N° Attribué par la bibliothèque

République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieure et de la recherche scientifique

Année univ.: 2016/2017

Sur l'estimation sous certainnes classes de fonction de perte équilibrée géneralisée

Mémoire présenté en vue de l'obtention du diplôme de

Master Académique

Université Dr Tahar Moulay - Saïda

Discipline : MATHÉMATIQUES

Spécialité : ASSPA

par

Cherifi Omar El Farouk [1]

1. e-mail : cherifi.omar74@gmail.com

Sous la direction de

Dr.Kouider Djerfi

Soutenue le 24 Mai 2017 devant le jury composé de

N. Ait Ouali	Université Dr Tahar Moulay - Saïda	Président
K. Djerfi	Université Dr Tahar Moulay - Saïda	Encadreur
F. Madani	Université Dr Tahar Moulay - Saïda	Examinateur
F. Mokhtari	Université Dr Tahar Moulay - Saïda	Examinatrice

Remerciements

*Je remercie, en premier lieu, notre **Dieu** qui nous a donné de la force pour effectuer le présent travail.*

*En second lieu, je tenée à adresser mon remerciement à mon encadreur **D.r.Kouider Djerfi** pour son aide et ses conseils, en saluant en lui son savoir faire, ses compétences et ses connaissances dont il m'a fait en profiter.*

*Je remercie mes enseignant examinateur de mon mémoire **N.Ait ouali**, **F.Madani** et **F.Mokhtari** qui nous a fait l'honneur de bien vouloir accepter de juger ce modeste travail.*

Je remercie mes très chers parents, qui ont toujours été là pour moi.

*J'adresse mes sincères remerciements à tous Les membres du laboratoire de mathématiques et a la tête le directeur **Pr.A.K**andouci.*

Je remercie également touts mes enseignants et les enseignantes de départements de Mathématiques qui ont Aidez-nous à succès, par leurs paroles, leurs écrits, leurs conseils et leurs critiques.

Merci à tous les collègues étudiantes et étudiants de Mathématiques. Enfin, à tous les personnes qui me aidé lors de la réalisation de ce modeste travail.

«Merci à tous»

Table des matières

Introduction

La formule de Bayes est apparue pour la première fois en 1761, dans le cadre de l'exemple binomial, exposé par Thomas Bayes, et publié de façon posthume par son ami R. Price en 1763. Pierre Simon Laplace redécouvrit ensuite cette formule dans une plus grande généralité en 1773, sans, semble-t-il, avoir connaissance des travaux précédents de Bayes. L'utilisation du principe bayésien devient alors courant pendant le siècle suivant, comme le rapporte Stigler (1986), mais des critiques commencèrent à èmerger vers la fin du **XIX**ème siècle, comme par exemple dans Venn (1886) ou Bertrand (1889), en particulier sur le choix de la loi a priori uniforme et des paradoxes de reparamétrisation qui en résultent.

Puis, malgré des formalisations plus poussées du paradigme bayésien par Edgeworth et Karl Pearson au tournant du siècle et plus tard par Keynes (1921), le début du **XX**ème siècle fut surtout marqué par

Tout d'abord,Kolmogorov, qui proposa dans les années 1920 une axiomatisation de la théorie des probabilités semblant contredire le paradigme bayésien et la notion de probabilité subjective, ensuite par Fisher qui s'éloigna de l'approche bayésienne en 1912 et définissant la fonction de vraissemblance en 1922, puis en développant la Statistique fiduciaire en 1930, et qui ne révisa jamais son opinion négative sur la Statistique bayésienne. Cette opposition paraît quelque peu paradoxale, car la Statistique fiduciaire tentait, en un certain sens de surmonter la difficulté de choisir une loi a priori en la construisant à partir de la fonction de vraisemblance (Seidenfeld en 1992), dans le même esprit que les approches non informatives de Jeffreys (1939) et Bernardo (1979).

Le phénomène de Stein est dû à l'utilisation de la fonction de coût quadratique, qui permet à l'estimateur dominant de tirer profit des autres composantes, même si celles-ci sont indépendantes et correspondent à des problèmes d'estimation sans rapport entre eux.

Stein démontré l'admissibilité de l'estimateur usuel pour (p=2) en (1955), et puis James et Stein en (1961) exhibèrent un estimateur qui domine uniformément $\beta_0(X) = X$ sous le cout quadratique pour $p \geq 3$ dans le cas gaussien.

La fonction de perte prend un nouveau sens en (1994), Zellner propose la fonction de perte équilibrée quadratique qui reliée entre les deux critère de comparaison (la qualité d'ajustement de modéle et la précision de l'estimateur), avec une fonction de pois reduite, et un estimteur a priori(reférense) de moindre carré, et puis en (1995) Shalabch proposé la fonction de perte prédictive, qui mesure la correlation entre les deux critère de comparaison, cette dernière fonction introduit le concept simultanée de la prévision des valeurs actuelles et valeurs moyennes, enfin en (2005) Shalabch proposé la fonction de perte équilibré prolongée.

On s'intéresse en génerale au l'estmateur de type Stein et la fonction de perte équilibré dans un cas géneralisée (le choix de perte est arbitraire).

Ce mémoire est partagé en cinq chapitres.

Dans le premier chapitre, nous présentons un rappel sur le paradigme bayésienne et l'estimateur de bayes, puis dans la théorie de décision on a définit la fonction de perte et la fonction de risque ou (coût) avec quelques exemples, et les deux notion principale d'optimalité (l'admissibilité et la minimaxité) avec quelques lemmes et propositions, a la fin de ce chapitre j'ai parlé de l'estimation de coût.

Le premier chapitre donne l'occasion de soulever le paradoxe de Stein, alors le deuxième chapitre expliquer l'idée de Stein en ce qui concerne l'admissibilité de l'estimateur usuel quand (p=2), et l'inadmissibilité quand (p>2) avec des démonstrations détaillées, ce qui explique l'idée fondamentale de l'amélioration de l'estimateur usuel, et j'ai donné la classe de ces estimateurs admissible (la famille de type Stein) et les estimateurs sphériquements symétriques.

les critique qu'il a subis Stein en ce qui concerne la fonction de perte quadratique, et la distribution gaussienne sont indéniable, donc le troisième chapitre illustrer le problème relier a la fonction de perte l'idée la ces donne une fonction ρ en fonction de la fonction de perte quadratique sous des conditions on va étudier l'admissibilité de l'estimateur de type Stein, et a la fin de ce chapitre on va donne une condition concerne l'absence de normalité.

l'utile principale dans les chapitres précédents est la fonction de perte, donc le quatrième chapitre représente le coeur de ce mémoire, il contiendra la nouvelle conception de la fonction de perte c'est la fonction de perte équilibrée quadratique pondérée proposé par Arnold Zellner en 1994, qui besoin d'un estimateur a priori "référence" l'idée ces l'amélioration de l'estimateur usuel et la projection des propriétés d'admissibilité et de minimaxité dans le cas particulière "déséquilibré" (w=0), se qu'on va faire ces la généralisation de cette fonction de perte, a la fin de ce chapitre pour leur utilisation dans l'application de dernier chapitre on a définit la fonction de perte équilibrée prolongée proposé par Shalabch en 2006.

Enfin, le dernier chapitre est consacré à la partie "Applications", dans lequel on va comparer l'estimateur de type Stein avec d'autre estimateur sous certaines classes de fonction de perte équilibrée généralisée, dans le cas ou on a l'absence de normalité.

Chapitre 1

Introduction à la statistique bayesienne

Comparée à la modélisation probabiliste, l'analyse statistique se ramène fondamentalement à une inversion, car elle doit déterminer les causes-réduites aux paramètres du mécanisme probabiliste générateur-à partir des effets-résumés par les observations [1]. En d'autres termes, quand nous observons un phénomène aléatoire contrôlé par le paramètre β, une méthode statistique permet de déduire de ces observations une inférence (c'est-à-dire, en résumé, une caractérisation) sur β, alors que la modélisation probabiliste caractérise le comportement des observations futures conditionnellement à β. Ce caractère d'inversion propre à la Statistique apparaît de façon évidente dans la notion de fonction de **vraissemblance**, car d'un point de vue formel, il s'agit simplement d'une densité réécrite dans le bon ordre,

$$L(\beta, x) \quad = \quad f(x, \beta) = \prod_{i=1}^{n} (f(x_i, \beta) \tag{1.1}$$

soit donc comme fonction de β qui est inconnu, dépendant de la valeur observée x.

1. À l'époque de Bayes et de Laplace, c'est-à-dire à la fin du **XVIII**éme siècle, la Statistique était souvent appelée Probabilités inverses, à cause de cette perspective.

Historiquement l'approche fiduciaire de Fisher (1956) se fonde aussi sur cette inversion . Une description générale de l'inversion des probabilités est donnée par le théorème de Bayes : Si A et E sont des événements tels que $\mathbb{P}(E) \neq 0, \quad \mathbb{P}(A/E) \quad et \quad \mathbb{P}(E/A)$ sont reliés par :

$$\begin{aligned}
\mathbb{P}(A/E) &= \frac{\mathbb{P}(E/A)\mathbb{P}(A)}{\mathbb{P}(E/A)\mathbb{P}(A) + \mathbb{P}(E/A^c)\mathbb{P}(A^c)} \\
&= \frac{\mathbb{P}(E/A)\mathbb{P}(A)}{\mathbb{P}(E)}
\end{aligned}$$

En particuler,

$$\frac{\mathbb{P}(A/E)}{\mathbb{P}(B/E)} = \frac{\mathbb{P}(E/A)}{\mathbb{P}(E/B)} \tag{1.2}$$

quand $\mathbb{P}(A) = \mathbb{P}(B)$. Obtenir ce résultat à partir des axiomes de la Théorie des Probabilités est trivial. Il s'agit cependant de l'étape conceptuelle la plus importante dans l'histoire de la Statistique, constituant la première inversion des probabilités. L'équation (1.2) exprime le fait fondamental que, pour deux causes équiprobables, le rapport des probabilités pour un effet donné est égal au rapport des probabilités de ces deux causes. Ce théorème est aussi un principe d'actualisation, car il décrit la mise à jour de la **vraissemblance** de A de $\mathbb{P}(A)$ vers $\mathbb{P}(A/E)$, une fois que E a été observé. Bayes (1763) donne en réalité une version continue de ces résultats, à savoir, pour deux variables aléatoires x et y, de distributions conditionnelle[2] $f(x/y)$ et marginale $g(y)$, la distribution conditionnelle de y sachant x est

$$g(y/x) = \frac{f(x/y)g(y)}{\int f(x/y)g(y)dy}$$

Bien que ce théorème d'inversion soit naturel d'un point de vue probabiliste, Bayes et Laplace sont allés plus loin et ont considéré que l'incertitude sur le

2. Souvent nous remplacerons distribution par densité, supposant que plus tard le concept sera mieux défini par rapport à la mesure naturelle dominante, comme la mesure de Lebesgue. C'est seulement dans un contexte plus avancé, comme pour la mesure de Haar (une connaissance plus approfondie de la théorie de la mesure sera nécessaire).

paramètre β d'un modèle peut être décrite par une distribution de probabilité π sur Θ, appelée distribution a priori. L'inférence est alors fondée sur la distribution de β conditionnelle à x, $\pi(\beta/x)$, appelée distribution a posteriori définie par

$$\pi(\beta/x) = \frac{f(x/\beta)\pi(\beta)}{\int f(x/\beta)\pi(\beta)d\beta} \tag{1.3}$$

Notons que $\pi(\beta/x)$ est ainsi proportionnelle à la distribution de x conditionnellement à β, qui est aussi la vraissemblance, multipliée par la distribution a priori de β.(la généralité de (1.3) été perçue par Laplace , qui la développera plus avant). La contribution principale apportée par un modèle statistique bayesien est donc de considérer en sus une distribution aléatoire pour les paramètres.

Un modèle statistique bayesien est constitué d'un modèle statistique paramétrique $f(x, \beta)$, et d'une distribution a priori pour les paramètres $\pi(\beta)$.

En termes statistiques, le théorème de Bayes actualise donc l'information sur β en extrayant l'information contenue dans l'observation x. Son impact provient de la décision audacieuse de mettre causes et effets sur le même niveau conceptuel, puisque les deux sont aléatoires. Du point de vue de la modélisation statistique, il y a donc peu de différences entre observations et paramètres, car les manipulations conditionnelles permettent l'échange de leurs rôles respectifs. Notons que historiquement, cette idée que les paramètres sont aléatoires peut être perçue comme allant à l'encontre du déterminisme athée de Laplace [3], ainsi que des conceptions religieuses de Bayes, qui était un ecclésiastique non-conformiste. En imposant cette modification fondamentale de la perception du phénomène aléatoire, ces deux mathématiciens ont créé l'analyse statistique moderne et plus particulièrement, l'analyse bayésienne.

3. "Nous devons envisager l'état présent de l'Univers comme un effet de l'état antérieur et comme la cause de l'état suivant." -Laplace (1795).

Notons qu'il existe aussi une approche bayésienne de la statistique non paramétrique. Elle met généralement en oeuvre des distributions a priori sur des espaces fonctionnels comme les processus de Dirichlet.

1.1 les principes bayesien

1.1.1 L'inférence bayésienne

La statistique bayésienne est une approche statistique fondée sur l'inférence bayésienne. On utilise le terme de bayésienne pour la différencier de la statistique fréquentiste (ou statistique classique) qui ne sait traiter que les grands échantillons (où elle donne les mêmes résultats que la bayésienne par des procédés moins coûteux en calcul). La statistique bayésienne est surtout utilisée lorsque l'on n'a que des petits échantillons, typiquement quand chaque observation est elle-même très coûteuse (par exemple campagne de prospection pétrolière par voie sismique). Contrairement à la statistique classique, elle n'exige pas au départ qu'on se fixe une hypothèse précise à confirmer ou infirmer, ce qui la rend utile en L'exploration de données [4] L'inférence statistique consiste à induire les caractéristiques inconnues d'une population à partir d'un échantillon issu de cette population, et L'inférence bayésienne est une méthode d'inférence permettant de déduire la probabilité d'un événement à partir de celles d'autres événements déjà évalués. Elle s'appuie principalement sur le théorème de Bayes,

Définition 1.1.1. *Un modèle statistique est une description mathématique approximative du mécanisme qui a généré les observations, que l'on suppose être un processus stochastique et non un processus déterministe. Il s'exprime*

4. L'exploration de données connue aussi sous l'expression de fouille de données, forage de données, prospection de données, data mining, ou encore extraction de connaissances à partir de données, a pour objet l'extraction d'un savoir ou d'une connaissance à partir de grandes quantités de données, par des méthodes automatiques ou semi-automatiques.

généralement à l'aide d'une famille de distributions et d'hypothèses sur les variables aléatoires $X_1, \ldots, X_n$. Chaque membre de la famille est une approximation possible, l'inférence consiste donc à déterminer le membre qui s'accorde le mieux avec les données.
Les types principaux : modèle linéaire, modèle linéaire généralisé, modèle multi-niveau , modèle d'équation structurelle , modèle mixte.

Définition 1.1.2. *(Modèle classique) On se place dans un espace probabilisé paramétrique classique avec, $\mathcal{X}$ désigne l'espace des données, Θ celui des paramètres β inconnu,*

$$X \in (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\beta, \beta \in \Theta\})$$

Exemple 1. *(Le modèle de regression lineaire)*
un modèle de régression linéaire est un modèle de régression qui cherche à établir une relation linéaire entre une variable, dite expliquée, et une ou plusieurs variables, dites explicatives, on rencontre principalement trois types de notations, simple (ou scalaire), vectoriel et matricielle, le modèle s'écrit alors :

$$Y = \beta X + \epsilon$$

La variable Y est appelée variable expliquée, variable dépendante, variable endogène ou encore réponse. Les variables X sont appelées variables explicatives, variable indépendante, variables exogènes ou encore prédicteurs. ϵ est appelé terme d'erreur ou perturbation.
On note généralement $\widetilde{\beta}$ le vecteur des paramètres estimés. On définit la valeur de prévision ajustée $\widehat{Y} = X\widetilde{\beta}$ et le résidu comme la différence entre la valeur observée et la valeur prédite $\widehat{\epsilon} = Y - \widehat{Y}$.

1.1.2 Loi jointe et loi a posteriori

Définition 1.1.3. *(La loi a priori) est une loi de probabilité π qui résume toute l'information disponible sur le paramètre d'intérêt β.*

Définition 1.1.4. *Dans le modèle classique, la loi **jointe** de (X, β) s'écrit :*

$$\lambda_\pi(X, \beta) = f(X, \beta)\pi(\beta)$$

La loi a ***posteriori*** *est définie par sa densité :*

$$\pi(\beta/X) \quad = \quad \underbrace{\frac{f(X,\beta)\pi(\beta)}{\int_{\Theta} f(X,\beta)\pi(\beta)d\beta}}_{*}$$

La quantité (*) est la loi marginale de X et est une constante de normalisation de la loi a ***posteriori****, indépendante de β. Nous ajoutons que par construction la loi a* ***posteriori*** *est absolument continue par rapport à la loi a priori π.*

Remarque 1.1.1. *Une loi impropre est une loi qui veifié :*

$$\int_{\Theta} \pi(\beta)d\nu(\beta) = +\infty$$

avec ν est une mesure σ-finie. La loi a priori peut être impropre, ce choix de type de loi n'a donc plus d'intérêt que calculatoire et s'interprète difficilement. Nous verrons par la suite que la construction de lois non informative peut conduire à des lois a priori de ce type.

Définition 1.1.5. *(Lois non informatives)*
Une loi non informative est une loi qui porte une information sur le paramètre à estimer dont le poids dans l'inférence est réduit. Certains auteurs la définissent également comme une loi a priori qui ne contient aucune information sur β ou encore comme une loi qui ne donne pas davantage de poids à telle ou telle valeur du paramètre.

Exemple 2. *Par exemple, supposons T un ensemble fini de taille q, une loi a priori non informative pourra être une loi de la forme :*

$$\mathbb{P}(\beta_i) = 1/q$$

On a l'équiprobabilité, les valeurs possibles de β se voit attribuer le même poids.

1.1.3 Principe de vraissemblance et d'exhaustivité

Deux principes fondamentaux sont respectés par le paradigme bayesien : le principe de vraissemblance et le principe d'exhaustivité.

Exhaustivité

Définition 1.1.6. *On appelle statistique exhaustive une statistique*[5] $T(X)$ *telle que la densité de x se décompose sous la forme :*

$$f(X,\beta) = h(X,T(X)) \cdot g(T(X),\beta)$$

cette égalité est connu sous le non "Théorème de factorisation".

Remarque 1.1.2. *Une statistique exhaustive* $T(x)$ *contient toute l'information apportée par x sur* β*. Selon le théorème de factorisation, sous certaines conditions de régularité, on peut écrit* $\pi(\beta/X) = \pi(\beta/T(X)$ *en termes informationnels.*

Principe d'exhaustivité[6] Deux observations x et y donnant la même valeur d'une statistique exhaustive T, c'est-à-dire telles que $T(x) = T(y)$, doivent conduire à la même inférence sur β.

Principe de vraissemblance

Ce deuxiéme principe est en partie une conséquence du principe d'exhaustivité. Il peut être attribué à Fisher (1959) ou même à Barnard (1949), mais il a été formalisé par Birnbaum (1962). Il est fortement défendu par Berger et Wolpert (1988) qui ont fourni une étude approfondie du sujet. Elle désigne de façon générale, l'ensemble des inférences possibles sur β.

Principe de vraissemblance :

L'information apportée par une observation de x sur β est entièrement contenue dans la fonction de vraissemblance $l(\beta, x)$ De plus, si $x_1 = x_1^1, \ldots, x_1^n$ et $x_2 = x_2^1, \ldots, x_2^n$ sont deux observations qui dépendent du même paramètre

5. C'est-à-dire une fonction des données

6. **L'importance :** Lorsqu'on utilise des lois impropres, le comportement d'une statistique peut être capricieux. Par exemple, si on considère un paramètre de la forme suivante $\beta(\beta_1, \beta_2)$ avec $\pi(\beta)$ est un loi impropre et pour T une statistique exhaustive pour β_1, nous pouvons écrire : $f(X/\beta_1, \beta_2) = g(X/\beta_2) \cdot f(T/X)$

β, et telles qu'il existe une constante c satisfaisant

$$l_1(\beta, x_1) = cl_2(\beta, x_2)$$

pour tout β, elles apportent la même information sur β et doivent conduire à la même inférence. Notons que le principe de vraissemblance n'est valide que lorsque

i) l'inférence concerne le même paramètre β.

ii) β prend en compte tous les facteurs inconnus du modèle.

Et cela accueillerait favorablement l'occasion de parler de l'estimateur de maximum de vraissemblance.

Définition 1.1.7. *(L'estimateur de maximum de vraissemblance (MV))*
Un estimateur $\widehat{\beta}_{MV}$ est appelé un estimateur du maximum de vraissemblance de β, si

$$\widehat{\beta}_{MV} = \arg\max_{\beta \in \Theta} L(\beta, X)$$

Noter qu'afin de resoudre l'equation que le gradien soit nulle, il faut determiner la dérivée d'un produit de n fonction, ce qui peut être un calcul compliqué, pour eviter ce probléme nous nous consentrons habituellement à maximiser la log-vraissemblance

$$L(\beta, X) = \log L(\beta, X)$$

au lieu de la vraissemblance puisque la fonction (log) est monotone, la vraisembalnce et la log-vraisembalnce ont les maximums et les minimums pour les même β, donc c'est un avantage.

Dérivation du principe de vraissemblance

Une justification du principe de vraissemblance à été avancée par Birnbaum (1962) qui a établi que le principe de vraissemblance est une conséquence du principe d'exhaustivité, à condition d'accepter un second principe.

Principe de conditionnement :

Si deux expériences sur le paramètre β, noté ξ_1 et ξ_2 sont possibles et si on choisit une de ces expériences avec probabilité $\mathbb{P}$, l'inférence sur β ne doit dependre que de l'experience choisie.

Il semble difficile de refuser ce principe quand l'expérience choisie est connue, comme on peut constater dans l'exemple classique de (Cox (1958)).

1.2 Une introduction à la théorie de la décision

1.2.1 La fonction de perte et risque

Définition 1.2.1. *La fonction de perte est une fonction mesurable de* $(\Theta \times \mathcal{D})$ *à valeurs réelles positives :*

$$L : \Theta \times \mathcal{D} \to \mathbb{R}^+$$

avec $\mathcal{D}$ *l'espace de décisions .Elle est définit selon le probléme statistique étudié.*

Définition 1.2.2. *La fonction de risque d'un estimateur* $\widetilde{\beta}$ *de* β*, est definit par :*

$$R(\beta, \widetilde{\beta}) = \mathbb{E}_\beta[L(\beta, \widetilde{\beta})] = \int \dots \int L(\beta, \widetilde{\beta}(x_1 \dots x_n)) d\mathbb{P}_\beta(x_1) \dots d\mathbb{P}_\beta(x_n)$$

Il existe plusieurs type de risque comme le risque fréquentiste, le risque à posteriorie et le risque intégré qu'on va prend comme un exemple.

Exemple 3. *(Le risque intégré)*
le risque intégré à une fontion de perte donnée est defini par :

$$r(\pi, \widetilde{\beta}) = \int_\Theta R(\beta, \widetilde{\beta}) d\pi(\beta)$$

Proposition 1.2.1. *Un estimateur bayésienne est un estimateur vérifié :*

$$r(\pi, \beta^\pi) = \inf_{\widetilde{\beta} \in \mathcal{D}} r(\pi, \widetilde{\beta})$$

1.2.2 Les fonctions de perte usuelles

La fonction de perte quadratique

Définition 1.2.3. *La fonction de perte quadratique est la fonction definit par :*

$$L(\beta, \widetilde{\beta}(x)) = (\beta - \widetilde{\beta}(x))^2$$

une variante de cette fonction est une fonction de perte quadratique pondérée de la forme, $L(\beta, \widetilde{\beta}(x)) = q(\beta)(\beta - \widetilde{\beta}(x))^2$*, avec* $q(\beta)$ *est une fonction de poids positive.*

Remarque 1.2.1. *La fonction de risque associé à la fonction de perte quadratique, c'est rien que l'erreur moyenne quadratique.*

Proposition 1.2.2. *Sous l'hypothèse de perte quadratique, l'estimateur de bayes* $\beta^{\pi}(x)$ *de* β *associé a la loi a priori* π *est la moyenne a posteriorie de* β*.*

$$\beta^{\pi}(x) = \mathbb{E}^{\pi(\cdot/x)}(\beta) = \int_{\beta\in\Theta} L(\beta, \widetilde{\beta}(x))\pi(\beta/x)d\beta$$

Démonstration :

Par définition, l'estimateur de bayes minimise le risque a posteriorie i.e $\rho(\pi, \widetilde{\beta}) = \mathbb{E}^{\pi(\cdot/x)}[L(\beta, \widetilde{\beta}(x))]$. Sous l'hypothèse de risque quadratique on a :

$$\begin{aligned}
\rho(\pi, \widetilde{\beta}) &= \mathbb{E}^{\pi(\cdot/x)}[(\beta - \widetilde{\beta}(x))^2] \\
&= \mathbb{E}^{\pi(\cdot/x)}(\beta^2) - 2\widetilde{\beta}(x)\mathbb{E}^{\pi(\cdot/x)}(\beta) + \widetilde{\beta}^2(x)
\end{aligned}$$

On remarque que la deuxiéme formule est un polynome de second degré en $\widetilde{\beta}(x)$, il sera minimum en $\mathbb{E}^{\pi(\cdot/x)}(\beta)$.

La fonction de perte absolue

Définition 1.2.4. *La fonction de perte absolue (perte* L^1*), est la fonction definit par :*

$$L(\beta, \widetilde{\beta}(x)) = |\beta - \widetilde{\beta}(x)|$$

Proposition 1.2.3. *sous l'hypothèse de perte L^1, l'estimateur de bayes $\beta^\pi(x)$ de β associé a la loi a priori π est la médiane de $\pi(\beta/X)$*

Démonstration :

On a par définition :

$$\begin{aligned}
\rho(\pi, \widetilde{\beta}) &= \int_{\beta\in\Theta} |\beta, \widetilde{\beta}(x)| \pi(\beta/x) d\beta \\
&= \int_{-\infty}^{\widetilde{\beta}} (\widetilde{\beta} - \beta(x)) \pi(\beta/x) d\beta + \int_{\widetilde{\beta}}^{+\infty} (\beta - \widetilde{\beta}(x)) \pi(\beta/x) d\beta
\end{aligned}$$

est comme $\rho \mapsto \rho(\pi, (\widetilde{\beta}/X))$ est convexe et dérivable $\mu-$presque partout, il suffit la encore de détermine les points critiques :

$$\frac{\partial \rho(\pi, \widetilde{\beta}(x))}{\partial \widetilde{\beta}} = \int_{-\infty}^{\widetilde{\beta}} \pi(\beta/x) d\beta + \int_{\widetilde{\beta}}^{+\infty} \pi(\beta/x) d\beta$$

Donc $$\frac{\partial \rho(\pi, \widetilde{\beta}(x))}{\partial \widetilde{\beta}} = 0 \iff \mathbb{P}^\pi(\beta \le \widetilde{\beta}/X) = \mathbb{P}^\pi(\beta \ge \widetilde{\beta}/X).$$

Corollaire 1.2.1. *L'estimateur de Bayes β^π assosié à π et au perte quadratique pondéré est*

$$\beta^\pi(x) = \frac{\mathbb{E}^\pi(\omega(\beta))\beta/x}{\mathbb{E}^\pi(\omega(\beta))/x}$$

La fonction de perte 0-1

Cette fonction de perte est utilisé dans le contexte des testes, un test est la donnée d'une partition de Θ en Θ_0 et Θ_1 avec $\beta \in H_i$ correspond à l'hypothèse H_i, H_0 est appelée l'hypothèse nulle, le principe du test (décisions) $\widetilde{\beta}$, est definit comme suit :

$$\begin{cases} \widetilde{\beta} = 0 & si \quad \beta \in \Theta_1 \\ \widetilde{\beta} = 1 & si \quad \beta \in \Theta_0 \end{cases}$$

Définition 1.2.5. *la fonction de perte 0-1 est définit par :*

$$L(\pi, \widetilde{\beta}) = \mathbb{I}_{\beta \in \Theta_1} \times \mathbb{I}_{\widetilde{\beta}=0} + \mathbb{I}_{\beta \in \Theta_0} \times \mathbb{I}_{\widetilde{\beta}=1}$$

Remarque 1.2.2. *Le risque à postiriori est alors le suivant :*

$$\rho(\pi, \widetilde{\beta}) = \mathbb{I}_{\widetilde{\beta}=0} \mathbb{P}^{\pi}(\Theta_1/X) + \mathbb{I}_{\widetilde{\beta}=1} \mathbb{P}^{\pi}(\Theta_0/X)$$

Ainsi que l'estimateur de bayes associé a cette fonction de perte, est definit par :

$$\beta^{\pi}(X) = 1 \Longleftrightarrow \mathbb{P}^{\pi}(\Theta_0/X) \leq \mathbb{P}^{\pi}(\Theta_1/X)$$

C'est à dire que l'estimation permet d'accepter H_0, si c'est l'hypothèse la plus probable à posteriori, ce qui est une reponse naturelle.

1.2.3 Fonction de perte intrinsèque

Nous nous placons dans le cas dans le cas ou $X/\beta \sim f(X, \beta)$, avec f est une fonction de densité, la fonction de perte est definit comme suit :

$$L(\beta, \widetilde{\beta}) = d(f_{\beta}, f_{\widetilde{\beta}})$$

La notion de perte intrinsèque provient de l'exigence d'invariance par transformation monotone inversible sur les données, les distances et donc les fonctions de perte, utilisées doivent donc être inchangées par l'action d'un C^1-defféomorphisme sur $\mathcal{X}$, ainsi pour $y = g(x)$ et $X \sim f_{\beta}$ où g est un C^1-defféomorphisme, on note $y \sim g_{\beta}(y) = f_{\beta}(g^{-1}(y))|\frac{dx}{dy}|$ pour une perte intrinsèque, la distance entre f_{β} et $f_{\beta'}$ est la même que celle entre g_{β} et $g_{\beta'}$, pour un g donné, ceci correspond à une distance entre distributions (et non plus entre densités). L'invariance par C^1-defféomorphisme est vérifiée, pour les distances précédentes à l'exeption de la norme L^2, nous donnons la démonstration pour la norme L^1 :

$$
\begin{aligned}
d_1(g_\beta, g_{\beta'}) &= \int |f_\beta(g^{-1}(y)) - f_{\beta'}(g^{-1}(y))||\frac{dx}{dy}|dy \\
&= \int |f_\beta(x) - \int |f_{\beta'}(x)|dx \\
&= d_1(f_\beta, f_{\beta'})
\end{aligned}
$$

Définition 1.2.6. *La fonction de perte intrinsèque est une fonction definie à partir d'une distance entre distributions, c'est à dire invariante par transformation monotone inversible, definit par :*

$$L(\beta, \widetilde{\beta}) = d(f_\beta, f_{\widetilde{\beta}})$$

Exemple 4. *(Distance d'Hellinger de lois normale)*
En notant Φ_{μ,σ^2} la de la loi $\mathcal{N}(\mu, \sigma^2)$ et $\beta = (\mu, \sigma^2)$ nous avons :

$$
\begin{aligned}
d_H^2(\Phi_\beta, \Phi_{\widetilde{\beta}}) &= \int_{\mathbb{R}} \left(\sqrt{\Phi_\beta} - \sqrt{\Phi_{\widetilde{\beta}}}\right)^2 dx \\
&= 2 - 2\int_{\mathbb{R}} \sqrt{\Phi_\beta - \Phi_{\widetilde{\beta}}} dx \\
&= 2 - \frac{2}{\sqrt{2\pi\sigma\widetilde{\sigma}}} \underbrace{\int_{\mathbb{R}} exp\left(-\frac{(x-\mu)^2}{4\sigma^2} - \frac{(x-\widetilde{\mu})^2}{4\widetilde{\sigma}^2}\right) dx}_{=I}
\end{aligned}
$$

$$I = \int_{\mathbb{R}} exp\left(-\frac{1}{4\Sigma^2}\left[x - (\frac{\mu}{\sigma^2} + \frac{\mu}{\widetilde{\sigma}^2})\Sigma^2\right]^2 - \frac{\Sigma^2}{4}(\frac{\mu}{\sigma^2} + \frac{\mu}{\widetilde{\sigma}^2}) + \frac{\mu}{4\sigma^2} + \frac{\widetilde{\mu}}{4\widetilde{\sigma}^2}\right) dx$$

en notant $\Sigma^2 = \left(\frac{1}{\sigma^2} + \frac{1}{\widetilde{\sigma}^2}\right)^{-1}$.D'ou en reconnaissant un intégrale gaussienne de variance $2\Sigma^2$ et en factorisant le facteur exponentiel :

$$I = \Sigma\sqrt{2\pi} exp - \left(\frac{(\mu - \widetilde{\mu})^2}{4(\sigma^2 - \widetilde{\sigma}^2)}\right)$$

Ainsi nous arrivons à

$$
\begin{aligned}
\frac{1}{2} d_H^2(\Phi_\beta, \Phi_{\widetilde{\beta}}) &= 1 - \sqrt{\frac{2\sigma\widetilde{\sigma}}{\sigma^2\widetilde{\sigma}^2}} exp - \left(\frac{(\mu - \widetilde{\mu})^2}{4(\sigma^2 - \widetilde{\sigma}^2)}\right) \\
&= 1 - \sqrt{\frac{2(u+1)}{(u+1)^2 + 1}} \left(1 - \frac{h^2}{4(\sigma^2 - \widetilde{\sigma}^2)}\right) + O(h^2)
\end{aligned}
$$

en notant $h = \mu - \widetilde{\mu}$ et $u = \dfrac{\sigma}{\widetilde{\sigma}} - 1$ (développement limité en (h, u)) c'est la somme de deux termes du second ordre : et

$$\sqrt{\frac{2(u+1}{(u+1)^2+1}}.\frac{h^2}{\widetilde{\sigma}^2(1+(u+1)^2)} \sim \frac{h^2}{8\widetilde{\sigma}^2}$$

d'ou

$$d_H^2(\Phi_\beta, \Phi_{\widetilde{\beta}}) = \frac{2(\sigma-\widetilde{\sigma})^2 + (\mu-\widetilde{\mu})^2}{4\widetilde{\sigma}^2} + O(\|h, u\|^2)$$

.

1.2.4 Estimateur bayesien (EB)

l'idée centrale de l'analyse bayésienne est de considére le paramètre inconnu β comme aléatoire : Θ (l'espase des paramètres β) est muni d'une probabilité π telque $(\Theta, \mathcal{A}, \pi)$ est un espace probabilisé. Nous noterons $\beta \sim \pi$. π est appelée loi a priori. Intuitivement et en termes informationnels, elle détermine ce qu'on sait et ce qu'on sait pas avant d'observer X.

Définition 1.2.7. *Soit $\widehat{\beta} = \widehat{\beta}(x)$ un estimateur de β, On appelle estimateur de Bayes associé à la fonction de perte $L(\beta, \widehat{\beta})$ et à une distribution a priori π, toute décision β^π qui minimise le risque de bayes définit par $\mathbb{E}_\pi[L(\beta, \widehat{\beta})]$, ce estimateur est definit par :*

$$\beta^\pi = \arg\min_{\beta \in \mathcal{D}} \mathbb{E}_\pi[L(\beta, \widehat{\beta})]$$

avec $\mathcal{D}$ l'espace de décisions possibles.

Définition 1.2.8. *(Estimateur biais)*
Un estimateur $\widehat{\beta}$ de β est dit sans biais si est seullement si

$$\mathbb{E}(\widehat{\beta}) = \beta$$

Remarque 1.2.3. *une propriété moins forte est le biais asymptotique.*

$$\lim_{n \mapsto +\infty} \mathbb{E}(\widehat{\beta}) = \beta$$

Définition 1.2.9. *(Convergence)*
un estimateur $\widehat{\beta}$ de β est dit converge si est seullement si

$$\lim_{n\mapsto+\infty} Var(\widehat{\beta}) = O$$

1.2.5 Deux optimalités minimaxité et admissibilité

Cette section est consacrée à deux notions fondamentales de la Théorie de la Décision fréquentiste, présentées par Wald (1950) et Neyman et Pearson (1933). Comme il a été mentionné auparavant, et contrairement à l'approche bayésienne, le paradigme fréquentiste n'est pas assez réducteur pour conduire à un seul estimateur optimal.

Admissibilité

Définition 1.2.10. *Soit $X \in (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\beta, \beta \in \Theta\})$ un modèle paramétrique, et L une fonction de perte sur $\Theta \times \mathcal{D}$ où $\mathcal{D}$ est l'ensemble de décision, on dit que $\widetilde{\beta} \in \Theta$ est **inadmissible**[7] si et seulement si :*

$$\begin{array}{lll} & \exists \widetilde{\beta}_0 \in \mathcal{D}, \quad \forall \beta \in \Theta, & R(\beta, \widetilde{\beta}) \geq R(\beta, \widetilde{\beta}_0) \\ \textit{et} & \exists \beta_0 \in \Theta & R(\beta_0, \widetilde{\beta}) > R(\beta_0, \widetilde{\beta}_0) \end{array}$$

*Un estimateur est dit **admissible** si et seulement s'il n'est pas **inadmissible**.*

Remarque 1.2.4. *Si*

$$\forall (\beta, \widetilde{\beta}), R(\beta, \widetilde{\beta}) \leq R(\beta, \widetilde{\beta}_0) \Longrightarrow \pi\left(\beta \in \Theta, R(\beta, \widetilde{\beta}) < R(\beta, \widetilde{\beta}_0)\right) = 0$$

On dit que $\widetilde{\beta}_0$ est ($\pi-$admissible), présque admissible par rapport à la mesure π.

Définition 1.2.11. *Une classe $\mathcal{D}$ d'estimateurs est dite complète si, quel que soit $\widetilde{\beta}'$ n'appartient pas a $\mathcal{D}$, il existe $\widetilde{\beta} \in \mathcal{D}$ qui domine $\widetilde{\beta}'$. La classe est essentiellement complète si, quel que soit $\widetilde{\beta}'$ n'appartient pas a $\mathcal{D}$, il existe $\widetilde{\beta} \in \mathcal{D}$ au moins aussi bon que $\widetilde{\beta}'$.*

7. Un estimateur $\widetilde{\beta}$ est inadmissible s'il existe un estimateur $\widetilde{\beta}0$ qui domine $\widetilde{\beta}$

Théorème 1.2.1. *(Estimateurs bayesiens admissibles) Si l'estimateur bayesien*[8] β^{π} *associé à une fonction de perte L et une loi à priori π est unique, alors il est* ***admissible****.*

Démonstration :

supposons β^{π} estimateur bayesien non admissible

$$\exists \widetilde{\beta}_0 \in \mathcal{D}, \forall \beta \in \Theta, \quad R(\beta, \widetilde{\beta}) \geq R(\beta, \widetilde{\beta}_0)$$

$$et \qquad \exists \beta_0 \in \Theta \quad R(\beta_0, \widetilde{\beta}) > R(\beta_0, \widetilde{\beta}_0)$$

En intégrant la premiére inégalité :

$$\int_{\Theta} R(\beta, \widetilde{\beta}_0) d\pi(\beta) \leq \int_{\Theta} R(\beta, \beta^{\pi}) d\pi(\beta) = r(\pi)$$

On remarque que $\widetilde{\beta}_0$ est aussi un estimateur bayesien associé à L et π avec $(\widetilde{\beta}_0 \neq \beta^{\pi})$, le théorème se déduit par contraposée.

Remarque 1.2.5. *Ce théorème s'applique notament dans le cas d'un risque fini et d'une fonction de perte convexe, en outre l'unicité de l'estimateur bayesien implque la finitude du risque : $r(\pi) = \int_{\Theta} R(\beta, \beta^{\pi}) d\pi(\beta) < \infty$ (sinon, tout estimateur minimise le risque).*

Proposition 1.2.4. *Tout estimateur bayesien telque $r(\pi) < \infty$ est $\pi -$ admissible.*

Démonstration :

Soit la derniére condition vérifiée pour $\widetilde{\beta}_0$ telque $\forall \beta$, om note :

$$A = \beta \in \Theta, R(\beta, \widetilde{\beta}) < R(\beta, \widetilde{\beta}_0)$$

8. Notons que que pour des fonctions de pertes strictement convexes, les esimateurs de bayes sont uniques.

nous avons alors

$$\int_{\Theta} R(\beta,\widetilde{\beta}_0)d\pi(\beta) - \int_{\Theta} R(\beta,\beta^{\pi})d\pi(\beta) = \int_{A} \Big(R(\beta,\widetilde{\beta}) - R(\beta,\widetilde{\beta}_0)\Big)\, d\pi(\beta)$$

si et seulement si $\pi(A) = 0$, or comme β^{π} est bayesien et le risque finie, $r(\beta,\widetilde{\beta}_0) \geq r(\beta,\beta^{\pi})$, donc l'intégrale est nulle (positive et négative), d'ou $\pi(A) = 0$.

Maintenant on va énoncer une condition suffisante d'admissbilité des estimateurs bayésiens.

Théorème 1.2.2. *(continuté et $\pi -$ admissible)si $\pi > 0$ sur Θ et $r(\pi) < \infty$ pour une fonction de perte L donnée, donc on a si β^{π} l'estimateur bayesien correspondant existe, avec la fonction $R : \beta \longrightarrow R(\beta,\widetilde{\beta})$ soit continue, alors β^{π} est* ***admissible****.*

Démonstration :

supposons que β^{π} est **inadmissible**, d'aprés la proposition (1.2.4) β^{π} est π-**admissible**. Ainsi $\exists\widetilde{\beta}_0$ telque pour tout β
$R(\beta,\widetilde{\beta}_0) \leq R(\beta,\beta^{\pi})$, et $\beta_0 \in \Theta$, $\quad R(\beta_0,\widetilde{\beta}_0) < R(\beta_0,\beta^{\pi})$, et on à $R : \beta \longrightarrow R(\beta,\widetilde{\beta}_0) - R(\beta,\beta^{\pi})$ est une foncton continue definit sur Θ. Donc il existe un voisinage (ouvert) de β_0, $\mathcal{V}_0 \subset \Theta$ telque $\forall\beta \in \mathcal{V}_0$, $\quad R(\beta,\widetilde{\beta}_0) \leq R(\beta,\beta^{\pi})$. En considére A la même région que dans la proposition précedent, $\pi(A) \geq \pi(\mathcal{V}_0)$, or π est supposée strictement positive que Θ, donc en prenant un modèle **dominé** par une mesure qui charge positivement les ouverts (la mesure de lebesque par exemple) $\pi(\mathcal{V}_0) > 0$, A est donc non néglegeable (de mesure non nulle), ce qui n'est pas conforme avec la π-**admissiblité**, on conclusion β^{π} est **admissible**.

Estimateur randomisé

L'idée a l'origine de cette notion est de rendre $\mathcal{D}$ convexe, pour pouvoir maximiser facilement l'estimateur de Neyman-Pearson.

Définition 1.2.12. *Soit le modèle $X \in (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\beta, \beta \in \Theta\})$ un modèle paramétrique, où $\mathcal{D}$ est l'ensemble des décisions, on definit $\mathcal{D}^*$ comme l'ensemble des probabilités sur $\mathcal{D}$, $\beta^* \in \mathcal{D}^*$ est appelé estimateur* ***randomisé****.*
On definit de même les fonctions de perte et risque ***randomisés*** *par :*

$$\begin{aligned} L^*(\beta, \beta^*) &= \int_{\mathcal{D}} L(\beta, a) d\beta^*(a) \\ R^*(\beta, \beta^*) &= \mathbb{E}_\beta \left[L^*(\beta, \beta^*) \right] \end{aligned}$$

Minimaxité

Risque et estimateur minimax

L'estimateur **minimax** correspond au point de vue (conservateur) de Judith Rousseau est faire le mieux dans le pire des cas, c'est à dire s'assurer contre le pire, il est utile dans des cadres complexes, mais trop conservateur dans certains cas où le pire trés peut probable, il peut étre judicieux de voir l'estimation comme un jeu entre le statistien (choix de β), et la nature (choix de β), l'estimation **minimax** rejoint alors celle de la théorie des jeux.

Définition 1.2.13. *Le risque* ***minimax*** *est defini par :*

$$\overline{R} = \inf_{\beta^* \in \mathcal{D}^*} \sup_{\beta \in \Theta} R(\beta, \beta^*)$$

On dit que $\widetilde{\beta}_0$ est un estimateur minimax si et seulement si :

$$\overline{R} = \sup_{\beta \in \Theta} R(\beta, \widetilde{\beta}_0)$$

.

Théorème 1.2.3. *Si $\mathcal{D} \in \mathbb{R}^k$ est convexe et compact, et si $L(\beta, \widetilde{\beta})$ est continue et convexe en tant que fonction de $\widetilde{\beta}$, pour chaque $\beta \in \Theta$ il existe un estimateur minimax non randomisé*[9].

Le résultat suivant met en avant la connexion entre approche bayésienne et principe minimax, dont la démonstration est immédiate.

9. Ce résultat est un cas particulier du théorème de Rao-Blackwell voir [2].

Lemme 1.2.1. *Le risque de Bayes est toujours plus petit que le risque minimax*[10]

$$\underline{R} = \sup_{\pi} r(\pi) = \sup_{\pi} \inf_{\widetilde{\beta} \in \mathcal{D}} r(\pi, \widetilde{\beta}) \leq \overline{R} = \inf_{\widetilde{\beta} \in \mathcal{D}^*} \sup_{\beta} R(\beta, \widetilde{\beta})$$

Un cas particuliérement intéressant correspond à la définition suivante.

Définition 1.2.14. *Un probléme d'estimation est dit admettre*[11] *une valeur si, $\overline{R} = \underline{R}$, c'est-à-dire quand*

$$\sup_{\pi} \inf_{\widetilde{\beta} \in \mathcal{D}} r(\pi, \widetilde{\beta}) \leq \overline{R} = \inf_{\widetilde{\beta} \in \mathcal{D}^*} \sup_{\beta} R(\beta, \widetilde{\beta})$$

Lemme 1.2.2. *Si $\widetilde{\beta}_0$ est un estimateur de Bayes pour π_0 et si $R(\beta, \widetilde{\beta}_0) \leq r(\pi_0)$ pour tout β dans le support de π_0, $\widetilde{\beta}_0$ est minimax et π_0 est la distribution la moins favorable.*

Puisque les estimateurs minimaxs correspondent généralement à des estimateurs de Bayes généralisés, on doit souvent recourir à un argument limite pour établir la minimaxité, plutôt que de calculer directement le risque de Bayes comme dans le Lemme (1.2.2).

Lemme 1.2.3. *S'il existe une suite (π_n) de loi a priori propres, donc si l'estimateur de Bayes généralisé $\widetilde{\beta}_0$ verifié*

$$R(\beta, \widetilde{\beta}_0) \leq \lim_{n \mapsto +\infty} r(\pi_n) < +\infty$$

pour tout $\beta \in \Theta$, alors $\widetilde{\beta}_0$ est minimax.

Exemple 5. *Quand $x \sim> \mathcal{N}(\beta, 1)$ l'estimateur de (MV) $\widehat{\beta}_{MV}(x) = x$ est un estimateur de bayes géneralisé par rapport à la mesure de lebesgue sur $\mathbb{R}$, pour le coût quadratique puisque*

$$R(\widehat{\beta}_{MV}, \beta) = \mathbb{E}_{\beta}(x - \beta)^2 = 1$$

10. La premiére valeur est dite risque maximin et une distribution π^*. telle que $r(\pi^*) = \underline{R}$ est appelée distribution a priori la moins favorable.

11. Quand le probléme admet une valeur, certains estimateurs minimax sont des estimateurs de Bayes correspondant aux lois a priori les moins favorables. Cependant, ils peuvent être randomisés.

ce risque est la limite du risque de bayes $r(\pi_n)$ quand π_n est égale à $\mathcal{N}(0,n)$, comme

$$r(\pi_n) = \frac{n}{n+1}$$

Par conséquent, l'estimateur de (MV) $\widehat{\beta}_{MV}$ est minimax.
Notons que cet argument peut être étendu directement au cas $x \sim> \mathcal{N}_p(\beta, I_p)$ pour établir que $\widehat{\beta}_{MV}$ est minimax pour tout p.

Proposition 1.2.5. *S'il existe un unique estimateur minimax, cet estimateur est admissible.*

Démonstration :

Si β^* est le seul estimateur minimax, pour tout estimateur $\widetilde{\beta} \neq \beta^*$:

$$\sup_\beta R(\beta, \widetilde{\beta}) > \sup_\beta R(\beta, \beta^*)$$

Donc $\widetilde{\beta}$ ne peut pas dominer β^*.

Remarque 1.2.6. *Notons que la réciproque de ce résultat est fausse, car il peut exister plusieurs estimateurs minimax admissibles. Par exemple, dans le cas $\mathcal{N}_p(\beta, I_p)$, il existe des estimateurs de Bayes réguliers minimax pour $p \geq 5$.*

Proposition 1.2.6. *Si $\widetilde{\beta}_0$ admissible de risque constant, alors $\widetilde{\beta}_0$ est l'unique estimateur minimax.*

Démonstration :

Pour tout $\beta_0 \in \Theta$, $\sup\limits_\beta R(\beta, \widetilde{\beta}_0) = R(\beta_0, \widetilde{\beta}_0)$. Aors s'il existe $\widetilde{\beta}_1$ telque

$$\overline{R} \leq \sup_\beta R(\beta, \widetilde{\beta}_1) < R(\beta_0, \widetilde{\beta}_0)$$

$\widetilde{\beta}_0$ ne peut pas être admissible. De la même façon, si

$$R = \sup_\beta R(\beta, \widetilde{\beta}_1) = R(\beta_0, \widetilde{\beta}_0)$$

et si $\exists \widetilde{\beta}_1$ telque $R(\beta_1, \widetilde{\beta}_1) < \bar{R}$, $\widetilde{\beta}_1$ domine $\widetilde{\beta}_0$. par conséquent quand $\widetilde{\beta}_0$ est admissible, le seul cas possible est qu'il existe $\widetilde{\beta}_1$ tel que $R(\beta, \widetilde{\beta}_1) = R(\beta, \widetilde{\beta}_0)$ pour tout $\beta \in \Theta$. Ce qui set aussi impossible quand $\widetilde{\beta}_0$ est admissible.

Remarque 1.2.7. *Remarquons à nouveau que la réciproque de ce résultat est fausse. Il peut exister des estimateurs minimax ayant un risque constant qui soient inadmissibles. En fait, ils sont inadmissibles dès qu'il existe d'autres estimateurs minimax. C'est le cas par exemple pour* $\widetilde{\beta}_0(x) = x$ *quand* $x \sim \mathcal{N}_p(\beta, I_p)$ *et* $p \geq 3$*, Il y a aussi des cas òu il néxiste pas d'estimateur minimax admissible.*

1.2.6 Le cas particulier du modèle normal

Lorsque Gauss introduisit la distribution normale aux alentours de 1810, Laplace estima qu'il s'agissait en fait de la loi d'erreur idéale. Par la suite, s'appuyant sur le Théorème Central Limit, les statisticiens de la premiére moitié du **XIX**iéme siécle se référaient presque toujours à la distribution normale. Il y'a bien sûr, des nombreux phénomènes pour lesquels un modèle normal n'est pas applicable, mais ce dernier reste considérablement utilisé, en particulier en économétrie et dans des domaines où on peut justifier l'approximation du Théorème Central Limit (physique particulaire, etc.). En réalité, l'approximation normale est souvent justifiée par des raisons asymptotiques. Il est donc intéressant d'étudier en détail cette distribution particulière d'un point de vue bayesien.

Pour l'observation d'une distribution normale multivariée $\mathcal{N}_p(\beta, \Sigma)$, de matrice de covariance connue Σ, et la loi à priori [12] $\pi(\beta) \sim \mathcal{N}_p(\mu, A)$ avec A est un hyperparaméte, et la loi a posteriori $\pi(\beta/x)$ est

$$\mathcal{N}_p(\underbrace{x - \Sigma(\Sigma + A)^{-1}(x - \mu)}, \overbrace{(A^{-1} + \Sigma^{-1})^{-1}})$$

12. on peut dit aussi la loi conjuguée

Sous un coût quadratique, l'estimateur de Bayes s'écrire comme une combinaison convexe de l'bservation x et la moyenne a priori μ

$$\begin{aligned} \beta^{\pi}(x) &= x - \Sigma(\Sigma + A)^{-1}(x - \mu) \\ &= (\Sigma^{-1} + A^{-1})^{-1}(\Sigma^{-1}x + A^{-1}\mu) \end{aligned}$$

Estimation de la variance

Dans la plupart des cas, la variance du modèle est partiellement ou totalement inconnue. Il est alors nécessaire de considérer des lois a priori pour le paramètre (β, Σ). Si la variance est connue à une constante multiplicative prés, σ^2, il est généralement possible de revenir à un cadre unidimensionnel, c'est-à-dire lorsque $x_1, \ldots, x_n$ sont i.i.d $\mathcal{N}(\beta, \sigma^2)$ pour des raisons d'exhaustivité. Si nous définissons les statistiques $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ et $s^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2$ la vraissemblance [13] peut s'écrire

$$l(\beta, \sigma, \overline{x}, s^2) = \sigma^{-n} exp\left[-\frac{1}{2\sigma^2}\{s^2 + n(\overline{x} - \beta)^2\}\right]$$

1.2.7 Espace des paramètres restreint

Berger (1985) remarque l'intérêt d'une approche bayésienne non informative pour des espaces des paramètres restreints, la loi a priori étant simplement la troncation d'une loi non informative sans contrainte. D'un point de vue classique, le calcul d'estimateurs du maximum de vraissemblance restreints est souvent compliqué, notamment quand les contraintes sont non linéaires. En revanche la mise en oeuvre d'une approche bayésienne via des méthodes de simulation de Monte Carlo permet un calcul aisé des estimateurs de Bayes.

13. l'estimateur de Bayes ne dépend que de $\overline{x}$ et s^2.

(Cet avantage peut même être utilisé pour calculer des estimateurs du maximum de vraissemblance restreints à travers des techniques bayésiennes.

1.2.8 Estimation du coût

pour un coût donné $L(\beta, \widetilde{\beta})$, d'un point de vue décisionnel comme l'estimation du coût $L(\beta, \beta^\pi(x))$ par $\gamma(x)$, sous une seconde fonction de coût, comme

$$\overline{L}(\beta, \beta^\pi, \gamma) = L\underbrace{[\gamma(x) - L(\beta, \beta^\pi(x))]}_{Q}$$

De nouveau, le coût quadratique (Q) n'est pas plus justifié comme choix automatique dans ce contexte que dans d'autres cas d'estimation. Mais en dehors de son côté pratique, le choix du coût quadratique peut se défendre par l'absence de justification en termes d'utilité et par conséquent, une perception plus proche de l'erreur comme une variance. Sous (Q), l'évaluation bayésienne des performances de β^π est donnée par le résultat suivant.

Proposition 1.2.7. *L'estimateur de Bayes du coût $L(\beta, \beta^\pi(x))$ sous (Q) pour la loi a priori π est*

$$\gamma^\pi(x) = \mathbb{E}^\pi(L(\beta, \beta^\pi(x))/x)$$

Remarque 1.2.8. *Le but est d'estimer une fonction particulière de β sous un coût quadratique. Notons que la dépendance de cette fonction à x n'a pas d'importance d'un point de vue bayesien car, une fois x observé, x est fixé. De même, pour un coût d'erreur absolue, l'estimateur de Bayes du coût est la médiane de la distribution a posteriori de $L(\beta, \beta^\pi(x))$, moins facile à obtenir. quand L est le coût quadratique, la variance a posteriori, $var^\pi(x)$, est par conséquent l'estimateur de Bayes du coût associé avec β^π.*

f(x,β)	$\pi(\beta)$	$\pi(\beta/x)$
Normal $\mathcal{N}(\beta,\sigma^2)$	Normal $\mathcal{N}(\widehat{\beta},\tau^2)$	$\mathcal{N}(\varphi(\sigma^2\widehat{\beta}+\tau^2 x),\varphi\sigma^2\tau^2)$ $\varphi^{-1}=\sigma^2+\tau^2$
Poisson $p(\beta)$	Gamma $\Gamma(\alpha,\widehat{\beta})$	$\Gamma(\alpha+x,\widehat{\beta}+1)$
Gamma $\Gamma(\nu,\beta)$	Gamma $\Gamma(\alpha,\widehat{\beta})$	$\Gamma(\alpha+\nu,\widehat{\beta}+x)$
Binomial $B(n,\beta)$	Bêta $Be(\alpha,\widehat{\beta})$	$Be(\alpha+x,\widehat{\beta}+n-x)$
Binomial negative $Neg(m,\beta)$	Bêta $Be(\alpha,\widehat{\beta})$	$Be(\alpha+m,\widehat{\beta}+x)$
Multinomial $M_k(\beta_1,\ldots,\beta_k)$	Dirichlet $D(\widehat{\beta}_1,\ldots,\widehat{\beta}_k)$	$D(\widehat{\beta}_1+x_1,\ldots,\widehat{\beta}_k+x_k)$
Normal $\mathcal{N}(\mu,\frac{1}{\beta})$	Gamma $\Gamma(\alpha,\widehat{\beta})$	$\Gamma(\alpha+0.5,\widehat{\beta}+\frac{(\mu-x)^2}{2})$

Tableau de quelques lois a priori et a posteriori

Chapitre 2

Les estimateurs de type Stein

Soit $X_1 \dots X_p$ une suite de variable aléatoire réele indépendant suit la distribution normal (Gauss) avec moyenne $\mu_1 \dots \mu_p$ inconnue et variance 1, c'est habituel de estimer μ_i avec $X_i \quad (i = 1 \dots p)$, ce estimateur est appelé l'esimateur usuel (de bayes), si la perte utiliser et la perte quadratique, alors Stein en (1956) montrer que X est un estimateur admissible pour p=1,2 et puis en (1961) James et Stein ont prouvé que X est innadmissible pour $p \geq 3$.
On consider le probleme de l'estimation de μ avec la fonction de perte L definit par,

$$l(\mu, \widehat{\mu}) = (\mu - \widehat{\mu})^2 = \sum_{i=1}^{p} (\mu_i - \widehat{\mu}_i)^2$$

on le note que $\widehat{\mu}$ definit le vecteur d'estimateur.

2.1 Innadmissibilite de l'estimateur usuel pour $p \geq 3$

2.1.1 la variance connue

Soit X suit la distribution normal de dimension p, avec une moyenne inconnue $\mu = \mathbb{E}(X)$ et une matrice de covariance égale à la matrice d'identité $\mathbb{E}(X - \mu)(X - \mu)^t = I$, nous sommes intéressé à estimer μ, par l'estimateur $\widehat{\mu}$ avec la perte quadratique definit par,

$$L(\mu,\widehat{\mu}) = (\mu-\widehat{\mu})(\mu-\widehat{\mu})^t = \|\mu-\widehat{\mu}\|^2 \tag{2.1}$$

L'estimateur usuel est definit par,

$$\overline{\mu}(X) = X$$

Le risque de l'estimateur usuel est definit par,

$$R(\mu,\overline{\mu}) = \mathbb{E}(L(\mu,X)) = \mathbb{E}(\|X-\mu\|^2) = p$$

ce resultat est obtenue puisque la quantité $\|X-\mu\|^2 \sim> \chi^2_p$ alors $R(\mu,\overline{\mu}) = p$ pour tout $\mu \in \mathbb{R}^p$. On definit l'estimateur de type Stein ou bien (James et Stein) par,

$$\mu_1^{JS}(X) = \left(1 - \frac{c}{\|X\|^2}\right) X \qquad c > 0 \tag{2.2}$$

Le risque de l'estimateur $\mu_1^{JS}(X)$ est definit par,

$$\begin{aligned} R(\mu,\mu_1^{JS}) &= \mathbb{E}\left[\| \left(1-\frac{c}{\|X\|^2}\right) X - \mu\|^2\right] \qquad (2.3)\\ &= \mathbb{E}[\|X-\mu\|^2] - 2c\mathbb{E}\left[\frac{(X-\mu)^t X}{\|X\|^2}\right] + c^2\mathbb{E}\left[\frac{1}{\|X\|^2}\right] \\ &= p - 2c\mathbb{E}\left[\frac{(X-\mu)^t X}{\|X\|^2}\right] + c^2\mathbb{E}\left[\frac{1}{\|X\|^2}\right] \end{aligned}$$

Il est bien connue que $\|X\|^2$ suit la loi de χ^2 noncentre avec p le degré de libertie et $\|\mu\|^2$ le paramètre de noncentralité, on prend

$$U = \frac{\mu^t X}{\|\mu\|} \qquad V = \|X - \frac{\mu^t X}{\|\mu\|^2}\mu\|^2 \tag{2.4}$$

alors,

$$W = \|X\|^2 = U^2 + V \tag{2.5}$$

avec $U \sim> \mathcal{N}(\|\mu\|, 1)$ et V independant de U avec $V \sim> \chi^2_{p-1}$, si on prend k suit la loi de poisson de paramètre $\frac{1}{2}\|\mu\|^2$, donc la distribution conditionnel de W donne k est distribué de la même maniére que $\|X\|^2$ avec la distribution centre de χ^2_{p+2k}, ainsi

$$\mathbb{E}\left[\frac{1}{\|X\|^2}\right] = \mathbb{E}\left[\frac{1}{\chi^2_{p+2k}}\right] = \mathbb{E}\left(\mathbb{E}\left[\frac{1}{\chi^2_{p+2k}}/k\right]\right) = \mathbb{E}\left(\frac{1}{p-2+2k}\right) \tag{2.6}$$

On va calculer le deuxiéme terme de l'égalite (2.3), on à la densité conjointe de U et V est donne comme suit,

$$\begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u-\|\mu\|)^2} \frac{1}{2^{\frac{(p-1)}{2}}\Gamma[(p-1)/2]} v^{\frac{(p-1)}{2}} e^{-\frac{v}{2}} & si \quad v \geq 0 \\ 0 & sinon \end{cases} \tag{2.7}$$

La densité conjointe de U et W est definit par,

$$\begin{cases} \frac{1}{\sqrt{2\pi} 2^{\frac{(p-1)}{2}}\Gamma[(p-1)/2]} (w-u^2)^{\frac{(p-3)}{2}} e^{-\frac{1}{2}\|\mu\|^2 + \|\mu\| u - \frac{1}{2}w} & si \quad u^2 \leq w \\ 0 & sinon \end{cases} \tag{2.8}$$

Il s'ensuite que,

$$\mathbb{E}(\frac{U}{W}) = \frac{e^{-\frac{1}{2}\|\mu\|^2}}{\sqrt{2\pi} 2^{\frac{(p-1)}{2}}\Gamma[(p-1)/2]} \underbrace{\int_0^{\infty} dw \int_{-\sqrt{w}}^{\sqrt{w}} \frac{u}{w}(w-u^2)^{\frac{(p-3)}{2}} e^{\|\mu\| u - \frac{1}{2}w} du} \tag{2.9}$$

avec le changement de variable suivant $t = \frac{u}{\sqrt{w}}$ on trouve,

$$\begin{aligned}
& \int_0^\infty dw \int_{-\sqrt{w}}^{\sqrt{w}} \frac{u}{w}(w-u^2)^{\frac{(p-3)}{2}} e^{\|\mu\|u-\frac{1}{2}w} du \qquad (2.10)\\
= & \int_0^\infty w^{\frac{(p-3)}{2}} e^{\|\mu\|u-\frac{1}{2}w} dw \int_{-1}^{1} t(1-t^2)^{\frac{(p-3)}{2}} e^{\|\mu\|t\sqrt{w}} dt \\
= & \int_0^\infty w^{\frac{(p-3)}{2}} e^{\|\mu\|u-\frac{1}{2}w} dw \sum_{i=0}^{\infty} \frac{(\|\mu\|\sqrt{w})^i}{i!} \int_{-1}^{1} t^{i+1}(1-t^2)^{\frac{(p-3)}{2}} dt \\
= & \sum_{j=0}^{\infty} \frac{\|\mu\|^{2j+1}}{(2j+1)!}) \int_0^\infty w^{\frac{p}{2}+j-1} e^{-\frac{1}{2}w} dw \int_{-1}^{1} t^{2(j+1)}(1-t^2)^{\frac{(p-3)}{2}} dt \\
= & \sum_{j=0}^{\infty} \frac{\|\mu\|^{2j+1}}{(2j+1)!} 2^{\frac{p}{2}+j} \Gamma(\frac{p}{2}+j) B(j+\frac{3}{2}, \frac{p-1}{2}) \\
= & \sum_{j=0}^{\infty} \frac{\|\mu\|^{2j+1} 2^{\frac{p}{2}+j} \Gamma(j+\frac{3}{2}) \Gamma(\frac{p-1}{2})}{(2j+1)!(\frac{p}{2}+j)} \\
= & \Gamma(\frac{p-1}{2} 2^{\frac{p}{2}} \sum_{j=0}^{\infty} \frac{2^j \Gamma(\frac{1}{2}) \|\mu\|^{2j+1}}{2^{2j+1}\Gamma(j+1)(\frac{p}{2}+j)}
\end{aligned}$$

de l'equations (2.4)(2.5)(2.9) et (2.10) on a

$$\begin{aligned}
\mathbb{E}\left[\frac{(X-\mu)^t X}{\|X\|^2}\right] &= 1 - \|\mu\| \mathbb{E}(\frac{u}{w}) \qquad (2.11)\\
&= 1 - \|\mu\| \frac{e^{-\frac{1}{2}\|\mu\|^2}}{\sqrt{\pi}} \sum_{j=0}^{\infty} \frac{\Gamma(\frac{1}{2})\|\mu\|^{2j+1}}{2^{j+1}\Gamma(j+1)(\frac{p}{2}+j)} \\
&= e^{-\frac{1}{2}\|\mu\|^2} \left[\sum_{i=0}^{\infty} \frac{(\frac{1}{2}\|\mu\|^2)^i}{i!} - \|\mu\|^2 \sum_{j=0}^{\infty} \frac{(\frac{1}{2}\|\mu\|^2)^j}{j!(p+2j)}\right] \\
&= e^{-\frac{1}{2}\|\mu\|^2} \sum_{i=0}^{\infty} \frac{(\frac{1}{2}\|\mu\|^2)^i}{i!} \frac{(p-2)}{p-2+2i} \\
&= (p-2)\mathbb{E}\left[\frac{1}{p-2+2k}\right]
\end{aligned}$$

avec $k \sim> p(\frac{\|\mu\|^2}{2})$, donc depuis les resultats de (2.3)(2.6)(2.11), on trouve,

$$\begin{aligned} R(\mu, \mu_1^{js}) &= \mathbb{E}\left[\| \left(1 - \frac{c}{\|X\|^2}\right) X - \mu\|^2\right] \\ &= p - 2(p-2)c\mathbb{E}\left[\frac{1}{p-2+2k}\right] + c^2\mathbb{E}\left[\frac{1}{p-2+2k}\right] \end{aligned}$$

Cette quantité est minimiser au point $c = p - 2$ pour tout μ, cela conduit à l'utilisation de l'estimateur

$$\mu_2^{JS}(X) = \left(1 - \frac{(p-2)}{\|X\|^2}\right) X$$

avec le risque,

$$R(\mu, \mu_2^{js}) = p - \mathbb{E}\left[\frac{(p-2)^2}{p-2+2k}\right] < p \qquad si \quad p \geq 3.$$

2.1.2 La variance inconnue

Maintenat, regardons le cas ou X à une moyenne μ et matrice de variance covariance donné par,

$$\mathbb{E}(X-\mu)(X-\mu)^t = \sigma^2 I$$

et nous observons que $S \sim> \sigma^2\chi_p^2$ independant de X, avec $\mu \quad et \quad \sigma^2$ soit inconnues, on considére l'estimateur suivant,

$$\mu_3^{js}(X, S) = \mu_3^{js} = \left(1 - \frac{aS}{\|X\|^2}\right) X \qquad a > 0. \tag{2.12}$$

on va calculer le risque de ce estimateur, on à

$$\begin{aligned} R(\mu, \mu_3^{js}) &= \mathbb{E}\left[\|\mu_3^{js} - \mu\|^2\right] \qquad (2.13) \\ &= \mathbb{E}\left[\|X-\mu\|^2\right] - 2a\mathbb{E}\left[\frac{S(X-\mu)^t}{\|X\|^2}X\right] + a^2\mathbb{E}\left[\frac{S^2}{\|X\|^2}\right] \\ &= \sigma^2\left\{p - 2a\mathbb{E}[\frac{S}{\sigma^2}]\mathbb{E}\left[\frac{(\frac{X}{\sigma} - \frac{\mu}{\sigma})^t\frac{X}{\sigma}}{\|\frac{X}{\sigma}\|^2}\right] + a^2\mathbb{E}[\frac{S}{\|X\|^2}]^2\mathbb{E}[\frac{1}{\|\frac{X}{\sigma}\|^2}]\right\} \\ &= \sigma^2\left\{p - 2an(p-2)\mathbb{E}\left[\frac{1}{p-2+2k}\right] + a^2n(n+2)\mathbb{E}\left[\frac{1}{p-2+2k}\right]\right\} \end{aligned}$$

Si on remplace (X par X/σ) et (μ par μ/σ) dans (2.6) et (2.11), alors $k \sim> (\|\mu\|^2/2\sigma^2)$, le choix qui minimise (2.13) est $a = \frac{(p-2)}{(n+2)}$ pour tout μ, donc on obtient,

$$R(\mu, \mu_3^{js}) = \sigma^2 \left\{ p - \frac{n}{(n+2)}(p-2)^2 \mathbb{E}[\frac{1}{p-2+2k}] \right\} \tag{2.14}$$

Définition 2.1.1. *(Loi de Wishart)*
En théorie de probabilité et en statistique la loi de Wishart est la généralisation multidimensionelle de la loi de χ^2, la loi est dénouée en l'honneur de John Wishart qui la formula pour la premiére fois en 1928, c'est une loi de probabilité definies sur les matrice (positives, symétrique), une variable aléatoire S de loi de Wishart est donc une matrice aléatoire, on notera $S \sim> W(v, p, n)$.
la loi de Wishart apparait comme loi d'une matrice de variance covariance d'un échontillon de valeurs suit la loi normal multidimensionnell, on definit ca densité par,

$$\begin{cases} \frac{1}{2^{\frac{np}{2}} |v|^{\frac{1}{2}} \Gamma_p(\frac{n}{2})} |X|^{\frac{n-p-1}{2}} e^{-\frac{1}{2} tr(v^{-1}X)} & si \quad n \geq p \\ 0 & sinon \end{cases}$$

Γ_p est la fonction gamma multidimensionnelle definie par,

$$\Gamma_p(\frac{n}{2}) = \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma(n/2 + (1-j)/2)$$

On a definit la loi de Wishart parceque, nous pouvons aussi traiter le cas ou la matrice de covariance inconnue, mais une estimation basée sur la matrice de Wishart, soit X et S indépendant, X est de dimension p suit la distribution normal $\mathcal{N}(\mu, \Sigma)$, et S est distribuer comme matrice de Wishart de dimension $(p \times p)$ avec n degre de libertie et l'esperance égale $n\Sigma$, ou les deux paramètres (μ, Σ) soit inconnue. Supposons que nous voulons estimer μ par μ par $\widehat{\mu}$ avec la fonction de perte suivant,

$$L(\mu, \widehat{\mu}) = (\mu - \widehat{\mu})^t \Sigma^{-1} (\mu - \widehat{\mu}) \tag{2.15}$$

On considère l'estimateur suivant

$$\mu_4^{js}(X,S) = \mu_4^{js} = \left(1 - \frac{e}{X^t S^{-1} X}\right) X \tag{2.16}$$

le risque de l'estimateur μ_4^{js} est donner par,

$$\begin{aligned} R(\mu, \mu_4^{js}) &= \mathbb{E}_{\mu\Sigma}\left[(\mu_4^{js} - \mu)^t \Sigma^{-1} (\mu_4^{js} - \mu)\right] \quad (2.17) \\ &= \mathbb{E}_{\mu\Sigma}\left\{[(1 - \frac{e}{X^t S^{-1} X})X - \mu]^t \Sigma^{-1} [(1 - \frac{e}{X^t S^{-1} X})X - \mu]\right\} \\ &= \mathbb{E}_{\mu^* I}\left\{[(1 - \frac{e}{X^t S^{-1} X})X - \mu^*]^t \Sigma^{-1} [(1 - \frac{e}{X^t S^{-1} X})X - \mu^*]\right\} \end{aligned}$$

telque $\mu^*[(\mu^t \Sigma^{-1} \mu)^{\frac{1}{2}}, 0, \ldots, 0]$, mais c'est bien connu que la distribution conditionnel de $X^t S^{-1} X$ donné X est de la forme $X^t X / S^*$, ou $S^* \sim> \chi^2_{n-p+1}$ indépendant de X.

Par la comparaison de (2.13) (2.17) on voi que le choix optimum de e est $\frac{(p-2)}{(n-p+3)}$ et donc la fonction de risque devient,

$$R(\mu, \mu_4^{js}) = p - \frac{n-p+1}{n-p+3}(p-2)^2 \mathbb{E}[\frac{1}{p-2+2k}] \tag{2.18}$$

avec $k \sim> p((\frac{1}{2})\mu^t \Sigma^{-1} \mu)$.

L'améléoration obtennue par ces estimateurs sur l'estimateur usuel peut être mieux compris, si on considére le cas ou la matrice de covariance est connue et égale à I_p, et si on considére que l'erreur orthogonal de tout estimateur X est $\mu - (\frac{\mu^t X}{\|X\|^2})X$, et sont moyenne quadratique est,

$$\begin{aligned} \mathbb{E}_\mu[\|\mu - (\frac{\mu^t X}{\|X\|^2})X\|^2] &= \|\mu\|^2 - \mathbb{E}_\mu[\frac{(\mu^t X)^2}{\|X\|^2}] \quad (2.19) \\ &= \|\mu\|^2 (1 - \mathbb{E}_\mu[\frac{1+2k}{p+2k}]) \\ &= (p-1)\|\mu\|^2 \mathbb{E}_\mu[\frac{1}{p+2k}] \\ &= (p-1)\left[1 - (p-2)\mathbb{E}_\mu[\frac{1}{p-2+2k}]\right] \end{aligned}$$

ainsi que la moyenne quadratique de l'erreur de $\mu_2^{js} = \left[1 - \frac{(p-2)}{\|X\|^2}\right] X$ est,

$$\left[p - (p-2)^2 \mathbb{E}_\mu(\frac{1}{p-2+2k})\right] - (p-1)\left[1 - (p-2)\mathbb{E}_\mu(\frac{1}{p-2+2k})\right] = 1 + (p-2)\mathbb{E}_\mu(\frac{1}{p-2+2k}) \leq 2 \tag{2.20}$$

2.2 Les estimateurs spheriquement symétriques

Définition 2.2.1. *Un estimateur $\widehat{\mu}$ est dit spheriquement symétrique (SS) (sur l'origine) si il ecrit de la forme suivante,*

$$\widehat{\mu}(X) = [1 - h(X^2)]X \tag{2.21}$$

Remarque 2.2.1. *On remarque que les estimateurs de type Stein sont spheriquement symétrique, on peut dit que l'estimateur de type Stein* ***Shrink*** *l'estimateur usuel X (vers 0) par un facteur* $[1 - h(X^2)]$.

maintenant on va démontrer que si un estimateur (SS), $\widehat{\mu}_2$ est admissible par comparaison avec touts les estimateurs spheriquement symétrique, donc il est admissible dans la classe de tout les estimateurs. On à par la transformation orthogonal g,

$$g \circ \widehat{\mu} \circ g^{-1} = g[\widehat{\mu}(g^{-1}(x))] = \widehat{\mu}(x) \qquad \textit{pour tout} \quad x \tag{2.22}$$

si $\widehat{\mu}$ est de la forme (2.21), alors

$$g[\widehat{\mu}(g^{-1}(x))] = g\left[(1 - h(x^2))g^{-1}(x)\right] = \widehat{\mu}(x)$$

Suppose inversement que $\widehat{\mu}$ verifié (2.22) pour tout g orthogonal, en particuliére pour $g[\widehat{\mu}(x)] = \widehat{\mu}(x)$, donc $\widehat{\mu}(x)$ se trouve au long de X, c'est definit par,

$$\widehat{\mu}(x) = [1 - h'(X)]X$$

pour quelques fonctions h' à valeur reél, tantque le vecteur X, peut prend dans un autre vecteur à la même norme au carré x^2 avec une transformation orthogonale.

La preuve est basé la capacité du groupe orthogonal G, et la continuté du probléme, c'est similaire à un groupe finie, avec la condition de la convexité du fonction de perte quadratique, nous pouvons limiter notre attention aux procedures non randomisées.

Suppose que l'estimateur $\widehat{\mu}$ est mieux que l'estimateur (SS) $\widehat{\mu}_2$, donc

$$R_{\widehat{\mu}}(\mu) = \mathbb{E}_\mu[\widehat{\mu}(X) - \mu]^2 \leq \mathbb{E}_\mu[\widehat{\mu}_2(X) - \mu]^2 \quad \forall \mu \tag{2.23}$$

avec la condition de continuté de $R_{\widehat{\mu}}\, et\, R_{\widehat{\mu}_2}$, l'innégalité précédent reste verifié pour $\mu \in S$ (S un ensemble ouvert non vide), tantque $\widehat{\mu}_2$ est (SS) alors (2.23) restera vraie si on remplace $\widehat{\mu}$ par $g \circ \widehat{\mu} \circ g^{-1}$ avec g orthogonal, on obtient pour un $\mu \in S$ fixé,

$$\begin{aligned} R_{g\circ\widehat{\mu}\circ g^{-1}}(\mu) &= \mathbb{E}_\mu[g\{\widehat{\mu}(g^{-1}(X))\} - \mu]^2 \qquad (2.24) \\ &= \mathbb{E}_\mu[\widehat{\mu}(g^{-1}(X)) - g^{-1}(\mu)]^2 = R_{\widehat{\mu}}(g^{-1}(\mu)). \end{aligned}$$

L'ensemble de g dans lequel $R_{g\circ\widehat{\mu}\circ g^{-1}}(\mu) < R_{\widehat{\mu}_2}(\mu)$, est un ensemble ouvert non vide.

Soit γ la mesure de probabilité invariant sur G, qui attribue une mesure strictement positive pour un ensemble ouvert non vide (pour l'existence de cette mesure voir Chapitre 2,[22]), alors,

$$\mu' = \int g \circ \mu \circ g^{-1} d\gamma(g) \tag{2.25}$$

est un estimateur (SS), et avec la condition du convexite du fonction de perte quadratique,

$$R_{\widehat{\mu}_2'}(\mu) \leq \int R_{g\circ\widehat{\mu}\circ g^{-1}} d\gamma(g) \leq R_{\widehat{\mu}_2}(\mu) \quad \textit{pour tout } \mu \in S \tag{2.26}$$

Donc on conclure que $\widehat{\mu}_2$ n'est pas admissible dans la classe des estimateurs (SS), alors on obtient le resultat par contraposé.

2.3 L'admissibilité de l'estimateur usuel pour p=2

Cette preuve à été démentrer avec Stein on (1956) dans [26] , avec l'aide des resultats obtenue sur les estimateurs (SS) dans la partie précedente, on va montrer que pour (p=2) l'estimateur usuel definit par $\overline{\mu}(X) = X$ est admissible.

Soit $\widehat{\mu}$ un estimateur de μ avec un risque finie, et soit b le biais de $\widehat{\mu}$ definit par,

$$b(\widehat{\mu}) = \mathbb{E}_{\mu}[\widehat{\mu}(X)] - \mu] \tag{2.27}$$

donc, avec l'innégalité d'information,

$$R(\widehat{\mu}) \geq b^2(\widehat{\mu}) + \sum_i \left\{ \sum_j \eta_{ij}[\delta_{ij} + b_{ij}(\widehat{\mu})] \right\}^2 \tag{2.28}$$

pour tout η avec $\sum_i \eta_{ij}^2 = 1$, pour tout i, et

$$\delta_{ij} = \begin{cases} 1 & si \quad i = j \\ 0 & \quad sinon \end{cases}$$

$$b_{ij}(\widehat{\mu}) = \frac{\partial}{\partial \mu_j} b_i(\widehat{\mu}) \tag{2.29}$$

avec $b_i(\widehat{\mu})$ est la i-éme coordonné de partie biais $b(\widehat{\mu})$ choisir,

$$\eta_{ij} = \frac{\delta_{ij} + b_{ij}(\widehat{\mu})}{\sqrt{\sum_j [\delta_{ij} + b_{ij}(\widehat{\mu})]^2}} \tag{2.30}$$

afin de maximiser la quantité à droit de l'innégalité (2.28), on obtient

$$\begin{aligned} R(\widehat{\mu}) &\geq b^2(\widehat{\mu}) + \sum_{i,j} [\delta_{ij} + b_{ij}(\widehat{\mu})]^2 \qquad (2.31) \\ &= b^2(\widehat{\mu}) + p + 2\sum_i b_{ii}(\widehat{\mu}) + \underbrace{\sum_{i,j} b_{ij}^2(\widehat{\mu})}_{} \end{aligned}$$

dans le cas où $\widehat{\mu}$ est un estimateur (SS) (écrit sous la forme (2.21)), alors le biais devient,

$$b(\widehat{\mu}) = -\varphi(\mu^2)\mu \tag{2.32}$$

ou φ est une fonction differentiable à valeur reélle, dans ce cas laisser tember le dernier terme de (2.31) donne,

$$R(\widehat{\mu}) \geq p + \mu^2\varphi^2(\mu^2) - 2p\varphi(\mu^2) - 4\mu^2\varphi'(\mu^2) \tag{2.33}$$

On va utiliser cette formule pour montrer notre resultat,si on pose que $\overline{\mu}(X)$ n'est pas admissible, alors il existe un estimateur (SS) $\widehat{\mu}$ qui est mieux que $\overline{\mu}$, on plus il existe une fonction φ (ne pas disparaître de façon identique), telque

$$2 \geq R(\widehat{\mu}) \geq 2 + \mu^2\varphi^2(\mu^2) - 4\varphi(\mu^2) - 4\mu^2\varphi'(\mu^2) \tag{2.34}$$

avec le changement de variable suivant, pour tout $\mu^2 > 0$, soit $t = \mu^2$ et $\psi(t) = t\varphi(t)$, on obtient donc,

$$0 \geq \frac{1}{t}\psi^2(t) - 4\psi'(t) \quad t > 0 \tag{2.35}$$

Ceci montre que $\psi(t)$ est une foncion croissante, on va montrer que (2.35), implique que ψ est identiquement nul.

Suppose que $\psi(t_0) < 0$ pour $t_O > 0$, donc l'intégrale de l'innégalité suivante sur l'intervalle $]0, t_0[$,

$$\frac{\psi'(t)}{\psi^2(t)} \geq \frac{1}{4t} \tag{2.36}$$

est donne,

$$-\frac{1}{\psi(t_0)} + \frac{1}{\psi(t)} \geq \frac{1}{4}\log(\frac{t_0}{t}) \tag{2.37}$$

la quntité à gauche est borné quand ($t \mapsto 0$), tandis que la quantité à droit tend vers l'infinie, et c'est une contraduction.

Suppose maintenant que $\psi(t_0) > 0$, pour tout $t_0 > 0$ alors

$$\frac{1}{\psi(t_0)} - \frac{1}{\psi(t)} \geq \frac{1}{4}\log(\frac{t}{t_0}) \quad t > t_0. \tag{2.38}$$

quand ($t \mapsto \infty$), la quantité à gauche est bornée, tandis que la quantité à droit tend vers l'infinie ($+\infty$), et encore c'est une contraduction, donc la fonction ($\psi = 0$), ce qui fallait demontrer.

Maintenant, on utilise (2.31), pour montrer que pour $p \geq 3$, il n'existe pas $c > (p-2)^2$ et $\overline{\mu}^2$, telque pour tout $\mu^2 \geq \overline{\mu}^2$, on a

$$R(\widehat{\mu}) \leq p - \frac{c}{\mu^2} \tag{2.39}$$

Il faut montrer que l'innégalité differentiable suivante,

$$p - \frac{c}{\mu^2} \geq p + \mu^2\varphi^2(\mu^2) - 2p\varphi(\mu^2) - 4\mu^2\varphi'(\mu^2) \tag{2.40}$$

obtenue de (2.33) et (2.39), n'a pas de solution pour tout $\mu^2 \geq \overline{\mu}^2$, pour avoir ce resultat soit,

$$\varphi(\mu^2) = \frac{(p-2)}{\mu^2} + f(\mu^2) \tag{2.41}$$

donc (2.40) devient,

$$\frac{c-(p-2)^2}{\mu^2} \geq \mu^2 f^2(\mu^2) - 4f(\mu^2) - 4\mu^2 f'(\mu^2) \tag{2.42}$$

a l'aide de changement de variable suivant, $t = \mu^2$ et $\psi(t) = tf(t)$, on obtient,

$$-\frac{c-(p-2)^2}{t} \quad \geq \quad \frac{1}{t}\psi^2(t) - 4\psi'(t) \tag{2.43}$$

$$\frac{\psi'(t)}{a^2 + \psi^2(t)} \quad \geq \quad \frac{1}{4t} \tag{2.44}$$

avec $a^2 = c - (p-2)^2$, depuis l'innegalité (2.35) pour $(t \geq t_0)$, qui est plus faible que (1.43), on conclure que $\psi(t) < 0$, on a comme resultat pour tout $t_0 < t$,

$$\tan^{-1}\left(\frac{\psi(t)}{a}\right) - \tan^{-1}\left(\frac{\psi(t_0)}{a}\right) \geq \frac{1}{4a}\log\left(\frac{t}{t_0}\right) \tag{2.45}$$

La quantitée à gauche est bornée (tantque ψ ne change pas le signe), et la quantité à droit tend vers $(+\infty)$ quand $t \mapsto \infty$, ce qui est contraduction.

Chapitre 3

Probleme reliée à la fonction de perte

3.1 Clarifiez le problème

En voyant les resultats donnée dans le chapitre précédent, plusieurs personnes ont exprimé la créante qu'elles étaient étroitement liées avec l'utilisation d'une fonction de perte illimitée, que beaucoup de gens considérent comme déraisonnable, Charles stein et James donne un exemple pour montrer qu'au moins qualitativement, ce n'est pas le cas pour simplifier on suppose X un p-vecteur de variable aléatoire distribuer selon la loi normal de moyenne μ, et matrice de covariance égale à I_p.
Supposons que nous sommes intéressés à estimer μ par $\widehat{\mu}$ avec la fonction de perte definie par

$$L(\mu, \widehat{\mu}) = \rho(\|\mu - \widehat{\mu}\|^2) \tag{3.1}$$

sous l'hypothèse que ρ a une dérivé borné et soit continument differentiable, positive croissant et concave ($\rho"(t) \leq 0, \quad \textit{pour tout } t > 0$).

3.2 Formulation du probleme

On va démontrer que pour un petit b, et grand a (indépendant de μ), l'estimateur [1] $\mu_5^{JS}(X)$ definie par

$$\mu_5^{JS} = (1 - \frac{b}{a + \|X\|^2})X \tag{3.2}$$

à un petit risque par rapport à le risque de l'estimateur usuel X.

on pose que $(Y = X - \mu)$ on obtient

$$\begin{aligned} R(\mu, \mu_5^{JS}) &= \mathbb{E}_\mu \rho \left[\|(1 - \frac{b}{a + \|X\|^2})X - \mu\|^2 \right] \\ &= \mathbb{E}_\mu \rho \left[\|X - \mu\|^2 - 2b \frac{X^t(X - \mu)}{a + \|X\|^2} + b^2 \frac{\|X\|^2}{(a + \|X\|^2)^2} \right] \\ &\leq \mathbb{E}_\mu[\rho(\|X - \mu\|^2)] - 2b\mathbb{E}_\mu \left[\frac{X^t(X - \mu) - \frac{b}{2}}{a + \|X\|^2} \right] \rho'(\|X - \mu\|^2) \\ &= \mathbb{E}_\mu[\rho(\|Y\|^2)] - \underbrace{2b\mathbb{E}_\mu \left[\frac{(Y + \mu)^t Y - \frac{b}{2}}{a + \|Y + \mu\|^2} \right] \rho'(\|Y\|^2)} \end{aligned} \tag{3.3}$$

juste comme Stein 1956 [26] on peut obtient avec le developpement de taylor de

$$\frac{1}{1 + x} = 1 - x + \frac{x^2}{1 + x} \tag{3.4}$$

puis on remplace (3.4) par $\frac{1}{a + \|Y + \mu\|^2}$, le dernier terme à droite de l'égalité (3.3) est donne

1. Ce estimateur est definit pour la premiére fois par Stein en 1956 [26], et il démontrer que $\mu_5^{JS}(X)$ admissible sous la perte quadratique si $(p > 2)$.

$$\begin{aligned}
\mathbb{E}_\mu\left[\frac{(Y+\mu)^tY-\frac{b}{2}}{a+\|Y+\mu\|^2}\right]\rho'(\|Y\|^2) &= \mathbb{E}_\mu\left[\frac{(Y+\mu)^tY-\frac{b}{2}}{a+\|Y\|^2+\|\mu\|^2}\right]\left(1-\frac{2\mu^tY}{a+\|Y\|^2+\|\mu\|^2}\right) \\
&\quad \rho'(\|Y\|^2)+O(\frac{1}{a+\|\mu\|^2}) \\
&= \mathbb{E}_\mu\left[\mathbb{E}_\mu\left\{\frac{(Y+\mu)^tY-\frac{b}{2}}{a+\|Y\|^2+\|\mu\|^2}\left(1-\frac{2\mu^tY}{a+\|Y\|^2+\|\mu\|^2}\right)/\|Y\|^2\right\}\right] \\
&\quad \rho'(\|Y\|^2)+O(\frac{1}{a+\|\mu\|^2}) \\
&= \mathbb{E}_\mu\left\{\frac{\|Y\|^2-\frac{b}{2}}{a+\|Y\|^2+\|\mu\|^2}\left(\frac{2\|\mu\|^2\|Y\|/2}{(a+\|Y\|^2+\|\mu\|^2)^2}\right)/\|Y\|^2\right\} \\
&\quad \rho'(\|Y\|^2)+O(\frac{1}{a+\|\mu\|^2}) \\
&\geq \mathbb{E}_\mu\left[\frac{(1-\frac{2}{p})\|Y\|^2-\frac{b}{2}}{a+\|Y\|^2+\|\mu\|^2}\right] \\
&\quad \rho'(\|Y\|^2)+O(\frac{1}{a+\|\mu\|^2})
\end{aligned} \tag{3.5}$$

Pour voir que cette quantité et positive, sous la condition que (b) est petit, et (a) est garnd nous regardons,

$$f(A) = A\mathbb{E}\left(\frac{\|Y\|^2\rho'(\|Y\|^2)}{A+\|Y\|^2}\right) \tag{3.6}$$

$$et \qquad g(A) = A\mathbb{E}\left(\frac{\rho'(\|Y\|^2)}{A+\|Y\|^2}\right) \tag{3.7}$$

C'est claire que f et g deux fonctions continues strictement positives sur $[1,\infty[$, avec une limite non nulle à l'infinie, il sensuit que

$$c = \inf_{A\geq 1} f(A) = 0 \tag{3.8}$$

$$d = \sup_{A\geq 1} g(A) > \infty \tag{3.9}$$

Si b est choisi de sorte que

$$(1-\frac{2}{p})c-\frac{b}{2}d > 0 \tag{3.10}$$

on a de (3.3) et (3.5), que pour un (a) suffisamment grand

$$R(\mu, \mu_5^{js}) \leq \mathbb{E}_\mu \rho(\|X-\mu\|^2) \quad \textit{pour tout} \;\; \mu \tag{3.11}$$

3.3 L'hypothèse de normalité

L'innadmissibilitéde l'estimateur usuel pour $p \geq 3$, n'exige pas l'hypothèse de normalité, nous donnerons ce resultats sons preuve, Soit X un p-vecteur de variables aléatoire, avec moyenne μ, et du moment d'ordre 4 finie,

$$\mathbb{E}_\mu (X_i - \mu_i)^4 \leq M < \infty \tag{3.12}$$

pour simplifier la notation, nous supposons que les X_i ne sont pas corrélés, est on ecrit

$$\mathbb{E}_\mu[(X_i - \mu_i)^2] = \sigma_i^2 \tag{3.13}$$

alors pour $p \geq 3$, $b < 2(p-2)\min \sigma_i^2$ et suffisamment grand a, depend seulment de M, on a

$$\mathbb{E}_\mu \sum_{i=1}^{p} \left[\left(1 - \frac{b}{\sigma_i^2 [a + \sum \frac{X_i^2}{\sigma_i^2}]} \right) X_i - \mu_i \right]^2 < \mathbb{E}_\mu [\sum_{i=1}^{p} (X_i - \mu_i)^2] = \sum_{i=1}^{p} \sigma_i^2 \tag{3.14}$$

Remarque 3.3.0.1. *Il serait souhaitable d'obtenir des formules explicites pour les estimateurs que l'on peut sérieusement recommander dans les deux cas précedents.*

3.4 Formulation du probléme générale de l'estimation admissible avec la perte quadratique

Soit l'espace $(Z, \mathcal{A}, \lambda)$ avec Z l'espace d'échantillon, et $\mathcal{A}$ une tribue et λ la mesure σ-finie sur $\mathcal{A}$, et soit l'espace $(\Theta, \mathcal{C}, \mathbb{P}(\cdot/\cdot))$, avec Θ l'espace des paramètres $\beta \in \Theta$, et $\mathcal{C}$ une tribue et $\mathbb{P}(\cdot/\cdot)$ la fonction $\mathcal{AC}$-mesurable à valeur positive dans $(Z \times \Theta)$, telque pour tout $\beta \in \Theta$ on a $\mathbb{P}(\cdot/\beta)$ est une densité de probabilité par rapport à λ, definit par

$$\int \mathbb{P}(z/\beta) d\lambda(z) = 1 \tag{3.15}$$

Soit A l'espace d'action ou (l'espace de decision) de dimension k avec des coordonnés reéles, et α la fonction $\mathcal{C}$-mesurable definit sur Θ a l'ensemble des matrices symétriques semidefinit positive de dimension $(k \times k)$, et η la fonction $\mathcal{C}$-mesurable definit sur Θ a A.

On remarque que z est distribuer sur Z, selon la densité de probabilité $\mathbb{P}(\cdot/\beta)$ par rapport à λ, ou β est inconnue, si on choisie une action $a \in A$, alors on definit la fonction de perte suivante

$$L(\beta, a) = [a - \eta(\beta)]^t \alpha(\beta)[a - \eta(\beta)] \tag{3.16}$$

Un estimateur $\widehat{\eta}$ de $\eta(\beta)$, est une fonction $\mathcal{A}$-mesurable definit sur Z à A, ca fonction de risque $R(\cdot, \widehat{\eta})$ est la fonction sur Θ, definit par

$$R(\beta, \widehat{\eta}) = \mathbb{E}_\beta \left([\widehat{\eta}(z) - \eta(\beta)]^t \alpha(\beta)[\widehat{\eta}(z) - \eta(\beta)] \right) \tag{3.17}$$

On veut choisir $\widehat{\eta}$ afin de garder $R(\beta, \widehat{\eta})$ petit, mais ce n'est pas une déclaration précise, pour tout estimateur $\widehat{\eta}$ donné, on peut modifiér $\widehat{\eta}$, pour diminuer $R(\beta, \widehat{\eta})$ pour quelque β, mais l'augmenter avec d'autres β, dans beaucoup de problemes il existe un estimateur couramment utilisé, par exemple, l'estimateur de maximum de vraicemblance, donc il est naturel de poser la question que ce estimateur est admissible au sens de Wald.

On donne un condition suffisente pour la classe des estimateurs presque admissible, il faut dit que dans [30], la condition de bornéture du risque a été omis par inadvertance dans les théorèmes de [29] ,(il est necessaire de justifié la formule (3.20)), si π est une mesure σ-finie sur $\mathcal{C}$, Nous devons définir

$$\widehat{\eta}_\pi \left[\int \alpha(\beta)\mathbb{P}(x/\beta)d\pi(\beta)\right]^{-1} \int \alpha(\beta)\eta(\beta)\mathbb{P}(x/\beta)d\pi(\beta) \tag{3.18}$$

avec la condition que les intégrales précédents soit finie, on remarque que si π est une mesure de probabilité, alors $\widehat{\eta}_\pi$ est l'estimateur de bayes de $\eta(\beta)$, telque $\widehat{\eta} = \widehat{\eta}_\pi$, qui minimiser

$$\int \mathbb{E}_\beta \left([\widehat{\eta}(X) - \eta(\beta)]^t \alpha(\beta)[\widehat{\eta}(X) - \eta(\beta)]\right) d\pi(\beta) \tag{3.19}$$

si q est une densité de probabilité par rapport à π, alors on peut ecrit,

$$(q, \pi)(S) = \int_S q(\beta)d\pi(\beta) \tag{3.20}$$

Théorème 3.4.1. *Si $\widehat{\eta}$ est un estimateur de $\eta(\beta)$ avec un risque borné, telque pour tout ensemble $\mathcal{C} \subseteq \mathcal{F}$ ($\mathcal{F}$ est une famille dénombrable des ensembles dont l'union est Θ, on a*

$$\inf_{q \in S(C)} \frac{\int \mathbb{E}_\beta \left([\widehat{\eta}(X) - \widehat{\eta}_\pi(X)]^t \alpha(\beta)[\widehat{\eta}(X) - \widehat{\eta}_\pi(X)]\right) q(\beta)d\pi(\beta)}{\int_C q(\beta)d\pi(\beta)} = 0 \tag{3.21}$$

ou $S(C)$ est l'ensembes des densités de probabilité par rapport à π, qui sont constant (mais pas nulle) sur C.
Alors, $\widehat{\eta}$ est un estimateur ***presque admissible*** *par rapport à π.*

Démonstration :

Suppose le contraire, c'est à dire il existe un estimateur $\overline{\eta}$, telque

$$\mathbb{E}_\beta \left([\overline{\eta}(X) - \eta(\beta)]^t \alpha(\beta)[\overline{\eta}(X) - \eta(\beta)]\right) \leq \mathbb{E}_\beta \left([\widehat{\eta}(X) - \eta(\beta)]^t \alpha(\beta)[\widehat{\eta}(X) - \eta(\beta)]\right) \tag{3.22}$$

Avec une inégalité stricte sur un ensemble S de partie positive π-mesurable, on choisir $\varepsilon > 0$ et $C \subseteq \mathcal{F}$, telque $\pi(C \cap S_\varepsilon) > 0$ ou S_ε est l'ensemble de β

qui verifié,

$$\mathbb{E}_\beta\left([\overline{\eta}(X)-\eta(\beta)]^t\alpha(\beta)[\overline{\eta}(X)-\eta(\beta)]\right) \leq \mathbb{E}_\beta\left([\widehat{\eta}(X)-\eta(\beta)]^t\alpha(\beta)[\widehat{\eta}(X)-\eta(\beta)]\right)-\varepsilon \tag{3.23}$$

donc pour tout $q \in S(C)$

$$\int \mathbb{E}_\beta\left([\overline{\eta}(X)-\eta(\beta)]^t\alpha(\beta)[\overline{\eta}(X)-\eta(\beta)]\right) q(\beta)d\pi(\beta) \leq$$

$$\int \mathbb{E}_\beta\left([\widehat{\eta}(X)-\eta(\beta)]^t\alpha(\beta)[\widehat{\eta}(X)-\eta(\beta)]\right) q(\beta)d\pi(\beta) - \varepsilon\int_{C\cap S_\varepsilon} q(\beta)d\pi(\beta)$$

il s'ensuit que $\dfrac{\varepsilon\pi(C\cap S_\varepsilon)}{\pi(C)}$ est égale,

$$\begin{aligned}
&= \varepsilon\frac{\displaystyle\int_{C\cap S_\varepsilon} q(\beta)d\pi(\beta)}{\displaystyle\int_C q(\beta)d\pi(\beta)} \\
&\leq \frac{1}{\displaystyle\int_C q(\beta)d\pi(\beta)}\int \mathbb{E}_\beta\left([\widehat{\eta}(X)-\eta(\beta)]^t\alpha(\beta)[\widehat{\eta}(X)-\eta(\beta)]\right) q(\beta)d\pi(\beta) \\
&- \int \mathbb{E}_\beta\left([\overline{\eta}(X)-\eta(\beta)]^t\alpha(\beta)[\overline{\eta}(X)-\eta(\beta)]\right) q(\beta)d\pi(\beta) \\
&\leq \frac{1}{\displaystyle\int_C q(\beta)d\pi(\beta)}\int \mathbb{E}_\beta\left([\widehat{\eta}(X)-\eta(\beta)]^t\alpha(\beta)[\widehat{\eta}(X)-\eta(\beta)]\right) q(\beta)d\pi(\beta) \\
&- \int \mathbb{E}_\beta\left([\widehat{\eta}_\pi(X)-\eta(\beta)]^t\alpha(\beta)[\widehat{\eta}_\pi(X)-\eta(\beta)]\right) q(\beta)d\pi(\beta) \\
&= \frac{\int \mathbb{E}_\beta\left([\widehat{\eta}(X)-\widehat{\eta}_\pi(X)]^t\alpha(\beta)[\widehat{\eta}(X)-\widehat{\eta}_\pi(X)]\right) q(\beta)d\pi(\beta)}{\displaystyle\int_C q(\beta)d\pi(\beta)}
\end{aligned}$$

on à contraduction avec (3.21).

Chapitre 4

Sur l'estimation sous certainnes classes de fonction de perte équilibrée généralisée

4.1 Sur l'estimation sous certainnes classes de fonction de perte équilibrée pondérée

Dans cette section, nous présentons une classe de fonction de perte, pour estimer un paramètre inconnue β, qui connue sous le non (la fonction de perte équilibré pondéré) definit par

$$L_{w,\beta_0}(\beta,\widehat{\beta}) = wq(\beta)(\widehat{\beta}-\beta_0)^2 + (1-w)q(\beta)(\widehat{\beta}-\beta)^2 \tag{4.1}$$

avec $0 \leq w \leq 1$, $q(\beta)$ est une fonction de poids positive, et β_0 un estimateur référence "priori", plusieurs exemples sont donnés en ce qui concerne (admissibilité, dominonce, bayésianité et minimaxité) dans plusieurs cas comme dans [11] , on va montrer que les resultats pour la perte L_{w,β_0}, peuvent être déduit directement du perte quadratique pondéré (w=0), des problemes spécifiques liés à l'espace des paramètres contraints, qui comprennent le choix de l'estimateur référence β_0, on prend β_0 est l'estimateur de maximum de vraicemblance de β, dans cette section on va estimer le paramètre inconnue

β, sous le modèle $(X_1, \ldots, X_p) \sim F_\beta$, et la perte (4.1).

Ce perte dépend au valeur observé de $\beta_0(X)$, et refléte la proximité de $\widehat{\beta}$ a β sous la perte quadratique pondéré, et la proximité de $\widehat{\beta}$ a β_0 sous le carré de distance pondéré, le premier terme à droit de (1.1) est analogue à un terme de pinalité pour manque de regularité dans la regression non paramétrique, le poids w déterminer l'importance relative de ces deux critéres, le choix de perte (4.1) est inspiré par Zellner 1994 [1] dans le cas $q(\beta) = 1$ (la fonction de perte équilibré), et β_0 est l'estimateur de moindre carré, en particuliére si $\beta_0(X) = \overline{X}$,alors la perte (4.1) devient

$$\frac{w}{p}\sum_{i=1}^{p}(X_i - \widehat{\beta})^2 + (1-w)(\widehat{\beta} - \beta)^2$$

cependant, il est utile de considérer β_0 comme un estimateur rééférence plus géneral que l'estimateur de moindre carré, comme un shéma unificateur pour établir des propriétés théorique de decision, de plus nous étudions l'application de (4.1) dans le contexte d'espace de paramètre restreints Θ, ou il sera souhaitable que l'estimateur rééférence prenne des valeurs uniquement dans Θ.

Comme le montre [Dey 1999 [11]] le probléme de l'admissibilité, la dominance et la bayésianité pour la fonction de perte équilibré (FPE) ne dépend pas de w, et peut être déduit par des resultats précédemment ou à être établis sous la perte quadratique $(w = 0)$.

4.1.1 Admissibilité et dominance

Pour l'estimateur $\widehat{\beta}(X)$, on note le risque associé au perte (4.1) par $R_{w,\beta_0}(\beta, \widehat{\beta})$, dans le cas $(w = 0)$ on écrit tout simplement $R_0(\beta, \widehat{\beta})$, le risque est donc independant de β_0, et on note la fonction de perte L_{0,β_0} par L_0, et la différence entre les pertes Δ_{0,β_0} par Δ_0, comme dans [11], rapellons que Δ_{w,β_0} est definit par

$$\Delta_{w,\beta_0}(\beta,\widehat{\beta}_1,\widehat{\beta}_2) = L_{w,\beta_0}(\beta,\widehat{\beta}_1) - L_{w,\beta_0}(\beta,\widehat{\beta}_2)$$

Lemme 4.1.1. *Se lemme est donne la relations entres les fonctions de pertes L_{w,β_0} ,R_{w,β_0} et Δ_{w,β_0} quand $0 \leq w < 1$, on à*

1. $\Delta_{w,\beta_0}(\beta,\beta_0+(1-w)g,\beta_0) = (1-w)^2\Delta_0(\beta,\beta_0+g,\beta_0)$
2. $\Delta_{w,\beta_0}(\beta,\beta_0+(1-w)g_1,\beta_0+(1-w)g_2) = (1-w)^2\Delta_0(\beta,\beta_0+g_1,\beta_0+g_2)$
3. $R_{w,\beta_0}(\beta,\beta_0(X)+(1-w)g(X)) = w(1-w)R_0(\beta,\beta_0(X))+(1-w)^2R_0(\beta,\beta_0(X)+g(X))$

Démonstration :

Prener d'abord l'esperance de l'égalité 1 on obtient alors,

$$R_{w,\beta_0}(\beta,\beta_0(X)+(1-w)g(X)) = R_{w,\beta_0}(\beta,\beta_0(X))+(1-w)^2\{R_0(\beta,\beta_0(X)+g(X))-R_0(\beta,\beta_0(X))\}$$

l'égalité 3 étant donné que

$$R_{w,\beta_0}(\beta,\beta_0(X)) = (1-w)R_0(\beta,\beta_0(X))$$

l'égalité 2 est déduit directement comme resultats de 1

$$\Delta_{w,\beta_0}(\beta,\beta_0+(1-w)g_1,\beta_0+(1-w)g_2) = \Delta_{w,\beta_0}(\beta,\beta_0+(1-w)g_1,\beta_0)-\Delta_{w,\beta_0}(\beta,\beta_0+(1-w)g_2,\beta_0)$$

finallement, l'égalité 1 implique de simlifier le terme suivant

$$\begin{aligned} L_{w,\beta_0}(\beta,\beta_0+(1-w)g) &= q(\beta)\{g^2(1-w)^2+2g(1-w)^2(\beta_0-\beta) \\ &+ (1-w)(\beta_0-\beta)^2\} \\ &= q(\beta)(1-w)^2\{g^2(1-w)^2+2g(\beta_0-\beta)\}+L_{w,\beta_0}(\beta,\beta_0) \end{aligned}$$

alors on à

$$\Delta_{w,\beta_0}(\beta,\beta_0+(1-w)g,\beta_0) = q(\beta)(1-w)^2\{g^2(1-w)^2+2g(\beta_0-\beta)\}$$

maintenat ce qui suit est immédiate.

Corollaire 4.1.1. .

1. *l'estimateur* $\beta_0(X) + (1-w)g(X)$ *est admissible sous la perte* L_{w,β_0}, *si est seulement si l'estimateur* $\beta_0(X) + g(X)$ *est admissible sous la perte* L_0.
2. $\beta_0(X) + (1-w)g_1(X)$ *domine* $\beta_0(X) + (1-w)g_2(X)$ *sous la perte* L_{w,β_0}, *si est seulement si* $\beta_0(X) + g_1(X)$ *domine* $\beta_0(X) + g_2(X)$ *sous la perte* L_0.

Remarque 4.1.1. *Pour compléter ce qui précéde en ce qui concerne l'admissibilité et la dominance, des connexions similaires entre les estimateurs de bayes pour la variable* w, *sont facilement derivées et données dans le lemme suivant, en particuliére l'estimateur de bayes sous la perte* L_{w,β_0}, *peut être exprimé simplement comme une combinaison lineaire convexe de l'estimateur de bayes sous la perte* L_0 *(poids* $(1-w)$*) et* β_0 *(poids* (w)*).*

Lemme 4.1.2. *sous la perte (4.1), l'estimateur de bayes de* β *avec la loi à priori* π, *admet la représentation suivante*

$$\beta_\pi(X) = w\beta_0(X) + (1-w)\frac{\mathbb{E}_\pi[\beta q(\beta)/X]}{\mathbb{E}_\pi[q(\beta)/X]}$$

de maniére équivalente,

$$\beta_\pi(X) = \beta_0(X) + (1-w)\left\{\frac{\mathbb{E}_\pi[\beta q(\beta)/X]}{\mathbb{E}_\pi[q(\beta)/X]} - \beta_0(X)\right\} \tag{4.2}$$

avec $\mathbb{E}_\pi[\beta^i q(\beta)/X] < \infty, \quad i = 0, 1,$ *avec probabilité 1.*

Démonstration :

La preuve est simillaire à celle donnée par [11] dans le cas ou $q(\cdot) = 1$, il découle même de resultat de [11] pour le cas $q(\cdot) = 1$, est donne que l'estimateur de bayes pour (q, π) est équivalant à l'estimateur de bayes pour $(q = 1,\ \pi' = q \times \pi)$.

Remarque 4.1.2. *Les resultats concernant la dominance, l'admissibilité et la bayésianité d'un paramètre sous l'erreur quadratique, ou ou la perte quadratique pondéré sont bien sûr abondants dans la littérature, par exemple de*

~~nombreux résultats peuvent être trouvés dans [2], ainsi que dans divers do-~~ cuments datant d'au moin 50 ans, le corollaire 1, et lemme 2, impliquent que pou arbitraire $\beta_0(X)$, l'inverse est également vraie, par conséquent, les problemes qui traitent les questions (dominance, admissibilité et bayésianité) par rapport à une une fonction de perte L_{w,β_0}, peut être déja traités par une analyse de perte L_0.

Exemple 6. *(Estimateurs linéaires affines) Ce exemple donne le resultat connu concernant la performance des estimateurs lineaires affines, sous les hypothéses :* $\mathbb{E}_\beta(X) = \beta, \quad var_\beta(X) = \sigma^2(\beta) = \sigma^2$ *voir [3] indiquent que pour estimer β sous la perte quadratique :*

i $cX + d$ *est inadmissible quand* $c > 1, \;\; c < 0,$ *ou* $c = 1, \;\; d \neq 0$

ii $cX + d$ *avec* $X \sim> \mathcal{N}(\beta, 1)$ *est admissible si et seulement si*

$$c = 1, \;\; d = 0 \quad ou \quad c \in [0, 1[, d \in \mathbb{R}$$

.

iii $cX + d$ *avec* $X \sim> \Gamma(\alpha, \beta/\alpha)$*, est admissible si et seulement si* $c = \alpha/(\alpha+1), d = 0$ *ou* $c \in]0, \alpha/(\alpha+1)], d > 0.$

le resultat d'inadmissibilité dans (i) se traduit, pour un estimateur référence arbitraire $\beta_0(X)$, au l'inadmissibilité sous la perte L_{w,β_0} de l'estimateur de la forme

$$\beta_0(X) + (1-w)(cX + d - \beta_0(X)) = w\beta_0(X) + (1-w)(cX + d)$$

pour $c > 1, \;\; c < 0,$ *ou* $c = 1, \;\; d \neq 0.$
en revanche, étant donné (ii) les estimateurs de la forme

$$w\beta_0(X) + (1-w)(cX + d) \quad avec \quad X \sim> \mathcal{N}(\beta, 1)$$

sont admissible si et seulement si $c = 1, \quad d = 0$ *ou* $c \in [0, 1[, d \in \mathbb{R}$*, en partculiére si $\beta_0(X) = X$, on obtient l'inadmissibilité de $c'X + d'$ (sous la perte L_{w,β_0}) pour $c' > 1, c' < w$, ou $c' = w, d' \neq 0$, et on obtient l'admissibilité si et seulement si $c' = 1, d' = 0$ ou $c' \in [w, 1[, d' \in \mathbb{R}$.*
On peut tirer des conclusions de (iii), avec $X \sim> \Gamma(\alpha, \beta/\alpha)$, $c'X + d'$ (sous la perte L_{w,β_0}, $\beta_0(X) = X$) est admissible (pour estimer β, si et seulement si $c' = (\alpha+w)/(\alpha+1), d' = 0$ ou $c' \in]w, (\alpha+w)/(\alpha+1)], d > 0$.

Exemple 7. *(estimateurs équivoques à risque minimum (ERM))*
On considére le modèle $X \sim f_0(x-\beta)$ avec $\mathbb{E}_0(X^2 < \infty)$, et on prend la perte (4.1) avec ($q \equiv 1$), il est bien connu que sous la perte L_0, les estimateurs invariants de position de la forme $X + c$ a un risque constante, et le choix optimal de (ERM) est donné par $\beta^*(X) = X + c^*$ (ou $c^* = -\mathbb{E}_0(X)$).
l'estimateur $\beta^*(X)$ (sous la perte L_0) est minimax, admissible et de bayes, par rapport à la la loi à priori de Haar (invariants à droit) $\pi^*(\beta) = 1$.

Remarque 4.1.3. *Les resultats précédents nous disent que sous la perte L_{w,β_0}, l'estimateur $\beta^{**}(X) = w\beta_0(X) + (1-w)\beta^*(X)$ est admissible et de bayes par rapport à ($\pi^*(\beta) = 1$), en particuliére si l'estimateur $\beta_0(X) = X + c_0$ (estimateur invariants de position), alors l'estimateur donné par $X + wc_0 + (1-w)c^*$, est égale à $\beta^{**}(X)$ si $c_0 = c^*$, et on a aussi,*

I *A un risque constant (partie donnée (3) de lemme (4.1.1)).*

II *Est optimale entre les estimateurs invariants de posiion (localisation) $X + c$.*

III *Est en vertue du théorème suivante minimax.*

4.1.2 Minimaxité

Maintenant nous capitalisons sur les resultats du lemme (4.1.1), pour présenter les resultats généreaux, et divers exemples concernant les estimateurs minimaxs sous la perte L_{w,β_0}, dans les deux cas ou l'estimateur référence $\beta_0(X)$ a un risque constante et n'a pas de risque constant.

β_0 a un risque constant sous L_0

.

Théorème 4.1.1. *Suppose que $\beta_0(X)$ a un risque constant γ sous la perte L_0, alors l'estimateur $(\beta_0(X) + (1-w)g(X))$ est minimax sous la perte L_{w,β_0}, avec un risque minimax égale $(1-w)(w\gamma + (1-w)r)$ si et seulement si, $\beta_0 + g(X)$ est minimax sous la perte L_0, avec un risque minimax égale r.*

Démonstration :

tant que $R_0(\beta, \beta_0(X)) = \gamma$ par hypothèse, on obtient de (3),

$$R_{w,\beta_0}(\beta, \beta_0(X) + (1-w)g(X)) = w(1-w)\gamma + (1-w)^2 R_0(\beta, \beta_0(X) + g(X)) \tag{4.3}$$

ce qui fallait que,

$$\sup_\beta R_{w,\beta_0}(\beta, \beta_0(X) + (1-w)g(X)) = w(1-w)\gamma + (1-w)^2 \sup_\beta R_0(\beta, \beta_0(X) + g(X)$$

et donné le resultat.

Corollaire 4.1.2. *Suppose que l'estimateur $\beta_0(X)$ a un risque constant sous la perte L_0, alors $\beta_0(X)$ est minimax sous la perte L_{w,β_0}, avec un risque minimax constant égale $(1-w)\gamma$ si et seulement si $\beta_0(X)$ est minimaxsous la perte L_0 avec un risque minimax constant γ.*

Démonstration :

(Applique le théorème precedent avec $g(X) = 0$), Nous poursuivons avec une selection d'application du théorème et corollaire précedents.

Exemple 8. *(Poison) soit $X \sim> P(\beta)$, $\beta \in]0, +\infty[$, avec la perte (perte d'information normalisé) definit par*

$$L_0(\beta, \widehat{\beta}) = \frac{(\widehat{\beta} - \beta)^2}{\beta}$$

tant que $\beta_0(X)$ est minimax sous la perte L_0, avec un risque minimax réduit, le corollaire (4.1.2) montrer que $\beta_0(X)$ et aussi minimax sous la perte L_{w,β_0}, $(w \in]0,1]$, poids $q(\beta) = \frac{1}{\beta}$), avec un risque minimax constant $(1-w)$.

Exemple 9. *(Binomiale)soit $X \sim> B_i(n, \beta)$, $\beta \in]0,1[$, avec la perte (perte d'information normalisé) definit par*

$$L_0(\beta, \widehat{\beta}) = \frac{(\widehat{\beta} - \beta)^2}{\beta(1-\beta)}$$

tant que $\beta_0(X) = \frac{X}{n}$, est minimax sous la perte L_0 avec un risque minimax constant $\frac{1}{n}$, le corollaire précedent montre que $\beta_0(X)$, est aussi minimax sous

la perte $L_{w,\beta_0}(w \in]0,1]$, poids $q(\beta) = \frac{1}{\beta(1-\beta)}$), avec un risque minimax constant $\frac{(1-w)}{n}$.
En revanche si on a la fonction de perte quadratique

$$L_0(\beta, \widehat{\beta}) = (\widehat{\beta} - \beta)^2$$

le corollaire (4.1.2) applique sur l'estimateur $\beta_0(X) = \frac{X+(\sqrt{n}/2)}{n+\sqrt{n}}$, avec un risque minimax constant, et implique la minimaxité de $\beta_0(X)$, sous la perte $L_{w,\beta_0}(w \in]0,1])$.
finallement, si on a l'estimateur $\beta_0(X) = \frac{X}{n}$ sous la perte quadratique, le théorème precedent ne s'applique pas, et on peut procéde directement en trouvant un risque de bayes (constant), et donc estimateur minimax voir [4] .

Dans l'exemple suivant, l'estimateur référence $\beta_0(X)$ n'est pas minimax, mais le théorème precedent reste appliquable.

Exemple 10. *(Binomiale negative),soit $X \sim> NB_i(\alpha, \beta)$, $\alpha > 0$ connu, $\beta \in]0,1[$ inconnu, avec $\beta = \mathbb{E}(X) = \alpha(\frac{1}{p} - 1)$ et $Var_\beta(X) = \alpha\beta(1+\beta)$ et ca densité est definit par :*

$$\mathbb{P}_\beta(X = x) = \frac{\Gamma(\alpha + x)}{x!\Gamma(\alpha)} \left(\frac{\beta}{\alpha + \beta}\right)^x \mathbb{I}_{\{0,1,\dots\}}(x)$$

On considére la fonction de perte definit par

$$L_0(\beta, \widehat{\beta}) = \frac{(\widehat{\beta} - \beta)^2}{Var_\beta(X)}$$

avec $\beta_0(X) = X$ est l'estimateur de maximum de vraissemblance, avec un risque constant égale a $\gamma = 1$, étant donné que sous la perte L_0, $\frac{X}{1+\alpha}(X - \frac{\alpha}{1+\alpha}X)$ est minimax, avec le risque minimax $r = \frac{\alpha}{1+\alpha}$ [5].
Le théorème precedent montrer que $X-(1-w)\frac{\alpha}{1+\alpha}X = (\frac{1+w\alpha}{1+\alpha})X$ est minimax sous la perte L_{w,β_0}, avec le risque minimax $(1-w)(w+(1-w)(\frac{\alpha}{1+\alpha})) = \frac{(1-w)(\alpha+w)}{(1+\alpha)}$, de plus tantque $\frac{X}{1+\alpha}$ est admissible et domine $\beta_0(X) = X$ sous la perte L_0, donc comme resultat de corollaire precedent, $(\frac{1+w\alpha}{1+\alpha})X$ est admissible

et domine $\beta_0(X) = X$ *sous la perte* $L_{w,\beta_0}(w \in]0,1])$.

Finallement, nous soulignons que d'autre resultats de d'admissibilité et de bayésianité sous la perte L_0*, tels que ceux donné par[[6]], peut être traduit via (corollaire et théorème precedent) au resultat d'admissibilité et de bayésianité sous la perte* L_{w,β_0}.

Dans l'exemple suivant, il existe plusieurs estimateurs minimaxs.

Exemple 11. *(Paramétre d'echelle à borne inferieur), soit le modèle* $X \sim> \Gamma(\alpha, \beta)$*, (densité proportionnelle à* $x^{\alpha-1}e^{-x/\beta}$*), on considére l'estimation d'une borne inférieur* β*,* $\beta \geq a > 0$ *(a connu), sous la perte (1.1) et* $q(\beta) = \beta^{-2}$ *soit*

$$h_\alpha(z) = \frac{z^{\alpha+1}e^{-z}}{\displaystyle\int_0^z t^{\alpha+1}e^{-t}dt}$$

[7] a montrer que sous la perte L_0*, l'estimateur de bayes* $\beta_a^*(X)$ *par rapport à la loi a priori* $\pi_a^*(\beta) = \beta^{-1}\mathbb{I}_{]a,\infty[}(\beta)$*, definit par*

$$\beta_a^*(X) = \frac{X}{\alpha+1}\left(1 + h_{\alpha+2}(\frac{X}{a})\right)$$

est admissible, minimax de β*, avec le risque minimax* $r = \frac{1}{\alpha+1}$ *voir [8].*
Pour un esstimateur de référence plus génerale $\beta_0(X)$ *on à montrer dans les resultats précedent que* $\beta_a^{**}(X) = w\beta_0(X) + (1-w)\beta_a^*(X)$*, est un estimateur generalisé de bayes admissible, par rapport à* π_a^*.
Si l'estimateur $\beta_0(X)$ *est un estimateur (ERM) definit par* $\frac{X}{\alpha+1}$*, donc comme resultat de théorème precedent, l'estimateur* $\beta_a^{**}(X)$ *definit par,*

$$\frac{X}{1+\alpha} + (1-w)\beta_a^*(X) = \frac{X}{\alpha+1}(1 + (1-w)h_{\alpha+2}(\frac{X}{a}))$$

n'est pas seulement admissible, mais aussi minimax avec le risque minimax $(\frac{1-w}{\alpha+1})$*(tant que* $\gamma = R(\beta, \frac{X}{\alpha+1}) = \frac{1}{\alpha+1}$*).*

Remarque 4.1.4. *On remarque qu'il existe de nombreux estimateurs minimax sous la perte* L_0 *et donc beaucoup sous la perte* L_{w,β_0}*, quand* $\beta_0(X)$ *est un (ERM), des developements similaires pour les paramètres d'echelle à borne inférieur sous la perte* L_0*, qui peuvent être appliquées aux pertes* L_{w,β_0}*, voir [9].*

Remarque 4.1.5. *En ce qui concerne l'exemple précédent, si l'estimateur de référence β_0 prend des valeurs en dehors de l'espace des paramètres $[a, \infty[$, il n'y a aucune garantie que $\mathbb{P}(\beta_a^{**}(X) \in [a, \infty[) = 1$, en effet si$\beta_0(X) = \frac{X}{\alpha+1}$ alors,*

$$\lim_{x \mapsto 0} \beta_a^{**}(x) = (1-w) \lim_{x \mapsto 0} \beta_a^{*}(x) = (1-w) \frac{(\alpha+2)}{(\alpha+1)} a$$

donc $\lim_{x \mapsto 0} \beta_a^{*}(x) = \dfrac{(\alpha+2)}{(\alpha+1)} a$ [7], on peut conclut que $\beta_a^{**}(x)$ prendrait des valeurs dans l'espace des paramètres si et seulement si, $w \leq \frac{1}{\alpha+2}$, ce phénomène n'est pas limité à cet exemple et se produit, puisque la fonction de perte refléte la proximité de $\widehat{\beta}$ à $\beta_0(X)$, plutôt que le critère $\widehat{\beta} \in \Theta$.
C'est motivation pour étudier, et présenter des exemples et avoir des résultats disponible pour un estimateur généralisée $\beta_0(X)$, qui n'est pas necessairement de moindre carré et qui prent des valeurs uniquement dans l'espace des paramètres.
Les propriétes d'admissibilité et bayesianité de $\beta_a^{**}(x)$ dans l'exemple précedent, reste verifié dans le cas ou $\beta_0(X)$ est la version tranqué de (ERM), definit par, $\beta_0(X) = \max(a, \frac{X}{\alpha+1})$, ou dans le cas de l'estimateur du maximum de vraissemblance definit par $\beta_0(X) = \max(a, X/\alpha)$.

$\beta_0(X)$ dans un cas plus géneralisé

Malgré tout les différentes applications d'estimation minimax du théorème et corollaire donné dans les exemples précedents, elles sont limitées aux cas où l'estimateur référence $\beta_0(X)$ présente un risque constant, en revanche le développement suivant se rapporte à des situations où $\beta_0(X)$, n'a pas de risque constant, mais posséde plutôt des propriétés de risque nous permettant d'exploiter les criteres pour la minimaxité, voir ([2] section 5.1).

Lemme 4.1.3. *Si β_π est un estimateur de bayes par rapport à la loi a priori propre π, et $S_\pi = \left\{ \beta \in \Theta,\ \sup_{\beta} R(\beta, \beta_\pi),\ \beta \in \Theta = R(\beta, \beta_\pi) \right\}$, alors β_π est minimax, quand $\mathbb{P}_\pi(\beta \in S_\pi) = 1$.*

Pour poursuivre, nous définissons sous la perte L_{w,β_0} et l'estimateur référence $\beta_0(X)$,

$$S_{w,\beta_0}(\widehat{\beta}) = \left\{ \beta \in \Theta,\ \sup_{\beta} R_{w,\beta_0}(\beta, \widehat{\beta}),\ \beta \in \Theta = R_{w,\beta_0}(\beta, \widehat{\beta}) \right\}$$

(c-à-d, l'ensemble des β telque $\widehat{\beta}(X)$ atteint sont risque maximum sous la perte L_{w,β_0}, on note $S_0(\widehat{\beta})$ au lieu de $S_{0,\beta_0}(\widehat{\beta})$.

Lemme 4.1.4. *(★)suppose que $\beta_\pi(X) = \beta_0(X) + g_\pi(X)$ est un estimateur de bayes unique sous la perte L_0 pour la loi a priori propre π, suppose aussi que $\mathbb{P}_\pi(\beta \in S_0(\beta_\pi)) = \mathbb{P}_\pi(\beta \in S_0(\beta_0)) = 1$, alors l'estimateur $\beta_\pi(X) = \beta_0(X) + g_\pi(X)$, est minimax et unique sous la perte[1] L_{w,β_0}.*

Démonstration :

Étant donné que sous la perte L_{w,β_0}, l'estimateur de bayes $\beta'_\pi(X) = \beta_0(X) + (1-w)g_\pi(X)$ est unique par rapport à π, comme resultat de (4.2), il suffit de montrer que

$$\mathbb{P}_\pi(\beta \in S_{w,\beta_0}(\beta'_\pi)) \tag{4.4}$$

puisque nous pourrions alors appliquer le lemme (4.1.3) de cette partie, pour déduir le résultat , mais (4.4) suit par le lemme (4.1.1) partie (3), étant donné que par la représentation du risque $R_{w,\beta_0}(\beta, \beta'_\pi)$ et par hypothèse, que les deux risques $R_0(\beta, \beta_0)$ et $R_0(\beta, \beta'_\pi)$, attaint leur sup avec la même probabilité 1.

4.2 Sur l'estimation de bayes sous certaines classes de fonction de perte équilibrée géneralisée

Dans cette section on va generaliser la fonction de perte équilibrée de la section précédent, on ne va pas donc se restreindre au cas de la fonction de

1. On note que ces hypothéses implique que $\beta_0(X) + g_\pi(X)$ est minimax sous la perte L_0

perte quadratique, on pose alors

$$L_{\rho,w,\beta_0}(\beta,\widehat{\beta}) = w\rho(\beta_0,\widehat{\beta}) + (1-w)\rho(\beta,\widehat{\beta}) \tag{4.5}$$

ainsi que la version pondéré $q(\beta)L_{\rho,w,\beta_0}$, avec $q(\cdot)$soit une fonction de poids positive, et $\rho(\beta,\widehat{\beta})$ est une fonction de perte arbitraire, β_0 est l'estimateur référence "cible" de β, et $w \in [0,1[$.
Un développement générale en ce qui concerne les estimateurs bayésiennes pour le cas déséquilibré ($w = 0$), des représentations sont donné pour différents choix de ρ, par exemple la perte absolu, entropie, linex, intrinsèque et géneralisation du perte quadratique.
Cette section concerne l'estimation ponctuelle bayésienne du paramètre inconnue β, avec le même modèle de section précedent sous la perte (4.5), et l'estimateur référence $\beta_0(X)$ comme un estimateur de maximum de vraicemblanse (MV) ou de moindre carré (MC), ou d'autre estimateurs.

4.2.1 Estimation de bayes sous la perte équilibrée

On considére l'estimation de bayes sous la perte (4.5) avec $w = 0$ (cas déséquilibré), on le note comme la section précédente L_0, on va montrer comment l'estimateur de bayes de β sous la perte L_{ρ,w,β_0} peut être derivé ou exprimé en termes d'un estimateur de bayes sous β sous la perte L_0, sous l'hypothèse que la perte $\rho(\widehat{\beta},\beta_0) < \infty$ pour au moin un $\widehat{\beta} \neq \beta_0$.

Lemme 4.2.1. *pour l'estimation de β sous la fonction de perte (4.5), avec la loi a priori $\pi(\beta)$, l'estimateur de bayes $\beta_{\pi,w}(X)$ correspond à la solution de bayes $\beta^*(X)$ par raport a $(\pi^*(\beta/X), L_0)$ pour tout x, alors que*

$$\pi^*(\beta/X) = w\mathbb{I}_{\beta_0(X)}(\beta) + (1-w)\pi(\beta/X)$$

c-à-d, un mélange d'une masse ponctuelle à $\beta_0(X)$ et la loi a posteriori $\pi(\beta/X)$.

Démonstration :

Indiquer $\mu_*(\cdot)$ et $\nu_*(\cdot)$, deux mesures dominants respectivement de $\pi(\beta/X)$ et $\pi^*(\beta/X)$, on a par définition de $\beta_{\pi,w}(X)$, $\beta^*(X)$ et L_0

$$\begin{aligned}
\beta_{\pi,w}(X) &= \arg\min_{\widehat{\beta}} \int_\Theta \left\{ w\rho(\beta_0,\widehat{\beta}) + (1-w)\rho(\beta,\widehat{\beta}) \right\} \pi(\beta/X) d\mu_*(\beta) \\
&= \arg\min_{\widehat{\beta}} \int_{\Theta \cup \{\beta_0(X)\}} L_0(\beta,\widehat{\beta}) \pi^*(\beta/X) d\nu_*(\beta) = \beta^*(X)
\end{aligned}$$

Remarque 4.2.1. *Nous avertissons que $\beta^*(X)$ n'est pas un estimateur de bayes sous la perte L_0, parce que $\pi^*(\beta/X)$ n'est pas la loi a posteriori associée avec π, tandis que $\beta_{\pi,w}(X)$ est en effet un estimateur de bayes sous la perte L_{ρ,w,β_0} et la loi a priori π.*

Remarque 4.2.2. *Le lemme precédent s'applique aussi dans le cas de la fonction de perte pondéré definit par*

$$L^q_{\rho,w,\beta_0}(\beta,\widehat{\beta}) = q(\beta) L_{\rho,w,\beta_0}(\beta,\widehat{\beta})$$

avec $q(\beta)$ une fonction de poids positive, en effet, il est facile de voir que l'estimateur de bayes de β sous la perte L^q_{ρ,w,β_0}, avec la loi a priori π_0 est équivalant à l'estimateur de bayes de β, sous la perte L_{ρ,w,β_0}, avec la loi a priori $\pi = q \times \pi_0$.

4.2.2 Quelques applications

Une généralistion de la perte quadratique

Le choix de ρ est le suivant

$$\rho(\beta,\widehat{\beta}) = \tau(\beta)(\beta - \widehat{\beta})^2 \quad \tau(\cdot) > 0$$

alors (4.5) devient

$$w\tau(\beta_0)(\beta-\beta_0)^2 + (1-w)\tau(\beta)(\beta-\widehat{\beta})^2$$

on a l'estimateur de bayes sous L_0, avec la loi a priori π, donnée par

$$\delta_{0,\pi}(x) = \frac{\mathbb{E}_\pi(\beta\tau(\beta)/x)}{\mathbb{E}_\pi(\tau(\beta)/x)}$$

Sous la condition $\mathbb{E}_\pi(\beta^i\tau(\beta)/x) < \infty,\ i = 0, 1$ por tout x, le lemme précedent donne,

$$\beta_{w,\pi}(x) = \frac{\mathbb{E}_{\pi^*}(\beta\tau(\beta)/x)}{\mathbb{E}_{\pi^*}(\tau(\beta)/x)} = \frac{w\beta_0(x)\tau(\beta_0(x)) + (1-w)\mathbb{E}_\pi(\beta\tau(\beta)/x)}{w\tau(\beta_0(x)) + (1-w)\mathbb{E}_\pi(\tau(\beta)/x)}$$

le cas ou ($\tau(\beta) = 1,\ \beta_0$ est un estimateur de (MC), est le cas de la fonction de perte quadratique équilibré et ($\beta_{w,\pi}(x) = w\beta_0(x) + (1-w))\mathbb{E}_\pi(\beta/x)$), comme présentée par [38], et [11], étudier le cas ou $\tau(\beta) = 1$ mais $\beta_0(x)$ est un choix arbitraire, ainsi que la version pondéré de cas précedent (quadratique), a été étudier par [17] par rapport aux critère de la théorie de decision, tels que la bayésianité, dominance, admissibilité et minimaxité, sachant que [14] et aussi étudier le cas (quadratique avec $\tau(\beta) = 1$).

Le perte d'entropie équilibré

Le choix de ρ est le suivant

$$\rho(\beta, \widehat{\beta}) = \frac{\beta}{\widehat{\beta}} - \log(\frac{\beta}{\widehat{\beta}}) - 1$$

alors (4.5) devient

$$w\left(\frac{\beta_0}{\widehat{\beta}} - \log(\frac{\beta_0}{\widehat{\beta}}) - 1\right) + (1-w)\left(\frac{\beta}{\widehat{\beta}} - \log(\frac{\beta}{\widehat{\beta}}) - 1\right) \tag{4.6}$$

la solution de bayes sous la perte L_0, avec la loi a priori π, est donné par

$$\beta_{0,\pi}(x) = \mathbb{E}_\pi(\beta/x)$$

sous la condition $\mathbb{E}_\pi(\beta/x) < \infty$ et $\mathbb{E}_\pi((\log(\beta)/x) < \infty,\ \forall x)$, le lemme precédent donne

$$\beta_{w,\pi}(x) = \mathbb{E}_{\pi^*}(\beta/x) = w\beta_0(x) + (1-w)\mathbb{E}_\pi(\beta/x)$$

comme exemple particuliére, en prend le modèle $X \sim> \Gamma(\alpha, \beta),\ \alpha$ est connu, $\beta > 0$, (densité proportionnelle à $x^{\alpha-1}e^{-x/\beta}$), on considére l'estimation de β

sous la perte (4.6), avec $\beta_0(x) = \beta_{MV}(x) = \frac{x}{\alpha}$, et la loi a priori conjugue π telque, $\beta^{-1} \sim> \Gamma(\gamma, \frac{1}{h}),\ \gamma > -\alpha,\ h \geq 0$.

La loi a posteriori de β^{-1}, suit la loi $\Gamma(\alpha+\gamma, \frac{1}{h+x})$, alors l'estimateur de bayes unique de β, sous la perte d'entropie équilibré (4.6), definit comme suit

$$\beta_{w,\pi}(x) = w\beta_0(x) + (1-w)\frac{h+x}{\alpha+\gamma-1}$$

Le perte de Stein

Ce perte est definit par

$$\rho(\beta, \widehat{\beta}) = (\frac{\widehat{\beta}}{\beta})^{\iota} - \iota \log(\frac{\widehat{\beta}}{\beta}) - 1 \ \iota \neq 0$$

cette classe de perte est une géneralisation du perte d'entropie équilibré ($\iota = 1$), ainsi que la perte de Stein dans le cas ($\iota = -1$), illustre bien le lemme précedent.

Sous la perte L_0, on a la solution de bayes $\beta_{0,\pi}(x) = (\mathbb{E}_{\pi}(\frac{1}{\beta^{\iota}}/x))^{-\frac{1}{\iota}}$, avec la condition que le risque soit finie, le lemme précedent montrer que

$$\beta_{w,\pi}(x) = (\mathbb{E}_{\pi^*}(\frac{1}{\beta^{\iota}}/x))^{-\frac{1}{\iota}} = \left\{\frac{w}{(\beta_0(x))^{\iota}} + (1-w)\mathbb{E}_{\pi}(\frac{1}{\beta^{\iota}}/x)\right\}^{-\frac{1}{\iota}}$$

Le perte $((\beta/\widehat{\beta}) - 1)^2$

La solution de bayes sous la perte L_0 avec la loi a priori π definit sur $]0, \infty[$, est definit par

$$\beta_{0,\pi}(x) = \frac{\mathbb{E}_{\pi}(\beta^2/x)}{\mathbb{E}_{\pi}(\beta/x)}$$

avec la condition que $\mathbb{E}_{\pi}(\beta^2/x) < \infty,\ \forall x$, le lemme précedent montrer que

$$\beta_{w,\pi}(x) = \frac{\mathbb{E}_{\pi^*}(\beta^2/x)}{\mathbb{E}_{\pi^*}(\beta/x)} = \frac{w\beta_0^2 + (1-w)\mathbb{E}_{\pi}(\beta^2/x)}{w\beta_0 + (1-w)\mathbb{E}_{\pi}(\beta/x)}$$

dans [29] Rodrigues et Zellner on 1994 étudié l'estimateur de bayes sous cette fonction de perte mais avec ($\beta_0(X)$ est l'estimateur de moindre carré) et sous le modèle exponentiel.

La fonction de perte de linex

Cette fonction de perte est definit par

$$\rho(\beta, \widehat{\beta}) = e^{a(\widehat{\beta}-\beta)} - a(\widehat{\beta} - \beta) - 1, \quad a \neq 0$$

la solution de bayes sous la perte L_0 est donné par

$$\beta_{0,\pi}(x) = -\frac{1}{a} \log \mathbb{E}_\pi(e^{-a\beta}/x)$$

avec la condition que le risque soit finie, par conséquent, en utilisant le lemme précedent donc on obtient

$$\begin{aligned}
\beta_{w,\pi}(x) &= -\frac{1}{a} \log \mathbb{E}_{\pi^*}(e^{-a\beta}/x) \\
&= -\frac{1}{a} \log(we^{-a\beta_0(x)} + (1-w)\mathbb{E}_\pi(e^{-a\beta}/x)) \\
&= -\frac{1}{a} \log(we^{-a\beta_0(x)} + (1-w)e^{-a\beta_{0,\pi}(x)})
\end{aligned}$$

Remarque 4.2.2.1. *Pour des exemples concerne la perte linex voir [39] [27].*

La fonction de perte intrinsèque équilibrée

On traivaille sous le modèle $X/\beta \sim> f(X/\beta)$ avec la fonction de perte

$$\rho(\beta, \widehat{\beta}) = d(f(\cdot/\beta), f(\cdot/\widehat{\beta}))$$

c-à-d la distance entre deux fonctions, donc la perte intrinsèque équilibrée est definit par

$$wd(f(\cdot/\beta_0), f(\cdot/\widehat{\beta})) + (1-w)d(f(\cdot/\beta), f(\cdot/\widehat{\beta})) \tag{4.7}$$

on a deux condidats intérêt pour d, sont la distance de Hillinger et la distance de Kulback-Leibler, donc on obtient

✠ Le perte équilibré de Kullback-Leibler

$$L^{KL}_{w,\beta_0}(\beta,\widehat{\beta}) = w\mathbb{E}_{\beta_0}\left[\ln\frac{f(X/\beta_0)}{f(X/\beta)}\right] + (1-w)\mathbb{E}_{\beta}\left[\ln\frac{f(X/\beta)}{f(X/\widehat{\beta})}\right] \quad (4.8)$$

✠ Le perte équilibré de Hillinger

$$L^{H}_{w,\beta_0}(\beta,\widehat{\beta}) = w\frac{1}{2}\mathbb{E}_{\beta_0}\left[\left(\sqrt{\frac{f(X/\widehat{\beta})}{f(X/\beta_0)}}-1\right)^2\right] - (1-w)\frac{1}{2}\mathbb{E}_{\beta}\left[\left(\sqrt{\frac{f(X/\widehat{\beta})}{f(X/\beta)}}-1\right)^2\right] \quad (4.9)$$

ou de maniére équivalente

$$L^{H}_{w,\beta_0}(\beta,\widehat{\beta}) = 1 - w\mathbb{E}_{\beta_0}\left[\sqrt{\frac{f(X/\widehat{\beta})}{f(X/\beta_0)}}\right] - (1-w)\mathbb{E}_{\beta}\left[\sqrt{\frac{f(X/\widehat{\beta})}{f(X/\beta)}}\right] \quad (4.10)$$

Par conséquent, avec les représentation bayésienne commeNT dans le cas déséquilibré ($w = 0$) [2] [30] et avec le lemme précedent, on pêut construire representation bayésienne pour L^H, L^{KL}.

On prend comme exemple la famille exponentielle de paramètre naturelle des destributions avec des densités

$$f(x,\beta) = e^{\beta T(x) - \psi(\beta)}h(x)$$

(par rapport a la la mesure σ-finie ν definit sur χ), avec le paramètre naturelle inconnu β, donc la fonction de perte de Kullback-Lielber devient

$$w\left[(\beta_0-\widehat{\beta})\psi'(\beta_0)+\psi(\widehat{\beta})-\psi(\beta_0)\right] + (1-w)\left[(\beta-\widehat{\beta})\psi'(\beta)+\psi(\widehat{\beta})-\psi(\beta)\right] \quad (4.11)$$

de plus, sous la perte L_0 ($w = 0$ dans (4.10)) avec la loi a priori π, l'estimateur de bayes $\beta_{0,\pi}(x)$ est donné comme solution de $\psi'(\beta_{0,\pi}(x)) = \mathbb{E}_\pi[\psi'(\beta)/x]$, alors comme resultat l'estimateur de bayes $\beta_{w,\pi}(x)$ admet la representation suivant

$$\psi'(\beta_{w,\pi}(x)) = w\psi'(\beta_0(x)) + (1-w)\psi'(\beta_{0,\pi}(x))$$

En particuliére si on a $\mathbb{E}_\beta[T(X)] = \psi'(\beta)$, pour $(X \sim> \mathcal{N}(\beta, 1), T(X) = X)$, on obtient

$$\beta_{w,\pi}(x) = w\beta_0(x) + (1-w)\beta_{0,\pi}(x)$$

Un combainaison lineaire convexe de β_0 et l'estimateur de bayes $\beta_{0,\pi}(x))$, également, pour la famille exponentielle naturelle des distributions comme ci-dissus, la fonction de perte équilibré de Hillinger (4.9) devient

$$1 - we^{\left\{\psi(\frac{\beta_0+\widehat{\beta}}{2}) - \frac{\psi(\beta_0)+\psi(\widehat{\beta})}{2}\right\}} - (1-w)e^{\left\{\psi(\frac{\beta+\widehat{\beta}}{2}) - \frac{\psi(\beta)+\psi(\widehat{\beta})}{2}\right\}} \tag{4.12}$$

sous le modèle $\mathcal{N}(\beta, \infty)$ la perte (4.11) devient

$$1 - we^{-\frac{(\beta_0-\widehat{\beta})^2}{8}} - (1-w)e^{-\frac{(\beta-\widehat{\beta})^2}{8}}$$

est qui rêfléte la perte normal équilibré borné, voir [32] pou le cas $w = 0$.

Remarque 4.2.3. *Le perte d'entropie équilibré est équivalant au perte équilibré de Kullback-Leibler sous le modèle $\Gamma(\alpha, \beta)$, avec $\alpha > 0$ connu, et le paramtre inconnue β, voir aussi [28] [30].*

Le perte équilibré absolue ou L^1

Cette fonction de perte est definit par

$$\rho(\beta, \widehat{\beta}) = |\widehat{\beta} - \beta|$$

donc la fonction de perte équilibrée absolue est definit par

$$L_{w,\beta_0}(\beta, \widehat{\beta}) = w|\widehat{\beta} - \beta_0| + (1-w)|\beta - \widehat{\beta}| \tag{4.13}$$

onà la solution de bayes associé à L_0, est la mediane a posteriorie (voir chapitre 1), le lemme précedent donne que l'estimateur de bayes $\beta_{w,\pi}(x)$ de β, par rapport au perte (4.13) et la loi a priori π, est la medianne de $\pi^*(\beta/x)$, pour tout x.

Maintenant, on va travailler sous le modèle $f(\cdot/\beta)$, donc avec la loi a priori π, on obtient un unique medianne a posteriori definit par

$$\begin{aligned}
\beta_{w,\pi}(x) &= F^{-1}_{\beta/x}\left(\frac{1-2w}{2(1-w)}\right)\mathbb{I}_{[0,\frac{1}{2}-w[}(\alpha(x)) + \beta_0(x)\mathbb{I}_{[\frac{1}{2}-w,\frac{1}{2}[}(\alpha(x)) \\
&+ F^{-1}_{\beta/x}\left(\frac{1}{2(1-w)}\right)\mathbb{I}_{[\frac{1}{2},1[}(\alpha(x)),\ \forall x \qquad (4.14)
\end{aligned}$$

avec $\alpha(x) = (1-w)\mathbb{P}_\pi(\beta \leq \beta_0(x)|x)$, et $F^{-1}_{\beta/x}$ est l'inverse de la fonction de repartition a posteriori, on remarque que si $w \geq \frac{1}{2}$ alors $\beta_{w,\pi}(x) = \beta_0(x)$ (c-à-d la masse ponctuelle de $\beta_0(x)$ a une probabilité assez grand pour forcer la mediane a $\beta_0(x)$).

Maintenant, nous donnons quelques application de (4.13).

✠ Le cas de la loi normal avec moyenne positive :

Soit $X/\beta \sim> \mathcal{N}(\beta, \sigma^2)$ avec $\beta > 0$ inconnu et σ^2 connu, et soit la fonction de repartition F_0, la théorie de decision concernant l'estimation de β a un longe histoire, voir [19].

Prenez $\beta_0(X) = \beta_{MV}(X) = \max(0, x)$ et $w < \frac{1}{2}$ dans (4.12), on considére l'estimateur $\beta_{w,\pi}(X)$ associé à la loi a priori (flat) $\pi(\beta) = \mathbb{I}_{]0,\infty[}(\beta)$, pour simplifier (4.14) on a

1. $\mathbb{P}_\pi(\beta \geq y|x) = \dfrac{F_0(x-y)}{F_0(x)}$ $\forall y \geq 0$.
2. $F^{-1}_{\beta|x}(z) = x - F_0^{-1}((1-z)F_0(x))$
3. $\alpha(x) = (1-w)\mathbb{P}_\pi(\beta \leq \beta_0(x)|x) = (1-w)(1-\frac{1}{2F_0(x)})\mathbb{I}_{]0,\infty[}(x)$

donc, tantque $\alpha(x) < \frac{1}{2}$ et $x \leq F_0^{-1}(1-w)$ on obtient,

$$\begin{aligned}
\beta_{w,\pi}(x) &= F^{-1}_{\beta|x}(1-\frac{1}{2(1-w)})\mathbb{I}_{]-\infty,F_0^{-1}(1-w)[}(x) + \beta_0(x)\mathbb{I}_{]F_0^{-1}(1-w),\infty[}(x) \\
&= \left\{x - F_0^{-1}(\frac{F_0(x)}{2(1-w)})\right\}\mathbb{I}_{]-\infty,F_0^{-1}(1-w)[}(x) + x\mathbb{I}_{]F_0^{-1}(1-w),\infty[}(x) \\
&= x - \left\{F_0^{-1}(\frac{F_0(x)}{2(1-w)})\right\}\mathbb{I}_{]-\infty,F_0^{-1}(1-w)[}(x)
\end{aligned}$$

~~Notons que cette expression reste verifié dans le cas desequilibre (w=0),~~ tandis qu'un développement similaire se présenter les modéles $X|\beta \sim> f_0(x-\beta)$ avec $\beta \geq 0$ et f_0 connue et symétrique.

✠ Soit $X|\beta \sim> \mathcal{N}(\beta, \sigma^2)$ avec $\beta \in \mathbb{R}$, σ^2 connue, avec la fonction de répartition definit par $\Phi(\frac{x-\beta}{\sigma})$, on a sous la loi a priori $\beta \sim> \pi(\beta) = \mathcal{N}(\mu, \tau^2)$, (μ, τ^2) connue), et la perte (4.12) avec $\beta_0(X) = X$, $w < \frac{1}{2}$, et tant que la loi a posteriori $X|\beta \sim> \mathcal{N}(\frac{\tau^2}{\tau^2+\sigma^2}x + \frac{\sigma^2}{\tau^2+\sigma^2}\mu, \frac{\tau^2\sigma^2}{\tau^2+\sigma^2})$ on a

$$\alpha(x) = (1-w)\mathbb{P}_\pi(\beta \leq x|x) = (1-w)\Phi(d(x-\mu))$$

avec

$$d = \frac{\sigma}{\tau\sqrt{(\tau^2+\sigma^2)}}$$

et

$$F^{-1}_{\beta|x}(\alpha) = \frac{\tau^2}{\tau^2+\sigma^2}x + \frac{\sigma^2}{\tau^2+\sigma^2}\mu + \sqrt{\frac{\tau^2\sigma^2}{\tau^2+\sigma^2}}\Phi^{-1}(\alpha)$$

donc de (4.14), on obtient l'estimateur lineaire par morceaux suivant

$$\beta_{w,\pi}(x)\begin{cases} \frac{x\tau^2+\mu\sigma^2}{\tau^2+\sigma^2} + \sqrt{\frac{\tau^2\sigma^2}{\tau^2+\sigma^2}}\Phi^{-1}(\frac{1-2w}{2(1-w)} & si \quad x \leq \mu + \frac{1}{d}\Phi^{-1}(\frac{1-2w}{2(1-w)} \\ x & si \quad x \in]\mu + \frac{1}{d}\Phi^{-1}(\frac{1-2w}{2(1-w)}, \mu + \frac{1}{d}\Phi^{-1}(\frac{1}{2(1-w)}[\\ \frac{x\tau^2+\mu\sigma^2}{\tau^2+\sigma^2} + \sqrt{\frac{\tau^2\sigma^2}{\tau^2+\sigma^2}}\Phi^{-1}(\frac{1}{2(1-w)} & si \quad x \geq \mu + \frac{1}{d}\Phi^{-1}(\frac{1}{2(1-w)} \end{cases}$$

4.3 La fonction de perte équilibrée prolongée

Dans cette section on va présenter une fonction de perte géneralisé, [Shalabch, H. Toutenburg et C.Heuman le 5 decembre 2007][33], sont prolongent et présentent une fonction de perte générale appelée la fonction de perte équilibrée prolongée, dans lequelle la fonction de perte prédictive proposer par Zellner 1994 [38], est une cas particuliéres.

On travaille sous le modèle de regression lineaire $Y = \beta X + \epsilon$ avec $\widehat{\beta}$ est l'estimateur de β. donc la fonction de perte quadratique qui refléte la qualité de l'ajustement du modéle est definit par

$$(X\widehat{\beta}-Y)^t(X\widehat{\beta}-Y) \tag{4.15}$$

on a $\widehat{Y}=X\widehat{\beta}$ (prévesion de Y), de même la précision de l'estimateur $\widehat{\beta}$ est mesurer par la fonction de perte pondéré definit par

$$(\widehat{\beta}-\beta)^t X^t X(\widehat{\beta}-\beta) \tag{4.16}$$

[Zellner 1994][38] a considéré a la fois les deux critéres et proposé la fonction de perte pondérée definit par

$$BL(\widehat{\beta})=w(X\widehat{\beta}-Y)^t(X\widehat{\beta}-Y)+(1-w)(\widehat{\beta}-\beta)^t X^t X(\widehat{\beta}-\beta) \tag{4.17}$$

avec $w\in[0,1]$ (qui fournit le poids attribué a la qualité d'ajustement du modèle). Si on considére la prédiction de l'echantillon, alors le prédicteur $\widehat{Y}=X\widehat{\beta}$ est utilisé pour prédire la valeur actuel Y ainsi que la valeur moyenne $\mathbb{E}(Y)$, dans certains cas, il peut être souhaitable d'envisager la prédiction simultanée de Y et $\mathbb{E}(Y)$, et definit la fonction cible suivante

$$T=\lambda Y+(1-\lambda)\mathbb{E}(Y) \quad \lambda\in[0,1] \tag{4.18}$$

si on utilisons le prédicteur $X\widehat{\beta}$ pour la prévision simultanée des valeurs actuel et moyenne de Y à travers la fonction cible, on obtient la fonction de perte prédictive definit par

$$\begin{aligned} PL(\widehat{\beta}) &= (X\widehat{\beta}-T)^t(X\widehat{\beta}-T) \qquad (4.19)\\ &= \lambda^2(X\widehat{\beta}-Y)^t(X\widehat{\beta}-Y)+(1-\lambda)^2(\widehat{\beta}-\beta)^t X^t X(\widehat{\beta}-\beta)\\ &+ 2\lambda(1-w)(X\widehat{\beta}-T)^t X(X\widehat{\beta}-T) \end{aligned}$$

si on prend la fonction de perte équilibrée et la fonction de perte prédicive [Shalabch 2007][33] proposons la fonction de perte pondérée (géneralisée)

$$\begin{aligned} WL(\widehat{\beta}) &= \lambda_1(X\widehat{\beta}-Y)^t(X\widehat{\beta}-Y)+\lambda_2(\widehat{\beta}-\beta)^tX^tX(\widehat{\beta}-\beta) \quad (4.20)\\ &+ (1-\lambda_1-\lambda_2)(X\widehat{\beta}-Y)^tX(\widehat{\beta}-\beta) \end{aligned}$$

avec $\lambda_1 \in [0,1]$ $\lambda_2 \in [0,1]$, cette fonction de perte prolongée englobe la perte (4.15)(4.16)(4.17)(4.18), comme ces cas particuliéres, elle est donc assez génerale et suffisamment souple.

Chapitre 5

Applications

5.1 L'estimateur de bayes minimax

On va donne une application du lemme ($\bigstar$), a la distribution normal dans le cas $X \sim> \mathcal{N}(\beta, \sigma^2)$ ou $\beta \in [-m, m]$, et on prend l'estimateur référence $\beta_0(X)$ comme l'estimateur de maximum de vraissemblance $\beta_{MV}(X) = (m \wedge X)sgn(X)$.

Incidement, [Marchand et Perron (2001)][24] fournit divers estimateurs de la forme $(\beta_0(X) + g(X))$ dominants $\beta_{MV}(X)$, sous la perte quadratique L_0, qui se traduisent par divers estimateurs de la forme $(\beta_0(X) + (1 - w)g(X))$ dominants $\beta_{MV}(X)$, sous la perte L_{w,β_0} avec $\beta_0(X) = \beta_{MV}(X)$.

Un exemple est donner par l'estimateur minimax du théorème suivant, quand $m \leq \sigma$, en revenant a l'estimation minimax de β, a l'aide du travaille de [Casella et Strawderman 1981][9] qui donne des conditions a l'estimateurs de bayes (on le note $\beta_{BU}(X)$), par rapport a la loi a priori bornée uniforme π_{BU} sur $\{-m, m\}$ soit minimax.

Théorème 5.1.1. *Pour l'estimaton de $\beta \in [-m, m]$ sout la perte definit dans (4.1), avec $q(\beta) = 1$ et $\beta_0 = \beta_{MV}$, sous le modèle $X \sim> \mathcal{N}(\beta, \sigma^2)$, avec la densité definit par $\frac{1}{\sigma}\Phi(\frac{x-\beta}{\sigma})$, et fonction de répartition definit par $\Phi(\frac{x-\beta}{\sigma})$,* *ou σ est connue, alors, l'estimateur*

$$(1-w)\beta_{BU}(X) + w\beta_{MV}(X)$$

est minimax et unique, tantque $m \leq m_0\sigma$, avec $m_0 \simeq 1,0567$, et

$$\beta_{BU}(X) = \frac{m}{\sigma}\tanh(mx/\sigma^2).$$

Démonstration :

On utilse le lemme ($\bigstar$) et lemme ce qui propose plus tard dans cette section, on a d'aprés [Casella et Strawderman 1981][9],

$$\mathbb{P}_\pi(\beta \in S_0(\beta_\pi)) = 1 \ \ pour \ \ \pi = \pi_{BU} \ \ et \ \ m \leq m_0\sigma \ \ \blacksquare$$

dans le lemme suivant on va montrer que la valeur maximum du risque $R(\beta, \beta_{MV}(X))$ nécessairement atteint a $\beta = (+-)m$, n'importe quand $m \leq \sqrt{2}\sigma$ puisque $m_0 \leq \sqrt{2}$, en d'autre terme la condition ($\blacksquare$) est verifié, donc le resultat suit directement comme une conséquence du lemme ($\bigstar$).

Lemme 5.1.1. *Pour $X \sim> \mathcal{N}(\beta, \sigma^2)$, avec $|\beta| \leq m$, le risque maximum sous la perte quadratique de l'estimateur $\beta_{MV}(X)$ est atteinte sur la frontiére $\{-m, m\}$, n'importe quand $m \leq \sqrt{2}\sigma$.*

Démonstration :

sons perte de généralité, soit $\sigma = 1$, d'abord on va montrer que pour tout $m \in]0, \sqrt{2}[$

$$(A) \qquad \sup_{\beta \in [-m,m]} \{R_0(\beta, \beta_{MV}(X))\} = \max\{R_0(0, \beta_{MV}(X)) R_0(m, \beta_{MV}(X))\}$$

puis on montrer que

$$(B) \qquad R_0(0, \beta_{MV}(X)) \leq R_0(m, \beta_{MV}(X)) \ \ ssi \ \ m \leq m_1 \simeq 1.96422$$

c'est clair que (A) et (B) (qui sont tous deux d'intérêts indépendants), pour établir (A), on procéde directement en montrant que pour $m \leq \sqrt{2}$.

$$\frac{\partial}{\partial^3 \beta} R_0(\beta, \beta_{MV}(X)) \geq 0$$

pour $\beta \in [0, m]$, en effet on observe que

i La convexité de $R_0(\beta, \beta_{MV}(X))$ pour $\beta \in [0, m]$, ou

ii Un changement de concavité à convexité de $R_0(\beta, \beta_{MV}(X))$ sur $[0, m]$.

Les deux impliquent (A), notons que la suffisance de (ii) exploite le fait que que le risque de β_{MV} est une fonction de β, ce qui implique que son dérivé à $\beta = 0^+$ est nulle, et que le risque soit décroissante pour un $\beta > 0$ dans un voisinage de 0.

avec le développement de risque $\mathbb{E}[(\beta_{MV}(X) - \beta)^2]$ on obtient

$$\begin{aligned} R_0(\beta, \beta_{MV}(X)) &= \int_{-m}^{m} (x-\beta)^2 \Phi(x-\beta) dx + (m-\beta)^2 \Phi(\beta - m) \quad (5.1) \\ &+ (m+\beta)^2 \Phi(-(\beta+m)) \end{aligned}$$

avec des calculs directs on obtient

$$\begin{aligned} \frac{1}{2}\frac{\partial}{\partial \beta} R_0(\beta, \beta_{MV}(X)) &= (m+\beta)\Phi(-(\beta+m)) + (\beta-m)\Phi(\beta-m) \\ \frac{1}{2}\frac{\partial}{\partial^2 \beta} R_0(\beta, \beta_{MV}(X)) &= \Phi(-\beta-m) + \Phi(\beta-m) - (m+\beta)\Phi(\beta+m) \\ &+ (\beta-m)\Phi(\beta-m) \\ \frac{1}{2}\frac{\partial}{\partial^3 \beta} R_0(\beta, \beta_{MV}(X)) &= \Phi(\beta-m)(2-(m-\beta)^2) - \Phi(\beta+m) \\ &\quad (2-(m+\beta)^2) \end{aligned}$$

Enfin, le resultat (A) suit de $\Phi(\beta - m) \geq \Phi(\beta + m) \geq 0$ et $2 - (m-\beta)^2 \geq 2 - (m+\beta)^2$ et $2 - (m-\beta)^2 \geq 0$ pour $\beta \in [0, m]$, avec $m \leq \sqrt{2}$.

Pour établir (B) utilisé (5.1)pou écrit

$$\begin{aligned} R_0(0, \beta_{MV}(X)) &= 2m^2\Phi(-m) + 2\int_0^m x^2 \Phi(x) dx \\ et \quad R_0(m, \beta_{MV}(X)) &= 4m^2\Phi(-2m) + \int_0^{2m} x^2 \Phi(x) dx \end{aligned}$$

donc, la différence des risques est definit par

$$\begin{aligned}\Delta(0,m,\beta_{MV}) &= R_0(m,\beta_{MV}(X)) - R_0(m,\beta_{MV}(X))\\ \frac{\partial\Delta(0,m,\beta_{MV})}{\partial m} &= 4m\underbrace{\{2\Phi(-2m) - m\Phi(2m) - \Phi(-m)\}}\\ &= 4mN(m)\end{aligned}$$

et on a

$$\frac{\partial N(m)}{\partial m} = \Phi(2m)\{e^{3m^2} + 4m^2 - 5\}$$

maintenant, on observe que $N(0) = \frac{1}{2} > 0$, $\lim\limits_{m\mapsto\infty} N(m) = 0$, et que $\frac{\partial}{\partial m}N(m)$ changer le sign du - au + dans $]0,\infty[$, cette condition implique que N(m) changer le sign une fois du + au - quand $m \in]0,\infty[$,
ce qui montre que $\Delta(0,m,\beta_{MV})$ a au plus un sign change de + au - pour $m \in]0,\infty[$, $(\Delta(0,0,\beta_{MV}) = 0)$, enfin, avec une évaluation numérique $\Delta(0,m,\beta_{MV}) = 0$ pour $m = m_1 = 1,96422$.

5.2 La performance de l'estimateur de type Stein par rapport a celle de moindre carré

La plupart de la litterature traitaient l'estimation de Stien des coéfficients de regression, suppose que les X_i distribuer selon la distribution de Gauss (Normal) comme on a vu dans le deuxiéme chapitre, mais en pratique cette hypothèse pêut ne pas être vraie, comment la performance de l'estimateur de Stein change quand la distribution n'est pas gaussien (ne suit pas la loi normal) ?
Dans cette section an va etudier la performance de l'estimateur de moindre carrée par rapport à l'estimateur de Stein sous la fonction de perte équilibré prolongér (qu'on a vu dans la derniére section de chapitre 4), dans la structure de perte quadratique, dans le cas ou la distribution n'est pas neécessairement

gaussien, sous le modèle de regression lineaire.

On va comparer les fonctions de risque associé aux deux estimateurs (MC-JS), et on cherche des conditions caractérisant les scalaires, pour la dominance de l'estimateur (JS) par rapport a l'estimateur de (MC), on va coincide avec plusieurs formes populaires de fonction de perte qui se présentent comme des cas particuliéres de la fonction de perte équilibré prolongé, rappellons qu'on a travailler sous le modèle de regression lineaire

$$Y = \beta X + \epsilon$$

5.2.1 L'estimateur de (MC)

l'estimateur de (MC) de coéfficient β est donné par

$$\beta_{MC} = (X^t X)^{-1} X^t Y \tag{5.2}$$

qui est bien connue pour son optimalité dans la classe des estimateurs linéaires et sons biais.

5.2.2 L'estimateur de (JS)

C'est une famille intéressante d'estimateurs non linéaires et biaisés de X, connus sous le non d'estimateur de Stein, definit par

$$\beta_{JS}^{6} = \left[1 - \left(\frac{k}{n-p+2}\right)\frac{Y^t(I-H)Y}{Y^t H Y}\right]\beta_{MC} \tag{5.3}$$

$$avec \qquad H = X(X^t X)^{-1} X^t \tag{5.4}$$

et k est un scalaire positive voir[Shalabh 2006][33].

5.2.3 Le risque de l'estimateur de (MC)

On va voir la performance de ces deux estimateurs sous la fonction de perte équilibré prolongé definit dans la derniére section de chaptre(5) (WL), on remarque que

$$WL(\widehat{\beta}) = \lambda\epsilon^t\epsilon - \sigma(1+\lambda_1-\lambda_2)\epsilon^t X(\widehat{\beta}-\beta) + (\widehat{\beta}-\beta)^t X^t X(\widehat{\beta}-\beta) \quad (5.5)$$

On remplace $\widehat{\beta}$ par β_{MC}, on obtient le risque de (MC) definit par

$$R(\beta_{MC}) = \mathbb{E}(WL(\beta_{MC})) = \sigma^2\lambda_1 n - \sigma^2 p(\lambda_1-\lambda_2) \quad (5.6)$$

5.2.4 Le risque de l'estimateur de (JS)

L'expresions exacte de la fonction de risque de l'esimateur (JS) dérivée de leur nature, serait suffisamment compliqué, nous utilisons donc la théorie d'approximation assymptotique des échantillons de grands tailles, pour dérive la fonction de risque.

Nous supposons que les variables explicatives sont asymptotiquement co-orpératives (c-à-d la forme limite de la matrice $n^{-1}X^tX$ est finie quand $n \mapsto +\infty$), cette hypothèse est nécessaire pour l'approximation assymptotique des grands échantillons),

maintenant, si on pose

$$\varepsilon = n^{\frac{1}{2}} X^t \epsilon$$

et

$$\delta = n^{\frac{1}{2}} \left(\frac{\epsilon^t\epsilon}{n} - 1 \right)$$

on obtient

$$\beta_{MC} - \beta = \frac{1}{n^{\frac{1}{2}}} S^{-1}\varepsilon \quad (5.7)$$

avec $S = \frac{1}{n}X^tX$

$$Y^t(I-H)Y = n + n^{\frac{1}{2}}\delta - \varepsilon^t S^{-1}\varepsilon$$

on calcul la quantité,

$$\begin{aligned}
\frac{Y^t(I-H)Y}{(n-p+2)Y^tHY} &= \frac{1}{n\beta^tS\beta}\left(1+\frac{\delta}{n^{\frac{1}{2}}}-\frac{1}{n}\varepsilon^tS^{-1}\varepsilon\right)\left(1-\frac{p-2}{n}\right)^{-1} \\
&\quad \left(1+\frac{2\beta^t\varepsilon}{n^{\frac{1}{2}}\beta^tS\beta}+\frac{\varepsilon^tS^{-1}\varepsilon}{n\beta^tS\beta}\right)^{-1} \\
&= \frac{1}{n\beta^tS\beta}+\frac{1}{n^{\frac{3}{2}}\beta^tS\beta}\left(\delta-\frac{2\beta^t\varepsilon}{\beta^tS\beta}\right)+O_p(n^{-2}) \qquad (5.8)
\end{aligned}$$

Si on remplace (5.7) (5.8) dans (5.3) on trouve

$$\begin{aligned}
(\beta^6_{JS}-\beta) &= \frac{1}{n^{\frac{1}{2}}}S^{-1}\varepsilon-\frac{k}{n^{\frac{3}{2}}\beta^tS\beta}[\delta\beta+(S^{-1}-\frac{2}{\beta^tS\beta}\beta\beta^t) \\
&\quad \varepsilon]+O_p(n^{-2}) \qquad (5.9)
\end{aligned}$$

avec l'utilisation des propriétés distributives de ϵ, ces resultats est on négligeant les termes d'ordre supérieur de petitesse que $O(n^{-1})$, on obtient avec (5.9), la forme suivante

$$\begin{aligned}
\mathbb{E}[\epsilon^tX(\beta^6_{JS}-\beta)] &= \sigma p-\frac{\sigma^2k}{n\beta^tS\beta}[\frac{\gamma_1}{n}e^tX\beta+\sigma(p-2)] \qquad (5.10) \\
\mathbb{E}[(\beta^6_{JS}-\beta)^tX^tX(\beta^6_{JS}-\beta)] &= \sigma^2p-\frac{\sigma^3k}{n\beta^tS\beta}[\frac{2\gamma_1}{n}e^tX\beta \\
&+ 2\sigma(p-2)-6k] \qquad (5.11)
\end{aligned}$$

on remplace $\widehat{\beta}$ par β^6_{JS} dans l'equation (5.5), en utilisant les resultats ci-dessus et en concervant les termes a l'ordre $O(n^{-1})$, on obtient,

$$\begin{aligned}
R(\beta^6_{JS}) &= \mathbb{E}[WL(\beta^6_{JS})] \qquad (5.12) \\
&= \sigma^2\lambda_1n-\sigma^2p(\lambda_1-\lambda_2)-\frac{\sigma^4k}{n\beta^tS\beta}[(1-\lambda_1+\lambda_2) \\
&\quad (\frac{\gamma_1}{\sigma}\overline{X}^t\beta+p-2)-k]
\end{aligned}$$

avec $\overline{X}$ est un vecteur de dimension p, représentent les moyennes des colonnes de matrice X (variable explicative).

On remarque de (5.6) et (5.12) que les deux critère de performance (qualité d'ajustement et précision de l'estimation), affectent la performance de risque de l'estimateur (MC) et (JS), et l'utilisation des deux critéres ensembles a plus d'intérêt que l'utilisation de l'un des deux.

5.2.5 La comparaison des risques

On observe que l'estimateur de (JS) a un risque plus faible, par rapport à l'estimateur de (MC), quand

$$k < (1 - \lambda_1 + \lambda_2)(\frac{\gamma_1}{\sigma}\overline{X}^t\beta + p - 2) \tag{5.13}$$

à condition que

$$(\lambda_1 - \lambda_2) < 1 \quad et \quad (\frac{\gamma_1}{\sigma}\overline{X}^t\beta + p - 2) > 0 \tag{5.14}$$

ou

$$(\lambda_1 - \lambda_2) > 1 \quad et \quad (\frac{\gamma_1}{\sigma}\overline{X}^t\beta + p - 2) < 0 \tag{5.15}$$

quand la distribution est symétrique et/ou $\overline{X} = 0$, (c-à-d les observations sur les variables explicatives sont considérées comme des écarts par rapport à leur moyennes correspondants).
Donc, la condition (5.13) ne depend pas du paramètre inconnue β, et est satisfaite lorsque l'une des deux conditions suivantes est vraie

$$k < (1 - \lambda_1 + \lambda_2)(p - 2) \quad et \quad (\lambda_1 - \lambda_2) < 1 \; si \; p > 2. \tag{5.16}$$

$$k < (\lambda_1 - \lambda_2 - 1)(2 - p) \quad et \quad (\lambda_1 - \lambda_2) > 1 \; si \; p = 1, 2. \tag{5.17}$$

maintenant,nous examinons la performance des deux estimateurs précedents sous certaines fonctions de perte intéressantes,

✠ On considére le critère de la qualité d'ajustemant du modéle, qui est un cas particuliére de (WL) quand $\lambda_1 = 1$ et $\lambda_2 = 0$, on remarque de (5.6) et (5.12),

l'estimateur de (MC) reste invaincu par tout les estimateurs de type stein, quelque soit la nature des observations sur les variables explecatives, et leur distributions, cela correspond au résultats obtenue par [Srivastava et Shalabch 1996 page 143] [34].

✠ On considére le critère de précision de l'estimation, qui obtenue comme cas particuliére de (WL) quand $\lambda_1 = 0$ et $\lambda_2 = 1$ donc

l'estimateur de (JS) est mieux que l'estimateur de (MC), quand

$$k < 2(\frac{\gamma_1}{\sigma}\overline{X}^t\beta + p - 2) \tag{5.18}$$

sachant que la quantité à droit est positive, cette condition se reduit à

$$k < 2(p-2),\ p > 2 \tag{5.19}$$

quand la distribution est symétrique indepépendamment de la nature des données sur les variables explicatives, ou $\overline{X} = 0$, si les distributions sont symétriques ou asymétriques.

De même la condition (5.18) est satisfaittant que (5.19) est verifiée, à condition que (γ_1 et $\overline{X}^t\beta$) a le même sign (c-à-d $\overline{X}^t\beta$ est positive pour la distribution asymétrique positive, et negative pour la distribution asymétrique negative), en effet, c'est possible de trouver un estimateur de (JS) avec une meilleure performance que l'estimateur de (MC) même pou $p = 1$ et $p = 2$, quand

$$\gamma_1\overline{X}^t\beta > 2\sigma \tag{5.20}$$

comme une remarque, la condition (5.19) est bien connue de la supériorité des estimateurs de type stein, sous la distribution normale.

✠ Si on remplace par $\lambda_1 = \theta$ et $\lambda_2 = (1-\theta)$ dans (WL), on obtient la fonction de perte proposer par Zellner 1994 pour $(0 \leq \theta < 1)$, une condition suffisante pour que l'estimateur de (JS) est mieux que l'estimateur de (MC), est quand

$$k < 2(1-\theta)(\frac{\gamma_1}{\sigma}\overline{X}^t\beta + p - 2) \qquad (5.21)$$

quand $\gamma_1 = 0$ et/ou $\overline{X} = 0$, la condition (5.21) prend la forme simple suivante

$$k < 2(1-\theta)(p-2), \quad p > 2 \qquad (5.22)$$

ceci constitue une condition suffisante pour la supériorité des estimateurs de (JS) par rapport à l'estimateur de (MC), dans le cas ou la distributionest asymétrique, a condition que $sgn(\gamma_1) = sgn(\overline{X}^t\beta)$.
De plus la condition (5.21) est vraie on peut trouver des estimateurs de (JS) mieux que l'estimateur de (MC) même si ($p = 1, 2$). Cette condition a été derivée par [15], sous distribution normal, voir aussi [26].

✠ Maintenant, on considére le cas ou $\lambda_1 = \lambda^2$ et $\lambda_2 = (1-\lambda)^2$ dans (WL), on obtient la fonction de perte suivante

$$\lambda^2(X\widehat{\beta}-Y)^t(X\widehat{\beta}-Y)+(1-\lambda)^2(\widehat{\beta}-\beta)^tX^tX(\widehat{\beta}-\beta)+2\lambda(1-\lambda)(X\widehat{\beta}-Y)^tX(\widehat{\beta}-\beta)$$

c'est un combination de la somme des carrés des residues, et la somme pondéré des carrés d'erreurs d'estimation, et leur produit.
Et c'est la somme des carrés des erreurs de prédiction lorsque $X\widehat{\beta}$ est utilisé pour la prédiction d'un combainaison convexe des valeurs actuels et moyennes.
Donc depuis (5.6) et (5.12), l'estimateur de (JS) à un risque petit par rapport à l'estimateur de (MC), quand

$$k < 2(1-\lambda)(\frac{\gamma_1}{\sigma}\overline{X}^t\beta + p - 2) \qquad (5.23)$$

qui est précisément le même que (1.21), obtenue a partir de la fonction de perte équilibré.

Enfin, la comparaison sous la condition (5.23) avec $\gamma_1 = 0$ correspond à la condition de [Shalabh 1995][35].

5.3 L'estimation de paramètre de la loi de Bernaulli sous la fonction de perte équilibrée de Zellner

Soit $X_1, X_2, \ldots, X_n$ une suite de variable aléatoire suit la loi de Bernaulli avec la densité

$$f(x,p) = p^x(1-p)^{1-x} \quad x = 0,1 \quad et \quad 0 < p < 1$$

on va estimer p sous la fonction de perte équilibrée

$$L(p,\widehat{p}) = \frac{w}{n}{}_{i=1}^{n}(X_i - \widehat{p})^2 + (1-w)(p-\widehat{p})^2 \quad 0 < w < 1 \tag{5.24}$$

Dans cette application on va étudier l'admissibilité de l'estimateur lineaire (affine) de la forme $c\overline{X} + d$ pour l'estimation de la moyenne de la loi de Bernaulli, ou la moyenne de telles distributions exponentielles, sous la fonction de perte (5.24).

5.3.1 L'estimateur de bayes

Proposition 5.3.1. *La fonction de risque de l'estimateur $c\overline{X} + d$, sous la fonction de perte (5.24) est*

$$\begin{aligned} R(p, c\overline{X} + d) &= [(c-1)p + d]^2 + \frac{p(p-1)}{n}[(c-w)^2 + w(n-w)] \\ &= [(c-1)^2 - \frac{(c-w)^2 + w(n-w)}{n}]p^2 \\ &+ [2d(c-1) + \frac{(c-w)^2 + w(n-w)}{n}]p + d^2 \end{aligned}$$

et le risque de bayes de $c\overline{X}+d$, sous la priori beta est

$$\begin{aligned} r(\pi, c\overline{X}+d) &= [(c-1)^2 - \frac{(c-w)^2 + w(n-w)}{n}]\frac{c(c+1)}{(\alpha+\beta)(\alpha+\beta+1)} \\ &+ [2d(c-1) + \frac{(c-w)^2 + w(n-w)}{n}]\frac{\alpha}{\alpha+\beta} + d^2 \end{aligned}$$

On peut facillement remarquer que la loi a priori de l'estimateur de bayes de p, suit la distribution beta $B(\alpha, \beta)$ avec la densité

$$\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \quad 0 < p < 1 \tag{5.25}$$

avec $\alpha > 0$ et $\beta > 0$,alors la loi a posteriori est definit par

$$B(\alpha + \sum_{i=1}^{n} x_i, n - \sum_{i=1}^{n} x_i + \beta)$$

le risque a posteriori de l'estimateur $\widehat{p}$ sous (5.24) est

$$\mathbb{E}(L(\widehat{p}, p)) = \frac{w}{n}\sum_{i=1}^{n}(X_i - \widehat{p})^2 + (1-w)\mathbb{E}[(p-\widehat{p})^2]$$

aprés resoudre l'équation suivante

$$\frac{\partial \mathbb{E}[L(\widehat{p}, p)]}{\partial \widehat{p}} = 0$$

on obtient

$$p^{\pi} = w\overline{X} + (1-w)\overline{p}$$

avec $\overline{p}$ est la moyenne a posteriori, l'estimateur de bayes sous la loi a priori beta est

$$\begin{aligned} p^{\pi} &= w\overline{X} + (1-w)\frac{a + \sum_{i=1}^{n} x_i}{n+\alpha+\beta} \\ &= \frac{n + \alpha w + \beta w}{n+\alpha+\beta}\overline{X} + (1-w)\frac{\alpha}{n+\alpha+\beta} \end{aligned} \tag{5.26}$$

telque $0 < \dfrac{n+\alpha w+\beta w}{n+\alpha+\beta} < 1$. On peut aussi conclut le risque de p^{π} depuis la proposition (5.3.1).

5.3.2 Admissibilité

Théorème 5.3.1. *L'estimateur* $c\overline{X}$+d *est innadmissible sous la perte (5.24), sous les conditions suivantes :*

(i) c>*1.*

(ii) c<w.

(iii) w<c<*1 et* c+d>*1.*

(iv) w<c<*1 et* d<*0.*

(v) c=*1 et* d≠*0.*

(vi) c=w *et* d<*0.*

Démonstration :

(i) Si c>1 alors $(c-w)^2 > (1-w)^2$ et daprés la proposition precédente

$$R(p, c\overline{X} + d) > R(p, \overline{X})$$

(ii) Si c<w alors $(c-1)^2 > (w-1)^2$ et

$$R(p, c\overline{X} + d) > R(p, w\overline{X} + \frac{d(w-1)}{(c-1)})$$

(iii) Si w<c<1 et c+d>1 alors

$$(c-1)p + d > (c-1)p + 1 - c = (c-1)(p-1) > 0$$

on obtient,

$$R(p, c\overline{X} + d) > R(p, c\overline{X} + 1 - c)$$

(iv) pour w<c<1 et d<0, l'estimateur $c\overline{X} + d$ est dominé par $c\overline{X} - d$.

(v) Dans se cas $c\overline{X} + d = \overline{X} + d$ est dominé par $\overline{X}$.

(vi) Il est facile de voir que l'estimateur $w\overline{X} + d$ est dominé par $w\overline{X}$.

Théorème 5.3.2. *L'estimateur* $c\overline{X} + d$ *est admissible sous la condition (★)* w<c<*1 et* $0 \leq d < 1 - c$.

Démonstration :

On considére

$$\alpha^* = \frac{nd}{c-w} \quad \beta^* = \frac{n(1-c-d)}{c-w}$$

la condition (★) assure que $\alpha^* > 0$ et $\beta^* > 0$ depuis (5.26), tanque

$$\begin{aligned} \frac{n+\alpha^* w+\beta^* w}{n+\alpha^*+\beta^*} &= c \\ \frac{(1-w)\alpha^*}{n+\alpha^*+\beta^*} &= d \end{aligned}$$

Il s'ensuit que $c\overline{X}+d$ est un estimateur de bayes de p, sous la priori $B(\alpha^*, \beta^*)$, on a aussi la perte (5.24) alors (5.26) est unique et admissible.

Il s'ensuit aussi que $c\overline{X}+d$ est admissible quand w<c<1 et 0<d<1-c.

Dans le cas d=0, on va montrer que l'estimateur $c\overline{X}$, est admissible quand $a \in]w,1[$, on considére l'estimateur

$$\widehat{p} = (\frac{n+\beta w}{n+\beta})\overline{X}$$

On remarque que $\widehat{p}$ est la limite de l'estimateur de bayes sous la priori $B(\alpha, \beta)$ quand $a \mapsto 0$, on va établir l'admissibilité de $\widehat{p}$, avec la technique de limite de Bayes et Blyth (1951) [3], on considére la priori impropre

$$\pi_k = p^{\frac{1}{k}-1}(1-p)^{\beta-1} \quad \beta > 0$$

si A est un sous ensemble convexe non dégénérée de]0,1[, il existe k_0 telque $\displaystyle\int_A \pi_k(p)dp \geq \varepsilon$ pour un $\varepsilon > 0$, et pour tout $k \geq k_0$.

L'estimateur de bayes par rapport a π_k depuis (5.26) est donné par

$$p_k^\pi = \frac{nk+w+\beta kw}{nk+1+\beta k}\overline{X} + \frac{1-w}{nk+1+\beta k}$$

On trouve que la limite de la différence entre le risque de bayes par rapport a p_k^π et p^π est nulle

$$\lim_{k \mapsto \infty} r(\pi_k, p^\pi) - r(\pi_k, p_k^\pi) = 0$$

par conséquent, $\frac{n+\beta w}{n+\beta}\overline{X}$ est admissible.

Théorème 5.3.3. *Sous la fonction de perte (5.24) $\overline{X}$ est admissible.*

Démonstration :

On considére la priori impropre

$$\pi_k = p^{\frac{1}{k}-1}(1-p)^{\frac{1}{k}-1} \quad k > 0$$

l'estimateur de bayes $\widetilde{p_k^{\pi}}$ par rapport a π_k, depuis (5.26) est donné par

$$\widetilde{p_k^{\pi}} = \frac{nk+2w}{nk+2}\overline{X} + \frac{1-w}{nk+2}$$

et avec la même méthode de démonstration précedente

$$\lim_{k \mapsto \infty} r(\pi_k, \overline{X}) - r(\pi_k, \widetilde{p_k^{\pi}}) = 0$$

alors, $\overline{X}$ est admissible avec le lemme de Blyth's (1951)[3].

5.3.3 Minimaxité

Pour determiner un estimateur minimax, on considére l'estimateur de bayes (5.26) et on fixe des valeurs $\alpha > 0$ et $\beta > 0$, telque la fonction de risque de l'estimateur de bayes soit constante, depuis la proposition (5.3.1) la fonction de risque de p^{π} est donné par

$$\begin{aligned} R(p, p^{\pi}) &= [\frac{(1-w)^2(\alpha+\beta)^2}{(n+\alpha+\beta)^2} - \frac{n(1-w)}{(n+\alpha+\beta)^2} - \frac{w(n-w)}{n}]p^2 \\ &+ [-2\alpha(\alpha+\beta)(1-w)^2 + \frac{n(1-w)^2}{(n+\alpha+\beta)^2} + \frac{w(n-w)}{n}]p \\ &+ \frac{(1-w)^2\alpha^2}{(n+\alpha+\beta)^2} \end{aligned} \tag{5.27}$$

On va montrer que (5.27) est constante, si et seullement si

$$\left\{ \begin{array}{rcl} n(\alpha+\beta)^2(1-w)^2 &=& n^2(1-w)^2 + w(n-w)(n+\alpha+\beta)^2 \\ 2n\alpha(\alpha+\beta)(1-w)^2 &=& n^2(1-w)^2 + w(n-w)(n+\alpha+\beta)^2 \end{array} \right. \tag{5.28}$$

maintenant, supposons que $0 < w < \dfrac{3n - \sqrt{n(5n-4)}}{2(n+1)}$.

Noté que $\dfrac{3n - \sqrt{n(5n-4)}}{2(n+1)} \leq \dfrac{1}{2}$, la résolution des équations (5.28) our α et β donne

$$\alpha = \beta = \frac{nw(n-w) + n(1-w)\sqrt{n(1-w)^2 + w(n-1)(n-w)}}{2n(1-w)^2 - 2w(n-w)} = C_n(w)$$

on a par hypothèse $C_n > 0$, et donc l'estimateur suivant

$$\delta = \frac{n + 2C_n(w)w}{n + 2C_n(w)}\overline{X} + \frac{(1-w)C_n(w)}{n + 2C_n(w)}$$

est l'unique estimateur minimax de p sous (5.24) avec

$$0 < w < \frac{3n - \sqrt{n(5n-4)}}{2(n+1)}$$

la fonction de risque de l'estimateur δ est

$$R(p, \delta) = \frac{(1-w)^2 C_n^2(w)}{(n + 2C_n)^2}$$

Dans le tableau suivant on va calculer les valeurs de $C_n(w)$ pour les différents valeurs de n quand $w = \frac{1}{3}, \frac{1}{4}\ et\ \frac{1}{5}$, si on veut choisir entre les deux estimateurs δ et $\overline{X}$ ce qui a un risque n'est pas constant definit par

$$R(p, \overline{X}) = \frac{p(1-p)}{n}[(1-w)^2 + w(n-w)]$$

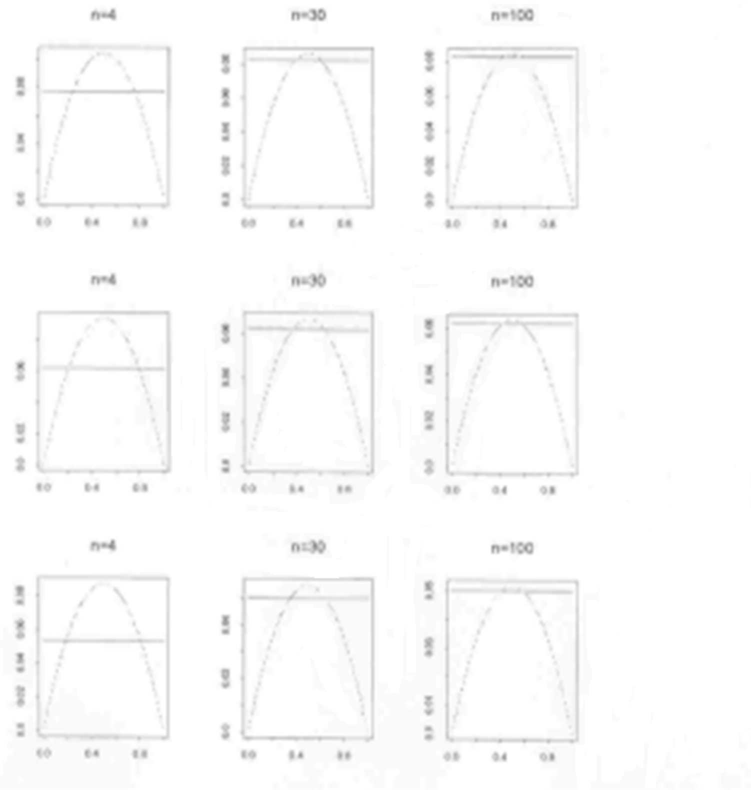

FIGURE 5.1 – La comparaison entre les deux estimateurs $\overline{X}$ et δ.

	w=$\frac{1}{3}$	w=$\frac{1}{4}$	w=$\frac{1}{5}$
n	$C_n(w)$	$C_n(w)$	$C_n(w)$
2	4.07	2	1.50
3	6.97	3	2.16
4	10	4	2.80
5	13.10	5	3.44
10	28.95	10	6.62
15	45.00	15	9.80
20	61.10	20	12.97
25	77.22	25	16.14
30	93.36	30	19.31
50	157.95	50	31.99
100	319.51	100	63.68

Dans (la figure 5.1) on comparer le risque de $\overline{X}$ par rapport à δ, pour $n = 4, 30, 100$ et $w = \frac{1}{3}, \frac{1}{4}\ et\ \frac{1}{5}$ (respectivement), on remarque que pour un petit n, δ est mieux que $\overline{X}$ et pour un grand n (même si n modéré) $\overline{X}$ est mieux que δ.

Conclusion et prespectives

Dans ce mémoire on a présenté quelques travaux sur l'amélioration des estimateurs usuels et de type "Stein", dans le cadre paramétrique, et paramétrique bayésienne.
Ce travail peut être généraliser, à une étude dans le cadre non paramétrique, pour l'amélioration de l'estimateur a noyau, en utilisant une fonction de perte génnéralisée inspirée de celle definit par Arnold Zellner en 1994.

Bibliographie

[1] **A.Weil**, L'intergration dans les groupes topologiques et ses applications, Paris, Hermann, (1938).

[2] **Brown, L.D**, Fonction of exponential families, IMS lecture Notes, Monographeseries 9, Hayward, California, (1986).

[3] **Blyth, C.R**, On minimax statistical decesion procedures and their admissibility, Ann. Math. Statist. Vol. 22 (22-42), (1951).

[4] **Chou, J.p**, Admissibility of conjugate bayes estimators for the mean of a negative binomial distribution, Statistics and Decisions, 13, 301-306, (1995).

[5] **Charles Stein**, The inadmissibility of the usual estimator for the mean of a multivariate normal distribution, University of stanford, 1956.

[6] **Charles M. Stein**, Estimation of the mean of a multivariate normal distribution, Institute of mathematical statistics, The annals of statistics, Vol.9,NO,6 (1135-1151), Nov 1981.

[7] **C.Stein**, A necessary and sufficient condition for admissibility, An, Math statist, Vol.26, p 518-522, (1955).

[8] **C.Blyth**, On minimax statistical decision procedures and their admissibility, An, Math statist, Vol.22, p 22-42, (1951).

[9] **Casella, G., Strawderman, W.E.** . Estimating a bounded normal mean. Annals of Statistics, 9, 870-878, (1981).

[10] **Christian P. Robert**, Le choix bayésien Principes et pratique, Université Paris Dauphine et CREST, INSEE, Paris, (2006).

[11] **Dey, D.Ghosh, M.Strawderman, W.E**, On estimation with balanced loss functions, Statistics and Probability letters, 45, 97-101, (1999).

[12] **Ferguson .T.S**, Mathematical Statistics ; A decision theoric approach, Academic press, Newyork London, (1967).

[13] **Gupta**, On the admissiblity of linear estimates for estimating the mean of distributions of the one paramètre exponential fmily, Calcul statistical association Bulletin, 15, 14-19, (1966).

[14] **Goméz Déniz, E**, On the use of the weighted balanced loss function to obtain credibility premiums, International conference of Enrique Castillo, June 28-30, (2006).

[15] **Giles, J.A., Giles, D.E.A. and Ohtani K.**, The exact risks of some pre-test and Stein-type regression estimates under balanced loss, Communications in Statistics - Theory and Methods, 25, pp 2901-2924, (1996).

[16] **Jafari Jozani,M., Nematollahi, N., Shafie, K.**, An admissible minimax estimators of a bounded scale-parameter in subclass of the exponential family under scale-invariant squared error loss, Statistics and Probability letters, 60, 437-444, (2002).

[17] **Jafari Jozani,M.Marchand,E.Parsian** ,On estimation with weighed balanced type loss function,Statistics and Probability letters, 76(773-780) 2006.

[18] **Lehman, E.L, Casella,G.**, Theory of point estimation, Springer-Verlag, Newyork, 2nd edition, (1998).

[19] **Marchand, É., Strawderman, W.E**, On improving on the minimum risk equivariant estimator of a location parameter which is constrained

to an Interval or a half-interval, Annals of the institueof statistical mathematics, 57, 129-143, (2005a).

[20] **Marchand, É., Strawderman, W.E**, On improving on the minimum risk equivariant estimator of a scale parameter under a lower-bound constraint, Journal of statistical planing and Inference, 134,90-101, (2005b).

[21] **Marchand, E., Strawderman.W.E** , Estimation in restricted parameter spaces : A review .AFests-chrift for Herman Robin, IMS lecture Notes-Monographe series 45, Institue of mathematical statistics, Heyward, Clifornia, 21-44, (2004).

[22] **Mohammad Jafari Jozani, Eric Marchand, Ahmad Parsian**, Bayes Estimation Under a General Class of Balanced Loss Functions, School of Mathematics, Statistics and Computer Science, University of Tehran, Tehran, IRAN, Université de Sherbrooke, Département de mathématiques, Sherbrooke, QC, CANADA, J1K 2R1.

[23] **Mohammad Jafari Jozani, Eric Marchand, Ahmad Parsian**, On estimation with weighted balanced type loss function, Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, IRAN, University of New Brunswick and Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc, CANADA, School of Mathematical Sciences, Isfahan University of Technology, Isfahan, 84156, IRAN.

[24] **Marchand, Eric, Perron, François**. Improving on the MLE of a Bounded Normal Mean. Annals of Statistics, 29, 1078-1093, (2001).

[25] **N.Sanjari Frasipour and A. Asgarsadzadeh**, Estimation of the parameter of a bernaulli distribution using a balanced loss function, the

Korean communication in statistics, Vol.9, NO. 3, 2002, pp(889-899), (2002).

[26] **Ohtani, K.**,The exact risk of a weighted average estimator of the OLS and Stein-rule estimators in regression under balanced loss, Statistics and Decisions, 16, pp. 35-45,(1998).

[27] **Parsian.A and Kirmani, S.N.U.A**, Estimation under linex loss function, Handbook of applied econometrics and statistical inference, statistics textbooks Monographe, 53-76, (2002).

[28] **Parsian, A., Nematollahi, N.**, Estimation of scale parametre under entropy loss function, journalof statistical planing and inference, 52, 77-91, (1996).

[29] **Rodrigues, J.Zellner, A.** , Weighted balanced loss function and estimation of the mean time to failure, communications in statistics : theory and methods, 23, 3609-3616, (1994).

[30] **Robert, C.P**, Intrinsic loss function, theory and decision, 40, 192-214, (1996).

[31] **Sanjari Farsipour, N. et Asgharzadeh, A.** , Estimation of the paramètre of a Bernoulli distribution using a balanced loss function, Personal communication, (2004).

[32] **Spring, F.A** , The reflected normal loss function canadian journal of statistics, 21, 321, (1993).

[33] **Shalabh and Helge Toutenburg and Christian Heumann**, Stein-Rule Estimation under an Extended, Balanced Loss Function, Department of Statistics University of Munich, Number 007, 2007, December 5, 2007.

[34] **Srivastava, A.K. and Shalabh**, Éfficiency properties of least squares and Stein-rule predictions in linear regression model, Journal of Applied Statistical Science, 4, pp. 141-145, (1996).

[35] **Shalabh**, Performance of Stein-rule procedure for simultaneous prediction of actual and average values of study variable in linear regression model, Jan Tinbergen Award Paper, Bulletin of the 50th Session of the International Statistical Institute, pp. 1375- 1390, (1995).

[36] **Van Eeden, C.**, Minimax estimation of a lower bounded scale parameter of a gamma distribution for scale-invariant square error loss, Canadian Journal of statistics, 23, 245-256, (1995).

[37] **W.James, Charles Stein**, Estimation with quadratic loss, Fresno state college, Stanford university, 1961.

[38] **Zellner.A** , Bayesian and non bayesian estimation using balanced loss function, Statistical decision theory and methods V,(J.O.Berger and S.S.Gupta EDS). Newyork : Springer-Verlag, 337-390, (1994).

[39] **Zellner.A**, Bayesian estimation and prediction using asymetric loss functions, Journal of the American statistical association, 81, 446-451, (1986).