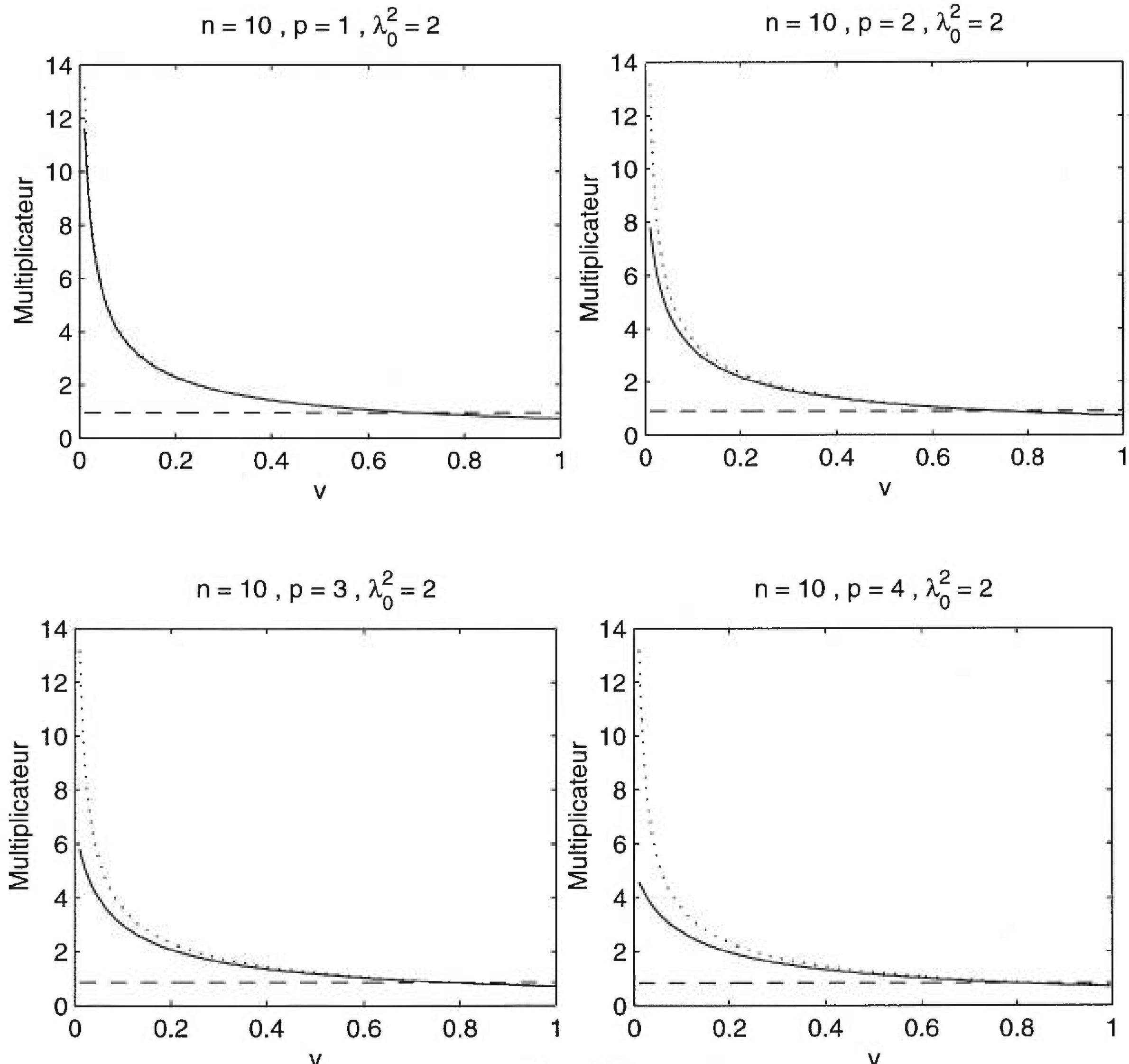
Graphique 9

ÉVALUATION DU MULTIPLICATEUR OPTIMAL LORSQUE $\lambda^2 \neq \lambda_0^2$ $(n = 10)$.

———: *Multiplicateur équivariant optimal.*

— — —: *Multiplicateur linéaire optimal.*

······ : *Multiplicateur du maximum de vraisemblance optimal.*



Graphique 10

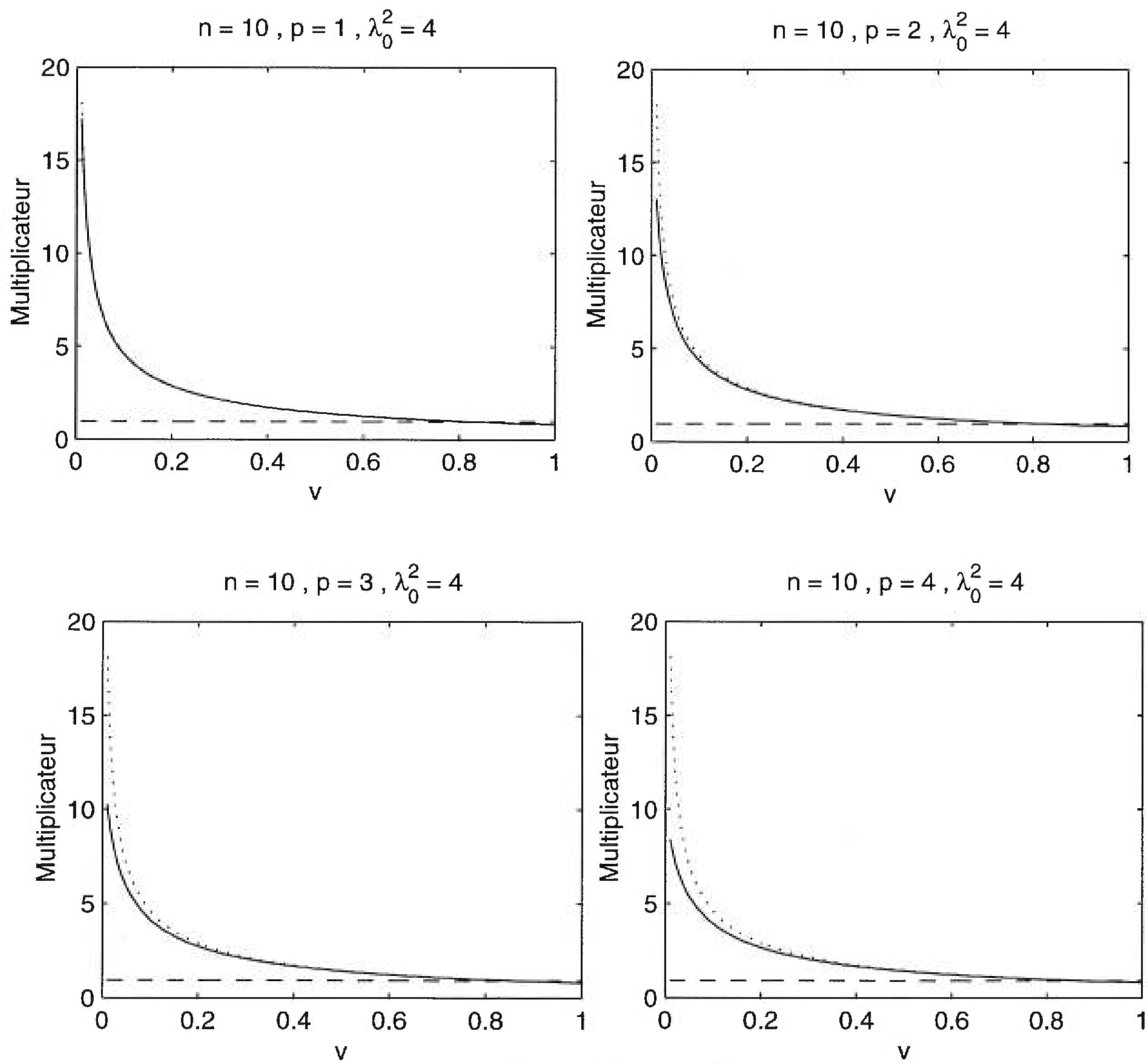
ÉVALUATION DU MULTIPLICATEUR OPTIMAL

LORSQUE $\lambda^2 \neq \lambda_0^2$ $(n = 10)$.

———: *Multiplicateur équivariant optimal.*

– – –: *Multiplicateur linéaire optimal.*

······: *Multiplicateur du maximum de vraisemblance optimal.*



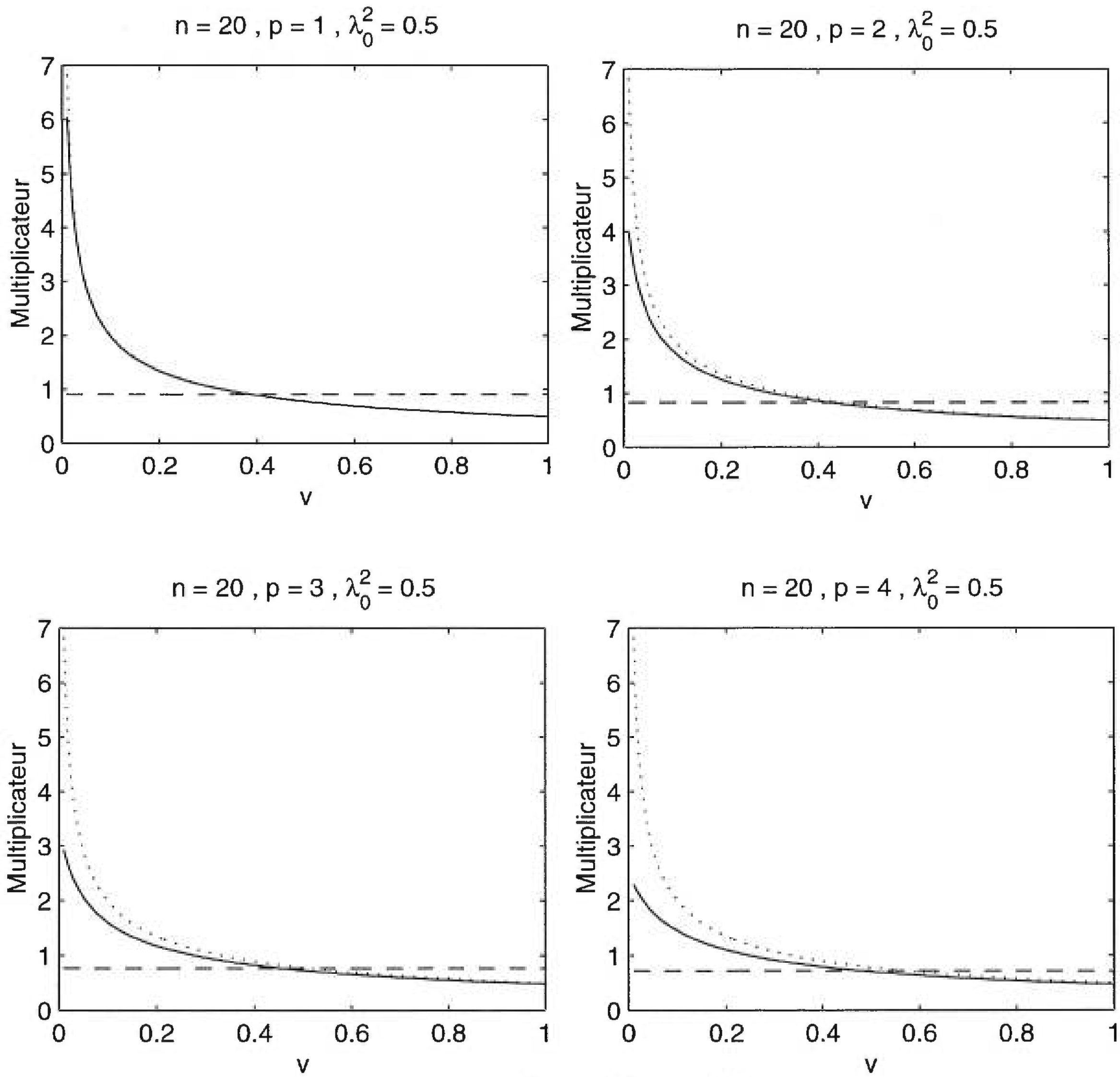
Graphique 11

ÉVALUATION DU MULTIPLICATEUR OPTIMAL LORSQUE $\lambda^2 \neq \lambda_0^2$ $(n = 20)$.

——— : *Multiplicateur équivariant optimal.*

— — — : *Multiplicateur linéaire optimal.*

⋯⋯ : *Multiplicateur du maximum de vraisemblance optimal.*



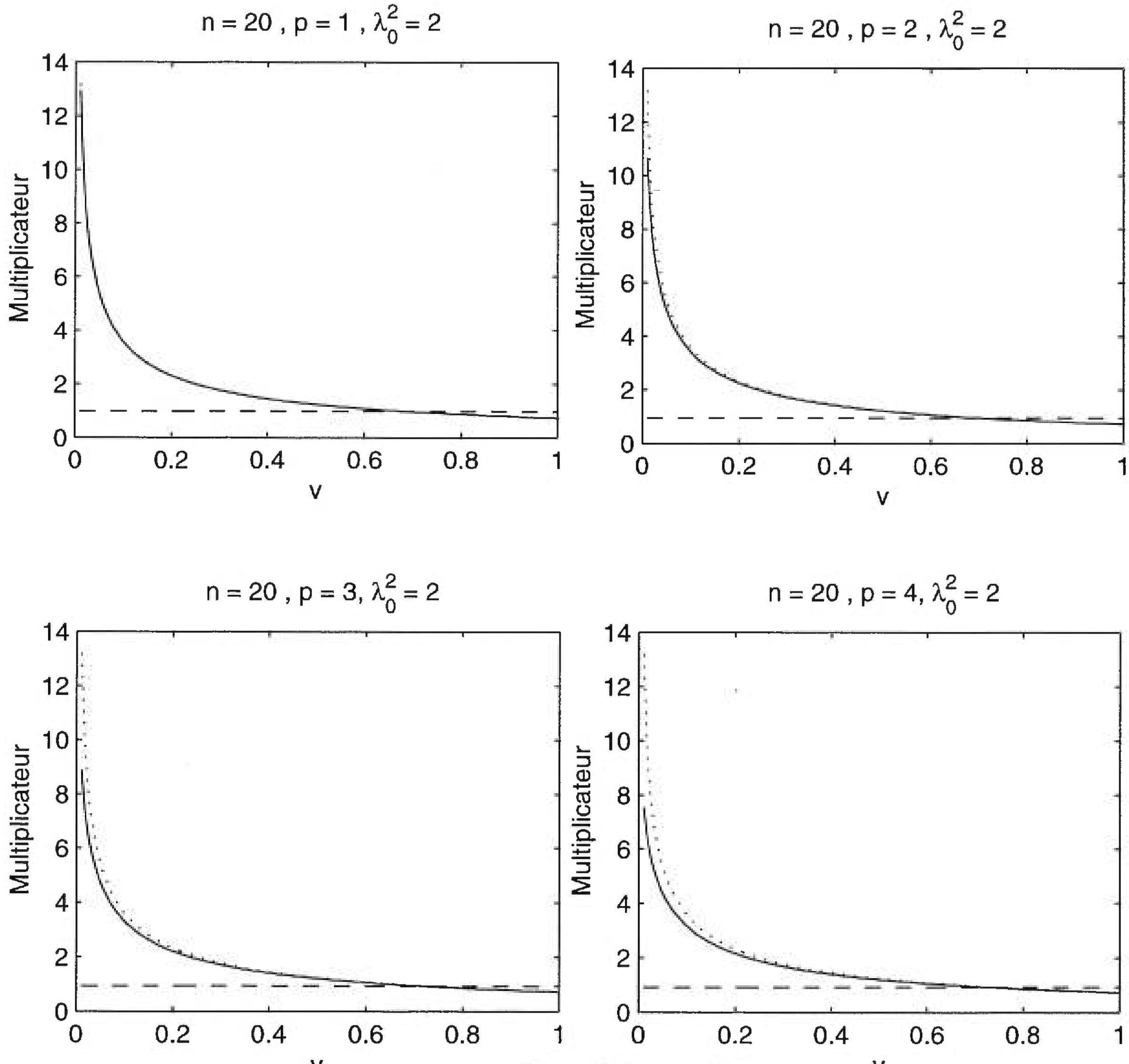
Graphique 12

ÉVALUATION DU MULTIPLICATEUR OPTIMAL LORSQUE $\lambda^2 \neq \lambda_0^2$ $(n = 20)$.

——— : *Multiplicateur équivariant optimal.*

– – – : *Multiplicateur linéaire optimal.*

······ : *Multiplicateur du maximum de vraisemblance optimal.*



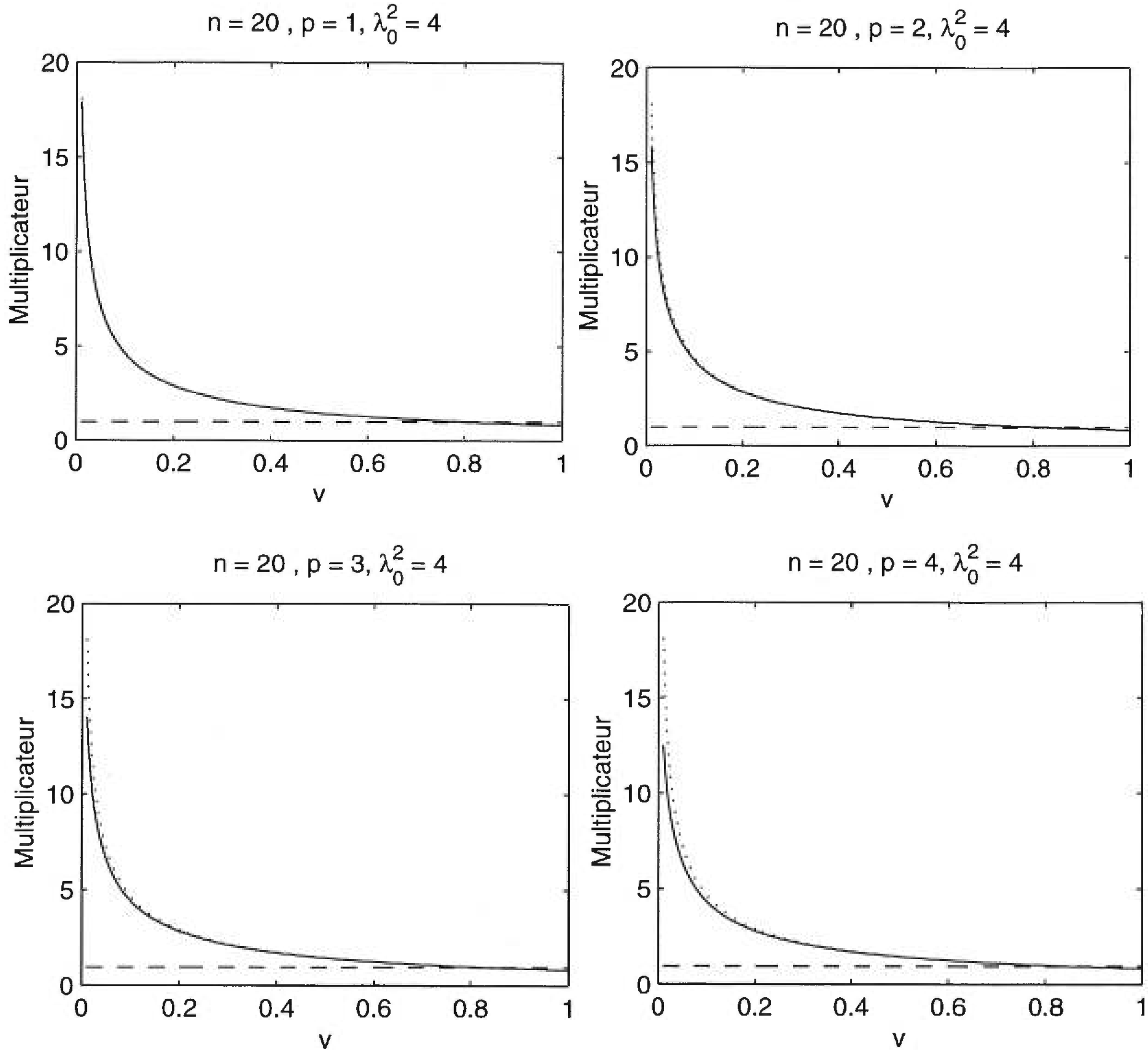
Graphique 13

ÉVALUATION DU MULTIPLICATEUR OPTIMAL LORSQUE $\lambda^2 \neq \lambda_0^2$ ($n = 20$).

———: *Multiplicateur équivariant optimal.*

— — —: *Multiplicateur linéaire optimal.*

······: *Multiplicateur du maximum de vraisemblance optimal.*



Graphique 14

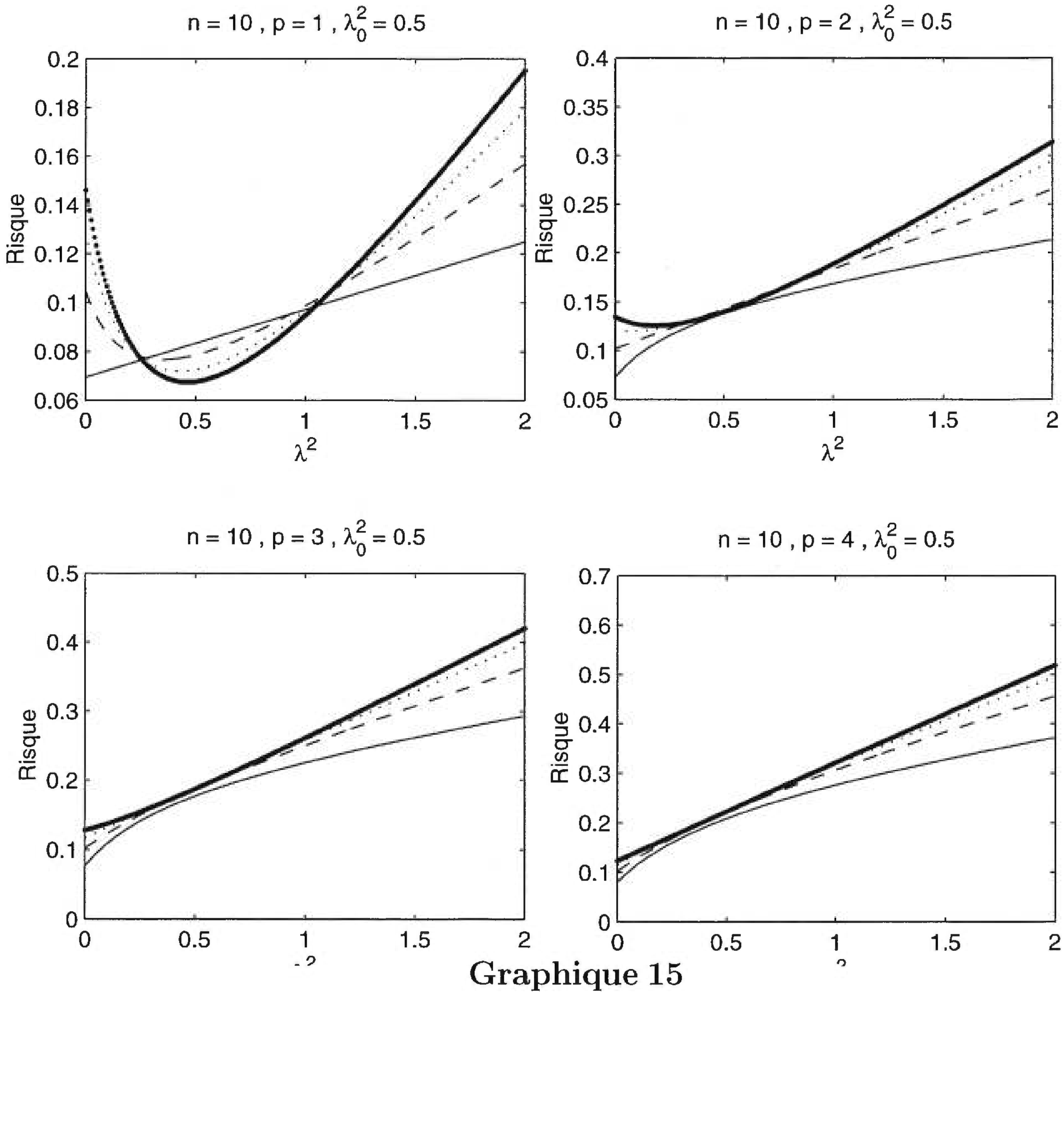
FONCTION DE RISQUE D'UN ESTIMATEUR ÉQUIVARIANT LOCALEMENT MEILLEUR EN $\lambda_0^2 = 0.5$. (cas d'une loi de student) $(n = 10)$.

———: *Risque de l'estimateur localement meilleur* $(m = 3)$.

– – –: *Risque de l'estimateur locallement meilleur* $(m = 4)$.

······: *Risque de l'estimateur localement meilleur* $(m = 5)$.

●●●●●●: *Risque de l'estimateur localement meilleur* $(m = 6)$.



Graphique 15

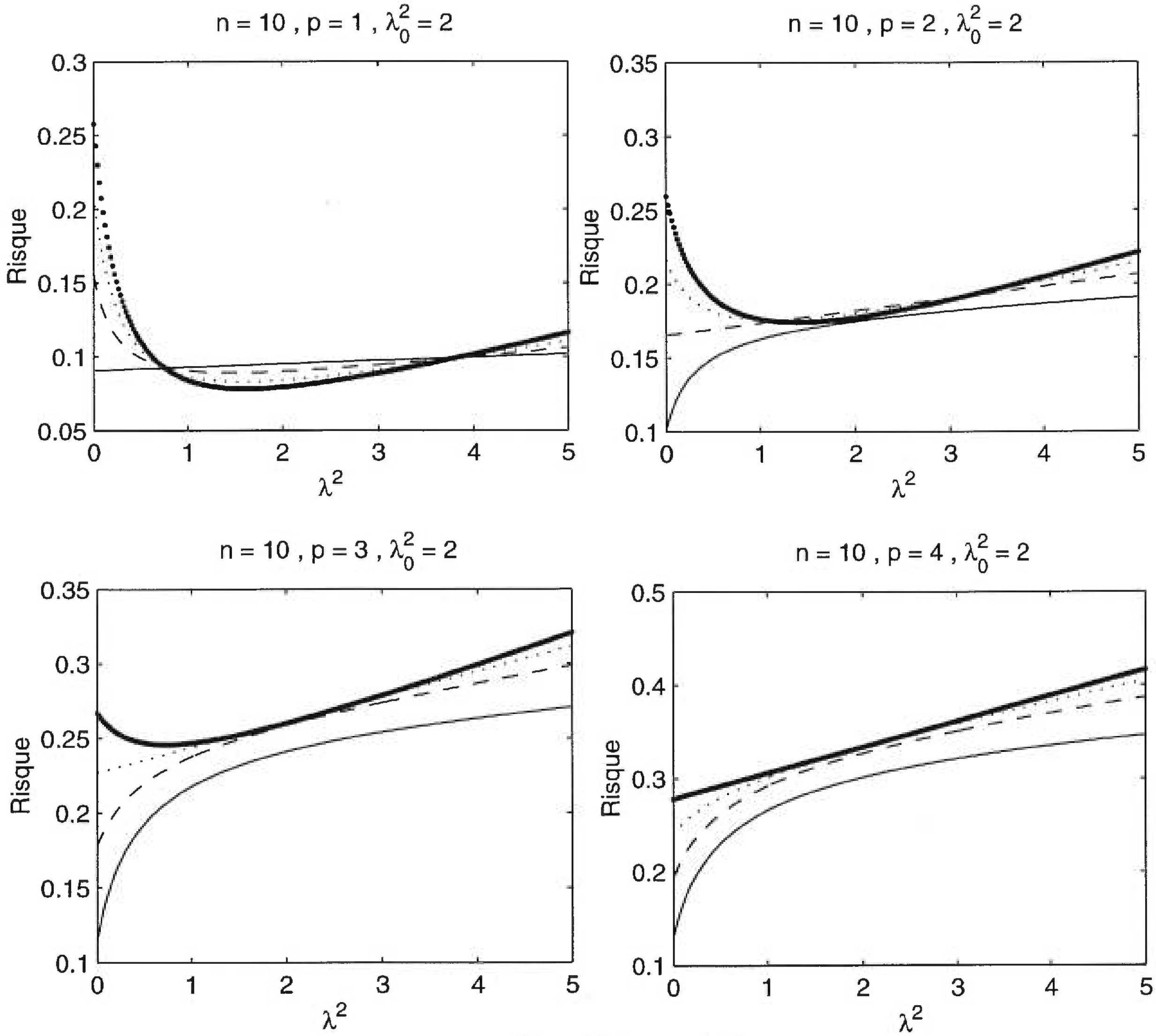
FONCTION DE RISQUE D'UN ESTIMATEUR ÉQUIVARIANT LOCALEMENT MEILLEUR EN $\lambda_0^2 = 2$. (cas d'une loi de student) $(n = 10)$.

———: *Risque de l'estimateur localement meilleur* $(m = 3)$.

– – –: *Risque de l'estimateur locallement meilleur* $(m = 4)$.

······: *Risque de l'estimateur localement meilleur* $(m = 5)$.

●●●●●●: *Risque de l'estimateur localement meilleur* $(m = 6)$.



Graphique 16

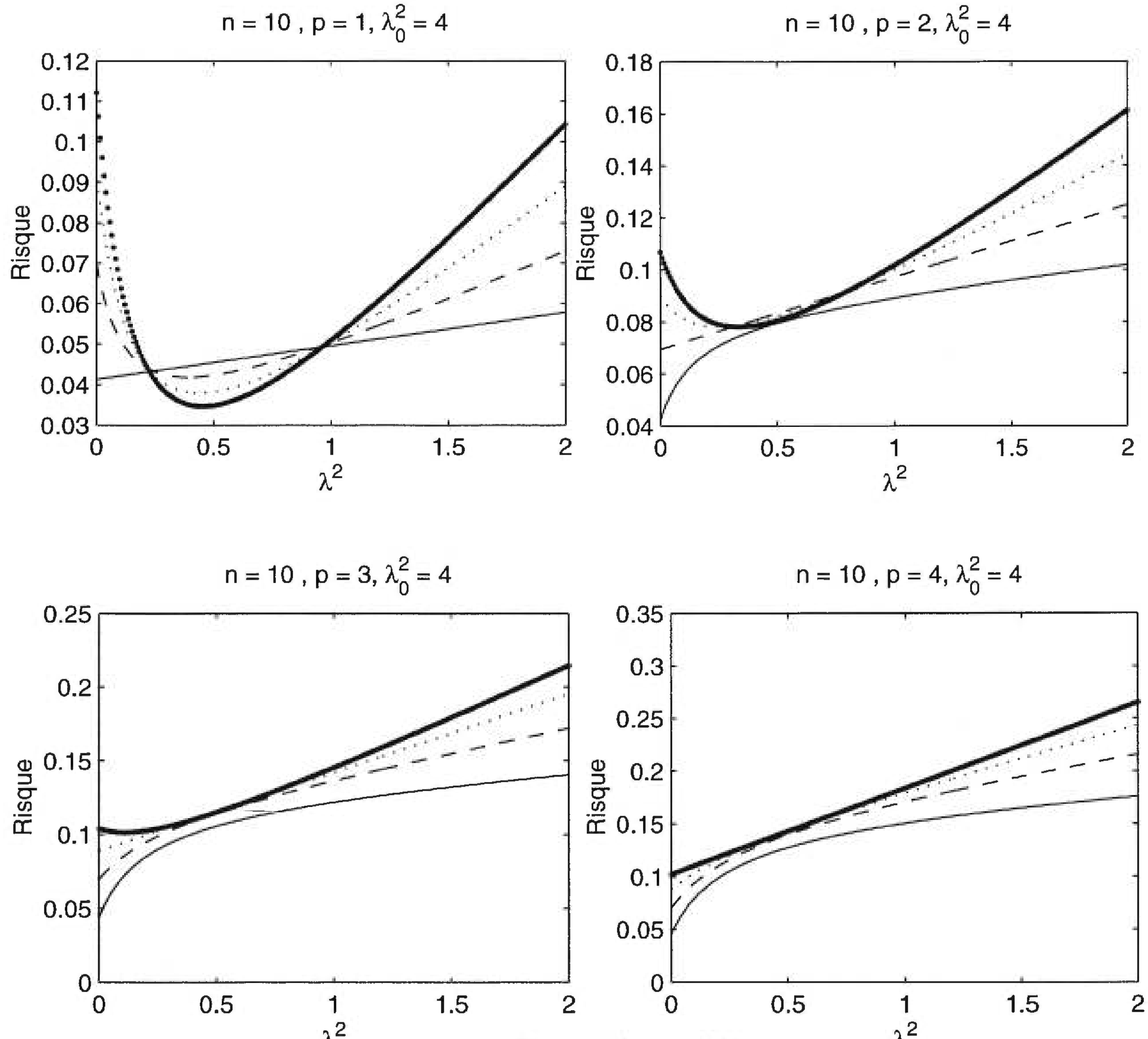
FONCTION DE RISQUE D'UN ESTIMATEUR ÉQUIVARIANT LOCALEMENT MEILLEUR EN $\lambda_0^2 = 4$. (cas d'une loi de student) $(n = 10)$.

———: *Risque de l'estimateur localement meilleur (m = 3).*

– – – : *Risque de l'estimateur locallement meilleur (m = 4).*

······ : *Risque de l'estimateur localement meilleur (m = 5).*

●●●●●● : *Risque de l'estimateur localement meilleur (m = 6).*



Graphique 17

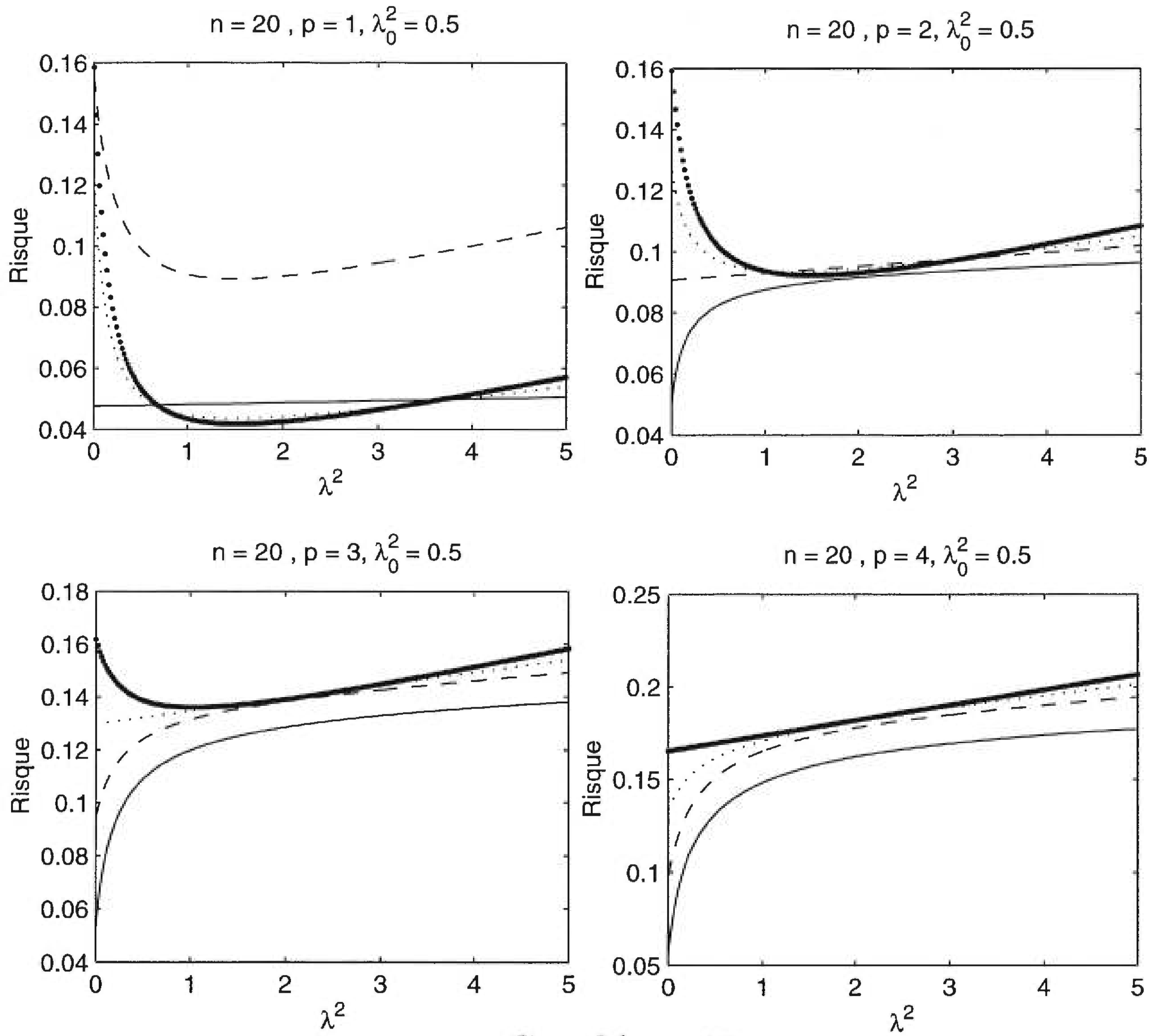
FONCTION DE RISQUE D'UN ESTIMATEUR ÉQUIVARIANT LOCALEMENT MEILLEUR EN $\lambda_0^2 = 0.5$. (cas d'une loi de student) ($n = 20$).

———: *Risque de l'estimateur localement meilleur ($m = 3$).*

– – – : *Risque de l'estimateur locallement meilleur ($m = 4$).*

$\cdots\cdots$: *Risque de l'estimateur localement meilleur ($m = 5$).*

●●●●●● : *Risque de l'estimateur localement meilleur ($m = 6$).*



Graphique 18

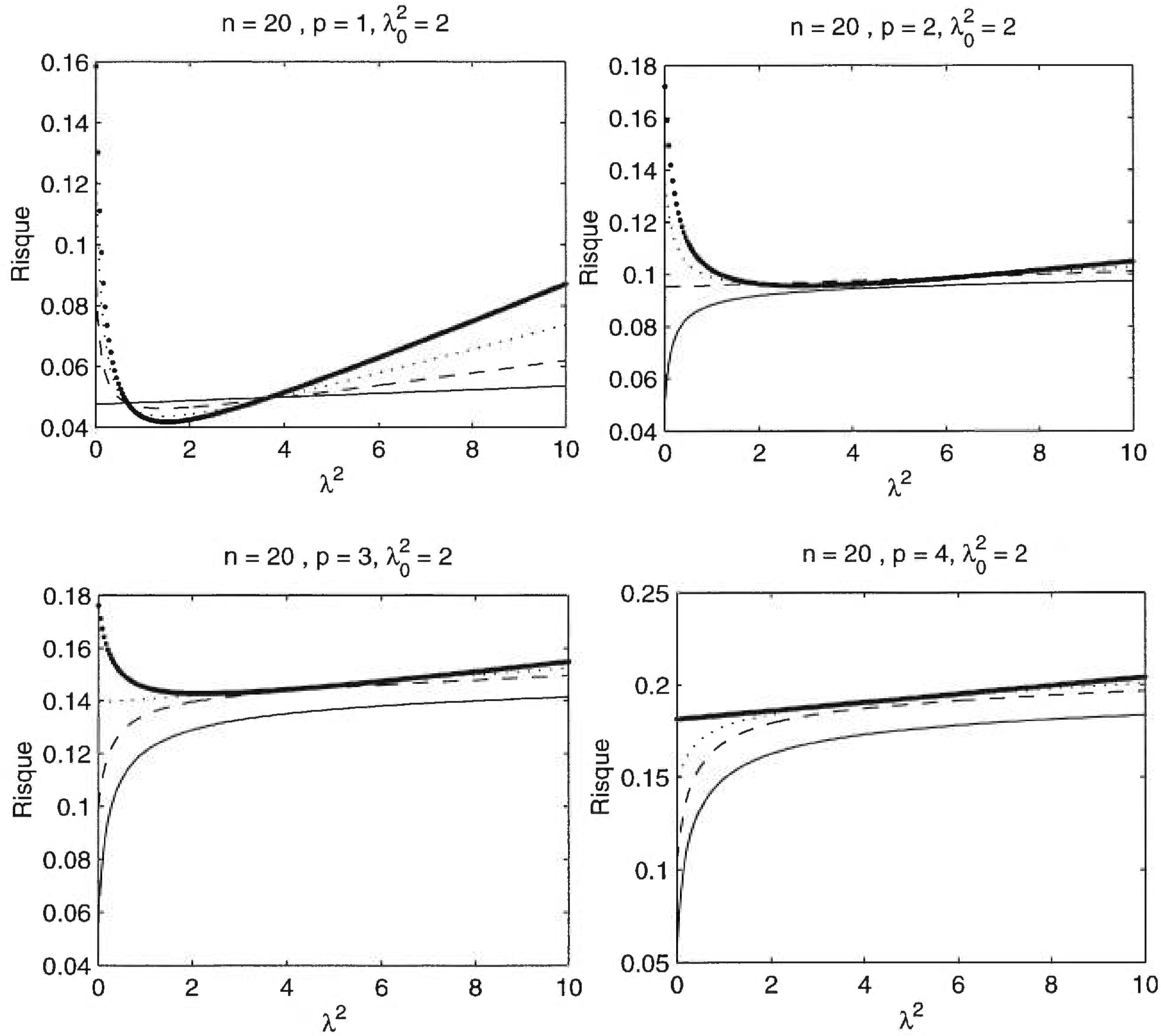
FONCTION DE RISQUE D'UN ESTIMATEUR ÉQUIVARIANT LOCALEMENT MEILLEUR EN $\lambda_0^2 = 2$. (cas d'une loi de student) $(n = 20)$.

———: *Risque de l'estimateur localement meilleur ($m = 3$).*

– – –: *Risque de l'estimateur locallement meilleur ($m = 4$).*

······: *Risque de l'estimateur localement meilleur ($m = 5$).*

●●●●●●: *Risque de l'estimateur localement meilleur ($m = 6$).*



Graphique 19

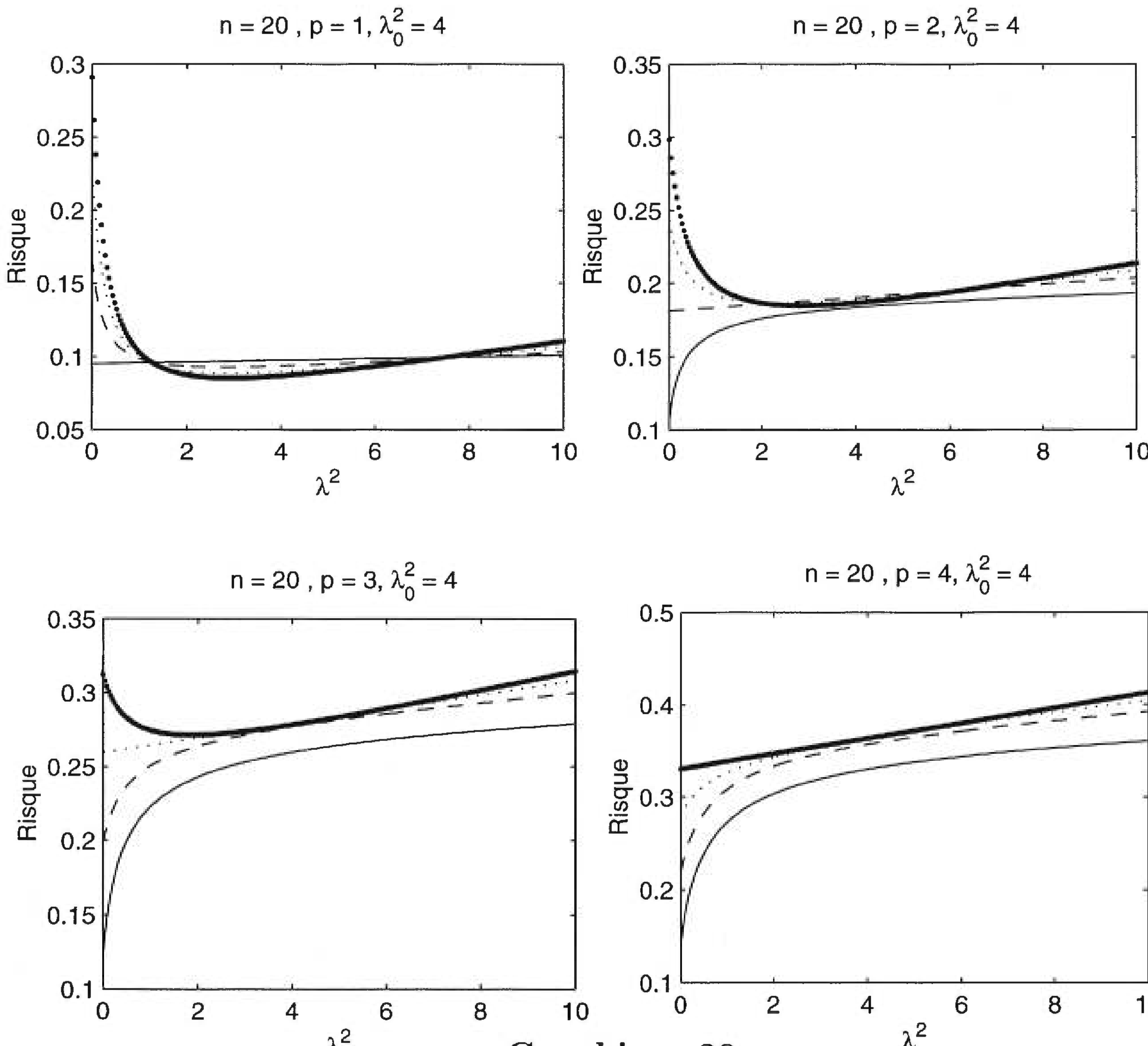
FONCTION DE RISQUE D'UN ESTIMATEUR ÉQUIVARIANT LOCALEMENT MEILLEUR EN $\lambda_0^2 = 4$. (cas d'une loi de student) ($n = 20$).

———: *Risque de l'estimateur localement meilleur ($m = 3$).*

— — —: *Risque de l'estimateur locallement meilleur ($m = 4$).*

······: *Risque de l'estimateur localement meilleur ($m = 5$).*

●●●●●●: *Risque de l'estimateur localement meilleur ($m = 6$).*



Graphique 20

RISQUE MINIMAL D'UN ESTIMATEUR ÉQUIVARIANT.

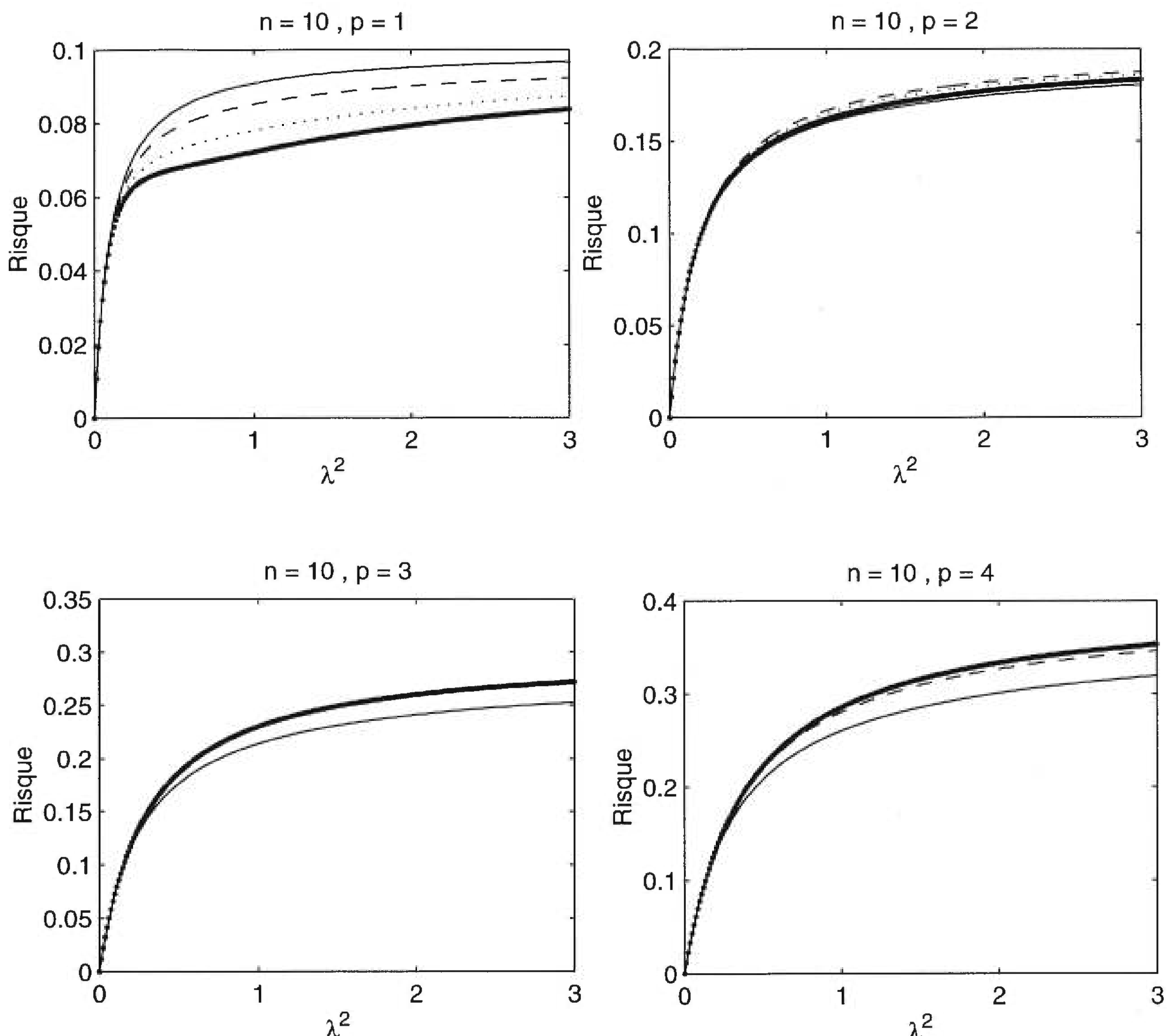
(cas d'une loi de student) ($n = 10$).

———: *Risque minimal de l'estimateur équivariant ($m = 3$).*

– – –: *Risque minimal de l'estimateur équivariant ($m = 4$).*

······: *Risque minimal de l'estimateur équivariant ($m = 5$).*

●●●●●●: *Risque minimal de l'estimateur équivariant ($m = 6$).*



Graphique 21

RISQUE MINIMAL D'UN ESTIMATEUR ÉQUIVARIANT.

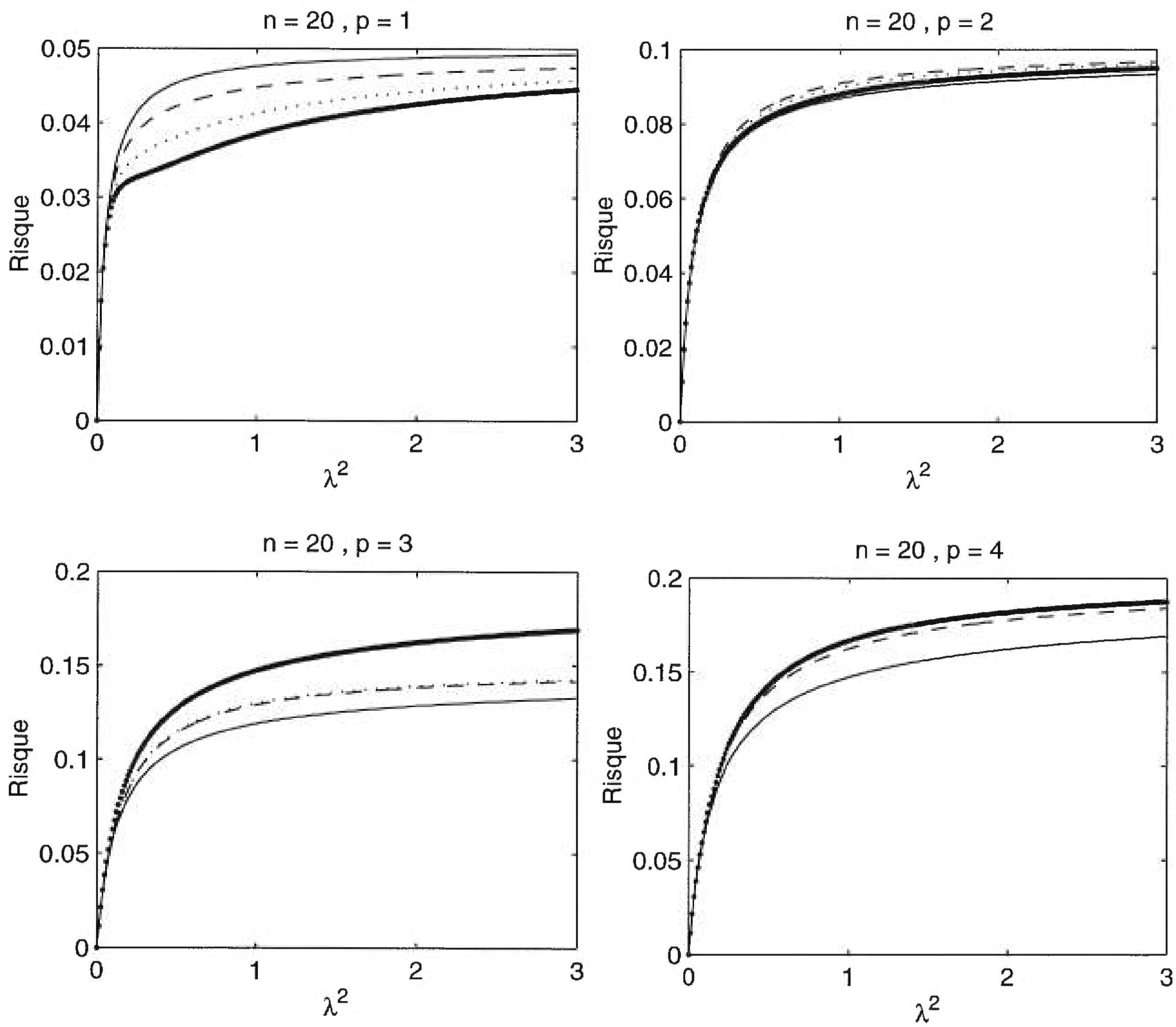
(cas d'une loi de student) ($n = 20$).

———: *Risque minimal de l'estimateur équivariant ($m = 3$).*

— — —: *Risque minimal de l'estimateur équivariant ($m = 4$).*

⋯⋯ : *Risque minimal de l'estimateur équivariant ($m = 5$).*

●●●●●● : *Risque minimal de l'estimateur équivariant ($m = 6$).*



Graphique 22

CONCLUSION

La recherche d'une solution aux problèmes d'estimation de paramètres est parfois très difficile. On s'est restreint à une sous classe des estimateurs, qui est la classe des estimateurs équivariants. Cette restriction permet d'évaluer explicitement le meilleur estimateur (l'estimateur qui rend le risque minimal) et le risque minimal (le risque du meilleur estimateur) dans cette classe. De plus elle permet de mieux dégager les propriétés de ces derniers.

On s'est servi de l'approche de Kariya (1989) pour caractériser les estimateurs équivariants, dans sa forme générale, en modèle restreint. Nous avons utilisé cette forme générale pour l'étude des problèmes d'estimation de la moyenne μ d'un mélange de lois normale (μ, Ω, Z), où Z est une variable aléatoire positive dont l'espérance existe, sous certaines contraintes.

Dans un premier temps, nous avons étudié l'estimation de la moyenne d'un mélange de lois normales (μ, Σ, Z), où Z est une variable aléatoire positive dont l'espérance existe, sous la contrainte $\mu' \Sigma^{-1} \mu = \lambda^2$. Cette contrainte représente la distance de Mahalanobis entre $N_p(\mu, \Sigma)$ et $N_p(0, \Sigma)$. Nous avons d'abord analysé le cas particulier d'une loi multinormale. Ensuite nous avons utilisé les résultats du cas normal pour déterminer le meilleur estimateur équivariant et le risque minimal d'un mélange de lois normales. Enfin, nous avons appliqué les résultats obtenus du cas général, au cas particulier d'une loi de Student multivariée (cette loi correspond à une mélange de lois normales et loi gamma).

Dans un second temps, nous avons étudié l'estimation de la moyenne d'un mélange de lois normales $N_p\left(\mu, \frac{\mu^t\mu}{C^2}I\right)$, où C est le coefficient de variation (connu). Nous avons fait la même procédure que le premier cas étudié.

Nous avons évalué le meilleur estimateur et le risque minimal dans la classe des estimateurs linéaires, dans la classe des estimateurs de James-Stein et l'estimateur du maximum de vraisemblance lorsque $\mu^t\Sigma^{-1}\mu = \lambda^2$ pour un mélange de lois normales de paramètres (μ, Ω, Z). Mais nos résultats nous montrent que le meilleur estimateur linéaire et le meilleur estimateur de James-Stein ne dépendent pas de la variable de mélange Z.

Nous avons aussi évalué les fonctions de risque de l'estimateur équivariant localement meilleur en λ_0^2, de l'estimateur linéaire localement meilleur en λ_0^2, de l'estimateur de James-Stein localement meilleur en λ_0^2 et de l'estimateur du maximum de vraisemblance localement meilleur en λ_0^2 d'une loi normale. La comparaison graphique de ces résultats nous montre que ces estimateurs ont des comportements assez semblables. En outre, on constate que l'estimateur linéaire localement meilleur en λ_0^2 est plus stable que les autres compétiteurs.

Nous avons aussi étudié numériquement et graphiquement le risque minimal des estimateurs localement meilleurs en λ_0^2 d'une loi $St_p(\mu, \Sigma, m)$ pour différentes valeurs du degré de liberté m. Cette étude montre, pour $p \neq 1$ et pour tout n et λ_0^2, que la fonction de risque évaluée en $m = 3$ est plus petite que les autres fonctions de risque.

BIBLIOGRAPHIE

Abramowitz, M. & Stegun, I. (1964), Handbook of mathematical functions with formlas, graphs and mathematical tabes. *National Bureau of Standard Applied Mathematics,* **serie 55**.

Anderson, T.W. (1984), *An introduction to multivariate statistical analysis*. Wiley. New York

Andersson, S.A. (1982), Distributions of maximal invariants using quotient measures, *Annals of Statistics* **10**, 955-961.

Berger, J. (1975), Minimax estimation of location vectors for a wide class of densities, *Annals of Statistics* **3**, 1318-1328.

Berger, J. (1985), *Statistical decision theory and bayesian analysis*. Springer-Verlag, New York.

Bondar, J. V. (1976), Borel cross-sections and maximal invariants, *Annals of Statistics* **4**, 866-877.

Bondar J.V. & Miles P. (1981), Amenability : A survey for statistical applications of Hunt-Stein and related conditions on groups, *Zeitschrift für wahrs-chein lichkeits theorie* **57**, 103-128.

Bourbaki, N. (1963), *Eléments de mathématique* 29, *Intégration chapitres 7, 8*. Hermann, Paris.

Brown, L (1980), *A Hunt-Stein theorem*. (Manuscrit).

Cox D.R & Hinkley, D.V. (1974), *Theoretical statistics*. Chapman & Hall, London.

Eaton, M. L. (1983), *Multivariate statistics : A vector approach*. Wiley, New York.

Eaton, M. L. (1989), *Group invariance application in statistics.* Regional Conference Series in Probability and Statistics. Institut Mathematical Statistics. Hayward, California.

Efron, B. (1975), Defining the curvature of a statistical; problem (with application second order efficiency), *Annals of Statistics* **3**. 1189-1242.

Fang K. T., Kotz S. & Ng K. W. (1990), Symmetric multivariate and related distributions. *Biometrics* **46**, 1235-1236.

Farrell, R. H. (1967), Weak limits of Bayes procedures in estimation theory, *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability* **1**, 83-111.

Fisher, R. (1956), *Statistical methods and scientific inference*. Oliver and Boyd, Edinburgh.

Ferguson, T. S. (1967), *Mathematical statistics : A decision theoretic approach.* Academic Press, New York.

Giri, N. C. (1975), Invariance and minimax statistical Tests. *Selecta Statistica. Canadiana* **3**, 1-104.

Giri, N. C. (1995), *Group invariance in statistical inference*. World Scientific. Singapore, New Jersey, London, Hong Kong.

Girshick M.A. & Savage, L.J. (1951), *Bayes and minimax estimates for quadratic loss functuions*. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. Univ. of California Press.

Greenleaf, F. (1969), Amenable actions of locally compact groups, *Journal of Functional Analysis* **4**, 295-315.

James, W. & Stein, C. (1961), Estimation with quadratic loss, *In Proc Fourth Berkeley Symp Math and Statist. Probab* **1**, University of California Press, Berkeley.

Johnson, N.L., Kotz S. & balarifhan N. (1995), *Distributions in statistics : continuous univariate distributions* **2**

Kariya, T. (1989), Equivariant estimation in a model with ancillary Statistics, *Annals of Statistics* **17**, 920-928.

Kariya T., Perron F. & Giri N. C. (1988), Invariant estimation of mean vector μ of $N_p(\mu,\Sigma)$ with $\mu'\Sigma^{-1}\mu=1$ or $\Sigma^{-1/2}\mu=C$ or $\Sigma=\delta^2\mu'\mu I$, *Journal of Multivariate Analysis* **27**, 270-283.

Kelker D. (1970), Distribution theory of spherical distributions and location-scale parameter generalization, *Sankhya A* **32**, 419-430.

Kiefer, J. (1957), Invariance, minimax sequential estimation and continuous time process, *Annals of Mathematical Statistics* **28**, 537-601.

LE Cam, L. (1986), *Asymptotic methods in statistical decision theory.* Springer-Verlag, New York.

LehmanN, E. L. (1986), *Testing statistical hypotheses.* Willey. New York.

Marchand, E. (1990), *Estimation de la moyenne multivariée avec contraintes.* Thèse de doctorat, Département de Mathématiques et de Statistique, Université de Montréal.

Marchand, E. (1993), Estimation of a multivariate mean with constrains on the norm, *Canadian Journal of Statistics* **21**, 359-366.

Marchand, E. (1994), On the estimation of the mean of a $N_p(\mu,\Sigma)$ population with $\mu'\Sigma^{-1}\mu$ known, *Statistics & Probability Letters* **21**, 69-75.

Marchand, E. (1994), *On the estimation of the mean of a* $N_p(\mu,\Sigma)$ *population with* $\mu'\Sigma^{-1}\mu$ *known.* Actes du Collogue Apics.

Muirhead, R. J. (1982), *Aspects of multivariate statistical theory.* Wiley. New York.

Nachbin L. (1965), *The Haar integral.* Van Nostrand, Princeton, New York.

Perron, F. (1987), *Application de l'invariance en théorie de la décision.* Thèse de doctorat, Département de Mathématiques et de Statistique, Université de Montréal.

Perron, F. & Giri N.C. (1990), On the best equivariant estimator of mean of a multtivariate normal population, *Journal of Multivariate Analysis* **32**, 1-16.

Pitman, E. (1938), The estimate of location and scale parameters of continuous population of any given form, *Biometrika* **30**, 391-421.

Rickert, W (1967), Amenable groups with the fixed point property, *Transactions of the American Mathematical Society* **127**, 221-232.

Robert C. (1992), *L'analyse statistique bayesienne*. Economica. Paris.

Segal & Kunze (1978), *Integral and operators*. Springer. New York.

Strasser, H. (1985), *Mathematical theory of statistics*. Walter de Gruyter, Berlin, New York.

Strawderman, W. E. (1974), Minimax estimation of location parameters certain spherically symmetric distributions, *Journal of Multivariate Analysis* **4**, 255-264.

Weil, A. (1951), *L'Intégration dans les groupes topologiques et ses applications*. Herman, Paris.

Wesler, O. (1959), Invariance theory and modified minimax principle, *Annals of Mathematical Statistics* **30**, 1-20.

Wijsman, R. A. (1985), Proper action in steps, with application to density ratios of maximal invariants, *Annals of. Statistics* **13**, 395-402.

Wijsman, R. A. (1990), *Invariant measure on groups and their use in statistics*. Institut Mathematical Statistics. Hayward, California.