

2m11.2862.10

Université de Montréal

ESTIMATEURS d'HORVITZ-THOMPSON
ROBUSTES BASÉS SUR DES MODÈLES
BAYÉSIENS

par

Fritz Pierre

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

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

décembre 2000

© Fritz Pierre, 2001



QA

3

154

2001

N.008

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

**ESTIMATEURS d'HORVITZ-THOMPSON
ROBUSTES BASÉS SUR DES MODÈLES
BAYÉSIENS**

présenté par

Fritz Pierre

a été évalué par un jury composé des personnes suivantes :

Urs Maag

(président-rapporteur)

Yves Lepage

(directeur de recherche)

Jean-François Angers

(co-directeur)

Martin Bilodeau

(membre du jury)

Mémoire accepté le :

19 mars 2001

SOMMAIRE

Dans ce mémoire, nous abordons le problème des valeurs aberrantes et influentes en théorie de l'échantillonnage en utilisant des estimateurs robustes. Plus précisément, nous développons deux estimateurs robustes de la moyenne d'une population asymétrique finie à l'aide de modèles bayésiens. Un des estimateurs robustes proposés est basé sur le modèle hiérarchique de Bayes tandis que l'autre est tiré du modèle de paramètre de dispersion augmenté. Une étude par simulation illustre les propriétés des nouveaux estimateurs.

REMERCIEMENTS

Je remercie ma mère Dieula tout particulièrement pour m'avoir initié au chemin de la réussite, mon fils Frentzen, mes journées seraient sombres sans son sourire et sa bonne humeur quotidiens.

Je remercie également tous mes amis qui ont participé de près ou de loin à la réussite de ce mémoire.

Je ne peux terminer sans remercier sincèrement mes directeurs Yves Lepage et Jean-François Angers. Ils constituent la base même de cet ouvrage. Je remercie Jean-François pour sa grande disponibilité et ses explications toujours claires et précises. Je termine en soulignant le travail d'Yves qui m'a servi de guide tout au long de ce travail.

Table des matières

Sommaire	iii
Remerciements	iv
Table des figures.....	viii
Liste des tableaux	ix
Introduction.....	1
Chapitre 1. Estimateurs d’Horvitz-Thompson robustes	5
1.1. Estimateur d’Horvitz-Thompson	6
1.1.1. Introduction et notation	6
1.1.2. Estimateur d’Horvitz-Thompson sous le plan PIPT	9
1.1.3. Estimateur Horvitz-Thompson comme fonctionnelle de moindres carrés.....	10
1.2. Robustification de l’estimateur HT	14
1.2.1. Introduction	14
1.2.2. Estimateurs HTR.....	16
1.2.3. Espérance et variance de HTR.....	18
1.2.4. Estimation de la variance de l’estimateur HTR	23
1.3. Estimateurs à risque estimé minimum (estimateurs REM).....	25

1.3.1. Motivation des estimateurs <i>REM</i>	25
1.3.2. Définition de l'estimateur <i>REM</i>	26
Chapitre 2. Estimateurs de Bayes	29
2.1. Modèles bayésiens	29
2.1.1. Introduction	29
2.1.2. Inférence statistique bayésienne : l'approche décisionnelle	34
2.1.3. Estimation bayésienne sous la fonction de perte quadratique ...	36
2.2. Modèles hiérarchiques	38
2.3. L'analyse bayésienne empirique	47
2.4. Modèles de régression linéaire bayésien	49
2.4.1. Modèle normal	50
2.4.2. Modèle normal hiérarchique	51
2.4.3. Modèle de paramètre de dispersion augmenté	54
2.5. Estimateurs bayésiens robustes de la moyenne	58
2.5.1. Estimateur hiérarchique de la moyenne	59
2.5.2. Estimateur de paramètre de dispersion augmenté de la moyenne	62
Chapitre 3. Analyse des résultats	66
3.1. Populations à l'étude	66
3.2. Descriptions des estimateurs	71
3.3. Mesures fréquentistes	73
3.4. Analyse des résultats	77
Conclusion	82

Annexe A. Les programmes S-PLUS	85
Bibliographie	106

Table des figures

3.1.1	Les huit populations à l'étude.....	67
3.1.2	Histogramme de la population M284.....	70
3.1.3	Histogramme de la population M281.....	72

Liste des tableaux

3.1.1	Statistiques descriptive sur les populations à l'étude.....	68
3.3.1	Biais relatif, écart type relatif, efficacité des estimateurs et moyenne de Monte Carlo de l'estimateur T_{HT} pour chacune des populations	75
3.3.2	Moyenne du temps de calcul, du nombre de valeurs influentes et du nombre de mauvais points de levier par échantillon pour chacune des populations.....	77

INTRODUCTION

Le problème des valeurs aberrantes est important dans toutes les branches de la statistique. En théorie d'échantillonnage, ce problème se rencontre lors de l'échantillonnage des populations dont les distributions sont étalées vers la droite. Les échantillons tirés de telles populations contiennent souvent des valeurs aberrantes qui sont beaucoup plus grandes que la plupart des valeurs échantillonnées. On tient compte habituellement de telles unités à l'étape de l'élaboration du plan de sondage. Toutefois, compte tenu des buts multiples de la plupart des sondages, on se trouve souvent aux prises avec des valeurs aberrantes à l'étape de l'estimation. Ces observations nuisent à la stabilité des estimateurs classiques de la moyenne comme l'estimateur d'Horvitz-Thompson.

Kish (1965) a abordé ce problème. En effet, il arrive souvent que les populations dans les enquêtes économiques et les enquêtes portant sur les personnes soient asymétriques et que quelques unités risquent d'être extrêmes par rapport aux autres. Le rejet de telles unités présentes dans un échantillon rendrait l'estimateur d'Horvitz-Thompson biaisé. Cependant, en conservant ces valeurs avec tout leur poids, on rendrait cet estimateur extrêmement instable puisque les valeurs aberrantes n'apparaîtraient, typiquement, que dans quelques-uns des échantillons possibles. Il y a donc un effet de compensation entre le biais et la variance dans des cas comme celui-ci qui comprennent, en particulier, les distributions

asymétriques dont une aile est importante. Il paraît donc utile d'étudier des estimateurs reposant sur un compromis entre le biais et la variance. En présence de valeurs aberrantes, le défi est de proposer des estimateurs robustes de la moyenne ayant un petit biais et un petit écart quadratique moyen qui soient peu influencés par certaines unités très différentes des autres.

Plusieurs approches ont été proposées pour obtenir de tels estimateurs. Entre autre, Kish (1965) propose la formation des strates séparées pour les valeurs aberrantes, lorsque possible ainsi que la troncature (précisée avec la moyenne «winsorisée» proposée par Searls, 1966), la transformation ou la modélisation. Chambers (1986) met au point une méthode d'accroissement de la robustesse de la prédiction s'appuyant sur l'utilisation des estimateurs M . Également, sous le plan à probabilités d'inclusion proportionnelles à une mesure de taille, Hulliger (1991 et 1995) propose des estimateurs robustes basés sur l'utilisation des estimateurs M . Fuller (1991) suggère pour sa part un estimateur de test préliminaire ayant pour effet de réduire l'incidence des valeurs les plus grandes dans le seul cas où un test des valeurs extrêmes donne des résultats significatifs. Lee (1995) discute de diverses propositions reposant sur la méthode de réduction de poids dans un tirage aléatoire simple. Duchesne (1999) développe un estimateur de calage robuste basé sur l'estimateur de calage de Deville et Särndal (1992). La modélisation est élaborée sur un angle prédictif qui s'appuie sur des modèles stochastiques de la population pour prévoir la moyenne de la réalisation actuelle.

On peut également aborder le problème des valeurs aberrantes d'un point de vue bayésien. Dans cette optique, le concept de robustesse concerne le choix

de la loi *a priori*, la fonction de perte et la fonction de vraisemblance des observations (voir Ramsay et Novick, 1980). Le choix de la loi *a priori* robuste a été étudié essentiellement à l'aide de deux méthodes : les lois ϵ -contaminées et les modèles hiérarchiques. Ghosh et Sinha (1990) ont développé un estimateur de la moyenne basé sur un modèle hiérarchique. Dans le contexte de l'échantillonnage, Dal Ho (1998) propose des estimateurs de Bayes robustes utilisant les lois *a priori* *ML - II* en présence d'informations auxiliaires.

On peut remarquer que les articles consacrés aux estimateurs bayésiens robustes (en échantillonnage) ne sont pas nombreux. En plus, la plupart des estimateurs proposés ignorent le plan d'échantillonnage sous lequel les unités ont été tirées; cet élément est important dans l'élaboration des estimateurs de la moyenne (ou du total). Nous considérons donc qu'il est primordial de développer des estimateurs bayésiens qui incorporent le plan d'échantillonnage dans l'estimation. Les estimateurs bayésiens robustes que nous allons développer dans ce mémoire auront en plus comme avantage d'être beaucoup plus facile à calculer tout en étant presque autant efficace (en terme de l'écart quadratique moyen) que les estimateurs classiques (basé sur la théorie fréquentiste) connus jusqu'à présent.

Dans le présent mémoire, nous considérons le problème des valeurs aberrantes en échantillonnage en présence d'information auxiliaire. La présentation des estimateurs robustes de Hulliger (1995) constitue essentiellement l'objet du premier chapitre de ce mémoire. Évidemment, nous ne pouvons pas introduire ces estimateurs sans faire un bref rappel du plan d'échantillonnage avec probabilités proportionnelles à la taille et de l'estimateur d'Horvitz-Thompson et de ses propriétés (biais et variance). Par la suite, nous exprimons cet estimateur sous

la forme d'une fonctionnelle de moindres carrés. Nous serons alors en mesure de présenter les deux estimateurs robustes de la moyenne basés sur les estimateurs M . Nous présentons également quelques propriétés de ces derniers.

Dans le deuxième chapitre, nous utilisons les éléments de la théorie bayésienne pour proposer deux estimateurs robustes de la moyenne. Par estimateurs robustes nous entendons des estimateurs qui ne sont pas influencés par des valeurs extrêmes. Nous présentons, en premier lieu, dans un cadre général, les modèles bayésiens. Nous y introduisons les concepts fondamentaux de la théorie bayésienne : les lois *a priori* et *a posteriori* ainsi que la loi marginale du modèle. Nous poursuivons avec les modèles hiérarchiques de Bayes pour présenter ensuite l'analyse bayésienne empirique. Nous enchaînons avec les modèles de régression linéaire bayésien où nous verrons un modèle hiérarchique particulier, le modèle normal hiérarchique et le modèle de paramètre de dispersion augmenté de Box et Tiao (1968). En dernier lieu, nous présentons deux nouveaux estimateurs robustes de la moyenne, un tiré du modèle hiérarchique normal et l'autre basé sur le modèle de paramètre de dispersion augmenté.

Le troisième et dernier chapitre de ce mémoire est une application des deux premiers chapitres. Nous présentons une analyse empirique des estimateurs étudiés basée sur une comparaison des propriétés d'efficacité empirique. Dans ce cadre, nous utilisons, d'une part des populations simulées par des méthodes aléatoires, d'autre part des populations réelles qui ont beaucoup été étudiées dans la littérature (voir Duchesne, 1999).

Chapitre 1

ESTIMATEURS D'HORVITZ-THOMPSON ROBUSTES

La moyenne d'une variable appartenant à une population finie est un indicateur important à cause de son rapport avec la somme. Elle est une caractéristique sensible puisqu'une seule observation démesurée peut suffire à en influencer la valeur. L'estimateur d'Horvitz-Thompson (estimateur HT) est un estimateur naturel de la moyenne d'une population lorsque le plan d'échantillonnage est assorti de probabilités d'inclusion inégales et que l'échantillonnage se fait sans remise. Il est sans biais, peu importe la distribution de la variable étudiée. Cependant, puisque c'est un estimateur linéaire, comme la valeur qu'il estime, il résiste mal aux valeurs aberrantes (unités très différentes des autres). En supposant que les valeurs aberrantes qui peuvent être présentes dans l'échantillon sont légitimes, il devient utile de développer des estimateurs qui ne seront pas trop affectés par ces valeurs extrêmes. Les estimateurs HT à robustesse accrue grâce aux M -estimateurs présentés dans ce chapitre sont des candidats valables à cette tâche difficile.

La première section de ce chapitre est consacrée à l'estimateur HT ; sous le plan assorti de probabilités d'inclusion proportionnelles à une mesure de taille, il est exprimé sous la forme d'une fonctionnelle de moindres carrés pour permettre

sa robustification à l'aide des M -estimateurs. Sa variance ainsi que deux estimateurs sans biais de celle-ci sont données également dans cette section. Nous présentons une première tentative de robustification de l'estimateur HT dans la deuxième section. Ces estimateurs s'obtiennent par la minimisation d'une certaine somme ou d'une intégrale qui dépend à la fois des probabilités d'inclusion et d'une fonction de poids η qui sert à diminuer l'influence des résidus aberrants. Ce nouvel estimateur est appelé estimateur HT à robustesse accrue et est noté HTR . Tout en donnant une formule d'estimation de sa variance, il est montré que l'estimateur HTR est convergent pour la moyenne d'une population finie. Enfin, la troisième et dernière section de ce chapitre est une généralisation de la deuxième section 2. En effet, on calcule plusieurs estimateurs HTR en ajustant deux constantes de mise au point afin de minimiser le risque estimé. L'estimateur obtenu, appelé estimateur à risque estimé minimum (estimateur REM), est l'estimateur HTR pour lequel le risque estimé est minimum.

1.1. ESTIMATEUR D'HORVITZ-THOMPSON

1.1.1. Introduction et notation

Nous considérons un ensemble appelé population, composé de N éléments, appelés unités de sondage ou tout simplement unités ; symboliquement, nous dénotons la population par $U = \{1, 2, \dots, N\}$, $0 < N < \infty$. Soit y_i la valeur prise par la variable Y pour la i -ième unité, une certaine caractéristique de la population. L'objectif consiste à estimer la moyenne ou le total des valeurs y_1, y_2, \dots, y_N . Nous dénotons cette moyenne par $\bar{y}_U = \sum_{i=1}^N y_i/N$.

Un échantillon s est un sous-ensemble quelconque de U . Comme ensemble, s est composé d'unités distinctes.

Définition 1.1.1. Soit S l'ensemble des échantillons. Par un **plan d'échantillonnage** (plan de sondage) $p(\cdot)$, nous entendons une loi de probabilité sur S , c'est-à-dire, pour tout $s \in S$, $p(s) \geq 0$ est connu et

$$\sum_{s \in S} p(s) = 1.$$

Ici $p(s)$ représente la probabilité de choisir s selon le plan en question.

Soit n_s la taille de s . Si $p(s) > 0$ entraîne $n_s = n \forall s$, une constante, on dit que le plan est de *taille fixe*. Si non, le plan est de *taille aléatoire*.

La probabilité d'inclusion de la i -ième unité, c'est-à-dire la probabilité que cette unité soit sélectionnée, peut être calculée à partir de $p(s)$ comme

$$\pi_i = \sum_{s \in C_i} p(s),$$

où $C_i = \{s : i \in s\}$; $i = 1, \dots, N$. De la même façon, la probabilité d'inclusion simultanée des unités i et j dans s , notée π_{ij} , peut se calculer comme

$$\pi_{ij} = \sum_{s \in C_{ij}} p(s),$$

où $C_{ij} = \{s : i \in s, j \in s\}$; $i \neq j = 1, \dots, N$. Dans ce mémoire nous supposons que $\pi_i > 0$ et $\pi_{ij} > 0$ pour tout $i, j \in U$.

On tire un échantillon s de taille n de la population selon un plan $p(S)$, où S est l'ensemble de tous les échantillons possibles de taille n .

Définition 1.1.2. *L'estimateur d'Horvitz-Thompson de la moyenne de Y sur la population, $\bar{y}_U = \sum_{i=1}^N y_i/N$, est défini par*

$$T_{HT} = N^{-1} \sum_{i \in s} y_i / \pi_i. \quad (1.1.1)$$

La variance de cet estimateur est donnée par

$$V_p(T_{HT}) = N^{-2} \left[\sum_{i \in U} \pi_i (1 - \pi_i) \frac{y_i^2}{\pi_i^2} + \sum_{i \neq j \in U} \pi_{ij} (1 - \pi_i \pi_j / \pi_{ij}) \frac{y_i y_j}{\pi_i \pi_j} \right], \quad (1.1.2)$$

et deux estimateurs sans biais de cette variance sont :

$$v_{HT} := \hat{V}_p(T_{HT}) = \frac{1}{N^2} \left[\sum_{i \in s} (1 - \pi_i) \frac{y_i^2}{\pi_i^2} + \sum_{i \neq j \in s} \left(1 - \frac{\pi_i \pi_j}{\pi_{ij}}\right) \frac{y_i y_j}{\pi_i \pi_j} \right], \quad (1.1.3)$$

$$v_{YGS} := \hat{V}_p(T_{HT}) = \frac{1}{2N^2} \sum_{i=1}^n \sum_{j>i}^n \frac{(\pi_i \pi_j - \pi_{ij})}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2, \quad (1.1.4)$$

où l'indice p représente la variance par rapport à $p(s)$.

L'estimateur d'Horvitz-Thompson est un estimateur sans biais de la moyenne de la population (voir Horvitz et Thompson, 1952). L'estimateur de variance (1.1.3) a été élaboré par Horvitz et Thompson (1952). L'autre estimateur (1.1.4) a été mis au point par Yates, Grundy et Sen (voir Cochran (1977), page 261). Les deux estimateurs sont définis seulement si $\pi_{ij} > 0 \forall i, j \in S$ et peuvent devenir négatifs. Si $\pi_{ij} < \pi_i \pi_j$ $i, j \in U$, alors $v_{YGS} > 0$ pour tout S . De ce point de vue, l'estimateur v_{YGS} est plus stable que l'estimateur v_{HT} dans plusieurs situations.

1.1.2. Estimateur d'Horvitz-Thompson sous le plan PIPT

Nous débutons cette sous-section en considérant la définition d'une variable auxiliaire, c'est-à-dire toute variable pour laquelle l'enquêteur possède de l'information avant une enquête (sondage).

Ordinairement, il est supposé *a priori* que l'information sur la variable auxiliaire est complète. La valeur de la variable, disons x , est connue pour chacune des N unités de la population. Le but est d'utiliser cette variable (ou ces variables) pour obtenir une meilleure estimation du total ou de la moyenne de la population de la variable d'intérêt y . Une (ou plusieurs) variable(s) auxiliaire(s) disponible(s) avant le sondage peut (peuvent) être utilisée(s) de différentes façons : création des plans d'échantillonnage dont les π_i en dépendent, formation de strates, formulation des estimateurs qui en dépendent explicitement (estimateur par la régression) ou encore aux deux étapes, c'est-à-dire à l'étape du plan d'échantillonnage et dans la formulation de l'estimateur (estimateur HTR, voir section 2).

Un plan d'échantillonnage bien connu qui utilise l'information d'une variable auxiliaire est le plan PIPT (probabilité d'inclusion proportionnelle à la taille). Dans ce plan, on suppose que les π_i sont proportionnelles à une certaine variable auxiliaire positive connue, x_i ($i \in U$) qui est souvent une mesure de taille. Sous ce plan, on devrait s'attendre à une réduction de la variance de l'estimateur T_{HT} . Dans un tel cas, on peut montrer que pour $\pi_i = nx_i/x_U$ ($x_U = \sum_{i=1}^N x_i$), alors l'estimateur HT peut s'écrire :

$$T_{HT} = \bar{x}_U \frac{\sum_{i=1}^n y_i/\pi_i}{\sum_{i=1}^n x_i/\pi_i},$$

où $\bar{x}_U = x_U/N$. Des estimateurs de la variance de cet estimateur sont immédiats en remplaçant les π_i dans l'équation (1.1.3) ou (1.1.4).

1.1.3. Estimateur Horvitz-Thompson comme fonctionnelle de moindres carrés

Lorsque les π_i sont exactement proportionnelles à y_i , la variance de l'estimateur HT est égale à zéro.

Lemme 1.1.1. *Si $\pi_i = Cy_i \forall i \in U$, où $C > 0$, alors $V_p(HT) = 0$.*

Démonstration. En remplaçant les $\frac{y_i}{\pi_i}$ par $\frac{1}{C}$ dans l'équation (1.1.2), on obtient

$$\begin{aligned} N^2V_p(HT) &= C^{-2} \left(\sum_{i \in U} \pi_i(1 - \pi_i) + \sum_{i \neq j \in U} \pi_{ij}(1 - \pi_i\pi_j/\pi_{ij}) \right) \\ &= C^{-2} \left(\sum_{i \in U} \pi_i - \sum_{i \in U} \pi_i^2 + \sum_{i \neq j \in U} \pi_{ij} - \sum_{i \neq j \in U} \pi_i\pi_j \right) \\ &= C^{-2} \left(\sum_{i \in U} \pi_i - \sum_{i \neq j \in U} \pi_{ij} - \left(\sum_{i \in U} \pi_i \right)^2 \right). \end{aligned}$$

Comme $\sum_{i \in U} \pi_i = n$ et $\sum_{i \neq j \in U} \pi_{ij} = n(n-1)$ (voir, Särndal, Swensson et Wretman (1992), page 38), nous obtenons $N^2V_p(HT) = C^{-2}(n + n(n-1) - n^2) = 0$. \square

Dans un tel cas, $T_{HT}(y_s) = \bar{y}_U$ pour chaque échantillon s de S . Tout en étant robuste en distortion (sans biais), HT est non robuste en ce qui concerne la variance pour les écarts par rapport à la proportionnalité entre y_i et π_i . Par exemple, si $\pi_i = y_i \forall i \neq 1$ et $\pi_1 \neq y_1$, en procédant de façon similaire que dans la démonstration du lemme 1.1.1, nous obtenons $N^2V_p(HT) = \pi_1(1 - \pi_1)[1 + (\frac{y_2}{\pi_1})^2] + \left(\sum_{1 \neq i} \pi_{1i} - \pi_1 \sum_{1 \neq i} \pi_i \right) (1 + \frac{y_1}{\pi_1})$. Après quelques manipulations algébriques, nous avons $N^2V_p(HT) = (1 - \pi_1)y_1(\frac{y_1}{\pi_1} - 1)$. Donc, il suffit d'avoir y_1 grand pour que la variance de l'estimateur HT soit élevée.

Hulliger (1995) a formulé l'estimateur HT sous la forme d'une fonctionnelle de moindres carrés (MC). À l'aide d'une estimation de la fonction de distribution de la population, on peut trouver une fonction d'influence analogue à celle pour les populations infinies (voir Hampel, 1974). Une telle idée permet d'incorporer le plan dans l'estimateur de cette fonction tout en gardant la proportionnalité de y_i et de x_i dans la fonctionnelle de moindres carrés. Nous verrons plus loin que l'estimateur HT est un estimateur de moindres carrés pondérés.

Avant d'établir une fonctionnelle de moindres carrés, nous avons besoin des définitions suivantes.

Définition 1.1.3. *Supposons des vecteurs aléatoires indépendants $(X_i, Y_i) \in \mathbb{R}^2$, $i=1, 2, \dots, N$. La fonction de distribution conjointe de la population des réalisations des vecteurs (x_i, y_i) , notée $F_U(r, t)$, est donnée par*

$$F_U(r, t) = \frac{1}{N} \sum_{i \in U} I\{x_i \leq r\} I\{y_i \leq t\},$$

où $I\{z \leq A\} = 1$ si $z \leq A$ et, zéro sinon.

Il existe divers moyens d'estimer $F_U(r, t)$ (voir Rao, Kovar et Mantel, 1990), mais l'estimateur le plus simple et le plus généralement applicable est la fonction de distribution échantillonnale (fonction de distribution empirique pondérée) qui est donnée dans la définition suivante.

Définition 1.1.4. *Soit $Y_i \in \mathbb{R}$, ($i=1, \dots, n$), des variables aléatoires indépendantes (un échantillon s de U) et soit X , une matrice de nombres réels connus de dimension $n \times p$, $p \leq n$ telle que X soit de plein rang p . Dénotons par x_i $i = 1, \dots, n$,*

les lignes de X écrites comme un vecteur colonne $p \times 1$. Supposons que

$$E[Y|X] = X\beta, \text{ et } \text{Var}(Y_i) = \sigma^2 v(x_i),$$

pour $Y^t = (Y_1, \dots, Y_n)$, $\beta \in \mathbb{R}^p$ inconnu, $\sigma^2 \in \mathbb{R}_+$ et une certaine fonction $v : \mathbb{R}^p \rightarrow \mathbb{R}$. La fonction de distribution empirique pondérée est définie comme

$$F_S(r, t) = \sum_{i=1}^n \frac{1}{\pi_i} I\{x_i \leq r\} I\{y_i \leq t\} / \sum_{i=1}^n \frac{1}{\pi_i},$$

où $I\{x \leq r\} = 1$ si $x_j \leq r_j \forall j = 1, \dots, p$ et, zéro sinon.

On peut montrer que l'estimateur des moindres carrés pondérés β_{MC} de β correspondant à F_S est la valeur de β qui minimise

$$\int \frac{(y - x^T \beta)^2}{v(x)} dF_S(x, y),$$

ou d'une façon équivalente, qui minimise

$$\sum_{i \in S} \frac{1}{\pi_i} \zeta \left(\frac{y_i - x_i^T \beta}{\sqrt{v(x_i)}} \right) / \sum_{i=1}^n \frac{1}{\pi_i} \quad (1.1.5)$$

avec $\zeta(r) = r^2$ comme fonction de poids. L'estimateur des moindres carrés (unique) est

$$\beta_{MC}(F_S) = (X^T A X)^{-1} X^T A Y, \quad (1.1.6)$$

où A est une matrice diagonale avec comme éléments diagonaux $1/[\pi_i v(x_i)]$ et S l'ensemble des échantillons.

Si les Y_i sont i.i.d. (indépendantes et identiquement distribuées) et $v(x) = 1$ alors β_{MC} est le meilleur estimateur linéaire sans biais de β (voir Royall, 1970).

On remarque que l'estimateur de F_U est lui-même une fonction de distribution. C'est un estimateur par ratio dont le bias est d'ordre $O(1/n)$ (voir Cochran, 1977, page 160)

Théorème 1.1.1. (*estimateur HT comme fonctionnelle de moindres carrés*) Soit une variable auxiliaire $x_i > 0$ connue pour tout $i \in U$. Soit S l'ensemble des échantillons de taille n d'un plan d'échantillonnage PIPT avec $\pi_i = nx_i/x_U$. Supposons que la caractéristique y est un vecteur de réalisations des variables du modèle (superpopulation) :

$$Y_i = \beta x_i + E_i, \mathbf{E}(E_i) = 0 \text{ et } \text{Var}(E_i) = \sigma^2 v(x_i), i \in U,$$

avec $\beta \in \mathbb{R}$ et σ^2 inconnus. Supposons la fonction de variance $v(x_i) = x_i$. L'estimateur HT est alors donné par :

$$T_{HT} = \bar{x}_U \beta_{HT}(F_S), \tag{1.1.7}$$

où $\beta_{HT}(F_S)$ est l'estimateur des moindres carrés pondérés de β correspondant à la fonction de distribution F_S .

Il faut noter que $\beta_{HT}(F_S)$ est tout simplement une nouvelle façon de noter β_{MC} par analogie avec $\beta_{HT}(F_S, \eta)$ (voir section 1.2).

Démonstration. Les poids de la fonction de distribution F_S sont les probabilités d'inclusion π_i . De plus, l'estimateur des moindres carrés pondérés correspondant

à F_S est

$$\begin{aligned}
 \beta_{HT}(F_S) &= \beta_{MC}(F_S) \\
 &= \sum_S \frac{y_i x_i}{\pi_i v(x_i)} / \sum_S \frac{x_i^2}{\pi_i v(x_i)} \\
 &= \sum_S \frac{y_i x_i x_U}{n x_i^2} / \sum_S \frac{x_i^2 x_U}{n x_i^2} \\
 &= \sum_S \frac{y_i}{n x_i} \\
 &= \sum_S \frac{y_i}{\pi_i x_U}.
 \end{aligned}$$

Finalement, $\bar{x}_U \beta_{HT}(F_S) = \frac{1}{N} \sum_S y_i / \pi_i = T_{HT}$. \square

Il faut noter que le modèle de superpopulation dans le théorème n'est pas nécessaire à l'obtention $T_{HT} = \bar{x}_U \beta_{HT}(F_S)$ lorsque $\beta_{HT}(F_S)$ est défini directement comme le minimum de $\sum_S \frac{1}{\pi_i} \zeta \left(\frac{y_i - \beta x_i}{\sqrt{x_i}} \right)$. Cependant, ce modèle met en évidence le rôle de l'estimateur HT : la pente $\beta(F_S)$ utilisée dans la formule de l'estimateur HT est un estimateur des moindres carrés généralisés qui incorpore l'information du plan par l'intermédiaire de F_S ainsi que l'information de la variable auxiliaire par l'intermédiaire du modèle de régression à l'origine.

1.2. ROBUSTIFICATION DE L'ESTIMATEUR HT

1.2.1. Introduction

En considérant le modèle linéaire du théorème 1.1.1 et après avoir séparé le plan d'échantillonnage de l'information auxiliaire, on peut accroître la robustesse de l'estimateur HT en remplaçant β_{MC} de l'équation (1.1.7) par un estimateur robuste de β . Pour y arriver, nous pouvons utiliser un des estimateurs robustes analysés dans Hampel, Ronchetti, Rousseeuw et Stahel (1986) ou Rousseeuw et

Leroy (1987).

Essentiellement, ces estimateurs sont obtenus en remplaçant $\zeta(r)$ de l'équation (1.1.5) par une autre fonction. Par exemple, on pourrait utiliser $\zeta(r) = |r|$ au lieu de $\zeta(r) = r^2$, ce qui revient à remplacer β_{MC} par l'estimateur robuste L_1 (estimateur des moindres valeurs absolues, voir Bloomfield et Steige, 1983) de β . Cependant, cet estimateur gagne en robustesse par rapport à y dans le sens qu'il est moins influencé par les valeurs aberrantes en y mais pas par rapport à x . De plus, l'efficacité, défini par le rapport d'écart quadratique moyen de l'estimateur L_1 diminue quand n augmente (voir Wilson, 1978).

Une deuxième classe d'estimateurs bien connus est la classe des M -estimateurs. Ces estimateurs robustes sont à la fois simples et souples (voir Hulliger, 1995).

Définition 1.2.1. *Un M – estimateur T_n est la solution de l'équation d'estimation*

$$\sum_{i=1}^n \psi(X_i - T_n) = 0,$$

pour une fonction ψ déterminée à l'avance.

En réalité, ceci revient à remplacer $\zeta(r)$ par $\rho(r)$ dans l'équation (1.1.5) et poser $\psi(r) = \partial\rho(r)/\partial r$.

Une fonction ψ souvent considérée est celle de Huber, $\psi_{Hub}(x, k) = \max(-k, \min(k, x))$, où k est une constante de mise au point.

Pour la robustification de l'estimateur HT , nous nous concentrons à présent sur cette classe de M -estimateurs.

1.2.2. Estimateurs HTR

La robustification de l'estimateur HT s'obtient de façon analogue à celle des M -estimateurs dans le modèle linéaire pour les populations infinies. Comme alternative à $\beta_{HT}(F_S)$, Hulliger (1991) a proposé de poser $\beta_{HT}(F_S, \eta)$ comme étant la solution de

$$\int \eta \left(\frac{x}{\sqrt{v(x)}}, \frac{y - \beta x}{\sqrt{v(x)}} \right) \frac{x}{\sqrt{v(x)}} dF_S(x, y) = 0,$$

ou d'une façon équivalente

$$\sum_S \frac{1}{\pi_i} \eta \left(\frac{x_i}{\sqrt{v(x_i)}}, \frac{y_i - \beta x_i}{\sqrt{v(x_i)}} \right) \frac{x_i}{\sqrt{v(x_i)}} = 0,$$

où $\eta(x, r)$ est une fonction définie à l'avance.

Définition 1.2.2. (*estimateur HTR*)

L'estimateur HT robustifié, noté HTR , est \bar{x}_U fois une solution $\beta_{HT}(F_S, \eta)$ de l'équation

$$\sum_S \frac{1}{\pi_i} \eta \left(\frac{x_i}{\sqrt{x_i}}, \frac{y_i - \beta x_i}{\sqrt{x_i}} \right) \frac{x_i}{\sqrt{x_i}} = 0, \quad (1.2.1)$$

où la fonction $\eta(x, r) : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ satisfait aux conditions de régularité de Hampel (voir Hullinger (1991), page 29).

Pour faciliter la notation, posons à présent $x' = x/\sqrt{v(x)}$, $y' = y/\sqrt{v(x)}$ et $r'(\beta) = (y - \beta x)/\sqrt{v(x)}$. Plusieurs fonctions η peuvent être considérées, mais toutes ces fonctions peuvent être écrites sous la forme de $\eta(x, r) = w(x, k_x)\psi(r(k_r) \cdot u(x))$, où $w(x)$ et $u(x)$ sont deux fonctions de pondération, et ψ est une fonction de définition d'un M -estimateur ponctuel (voir Hampel, Ronchetti, Rousseeuw et Stahel (1986), page 315). Dans tout ce qui suit, la forme de "Mallows" avec

$u(x) \equiv 1$ est utilisée. Ce choix de $u(x)$ permet de donner un poids moindre aux valeurs x aberrantes et aux valeurs résiduelles aberrantes de façon indépendante (voir Hulliger, 1995). L'estimateur HT est un cas particulier de l'estimateur HTR pour laquelle $\eta(x, r) \equiv r$ pour tout x . Ainsi, si on prend la fonction de Huber $\eta(x, r) = \psi_{Hub}(r, k_r) = \max(-k_r, \min(k_r, r))$ ($w(x, k_x) \equiv 1$), pour une certaine constante de mise au point k_r , il est possible d'obtenir des estimateurs de plus en plus robustes en ajustant cette constante à la baisse.

Hulliger (1991) a proposé d'utiliser des fonctions $w(x, k_x)$ et $\psi(r, k_r)$ qui dépendent des constantes d'ajustement k_x et k_r respectivement. Alors la robustification de l'estimateur HT du type de Mallows est $\bar{x}_u \beta_{HT}(F_S, \eta)$, où $\beta_{HT}(F_S, \eta)$ est une solution de $\sum_S \frac{1}{\pi_i} w(x'_i, k_x) \psi(r'(\beta), k_r) x'_i = 0$. Pour que $\beta_{HT}(F_S, \eta)$ soit équivariant aux changements d'échelle, il faut introduire des estimateurs d'échelle dans les fonctions $w(x, k_x)$ et $\psi(r, k_r)$. On peut prendre le MAD ($\text{med}_i\{|x_i - \text{med}_j(x_j)|\}$) des x_i comme estimateur d'échelle pour la fonction de pondération $w(x, k_x)$ et le MAD des résidus absolus pour estimer l'échelle des résidus. Pour faciliter les calculs dans le développement théorique qui suit, nous supposons que l'échelle est connu.

Sous certaines conditions, d'autres formulations de l'estimateur HT peuvent s'avérer appropriées. Par exemple, si on présume que y_i est proportionnel à x_i , mais avec une variance proportionnelle au carré de x_i (modèle sur lequel l'estimateur HT est optimal, voir Hulliger (1995), page 92), l'accroissement de la robustesse correspondant est β solution de $\sum_{i \in S} \eta(x_i, y_i/x_i - \beta) = 0$. Cet accroissement ignore le plan d'échantillonnage et si on intègre le plan dans l'équation

en résolvant l'équation $\sum_{i \in S} \eta(x_i, y_i/x_i - \beta)/\pi_i = 0$, l'estimateur HT n'est pas récupéré quand $\eta(x, r) \equiv r$.

1.2.3. Espérance et variance de HTR

Pour obtenir une variance d'échantillonnage de l'estimateur HTR , on a besoin d'utiliser une fonction d'influence pour les populations finies analogue à la fonction d'influence pour les populations infinies (voir Hampel, 1974).

Définition 1.2.3. *La fonction d'influence (IF) d'un estimateur T par rapport à F_U est donnée par*

$$IF(y; T, F_U) = \lim_{t \rightarrow 0^+} \frac{T[(1-t)F_U + t\Delta_y] - T(F_U)}{t},$$

où $T(F_U)$ est un estimateur de β basé sur F_U et Δ_y est une mesure de Dirac à y .

En posant $t = \frac{1}{N+1}$, on réalise que $IF(y; T, F_U)$ mesure approximativement $N + 1$ fois le changement de l'estimateur T causé par une observation additionnelle y .

On peut partir directement du modèle de superpopulation et essayer de travailler avec la fonction d'influence, mais on perdra la concordance avec le plan d'échantillonnage. Pour préserver cette concordance avec les poids échantillonnage, il convient de définir une courbe de sensibilité (version renommée de la fonction d'influence) de la fonction de distribution de la population, F_U (voir Hulliger, 1991).

Définition 1.2.4. *Soit $\beta_{HT}(F_U, \eta)$ une solution de*

$$\int \eta(x', r'(\beta)) x' dF_U(x, y) = 0,$$

ou de façon équivalente

$$\sum_U \frac{1}{N} \eta(x'_i, r'_i(\beta)) x'_i = 0. \quad (1.2.2)$$

La courbe de sensibilité de $\beta_{HT}(F_U, \eta)$, notée $CS(x, y, \beta_{HT}(F_U, \eta))$, est donnée par :

$$CS(x, y, \beta_{HT}(F_U, \eta)) = (N + 1)[\beta_{HT}(F_{U+}, \eta) - \beta_{HT}(F_U, \eta)], \quad (1.2.3)$$

où $U+$ est la population U augmentée d'une unité présentant les caractéristiques (x, y) .

Cette courbe mesure l'influence de l'ajout d'une nouvelle observation sur l'estimateur $\beta_{HT}(F_U, \eta)$.

Proposition 1.2.1. *La courbe de sensibilité de $\beta_{HT}(F_U, \eta)$, $CS(x, y, \beta_{HT}(F_U, \eta))$, peut s'écrire sous la forme de*

$$\begin{aligned} CS(x, y, \beta_{HT}(F_U, \eta)) &= (N + 1)[\beta_{HT}(F_{U+}, \eta) - \beta_{HT}(F_U, \eta)] \\ &\approx \frac{\sigma \eta(x', r'(\beta_{HT}(F_U, \eta))) x'}{\frac{1}{N} \sum_U \eta_2(x'_i, r'_i(\beta_{HT}(F_U, \eta))) x_i'^2}, \end{aligned} \quad (1.2.4)$$

où $\eta_2(x, r) = \partial \eta(x, r) / \partial r$.

Démonstration. En utilisant Hullinger (1991), section 4.2.2, l'équation (1.2.3) peut s'écrire sous la forme de l'équation implicite suivante :

$$(N + 1) \left[\frac{1}{N + 1} \left\{ \sum_U \eta(x'_i, r'_i(\beta_{HT}(F_{U+}, \eta))) x'_i + \eta(x', r'(\beta_{HT}(F_{U+}, \eta))) x' \right\} - \frac{1}{N} \sum_U \eta(x'_i, r'_i(\beta_{HT}(F_U, \eta))) x'_i \right] = 0.$$

En procédant de façon similaire à l'approximation de la courbe de sensibilité de l'estimateur du risque estimé minimum (estimateur REM) pour les populations infinies (voir Hulliger (1991), Annexe A.1), nous obtenons l'approximation suivante :

$$\begin{aligned}
& (N+1) \sum_U \frac{1}{N+1} \eta(x'_i, r'_i(\beta_{HT}(F_{U+}, \eta))) x'_i \\
& - \frac{N+1}{N} \sum_U \eta(x'_i, r'_i(\beta_{HT}(F_U, \eta))) x'_i \\
& + \eta(x', r'(\beta_{HT}(F_{U+}, \eta))) x' \\
\approx & (N+1) \sum_U \frac{1}{N+1} [\eta(x'_i, r'_i(\beta_{HT}(F_U, \eta))) x'_i \\
& - [\beta_{HT}(F_{U+}, \eta) - \beta_{HT}(F_U, \eta)] \eta_2(x'_i, r'_i(\beta_{HT}(F_U, \eta))) x'^2 / \sigma] - 0 \\
& + \eta(x', r'(\beta_{HT}(F_U, \eta))) x' - (N+1) \frac{1}{N+1} [\beta_{HT}(F_{U+}, \eta) - \beta_{HT}(F_U, \eta)] \times \\
& \eta_2(x', r'(\beta_{HT}(F_U, \eta))) x'^2 / \sigma \\
= & 0
\end{aligned}$$

ou d'une façon équivalente

$$\begin{aligned}
(N+1)(\beta_{HT}(F_{U+}, \eta) - \beta_{HT}(F_U, \eta)) = \\
\frac{\sigma \eta(x', r'(\beta_{HT}(F_U, \eta))) x'}{\frac{1}{N+1} \sum_U \eta_2(x'_i, r'_i(\beta_{HT}(F_U, \eta))) x_i'^2 + \frac{1}{N+1} \eta_2(x', r'(\beta_{HT}(F_U, \eta))) x'^2}.
\end{aligned}$$

Quand N est grand, $\frac{1}{N+1} \eta_2(x', r'(\beta_{HT}(F_U, \eta))) x'^2$ est négligeable par rapport à $\frac{1}{N+1} \sum_U \eta_2(x'_i, r'_i(\beta_{HT}(F_U, \eta))) x_i'^2$ et $\frac{1}{N+1} \approx \frac{1}{N}$. Alors nous avons

$$(N+1)(\beta_{HT}(F_{U+}, \eta) - \beta_{HT}(F_U, \eta)) \approx \frac{\sigma \eta(x', r'(\beta_{HT}(F_U, \eta))) x'}{\frac{1}{N} \sum_U \eta_2(x'_i, r'_i(\beta_{HT}(F_U, \eta))) x_i'^2}. \quad \square$$

L'espérance du membre de gauche de l'équation de définition (1.2.1) de $\beta_{HT}(F_S, \eta)$ est égale à la partie de gauche de l'équation de définition (1.2.2)

de $\beta_{HT}(F_U, \eta)$. Alors $\beta_{HT}(F_S, \eta)$ est convergent et son biais est donné dans le théorème suivant.

Théorème 1.2.1. *Soit la variance résiduelle $\sigma = 1$, et supposons que $\beta_{HT}(\cdot, \eta)$ est continue à F_U . L'estimateur HTR est convergent sous le plan d'échantillonnage, i.e. $\mathbf{E}_S \beta_{HT}(F_S, \eta) - \beta_{HT}(F_U, \eta) = O(1/n)$.*

Pour faciliter la notation, nous utilisons les expressions suivantes

$$\begin{aligned}\eta_i(U) &= \eta(x'_i, r'_i(\beta_{HT}(F_U, \eta))), \\ \eta_{2i}(U) &= \eta_2(x'_i, r'_i(\beta_{HT}(F_U, \eta))), \\ \tilde{\eta}_i(U) &= \eta_i(U)x'_i, \\ \tilde{\eta}_{2i}(U) &= \eta_{2i}(U)x_i'^2, \\ \check{\eta}_i(U) &= \tilde{\eta}_i/\pi_i, \\ \check{\eta}_{2i}(U) &= \tilde{\eta}_{2i}(U)/\pi_i.\end{aligned}$$

Une notation analogue sera utilisée avec l'indice S à la place de U pour indiquer le remplacement de F_U par F_S .

Démonstration. Considérons une approximation linéaire de (1.2.1) autour de $\beta_{HT}(F_U, \eta)$:

$$\begin{aligned}0 &= \sum_S \frac{1}{\pi_i} [\tilde{\eta}_i(U) - (\beta_{HT}(F_S, \eta) - \beta_{HT}(F_U, \eta))\tilde{\eta}_{2i}(U)] \\ &\quad + o(\beta_{HT}(F_S, \eta) - \beta_{HT}(F_U, \eta)).\end{aligned}$$

Maintenant, F_S étant un estimateur convergent de F_U et puisque $\beta_{HT}(\cdot, \eta)$ est supposé continu en F_U , alors la différence $\beta_{HT}(F_S, \eta) - \beta_{HT}(F_U, \eta)$ tend vers zéro quand n est grand.

Quand $\sum_S \frac{1}{\pi_i} \tilde{\eta}_{2i}(U) \neq 0$, il s'en suit que

$$\begin{aligned} \mathbf{E}_S[(\beta_{HT}(F_S, \eta) - \beta_{HT}(F_U, \eta))] &= \mathbf{E}_S \left[\frac{\sum_S \frac{1}{\pi_i} \tilde{\eta}_i(U)}{\sum_S \frac{1}{\pi_i} \tilde{\eta}_{2i}(U)} \right] \\ &+ \mathbf{E}_S \left[\frac{O(\beta_{HT}(F_S, \eta) - \beta_{HT}(F_U, \eta))}{\sum_S \frac{1}{\pi_i} \tilde{\eta}_{2i}(U)} \right]. \end{aligned}$$

Le premier terme du membre de droite est un estimateur par ratio de $0 = \sum_U \tilde{\eta}_i(U) / \sum_U \tilde{\eta}_{2i}(U)$ et son biais est d'ordre $O(1/n)$ (voir Cochran (1977), page 160) et le second terme du membre de droite est d'ordre plus petit. Par conséquent le résultat découle. \square

Théorème 1.2.2. Soit $V_r(F_U, \eta)$ la variance de l'estimateur HTR. Alors

$$V_r(F_U, \eta) / \bar{x}_U^2 \approx \frac{\sigma^2 \left(\sum_U \frac{1}{\pi_i} \tilde{\eta}_i(U)^2 + \sum_U \frac{\pi_{ij}}{\pi_i \pi_j} \tilde{\eta}_i(U) \tilde{\eta}_j(U) \right)}{\sum_U \tilde{\eta}_{2i}(U)^2 + \sum_U \tilde{\eta}_{2i}(U) \tilde{\eta}_j(U)}. \quad (1.2.5)$$

Démonstration. Comme dans Hampel, Ronchetti, Rousseeuw et Stahel (1986), page 85, en remplaçant la fonction d'influence par la courbe de sensibilité donnée à l'équation (1.2.4), on peut estimer $\beta_{HT}(F_S, \eta) - \beta_{HT}(F_U, \eta)$ par

$$\begin{aligned} \beta_{HT}(F_S, \eta) - \beta_{HT}(F_U, \eta) &\approx \int CS(x, y, F_U, \eta) dF_S(x, y) \\ &= \sigma \left(\frac{\sum_S \frac{1}{\pi_i} \eta(x'_i, r'_i(\beta_{HT}(F_U, \eta))) x'_i}{\sum_S \frac{1}{\pi_i}} \right) \\ &\quad / \left(\frac{\sum_U \frac{1}{N} \eta_{2i}(x'_i, r'_i(\beta_{HT}(F_U, \eta))) x_i'^2}{\sum_U \frac{1}{N}} \right) \\ &= \sigma \left(\frac{\sum_S \tilde{\eta}_i(U) / \sum_S \frac{1}{\pi_i}}{\sum_U \tilde{\eta}_{2i}(U) / N} \right). \end{aligned}$$

La somme $\sum_S \frac{1}{\pi_i}$ est la constante de normalisation de la fonction de distribution échantillonnale. Son espérance est $\mathbf{E}_S \sum_S \frac{1}{\pi_i} = N$. Supposons que $\sum_S \frac{1}{\pi_i} \approx N$, alors $Var_S(\beta_{HT}(F_S, \eta))$ est donné par

$$\begin{aligned} \mathbf{E}_S[(\beta_{HT}(F_S, \eta) - \beta_{HT}(F_U, \eta))^2] &\approx \mathbf{E}_S \left[\sigma \left(\frac{\sum_S \tilde{\eta}_i(U)}{\left(\sum_U \tilde{\eta}_{2i}(U) \right)} \right)^2 \right] \\ &= \sigma^2 \mathbf{E}_S \left(\frac{\left(\sum_S \tilde{\eta}_i(U) \right)^2}{\left(\sum_U \tilde{\eta}_{2i}(U) \right)^2} \right) \\ &= \frac{\sigma^2 \left(\sum_U \frac{1}{\pi_i} \tilde{\eta}_i(U)^2 + \sum_U \frac{\pi_{ij}}{\pi_i \pi_j} \tilde{\eta}_i(U) \tilde{\eta}_j(U) \right)}{\sum_U \tilde{\eta}_{2i}(U)^2 + \sum_U \tilde{\eta}_{2i}(U) \tilde{\eta}_{2j}(U)}. \square \end{aligned}$$

Ce cas diffère surtout de celui de la variance asymptotique d'un M – estimateur avec des variables aléatoires indépendantes et identiquement distribuées du fait que les produits croisés du numérateur de (1.2.5) ne disparaissent pas. Si $\eta(x, r) \equiv r$, alors $V_r(F_U, \eta)$ donne la variance de l'estimateur HT .

1.2.4. Estimation de la variance de l'estimateur HTR

L'estimateur naturel du numérateur de $V_r(F_U, \eta)$ pourrait être $[\sum_S \frac{1}{\pi_i} \eta(x'_i, r_i(\beta(F_S, \eta)))x'_i]^2$. Mais cette somme est nulle étant donné la définition de l'estimateur HTR . L'astuce pour arriver à un estimateur du numérateur de $V_r(F_U, \eta)$ est de soustraire $0 = [\sum_U \tilde{\eta}_i(U)]^2$ pour arriver à

$$\begin{aligned} \sum_U \frac{1}{\pi_i} \tilde{\eta}_i(U)^2 + \sum_U \frac{\pi_{ij}}{\pi_i \pi_j} \tilde{\eta}_i(U) \tilde{\eta}_j(U) - \left(\sum_U \tilde{\eta}_i(U) \right)^2 \\ = \sum_U \frac{1 - \pi_i}{\pi_i} \tilde{\eta}_i(U)^2 + \sum_U \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \tilde{\eta}_i(U) \tilde{\eta}_j(U). \end{aligned}$$

C'est la variance de «l'estimateur» $\sum_S \tilde{\eta}_i(U)/\pi_i$, lequel est une somme pondérée comme l'estimateur HT . Par conséquent, on peut utiliser l'estimateur de la

variance (1.1.3) de l'estimateur HT . L'estimateur de variance de «l'estimateur» $\sum_S \tilde{\eta}_i(U)/\pi_i$ est donc

$$\sum_S \frac{1 - \pi_i}{\pi_i^2} \tilde{\eta}_i(U)^2 + \sum_S \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j \pi_{ij}} \tilde{\eta}_i(U) \tilde{\eta}_j(U).$$

Après avoir remplacé $\beta_{HT}(F_U, \eta)$ par $\beta_{HT}(F_S, \eta)$ et une réorganisation des termes, nous avons

$$\begin{aligned} \left(\sum_S \frac{\tilde{\eta}_i(S)}{\pi_i} \right)^2 - \left(\sum_S \frac{\tilde{\eta}_i(S)}{\pi_i} + \sum_S \frac{\tilde{\eta}_i(S) \tilde{\eta}_j(S)}{\pi_{ij}} \right) \\ = 0 - \sum_{i \in S} \frac{1}{\pi_i} \eta(x'_i, r'_i)^2 x_i'^2 + \sum_{i \neq j \in S} \frac{1}{\pi_{ij}} \eta(x'_i, r'_i) x'_i \eta(x'_j, r'_j) x'_j. \end{aligned}$$

Le dénominateur de $V_r(F_U, \eta)$ peut être estimé par

$$\sum_{i \in S} \frac{1}{\pi_i} \eta_2(x'_i, r'_i)^2 x_i'^4 + \sum_{i \neq j \in S} \frac{1}{\pi_{ij}} \eta_2(x'_i, r'_i) x_i'^2 \eta_2(x'_j, r'_j) x_j'^2.$$

L'estimateur de la variance de l'estimateur HTR est donc

$$v_{rHTR} = -\bar{x}_U^2 \frac{\sum_{i \in S} \frac{1}{\pi_i} \eta(x'_i, r'_i)^2 x_i'^2 + \sum_{i \neq j \in S} \frac{1}{\pi_{ij}} \eta(x'_i, r'_i) x'_i \eta(x'_j, r'_j) x'_j}{\sum_{i \in S} \frac{1}{\pi_i} \eta_2(x'_i, r'_i)^2 x_i'^4 + \sum_{i \neq j \in S} \frac{1}{\pi_{ij}} \eta_2(x'_i, r'_i) x_i'^2 \eta_2(x'_j, r'_j) x_j'^2}, \quad (1.2.6)$$

où $\eta_2 = \frac{\partial}{\partial r} \eta(x, r)$. Le signe moins dans l'équation (1.2.6) est approprié car les produits (négatifs) croisés dans le numérateur dominant habituellement (voir Hülliger, 1995). Néanmoins, v_{rHTR} peut devenir négatif puisque les termes $(\pi_i \pi_j - \pi_{ij})$ varient beaucoup et peuvent devenir négatifs pour certains échantillons, tout comme pour l'estimateur de variance HT (1.1.3) (voir Cochran (1977), page 261). L'estimateur de variance v_{rHTR} ne donne pas l'estimateur de variance (1.1.2) si $\eta(x, r) \equiv r$.

On peut évidemment utiliser un deuxième estimateur, soit celui de Yates-Grundy-Sen (1.1.4) pour faire l'estimation du numérateur de $V_r(F_U, \eta)$. Un troisième estimateur de variance peut être obtenu en assimilant l'estimateur HTR à un estimateur par les moindres carrés pondérés dont les poids dépendent de l'estimation (voir Hulliger (1991), page 166).

1.3. ESTIMATEURS À RISQUE ESTIMÉ MINIMUM

(ESTIMATEURS *REM*)

1.3.1. Motivation des estimateurs *REM*

L'estimateur HT est exprimé sous la forme d'une fonctionnelle de moindres carrés pondérés d'un modèle de régression passant par l'origine pour permettre sa robustification. Pour l'estimateur $\beta(F_S, \eta)$ du type de Mallows, deux fonctions ont besoin d'être spécifiées, une fonction de poids w pour la variable auxiliaire et une autre fonction ψ pour les résidus. On doit choisir les deux constantes de mise au point k_x et k_r puisque chacune des deux fonctions w et ψ peuvent en dépendre. L'estimateur de l'écart quadratique moyen (EQM) de la pente $\beta(F_S, \eta)$ de l'estimateur HTR dépend alors de ces deux constantes : $r(F_S, k_x, k_r) = v_r(F_S, k_x, k_r) + (\beta(F_S, \eta) - \beta_{MC}(F_S))^2$. L'estimateur HTR résultant de ces deux fonctions est généralement biaisé. L'EQM, $\mathbf{E}_S(\bar{x}_U \beta_{HT}(F_S, \eta) - \bar{y}_U)^2$, de l'échantillonnage constitue donc un critère de rendement pratique. Pour les échantillons de petite à moyenne taille, les gains obtenus avec les estimateurs HTR par rapport à l'estimateur HT ne sont pas très sensibles à la méthode d'accroissement de la robustesse retenue lorsque l'échantillon comporte des valeurs aberrantes (voir Hulliger, 1995). Par contre, les pertes en matière d'EQM de certains estimateurs HTR risquent d'être énormes quand les données sont adéquates et que les échantillons sont de taille moyenne à grande. Alors, pour choisir un bon estimateur HTR ,

Hulliger (1991) a proposé les *estimateurs à risque estimé minimum* (*estimateur REM*) qui adaptent les constantes de mise au point d'un estimateur *HTR* à l'échantillon.

1.3.2. Définition de l'estimateur *REM*

Considérons l'ensemble des fonctions paramétriques $\Omega = \{\eta_k(x, r) : k \in \mathcal{K}\}$ où $\mathcal{K} \subset \mathbb{R}_+^p$ est l'ensemble des constantes de mise au point (paramètres de nuisance). Ici on prend $p = 2$ pour rendre la minimisation possible et éviter la perte d'efficacité dû à l'estimation de ces paramètres. Un sous-ensemble de fonctions de Ω (sous les mêmes conditions que les fonction η de l'estimateur *HTR*) incluant les *M-estimateurs* est l'ensemble $\mathcal{B} = \{\beta(F_S, \eta_k) : k \in \mathcal{K}\}$, où $\beta(F_S, \eta_k)$ est la pente d'un estimateur *HTR*. Posons $k = (k_x, k_r) \in \mathcal{K}$ et supposons que

$$\lim_{k_x \rightarrow \infty} \eta_{k_x}(x, r) = r,$$

$\forall (x, r)$ et k_r fixé (même hypothèse quand k_x est fixé), de sorte que l'estimateur *HT* est un élément de \mathcal{B} . L'EQM de $\beta(F_S, \eta_k)$ est estimé par:

$$r(F_S, k) = \max(v_r(F_S, k), 0) + (\beta(F_S, k) - \beta_{MC}(F_S))^2,$$

où $v_r(F_S, k)$ est l'estimateur de la variance (1.2.6) ou un autre estimateur de la variance de $\beta(F_S, \eta_k)$. Le terme $\max(v_r, 0)$ utilisé dans $r(F_S, k)$ est motivé par le fait que l'estimation de la variance (1.2.6) pourrait devenir négative.

Définition 1.3.1. *Supposons que $r(F_S, \cdot)$ présente un minimum global à $k_m = (k_{xm}, k_{rm}) \in \mathcal{K}$. Dans ce cas, l'estimateur de la moyenne de la population, \bar{y}_U est*

$$M(F_S) = \bar{x}_U \beta(F_S, \eta_{k_m}). \quad (1.3.1)$$

La recherche d'un minimum à deux dimensions dans le plan (k_x, k_r) est possible, mais plus difficile que celle à une dimension (voir Hulliger, (1991) page

170). Plusieurs voies pour restreindre la recherche en une dimension peuvent être considérées. Il peut être suffisant de choisir k_x fixe. C'est raisonnable parce qu'on connaît tous les x_i , ($i = 1, \dots, N$) et un choix rationnel de k_x avec $\mathbf{x}_S = (x_1, \dots, x_n)$ peut donner de bons résultats. Il est difficile d'utiliser la connaissance de $\mathbf{y}_S = (y_1, \dots, y_n)$ pour le choix de k_x . Une deuxième façon est de relier les deux constantes de mise au point. Le plus simple est de supposer que k_x et k_r sont liées par une relation linéaire. On peut y arriver en fixant séparément l'étendue de k_x et k_r et chercher le minimum sur la pente positive de la diagonale du rectangle résultant.

Bien entendu, les deux méthodes (recherche verticale seulement avec k_x fixé ou celle de la diagonale d'un rectangle) peuvent être utilisées en même temps afin d'optimiser la combinaison (k_{xm}, k_{rm}) .

Les estimateurs *REM* sont en général des estimateurs convergents de la moyenne d'une population (voir Hulliger (1991), chapitre 2). Ils sont équivariants d'échelle et n'ont pas besoin d'un estimateur d'échelle préliminaire si les fonctions de définition sont définies adéquatement.

En pratique, on peut résoudre le problème du non-unicité du minimum ou non atteinte du minimum sur \mathcal{K} en étudiant la fonction $r(F_S, k)$. S'il existe plusieurs minimums globaux, on peut choisir celui assorti des plus petites constantes de mise au point pour obtenir une plus grande robustesse.

Étant donné que $r(F_S, k)$ dépend de $\beta_{MC}(F_S)$, la sensibilité par rapport aux valeurs aberrantes de l'estimateur *HT* est transférée aux estimateurs *REM* et la

robustesse des estimateur HTR est encore une fois perdue. Toutefois, si l'estimateur REM doit être convergent pour la moyenne de la population, il y a aucun moyen d'éviter la non-robustesse dans la partie biaisée du risque estimé. Néanmoins, les estimateurs REM sont quantitativement moins sensibles aux valeurs aberrantes et plus efficaces que l'estimateur HT en présence des valeurs aberrantes (voir Hulliger, 1995).

Dans ce chapitre, nous avons développé deux estimateurs robustes de la moyenne d'une caractéristique y d'une population finie asymétrique. Ces estimateurs sont obtenus par une robustification de l'estimateur d'Horvitz-Thompson sous le plan d'échantillonnage à probabilités d'inclusion proportionnelles à une mesure de taille. Le prochain chapitre est consacré au développement d'estimateurs bayésiens robustes de la moyenne.

Chapitre 2

ESTIMATEURS DE BAYES

Pour atteindre notre but qui est de faire de l'inférence sur la distribution de la variable Y , une caractéristique de la population U , nous avons vu au chapitre 1 comment estimer de façon robuste le total ou la moyenne d'une population. Nous qualifions ces méthodes de *méthodes classiques*. Nous présentons maintenant d'autres méthodes pouvant servir à estimer efficacement le total ou la moyenne d'une population finie en présence de valeurs aberrantes. Ces méthodes sont obtenus à l'aide de la théorie bayésienne. Avant de présenter dans la section 5, les estimateurs bayésiens qui sont des compétiteurs des estimateurs étudiés au chapitre 1, nous donnons une brève description des principaux modèles de Bayes dans les quatre premières sections du présent chapitre. Nous commençons avec la présentation des modèles bayésiens dans la première section. Nous poursuivons avec des modèles hiérarchiques de Bayes dans la deuxième section. La troisième section est consacrée à l'analyse bayésienne empirique alors que la quatrième section traite des modèles de régression bayésiens.

2.1. MODÈLES BAYÉSIENS

2.1.1. Introduction

De façon générale, la théorie de la statistique bayésienne est basée sur le théorème de Bayes dont la démonstration a été publiée par R. Price dans Bayes

(1763) (redémontré par P.S. Laplace (1773) de manière indépendante et dans un cadre plus général). De manière générale, l'inversion des probabilités est décrite par le théorème suivant.

Théorème 2.1.1. (*Théorème de Bayes*)

Si A et E sont des événements tels que $P(E) \neq 0$ et si on connaît $P(A)$, $P(E|A)$ et $P(E|A^c)$, alors

$$P(A|E) = \frac{P(E|A)P(A)}{P(E|A)P(A) + P(E|A^c)P(A^c)},$$

où A^c désigne le complémentaire de A .

Partant de la définition (moderne) de la probabilité conditionnelle, ce théorème n'offre aucune difficulté mathématique. Il représente néanmoins un saut conceptuel majeur dans l'histoire de la statistique en tant que première formule d'inversion de probabilités (voir Robert, 1992, section 1.2).

L'innovation principale du modèle statistique bayésien dont la définition est donnée ci-après par rapport aux modèles classiques, est de proposer en plus une loi de probabilité sur les paramètres.

Définition 2.1.1. *Un modèle statistique paramétrique consiste en l'observation d'une variable aléatoire y , de loi $f(y|\beta)$, où seul le paramètre β , appartenant à un sous-ensemble Ω d'un espace vectoriel de dimension finie, n'est pas connu.*

Définition 2.1.2. *On appelle modèle statistique bayésien, la donnée d'un modèle statistique paramétrique $f(y|\beta)$, et d'une loi a priori sur le paramètre β , notée $\pi(\beta)$.*

Supposons un modèle statistique bayésien. Nous sommes maintenant en mesure de définir la notion de loi *a posteriori*.

Définition 2.1.3. *Nous appelons loi a posteriori, la loi du paramètre étant donné les observations, c'est-à-dire*

$$\pi(\beta|y) = \frac{f(y|\beta)\pi(\beta)}{m(y)}, \quad (2.1.1)$$

où $m(y)$ désigne la loi marginale de y et est donnée par

$$m(y) = \int_{\Omega} f(y|\beta)\pi(\beta)d\beta, \quad (2.1.2)$$

$f(y|\beta)$ désigne la fonction de vraisemblance des observations, $\pi(\beta)$ désigne la loi *a priori* des paramètres β , Ω désigne le support de β .

Cette loi de β conditionnellement à y est proportionnelle à la loi de y conditionnellement à β multipliée par la loi *a priori* (la marginale) de β . De plus, cette définition nous indique un peu la façon de procéder quant aux choix possibles des lois *a priori*. En effet, nous venons de voir que la loi *a posteriori* dépend de la loi marginale des observations, cette dernière s'obtient en intégrant le paramètre β de la fonction de vraisemblance des observations $f(Y|\beta)$, multipliée par la loi *a priori* des paramètres $\pi(\beta)$. Souvent le choix de la loi *a priori* est fait de façon à pouvoir calculer facilement la densité marginale des observations. Une classe de densités qui facilite énormément les calculs est donnée dans la définition suivante.

Définition 2.1.4. *Pour un modèle bayésien, une famille, \mathcal{F} , de lois sur β , est dite conjuguée si, pour toute loi a priori $\pi(\beta) \in \mathcal{F}$, la loi a posteriori, $\pi(\beta|y)$ appartient également à \mathcal{F} .*

L'approche à privilégier pour l'inférence bayésienne est celle passant par la loi *a posteriori* (voir Robert (1992), section 1.4). En effet, travaillant conditionnellement aux observations, elle suit l'idée d'inversion des probabilités qui est décrite

par le théorème de Bayes et évite un moyennage sur des valeurs non observées, ce qui est l'essence même de l'approche fréquentiste. En réalité, la loi *a posteriori* représente l'actualisation de l'information *a priori*, $\pi(\beta)$, au vu de l'information contenue dans les observations, c'est-à-dire celle contenue dans $f(y|\beta)$.

Si on dispose d'un échantillon y_1, y_2, \dots, y_n , le principe d'exhaustivité (voir Robert (1992), p. 22) permet en général de se ramener à l'équation (2.1.1). En effet si T est une statistique exhaustive pour β avec la densité $g(t|\beta)$, nous avons le résultat suivant. La démonstration est donnée dans Berger (1985), section 4.2.

Lemme 2.1.1. *Supposons que $m(t)$, la densité marginale de t , est strictement positive dans le domaine de t . Si $T(y)=t$, alors*

$$\pi(\beta|y) = \pi(\beta|t) = \frac{\pi(\beta)g(t|\beta)}{m(t)}.$$

Ce lemme se démontre facilement en utilisant le théorème de factorisation (voir Lehmann, 1983).

La principale raison de déterminer $\pi(\beta|y)$ à partir de la statistique exhaustive T est que $g(t|\beta)$ et $m(t)$ sont souvent plus faciles à manier que $f(y|\beta)$ et $m(y)$.

Exemple 2.1.1. *Supposons $Y|\beta \sim \mathcal{N}(\beta, \sigma^2)$, soit $\pi(\beta) = \mathcal{N}(\mu, \tau^2)$, où μ , σ^2 et τ^2 sont connus. Alors*

$$h(y, \beta) = \pi(\beta)f(y|\beta) = \frac{1}{2\pi\sigma\tau} \exp \left\{ -\frac{1}{2} \left[\frac{(\beta - \mu)^2}{\tau^2} + \frac{(y - \beta)^2}{\sigma^2} \right] \right\}.$$

Pour trouver $m(y)$, définissons $\rho = \frac{\tau^2 + \sigma^2}{\tau^2 \sigma^2}$. Après avoir complété le carré, nous avons

$$\frac{1}{2} \left[\frac{(\beta - \mu)^2}{\tau^2} + \frac{(y - \beta)^2}{\sigma^2} \right] = \frac{1}{2} \rho \left[\beta - \frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{y}{\sigma^2} \right) \right]^2 + \frac{(\mu - y)^2}{2(\sigma^2 + \tau^2)}.$$

On obtient alors

$$h(y, \beta) = \pi(\beta)f(y|\beta) = (2\pi\sigma\tau)^{-1} \exp \left\{ -\frac{1}{2}\rho \left[\beta - \frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{y}{\sigma^2} \right) \right]^2 \right\} \times \\ \exp \left\{ -\frac{(\mu - y)^2}{2(\sigma^2 + \tau^2)} \right\}$$

et $m(y) = \int_{-\infty}^{\infty} h(y, \beta) d\beta = (2\pi\rho)^{-1/2} (\sigma\tau)^{-1} \exp \left\{ -\frac{(\mu - y)^2}{2(\sigma^2 + \tau^2)} \right\}$, c'est à dire que $Y \sim \mathcal{N}(\mu, \tau^2 + \sigma^2)$. Comme $\pi(\beta|y) = \frac{h(y, \beta)}{m(y)}$, il s'ensuit que la loi conditionnelle de β étant donné $Y = y$ suit une loi $\mathcal{N}(\mu(y), \rho^{-1})$, où $\mu(y) = y - \frac{\sigma^2}{\sigma^2 + \tau^2}(y - \mu)$ et $\rho = \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}$.

Supposons maintenant que nous avons un échantillon $Y = (Y_1, \dots, Y_n)$ provenant d'une population de loi $\mathcal{N}(\beta, \sigma^2)$, σ^2 connu. Puisque \bar{Y} est exhaustif pour β , par le lemme 2.1.1, nous pouvons obtenir $\pi(\beta|y)$ de la façon suivante. En notant que $\bar{Y} \sim \mathcal{N}(\beta, \sigma^2/n)$, on peut conclure que la distribution a posteriori de β étant donné $\mathbf{y} = (y_1, \dots, y_n)$ est $\mathcal{N}(\mu(y), \rho^{-1})$, où $\mu(y) = \bar{y} - \frac{\sigma^2/n}{\sigma^2/n + \tau^2}(\bar{y} - \mu)$ et $\rho = \frac{n\tau^2 + \sigma^2}{\tau^2\sigma^2}$.

Disposant ainsi de cette distribution de probabilités sur β (ou d'une mesure σ -finie lorsque π est une distribution impropre, c'est-à-dire lorsque $\int_{\Omega} \pi(\beta) d\beta = +\infty$), le champ de l'inférence est beaucoup plus vaste que dans la théorie classique (voir Robert (1992), section 1.4). La connaissance de la loi permet de calculer plus facilement moyenne, médiane, mode et aussi des estimateurs de la précision des estimateurs (variance et moments d'ordre plus élevés). Nous verrons à la prochaine section que les lois *a priori* peuvent se décomposer en plusieurs niveaux suivant une structure hiérarchique.

2.1.2. Inférence statistique bayésienne : l'approche décisionnelle

Toutes décisions basées sur l'inférence statistique prises soit par le statisticien, soit par le client doivent être évaluées. Un des critères d'évaluation le plus utilisé est ce qu'on appelle coût en théorie de la décision.

Définition 2.1.5. On appelle *coût* toute fonction $L(\beta, d) : \Omega \times \mathcal{D} \rightarrow [0, \infty)$, où \mathcal{D} est l'espace des décisions et Ω est l'espace des paramètres.

Cette fonction évalue la «pénalité» résultant de l'emploi de la décision d lorsque la vraie valeur des paramètres est β . Hormis des exceptions triviales, il est impossible de minimiser une fonction de coût $L(\beta, d)$ lorsque β est inconnu (voir Robert (1992), section 2.1). Pour pallier à ce problème, il convient de définir le coût moyen *a posteriori*.

Définition 2.1.6. Le coût moyen *a posteriori*, noté $\rho(\pi, d|y)$, est donné par

$$\rho(\pi, d|y) = \mathbf{E}^{\pi(\beta|y)}[L(\beta, d)] = \int_{\Omega} L(\beta, d)\pi(\beta|y)d\beta.$$

Cette expression moyenne le coût par rapport à la loi *a posteriori* de β , conditionnellement à y . On peut également définir le risque de Bayes de la façon suivante.

Définition 2.1.7. Le risque de Bayes, noté $r(\pi, \delta)$, est donné par

$$r(\pi, \delta) = \int_{\Omega} \int_{\mathcal{Y}} L(\beta, \delta(y))f(y|\beta)dy\pi(\beta)d\beta.$$

Cette expression donne une valeur réelle et non une fonction de β ; elle permet donc la comparaison directe des estimateurs. De plus, le risque de Bayes peut être obtenu à partir du coût *a posteriori*.

Proposition 2.1.1. Un estimateur minimisant le risque bayésien $r(\pi, \delta)$ peut être obtenu en choisissant, pour chaque $y \in \mathcal{Y}$, la valeur $\delta(y)$ qui minimise le

coût *a posteriori*, $\rho(\pi, \delta|y)$, puisque

$$r(\pi, \delta) = \int_{\mathcal{Y}} \rho(\pi, d|y) m(y) dy,$$

où $\delta(y)$ est la valeur qui minimise $\rho(\pi, d|y)$, $m(y)$ est donné par l'équation (2.1.2) et \mathcal{Y} le support de y .

La démonstration de cette proposition se retrouve dans Robert (1992), page 51.

On peut maintenant définir la notion d'estimateur de Bayes associée à un coût et une distribution *a priori* de la manière suivante.

Définition 2.1.8. *On appelle estimateur de Bayes associé à une loi a priori π et un coût L , tout estimateur δ^π qui minimise $r(\pi, \delta)$.*

Corollaire 2.1.1. *Pour $y \in \mathcal{Y}$ tel que $m(y) > 0$, l'estimateur de Bayes est aussi donné par $\delta(y)$, l'argument de $\min_d \rho(\pi, d|y)$.*

La démonstration de ce corollaire découle directement de proposition précédente.

On note que dans le cas d'une loi impropre, le risque bayésien peut être infini. On définit donc l'estimateur de Bayes associé, dit *estimateur de Bayes généralisé*, comme celui qui minimise, pour chaque y le coût *a posteriori*.

La définition 2.1.8 donne ainsi une méthode constructive de détermination des estimateurs de Bayes. Les propriétés des estimateurs de Bayes sont discutées dans Robert (1992), section 2.4 et Berger (1985), chapitres 5 et 6.

2.1.3. Estimation bayésienne sous la fonction de perte quadratique

Dans l'estimation réelle du paramètre β , le coût quadratique $L(\beta, \delta) = (\beta - \delta)^2$ est le critère le plus souvent utilisé pour comparer des estimateurs. Sa popularité résulte d'une part, de sa relation avec la théorie classique des moindres carrés, d'autre part de sa simplicité car il donne en général des solutions bayésiennes plus facilement maniables que celles données par les autres fonctions de perte telle que la perte « 0-1 » (voir Berger, 1985, p. 54-55). Sous la fonction de perte quadratique, nous avons le résultat suivant.

Proposition 2.1.2. *L'estimateur de Bayes δ^π de $\beta \in \mathbb{R}$, associé à la loi a priori π et au coût quadratique, est la moyenne a posteriori, c'est-à-dire*

$$\delta^\pi(y) = \mathbf{E}^{\pi(\beta|y)}[\beta] = \frac{\int_{\Omega} \beta f(y|\beta) \pi(\beta) d\beta}{\int_{\Omega} f(y|\beta) \pi(\beta) d\beta}.$$

Démonstration. Comme

$$\mathbf{E}^{\pi(\beta|y)}[(\beta - \delta(y))^2] = \mathbf{E}^{\pi(\beta|y)}[\beta^2] - 2\delta(y)\mathbf{E}^{\pi(\beta|y)}[\beta] + \delta(y)^2,$$

le minimum du coût *a posteriori* est effectivement atteint par $\delta^\pi = \mathbf{E}^{\pi(\beta|y)}[\beta]$. \square

Exemple 2.1.2. (exemple 2.1.1 suite) Comme $\pi(\beta|y) = \mathcal{N}(\mu(y), \rho^{-1})$, l'estimateur de Bayes est $\delta^\pi = \mu(y) = \bar{y} - \frac{\sigma^2}{\sigma^2 + n\tau^2}(\bar{y} - \mu)$.

Notons que des décompositions comme celles du lemme 2.1.1 ne sont pas toujours possibles. Il faut alors appliquer des méthodes d'approximation comme celles décrites dans Robert (1992), chapitre 9. Si la loi marginale est disponible, le résultat suivant permet le calcul direct de l'estimateur de Bayes dans le cas des familles exponentielles naturelles (voir Robert (1992), section 4.2).

Lemme 2.1.2. *Soit $f(y|\beta) = h(y)e^{\beta^t y - \psi(\beta)}$, une densité appartenant à une famille exponentielle à paramétrisation naturelle, Y et $\beta \in \mathbb{R}^p$. Pour toute loi a*

priori π , la moyenne a posteriori est donnée par

$$\delta^\pi(y) = \nabla \log m(y) - \nabla \log h(y),$$

où $m(y)$ est la loi marginale, $\nabla g(\cdot)$ est le gradient de $g(\cdot)$.

Pour démontrer ce lemme, nous avons besoin d'énoncer la proposition suivante dont la démonstration est donnée dans Billingsley (1986), page 215 .

Proposition 2.1.3. *Supposons que $g(\beta, y)$ est une fonction mesurable de β pour chaque $y \in \mathcal{Y}$. Si la dérivée de $g(\beta, y)$ par rapport à y , $g'(\beta, y)$, existe dans \mathcal{Y} et $|g'(\beta, y)| \leq G(\beta)$ pour $\beta \in \Omega$ et $y \in \mathcal{Y}$, où G est intégrable, alors*

$$\frac{\partial}{\partial y} \int g(\beta, y) d\beta = \int g'(\beta, y) d\beta.$$

Démonstration du lemme 2.1.2 : Comme

$$\frac{\partial}{\partial y_i} f(y|\beta) = \beta_i h(y) e^{\beta^t y - \psi(\beta)} + \frac{\partial}{\partial y_i} h(y) e^{\beta^t y - \psi(\beta)},$$

on a

$$\begin{aligned} \mathbf{E}^{\pi(\beta|y)}[\beta_i] &= \frac{\int_{\Omega} \beta_i h(y) e^{\beta^t y - \psi(\beta)} \pi(\beta) d\beta}{m(y)} \\ &= \frac{\int_{\Omega} [\frac{\partial}{\partial y_i} f(y|\beta) - \frac{\partial}{\partial y_i} h(y) e^{\beta^t y - \psi(\beta)}] \pi(\beta) d\beta}{m(y)} \\ &= \frac{\int_{\Omega} \frac{\partial}{\partial y_i} f(y|\beta) \pi(\beta) d\beta}{m(y)} - \frac{\partial}{\partial y_i} h(y) \frac{\int_{\Omega} e^{\beta^t y - \psi(\beta)} \pi(\beta) d\beta}{m(y)} \\ &= \frac{\int_{\Omega} \frac{\partial}{\partial y_i} f(y|\beta) \pi(\beta) d\beta}{m(y)} - \frac{\partial}{\partial y_i} h(y) \frac{m(y)/h(y)}{m(y)}. \end{aligned}$$

Si $f(y|\beta)$ est bornée et si $\pi(\beta)$ est une loi propre, on peut facilement démontrer que $f(y|\beta)\pi(\beta)$ vérifie les conditions de la proposition précédente. Nous trouvons

alors que,

$$\begin{aligned} \mathbf{E}^{\pi(\beta|y)}[\beta_i] &= \frac{\frac{\partial}{\partial y_i} \int_{\Omega} h(y) e^{\beta^t y - \psi(\beta)} \pi(\beta) d\beta}{m(y)} - \frac{\frac{\partial}{\partial y_i} h(y)}{h(y)} \\ &= \frac{\partial}{\partial y_i} (\log m(y) - \log h(y)). \end{aligned}$$

□

Disposant de $\pi(\beta|y)$, on peut aussi définir l'estimateur du maximum de vraisemblance généralisée de β qui est défini comme étant le maximum de $f(y|\beta)\pi(\beta)$. Il apparaît donc comme un estimateur de maximum de vraisemblance pour la vraisemblance pénalisée (au sens classique). Les propriétés de convergence et d'efficacité de l'estimateur du maximum de vraisemblance classique sont donc conservées par cet estimateur sous quelques hypothèses de régularité sur f et π (voir Lehmann (1983), chapitre 6). Par conséquent, sous la fonction de perte quadratique, les estimateurs de Bayes sont asymptotiquement équivalents à l'estimateur du maximum de vraisemblance (voir Robert (1992), section 4.1).

2.2. MODÈLES HIÉRARCHIQUES

L'analyse bayésienne est suffisamment directive pour permettre une décision effective, mais cette efficacité n'est pas sans danger. Par exemple, on peut modifier l'apport subjectif (modélisation de la loi *a priori*) de manière à produire les conclusions désirées. De plus, le choix d'une loi *a priori* doit toujours être justifiable aux yeux d'un observateur extérieur (voir Robert (1992), p. 263). Plus fondamentalement, les informations disponibles *a priori* sont rarement suffisamment précises pour définir exactement la loi *a priori*. Il convient alors de pouvoir incorporer cette imprécision dans le modèle bayésien.

Une première façon d'arriver est de considérer que l'incertitude peut se modéliser en supposant que la loi *a priori* π appartient à une classe Γ de lois *a priori* au lieu d'être complètement spécifiée. Ces lois peuvent être choisies pour des raisons pratiques ou subjectives. Des exemples de classes Γ sont données dans Robert (1992), page 112. Mais celle qui est la plus connue et la plus utilisée est la classe des lois ϵ -contaminées d'une loi *a priori* π_0 :

$$\Gamma_{\epsilon, \mathcal{Q}} = \{\pi = (1 - \epsilon)\pi_0 + \epsilon q; q \in \mathcal{Q}\},$$

où \mathcal{Q} est une classe de distribution plus ou moins grande. Un problème majeur lié à l'emploi de $\Gamma_{\epsilon, \mathcal{Q}}$ demeure la détermination de ϵ et de \mathcal{Q} (Robert (1992), section 3.5). De plus, l'emploi de la classe Γ n'est pas suffisamment robuste.

Une autre approche plus restrictive des problèmes de robustesse est de chercher à caractériser des familles de lois robustes. Par exemple, on montre que, pour le modèle de régression normale, les lois de Student seront plus robustes que les lois normales (Zellner, 1976; Angers et Berger, 1991 et Angers et MacGibbon, 1996). De façon plus générale, il a été établi que les structures hiérarchiques renforcent la robustesse d'une loi *a priori* (voir Robert (1992), section 8.1).

L'approche que nous présentons dans cette section est celle qu'on qualifie de « bayésienne hiérarchique ». Un des buts de l'analyse bayésienne hiérarchique est d'incorporer dans le modèle bayésien l'imprécision de l'information *a priori*. Elle modélise l'information en niveaux successifs et conditionnels en distinguant entre apports structurels et apports proprement subjectifs et en modélisant l'incertitude sur la loi *a priori* par une « loi sur cette loi ». La stratification de la modélisation *a priori* favorise également la robustesse des estimateurs résultants et autorise la prise en compte d'informations subjectives tout en restant performant d'un

point de vue fréquentiste (voir Robert, 1992, p. 113). Ce modèle est décrit dans la définition suivante.

Définition 2.2.1. *On appelle modèle bayésien hiérarchique un modèle statistique bayésien, $(f(y|\beta), \pi(\beta))$, où la loi a priori $\pi(\beta)$ est décomposée en distributions conditionnelles $\pi_1(\beta|\beta_1)$, $\pi_2(\beta_1|\beta_2)$, ..., $\pi_k(\beta_{k-1}|\beta_k)$ et une distribution marginale $\pi_{k+1}(\beta_k)$. Les paramètres β_i $i = 1, \dots, k$ sont dits hyperparamètres.*

Remarquons qu'un modèle bayésien hiérarchique est un cas particulier d'un modèle bayésien. En effet, si

$$y \sim f(y|\beta), \beta \sim \pi_1(\beta|\beta_1), \dots, \beta_k \sim \pi_{k+1}(\beta_k),$$

on a le modèle bayésien usuel

$$\begin{aligned} y &\sim f(y|\beta), \\ \beta &\sim \pi(\beta), \text{ avec} \\ \pi(\beta) &= \int_{\Omega_1 \times \dots \times \Omega_k} \pi_1(\beta|\beta_1) \pi_2(\beta_1|\beta_2) \dots \pi_{k+1}(\beta_k) d\beta_1 \dots d\beta_k. \end{aligned} \quad (2.2.1)$$

Cette réduction montre que cette approche particulière bénéficie des propriétés générales de l'approche bayésienne avec, en plus, des avantages propres à la décomposition de la loi *a priori*. Quand les hyperparamètres β_1, \dots, β_k ne présentent pas d'intérêt pour l'inférence (sur β), on considère souvent que deux niveaux du modèle hiérarchique, soit

$$\begin{aligned} y|\beta &\sim f(y|\beta), \\ \beta|\beta_1 &\sim \pi_1(\beta|\beta_1), \\ \beta_1 &\sim \pi_2(\beta_1) = \int_{\Omega_2 \times \dots \times \Omega_k} \pi_2(\beta_1|\beta_2) \dots \pi_{k+1}(\beta_k) d\beta_2 \dots d\beta_k, \end{aligned}$$

qui élimine des étapes intermédiaires. En dépit de la réduction (2.2.1), qui montre que toute écriture hiérarchique n'est qu'un cas particulier du modèle bayésien,

la décomposition de la loi *a priori* peut être préférable parce que l'analyse bayésienne hiérarchique amène, entre autre, une robustification de l'approche bayésienne usuelle d'un point de vue fréquentiste, car elle diminue l'arbitraire du choix des paramètres (éventuellement reporté au second niveau) et moyenne les réponses bayésiennes (voir Robert, 1992, section 8.2).

Un intérêt pratique des modèles hiérarchiques est d'autoriser le conditionnement à tous les niveaux, ce qui peut compenser l'apparence de complexité introduite par ces niveaux. Ainsi, en supposant le modèle à deux niveaux, on a les résultats suivants.

Théorème 2.2.1. *Si $y|\beta \sim f(y|\beta)$, $\beta|\beta_1 \sim \pi_1(\beta|\beta_1)$, $\beta_1 \sim \pi_2(\beta_1)$ et si toutes les quantités nécessaires existent, alors la loi a posteriori de β est*

$$\pi(\beta|y) = \int_{\Omega_1} \pi_1(\beta|\beta_1, y) \pi_2(\beta_1|y) d\beta_1$$

et

$$\mathbf{E}^{\pi(\beta|y)}[g(\beta)] = \int_{\Omega_1} \mathbf{E}^{\pi_1(\beta|\beta_1, y)}[g(\beta)] \pi_2(\beta_1|y) d\beta_1, \quad (2.2.2)$$

où

$$\begin{aligned} \pi_1(\beta|\beta_1, y) &= \frac{f(y|\beta) \pi_1(\beta|\beta_1)}{m_1(y|\beta_1)}, \\ m_1(y|\beta_1) &= \int_{\Omega} f(y|\beta) \pi_1(\beta|\beta_1) d\beta, \\ \pi_2(\beta_1|y) &= \frac{m_1(y|\beta_1) \pi_2(\beta_1)}{m(y)}, \\ m(y) &= \int_{\Omega_1} m_1(y|\beta_1) \pi_2(\beta_1) d\beta_1. \end{aligned}$$

Démonstration. Par définition, on a

$$\pi(\beta|y) = \frac{\pi(\beta)f(y|\beta)}{\int_{\Omega} \pi(\beta)f(y|\beta)d\beta}.$$

Puisque

$$\pi(\beta) = \int_{\Omega_1} \pi_1(\beta|\beta_1)\pi_2(\beta_1)d\beta_1,$$

on peut écrire

$$\pi(\beta|y) = \frac{\left[\int_{\Omega_1} \pi_1(\beta|\beta_1)\pi_2(\beta_1)d\beta_1 \right] f(y|\beta)}{\int_{\Omega} \left[\int_{\Omega_1} \pi_1(\beta|\beta_1)\pi_2(\beta_1)d\beta_1 \right] f(y|\beta)d\beta}.$$

Alors, si on peut interchanger les intégrales (voir le théorème 18.3 de Billingsley (1986), page 238), on trouve

$$\begin{aligned} \pi(\beta|y) &= \frac{\int_{\Omega_1} \frac{\pi_1(\beta|\beta_1)f(y|\beta)}{m_1(y|\beta_1)} m_1(y|\beta_1)\pi_2(\beta_1)d\beta_1}{\int_{\Omega_1} \left[\int_{\Omega} \pi_1(\beta|\beta_1)f(y|\beta)d\beta \right] \pi_2(\beta_1)d\beta_1} \\ &= \frac{\int_{\Omega_1} \pi_1(\beta|\beta_1, y)m_1(y|\beta_1)\pi_2(\beta_1)d\beta_1}{\int_{\Omega_1} m_1(y|\beta_1)\pi_2(\beta_1)d\beta_1} \\ &= \frac{\int_{\Omega_1} \pi_1(\beta|\beta_1, y)\pi_2(\beta_1|y)m(y)d\beta_1}{\int_{\Omega_1} \pi_2(\beta_1|y)m(y)d\beta_1} \\ &= \int_{\Omega_1} \pi_1(\beta|\beta_1, y)\pi_2(\beta_1|y)d\beta_1. \end{aligned}$$

De plus, on obtient

$$\begin{aligned} \mathbf{E}^{\pi(\beta|y)}[g(\beta)] &= \int_{\Omega} g(\beta)\pi(\beta|y)d\beta \\ &= \int_{\Omega} g(\beta) \int_{\Omega_1} \pi_1(\beta|\beta_1, y)\pi_2(\beta_1|y)d\beta_1 \\ &= \int_{\Omega_1} \left[\int_{\Omega} g(\beta)\pi_1(\beta|\beta_1, y)d\beta \right] \pi_2(\beta_1|y)d\beta_1 \\ &= \int_{\Omega_1} \mathbf{E}^{\pi_1(\beta|\beta_1, y)}[g(\beta)]\pi_2(\beta_1|y)d\beta_1. \end{aligned}$$

□

Les quantités utilisées dans le théorème 2.2.1 doivent être bien définies pour que les résultats soient valides. Ceci n'est pas toujours le cas puisque les lois de second niveau sont généralement impropres. Berger et Robert (1990) ont donné des conditions suffisantes pour l'existence des moments *a posteriori* dans le cas où $Y|\beta \sim N_p(\beta, \Sigma)$, c'est-à-dire, la loi *a posteriori* $\pi(\beta|y)$ admet une moyenne ou une variance pourvu que la loi marginale $m(y) = \int_{\Omega} f(y|\beta)\pi(\beta)d\beta$ soit finie pour $Y \in \mathbb{R}^p$.

Il convient de noter que les modèles hiérarchiques autorisent rarement les calculs explicites des estimateurs de Bayes, même si les divers niveaux sont conjugués. Nous devons donc souvent utiliser des approximations numériques pour évaluer l'équation (2.2.2).

Exemple 2.2.1. Soient $Y|\beta \sim \mathcal{N}(\beta, \sigma^2)$ avec σ^2 connue et $\beta|C \sim \mathcal{U}(-C, C)$, avec $C > 0$. On a

$$\begin{aligned} \pi_1(\beta|Y, C) &= \frac{f(Y|\beta)\pi_1(\beta|C)}{\int_{\Omega} f(Y|\beta)\pi_1(\beta|C)d\beta} \\ &= \frac{\exp\{-(Y-\beta)^2/2\sigma^2\}(1/2C)I_{(-C, C)}(\beta)}{\int_{-C}^C \exp\{-(Y-\beta)^2/2\sigma^2\}(1/2C)d\beta} \\ &= \frac{\exp\{-(Y-\beta)^2/2\sigma^2\}I_{(-C, C)}(\beta)}{\int_{-C}^C \exp\{-(Y-\beta)^2/2\sigma^2\}d\beta} \\ &= \frac{\frac{1}{\sqrt{2\pi\sigma^2}}(\exp\{-(Y-\beta)^2/2\sigma^2\})I_{(-C, C)}(\beta)}{\phi\left(\frac{Y+C}{\sigma}\right) - \phi\left(\frac{Y-C}{\sigma}\right)}, \end{aligned}$$

où $I_{(-C, C)}(\beta)$ est la fonction indicatrice d'intervalle $(-C, C)$ de β et ϕ représente la distribution d'une loi normale centrée réduite. De plus,

$$\pi_2(C|Y) = \frac{\int_{-C}^C (2\pi\sigma^2)^{-1/2} \exp\{-(Y-\beta)^2/2\sigma^2\} (1/2C) d\beta \pi_2(C)}{\int_0^{\infty} \left[\int_{-C}^C (2\pi\sigma^2)^{-1/2} \exp\{-(Y-\beta)^2/2\sigma^2\} (1/2C) d\beta \right] \pi_2(C) dC}.$$

Si on prend $\pi_2(C) \propto 1/C$, la loi a posteriori devient

$$\begin{aligned}\pi(\beta|Y) &= \int_0^\infty \pi_1(\beta|Y, C)\pi_2(C|Y)dC \\ &= \frac{\int_0^\infty \exp\{-(Y-\beta)^2/2\sigma^2\}(1/2C^2)I_{(-C, C)}(\beta)dC}{\int_0^\infty [\phi(\frac{Y+C}{\sigma}) - \phi(\frac{Y-C}{\sigma})](1/2C^2)dC} \\ &= \frac{\frac{1}{2|\beta|} \exp\{\frac{-1}{2\sigma^2}(Y-\beta)^2\}}{\int_0^\infty [\phi(\frac{Y+C}{\sigma}) - \phi(\frac{Y-C}{\sigma})](1/2C^2)dC}.\end{aligned}$$

Il faut alors recourir à des approximations numériques pour calculer $\mathbf{E}^{\pi(\beta|Y)}[\beta]$.

Mais dans le cas particulier de la loi normale, $Y|\beta \sim \mathcal{N}_p(\beta, \Sigma)$, comme dans Lindley et Smith (1972), Smith (1973) ou Berger (1985), avec une loi conjuguée de premier niveau, $\beta \sim \mathcal{N}_p(\mu, \Sigma_\pi)$ (modèle normal conjugué), on obtient une décomposition plus aisée des estimateurs, comme le montre le résultat suivant.

Lemme 2.2.1. Dans le modèle normal conjugué, l'estimateur de Bayes hiérarchique est

$$\delta^\pi(y) = \mathbf{E}^{\pi_2(\mu, \Sigma_\pi|y)} \mathbf{E}[\delta(y|\mu, \Sigma_\pi)],$$

avec $\delta(y|\mu, \Sigma_\pi) = y - \Sigma W(y - \mu)$, $W = (\Sigma + \Sigma_\pi)^{-1}$ et

$$\pi_2(\mu, \Sigma_\pi|y) \propto (\det W)^{-1/2} \exp\{-(1/2)(y - \mu)^t W(y - \mu)\} \pi_2(\mu, \Sigma_\pi).$$

Démonstration. En prenant $g(\beta) = \beta$ dans le théorème 2.2.1, on a

$$\delta^\pi(y) = \mathbf{E}^{\pi(\beta|y)}[\beta] = \int_{\Omega_1} \mathbf{E}^{\pi_1(\beta|\mu, \Sigma_\pi)}[\beta] \pi_2(\mu, \Sigma_\pi|y) d\mu d\Sigma_\pi.$$

Posons

$$\delta(y|\mu, \Sigma_\pi) = \mathbf{E}^{\pi_1(\beta|\mu, \Sigma_\pi)}[\beta] = y - \Sigma W(y - \mu). \quad (2.2.3)$$

Pour trouver $\delta^\pi(y)$, il faut trouver $\pi_1(\beta|y)$ et $\pi_2(\mu, \Sigma_\pi|y)$.

Pour trouver $\pi_1(\beta|y)$, on utilise le théorème de Bayes :

$$\pi_1(\beta|y) \propto \pi_1(y|\beta)\pi_1(\beta|\mu, \Sigma_\pi).$$

Dans un premier temps, le produit du membre de droite est $e^{-(1/2)Q}$, où

$$\begin{aligned} Q &= (y - \beta)^t \Sigma^{-1} (y - \beta) + (\beta - \mu)^t \Sigma_\pi^{-1} (\beta - \mu), \\ &= \beta^t B^{-1} \beta - 2b^t \beta + \{y^t \Sigma^{-1} y + \mu^t \Sigma_\pi^{-1} \mu\} \end{aligned}$$

et $B^{-1} = \Sigma^{-1} + \Sigma_\pi^{-1}$, $b = \Sigma^{-1} y + \Sigma_\pi^{-1} \mu$.

En rassemblant les termes quadratiques et les termes linéaires en β et en complétant le carré, Q peut finalement s'écrire comme suit

$$Q = (\beta - Bb)^t B^{-1} (\beta - Bb) + \{y^t \Sigma^{-1} y + \mu^t \Sigma_\pi^{-1} \mu - b^t Bb\}.$$

Les termes entre les accolades sont des constantes par rapport à la distribution de β . On peut donc conclure que $\pi_1(\beta|\mu, \Sigma_\pi, y) = \mathcal{N}(Bb, B^{-1})$.

Mais comme

$$\begin{aligned} Bb &= [\Sigma^{-1} + \Sigma_\pi^{-1}]^{-1} [\Sigma^{-1} y + \Sigma_\pi^{-1} \mu] \\ &= \Sigma [\Sigma^{-1} - \Sigma^{-1} + \Sigma^{-1} (\Sigma^{-1} + \Sigma_\pi^{-1})^{-1} \Sigma^{-1}] y \\ &\quad + \Sigma [(\Sigma_\pi^{-1} + \Sigma^{-1}) - \Sigma_\pi^{-1}] (\Sigma^{-1} + \Sigma_\pi^{-1})^{-1} \Sigma_\pi^{-1} \mu, \end{aligned}$$

on peut montrer (voir Lindley et Smith, 1972) que

$$\Sigma^{-1} - \Sigma^{-1} (\Sigma^{-1} + \Sigma_\pi^{-1})^{-1} \Sigma^{-1} = (\Sigma + \Sigma_\pi)^{-1} = W.$$

De la même façon, $\Sigma_\pi^{-1} - \Sigma_\pi^{-1} (\Sigma^{-1} + \Sigma_\pi^{-1})^{-1} \Sigma_\pi^{-1} = (\Sigma_\pi + \Sigma)^{-1} = W$.

Alors, on obtient

$$\begin{aligned} Bb &= \Sigma [\Sigma^{-1} - W] y + \Sigma W \mu \\ &= [I - \Sigma W] y + \Sigma W \mu \\ &= y - \Sigma W (y - \mu). \end{aligned}$$

Pour trouver $\pi_2(\mu, \Sigma_\pi | y)$, on écrit

$$\begin{aligned} y &= \beta + v, v \sim \mathcal{N}(0, \Sigma), \\ \beta &= \mu + w, W \sim \mathcal{N}(0, \Sigma_\pi). \end{aligned}$$

En mettant ces deux équations ensemble, on a $y = u + v + w$. Puisque $v + w$ est une fonction linéaire de variables aléatoires normales et indépendantes, on a $v + w \sim \mathcal{N}(0, \Sigma + \Sigma_\pi)$. Il s'ensuit que

$$m_1(y | \mu, \Sigma_\pi) = \mathcal{N}(\mu, W^{-1}), \text{ où } W = (\Sigma + \Sigma_\pi)^{-1}.$$

Comme $\pi_2(\mu, \Sigma_\pi | y) \propto m_1(y | \mu, \Sigma_\pi) \pi_2(\mu, \Sigma_\pi)$, on a

$$\pi_2(\mu, \Sigma_\pi | y) \propto \mathcal{N}(\mu, W^{-1}) \pi_2(\mu, \Sigma_\pi).$$

Pour trouver l'estimateur de Bayes, on a qu'à intégrer l'équation (2.2.3) par rapport à $\pi_2(\mu, \Sigma_\pi | y)$. \square

Étant donné que le modèle bayésien hiérarchique est un cas particulier du modèle bayésien ordinaire, il en résulte que les estimateurs de Bayes hiérarchiques possèdent toutes les propriétés des estimateurs de Bayes «classiques».

Par contre, dans le cas particulier de la loi normale conjuguée, on peut exhiber une condition de minimaxité spécifique aux estimateurs hiérarchiques (voir Robert (1992), pages 275-277). En écrivant la loi des hyperparamètres (μ, Σ_π) , le résultat précédent donne une condition suffisante de minimaxité qui ne dépend pas de $\pi_2^2(\mu)$ où $\pi_2(\mu, \Sigma_\pi) = \pi_2^1(\Sigma_\pi | \mu) \pi_2^2(\mu)$, $\pi_2^1(\Sigma_\pi | \mu)$ est la loi *a priori* de Σ_π sachant μ , et autorise donc toutes les modélisations subjectives possibles sur l'hyperparamètre μ .

2.3. L'ANALYSE BAYÉSIENNE EMPIRIQUE

L'analyse bayésienne empirique fut introduite par Robbins (1955) dans le contexte d'estimation non paramétrique d'une loi *a priori* complètement inconnue. Depuis, elle a été explorée avec succès dans une série d'articles par Efron et Morris (1973 et 1975) dans le contexte paramétrique. Dans la présente section, nous donnons une brève description de la méthode appelée bayésienne empirique paramétrique par Morris (1983).

Comme dans l'approche hiérarchique, nous nous plaçons dans un contexte non informatif où la loi *a priori* n'est pas disponible. Un choix naturel est donc de prendre la loi conjuguée associée à $f(y|\beta), \pi(\beta|\beta_1)$ (voir Robert, 1992, section 8.3). Au lieu de proposer une loi *a priori* sur l'hyperparamètre β_1 , comme dans la méthode hiérarchique, l'analyse empirique suggère d'estimer ce paramètre par $\hat{\beta}_1$ à partir de la loi marginale

$$m(y|\beta_1) = \int_{\Omega} f(y|\beta)\pi(\beta|\beta_1)d\beta$$

et d'utiliser $\pi(\beta|\hat{\beta}_1(y), y)$ comme pseudo-loi *a posteriori*.

Le fait de considérer la loi marginale $m(y|\beta_1)$ nous ramène donc à un problème fréquentiste d'estimation du paramètre β_1 (voir Robert (1992), section 8.3). Il faut alors sélectionner un estimateur à partir de l'une ou l'autre des différentes méthodes disponibles : moments, maximum de vraisemblance, etc. Pour des raisons pratiques et théoriques, la méthode du maximum de vraisemblance est la plus fréquemment retenue. Dans le cas particulier de l'estimation du paramètre

naturel d'une famille exponentielle sous le coût quadratique, nous avons le résultat suivant.

Lemme 2.3.1. *Soit $y \sim f(y|\beta) = e^{\beta^t y - \psi(\beta)} h(y)$, $y \in \mathbb{R}^k$. Si β a comme loi a priori $\pi(\beta|\beta_1)$, $\beta_1 \in \mathbb{R}^p$, et $\hat{\beta}_1$ est la solution des équations de vraisemblance associées à $m(y|\beta_1)$, l'estimateur de Bayes empirique est*

$$\begin{aligned} \delta^{EB}(y) &= (\nabla_y \log m(y|\beta_1)) \Big|_{\beta_1=\hat{\beta}_1(y)} - \nabla_y \log h(y) \\ &= \nabla_y [\log m(y|\hat{\beta}_1(y))] - \nabla_y \log h(y), \end{aligned}$$

où $\nabla_y = (\frac{\partial}{\partial y_1}, \dots, \frac{\partial}{\partial y_k})^t$.

Démonstration. On a en effet

$$\nabla_y \log m(y|\hat{\beta}_1(y)) = \nabla_y \log m(y|\beta_1) \Big|_{\beta_1=\hat{\beta}_1(y)} + (\nabla_y \hat{\beta}_1(y))^t \nabla_{\beta_1} m(y|\beta_1) \Big|_{\beta_1=\hat{\beta}_1(y)},$$

où $\nabla_{\beta_1} m(y|\beta_1) \Big|_{\beta_1=\hat{\beta}_1(y)}$ est le vecteur de composantes $\frac{\partial m(y|\beta_1)}{\partial \beta_i}$ ($1 \leq i \leq p$) et $\nabla_y \hat{\beta}_1(y)$ est la matrice $k \times p$ de composantes $\frac{\partial \hat{\beta}_i(y)}{\partial y_j}$ ($1 \leq i \leq p, 1 \leq j \leq k$). Par définition de $\hat{\beta}_1(y)$, le second terme est nul. \square

Exemple 2.3.1. *Supposons les variables aléatoires y_i étant donné β_i , indépendantes et de lois normales, $\mathcal{N}(\beta_i, \sigma^2)$, σ^2 connu, $i = 1, 2, \dots, n$. Soit $\pi(\beta|\beta_0, \tau^2) = \mathcal{N}(\beta_0, \tau^2)$. Alors, on obtient*

$$\begin{aligned} m(y_i|\beta_0, \tau^2) &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma\tau} \exp\{-(y_i - \beta)^2/(2\sigma^2)\} \exp\{-(\beta - \beta_0)^2/(2\tau^2)\} d\beta \\ &= \frac{1}{\sqrt{2\pi(\sigma^2 + \tau^2)}} \exp\left\{-\frac{(y_i - \beta_0)^2}{2(\sigma^2 + \tau^2)}\right\}. \end{aligned}$$

Donc $Y_i|\beta_0, \tau^2 \sim \mathcal{N}(\beta_0, \sigma^2 + \tau^2)$. Les estimateurs du maximum de vraisemblance de β_0 et de τ^2 sont respectivement $\hat{\beta}_0(y_1, \dots, y_n) = \bar{y}$ et

$$\hat{\tau}^2 = \left[n^{-1} \sum_{i=1}^n (y_i - \bar{y})^2 - \sigma^2 \right]^+.$$

Si on observe $Y_{n+1}|\beta_{n+1}, \sigma^2 \sim \mathcal{N}(\beta_{n+1}, \sigma^2)$ et si on estime β_{n+1} sous le coût quadratique, l'estimateur empirique de Bayes de β_{n+1} devient

$$\delta^{EB}(y_{n+1}) = \begin{cases} y_{n+1} - \frac{\sigma^2}{\sigma^2 + \hat{\tau}^2}(y_{n+1} - \bar{y}) & \text{si } \hat{\tau}^2 > 0, \\ \bar{y} & \text{si } \hat{\tau}^2 = 0, \end{cases}$$

où $\bar{y} = n^{-1} \sum_{i=1}^n y_i$ et $\hat{\tau}^2 = n^{-1} \sum_{i=1}^n (y_i - \bar{y})^2$.

Cette approche peut produire des estimateurs plus performants d'un point de vue fréquentiste, non seulement pour l'estimation ponctuelle mais aussi pour les tests et les régions de confiance (voir, Robert (1992), section 8.7). Cependant les tests bayésiens empiriques ont une portée très limitée dans le cas d'un seul échantillon (voir Robert (1992), p. 263).

On note que dans le cas des familles exponentielles, l'inférence bayésienne menée à partir de la loi approchée $\pi(\beta|\hat{\beta}_1(y))$ donne les mêmes résultats que l'approche bayésienne empirique naïve, où on remplace β_1 par $\hat{\beta}_1(y)$ après avoir obtenu l'estimateur de Bayes.

Pour conclure cette section, nous soulignons que l'analyse bayésienne empirique constitue une approche duale de l'analyse hiérarchique développée dans la section 2 et peut également constituer une approximation acceptable pour des problèmes où une modélisation hiérarchique est trop complexe ou trop coûteuse.

2.4. MODÈLES DE RÉGRESSION LINÉAIRE BAYÉSIEN

Comme nous avons vu au chapitre 1, les informations dont on peut disposer sur les variables auxiliaires avant le sondage peuvent aider à mieux estimer

le total ou la moyenne d'une caractéristique d'une population. Dans cette section, nous allons donc voir une formulation générale d'un modèle de régression linéaire bayésien qui permettra d'incorporer les informations auxiliaires dans le but d'améliorer l'estimation du total ou de la moyenne. Plus spécifiquement, nous donnons l'estimateur hiérarchique de Bayes pour une structure de variance particulière. Nous présentons également un autre estimateur du paramètre β , basé cette fois, sur le modèle de Box et Tiao (1968).

2.4.1. Modèle normal

En se basant sur Pettit et Smith (1985), une formulation générale d'un modèle de régression bayésien peut être décrite de la façon suivante :

$$Y|\beta \sim \mathcal{N}(\mathbf{X}\beta, \Sigma), \beta|\mu \sim \mathcal{N}(A\mu, \Sigma_\pi), \quad (2.4.1)$$

où $Y^t = (Y_1, \dots, Y_n)$, β un vecteur de p paramètres, $\mu \in \mathbb{R}^k$ ($k \leq p$), \mathbf{X} est une matrice de dimension $n \times p$ (nous appelons \mathbf{X} , matrice des informations auxiliaires), A est une matrice $p \times k$, Σ et Σ_π sont des matrices symétriques et définies positives.

La loi des observations conditionnées aux paramètres β est appelée fonction de vraisemblance. Remarquons que cette loi est la même dans un contexte bayésien que dans un contexte classique. La loi de probabilité *a priori* que suit le paramètre β dans le modèle (2.4.1) est $\beta|\mu \sim \mathcal{N}(A\mu, \Sigma_\pi)$. Évidemment, il existe plusieurs choix quant aux lois *a priori* qu'on peut donner aux paramètres. De plus, dans le cadre du modèle hiérarchique, on peut considérer une loi π_2 sur les hyperparamètres μ et Σ_π qui aura pour effet de robustifier l'estimateur de Bayes.

Si on pose $\mathbf{X} = A = I$, on peut remarquer que le modèle (2.4.1) est une généralisation du modèle normal conjugué. Aussi, dans le cas où la variance est connue à un facteur multiplicatif près, σ^2 ,

$$Y|\beta \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2 I_n),$$

(ce qui est équivalent à considérer $Y = \mathbf{X}\beta + \epsilon$, $\epsilon \in \mathbb{R}^n \sim \mathcal{N}(0, \sigma^2 I_n)$) on peut obtenir des estimateurs de Bayes explicites du paramètre β . Dans la suite de ce mémoire, nous utiliserons cette structure de variance particulière pour avoir des estimateurs de Bayes hiérarchiques de β .

2.4.2. Modèle normal hiérarchique

Dans la présente sous-section, nous considérons donc le modèle hiérarchique suivant :

$$Y|\beta \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2 I_n), \quad (2.4.2)$$

$$\theta|\mu, S \sim \mathcal{N}(\mathbf{0}, S), \quad (2.4.3)$$

$$\pi_2(S|\sigma^2) \propto |\sigma^2 I_p + S\mathbf{X}^t \mathbf{X}|^{-l}, \quad (2.4.4)$$

où σ^2 est considéré connu et où $S = \text{diag}(s_1^2, \dots, s_n^2)$ et l est choisi assez grand pour assurer que la densité de $\pi_2(S|\sigma^2)$ s'intègre à 1. Il est évident qu'il n'y a pas de perte de généralité en considérant le modèle donné par l'équation (2.4.3) car par une transformation linéaire de l'équation (2.4.2), on peut se ramener à l'équation (2.4.3). Plus spécifiquement, si la moyenne *a priori* de β est μ , alors l'équation (2.4.2) peut s'écrire comme

$$Y = \mathbf{X}\beta + \epsilon = \mathbf{X}(\beta - \mu + \mu) + \epsilon \Leftrightarrow Y - \mathbf{X}\mu = \mathbf{X}(\beta - \mu) + \epsilon.$$

Soit $Y_* = Y - \mathbf{X}\mu$, $\beta_* = (\beta - \mu)$, la moyenne *a priori* de β_* est égale à 0. Si μ est inconnu et si la loi *a priori* de μ est $\mu \sim \mathcal{N}(0, T)$, où $T = \text{diag}(t_1^2, t_2^2, \dots, t_p^2)$, alors le

modèle *a priori* de β devient, après intégration sur β , $\beta \sim \mathcal{N}(0, S_*)$ où $S_* = S + T$.

Les estimateurs obtenus à partir du modèle spécifié par les équations (2.4.2), (2.4.3) et (2.4.4) sont dites robustes parce qu'ils performant mieux que l'estimateur des moindres carrés de β quand les données sont incompatibles à la distribution *a priori* (voir Angers et McGibbon, 1996). Si σ^2 est connu, l'estimateur a une forme analytique assez simple.

Dans le cas où σ^2 est connu, Angers et McGibbon (1996) ont démontré que l'estimateur hiérarchique de Bayes correspondant à ce modèle est

$$\hat{\beta}^H = Q[I_p - G_l]Q^{-1}\hat{\theta}^{LS} = [I_p - QG_lQ^{-1}]\hat{\beta}^{LS}, \quad (2.4.5)$$

où $\hat{\beta}^{LS}$ est l'estimateur des moindres carrés de β ,

$$G_l = \text{diag}[g_l(z_1^2/(2\sigma^2)), \dots, g_l(z_p^2/(2\sigma^2))],$$

$$g_l(x) = \begin{cases} \frac{l-1/2}{l+1/2}, & \text{si } x = 0, \\ \frac{\gamma(l+1/2, x)}{x\gamma(l-1/2, x)}, & \text{sinon,} \end{cases}$$

où $\gamma(m, a) = \int_0^a x^{m-1}e^{-x}dx$, $\mathbf{z}_{(1)} = (z_1, \dots, z_p)^t = Q^{-1}\hat{\beta}^{LS}$ et Q correspond à la matrice de décomposition de Cholesky de $(\mathbf{X}^t\mathbf{X})^{-1}$.

Le risque fréquentiste de cet estimateur correspond à l'écart quadratique moyen de β et il est donné par

$$MSE(\beta, \hat{\beta}^H) = \sigma^2 \text{tr}(\mathbf{X}^t\mathbf{X})^{-1} + 2\sigma^2 \text{tr}(Q^tQE), \quad (2.4.6)$$

où E est une matrice $p \times p$ avec les éléments (i, j) définis par

$$E_{i,j} = \begin{cases} e_{1,1}(\zeta_i)e_{1,1}(\zeta_j), & \text{si } i \neq j, \\ e_{2,2}(\zeta_i) - 2e_{1,2}(\zeta_i) + 2\zeta_i e_{1,1}(\zeta_i), & \text{si } i = j, \end{cases}$$

$$e_{a,b}(\zeta_i) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} g_i^a([u + \zeta_i]^2)(u + \zeta_i)^b e^{-u^2} du,$$

$$(\zeta_1, \dots, \zeta_p)^t = \frac{1}{\sqrt{2\sigma^2}} Q^{-1} \beta,$$

et $tr A$ désigne la trace de la matrice A .

L'équation (2.4.6) permet de comparer la précision de l'estimateur hiérarchique de Bayes avec celle de l'estimateur des moindres carrés. Le théorème suivant montre que $\hat{\beta}^H$ performe mieux que $\hat{\beta}^{LS}$ dans un voisinage de $\mathbf{0}$.

Théorème 2.4.1. $MSE(\mathbf{0}, \hat{\beta}^H) < \sigma^2 tr(\mathbf{X}^t \mathbf{X})^{-1} = MSE(\mathbf{0}, \hat{\beta}^{LS})$.

La démonstration de ce théorème se trouve dans Angers et McGibbon (1996).

Quand σ^2 est inconnu, on peut utiliser une loi non informative *a priori* de la forme $\pi_3(\sigma^2) \propto (\sigma^2)^{-k}$. Ainsi, la loi *a posteriori* de σ^2 peut être écrite comme suit :

$$\pi_3(\sigma^2 | \mathbf{y}) \propto \frac{1}{(\sigma^2)^{(n-2[p-l-k])/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=p+1}^n z_j^2\right) \prod_{j=1}^p \gamma\left(l - \frac{1}{2}, \frac{z_j^2}{2\sigma^2}\right),$$

où les $\mathbf{z} = (z_1, \dots, z_n)^t = H\mathbf{y}$, avec H qui est donné par

$$\mathbf{X} S \mathbf{X}^t = H \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} H^t,$$

avec $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$, où λ_j est le j -ième valeur propre de $S\mathbf{X}^t \mathbf{X}$. On a qu'à intégrer la fonction $g_l(z_j^2/(2\sigma^2))$ par rapport à $\pi_3(\sigma^2 | \mathbf{y})$ pour obtenir l'estimateur généralisé de Bayes. Cette intégration peut se faire analytiquement si on choisit

$l = (2m+1)/2$, où m est un entier positif, mais l'expression est un peu compliquée. Une alternative à ce problème est d'utiliser l'approche empirique (voir section 2.3) et de trouver l'estimateur du maximum de vraisemblance généralisé de σ^2 en considérant $\pi_3(\sigma^2|\mathbf{y})$. Une approximation de l'estimateur généralisé de σ^2 est donnée par

$$\hat{\sigma}^2 = \frac{\sum_{j=p+1}^n z_j^2}{n + 2k - 2pl}.$$

Contrairement aux autres estimateurs robustes, l'estimateur hiérarchique de Bayes a l'avantage de tenir compte de toutes les données et tend à assurer sa robustesse en diminuant la fonction de poids associées aux valeurs aberrantes (voir Angers et McGibbon, 1996).

2.4.3. Modèle de paramètre de dispersion augmenté

Le modèle bayésien de paramètre de dispersion augmenté présenté dans cette sous-section a été introduit par Box et Tiao (1968). Pour ce modèle, nous supposons qu'avec une faible probabilité α , certaines observations ne sont pas générées à partir du modèle linéaire général et qu'avec une probabilité $(1 - \alpha)$, les autres observations sont générées à partir du modèle général.

Supposons que nous croyons qu'il y a dans l'échantillon r valeurs aberrantes et notons par $Y_{(r)}$, le sous vecteur de Y désignant celles-ci et par $\mathbf{X}_{(r)}$, la matrice du devis du plan correspondant à ces valeurs aberrantes. De plus, notons par $Y_{(n-r)}$ et $\mathbf{X}_{(n-r)}$, le sous-vecteur de Y et la sous-matrice du devis du plan désignant le reste des observations respectivement. La méthodologie utilisée par Box et Tiao (1968) est de supposer que ces r observations proviennent d'une loi normale de moyenne $\mathbf{X}_{(r)}\beta$ et de matrice de variance-covariance $k\sigma^2\mathbf{I}_{(r)}$, k une constante

connue supérieure à 1. Alors que les $(n-r)$ autres observations proviennent d'une loi normale de moyenne $\mathbf{X}_{(n-r)}\beta$, mais de matrice de variance-covariance $\sigma^2\mathbf{I}_{(n-r)}$. Comme nous l'avons mentionné auparavant dans ce chapitre, toute inférence sur les paramètres doit être basée sur la loi *a posteriori*. Il faut donc trouver la forme de cette loi *a posteriori*. Pour ce modèle, les auteurs ont montré que la loi de β étant donné les observations $\pi(\beta|Y)$ peut s'écrire comme une moyenne pondérée de 2^n lois t de Student multidimensionnelles avec $(n-p)$ degrés de liberté chacune. Nous énonçons la forme de cette loi *a posteriori* dans le théorème suivant.

Théorème 2.4.2. *Pour β et $\log \sigma$ indépendants et uniformément distribués, k une constante connue et fixée, la loi *a posteriori* de β est donnée par :*

$$\pi(\beta|Y) = \sum_{(r)} w_{(r)} \pi_{(r)}(\beta|Y), \quad (2.4.7)$$

où $\sum_{(r)}$ désigne la somme sur le 2^n sous-ensembles de r éléments, $r = 0, 1, \dots, n$,

$$\begin{aligned} w_{(r)} &= C \left(\frac{\alpha}{1-\alpha} \right)^r k^{-r} \frac{|\mathbf{X}^t \mathbf{X}|^{1/2}}{|\mathbf{X}^t \mathbf{X} - \phi \mathbf{X}_{(r)}^t \mathbf{X}_{(r)}|^{1/2}} \left(\frac{S_{(r)}^2}{S^2} \right)^{-\nu/2}, \\ \pi_{(r)}(\beta|Y) &= \frac{\Gamma(\frac{n}{2}) |\mathbf{X}^t \mathbf{X} - \phi \mathbf{X}_{(r)}^t \mathbf{X}_{(r)}|^{1/2}}{\Gamma(\frac{\nu}{2}) (\pi \nu S_{(r)}^2)^{p/2}} \\ &\quad \times \left[1 + \frac{(\beta - \hat{\beta}_{(r)})^t (\mathbf{X}^t \mathbf{X} - \phi \mathbf{X}_{(r)}^t \mathbf{X}_{(r)}) (\beta - \hat{\beta}_{(r)})}{\nu S_{(r)}^2} \right]^{-n/2}, \\ \hat{\beta}_{(r)} &= \hat{\beta}^{LS} - \phi (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}_{(r)}^t [\mathbf{I}_{(r)} - \phi \mathbf{X}_{(r)} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}_{(r)}^t]^{-1} \\ &\quad \times (Y_{(r)} - \mathbf{X}_{(r)} \hat{\beta}^{LS}), \end{aligned}$$

$$\begin{aligned}\nu S_{(r)}^2 &= \nu S^2 - \phi \left(Y_{(r)} - \mathbf{X}_{(r)} \beta^{\hat{LS}} \right)^t \left[\mathbf{I}_{(r)} - \phi \mathbf{X}_{(r)} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}_{(r)}^t \right]^{-1} \\ &\quad \times \left(Y_{(r)} - \mathbf{X}_{(r)} \beta^{\hat{LS}} \right), \\ \phi &= 1 - k^2, \\ \nu &= n - p,\end{aligned}$$

$S^2 = \hat{\sigma}^2$ et $\hat{\beta}^{LS}$ désignent les estimateurs des moindres carrés de σ^2 et de β respectivement et C est une constante qui assure que les poids $w_{(r)}$ somment à 1.

La démonstration de ce théorème se trouve dans Box et Tiao (1968).

Avant d'aller plus loin, remarquons que $\pi_{(r)}(\beta|y)$ est la fonction de densité d'une loi t de Student multidimensionnelle de moyenne $\hat{\beta}_{(r)}$, de matrice de variance-covariance $S_{(r)}^2 (\mathbf{X}^t \mathbf{X} - \phi \mathbf{X}_{(r)}^t \mathbf{X}_{(r)})^{-1}$ et $\nu = n - p$ degrés de liberté. Ainsi, la fonction de densité marginale de β_l , le l -ième élément de β , $l = 1, 2, \dots, p$ peut s'écrire comme :

$$\pi(\beta_l|Y) = \sum_{(r)} w_{(r)} \pi_{(r)}(\beta_l|Y), \quad (2.4.8)$$

où

$$\pi_{(r)}(\beta_l|Y) = S_{(r)}^{-1} (v_{(r)}^l)^{-1/2} g_{n-p} \left(\frac{\beta_l - \hat{\beta}_{l(r)}}{S_{(r)} (v_{(r)}^l)^{1/2}} \right),$$

g_{n-p} est la fonction de densité de Student avec $n - p$ degrés de liberté, $\hat{\beta}_{l(r)}$ est le l -ième élément de $\hat{\beta}_{(r)}$ et $v_{(r)}^l$ le l -ième élément diagonal de la matrice $(\mathbf{X}^t \mathbf{X} - \phi \mathbf{X}_{(r)}^t \mathbf{X}_{(r)})^{-1}$. Ainsi, l'estimateur de Bayes associé à cette loi *a posteriori*

et au coût quadratique (voir proposition 2.1.1) de β_l ainsi que la variance de celui-ci sont respectivement

$$\begin{aligned} \mathbf{E}(\beta_l|Y) &= \bar{\beta}_l = \sum_{(r)} w_{(r)} \hat{\beta}_{l(r)}, \\ \mathbf{V}(\beta_l|Y) &= \frac{n-p}{n-p-2} \sum_{(r)} w_{(r)} [S_{(r)}^2 v_{(r)}'' + (\hat{\beta}_{l(r)} - \bar{\beta}_l)^2]. \end{aligned} \quad (2.4.9)$$

Nous remarquons que l'estimateur (2.4.9) est une somme pondérée des $\hat{\beta}_{l(r)}$. Pour voir comment cet estimateur tient compte des valeurs aberrantes, réécrivons l'équation (2.4.7) comme :

$$\begin{aligned} \pi(\beta|Y) &= w_{(0)}\pi_0(\beta|Y) + \sum_{i=1}^n w_{(i)}\pi_i(\beta|Y) + \sum_{i<j}^n w_{(ij)}\pi_{ij}(\beta|Y) \\ &\quad + \sum_{i<j<l}^n w_{(ijl)}\pi_{ijl}(\beta|Y) + \dots, \end{aligned} \quad (2.4.10)$$

où le poids $w_{(0)}$ est la probabilité *a posteriori* qu'il n'y ait aucune valeur aberrante, le poids $w_{(i)}$ est la probabilité *a posteriori* que l'observation i (seulement) soit aberrante, les poids $w_{(ij)}$ sont les probabilités *a posteriori* que la i -ième et la j -ième observations soient aberrantes simultanément, ... De plus, les distributions $\pi_i(\beta|Y)$ allouent la possibilité que chaque observation soit aberrante. Les π_{ij} allouent la possibilité d'avoir deux valeurs aberrantes, ... Enfin, soulignons que la densité $\pi_0(\beta|Y)$ serait appropriée comme estimateur de $\pi(\beta|Y)$ s'il n'y avait aucune valeur aberrante (voir Box et Tiao, 1968). Ceci nous renseigne sur la façon dont les valeurs aberrantes sont considérées dans l'estimation du paramètre β basée sur cette densité. En effet, pour des échantillons ayant au plus une valeur aberrante, il est clair que l'apport des termes possédant « plus de deux indices de sommations » dans l'estimation de $\pi(\beta|Y)$ est négligeable. Ainsi, une bonne

approximation (voir Lemire, 1997, section 2.3) de l'équation (2.4.10) est :

$$\pi(\beta|Y) \approx w_{(0)}\pi_{(0)}(\beta|Y) + \sum_{i=1}^n w_{(i)}\pi_{(i)}(\beta|Y).$$

Alors, nous pouvons approximer $\mathbf{E}(\beta_l|Y)$ et $\mathbf{V}(\beta_l|Y)$ par

$$\begin{aligned} \mathbf{E}(\beta_l|Y) &= \sum_{i=0}^n w_{(i)}\hat{\beta}_{l(i)}, \\ \mathbf{V}(\beta_l|Y) &= \frac{n-p}{n-p-2} \sum_{i=0}^n w_{(i)}[S_{(i)}^2 v_{(i)}^l + (\hat{\beta}_{l(i)} - \bar{\beta}_l)^2]. \end{aligned} \quad (2.4.11)$$

Notons que pour le cas particulier $r = 0$, c'est-à-dire qu'aucune des données n'est considérée comme aberrante, $w_{(0)} = C$, $\hat{\beta}_{(0)} = \hat{\beta}^{LS}$ et $S_{(0)}^2 = S^2$.

2.5. ESTIMATEURS BAYÉSIENS ROBUSTES DE LA MOYENNE

Ayant développé, dans la section précédente, des estimateurs robustes du paramètre β , nous sommes maintenant en mesure de construire des estimateurs robustes bayésiens de la moyenne \bar{y}_U d'une caractéristique Y de la population. Nous supposons que les couples (\mathbf{x}_i, y_i) où $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ $i = 1, 2, \dots, n$ ont été prélevés par le plan $p(s)$ dont la probabilité d'inclusion de l'unité i est $p_i > 0$ et la probabilité d'inclusion simultanée des unités i et j est $p_{ij} > 0$. De plus, nous admettons que $\bar{\mathbf{x}}_U = (\bar{x}_{1U}, \bar{x}_{2U}, \dots, \bar{x}_{pU})$, $\bar{x}_{jU} = N^{-1} \sum_{i=1}^N x_{ij}$ pour $j = 1, 2, \dots, p$ est connu.

Avant d'aller plus loin, il convient de noter que les estimateurs de β donnés par les équations (2.4.5) et (2.4.11) de la section précédente dépendent tous les deux de l'estimateur des moindres carrés de β . Cependant, ces estimateurs ne tiennent pas compte du plan d'échantillonnage par lequel les couples (\mathbf{x}_i, y_i) ont été tirés. Pour tenir compte du plan d'échantillonnage dans l'estimation de la

moyenne \bar{y}_U , nous proposons d'utiliser l'estimateur des moindres carrés pondérés de β (voir chapitre 1, équation 1.1.6) à la place de l'estimateur des moindres carrés ordinaire dans les calculs de ces estimateurs.

2.5.1. Estimateur hiérarchique de la moyenne

Dans cette sous-section, nous présentons un estimateur robuste de la moyenne de la caractéristique Y basé sur le modèle linéaire hiérarchique que nous avons étudié dans la section précédente.

Définition 2.5.1. *Supposons que l'on a le modèle défini par les équations (2.4.2), (2.4.3) et (2.4.4), l'estimateur hiérarchique pondéré de β , noté $\hat{\beta}^{HP}$, est donné par*

$$\hat{\beta}^{HP} = [I_p - Q^* G_l Q^{*-1}] \hat{\beta}^{MP}, \quad (2.5.1)$$

où

$$\begin{aligned} \hat{\beta}^{MP} &= (\mathbf{X}^{*t} \mathbf{X}^*)^{-1} \mathbf{X}^{*t} Y^*, \\ G_l &= \text{diag}[g_l(z_1^2/(2\sigma^2)), \dots, g_l(z_p^2/(2\sigma^2))], \\ g_l(x) &= \begin{cases} \frac{l-1/2}{l+1/2}, & \text{si } x = 0, \\ \frac{\gamma(l+1/2, x)}{x\gamma(l-1/2, x)}, & \text{sinon,} \end{cases} \end{aligned}$$

où A est une matrice diagonale avec les éléments diagonaux égaux à $1/[p_i\sigma^2]$, $\gamma(m, a) = \int_0^a x^{m-1} e^{-x} dx$, $\mathbf{z}_{(1)} = (z_1, \dots, z_p)^t = Q^{*-1} \hat{\beta}^{MP}$, Q^* correspond à la matrice de décomposition de Cholesky de $(\mathbf{X}^{*t} \mathbf{X}^*)^{-1}$, avec $\mathbf{X}^* = A^{1/2} \mathbf{X}^*$ et $Y^* = A^{1/2} Y$.

Ayant cet estimateur robuste de β , nous pouvons à présent définir un estimateur de la moyenne de façon analogue avec l'estimateur *HTR* de Hulliger.

Définition 2.5.2. *L'estimateur bayésien hiérarchique de la moyenne \bar{y}_U , noté T_{BH} est donné par*

$$T_{BH} = \bar{\mathbf{x}}_U \hat{\beta}^{HP}, \quad (2.5.2)$$

où $\bar{\mathbf{x}}_U = (\bar{x}_{1U}, \bar{x}_{2U}, \dots, \bar{x}_{pU})$, $\bar{x}_{jU} = N^{-1} \sum_{i=1}^N x_{ij}$ pour $j = 1, 2, \dots, p$ et $\hat{\beta}^{HP}$ est donné par l'équation (2.5.1).

Dans le cas particulier $p = 1$, l'estimateur T_{BH} se résume à

$$\begin{aligned} T_{BH} &= N^{-1} \sum_{i=1}^n w_i y_i, & (2.5.3) \\ w_i &= \left[\frac{N \bar{x}_U (1 - G_l(z^2/2\sigma^2))}{\sum_{i=1}^n x_i^2/p_i} \right] x_i/p_i, \\ G_l(x) &= \begin{cases} \frac{l-1/2}{l+1/2}, & \text{si } x = 0, \\ \frac{\gamma(l+1/2, x)}{x\gamma(l-1/2, x)}, & \text{sinon,} \end{cases} \\ z &= \frac{\sum_{i=1}^n x_i y_i/p_i}{\sqrt{\sigma^{-2} \sum_{i=1}^n x_i^2/p_i}}. \end{aligned}$$

Tout comme l'estimateur *HTR*, l'estimateur T_{BH} tient sa robustesse à travers la pente $\hat{\beta}^{HB}$ qui est robuste. Toutes les propriétés de $\hat{\beta}^{HB}$ lui sont donc conférées. D'une part, il performe mieux que l'estimateur $T_{HT} = \bar{\mathbf{x}}_U \beta_{HT}$ dans un voisinage de 0.

Théorème 2.5.1. *Si $p = 1$,*

$$MSE(0, \bar{x}_U \hat{\beta}^{HP}) < \bar{x}_U^2 \sigma^2 \text{tr}(\mathbf{X}^{*t} \mathbf{X}^*)^{-1} = MSE(0, T_{HT}). \quad (2.5.4)$$

Démonstration. Par définition de MSE , nous avons

$$\begin{aligned} MSE\left(0, \bar{x}_U \hat{\beta}^{HP}\right) &= \mathbf{E}\left(\bar{x}_U \hat{\beta}^{HP}\right)^2 \\ &= \bar{x}_U^2 \mathbf{E}\left(\hat{\beta}^{HP}\right)^2 \\ &= \bar{x}_U^2 MSE\left(0, \hat{\beta}^{HP}\right). \end{aligned}$$

En utilisant le théorème 2.4.1 où $MSE\left(0, \hat{\beta}^{HP}\right) < MSE\left(0, \beta_{HT}\right)$.

Nous avons donc

$$MSE\left(0, \bar{x}_U \hat{\beta}^{HP}\right) < \bar{x}_U^2 \left(\sigma^2 \sum_{i=1}^n x_i^2 / p_i\right) = MSE(0, T_{HT}).$$

□

D'autre part, il a l'avantage de tenir compte de toutes les données et tend à assurer sa robustesse en diminuant la fonction de poids associées aux valeurs aberrantes, en plus d'être facile à calculer.

Comme estimateur robuste de \bar{y}_U , Duchesne (1999) a développé un estimateur de calage robuste (T_{RQR}). Un estimateur de calage de \bar{y}_U s'écrit sous la forme de $N^{-1} \sum_s w_i y_i$, où les poids w_i doivent être aussi près que possible des poids d'échantillonnage $d_i = p_i^{-1}$ mais en respectant les contraintes de calage, notées CE , $\sum_s w_i x_i = T_x$, où $T_x = N \bar{x}_U$. La démarche de Duchesne (1999) consiste à trouver des poids $w_i \in [U, L]$ qui assurent en même temps la robustesse et le respect des contraintes CE . Plus spécifiquement, il considère la minimisation de la somme $\sum_s G(w_i; q_i; r_i)$, tel que $\sum_s w_i x_i = T_x$ et $w_i \in [U, L]$, où G est une métrique quadratique restreinte et q_i et r_i sont des constantes obtenues de façon robuste à partir de l'échantillon. Il pose alors $T_{RQR} = \sum_s w_i y_i$. Un des problèmes de cette méthodologie est que l'existence d'une solution au problème de minimisation précédent n'est pas garantie. Il peut arriver par exemple des réalisations de

l'échantillon pour lequel on ne peut même pas satisfaire les contraintes CE (voir l'étude de simulation de Stukel, Hidiraglou et Särndal, 1996). Un autre désavantage est que les poids w_i obtenus par cette procédure peuvent être négatifs (voir Duchensne, 1999).

Pour comparer l'estimateur T_{BH} avec l'estimateur T_{RQR} , prenons l'expression de T_{BH} donnée dans l'équation (2.5.3) et supposons que x est une mesure de taille. Puisqu'on peut facilement montrer que $0 \leq G_l \leq 1$, alors au contraire de l'estimateur T_{RQR} , l'estimateur T_{BH} n'est pas sujet à des problèmes de pondération négative tout en utilisant des poids qui sont obtenus de façon robuste. Cependant, il faut noter que l'estimateur T_{BH} n'est pas pour autant un estimateur de calage car rien n'assure *a priori* que les contraintes CE sont respectées.

2.5.2. Estimateur de paramètre de dispersion augmenté de la moyenne

Cette sous-section est consacrée à un estimateur de la moyenne basé sur le modèle de dispersion augmenté de Box et Tiao (1968). Cet estimateur de la moyenne est défini de façon analogue avec l'estimateur T_{BH} . Par souci de simplicité, nous supposons que l'échantillon ne contient qu'une seule valeur aberrante ($r = 1$). Pour commencer, nous donnons la définition d'un estimateur de la pente β basé sur les moindres carrés pondérés.

Définition 2.5.3. *Supposons qu'on a le modèle de dispersion augmenté de la sous-section 2.4.2, alors l'estimateur pondéré du modèle de dispersion augmenté de β , noté $\hat{\beta}^{DP}$, est donné par*

$$\hat{\beta}^{DP} = (\bar{\beta}_1, \bar{\beta}_2, \dots, \bar{\beta}_p)^t \quad (2.5.5)$$

où $\bar{\beta}_l$ est donné dans l'équation (2.4.11) dans laquelle \mathbf{X} et Y sont remplacés par \mathbf{X}^* et Y^* respectivement.

Définition 2.5.4. Soit $\hat{\beta}^{DP}$, l'estimateur donné par l'équation (2.5.5), alors l'estimateur du modèle de dispersion pondéré de \bar{y}_U , noté T_{DA} , est donné par

$$T_{DA} = \bar{\mathbf{x}}_U \hat{\beta}^{DP}, \quad (2.5.6)$$

c'est-à-dire,

$$T_{DA} = \bar{\mathbf{x}}_U \left(\sum_{i=0}^n w_{(i)} \hat{\beta}_{(i)} \right),$$

où

$$\begin{aligned} \hat{\beta}_{(i)} &= \hat{\beta}^{LS} - \phi(\mathbf{X}^{*t} \mathbf{X}^*)^{-1} \mathbf{X}_{(i)}^{*t} [\mathbf{I}_{(i)} - \phi \mathbf{X}_{(i)}^* (\mathbf{X}^{*t} \mathbf{X}^*)^{-1} \mathbf{X}_{(i)}^{*t}]^{-1} \\ &\quad \times (Y_{(i)}^* - \mathbf{X}_{(i)}^* \hat{\beta}^{LS}), \\ w_{(i)} &= C \left(\frac{\alpha}{1-\alpha} \right) k^{-1} \frac{|\mathbf{X}^{*t} \mathbf{X}^*|^{1/2}}{|\mathbf{X}^{*t} \mathbf{X}^* - \phi \mathbf{X}_{(i)}^{*t} \mathbf{X}_{(i)}^*|^{1/2}} \left(\frac{S_{(i)}^2}{S^2} \right)^{-\nu/2}, \\ \nu S_{(i)}^2 &= \nu S^2 - \phi (Y_{(i)}^* - \mathbf{X}_{(i)}^* \hat{\beta}^{LS})^t [\mathbf{I}_{(i)} - \phi \mathbf{X}_{(i)}^* (\mathbf{X}^{*t} \mathbf{X}^*)^{-1} \mathbf{X}_{(i)}^{*t}]^{-1} \\ &\quad \times (Y_{(i)}^* - \mathbf{X}_{(i)}^* \hat{\beta}^{LS}), \\ \hat{\beta}^{LS} &= (\mathbf{X}^{*t} \mathbf{X}^*)^{-1} \mathbf{X}^{*t} Y^*, \end{aligned}$$

où $\mathbf{X}_{(i)}^* = A_i^{1/2} \mathbf{X}_{(i)}$, $Y_{(i)}^* = A_i^{1/2} Y$, $A_i^{1/2}$ désigne la ligne de la matrice A correspondant à la valeur aberrante i et $\hat{\beta}^{LS}$ et S^2 représentent l'estimateur des moindres carrés pondéré de β et de σ^2 respectivement.

Cet estimateur est robuste puisqu'on met, à travers les poids $w_{(i)}$, les poids appropriés dans le bon modèle. En effet, si i est la bonne valeur aberrante, $S_{(i)}^2$

sera alors petit. On aura donc un poids w_i grand. La variance de cet estimateur est donnée dans le théorème suivant.

Théorème 2.5.2. *La variance de l'estimateur T_{DA} est donnée par*

$$\mathbf{V}(T_{DA}) = \bar{\mathbf{x}}_U \Sigma_D \bar{\mathbf{x}}_U^t, \quad (2.5.7)$$

où les éléments lm de Σ_D sont donnés par

$$\frac{n-p}{n-p-2} \sum_{i=0}^n w_{(i)} [S_{(i)}^2 v_{(i)}^{lm} + (\hat{\beta}_{l(i)} - \bar{\beta}_l)(\hat{\beta}_{m(i)} - \bar{\beta}_m)^t]$$

et $v_{(i)}^{lm}$ désigne l'élément lm de la matrice $(\mathbf{X}^{*t} \mathbf{X}^* - \phi \mathbf{X}_i^{*t} \mathbf{X}_i^*)^{-1}$.

Démonstration : Par la définition de l'estimateur T_{DA} , nous trouvons

$$\begin{aligned} \mathbf{V}(T_{DA}) &= \mathbf{V}(\bar{\mathbf{x}}_U \hat{\beta}^{DP}) \\ &= \bar{\mathbf{x}}_U \mathbf{V}(\hat{\beta}^{DP}) \bar{\mathbf{x}}_U^t. \end{aligned}$$

Et le résultat s'ensuit en utilisant le fait que la loi *a posteriori* de β est une fonction de densité d'une loi t de Student multidimensionnelle avec matrice de variance-covariance $S_{(i)}^2 (\mathbf{X}^{*t} \mathbf{X}^* - \phi \mathbf{X}_i^{*t} \mathbf{X}_i^*)^{-1}$. \square

Si on pense qu'il y a plus d'une observation aberrante, disons $r > 1$, on peut accroître la robustesse de l'estimateur T_{DA} en considérant les $r + 1$ premières sommations de l'équation (2.4.10). Cependant, plus le nombre d'observations aberrantes est élevé plus le calcul de l'estimateur T_{DA} sera long et fastidieux.

Avant de conclure cette section, notons que des estimateurs du total $\sum_U y_i$ s'obtient facilement en multipliant les estimateurs T_{HP} et T_{DP} par N .

Contrairement aux estimateurs T_{HTR} et T_{REM} de Hulliger, les estimateurs T_{BH} et T_{DA} ont des expressions analytiques explicites, ce qui fait qu'ils sont plus facile à calculer, en plus d'être robustes par rapport aux valeurs éloignées en x et partiellement robustes par rapport aux valeurs aberrantes en y .

Nous avons présenté, dans ce chapitre, un bref rappel de la théorie bayésienne. Une modélisation générale des modèles de régression linéaire bayésiens y est étudiée de façon sommaire. De cette modélisation, nous dérivons deux estimateurs bayésiens de la moyenne, un tiré du modèle hiérarchique et l'autre obtenu à l'aide d'un modèle de paramètre de dispersion augmenté. Nous exposons dans le chapitre suivant, les résultats obtenus lors d'une étude de simulation à l'aide des méthodologies introduites dans ce chapitre. Nous y relatons aussi les résultats fournis par les estimateurs du chapitre 1.

Chapitre 3

ANALYSE DES RÉSULTATS

Dans ce chapitre, nous étudions la performance des estimateurs bayésiens vus au chapitre précédent à l'aide d'une étude par simulation de Monte Carlo. Nous avons considéré des populations où les informations d'une seule variable auxiliaire x est disponible pour toute la population. Nous avons aussi considéré huit populations composées de données provenant d'une part, de l'article de Hulliger (1995) et d'autre part, du livre de Särndal, Swensson et Wretman (1992). Pour chacune des populations, deux mille ($K=2000$) échantillons de taille $n = 16$ ont été tirés par le plan d'échantillonnage proposé par Dey et Srivastava (1987), c'est un plan à probabilités d'inclusion proportionnelles à une mesure de taille. Notre objectif est de savoir si les estimateurs bayésiens décrits à la section 2.5 possèdent de bonnes propriétés empiriques (biais, écart quadratique moyen). Notons que tous les programmes ont été écrits en S-PLUS (Statistical Sciences, 1991) et ils sont donnés dans l'annexe A.

3.1. POPULATIONS À L'ÉTUDE

Les graphiques des populations se retrouvent dans la Figure 3.1.1. Quelques statistiques descriptives des populations sont aussi données dans le Tableau 3.1.1 où r désignent le coefficient de corrélation entre x et y . La taille des six premières

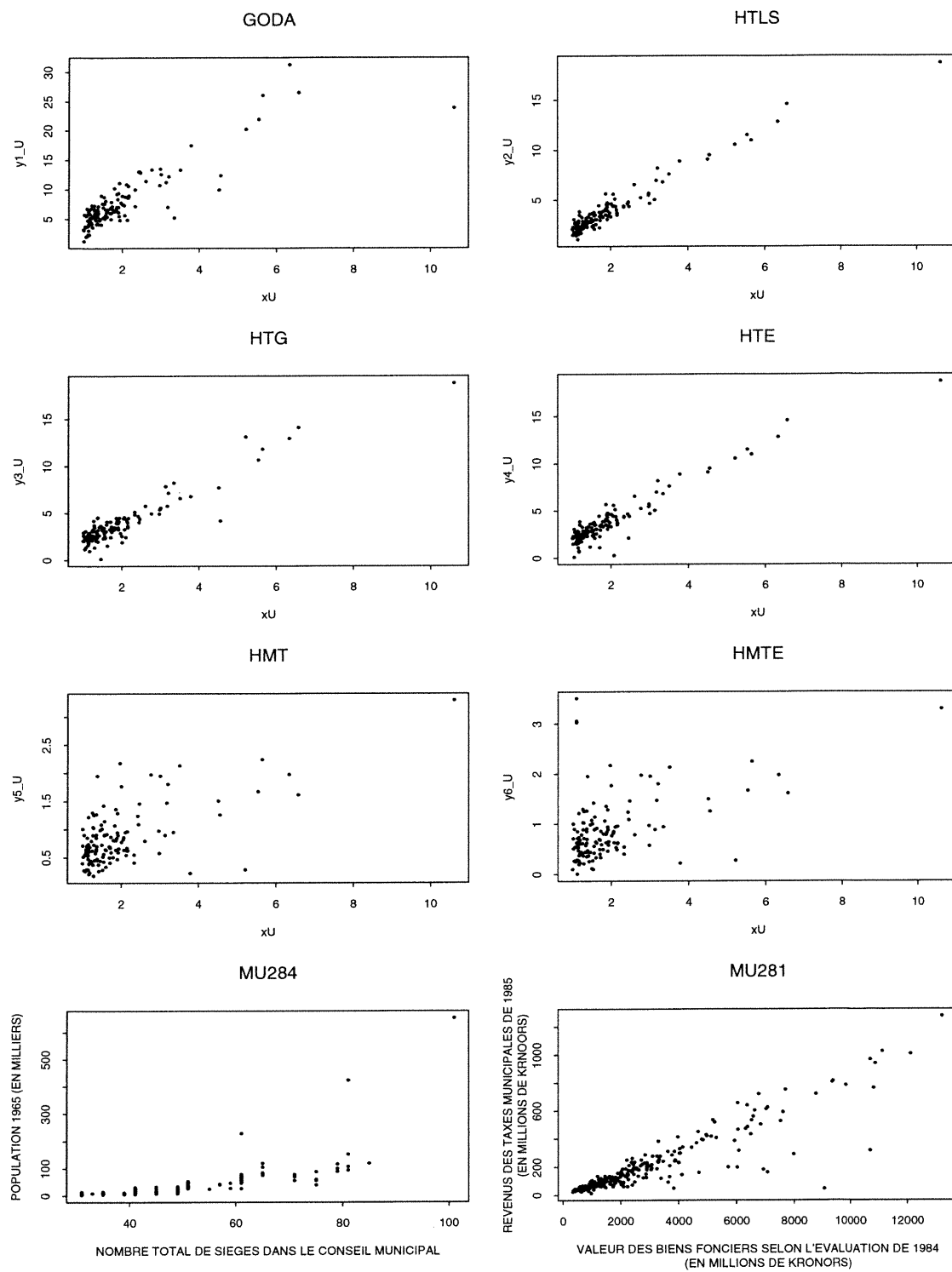


FIGURE 3.1.1 – Les huit populations à l'étude

populations est de $N=130$ tandis que celle de la septième population est de 284 et celle de la dernière est de 281.

TABLEAU 3.1.1 – Statistiques descriptive sur les populations à l'étude

Population	\bar{x}_U	\bar{y}_U	N	r
GODA	1,938	7,609	130	0,872
HTLS	1,938	3,908	130	0,969
HTG	1,938	3,806	130	0,938
HTE	1,938	3,787	130	0,950
HMT	1,938	0,833	130	0,636
HMTE	1,938	0,872	130	0,459
MU284	47,535	29,363	284	0,685
MU281	2695,889	189,301	281	0,912

Pour les six premières populations, les $x_i (i = 1, \dots, N)$ sont des réalisations indépendantes selon une loi exponentielle à contamination d'échelle de 5% ayant son origine en 1, c'est-à-dire, $(X_i - 1) \sim 0,95Exp(1) + 0,05Exp(3)$, où $Exp(\beta)$ désigne la fonction de distribution exponentielle : $F(x) = 1 - \exp(-x/\beta)$. Le facteur +1 est introduit pour abaisser la probabilité de réponses négatives dans la régression par l'intermédiaire du modèle d'origine à erreurs symétriques.

La première réponse, $y_U^{(1)}$ désignée par l'acronyme GODA, est une réalisation des variables normales indépendantes distribuées selon la formule $Y_i \sim \mathcal{N}(10x_i, x_i^2)$. Il s'agit du modèle en vertu duquel l'estimateur HT est optimal (voir Godambe, 1955). La réponse $y_U^{(2)}$ (HTLS) est une réalisation des variables

indépendantes distribuées selon la formule $Y_i \sim \mathcal{N}(2x_i, x_i/4)$. Pour ce modèle, l'estimateur T_{HT} correspond à un estimateur de moindres carrés (voir Hulliger, 1995). Une troisième réponse, $y_U^{(3)}$ (HTG) est créée par le modèle $Y_i \sim 0,95\mathcal{N}(2x_i, x_i/4) + 0,05\mathcal{N}(2x_i, 9x_i/4)$. Les valeurs aberrantes résiduelles ont un paramètre d'échelle trois fois plus grand. La réponse $y_U^{(4)}$ (HTE) comporte des valeurs aberrantes asymétriques qui ne sont pas liées à la variable x . Les données (122 observations) proviennent en majorité de la distribution $Y_i \sim \mathcal{N}(2x_i, x_i/4)$ de $y_U^{(2)}$, mais 8 observations choisies au hasard viennent de la distribution $Exp(2,5)$. La population $y_U^{(5)}$ (HMT) vient d'une distribution assortie d'une moyenne $0,4 + 0,25x_i$; elle a une distribution gamma dont la variance est proportionnelle à $x^{3/2}$. Ainsi la variable y présente la distribution proposée par Hansen, Madow et Tepping (1983, page 781). Finalement, une population $y_U^{(6)}$ (HMTE) est générée avec 122 observations tirées de la même distribution que $y_U^{(5)}$, mais comportant 8 observations choisies au hasard de la distribution $Exp(2)$. Les six populations décrites ci-haut sont choisies pour leur réalisme (voir, Hulliger, 1995). Elles utilisent toutes la même population de valeurs x (voir Figure 3.1.1).

L'avant dernière population est la population MU284 de Särndal, Swensson et Wretman (1992) qui porte sur les 284 municipalités de Suède. Nous considérons les variables $x = S82$ concernant le nombre total de sièges dans le conseil municipal et $y = P85$ la population des villes suédoises en 1985. En dernier lieu, nous analysons la population MU281 consistant de MU284 de laquelle les trois plus grosses municipalités ont été exclues. Les variables considérées sont $x = REV84$ qui représente les valeurs des biens fonciers suivant l'évaluation de 1984 et $y = RMT85$ qui représente les revenus des taxes municipales en 1985.

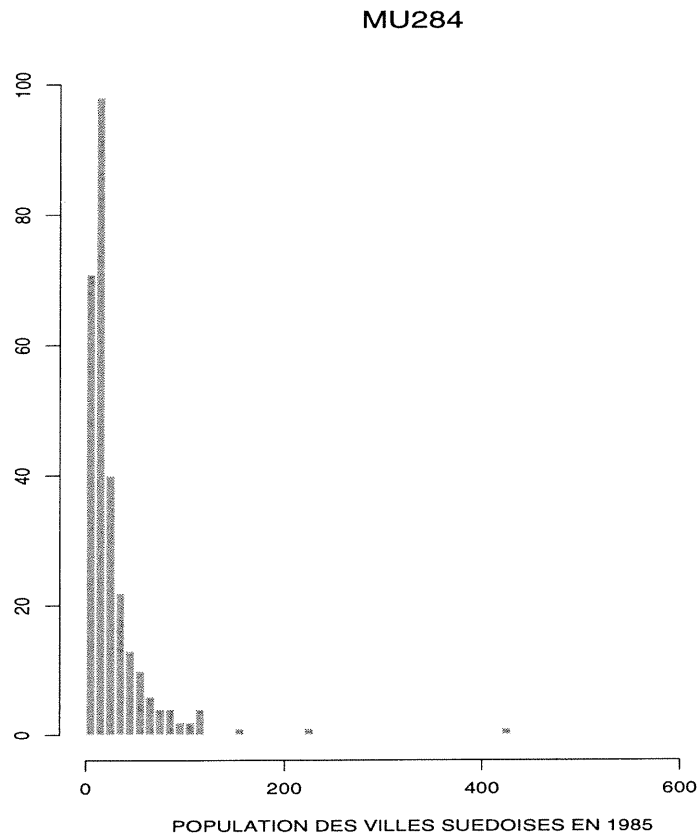


FIGURE 3.1.2 — *Histogramme de la population M284*

L'unité de mesure est le million de kronors pour ces deux variables. Les histogrammes des caractéristiques y de ces deux populations se retrouvent dans les Figures 3.1.2 et 3.1.3.

Avant de continuer à présenter l'analyse des résultats, notons que pour étudier le comportement des estimateurs, il convient de distinguer les valeurs aberrantes influentielles (observations possédant un grand résidus) des mauvais points de levier. Ces derniers sont des observations possédant une valeur h_{ii} élevée, où h_{ii}

désigne le i -ième élément de la diagonale de la matrice $\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$ et \mathbf{X} est la matrice des informations auxiliaires (voir section 2.5). Nous poursuivons notre discussion en motivant le choix de ces 8 populations. Il est clair qu'elles comportent toutes des valeurs aberrantes en x . Des mauvais points de levier semblent être présents dans toutes les populations. Les valeurs aberrantes en y sont présentes dans les populations HMTE, MU284 et MU281. Les quatre premières populations proviennent en majorité de lois normales. Pour étudier le comportement des estimateurs T_{BH} et T_{DA} sous des données ne provenant pas de lois normales, nous avons donc choisi les populations HMT et HMTE dont les données proviennent en majorité de loi gamma, et MU284 et MU281 qui sont très asymétriques vers la droite (voir Figures 3.1.2 et 3.1.3).

Comme nous l'avons déjà mentionné, nous utilisons le plan d'échantillonnage proposé par Dey et Srivastava (1987) (à noter qu'il manque un facteur de 2 dans leur formule (2.3)). Dey et Srivastava (1987) proposent de constituer des groupes $m > n/2$. Les totaux de groupes U_1, \dots, U_m où $\sum_{U_j} x_i (j = 1, \dots, m)$ doivent répondre à l'inégalité $\sum_{U_j} x_i / \sum_U x_i > (n - 2) / (n(m - 1))$. Ainsi, on ne tolère pour ces groupes qu'une très faible variabilité, et ils sont difficiles à former, en particulier dans le cas des grands échantillons (voir Hulliger 1991, page 179). C'est un plan d'échantillonnage sans remise avec probabilités inégales de sélection. Ce plan assure que les probabilités de sélection sont proportionnelles à la variable x .

3.2. DESCRIPTIONS DES ESTIMATEURS

Les trois estimateurs de base de la moyenne \bar{y}_U sont l'estimateur d'Horvitz-Thompson (T_{HT} , voir l'équation 1.1.1) et les deux autres estimateurs étudiés dans le chapitre 1. Ces derniers sont notés par T_{HTR} (voir l'équation 1.2.1) et T_{REM}

MU281

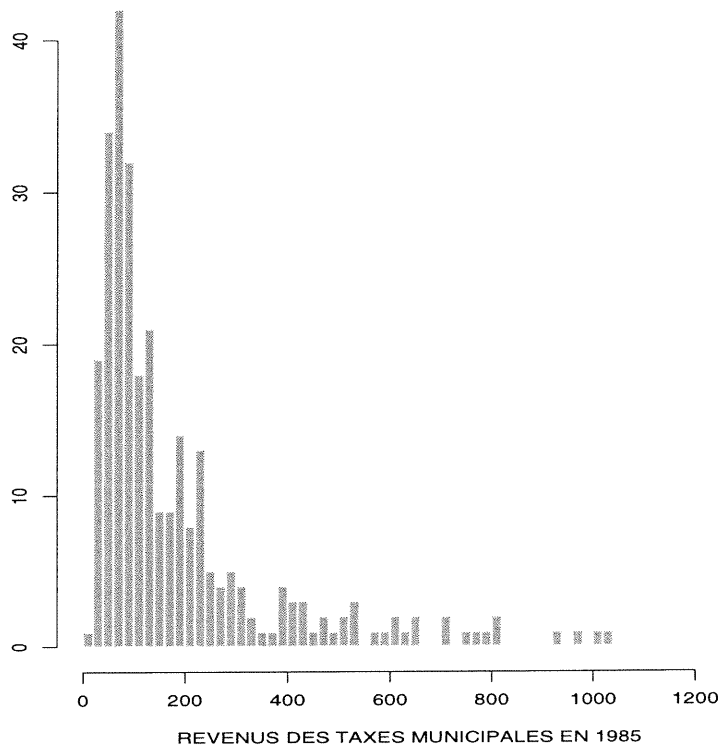


FIGURE 3.1.3 -. Histogramme de la population M281

(voir l'équation 1.3.1). Pour l'estimateur T_{HTR} , nous avons utilisé

$$\eta(x'_i, r'_i) = w(x'_i, k_x) \psi(r'_i, k_r \text{med}_S |r'_i|),$$

avec $w(x'_i, k_x) = \min(1, k_x \text{med}_U |x_i| / |x'_i|)$, $k_x = k_r = 2$, $\text{med}_S |r'_i|$ est la médiane des résidus absolus divisés par $\sqrt{x_i}$ sur l'échantillon S et $\text{med}_U |x_i|$ est la médiane des valeurs absolues des x_i sur la population U . La fonction de pondération $w(x'_i, k_x)$ correspond à une fonction de Huber asymétrique $\psi_{Hub}(x'_i, k_x) = \min(x'_i, k_x)$ qui donne un poids moins grand aux grandes valeurs de x'_i . Le terme $\text{med}_S |r'_i|$ d'échelle

est la médiane des résidus absolus; ces résidus sont calculés à l'aide de la solution de l'itération précédente de l'algorithme des moindres carrés pondérés par itérations. L'estimateur T_{REM} utilise la même fonction η avec les constantes de miss au point k_x et k_r , évaluées en 20 points, qui se trouvent sur la diagonale de l'étendue de k_x et de k_r , c'est-à-dire k_x et k_r sont reliées de façon linéaire avec une pente positive. Comme les estimateurs T_{REM} (voir chapitre, 1 section 3) ont donné des résultats légèrement meilleurs avec v_{rHTR} (voir l'équation 1.2.6) qu'avec les autres estimateurs de variance (voir Hulliger (1995), page 93), les simulations ont été faites avec v_{rHTR} . Les estimateurs bayésiens étudiés dans le chapitre précédent sont notés respectivement par T_{BH} et T_{DA} pour le modèle hiérarchique (voir l'équation 2.5.2) et le modèle de paramètre de dispersion augmenté (voir l'équation 2.5.6). Pour l'estimateur T_{BH} , nous avons utilisé la constante $l = 2,5$ et σ^2 a été remplacé par son estimateur des moindres carrés pondérés. Ce choix de l devrait conduire à une diminution de l'écart quadratique moyen quand la pente de l'estimateur T_{BH} est près zéro (voir Angers et Macgibbon, 1996). Ce choix pour σ^2 est fait dans un souci de faciliter les calculs de l'estimateur T_{BH} lors des simulations en utilisant une approche empirique (voir section 2.3). Finalement, nous avons pris des constantes α et k égale à 0,05 et 5 respectivement dans les calculs de l'estimateur T_{DA} . Le choix de la constante α est basé sur la suggestion de Box et Tiao (1968); ils proposent de prendre α dans l'intervalle $[0,01; 0,1]$. En outre, l'estimateur T_{DA} devrait être insensible au changement de la valeur de k dans l'intervalle $3 \leq k \leq 10$ (voir Box et Tiao, 1968).

3.3. MESURES FRÉQUENTISTES

Pour chaque échantillon, les 4 estimateurs de la section 3.2 sont calculés. Les résultats sont donnés dans le Tableau 3.3.1. L'aspect important auquel nous nous

intéressons est relié au biais et à l'efficacité des deux méthodes proposées. Soit \hat{T} un estimateur de la moyenne \bar{y}_U . Posons \hat{T}_k l'estimateur de la moyenne calculé à l'aide de l'échantillon k , $k = 1, \dots, K$. Pour l'estimateur \hat{T} , le biais relatif de Monte Carlo $BR(\hat{T})$ par rapport à \hat{T}_{HT} , la moyenne $E(\hat{T})$ et la variance $V(\hat{T})$ sont donnés par les formules usuelles, c'est-à-dire

$$BR(\hat{T}) = \frac{E(\hat{T}) - \bar{y}_U}{\hat{T}_{HT}},$$

$$E(\hat{T}) = \frac{1}{K} \sum_{k=1}^K \hat{T}_k,$$

$$V(\hat{T}) = \frac{1}{K} \sum_{k=1}^K (\hat{T}_k - E(\hat{T}))^2,$$

où \hat{T}_{HT} est la moyenne HT . Nous définissons l'écart type relatif de \hat{T} par rapport à l'estimateur \hat{T}_{HT} $ETR(\hat{T})$ par

$$ETR(\hat{T}) = \frac{\sqrt{V(\hat{T})}}{\hat{T}_{HT}}.$$

Le critère principal d'efficacité sera l'écart quadratique moyen de Monte Carlo, défini par

$$EQM(\hat{T}) = \frac{1}{K} \sum_{k=1}^K (\hat{T}_k - \bar{y}_U)^2.$$

Ainsi, nous définissons l'efficacité d'un estimateur, noté $Eff(\hat{T})$, par le quotient de $EQM(\hat{T}_{HT})$ de l'estimateur T_{HT} et de $EQM(\hat{T})$ de l'estimateur \hat{T} . Il convient de noter que toute l'analyse qui suit se fait par rapport à l'estimateur d'Horvitz-Thompson.

Le Tableau 3.3.1 contient le biais et l'écart type relatif par rapport à l'estimateur T_{HT} de chacun des estimateurs ainsi que leur efficacité. Dans le Tableau 3.3.2,

nous donnons le temps de calcul par échantillon (en centièmes de secondes) des estimateurs T_{BH} , T_{DA} et T_{REM} . Ils sont notés par TC_{BH} , TC_{DA} et TC_{REM} . On retrouve aussi dans ce même Tableau le nombre de valeurs influentes (observations possédant un grand résidu) et le nombre de mauvais points de levier par échantillon. Nous utilisons la méthodologie de Chatterjee et Hadi (1986) pour détecter les valeurs influentes et les mauvais points de levier.

TABLEAU 3.3.1 – Biais relatif, écart type relatif, efficacité des estimateurs et moyenne de Monte Carlo de l'estimateur T_{HT} pour chacune des populations

Pop.	Est.	$BR(\hat{T})$	$ETR(\hat{T})$	$Eff(\hat{T})$	T_{HT}
GOGA	T_{BH}	-0,018	0,056	0,910	7, 610
	T_{DA}	-0,052	0,067	0,443	
	T_{HTR}	0,022	0,067	0,769	
	T_{REM}	0,002	0,057	0,991	
	T_{HT}	0,000	0,057	1,000	
HTLS	T_{BH}	-0,003	0,035	1,236	3,908
	T_{DA}	-0,009	0,031	1,515	
	T_{HTR}	-0,005	0,040	0,949	
	T_{REM}	0,003	0,039	0,994	
	T_{HT}	0,000	0,039	1,000	
HTG	T_{BH}	-0,004	0,047	1,187	3,809
	T_{DA}	-0,014	0,044	1,265	
	T_{HTR}	0,009	0,048	1,122	
	T_{REM}	0,005	0,050	1,062	
	T_{HT}	0,001	0,052	1,000	

Pop.	Est.	$BR(\hat{H})$	$ETR(\hat{H})$	$Eff(\hat{T})$	T_{HT}
HTE	T_{BH}	0,002	0,048	1,268	3,791
	T_{DA}	0,004	0,040	1,892	
	T_{HTR}	0,011	0,047	1,258	
	T_{REM}	0,007	0,051	1,135	
	T_{HT}	-0,001	0,055	1,000	
HMT	T_{BH}	-0,042	0,108	1,021	0,833
	T_{DA}	-0,126	0,098	0,536	
	T_{HTR}	-0,016	0,117	0,977	
	T_{REM}	-0,011	0,116	1,001	
	T_{HT}	0,000	0,117	1,000	
HMTE	T_{BH}	-0,050	0,162	1,270	0,871
	T_{DA}	-0,150	0,114	1,033	
	T_{HTR}	-0,076	0,132	1,589	
	T_{REM}	-0,048	0,146	1,557	
	T_{HT}	0,001	0,191	1,000	
MU284	T_{BH}	0,063	0,321	0,668	29,364
	T_{DA}	0,184	0,439	0,334	
	T_{HTR}	-0,228	0,152	1,010	
	T_{REM}	-0,088	0,210	1,466	
	T_{HT}	0,003	0,275	1,000	
MU281	T_{BH}	0,013	0,076	0,865	189,609
	T_{DA}	0,036	0,091	0,541	
	T_{HTR}	0,019	0,073	0,914	
	T_{REM}	0,008	0,072	0,969	
	T_{HT}	0,002	0,072	1,000	

TABLEAU 3.3.2 – Moyenne du temps de calcul, du nombre de valeurs influentes et du nombre de mauvais points de levier par échantillon pour chacune des populations

	GODA	HTLS	HTG	HTE	HMT	HMTE	MU284	MU281
TC_{BH}	2,40	2,40	2,40	2,40	2,40	2,50	2,40	3,30
TC_{DA}	10,10	10,00	10,20	10,20	10,30	10,30	10,60	14,50
TC_{REM}	127,70	121,70	124,90	126,00	133,00	138,30	175,80	182,30
INF	1,17	1,19	1,32	1,29	1,32	1,34	1,14	1,38
LEV	2,26	2,26	2,25	2,27	2,28	2,25	1,04	2,79

3.4. ANALYSE DES RÉSULTATS

L'estimateur bayésien hiérarchique est moins variable que les estimateurs robustes classiques dans les quatre populations GODA, HTLS, HTG et HMT alors que ces derniers estimateurs sont plus variables que l'estimateur T_{DA} dans les populations HTLS, HTG, HTE, HMT et HMTE. En outre, le biais relatif (en valeur absolue) de l'estimateur T_{BH} est comparable à celui de l'estimateur T_{REM} sauf pour les populations GODA, HMT et MU284. Exception faite des populations HTE et MU284, le biais relatif (en valeur absolue) de l'estimateur T_{DA} est plus grand que celui des autres estimateurs. Comme dans Hulliger (1995), le biais relatif de l'estimateur T_{REM} est toujours plus petit en valeur absolue que celui de l'estimateur T_{HTR} .

Dans la population GODA, les estimateurs bayésiens sont plus biaisés que les estimateurs de Hulliger. Les écarts types des estimateurs T_{HTR} et T_{DA} sont égaux (0,067) tandis que les autres estimateurs varient autant que l'estimateur

T_{HT} (0,057). Si l'estimateur T_{DA} perd 56% de son efficacité avec cette population, alors que l'estimateur T_{HT} devrait être optimal, l'estimateur T_{BH} n'en perd que 9%. L'estimateur T_{HTR} perd 14% d'efficacité de plus que ce dernier pendant que l'estimateur T_{REM} garde presque toute la sienne. Avec la population HTLS, où l'estimateur T_{HT} correspond à l'estimateur des moindres carrés, les deux estimateurs bayésiens gagnent respectivement 24% et 51% d'efficacité alors que les deux autres estimateurs en perdent 5% et 1% respectivement. De plus, avec un biais relatif de -0,003, l'estimateur T_{BH} possède autant de biais que l'estimateur T_{REM} .

Les estimateurs bayésiens sont plus efficaces ($Eff(\hat{T})=1,187$; $Eff(\hat{T})=1,265$) que les estimateurs de Hulliger ($Eff(\hat{T})=1,122$; $Eff(\hat{T})=1,062$) dans la population HTG qui contient des valeurs aberrantes résiduelles asymétriques. Les estimateurs robustes n'occasionnent pas d'important biais dans cette population (moins de 1,4%). Avec les valeurs aberrantes asymétriques de la population HTE, l'estimateur T_{DA} se classe au premier rang avec un gain d'efficacité de plus de 89% par rapport à l'estimateur T_{HT} alors que l'estimateur bayésien hiérarchique occupe le deuxième rang avec un gain de 27%. L'efficacité de l'estimateur T_{REM} affiche un retard de 10% par rapport à celui de l'estimateur T_{HTR} . Aussi, les nouveaux estimateurs proposés au chapitre 2 montrent un biais moins élevé que ceux de Hulliger.

Dans la population HMT, le coefficient de corrélation est très faible ($r=0,64$; voir le Tableau 3.1.1) et les erreurs ne sont pas proportionnelles à leurs variances par rapport à la variable explicative. Pour une telle population, les estimateurs T_{DA} et T_{HTR} sont respectivement 46% et 2% moins efficaces que l'estimateur d'Horvitz-Thompson. Pendant que l'estimateur à risque estimé minimal conserve

toute son efficacité, celui du modèle modèle hiérarchique bayésien en acquiert 2%. Les quatre estimateurs robustes présentent tous un biais relatif négatif non négligeable. Si dans une telle population on vient à relever quelques valeurs aberrantes asymétriques comme dans la population HMTE, tous les estimateurs robustes tirent un avantage d'efficacité par rapport à l'estimateur T_{HT} : 27%, 3%, 59% et 56%. En outre, l'estimateur T_{DA} obtient une importante réduction de variance au prix d'un biais relatif de l'ordre de -15%. Les distorsions des estimateurs T_{BH} et T_{REM} sont comparables (5%).

La population MU284 contient des aberrants verticaux et des mauvais points de levier. Les estimateurs robustes de Hulliger réduisent de manière radicale la variance, puisqu'ils ne sont pas affectés par les trois unités extrêmes en y qui s'éloignent de façon marquée de la tendance linéaire. Quant aux estimateurs bayésiens, il semble que la réduction des poids accordée à ces valeurs aberrantes ne soit pas suffisante pour diminuer leur variance par rapport à l'estimateur T_{HT} . De plus, l'estimateur T_{DA} demeure sévèrement biaisé (0,184). Toutefois, le biais relatif de l'estimateur T_{BH} est le moins élevé parmi ceux des autres estimateurs robustes classiques mais ne suffit pas pour lui transmettre un gain d'efficacité par rapport à l'estimateur T_{HT} . Bien que le biais relatif de l'estimateur T_{HTR} soit beaucoup plus grand que celui de ce dernier, la réduction de sa variance par rapport à celle de T_{HT} est suffisamment importante pour lui procurer un gain d'efficacité de 1%. Le faible biais relatif de l'estimateur T_{REM} conjugué à sa réduction de variance lui procure un gain d'efficacité de plus de 46 % par rapport à l'estimateur T_{HT} .

La population MU281 contient un nombre assez grand de mauvais points de levier (voir Tableau 3.3.2) mais le coefficient de corrélation est assez élevé ($r=0,91$; voir Tableau 3.1.1). La variance domine la part de l'EQM dans cette population. Elle est supérieure ou égale à celle de l'estimateur T_{HT} . Par conséquent, les estimateurs robustes sont moins efficaces que l'estimateur d'Horvitz-Thompson. Les estimateurs bayésiens sont les plus variables ($ETR(\hat{T}_{BH})=0,076$; $ETR(\hat{T}_{DA})=0,091$; voir Tableau 3.3.1).

Une analyse du nombre de valeurs influentes par échantillon du Tableau 3.3.2 relève que l'estimateur bayésien hiérarchique est plus efficace que l'estimateur T_{HT} dans les populations où le nombre de valeurs influentes ne dépasse pas 1% (.146/16) de la taille de l'échantillon ($n=16$). Dans ce cas, à part dans la population HMTE où le coefficient de corrélation est très faible, cet estimateur est surtout plus avantageux (en terme d'efficacité) que l'estimateur T_{REM} de Hülliger. Dans le cas contraire, mis à part la population MU284, sa perte d'efficacité par rapport à l'estimateur T_{HT} est comparable à celui de l'estimateur T_{REM} . Quant la relation entre x et y est très forte ($r>0,91$), comme dans les populations HTLS, HTG et HTE, l'estimateur basé sur le modèle de paramètre de dispersion augmenté gagne beaucoup d'efficacité par rapport à l'estimateur T_{HT} . Son gain d'efficacité dépasse largement celui des autres estimateurs robustes pour de telles populations. Toutefois, sa variabilité combinée avec un biais relatif assez considérable lui occasionne une perte d'efficacité dramatique dans le cas des autres populations.

On remarque que le temps de calcul de l'estimateur T_{REM} dépasse largement celui des estimateurs bayésiens; il est plus de 50 fois plus élevé que celui de l'estimateur bayésien hiérarchique et plus de 10 fois plus élevé que celui de l'estimateur T_{DA} .

On peut conclure de cette simulation que les estimateurs bayésiens gagnent énormément en terme d'efficacité dans les populations provenant de loi normale où x et y sont très corrélées (HTLS, HTG et HTE). L'estimateur T_{HB} perd peu en terme d'EQM comparativement à l'estimateur T_{HT} dans les populations comportant de mauvais points de levier et où le coefficient de corrélation est élevé (GODA, MU281). En présence de valeurs aberrantes, même si la population n'est pas normale (HMT), l'estimateur bayésien hiérarchique obtient un gain considérable d'efficacité sur l'estimateur d'Horvitz-Thompson. Pour sa part, l'estimateur T_{DA} est très biaisé pour les populations non normales. Cependant, il demeure moins variable que les autres estimateurs dans la plupart des populations.

Pour l'estimation de la moyenne (ou du total) d'une caractéristique y , nous suggérons d'utiliser les estimateurs bayésiens pour les populations normales où le coefficient de corrélation entre x et y est très élevé ($r > 0,90$). Ces estimateurs peuvent également être employés dans les populations non normales où le lien linéaire entre x et y est très faible ($r < 0,46$). Cependant, nous privilégions l'emploi de l'estimateur bayésien hiérarchique par le fait que l'estimateur basé sur le modèle de paramètre de dispersion augmenté peut être très biaisé. Advenant une population non normale où le coefficient de corrélation entre x et y est plus élevé que celui du cas précédent, il est raisonnable d'utiliser l'estimateur T_{REM} comme estimateur de la moyenne.

CONCLUSION

Dans ce mémoire, nous avons analysé le problème d'estimation de la moyenne dans le cas où la population à l'étude présente des valeurs aberrantes. Nous nous sommes familiarisés avec les estimateurs robustes de la moyenne de Hulliger (1995). Nous avons aussi introduit des estimateurs bayésiens trouvés à l'aide du modèle bayésien hiérarchique et du modèle de paramètre de dispersion augmenté.

Après avoir examiné les principales propriétés de l'estimateur d'Horvitz-Thompson sous le plan à probabilités d'inclusion proportionnelles à la taille, nous avons vu des estimateurs robustes de la moyenne obtenus à l'aide des M -estimateurs. Nous avons passé en revue la notion de plan d'échantillonnage pour définir ensuite l'estimateur d'Horvitz-Thompson ainsi que deux estimateurs de sa variance. Cet estimateur de la moyenne est ensuite exprimé sous la forme d'une fonctionnelle des moindres carrés pour permettre sa robustification à l'aide des M -estimateurs. Une première robustification (l'estimateur T_{HTR}) de ce dernier est obtenu par la minimisation d'une certaine somme ou d'une intégrale qui dépend à la fois des probabilités d'inclusion et d'une fonction η qui sert à diminuer l'influence des valeurs extrêmes. Une formule d'estimation de sa variance est aussi donnée. La dernière section de ce chapitre est consacré à une généralisation de l'estimateur T_{HTR} où l'on ajuste deux constantes de mise au point afin de minimiser le risque estimé. Ainsi, on arrive à l'estimateur T_{REM} qui est l'estimateur

T_{HTR} pour lequel ce risque estimé est minimum.

Au deuxième chapitre, nous avons introduit des éléments de la théorie bayésienne dans le but de formuler le modèle de régression linéaire bayésien. À partir de celui-ci, nous avons formulé deux méthodes qui permettent d'obtenir des estimateurs bayésiens robustes de la moyenne ou du total : l'estimateur bayésien hiérarchique de la moyenne (T_{BH}) et l'estimateur de paramètre de dispersion augmenté de la moyenne (T_{DA}). Nous avons élaboré une formule pour le risque fréquentiste de l'estimateur hiérarchique; la variance de l'estimateur T_{DA} est également donnée.

Au dernier chapitre, nous avons présenté une analyse et une étude comparative des propriétés empiriques de tous les estimateurs qui furent étudiés. Nous avons pu constater que les estimateurs bayésiens sont généralement moins variables que les estimateurs de Hulliger (1995) mais au prix d'une distorsion assez élevée dans la plupart des populations considérées. Toutefois, ils acquièrent une bonne efficacité surtout dans les populations où le nombre de valeurs aberrantes dans l'échantillon ne dépasse pas 1%. L'emploi des estimateurs bayésiens est surtout recommandé dans les populations normales où le coefficient de corrélation entre la variable auxiliaire x et la caractéristique y est très élevé.

Bien évidemment, nous avons consacré notre étude seulement sur le plan d'échantillonnage à probabilités d'inclusion proportionnelles à une mesure de taille. Même si nous avons obtenu des résultats que sur ce plan, la théorie présentée est suffisamment complète pour une généralisation aux autres plans d'échantillonnage. Pour le modèle hiérarchique, nous avons considéré la loi normale

comme loi de premier niveau du paramètre (la moyenne de la loi normale), on pourrait développer d'autres estimateurs bayésiens plus robustes en considérant d'autre loi avec des ailes plus aplaties que celle de la loi normale. Par conséquent, nous considérons l'estimateur hiérarchique bayésien comme un premier pas dans la robustification des estimateurs de la moyenne à l'aide de la théorie bayésienne dans le cadre des populations asymétriques. Néanmoins, les estimateurs bayésiens que nous avons vu sont intéressants étant donné leur facilité d'emploi dues à la rapidité et la simplicité de leurs calculs.

Lors des simulations du chapitre 3, nous avons considéré un taux d'échantillonnage variant entre 6 et 12 pour cent selon les populations. Il serait intéressant de pousser les études empiriques du chapitre 3, en étudiant l'effet de la taille car on sait que le lien entre la taille de l'échantillon (n) et celle de la population (N) est très étroite dans les populations de petite taille comme celles utilisées pendant les simulations.

Annexe A

LES PROGRAMMES S-PLUS

Les fonctions G10, pop2 et pipts permettent de tirer un échantillon selon le lemme de Dey et Srivastava (1987).

```
#####  
#                                                                 #  
# DIVISION DE LA POPULATION MU284 (par défaut) EN m=10 GROUPES #  
#                                                                 #  
#####  
  
# Cette fonction sert à former les m=10 groupes.  
# XY est une matrice où la première colonne est la variable  
# auxiliaire x  
# et la deuxième colonne est la variable y  
  
G10<- function( XY = X284[,c(6,1)]) {  
  XY <- XY  
  Yu <- mean(XY[,2]) ;      # c'est la caractéristique y  
  PopXu <- XY[,1]  
  X <- XY[,1]              # c'est la variable auxiliaire  
  ordre <- order(X)  
  Xod <- X[ordre]  
  YY <- XY[ordre,]  
  m <- 10; n <- 16; N <- nrow(XY) ; M <- N/m  
  epse <- (n-2)/( n*(m-1) ) # condition sur les Xi/somme(X)  
  Xu <- sum(X)  
  Y <- XX <- matrix(, M, m)  
  P <- rep(epse+.005,m)  
  
  for (i in 1:(m-1) )
```

```

{
  while ( P[i]<= epse || P[i] > epse + .001 )
  {
    id <- sample(1:(N-(i-1)*M), M, replace=F)
    P[i] <- sum(YY[id,1])/Xu
  }
  Y[,i] <- YY[id,2]; XX[,i] <- YY[id,1]
  YY <- YY[-id,]

}
P[m] <- sum(YY[,1])/Xu
Y[,m] <- YY[,2] ; XX[,m] <- YY[,1]

list( XX=XX, Y=Y, Xu=PopXu, Yu=Yu, P=P)

}

#####
# Selection de 8 groupes parmi 10 #
#####

# XX : une matrice dont chaque colonne correspond a un groupe pour la
# variable auxiliaire
# YY : une matrice dont chaque colonne correspond a un groupe pour la
# variable y
pop2<- function(XX=XX284, Y=Y284, PopXu=Xu284, Yu=Yu284,
                P=P284, n=16, m=10){
Pi <-P
Pii <- ( n*(m-1)*Pi-(n-2) )/ (2*m-n)

PPY <- Y ; PPX <- XX
id1 <- sample(1:m,1,replace=F,Pii)
Yp1 <- cbind(PPX[,id1], PPY[,id1])
Pm1 <- Pi[id1]
Pi <- Pi[-id1]
PPY <- PPY[,-id1]; PPX <- PPX[,-id1]
id2 <- sample(1:(m-1), (n/2 -1), replace=F)
GY <- PPY[,id2]; GX <- PPX[,id2]
Pm2 <- Pi[id2]

```



```

Pm <- c(Pm1,Pm2) # pour les calcul des \pi_ij #

Yp2 <- cbind( GX[,1], GY[,1]); Yp3 <- cbind( GX[,2], GY[,2] )
Yp4 <- cbind( GX[,3], GY[,3]); Yp5 <- cbind( GX[,4], GY[,4] )
Yp6 <- cbind( GX[,5], GY[,5]); Yp7 <- cbind( GX[,6], GY[,6] )
Yp8 <- cbind( GX[,7], GY[,7]);

XY <- list( Yp1=Yp1, Yp2=Yp2, Yp3=Yp3, Yp4=Yp4, Yp5=Yp5,
            Yp6=Yp6, Yp7=Yp7, Yp8=Yp8 )

list( XY = XY, Xu = PopXu, Pm = Pm, Yu = Yu )

}

#####
# TIRAGE par PIPT d'un échantillon de taille n=16 #
#                                                                 #
#####

pipts_function( l=2, XX=XX284, Y=Y284, PopXu=Xu284, Yu=Yu284,
                P=P284, n=16, m=10) {
# les groupes echantillonnes #
pop <- pop2(XX, Y, PopXu, Yu, P, n=16, m=10)
# les groupes echantillonnes #

Pm <- pop$Pm
Ymoyen <- pop$Yu ; Xu <- pop$Xu
XY <- pop$XY
Yp1 <- XY$Yp1; Yp2 <- XY$Yp2; Yp3 <- XY$Yp3; Yp4 <- XY$Yp4;
Yp5 <- XY$Yp5;
Yp6 <- XY$Yp6; Yp7 <- XY$Yp7; Yp8 <- XY$Yp8;

X1_sum(Yp1[,1]); X2_sum(Yp2[,1]); X3_sum(Yp3[,1]); X4_sum(Yp4[,1]);
X5_sum(Yp5[,1]); X6_sum(Yp6[,1]); X7_sum(Yp7[,1]); X8_sum(Yp8[,1]);

N <- nrow(Yp1); n <- 16
Dm <- rep(0,n)          ##### pour les calculs des \pi_ij ###
s <- matrix(,n,2)

```

```

YY <- Yp1
XX <- X1
id <- sample(1:N,1, replace=F, prob=YY[,1]/XX)
s[1,] <- YY[id,]
u <- s[1,1]
D <- 1 + sum( YY[,1]/(XX-2*YY[,1]))
Dm[1] <- Dm[2] <- D ;
YY <- YY[-id,]
p <- YY[,1]*( 1/(XX-2*YY[,1]) + 1/(XX-2*u) )/D
id <- sample( 1:(N-1), 1, replace=F, p)
s[2,] <- YY[id,]

```

```

YY <- Yp2
XX <- X2
id_sample(1:N,1, replace=F, prob=YY[,1]/XX)
s[3,]_YY[id,]
u<-s[3,1]
D <- 1+sum(YY[,1]/(XX-2*YY[,1]))
Dm[3] <- Dm[4] <- D
YY <- YY[-id,]
p <- YY[,1]*( 1/(XX-2*YY[,1]) + 1/(XX-2*u) )/D
id <- sample( 1:(N-1),1,replace=F,p)
s[4,]<-YY[id,]

```

```

YY <- Yp3
XX <- X3
id <- sample(1:N,1, replace=F, prob=YY[,1]/XX)
s[5,] <- YY[id,]
u <- s[5,1]
D <- 1 + sum( YY[,1]/(XX-2*YY[,1]) )
Dm[5] <- Dm[6] <- D
YY <- YY[-id,]
p <- YY[,1]*( 1/(XX-2*YY[,1]) + 1/(XX-2*u) )/D
id <- sample( 1:(N-1),1,replace=F,p)
s[6,] <- YY[id,]

```

```

YY <- Yp4
XX <- X4
id <- sample(1:N,1, replace=F, prob=YY[,1]/XX)

```

```

s[7,] <- YY[id,]
u <- s[7,1]
D <- 1+ sum( YY[,1]/(XX-2*YY[,1]))
Dm[7] <- Dm[8] <- D
YY <- YY[-id,]
p <- YY[,1]*( 1/(XX-2*YY[,1]) + 1/(XX-2*u) )/D
id <- sample(1:(N-1),1,replace=F,p)
s[8,] <- YY[id,]

```

```

YY <- Yp5
XX <- X5
id <- sample(1:N,1, replace=F,prob=YY[,1]/XX)
s[9,] <- YY[id,]
u <- s[9,1]
D <- 1+sum(YY[,1]/(XX-2*YY[,1]))
Dm[9] <- Dm[10] <- D
YY <- YY[-id,]
p <- YY[,1]*( 1/(XX-2*YY[,1]) + 1/(XX-2*u) )/D
id <- sample(1:(N-1),1,replace=F,p)
s[10,]<-YY[id,]

```

```

YY <- Yp6
XX <- X6
id <- sample(1:N,1, replace=F, prob=YY[,1]/XX)
s[11,] <- YY[id,]
u <- s[11,1]
D <- 1+ sum(YY[,1]/(XX-2*YY[,1]))
Dm[11] <- Dm[12] <- D
YY <- YY[-id,]
p <- YY[,1]*( 1/(XX-2*YY[,1]) + 1/(XX-2*u) )/D
id <- sample(1:(N-1),1,replace=F,p)
s[12,] <- YY[id,]

```

```

YY <- Yp7
XX <- X7
id <- sample(1:N,1, replace=F,prob=YY[,1]/XX)
s[13,] <- YY[id,]
u <- s[13,1]
D <- 1+ sum( YY[,1]/(XX-2*YY[,1]) )

```

```

Dm[13] <- Dm[14] <- D
YY <- YY[-id,]
p <- YY[,1]*( 1/(XX-2*YY[,1]) + 1/(XX-2*u) )/D
id <- sample(1:(N-1),1,replace=F,p)
s[14,] <- YY[id,]

YY <- Yp8
XX <- X8
id <- sample(1:N,1, replace=F,prob=YY[,1]/XX)
s[15,] <- YY[id,]
u <- s[15,1]
D <- 1+sum(YY[,1]/(XX-2*YY[,1]))
Dm[15] <- Dm[16] <- D
YY <- YY[-id,]
p <- YY[,1]*( 1/(XX-2*YY[,1]) + 1/(XX-2*u) )/D
id <- sample(1:(N-1),1, replace=F, p)
s[16,] <- YY[id,]

list(Xs = s, Xu = Xu, Dm = Dm, Pm = Pm, Ymoyen = Ymoyen)
}

dis <- function(x, y, Xu,w=rep(1,length(x)), alpha=.05, k=5, r=1)
{
# Cette fonction calcule l'estimateur de la moyenne pour
# le modèle de paramètre de dispersion augmente
# x : vecteur des informations auxiliaires
# Xu : le vecteur des variables aux. pour la pop.
# y : vecteur de reponses pour l'échantillon
# alpha : pourcentage de valeurs aberrantes
# r nombre de valeurs aberrantes dans l'échantillon doit etre 1

n <- length(x)
phi <- 1-k^2
coef.res <- lsfit(x, y, wt=w, int=F)
betaLS <-coef.res$coef
res <- coef.res$residuals
S <- var(res)
x <- x*sqrt(w/S)      # X devient SIGMA^-1/2 X
y <- y*sqrt(w/S)     # Y devient SIGMA^-1/2 Y

```

```

A <- sum(x^2)                                # X'X
B <- sum(x*y)
w.alpha <- ( alpha/(1-alpha))^r *k^(-r)
gamma.n <- gamma(n/2)
nu <- n-1
poids.wr <- function(nu, S, phi, xr,yr, betaLS, A, w.alpha){
Sr <- ( nu*S - phi*(yr-xr*betaLS)*(1 -phi*xr^2/A )^(-1)*(yr-
                                             xr*betaLS))/nu
betar <- betaLS-phi*A^(-1)*xr*( 1-phi*xr^2/A )^(-1)*(yr-xr*betaLS)
wr <- (w.alpha*sqrt(A)/sqrt(A-phi*xr^2))* (Sr/S)^(-nu/2)
list( betar=betar, wr=wr)
}
wr <- betar <- c()
for ( i in 1: n )
{
  xr <- x[i]
  yr <- y[i]
  betar[i] <-poids.wr(nu, S, phi, xr, yr, betaLS, A,w.alpha)$betar
  wr[i] <- poids.wr(nu, S, phi, xr, yr, betaLS, A,w.alpha)$wr
}
betar<-c(betaLS,betar)
wr <- c(1,wr)
beta <- sum(wr*betar)/sum(wr)
T.ME <- mean(Xu)*beta
T.REG <-T.ME +sum(beta*x)/length(Xu)
list(T.ME=T.ME, T.REG=T.REG)
}

#####
# calcul de l'estimateur hierarchique de Bayes) #
#####
glmod_fonction(z,m=2) {

# Cette fonction calcule la matrice G1 qui sert dans le calcul de
# l'estimateur bayesien hierarchique

if ( z < 7e-14 ) G1 <- m/(m+1)
else {   if ( z > 1e154) G1 <- 1
         else {

```

```

        G1 <- pgamma(z,m+1)*gamma(m+1)/( z*
        gamma(z,m)*gamma(m))
    }
}
G1
}

```

```

B.H <- function (x, y, Xu, l= 2.5, w=rep(1,length(x)) ) {

# Cette fonction calcule l'estimateur bayesien hierarchique
# de la moyenne
# x, y sont des vecteurs, w: poids

mu.chap <- mean(y/x) # pour avoir le model developpe #
yy <- y - x *mu.chap
coef.res <- lsfit(x, yy, wt=w, int=F)
tethaLS <- coef.res$coef
res <- coef.res$residuals
sigma <- var(res)
z2 <- tethaLS^2/(2*sum(x^2*w))
m <- l-1/2
G1 <- glmod(z2)
tethaH <- (1-G1)*tethaLS
T.BH <- mean(Xu)*(tethaH+ mu.chap)

list( T.BH=T.BH, gl=1-G1, tet=tethaLS, mu=mu.chap,
      z2=z2, sigma=sigma)}

```

```

WX_fonction( x, Xu, Kx = 2) {

# fonction de poids pour le calcul de l'estimateur
# HTR et REM de Hulliger
# x : variables auxiliaires sur l'echantillon
# Xu: variables auxiliaires sur la population

medu <- median(sqrt( abs(Xu) ));
wx <- x;
a <- Kx*medu/sqrt(abs(x))
U <- ( a > 1 )

```

```

wx[U] <- 1
wx[!U] <- a[!U]
wx <- wx
wx
}

wt.FV <-function(u, c = 1.345)
{

#####
# Psi de Hubert modifie #
#####

a<- c * median(abs(u))
Ugtc <- (abs(u) < a)
w <- u
w[Ugtc] <- u[Ugtc]
w[!Ugtc] <- sign(u[!Ugtc])*a
abs(w)
}

prob <- function(n=16, m=10, p=rep(1,n), Pm=rep(1,n/2), Dm=rep(1,n) )
{
# Cette fonction permet de trouver les probabilités d'inclusion des
# unités i et j (plan de Dey et de Siravastava)
D <- Dm
PIJ <- matrix(,n,n)
pp <- p/n
P <- Pm[c(1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8)]
for ( i in 1:(n-1) )
{
PIJ[i,i] <- p[i]
for( j in i+1:(n-i) )
{
if ( i%%2 != 0 & j==i+1) PIJ[i,j] <- PIJ[j,i] <- 2*n*
pp[i]*pp[j]*( P[i]-pp[i]-pp[j] )/( D[i]*(P[i]-2*pp[i])*
(P[i]-2*pp[j]) )

else PIJ[i,j] <- PIJ[j,i] <- n*(n-2)*pp[i]*pp[j]*( (m-1)*

```

```

                                (P[i]+P[j])-1 )/( (m-1)*(m-2)*P[i]*P[j] )
                                }
                                }

PIJ[n,n] <- p[n]
PIJ
}

rreg1 <- function(x, y, w = rep(1, n), int = TRUE,
init = lsfit.simp(x, y, w, n, p)$coef, bob=2,
method = wt.default, wx, iter = 20,
acc = 10 * .Machine$single.eps^0.5,
test.vec = "resid")
{

  irls.delta <- function(old, new)
  {
    a <- sum((old - new)^2)
    b <- sum(old^2)
    if(b >= 1 || a < b * .Machine$double.xmax)
      sqrt(a/b)
    else .Machine$double.xmax
  }
  irls.rrxwr <- function(x, w, r)
  {
    w <- sqrt(w)
    max(abs((as.vector(r * w) %*% x)/
      sqrt(as.vector(w) %*% (x^2))))/
      sqrt(sum(w * r^2))
  }
  lsfit.simp <- function(x, y, wt, n, p)
  {
    wt.factor <- as.vector(wt^0.5)
    wt.zero <- wt.factor == 0
    x0 <- x[wt.zero, , drop = F]
    y0 <- y[wt.zero]
    x <- x * wt.factor

```



```

y <- y * wt.factor
inv.wt.factor <- 1/ifelse(wt.zero, 1, wt.factor)
z <- .Fortran("dqrls",
             qr = as.double(x),
             as.integer(c(n, p)),
             pivot = as.integer(1:p),
             qraux = double(p),
             y,
             as.integer(c(n, 1)),
             coef = double(p),
             residuals = y,
             qt = y,
             tol = as.double(1e-07),
             double(2 * p),
             rank = as.integer(p))[c("coef", "
             residuals", "pivot", "rank")]
if(z$rank < p) {
  xn <- names(z$coef)
  z$coef <- z$coef[z$pivot]
  names(z$coef) <- xn
}
z$residuals <- z$residuals * inv.wt.factor
if(any(wt.zero)) {
  z$residuals[wt.zero] <- y0 - x0 %*% z$coef
}
z
}
if(!(any(test.vec == c("resid", "coef", "w", "NULL")) ||
     is.null( test.vec)))
  stop("invalid testvec")
if(int)
  x <- cbind("(Intercept)" = 1, x)
else x <- as.matrix(x)
cnames <- dimnames(x)[[2]]
n <- dim(x)[1]
p <- dim(x)[2]
if(length(y) != n)
  stop("length of y is not equal to number of rows in x")
specials.x <- !is.finite(x %*% rep(1, p))

```

```

specials.y <- !is.finite(y)
if(missing(wx)) {
  specials.wt <- F
}
else {
  if(length(wx) != n)
    stop("Length of wx must equal number of observations")
  specials.wt <- !is.finite(wx)
  if(any(wx[!specials.wt] < 0))
    stop("Negative wx value")
  w <- w * wx
}
ok <- !(specials.x | specials.y | specials.wt)
if((bad.obs <- sum(!ok)) > 0)
  warning(paste(bad.obs,
    "observations with NA/NaN/Inf in x, y, or wx removed."))

fitted.out <- y
fitted.out[!ok] <- NA
resid.out <- wt.out <- y * NA
y <- y[ok]
x <- x[ok, , drop = F]
w <- w[ok]
if(!missing(wx))
  wx <- wx[ok]
n <- dim(x)[1]
if(n < p)
  stop("not enough usable observations")
coef <- init
if(p != length(coef))
stop("Must have same number of initial values as
      coefficients")

resid <- y - x %*% coef
converged <- FALSE
status <- "converged"
conv <- NULL
method.in.control <- method.exit <- FALSE
if(iter > 0) {
  for(iiter in 1:iter) {

```

```

if(!is.null(test.vec))
  previous <- get(test.vec)
scale <- median(abs(resid))/0.6745
if(scale == 0) {
  convi <- 0
  method.exit <- TRUE
  status <-
"could not compute scale of residuals"
}
else {
  w <- method(resid/scale,bob)
  if(!missing(wx))
    w <- w * wx
  temp <- lsfit.simp(x, y, w, n, p)
  coef <- temp$coef
  resid <- temp$residuals
  if(!is.null(test.vec))
    convi <- irls.delta(previous, get(test.vec))
  else convi <- irls.rrxwr(x, w, resid)
}
conv <- c(conv, convi)
converged <- convi <= acc
done <- method.exit ||
(converged && !method.in.control )
  if(done)
    break
}
if(!done)
  warning(status <- paste("failed to converge in",
iter,"steps"))
}
if(!missing(wx)) {
  tmp <- (wx != 0)
  w[tmp] <- w[tmp]/wx[tmp]
}
resid.out[ok] <- resid
fitted.out[ok] <- fitted.out[ok] - resid
wt.out[ok] <- w
names(coef) <- cnames

```

```

        list(coefficients = coef, residuals = resid.out,
              fitted.values = fitted.out, w = wt.out, int = int,
              conv = conv, status = status
              )
    }

rems1 <-function(xs, ys, xu, Pm=rep(1,length(xs)/2),
                Dm=rep(1,legtgh(xs)), kxtype = 1,
                kx0 = 2, nstep = 20, kr0 = 2, graph = T, wtype = 0,
                vtype = 3)
{
# Fn.name:  f.mers4
# Purpose: MER estimator for IPPS sampling,
# Author:  Beat Hulliger  , Date: 30/Jul/90
# -----
# Arguments:if kx0==0 kx and kr are linked linearly, else
# kx==kx0
# wtype==4 : x-weights according |xup-0|, d.h. xscale=med(xup)
# wtype!=4 : x-weights according |xsp-0|, d.h. xscale=med(xsp)
# kxtype=1 : running kx, =2 : kx fixed at kx0, =3 : kx=infty
# 28.6.91 : kxtype=3 fixiert auch kx=infty bei rrbeta
# -----
# if(Browse) on.exit(browser()) #
# calculations
    np <- length(xu)
    mxu <- mean(xu)
    # xs <- xu[s]
    ns <- length(xs)      #
# Transformation for unequal variance
    ysp <- ys/sqrt(xs)
    xup <- xu/sqrt(xu)
    xsp <- xs/sqrt(xs)    #xsp <- xup[s] #
# Weights for xsp
    if(wtype == 4)
        xscale <- median(abs(xup))
    else xscale <- median(abs(xsp))
    if(kxtype == 1)
        swxp <- c(median(xsp), max(xsp))/xscale *
                c(0.95, 1.05)

```

```

else swxp <- 0
if(kxtype == 2) wxsp <- pmin(1, kx0/abs(xsp) * xscale)
else wxsp <- rep(
  1, ns) #
# sampling weights
  m <- ns/2 + 2
  p <- ns*xs/sum(xu)
  Ps <- prob(n=ns, m=m, p=p, Pm=Pm,Dm=Dm)
  ips <- diag(Ps)
  ws <- 1/ips
  ht <- sum(ws * ys)/np #
# Regression
  lsresult <- lsfit(xsp, ysp, int = F, w = {
    if(kxtype == 2)
      wxsp * ws
    else ws
  }
)
  lsbeta <- lsresult$coef
  rresult <- rreg1(xsp, ysp, wx = wxsp * ws, int = F,
  bob=kr0, method = function(u,bob)
  wt.huber(u, c = bob* 0.6745), init = lsbeta,
  acc = 0.0001, test.vec =
    "coef", iter = 100)
  rrbeta <- rresult$coef #
# 1. variance-estimation : a la vht (vr1)
  etas <- wxsp * rresult$w * rresult$resid
  etaps <- wxsp * ifelse(rrresult$w == 1, 1, 0)
  Qs <- 1/Ps
  vsrr <- .Uminus(((etas * xsp) %*% Qs %*% (etas * xsp))
    )/((etaps * xsp^2) %*% Qs %*% (etaps * xsp^2))
# 2. variance-estimation : a la W-est. (vr2)
  wtot <- ws * wxsp * rresult$w
  as <- sum(wtot * xsp^2) # fuer den Nenner von vsm2
  wds <- rresult$resid * xsp * wtot
  vsm2rr <- (sum(((ips %o% ips - Ps)/Ps) * outer(wds,
    wds, "-")^2))/(2 *
    as^2) #
# 3. variance-estimation : a la vygs (vr3)

```

```

etixpi <- etas * xsp * ws
vsr3 <- (sum(((ips %>% ips - Ps)/(2 * Ps)) *
  outer(etixpi, etixpi, "-")^
    2))/((etaps * xsp^2) %>% Qs %>%
  (etaps * xsp^2)) #
swr <- c(1, max(1, abs(lsresult$resid)/
  median(abs(lsresult$resid)))) *c(0.95, 1.05)
if(kxtype == 1)
  kx <- c( seq(swxp[1], swxp[2], length = nstep)
    , kx0)
else kx <- c( rep(kx0, length = nstep),kx0)
if(kxtype == 3)
  wxsp <- rep(1, ns)
kr <- c( seq(swr[1], swr[2], length = nstep), kx0)
tk <- kx
rk <- kx
for(i in 1: (nstep+1)) {
#   print(paste("step: ", i))
   if(i == 1) start <- rrbeta
   else start <- tk[i - 1]
   if(kxtype == 1)
     wxsp <- pmin(1, kx[i]/abs(xsp) * xscale)
   rri <- rreg1(xsp, ysp, int = F,
     wx = wxsp * ws, bob=kr[i],
     method = function(u,bob)
     wt.huber(u, c = bob * 0.6745),
     test.vec = "coef", init = start, iter = 500)
   tk[i] <- rri$coef #
   if(vtype == 1) {
# 1. variance-estimation
     etas <- wxsp * rri$w * rri$resid
     etaps <- wxsp * ifelse(rri$w == 1, 1, 0)
     vr <- .Uminus(((etas * xsp) %>% Qs %>%
       (etas * xsp)))/((
       etaps * xsp^2) %>% Qs %>%
       (etaps * xsp^2))
   }
   else {
     if(vtype == 2) {

```

```

# 2. Variance estimator
      rscale <- median(abs(rri$resid))
      wtot <- ws * wxsp * rri$w
      as <- sum(wtot * xsp^2)
      # fuer den Nenner von vsm2
      wds <- rri$resid * xsp * wtot
      vr <- (sum(((ips %>% ips - Ps)/Ps) *
                outer(wds,
                      wds, "-"^2)))/(2 * as^2)
      }
      else {
# 3. variance-estimator
      etas <- wxsp * rri$w *
              rri$resid
      etaps <- wxsp *
              ifelse(rri$w == 1, 1, 0)
      etixpi <- etas * xsp * ws
      vr <- (sum(((ips %>% ips - Ps)
                  /(2 * Ps)) *
                outer(etixpi, etixpi, "-"^2
                      ))/((etaps * xsp^2) %>%
                          Qs %>% (etaps * xsp^2))
      }
      }
      rk[i] <- max(vr, 0) + (tk[i] - ht/mxu)^2
    }
    rkm <- min(rk)
    mers <- median(tk[rk == rkm]) * mxu
      # if(is.vector(mers))
#     mers <- mers[min(abs(mers - median(ys))) ==
#                   abs(mers - median(ys))][1]
    if(graph) {
#       par(mfrow = c(2, 1))
       plot(kr, rk)
       abline(h = rkm) #
#       plot(c(0, xsp), c(0, ysp), type = "n")
#       text(xsp, ysp, label = (1:5)[s])
#       abline(c(0, ht/mxu), lty = 1)
#       abline(c(0, mers/mxu), lty = 2) #

```

```

#           abline(rresult, lty = 3)
#       }
#       browser()
#       on.exit()           #
# result
V.mers4 <- list(wtype = wtype, vtype = vtype,
               kxtype = kxtype, xup =
               xup, xscale = xscale, mers = mers, kr = kr,
               tk = tk, rk = rk,
               rkm = rkm, ht = ht, rht = mxu * rrbeta,
               kx0 = kx0, kr0 = kr0,
               ls = lsbeta * mxu, vsm2rr = vsm2rr,
               vsrr = vsrr, vsr3 = vsr3)
      c(mers, rrbeta * mxu, vsrr * mxu^2, vsm2rr * mxu^2,
        vsr3 * mxu^2,
        lsbeta * mxu, V.mers4$ht)
}

simul2 <-function(XX=XX284, Y=Y284, PopXu=Xu284, Yu=Yu284,P=P284,
  n=16,m=10, B=2, l=2.5,Ke=1, Kx=2, Psi=wt.FV, kM= 5, alpha=.05,
  graph="F")
{
  # pop: la population à échantillonner           #
  # B  : nombre de répliques                       #
  # l  : constante pour l'estimateur hierarchique de Bayes#
  # Kx : constante de mise au point pour HTR et REM      #
  # Psi: fonction de poids pour le calcul de HTR et REM  #
  # kM : constante pour le calcul de l'estimateur du PDA #
  # alpha : pourcentage de valeurs aberrantes dans PDA  #

  T.chapBH<-T.chapBE<-T.chapHTR <-T.chapREM <-HT <-tempsREM<-c()
    tempsJF<-c()
  tempsBE<-mui.BE<-mui.BH<-mui.HTR<- mui.REM <-N.resAB<- c()
  N.Lev <-c()
  T.chapBH<-c(rep(0,B))
  for ( k in 1: B )

```



```

{ if (k==.75*B || k==.5*B || k==.25*B || k==.20*B ||
      k==.10*B || k==.05*B ) cat(k)
  # Les quantites pour les calculs #

  s <- pipts(k,XX, Y,PopXu,Yu,P,n=16,m=10)
  x <- s$Xs[,1]

  n0 <-min(x)
  y <- s$Xs[,2]
  Xu <- s$Xu
  ybarU <- s$Ymoyen
  p <- length(x)*x/sum(Xu)
  Dm <-s$Dm
  Pm <- s$Pm
  idres <- lsfit(x,y, wt=1/p, intercept=F)$res
  H <- x^2/sum(x^2)
  ti<- abs( idres/(sqrt( var(idres)))) )
  N.resAB<-c(N.resAB,rep(1,length(ti[ti >3])))
  N.Lev<-c(N.Lev,rep(1,length(H[H>2/n])))

  tempsREM[k]<-unix.time( temp <- rems1(x, y, Xu, Pm=Pm, Dm=Dm,
                                       graph=F)) [1]
  tempsJF[k] <-unix.time(T.chapBH[k] <- B.H(x,y, Xu, l=1,
                                             w=1/p)$T.BH) [1]
  tempsBE[k] <-unix.time(T.chapBE[k]<-dis(x,y,Xu,w=1/p,
                                          k=kM, alpha=alpha)$T.ME) [1]
  mui.BH[k] <- mean(T.chapBH[1:k])
  mui.BE[k] <- mean(T.chapBE[1:k])

  T.chapHTR[k] <- temp[2]
  T.chapREM[k] <- temp[1]
  HT[k] <- temp[7]

  mui.HTR[k] <-mean(T.chapHTR[1:k])
  mui.REM[k] <-mean(T.chapREM[1:k])

}

##### les quantites de Monte Carlo ! #####

```

```

THT <- mean(HT)
VHT <- (length(HT)-1)*var(HT)/length(HT)
EQM <- var(HT)+ ( mean(HT)- ybarU)^2

montec <- matrix(0, 5, 3)
montec[1,1] <- mean(T.chapBH)-ybarU
montec[1,2] <- (length(T.chapBH)-1)*var(T.chapBH)/
               length(T.chapBH)
montec[1,3] <- EQM /( montec[1,1]^2 + montec[1,2] )

montec[2,1] <- mean(T.chapBE)-ybarU
montec[2,2] <- (length(T.chapBE)-1)*var(T.chapBE)/
               length(T.chapBE)
montec[2,3] <- EQM /( montec[2,1]^2 + montec[2,2] )

montec[3,1] <- mean(T.chapHTR)-ybarU
montec[3,2] <- (length(T.chapHTR)-1)*var(T.chapHTR)/
               length(T.chapHTR)
montec[3,3] <- EQM/ ( montec[3,1]^2 +montec[3,2] )

montec[4,1] <- mean(T.chapREM)-ybarU
montec[4,2] <- (length(T.chapREM)-1)*var(T.chapREM)/length(T.
               chapREM)
montec[4,3] <- EQM/( montec[4,1]^2 + montec[4,2] )

montec[5,1] <- mean(THT)-ybarU
montec[5,2] <- VHT
montec[5,3] <- EQM / EQM

montec[,1] <- montec[,1]/THT
montec[,2] <- sqrt(montec[,2])/THT

dimnames(montec) <- list(c("BH","DA", "HTR","REM","HT"),
  c("biaisrel","ecart_typerel","efficacite"))

if (graph=="T") {
  postscript(file="G1.ps")
  par(mfrow=c(2,1))
  plot(mui.BH)

```

```
        plot(mui.BE)
        dev.off() }                    # fermer postscript #
TMM<-matrix(0,nrow=1,5)
dimnames(TMM) <- list(NULL, c("tempsBH","tempsDA", "tempsREM",
    "Influent","Levier"))
TMM[1,1] <-mean(tempsJF); TMM[1,2] <-mean(tempsBE);
TMM[1,3]<-mean(tempsREM);
TMM[1,4]<-sum(N.resAB)/B; TMM[1,5]<-sum(N.Lev)/B

list(THT=round(THT,3), REs=round(montec,3), TMM=round(TMM,3) )
    }
```

BIBLIOGRAPHIE

- Angers, J.-F. et Berger, J. (1991). Robust Hierarchical Bayes Estimation of Exchangeable Means. *Canadian Journal of Statistics*, vol. **19**, 36-56.
- Angers, J.-F. et MacGibbon, B. (1996). Protection against outliers in Bayesian linear models with econometric applications. *Advances in Econometrics*, vol. **11**, 211-229.
- Bayes, Th. (1763). An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc.* vol. **53**, 370-418.
- Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis* (2e édition). Springer-Verlag, NY.
- Berger, J. et Robert, C. (1990) Subjective hierarchical Bayes estimation of a multivariate normal mean: on the frequentist interface, *The Annals of Statistics*. vol. **18**, 617-651
- Billingsley, P. (1986). *Probability and Measure* (2e édition). New York: Wiley.
- Bloomfield, P. et Steiger, W. (1983). *Least Absolute Deviations: Theory, Applications, and Algorithms*. Boston, Birkhäuser Verlag.
- Box, G.E.P. et Tiao, G.C. (1968). A Bayesian approach to some outlier problems. *Biometrika*, vol. **55**, 119-129.
- Chambers, R.L. (1986). Outlier robust finite population estimation. *Journal of the American Statistical Association*, vol. **81**, 1063-1069.
- Chatterjee, S. et Hadi, A.S. (1986). Influential observations, high leverage points, and outliers in linear regression. *Statistical Science*, vol. **1**, 379-416.
- Cochran, W. G. (1977). *Sampling Techniques*. New York: Wiley.
- Dal Ho, K. (1998). Robust Bayes and empirical Bayes analysis in finite population sampling with auxiliary information. *Journal of the Korean Statistical Society*, vol. **27**, 331-348.
- Deville, J.-C. et Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, vol. **87**, 376-382.
- Dey, A et Srivastava, A.K (1987). Méthode d'échantillonnage avec probabilités de sélection proportionnelles à la taille. *Techniques d'enquête*, vol. **13**, 93-100.
- Duchesne, P. (1999). Estimateurs de calage robustes. *Techniques d'enquête*, vol. **25**, 47-60.
- Efron, B. et Morris, C. (1973). Stein's estimation rule and its competitors - an empirical Bayes approach. *Journal of the American Statistical Association*, vol. **68**, 117-130.
- Efron, B. et Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*, vol. **70**, 311-319.
- Fuller, W.A. (1991). Simple estimators of the mean of skewed populations. *Statistica Sinica*, vol. **1**, 137-158.
- Ghosh, M. et Sinha B.K. (1990). On the consistency between model and design-based

- estimators in survey sampling. *Communications in Statistics-Theory and Methods*, vol. **19**, 668-702.
- Godambe, V.P (1955). A unified theory of sampling from finite populations. *Journal of the Royal Statistical Society*, (Ser. B), vol. **17**, 269-278.
- Hampel, F.R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, vol. **69**, 383-393.
- Hampel, F.R.; Ronchetti, E.M.; Rousseeuw, P.J.; Stahal, W.A. (1986). *Robust Statistics; The Approach Based on Influence Functions*. New York: Wiley.
- Hansen, M.H., Madow, W.G., et Tepping, B.J (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association*, vol. **78**, 776-807.
- Horvitz, D.G. et Thompson, D.G (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, vol. **47**, 663-685.
- Hulliger, B. (1991). *Nonparametric M-estimation of a population mean*. Thèse de doctorat ETH No. 9943, ETH Zürich.
- Hulliger, B. (1995). Estimateurs Horvitz-Thompson à l'épreuve des valeurs aberrantes. *Techniques d'enquête*, vol. **21**, 89-97.
- Kish, L. (1965). *Survey Sampling*. New York: Wiley.
- Laplace, P.S. (1773). Mémoire sur la probabilité des causes par les événements. Mémoires de l'Académie Royale des Sciences présentés par divers savants, vol. **6**, 621-656.
- Lee, H. (1995). Outliers in business surveys. Dans *Business Survey Methods*, (Éds. B.G. Cox, D.A. Binder, B.N. Chinnappa, A. Christianson, M.J. Colledge et P.S. Kott). New York: Wiley.
- Lindley, D, et Smith, A.F (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society*. (Ser. B), vol. **34**, 1-41.
- Lehman, E. (1983). *Theory of Point Estimation*. New York: Wiley.
- Lemire, M.E. (1997). *Modèle de régression bayésien alternatif permettant la détection des valeurs aberrantes*. Mémoire de maîtrise, Dép. de mathématiques et de statistique. Université de Montréal.
- Morris, C. (1983). Parametric empirical Bayes inference: theory and applications. *Journal of the American Statistical Association*, vol. **78**, 47-65
- Pettit, L.I. et Smith, A.F.M. (1985), Outliers and influential observations in linear models. *Bayesian Statistics 2*, 473-494.
- Ramsay, J.O. et Novick, M.R.(1980). PLU robust Bayesian decision theory: point estimation. *Journal of the American Statistical Association*, vol. **75**, 901-907.
- Rao, J.N.K.; Kovar, J.G. et Mantel. H.J. (1990). On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika*, vol. **77**, 365-375.
- Robbins, H. (1955). An empirical Bayes approach to statistics. *In Proceeding of the Third Berkeley Symposium on Mathematical Statistics and Probability* vol. 1, University of California Press, Berkeley, 157-164.
- Robert, C. P. (1992). *L'analyse statistique bayésienne*. Paris: Economica.
- Rousseeuw, P. J. et Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. New York: Wiley.

- Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, vol. **57**, 337-387.
- Särndal, C.-E.; Swensson, B. et Wretman, J. H. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- Searls, D.T. (1966). An estimator which reduces large true observations. *Journal of the American Statistical Association*, vol. **61**, 1200-1204.
- Smith, A.F., (1973). A general Bayesian linear model. *Journal of the Royal Statistical Society*. (Ser. B), vol. **35**, 67-75.
- Statistical Sciences, INC. (1991). *S-PLUS Software*, Seattle : Statistical Science, Inc.
- Stukel, D.M., Hidiraglou, M.A. et Särndal, C.-E. (1996). Estimation de la variance des estimateurs de calage : comparaison des méthodes de jackknife et de la linéarisation de Taylor. *Techniques d'enquêtes*, vol. **22**, 117-126.
- Wilson, H. C. (1978). Last squares versus minimum absolute deviations estimation in linear models. *Decision Sciences*, vol. **X**, 322-335.
- Yates, F. et Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society*, (Ser. B), vol. **15**, 235-261.
- Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error term. *Journal of the American Statistical Association*, vol. **71**, 400-405.