République Algérienne Démocratique et Populaire.
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique.
Université de Mouloud Mammeri, Tizi-Ouzou

**Faculté des Sciences**
**Département des Mathématiques**

Mémoire de Master

Spécialité : MATHÉMATIQUES

Option : Probabilités et statistiques

Intitulé du mémoire

---

# Eléments de robustesse Bayésienne

---

*Présenté par :*

**Kamilia HAMROUN**

*Dirigé par :*

**Hocine FELLAG** **Professeur**

## *Remerciements*

*Tout d'abord, je souhaite remercier mon promoteur M. Fellag Hocine pour avoir proposé et dirigé ce travail, pour sa disponibilité, ses remarques et son soutien, ainsi que pour sa motivation de faire en sorte que ce Master continue d'être et soit amélioré.*

*Mes vifs remerciements iront aussi aux membres du jury qui ont accepté d'évaluer notre modeste travail et qui nous ferons l'honneur de le juger et de l'enrichir par leurs propositions.*

*Je tiens aussi à remercier l'ensemble des enseignants du département des mathématiques, qui ont contribué à notre formation chacun dans son domaine.*

*Je remercie mes parents pour leur soutien au quotidien, je remercie ma famille, mes amis(es) ainsi que touts mes camarades de classe.*

*Enfin, un remerciement spécial à "Naim" qui m'a aidé du mieux qu'il pouvait et qui a toujours été là pour moi durant toute la période de ce travail.*

# Table des matières

# Introduction générale

La statistique est définie comme étant l'étude de la collecte des données, leur traitement (appelé la statistique descriptive) et enfin l'interprétation des résultats et leur représentation afin de rendre les données compréhensibles (l'inférence statistique). Elle est à la fois une science, une méthode et un ensemble de techniques. Pour des uns, la statistique est un domaine des mathématiques, tandis que pour d'autres, elle représente une discipline distincte des mathématiques, mais de nos jours, elle fait partie de ce qu'on appelle la science des données (data science).

La statistique est utilisée dans tous les secteurs de l'activité humaine où se mêlent savoir et données ; elle fait partie des connaissances de base de l'ingénieur, du physicien, du médecin, du biologiste, du psychologue, de l'économiste, de l'assureur, du gestionnaires de l'entreprise, du géographe, etc.

Dans la littérature, deux approches statistiques sont utilisées : la modélisation statistique non paramétrique et la modélisation statistique paramétrique. Dans la modélisation non paramétrique, l'inférence tient compte de la complexité autant que possible et estime la distribution de l'ensemble du phénomène (estimation d'une densité par exemple). La modélisation paramétrique, considère que le modèle de départ est paramétré par un vecteur $\theta$ de dimension finie et la résolution de ce modèle se résume à l'identification de ce paramètre.

La démarche statistique est fondée sur une démarche d'inversion : remonter des "effets" (observations) aux "causes" (paramètres). Pour réaliser ce lien "effets-causes", plusieurs approches ont été proposer ; on site l'approche statistique classique et l'approche statistique Bayésienne :

> Le mode du raisonnement de l'approche statistique classique est toujours le même, quel que soit le paramètre inconnu $\theta$ à estimer : pour un statisticien classique "tout est dans les données", ces données lui permettent de décrire l'incertitude sur le paramètre $\theta$ et de calculer un intervalle de confiance à un risque $\alpha$ fixé en utilisant plusieurs techniques pour l'estimation de ce paramètre (méthode des moments, maximum de vraisemblance, etc.). L'inconvénient de cette approche est que les différentes techniques d'estimation peuvent mener à des résultats différents, des intervalle de confiances différents.
>
> La démarche statistique Bayésienne se distingue de celle de la statistique classique par le fait qu'elle représente l'incertitude portant sur le paramètre $\theta$ du modèle par une distribution de probabilité nommée "loi a priori" de $\theta$, elle s'interprète

comme la représentation formelle sous forme probabiliste de la connaissance sur le paramètre détenue par l'expert. Cette loi a priori doit être claire et indépendantes des donnée (Berry, 1996), sinon la même source d'information interviendrai deux fois.

Le fondement de la statistique Bayésienne s'est basé sur le théorème d'inversion des probabilité connu sous le nom du théorème de Bayes (Bayes, 1763). Les résultats de Bayes ont été redécouvert et étendu par Laplace (1774), lequel n'était pas apparemment au fait du travail de Bayes.
Bien que les premiers travaux d'inspiration Bayésienne datent du $XVII^{ieme}$ siècle, cette méthode connait un regain de popularité depuis quelques décennies. Ce renouveau est sensible dans des domaines très variés, en partie grâce à la disponibilité de calculateurs puissants, mais aussi à une évolution de la pensée statistique et des problèmes abordés.
Les statistiques Bayésiennes sont très utilisées en sciences sociales et politiques, car les données y sont rares et coûteuses à collecter (Gelman et al., 2004). Elles servent aussi en physique des particules (Cousins, 1995 ; Demortier, 2006), en thermodynamique (Chatterjee et al., 1998), en mécanique statistique (Jaynes, 1957), en chimie (Vines et al., 1993 ; Pohorille and Darve, 2006), en génétique (Smyth, 2004 ; Chan et al., 2006), et en bioinformatique (Wilkinson, 2007). La méthode Bayésienne est également employée en sciences cognitives, pour modéliser les comportements animaux et humains comme des prises de décisions rationnelles (Kording, 2004). En intelligence artificielle, la proposition de Bessière et al. (1998a,b) d'une théorie probabiliste des systèmes cognitifs sensi-moteurs a conduit à une méthode de programmation Bayésienne des robots (Lebeltel et al., 2003). Tous ces travaux ainsi que d'autre reposent sur la contribution fondamentale de Jaynes résumée dans son livre posthume Probability Theory : The Logic of Science (Jaynes, 2003).

Pour un modèle paramétré bien défini de densité $f(x|\theta)$ où $\theta$ est le paramètre indiçant la densité, à valeurs dans l'espace $\Theta$, la loi a priori de $\theta$ notée $\pi(\theta)$ représente pour un statisticien Bayésien, l'ensemble des informations a priori disponibles sur le paramètre ainsi que les imprécisions qui s'y rattachent, et dans un contexte pratique, elle regroupe aussi l'ensemble des opinions d'experts.
Le choix de cette loi a priori constitue le point le plus critiqué de l'analyse Bayésienne par les non Bayésiens. Sur le plan pratique de l'approche Bayésienne, il n'existe jamais une unique loi a priori pour $\theta$, mais plutôt un ensemble de lois compatibles avec les informations a priori disponibles et les opinions des modélisateurs. Et c'est bien pour cette cause qu'on a chercher à réduire l'influence de ce choix en retenant des lois a priori à faible contenu informatif, ou des lois dites non informatives. Une fois la loi a priori est construite, le théorème de Bayes rassemble l'information apportée par cette loi a priori avec celle apportée par les données dans une nouvelle distribution dite la distribution a posteriori notée $\pi(\theta|x)$ et qui est la base de toute inférence Bayésienne.

Comme nous venons de le dire, il n'existe jamais une unique loi a priori pour $\theta$ dans la pratique, mais plutôt un ensemble de lois a priori qui sont compatibles avec les informations a priori disponibles. La robustesse Bayésienne consiste à mettre toutes les lois compatibles avec l'information a priori dans une classe et évaluer les changements effectués sur les quantités a posteriori quand la loi a priori varie dans cette classe, afin de mener

une inférence robuste et construire des estimateurs plus robustes.
L'approche statistique paramétrique Bayésienne et la robustesse de ses méthodes feront l'objet de notre mémoire qui sera organisé comme suit :

- Un premier chapitre qui donnera un aperçu général sur l'approche statistique Bayésienne, abordera quelques méthodes de la construction d'une loi a priori et la démarche Bayésienne dans la mise en place d'une inférence sur un paramètre $\theta$ ;

- Un second chapitre où nous ouvrirons une fenêtre sur un sujet qui jusqu'ici fait un objet de recherche chez tous les statisticiens (en particulier les Bayésiens) : faire une analyse (Bayésienne) robuste pour pouvoir construire les estimateurs les plus robustes.

- Enfin, un dernier chapitre qui mettra en œuvre les résultats du deuxième chapitre, ce dernier présentera l'article intitulé *Robust Bayesian estimation in a normal model with asymmetric loss function* où une analyse Bayésienne robuste est menée dans un modèle normale avec la fonction de perte LINEX , et qui a pour objectif la construction d'estimateurs robustes (estimateur stable et estimateur conditionnel $\Gamma$-minimax).

# Chapitre 1

# Approche statistique Bayésienne

## 1.1 Le modèle statistique

Dans notre présent travail, nous ne considérons que l'approche statistique paramétrique. L'analyse statistique exige en général de disposer d'un ensemble d'observations. Pour les modéliser mathématiquement, nous considérons le modèle paramétrique

$$X \in (\Omega, \mathcal{A}, \{P_\theta, \theta \in \ \Theta\})$$

défini comme suivant :

- l'ensemble $\Omega$ désigne tous les résultats possibles, considérés individuellement comme des réalisations d'une grandeur aléatoire observable, décrite par une variable aléatoire $X$.
- l'ensemble $\mathcal{A}$ est formé des événements possibles, représentés par des sous-ensembles de $\Omega$, constituant une $\sigma$-algèbre de $\Omega$.
- pour chaque $\theta$ de $\Theta$, $P_\theta$ est une distribution de probabilité sur l'espace probabilisable $(\Omega, \mathcal{A})$ , attribuant une probabilité à chaque événement.
- $\Theta$ l'ensemble des valeurs du paramètre $\theta$, il est souvent multidimensionnel.
  le paramètre $\theta$ est appelé aussi l'état de la nature dans la théorie statistique de la décision.

## 1.2 Les lois de probabilités qui interviennent dans la statistique Bayésienne

Le but de l'analyse statistique est de faire de l'inférence sur $\theta$, c'est-à-dire décrire un phénomène passé ou à venir dans un cadre probabiliste.
L'idée centrale de l'analyse Bayésienne est de considérer le paramètre inconnu $\theta$ comme aléatoire : l'espace des paramètres $\Theta$ est muni d'une probabilité $\pi$ tel que $(\Theta, \mathcal{A}, \pi)$ est un espace probabilisé. Nous noterons $\theta \sim \pi$. $\pi$ est appelée "*loi a priori*", elle illustre toute information disponible sur le paramètre $\theta$ en dehors des observations $x_i$.
Le modèle statistique Bayésien consiste donc en la donnée d'une *loi a priori* et de la *loi des observations*. On appelle *loi des observations* ou *fonction de vraisemblance*, la loi conditionnelle de $X$ sachant $\theta$ où $X$ est l'ensemble des observations $x_i$ de $X$

$\big(X = (x_1, x_2, ..., x_i, ..., x_n)\big)$, les observations $x_i$ sont donc considérées comme des réalisations de la variable aléatoire $X$, sa densité est notée $f(x|\theta)$, que la variable aléatoire soit discrète ou continue. Si $X$ est discrète, $f(x|\theta)$ représente $P(X = x|\theta)$.
Dans notre travail, on fera l'hypothèse que, sachant $\theta$ les variables aléatoires $X_i$ sont indépendantes. Autrement dit on aura :

$$f(x|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

On a aussi *la loi du couple* $(\theta, X)$ appelée aussi loi jointe, sa densité est notée $f(\theta, x)$. On a donc :

$$f(\theta, x) = f(x, \theta)\pi(\theta),$$

*la loi marginale* de $X$ (appelée aussi la prédiction), sa densité est notée $f(x)$ ou $m(x)$ et est donnée par :

$$f(x) = \int_{\Theta} f(x, \theta)\pi(\theta)d\theta$$

et enfin, on a *la loi a posteriori* qui est la loi conditionnelle de $\theta$ sachant $x$. Sa densité est notée $\pi(\theta)$. En vertu de la formule de Bayes (qu'on verra plus loin), on a :

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x, \theta)\pi(\theta)d(\theta)}$$

Une fois le modèle statistique identifié, l'objectif principal de l'analyse statistique est de nous conduire à une inférence sur le paramètre $\theta$.

## 1.3 Introduction

La méthode Bayésienne est un ensemble de techniques statistiques utilisées pour modéliser des problèmes, extraire de l'information de données brutes et prendre des décisions de façon cohérente et rationnelle. Son cadre d'application est général, mais ses avantages sont déterminants lorsque l'information disponible est incertaine ou incomplète. Cette dernière s'appuie principalement sur le théorème de Bayes.
A l'opposition de l'approche statistique classique qui repose sur la loi des observations $f(x|\theta)$, la statistique Bayésienne repose sur la loi a posteriori $\pi(\theta|x)$. Cependant, ces deux démarches peuvent coïncider au niveau des résultats, en particulier lorsque l'on dispose de très peu d'information a priori et que $\pi(\theta)$ tend vers une loi uniforme. Ceci est relativement naturel, puisque dans l'optique classique, on fait précisément l'hypothèse, implicite, que l'on ne sait rien a priori et que toute l'information est apportée par l'échantillon.
Le but de notre premier chapitre n'est pas de faire une étude approfondie sur l'analyse statistique Bayésienne. Nous allons donner les principaux concepts de cette analyse particulière, nous nous intéresserons au point crucial de l'approche Bayésienne : le choix de la loi a priori, puis nous abordons un aperçu sur la démarche Bayésienne dans la mise en place d'une inférence sur le paramètre $\theta$.

## 1.4 Paradigme Bayésien

Étant donné un modèle paramétrique d'observation $X$ suivant une loi $f(x|\theta)$, $X \sim f(x|\theta)$, où $\theta \in \Theta$, un espace de dimension fini.
L'analyse statistique Bayésienne vise à exploiter le plus efficacement possible l'information apportée par $X$ sur le paramètre $\theta$, pour en suite construire des procédures d'inférence sur $\theta$. Bien que $X$ ne soit qu'une réalisation aléatoire d'une loi gouvernée pas $\theta$, elle apporte une actualisation aux informations préalablement recueillies par l'expérimentateur. Dans le paradigme Bayésien, contrairement a l'approche classique, le paramètre inconnu $\theta$ n'est plus considéré comme étant totalement inconnu et déterministe, il est devenue une variable aléatoire dont le comportement est supposé connu. On considère ainsi que l'incertitude sur le paramètre $\theta$ d'un modèle peut être décrite par une distribution de probabilité $\pi$ sur $\Theta$, qui est la distribution a priori, ce qui revient a supposer que $\theta$ est distribué suivant $\pi(\theta)$, $\theta \sim \pi(\theta)$, avant que $X$ ne soit généré suivant $f(x|\theta)$.
Notons que le rôle central de la distribution a priori ne réside pas dans le fait que $\theta$ puisse ou ne puisse pas être aperçu comme étant distribué selon $\pi$, ou même comme étant une variable aléatoire, mais plutôt dans la démonstration que l'utilisation d'une distribution a priori et de l'appareillage probabiliste qui l'accompagne reste la manière la plus efficace (au sens de nombreux critères) de résumer l'information disponible ou le manque d'informations sur ce paramètre. Un point plus technique est que le seul moyen de construire une approche mathématiquement justifiée opérant conditionnellement aux observations, tout en restant dans un schéma probabiliste, est d'introduire une distribution correspondante pour les paramètres (Bernardo et Smith(1994), Robert(2007, chapitre 10)).
Le concept fondamental du paradigme Bayésien est la distribution a posteriori $\pi(\theta|x)$ qui est un résumé complet de l'information sur le paramètre $\theta$. En effet, une fois $X$ observé, l'approche Bayésienne réalise en quelque sorte l'actualisation de l'information a priori par l'observation $X$, au travers de $\pi(\theta|x)$.
Maintenant nous allons énoncer le théorème de Bayes puis nous allons spécifier quelques méthodes pour le choix d'une loi a priori.

## 1.5 Théorème de Bayes

Soit $A$ et $B$ deux événements aléatoires tels que, $P(B) \neq 0$.
$P(A|B)$ et $P(B|A)$ se définissent respectivement comme étant la probabilité de $A$, conditionnellement à la réalisation de $B$ et la probabilité de $B$, conditionnellement à la réalisation de $A$ tels que :

$$P(A|B) = \frac{P(A,B)}{P(B)} \qquad et \qquad P(B|A) = \frac{P(B,A)}{P(A)}$$

où $P(A,B) = P(B,A)$ est la probabilité que les deux événements $A$ et $B$ aient lieu simultanément.
En remplaçant $P(A|B)$ par $P(B|A)$ qui est égale à $P(B|A)P(A)$, on déduit la relation entre les deux probabilités conditionnelles $P(A|B)$ et $P(B|A)$ :

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

Cette relation appelée théorème de Bayes est en fait un principe d'actualisation : elle décrit la mise à jour de la vraisemblance de $P(A)$ vers $P(A|B)$ une fois que $B$ a été observé.
Bayes(1763) donne une version continue de ce résultat : pour deux variables aléatoires $X$ et $Y$ de distribution conditionnelle $f(x|y)$ et marginale $g(y)$, la distribution conditionnelle de $y$ sachant $x$ est :

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

Ce théorème d'inversion est naturel d'un point de vue probabiliste mais Bayes et Laplace sont allés plus loin et ont considéré que l'incertitude sur le paramètre $\theta$ d'un modèle peut être décrite par une distribution de probabilité $\pi$ sur $\Theta$ appelée distribution a priori. L'inférence est alors fondée sur la distribution de $\theta$ conditionnelle à $x$, $\pi(\theta|x)$, appelée distribution a posteriori et est définie par :

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_\Theta f(x|\theta)\pi(\theta)d\theta} \tag{1.1}$$

Cette équation fût appelée la formule de Bayes. Le dénominateur étant indépendant de $\theta$, la Formule de Bayes peut s'écrire de la façon suivante :

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta)$$

Le passage de la distribution a priori à la distribution a posteriori des paramètres du modèle statistique, exprimé par la formule de Bayes peut être interprété comme la mise à jour (actualisation) de la connaissance (information) sur la base des observations.
Le schéma ci-dessous résume la démarche Bayésienne dans le cadre de la statistique paramétrique inférentielle.

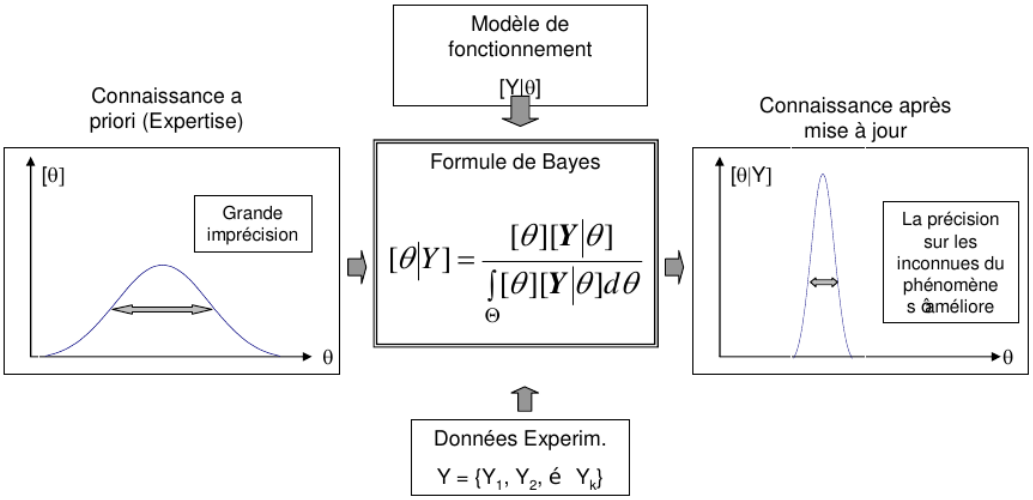


**Figure 1.1** : Mise à jour Bayésienne de la connaissance.

## 1.6 Choix de la loi a priori

La loi a priori est le moteur de l'approche Bayésienne, et sa détermination est l'étape fondamentale de cette approche. En effet, une fois que cette loi est connue l'inférence peut être menée de manière quasi-systématique.
L'appellation "a priori" exprime le fait qu'elle a été établie préalablement à l'observation de données $x$. Dans la pratique, il est rare que l'information a priori soit suffisamment précise pour conduire à une détermination exacte d'une loi a priori. Le statisticien est donc amené à faire un choix arbitraire de loi a priori, ce qui peut modifier considérablement l'inférence qui en découle. Ce choix de la loi a priori peut avoir différentes motivations et les stratégies sont diverses, elles peuvent se baser sur des expériences du passé, sur une intuition, un opinion personnel ou sur une idée que le praticien (statisticien) a du phénomène qu'il est entrain d'observer. Des exemples de mise en œuvre de ces stratégies sont fournis par Kadane et Wolfson (1998), Garthwaite et D'Hogan (2000), Parent et Prevost (2003). En fin, ces stratégies peuvent également tenir compte du fait qu'on ne sait rien par le truchement des loi non informative.
Cette étape, qui est la clé de voûte de l'analyse Bayésienne est aussi celle à laquelle l'approche Bayésienne doit toutes ses critiques. En effet, les détracteurs de celle-ci attirent l'attention sur le fait qu'il n'y a pas une façon unique de choisir une loi a priori, et que ce choix a un impact sur l'inférence résultante.
En pratique, l'information a priori peut être codée selon une des trois façons suivantes :

- Choisir une loi a priori conjuguée à la vraisemblance.
- Prendre une loi a priori impropre (généralisée).
- Choisir une loi a priori non-informative.

### 1.6.1 Lois a priori conjuguées

Une des difficulté majeur de l'approche Bayésienne, le calcul de la loi a posteriori. ce calcul est facilité lorsque la loi a priori et la loi a posteriori sont de même forme. Dans ce cas, on parle de loi a priori conjuguée.

**Définition :**
La loi des observations étant supposée connue, on se donne une famille $F$ de loi de probabilité sur $\Theta$. on suppose que la loi a priori appartient à $F$. Si dans ces conditions, la loi a posteriori appartient encore à $F$, on dit que la loi a priori est conjuguée.

**Remarque :** Une loi conjuguée peut être déterminée en considérant la forme de la vraisemblance $f(x|\theta)$ et en prenant une loi a priori de la même forme que cette dernière.
Les lois a priori conjuguées obtenues par ce procédé sont dite *conjuguées naturelles.*

**Exemple 1 :** Soit $X$ une variable aléatoire d'une loi de Pareto de paramètres $(\theta, a)$, $a > 0$, de densité de probabilités :

$$f(X = x|\theta, a) = \frac{\theta a^{\theta}}{x^{\theta+1}} I_{[a,+\infty[}(x) \qquad a > 0,\ \theta > 0$$

On suppose que $a$ est connu

$$\begin{aligned} f(X = x|\theta, a) &= e^{-(\theta+1)\ln(x)} e^{\theta \ln(a)}\ \theta \\ &= e^{-\theta(\ln(x)-\ln(a))-\ln(x)}\ \theta \\ &= \frac{e^{-\theta \ln(\frac{x}{a})}}{x}\ \theta \end{aligned}$$

L'écriture ressemble a celle d'une loi Gamma.
Donc on peut prendre la loi $\pi(\theta) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} e^{\beta\theta}\theta^{\alpha-1}$ qui est une loi Gamma de paramètre $(\alpha, \beta)$ comme loi a priori.

**Exemple 2 :** (lois a priori conjuguées pour quelques familles exponentielles)

| $f(x\|\theta)$ | $\pi(\theta)$ | $\pi(x\|\theta)$ | $E(\theta\|x)$ |
|---|---|---|---|
| Normale $N(0, \sigma^2)$ | Normale $N(\mu, \tau^2)$ | $N(\frac{\sigma^2\mu + \tau^2 x}{\sigma^2+\tau^2}, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2})$ | $\frac{\sigma^2\mu + \tau^2 x}{\sigma^2+\tau^2}$ |
| Poisson $P(\theta)$ | Gamma $\Gamma(\alpha, \beta)$ | $\Gamma(\alpha + x, \beta + 1)$ | $\frac{\alpha + x}{\beta + 1}$ |
| Gamma $\Gamma(\nu, \theta)$ | Gamma $\Gamma(\alpha, \beta)$ | $\Gamma(\nu + \alpha, \beta + x)$ | $\frac{\nu + \alpha}{\beta + x}$ |
| Binomiale $B(n, \theta)$ | Bêta $Be(\alpha, \beta)$ | $Be(\alpha + x, \beta + n - x)$ | $\frac{\alpha + x}{\alpha + \beta + n}$ |
| Binomiale Négative $Neg(m, \theta)$ | Bêta $Be(\alpha, \beta)$ | $Be(\alpha + m, \beta + x)$ | $\frac{\alpha + m}{\alpha + m + \beta + x}$ |
| Multinomiale $M_k(\theta_1, \cdots, \theta_k)$ | Dirichlet $D(\alpha_1, \cdots, \alpha_k)$ | $D(\alpha_1 + x_1, \cdots, \alpha_k + x_k)$ | $\frac{\alpha_i + x_i}{(\sum_j \alpha_j) + n}$ |
| Normale $N(\mu, 1/\theta)$ | Gamma $\Gamma(\alpha, \beta)$ | $\Gamma(\alpha + 0.5, \beta + (\mu - x)^2/2)$ | $\frac{\alpha + 0.5}{\beta + (\mu - x)^2/2}$ |

**tab.1.1** : lois a priori conjuguées pour quelques familles exponentielles

### 1.6.2 Lois a priori impropres (ou généralisées)

Dans de nombreux cas, la distribution a priori est déterminée par des critères subjectifs et théoriques qui conduisent à une mesure infinie sur l'espace des paramètres $\Theta$ plutôt qu'à une mesure de probabilité, c'est-à-dire à une mesure $\pi$ telle que :

$$\int_\Theta \pi(\theta) = +\infty$$

Dans de tels cas, on dit que la distribution a priori $\pi(\theta)$ est impropre. Cette terminologie est bien-sûr un abus de langage puisque $\pi(\theta)$ n'est pas une densité de probabilité. Ce type de lois n'est utile que si la loi a posteriori $\pi(\theta|x)$ existe, aussi on se limite aux lois impropres telle que :

$$\int_\Theta f(x|\theta)\pi(\theta) < \infty$$

Vérifier que $\int_\Theta f(x|\theta)\pi(\theta) < \infty$ est une difficulté pratique dans l'utilisation de ces lois, cependant cette difficulté ne doit pas être considérée comme un inconvénient car les nouvelles techniques de calcul Bayésien comme les algorithmes MCMC (Méthodes Monte-Carlo par chaîne de Markov) ne nécessite pas dans la pratique de vérifier cette condition.

D'après les fondements des statistiques Bayésiennes, associés à Ramsey, de Finetti et Savage , les lois a priori impropres ne devraient pas être utilisées, de plus, les dysfonctionnements liés à l'utilisation de ces lois, à voire qu'elles peuvent mener à des procédures inadmissibles, enrichissent les critiques contre le Bayésien (Wilkinson, 1971) ou tout simplement contre ce type de lois. Cependant, ces lois admettent un certains nombre d'avantages, ainsi, de nombreux statisticiens tentent de les faire accepter.
Soulignons quelques uns des avantages de ces lois a priori impropres :

Ces approches automatiques sont souvent la seule façon d'obtenir une distribution a priori dans un cadre non-informatif. Cette généralisation du paradigme Bayésien rend ainsi possible une extension supplémentaire de l'applicabilité des techniques Bayésiennes.

Une perspective récente (Berger, 2000) est que les lois a priori impropres devraient être privilégiées par rapport aux lois a priori propres vagues, comme une distribution $N(0, 100^{2)}$, car ces dernières donnent une fausse impression de sécurité due à leur caractère propre tout en manquant de robustesse en terme d'influence sur les résultats d'inférence.

Les lois a priori généralisées (impropres) se situent souvent à la limite des distributions propres.

Les performances des estimateurs obtenus à partir de ces distributions généralisées sont en général suffisamment bonnes pour justifier leur utilisation.

### 1.6.3 Lois a priori non-informatives

Une loi a priori non-informative représente une ignorance sur le problème considéré, mais ne signifie pas que l'on ne sache absolument rien sur la distribution statistique du paramètre. Les lois a priori non-informatives sont des lois qui portent une information sur le paramètre à estimer dont le poids dans l'inférence est réduit, Ce sont des distributions vagues qui, a priori ne favorisent aucune valeur, aucune hypothèse particulière. En conséquence elles laissent les données "parler pour elles-mêmes". Ces lois a priori ont des avantages : elles sont faciles à formuler, elles donnent l'apparence de l'objectivité et nous évitent de travailler avec des lois a priori subjectives mal formulées. Les différentes méthodes proposées pour obtenir ce type de lois a priori ont pour point commun de n'utiliser comme source d'information que la forme de la fonction de vraisemblance $f(x|\theta)$ définie par le modèle.
Les lois non-informatives peuvent être considérées comme des lois de référence auxquelles chacun pourrait avoir recours quand toute information a priori est absente ou minime. Certaines de ces lois sont plus utiles ou plus efficaces que d'autres mais ne peuvent être perçues comme moins informatives que d'autres. Il est désormais largement admis qu'il

n'existe pas d'a priori absolument non-informative (Kass et Wasserman, 1996), comme le fait aussi remarquer Lindley (1990) : *"l'erreur est de les interpréter [les lois a priori non-informative] comme des représentations d'une complète ignorance"*.

Nous donnons dans ce qui suit deux techniques de construction de lois non-informatives.

#### 1.6.3.1 Lois a priori uniformes (ou lois a priori de Laplace )

La loi a priori la plus simple et la plus utilisée est la loi uniforme. En effet, ce choix repose sur l'équiprobabilité des valeurs possible du paramètre $\theta$ sur $\Theta$ : si $\Theta$ est un ensemble de taille $k$, alors $\pi(\theta_i) = 1/k \ \forall i$. Cette méthode a été utilisé pour la première fois par Laplace.
Laplace été le premier à utiliser des techniques non-informatives puisque, bien que ne disposant pas d'information a priori pour les paramètres qu'il étudiait, il munit ces paramètres d'une loi qui prend en compte son ignorance en donnant la même vraisemblance à chaque valeur possible, soit donc en utilisant une loi uniforme. Son raisonnement, appelé plus tard *principe de la raison insuffisante*, se fondait sur l'équiprobabilité des événements élémentaires.
Plusieurs critiques (détaillées dans N.Belkacem, 2012) ont été plus tard avancées sur ce choix.

#### 1.6.3.2 Lois a priori de Jeffreys

Les lois a priori de Jeffreys (1946,1961) sont fondées sur l'information de Fisher donnée par :

$$I(\theta) = E_\theta \left[ \left( \frac{\partial \log f(x|\theta)}{\partial \theta} \right)^2 \right]$$

Dans le cas ou $\theta$ est est un paramètre unidimensionnel, si le domaine de $X$ ne dépend pas de $\theta$, cette information est aussi égal à :

$$I(\theta) = -E_\theta \left[ \left( \frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} \right) \right] \tag{1.2}$$

**Définition :**
Soit $\theta$ un paramètre réel, on appelle loi a priori non-informative de Jeffreys, la loi a priori de la forme :

$$\pi(\theta) = C\sqrt{I(\theta)}$$

avec, $C$ une constante de normalisation et, $I(\theta)$ l'information de Fisher apportée par $X$ sur $\theta$.

Dans le cas où $\theta$ est un paramètre multidimensionnel, on définit la matrice d'information de Fisher $I(\theta)$ par la généralisation de l'équation (1.2). Pour $\theta \in \mathbb{R}$, les éléments de $I(\theta)$ sont donnés par :

$$I_{ij}(\theta) = -E_\theta \left[ \frac{\partial^2 \log f(x|\theta)}{\partial \theta_i \partial \theta_j} \right], \qquad i, j = 1, \ldots, k$$

Et la loi non-informative de Jeffreys est alors définie par :

$$\pi(\theta) = \big[ det\big(I(\theta)\big) \big]^{1/2}$$

Le choix d'une loi a priori dépendant de l'information de Fisher se justifie par le fait que $I(\theta)$ est largement accepté comme un indicateur de la quantité d'information apportée par le modèle ou l'observation sur $\theta$ (Fisher,1956).

**Remarques :**

La loi a priori de Jeffreys offre une méthode automatique pour obtenir une loi a priori non-informative pour n'importe quel modèle paramétrique, et permet de retrouver des estimateurs classiques.

Ce type de construction de lois non-informatives conduit très souvent à des lois a priori impropres.

Dans le cas d'un échantillon de taille $n$, la loi de Jeffreys est donnée par :

$$\pi(\theta) \propto \big(I(\theta)\big)^{1/2}$$

**Exemples :**

**1.** Soit X une variable aléatoire telle que : $X \sim Binomiale(n, \theta)$,

$$P[X = x|\theta] = C_n^x \theta^x (1-\theta)^{n-x}$$

calculons la loi a priori $\pi(\theta)$ de Jeffreys :
on a :

$$I(\theta) = -E\left[\frac{\partial^2 \log P[X = x|\theta]}{\partial \theta^2}\right],$$

et après calcul on aura :

$$I(\theta) = \frac{n}{\theta(1-\theta)}$$

d'où : $\pi(\theta) = \sqrt{n}\big(\theta(1-\theta)\big)^{-1/2} \Rightarrow \pi(\theta) \propto \big(I(\theta)\big)^{1/2}$

Cette distribution a posteriori de Jeffreys s'agit de la loi bêta$(1/2, 1/2)$.

**2.** Soit $X$ une variable aléatoire telle que : $X \sim Exponentielle(\theta)$,

$$f(x|\theta) = \theta \exp\{-\theta x\}, \qquad x \geq 0$$

Calculons l'a priori de Jeffreys :
Comme dans l'exemple précédent on a :

$$I(\theta) = -E\left[\frac{\partial^2 \log f(x|\theta}{\partial \theta^2}\right]$$

Et après calcul on aura :

$$I(\theta) = \frac{1}{\theta^2}$$

D'où :

$$\pi(\theta) \propto \left(\frac{1}{\theta^2}\right)^{1/2}$$

$$\Rightarrow \pi(\theta) \propto \frac{1}{\theta}$$

On aura une loi $\pi(\theta)$ impropre, on peut normaliser cette distribution a condition de donner une information sur $\theta$ (contradictoire), c'est-à-dire, on suppose que $\theta \in [a, b]$, $a > 0$ et on détermine une constante $C$ telle que :

$$C \int_a^b \pi(\theta) d\theta = 1$$

$$\Rightarrow C(\ln(b) - \ln(a)) = 1$$

$$\Rightarrow C = \frac{1}{\ln(\frac{a}{b})}$$

## 1.7 Inférence statistique Bayésienne

Nous présentons dans le reste de ce chapitre, quelques idées générales sur comment mener une inférence Bayésienne sur le paramètre $\theta$ : faire une inférence (décision, estimation et tests) sur $\theta$ en utilisant la loi a posteriori que nous supposons connue.

### 1.7.1 Approche décisionnelle, Estimation ponctuelle

En pratique, l'inférence statistique conduit à une décision finale prise par le "décideur", et il est important de pouvoir comparer les différentes décisions au moyen d'un critère d'évaluation, qui va apparaître sous forme d'un coût.
Un problème de décision en général est fondé sur les trois éléments suivant :

- L'ensemble des décisions possibles $D$.
- Un espace des paramètres $\Theta$.
- Une fonction de coût (de perte) $L(\theta, \delta)$ qui décrit la perte à prendre la décision $\delta$ lorsque le paramètre est $\theta \in \Theta$.

Si nous notons $D$ l'ensemble des décisions possibles, on appellera coût (perte) une fonction mesurable de $(\Theta \times D)$ dans $\mathbb{R}^+$, elle est notée $L(\theta, \delta)$ :

$$\begin{aligned} L : \Theta \times D &\longrightarrow \mathbb{R}^+ \\ (\theta, \delta) &\longmapsto L(\theta, a) = L(\theta, \delta) \end{aligned}$$

Où $\delta$ est une règle de décision qui elle aussi est une application de $\Omega$ (l'espace des données) dans $D$ :

$$\begin{aligned} \delta : \Omega &\longrightarrow D \\ x_i &\longmapsto \delta(x_i) = a_i \end{aligned}$$

Une fonction coût $L(\theta, \delta)$ est définie telle que :

**1.** $\forall(\delta, \theta),\ L(\theta, \delta) \geq 0$

**2.** $\forall\theta, \exists\delta^*$ *tel que* $L(\theta, \delta^*) = 0$

Si la première propriété ci dessus n'est pas vérifier, c'est-à-dire, si la fonction coût est négative, $L(\theta, \delta) < 0$, alors elle correspond à un gain non à une perte.
La deuxième propriété est interprétée comme suite : une décision $\delta^*$ est considérée comme une bonne décision si elle conduit à un coût nul, c'est-à-dire, $L(\theta, \delta^*(x)) = 0$, mais le paramètre $\theta$ étant inconnu, la résolution de cette equation est impossible.

Classer les décision par la seule considération du coût est impossible, car il ne prend pas en considération l'information apportée par la fonction de vraisemblance $f(x, \theta)$. Et c'est pour cela qu'on est amené à considérer la moyenne du coût. Et c'est ce qu'on appelle le *risque fréquentiste*

**Définition :** (Risque fréquentiste)
Pour une fonction de perte $L(\theta, \delta)$ donnée, on appelle risque fréquentiste la moyenne de cette fonction : le coût moyen du coût d'une règle de décision. On le note $R(\theta, \delta)$ :

$$\begin{aligned} R(\theta, \delta) &= E_\theta\big[L\big(\theta, \delta(x)\big)\big] \\ &= \int_\Omega L\big(\theta, \delta(x)\big) f(x|\theta) dx \end{aligned}$$

Ce risque nous permet d'établir un critère (admissibilité) de comparaison entre les décisions : on dira que la décision $\delta_1$ est préférable que la décision $\delta_2$ et on note $\delta_1 < \delta_2$ si :

$$\begin{cases} R(\theta, \delta_1) \leq R(\theta, \delta_2), \forall \theta \in \Theta \\ \exists \theta_0 \in \Theta, tel\ que\ :\ R(\theta_0, \delta_1) < R(\theta_0, \delta_2) \end{cases}$$

Cela permet d'établir un pré-ordre sur l'ensemble des décisions $D$, mais malheureusement il n'est que partiel puisqu'il ne permet pas de comparer deux décisions telles que :

$$R(\theta_1, \delta_1) < R(\theta_1, \delta_2),\ pour\ \theta_1 \in \Theta \qquad et \qquad R(\theta_2, \delta_1) > R(\theta_2, \delta_2),\ pour\ \ \theta_2 \in \Theta$$

Dans l'approche classique, la règle de décision $\delta$ doit minimiser $R(\theta, \delta)$ pour tout $\theta \in \Theta$, mais en fait, comme nous venons de le voir, il est impossible de trouver une telle décision, dans l'approche Bayésienne, le paramètre $\theta$ suit une loi a priori $\pi(\theta)$, et suite a une observation $x$, cette loi a priori est réactualisée et transformée en une loi a posteriori $\pi(\theta|x)$
On définit alors la coût a posteriori.

**Définition :**(Coût a posteriori)
On définit le coût a posteriori $\rho(\pi, \delta|x)$ comme étant la moyenne du coût par rapport à la loi a posteriori, il est également nommé risque a posteriori :

$$\begin{aligned} \rho(\pi, \delta|x) &= E^\pi\big[L\big(\theta, \delta(x)\big)\big] \\ &= \int_\Theta L\big(\theta, \delta(x)\big) \pi(\theta|x) d\theta \end{aligned}$$

Pour chaque valeur de $x$, on cherche donc la décision $\delta(x)$ qui minimise ce coût et on construit ici une règle de décision.

On considère maintenant la moyenne du risque fréquentiste suivant la loi a priori, il s'agit du risque de Bayes (ou risque intégré)

**Définition :**(Risque de Bayes)
Pour une fonction de perte donnée, le risque Bayésien est défini par :

$$\begin{aligned} r(\pi,\delta) &= E^{\pi}\big[R\big(\theta,\delta\big)\big] \\ &= \int_{\Theta} R(\theta,\delta)\pi(\theta)d\theta \end{aligned}$$

Remarquons que le risque de Bayes défini par :

$$r(\pi,\delta) = E^{\pi}\big[R\big(\theta,\delta\big)\big] = \int_{\Theta} R(\theta,\delta)\pi(\theta)d\theta$$

satisfait :

$$r(\pi,\delta) = \int_{\Omega} \rho\big(\theta,\delta|x\big) f(x)dx$$

En effet, comme $L(\theta,\delta) \geq 0$, d'après le théorème de fubini, on obtient :

$$\begin{aligned} r(\pi,\delta) &= \int_{\Theta} R(\theta,\delta)\pi(\theta)d\theta \\ &= \int_{\Theta} \left[\int_{\Omega} L\big(\theta,\delta(x)\big) f(x|\theta)dx\right] \pi(\theta)d\theta \\ &= \int_{\Omega} \left[\int_{\Theta} L\big(\theta,\delta(x)\big) f(x|\theta)\pi(\theta)d\theta\right] dx \\ &= \int_{\Omega} \left[\int_{\Theta} L\big(\theta,\delta(x)\big)\pi(\theta|x)d\theta\right] f(x)dx \\ &= \int_{\Omega} \rho\big(\theta,\delta|x\big) f(x)dx = E\big[\rho(\theta,\delta|x)\big] \end{aligned}$$

D'où, le risque de Bayes $r(\pi,\delta)$ est aussi la moyenne du risque a posteriori $\rho(\theta,\delta|x)$ suivant la loi marginale $f(x)$.

Dans le cadre de l'estimation, l'ensemble des décisions est l'espace des paramètres $\Theta$, et la règle de décision $\delta$ est un estimateur.

**définition :** (Estimateur de Bayes)
Toute solution $\delta^{\pi}$ minimisant le risque de Bayes $r(\pi,\delta)$, si elle existe, est appelée estimateur de Bayes associé au coût $L(\theta,\delta)$ , et à la distribution a priori $\pi$.
On a :

$$\delta^{\pi} = r(\pi) = r(\pi,\delta^{\pi}) = \underset{\delta\in D}{inf} r(\pi,\delta)$$

L'estimatieur $\delta^\pi$ sera déterminé analytiquement ou numériquement (en utilisant des méthodes de simulation), selon la complexité du coût $L$ et de la loi a posteriori $\pi(\theta|x)$.
Pour des fonctions de coût classiques, les estimateurs de Bayes correspondant sont des caractéristiques usuelles de la distribution a posteriori, dans ce qui suit, nous donnons l'exemple des trois estimateurs Bayésiens les plus utilisés.

**Exemples :**

1. *La fonction de Coût quadratique* :
   Introduit par Légende (1805) et Gauss (1810), ce coût est sans conteste le critère d'évaluation le plus commun. Considérons le coût quadratique : $L(\theta, \delta) = (\theta - \delta)^2$.
   L'estimateur de Bayes $\delta^\pi$ du paramètre $\theta$, associé à une distribution a priori $\pi$, est donné par la moyenne a posteriori $E^\pi(\theta|x)$.

   **Preuve :**
   Par définition, l'estimateur de Bayes minimise le coût a posteriori :

   $$\rho(\pi, \delta|x) = \int_\Theta L\big(\theta, \delta(x)\big)\pi(\theta|x)d\theta$$

   On remplace $L\big(\theta, \delta(x)\big)$ par sa valeur :

   $$\rho(\pi, \delta|x) = \int_\Theta (\theta - \delta)^2\pi(\theta|x)d\theta$$

   Puis on minimise :

   $$\begin{aligned}
   \frac{\partial\rho(\pi, \delta|x)}{\partial\delta} &= \int_\Theta \frac{\partial}{\partial\delta}(\theta - \delta)^2\pi(\theta|x)d\theta = 0 \\
   &\Rightarrow \int_\Theta -2(\theta - \delta)\pi(\theta|x)d\theta = 0 \\
   &\Rightarrow -2\int_\Theta \theta\pi(\theta|x)d\theta + 2\delta\int_\Theta \pi(\theta|x)d\theta = 0 \\
   &\Rightarrow -2E^\pi(\theta|x) + 2\delta = 0 \\
   &\Rightarrow \delta = E^\pi(\theta|x)
   \end{aligned}$$

   Le minimum du coût a posteriori est effectivement atteint par :

   $$\delta^\pi = E^\pi(\theta|x)$$

   **Conclusion :** Dans ce cas l'estimateur de Bayes est la moyenne à posteriori, il est appelé estimateur MMSE (minimum mean square error).

   **Exemple :** Si $X$ est une variable aléatoire suivant une loi de Pareto de paramètre $\theta$, $X \sim P(\theta)$, et si $\pi(\theta)$ est la loi a priori qui est une loi Gamma de paramètres $(\alpha, \beta)$,

alors la loi a posteriori $\pi(\theta|x)$ est aussi une loi gamma de paramètres $(x+\alpha, \beta+1)$. Sous l'hypothèse d'un coût quadratique, un estimateur de Bayes $\delta^\pi$ de $\delta$ sera l'espérance a posteriori de $\delta$ . Puisque la loi a posteriori est une loi Gamma, l'espérance est le rapport des paramètres et on a :

$$\delta^\pi = \frac{x+\alpha}{\beta+1}$$

.

**2.** *La fonction de coût absolue* :
La fonction de coût absolue est la fonction définie par :

$$L(\theta, \delta(x)) = \begin{cases} k_2\big(\theta - \delta(x)\big), \ si\ \theta > \delta(x) \\ k_1\big(\delta(x) - \theta\big), \ si\ \theta \leqslant \delta(x) \end{cases}$$

L'estimateur de Bayes associé à la la loi a priori $\pi(\theta)$ et à la fonction de coût absolu est le fractile :

$$\frac{k_2}{(k_1+k_2)} \ de\ \pi(\theta|x).$$

En particulier, si $k_1 = k_2$, l'estimateur de Bayes est la médiane de $\pi(\theta|x)$.

**Preuve :**
Par définition, l'estimateur de Bayes minimise le coût a posteriori :

$$\rho(\pi, \delta|x) = \int_\Theta L\big(\theta, \delta(x)\big)\pi(\theta|x)d\theta$$

On remplace $L\big(\theta, \delta(x)\big)$ par sa valeur :

$$\begin{aligned}\rho(\pi, \delta|x) &= \int_{-\infty}^{\delta} k_1(\delta-\theta)\pi(\theta|x)d\theta + \int_{\delta}^{+\infty} k_2(\theta-\delta)\pi(\theta|x)d\theta \\ &= k_1\left[\int_{-\infty}^{\delta}\delta\pi(\theta|x)d\theta - \int_{-\infty}^{\delta}\theta\pi(\theta|x)\right] + k_2\left[\int_{\delta}^{+\infty}\theta\pi(\theta|x)d\theta - \int_{\delta}^{+\infty}\delta\pi(\theta|x)d\theta\right]\end{aligned}$$

Puis on minimise :

$$\begin{aligned}\frac{\partial\rho(\pi,\delta|x)}{\partial\delta} &= k_1\left[\frac{\partial}{\partial\delta}\left(\delta\int_{-\infty}^{\delta}\pi(\theta|x)d\theta\right) - \frac{\partial}{\partial\delta}\int_{-\infty}^{\delta}\theta\pi(\theta|x)\right] \\ &\quad + k_2\left[\frac{\partial}{\partial\delta}\int_{\delta}^{+\infty}\theta\pi(\theta|x)d\theta - \frac{\partial}{\partial\delta}\left(\delta\int_{\delta}^{+\infty}\pi(\theta|x)d\theta\right)\right] = 0 \\ &\Rightarrow k_1\left[\int_{-\infty}^{\delta}\pi(\theta|x)d\theta + \delta\pi(\delta|x) - \delta\pi(\delta|x)\right] \\ &\quad + k_2\left[-\delta\pi(\delta|x) - \int_{\delta}^{\infty}\pi(\theta|x)d\theta + \delta\big(-\pi(\delta|x)\big)\right] = 0\end{aligned}$$

$$\Rightarrow k_1 \int_{-\infty}^{\delta} \pi(\theta|x)d\theta - k_2 \int_{\delta}^{+\infty} \pi(\theta|x)d\theta = 0$$

$$\Rightarrow k_1 \int_{-\infty}^{\delta} \pi(\theta|x)d\theta - k_2 \left[1 - \int_{-\infty}^{\delta} \pi(\theta|x)d\theta\right] = 0$$

$$\Rightarrow \int_{-\infty}^{\delta} \pi(\theta|x)d\theta = \frac{k_2}{(k_1+k_2)}$$

$$\Rightarrow P(\theta \leqslant \delta|x) = \frac{k_2}{(k_1+k_2)}$$

$$\Rightarrow \delta^{\pi} \; est \; le \; fractile \; \frac{k_2}{(k_1+k_2)} \; de \; \pi(\theta|x)$$

**Conclusion :** Dans ce cas l'estimateur de Bayes et la médiane à posteriori.

**3.** *La fonction coût* 0-1 *:*
Ce coût est surtout utilisé dans l'approche classique des tests d'hypothèse, il a été proposé par Neyman et Pearson.
La fonction coût 0-1 est la fonction définie par :

$$L(\theta,\delta) = \begin{cases} 0, \; si \; \delta - \epsilon < \theta < \delta + \epsilon \\ 1, \; si \; \theta \geqslant \delta + \epsilon \; ou \; \theta \leqslant \delta - \epsilon \end{cases}$$

avec : $\epsilon$ très petit ($\epsilon$ est l'erreur commise lors de l'estimation).

Plus généralement, ce coût est un exemple typique d'un coût non quantitatif. En effet, pour ce coût, la pénalité associée à un estimateur $\delta$ est 0 si la réponse est correcte et 1 sinon.
L'estimateur de bayes associé à la loi $\pi(\theta)$ et à la fonction de coût 0-1 est le mode de la distribution a posteriori $\pi(\theta|x)$.

**Preuve :**
On a :

$$\rho(\pi,\delta|x) = \int_{\Theta} L\big(\theta,\delta(x)\big)\pi(\theta|x)d\theta$$

On remplace $L\big(\theta,\delta(x)\big)$ par sa valeur :

$$\rho(\pi,\delta|x) = \int_{-\infty}^{\delta-\epsilon} L\big(\theta,\delta(x)\big)\pi(\theta|x)d\theta$$

$$
\begin{aligned}
&+ \int_{\delta-\epsilon}^{\delta+\epsilon} L\big(\theta, \delta(x)\big)\pi(\theta|x)d\theta \\
&+ \int_{\delta+\epsilon}^{+\infty} L\big(\theta, \delta(x)\big)\pi(\theta|x)d\theta \\
&= \int_{-\infty}^{\delta-\epsilon} \pi(\theta|x)d\theta + \int_{\delta+\epsilon}^{+\infty} \pi(\theta|x)d\theta \\
&\Rightarrow \rho(\pi, \delta|x) = 1 - \int_{\delta-\epsilon}^{\delta+\epsilon} \pi(\theta|x)d\theta
\end{aligned}
$$

Donc pour minimiser le risque a posteriori il faut maximise l'intégrale $\int_{\delta-\epsilon}^{\delta+\epsilon} \pi(\theta|x)d\theta$
.
Pour un $\epsilon$ très petit, $\epsilon \longrightarrow 0$, on obtient la limite suivante :

$$\lim_{\epsilon \to 0} argmax \int_{\delta-\epsilon}^{\delta+\epsilon} \pi(\theta|x)d\theta = argmax\pi(\theta|x)$$

.
**Conclusion :** Dans ce cas l'estimateur de Bayes est appelé le maximum a posteriori (MAP), il représente le mode de la distribution a posteriori.

## 1.7.2 Admissibilité et minimaxité d'un estimateur

### 1.7.2.1 Admissibilité :

**définition :**(Estimateur admissible)
On dit que $\delta \in D$ est inadmissible si et seulement si :

$$\exists \delta_0 \in D, \forall \theta \in \Theta : R(\theta, \delta) \geq R(\theta, \delta_0) \; et \; \exists \theta_0 \in \Theta : R(\theta_0, \delta) > R(\theta_0, \delta_0)$$

de ce fait, $\delta$ est dit admissible s'il n'est pas inadmissible et par conséquence, un estimateur est dit admissible si et seulement si il n'est pas inadmissible.

**Théorème :** (Estimateur Bayésien admissible)

Si l'estimateur Bayésien $\delta^\pi$ associé à une fonction de perte $L$ et à la loi a priori $\pi$ est unique, alors il est admissible.

***Démonstration :***
Supposons que $\delta^\pi$ est un estimateur Bayésien non admissible :

$$\exists \delta_0 \in D, \forall \theta \in \Theta, R(\theta, \delta) \geqslant R(\theta, \delta_0) \; et \; \exists \theta_0 \in \theta, R(\theta_0, \delta) > R(\theta_0, \delta_0)$$

En intégrant la première inégalité :

$$\int_\Theta R(\theta, \delta_0) \leqslant \int_\Theta R(\theta, \delta^\pi \pi(\theta)d\theta) = r(\pi)$$

donc $\delta_0$ est aussi un estimateur Bayésien associé à $L$ et $\pi$ et $\delta_0 \neq \delta^\pi$ d'après la seconde inégalité, donc le théorème se déduit par contraposée. Ce théorème s'applique notamment

dans le cas d'un risque fini et d'une fonction de coût convexe. En outre, l'unicité de l'estimateur Bayésien implique la finitude du risque :
$r(\pi) = \int R(\theta, \delta^\pi)\pi(\theta) < \infty$ (sinon, tout estimateur minimise le risque).

**Définition :** ($\pi$-admissibilité)

Un estimateur $\delta_0$ est $\pi$-admissible si et seulement si :

$$\forall(\delta, \theta), R(\theta, \delta) \geqslant R(\theta, \delta_0) \Rightarrow \pi\{\theta \in \Theta, R(\theta, \delta) < R(\theta, \delta_0\}) = 0$$

Autrement dit, cette définition implique que chaque estimateur non admissible est $\pi$-admissible.

**Proposition :** Tout estimateur Bayésien tel que $r(\pi) < \infty$ est $\pi$-admissible.

***Démonstration :*** (Robert, 2006)
Soit $\delta^\pi$ un estimateur Bayésien à risque fini. Pour $\delta_0$ tel que $\forall\theta R(\theta, \delta) \leqslant R(\theta, \delta_0)$, on note $A = \theta \in \Theta, R(\theta, \delta) < R(\theta, \delta_0)$.
Nous avons alors :

$$\begin{aligned}\int_\Theta R(\theta, \delta_0)\pi(\theta)d\theta - \int_\Theta R(\theta, \delta^\pi)\pi(\theta)d\theta \\ = \int_A \big(R(\theta, \delta) - R(\theta, \delta_0)\big)\pi(\theta)d\theta \leqslant 0\end{aligned}$$

avec égalité si et seulement si $\pi(A) = 0$. Or, comme $\delta^\pi$ est Bayésien et le risque fini, $r(\theta, \delta_0) \geqslant r(\theta, \delta^\pi)$ donc l'intégrale est nulle (positive et négative !), d'où $\pi(A) = 0 : \delta^\pi$ est $\pi$-admissible. .Nous pouvons maintenant énoncer une condition suffisante d'admissibilité des estimateurs Bayésiens.

**Téorème :** Si $\pi > 0$ sur $\Theta$, $r(\pi) < \infty$ pour une fonction de perte $L$ donnée, si $\delta^\pi$ l'estimateur Bayésien correspondant existe et si $\theta \longmapsto R(\theta, \delta)$ est continue. Alors $\delta^\pi$ est admissible.

***démonstration*** **:** (Robert, 2006)
Supposons que $\delta^\pi$ est non admissible. D'après la définition précédente, $\delta^\pi$ est $\pi$ admissible. Ainsi, il existe $\delta_0$ tel que $R(\theta, \delta_0) \geqslant R(\theta, \delta^\pi)$ et $\theta_0 \in \Theta : R(\theta_0, \delta_0) < R(\theta_0, \delta^\pi)$. De tant que la fonction $\theta \longmapsto R(\theta, \delta)$ est continue, la fonction définie par $\theta \longmapsto R(\theta, \delta_0) - R(\theta, \delta^\pi)$ est aussi continue. Donc il existe un voisinage ouvert de $\theta_0$ , $v_0 \subset \Theta$ tel que $\forall\theta \in v_0, R(\theta, \delta_0) < R(\theta, \delta^\pi)$. En considérant $A = \{\theta \in \Theta : R(\theta, \delta) < R(\theta, \delta_0)\}$ , il en résulte que $\pi(A) > \pi(v_0)$. (car $\pi$ est supposée strictement positive et $v_0 \subset A$). Donc en prenant un modèle dominé par une mesure qui charge positivement les ouverts (la mesure de Lebesgue par exemple), $\pi(v_0) > 0$. $A$ est donc non négligeable (de mesure non nulle), ce qui n'est pas conforme avec la $\pi$-admissibilité. En conclusion, $\delta(\pi)$ est admissible.

#### 1.7.2.2 Minimaxité :

**Définition :** (Risque minimax et estimateur minimax)
On appelle risque minimax associé à la fonction du coût $L$ la valeur :

$$\overline{R} = \inf_{\delta \in D} \sup_{\theta \in \Theta} R(\theta, \delta)$$

On dit que $\delta_0$ est un estimateur minimax si et seulement si :

$$\overline{R} = \sup_{\theta \in \Theta} R(\theta, \delta_0) = \inf_{\delta \in D} \sup_{\theta \in \Theta} R(\theta, \delta)$$

L'estimateur minimax correspond au point de vue (conservateur !) de faire le mieux dans le pire des cas, c'est-à-dire à s'assurer contre le pire. Il est utile dans des cadres complexes mais trop conservateur dans certains cas où le pire est très peu probable. Il peut être judicieux de voir l'estimation comme un jeu entre le statisticien (choix de $\delta$) et la Nature (choix de $\theta$), l'estimation minimax rejoint alors celle de la Théorie des Jeux.

**Théorème :**
Le risque de Bayes est toujours plus petit (ou égal) que le risque minimax.

***Démonstration :***
On a :

$$\delta^\pi = r(\pi, \delta^\pi) = \underset{\delta \in D}{inf} \int_\Theta R(\theta, \delta)\pi(\theta)d\theta \leqslant \underset{\delta \in D}{inf} \underset{\theta}{sup} R(\theta, \delta)$$

d'où la démonstration du théorème.
Énonçons maintenant quelques propriétés l'estimateur minimax.

**Propriété 1 :** Si l'estimateur minimax $\delta_0$ est unique alors, il est admissible (la démonstration se fait par absurde).
**Propriété 2 :** Si l'estimateur $\delta_0$ est admissible et de risque constant alors $\delta_0$ est l'unique estimateur minimax.

### 1.7.3 Approche Bayésienne des tests d'hypothèses

#### 1.7.3.1 Introduction à la théorie des tests

La théorie des tests consiste à tester si une hypothèse est vraie. On dit qu'on teste $H_0$ (l'hypothèse nulle) contre $H_1$ (l'hypothèse alternative). Dans cette situation, il est possible de commettre deux erreurs : conclure que $H_0$ est vraie alors qu'en réalité c'est $H_1$ qui est vérifiée, et vice versa. On définit ces deux erreurs :

**Définition :** (Erreur de première et seconde espèce).
On appelle erreur de première espèce ou erreur de type 1 la quantité :

$$\alpha = P\{\text{accepter} H_0 | H_1 \text{ est vraie}\} .$$

On appelle erreur de seconde espèce ou erreur de type 2 la quantité :

$$\beta = P\{\text{accepter } H_1 | H_0 \text{ est vraie}\} .$$

Pour construire un test statistique au risque $\alpha$, on fixe l'erreur de première espèce à $\alpha$, avec $\alpha$ "petit" (de l'ordre de 5%, 1% voire moins). Une fois cette erreur fixée, on n'a plus aucun contrôle sur l'erreur de seconde espèce $\beta$.
On réalise un test selon les étapes suivantes :

1. Définition des hypothèse $H_0$ et $H_1$ . Cela implique de faire un choix : quelle est l'hypothèse privilégiée, suivant l'erreur qu'on veut contrôler ;
2. Choix du niveau $\alpha$ (petit) ;
3. Calculer la statistique du test. Ce calcul se fait à l'aide des observations statistiques à notre disposition, et du test choisis ;
4. Conclusion au vu de l'échantillon selon la règle de décision associée au test. La conclusion étant le rejet ou l'acceptation de $H_0$ .

**Remarque :** Si, à l'issue du test, on accepte $H_1$, alors l'erreur commise est $\alpha$ (par définition). Puisque $\alpha$ est fixé "petit", l'erreur commise est "petite". Si on rejette $H_0$ , l'erreur commise est $\beta$ (toujours par définition). Cette erreur n'est pas contrôlée et peut être très grande : l'acceptation de $H_0$ ne permet donc pas, a priori, de conclure que $H_0$ est effectivement vraie.
Un test a le plus souvent pour but de rejeter $H_0$ plutôt que de choisir entre deux hypothèses, puisque accepter $H_0$ ne veut pas dire que cette hypothèse est vraie mais qu'il n'y a pas de raison suffisante pour la rejeter. Rejeter $H_0$ est donc beaucoup riche en information que l'accepter. C'est pour cela que dans le cas de rejet on dit que le test est significatif.

Dans l'approche classique des tests statistiques, il s'agit d'identifier des événements improbables, c'est-à-dire ; des observations incompatibles avec une certaine loi de probabilité. Dans cette approche, l'hypothèse alternative y joue un rôle mineur voire négligeable. L'approche Bayésienne vise à une résolution complète en modélisant les deux hypothèses simultanément, permettant d'une part la résolution du problème décisionnel du choix de l'hypothèse la plus probable, et d'autre part, la construction d'une loi a posteriori sur les paramètres du modèle retenu, le choix d'une hypothèse $H_0$ se fait contre ou relativement à une hypothèse $H_1$ ($H_0$ versus $H_1$).
Cependant, un test d'hypothèses soit au sens Bayésien ou au sens classique peut être considéré comme un procédé statistique, c'est-à-dire, comme une fonction définie sur l'espace des observations à valeurs dans un espace à deux points que l'on appela arbitrairement "accepter" et "rejeter" une hypothèse, ou alors, il peut être considéré comme une façon pour le statisticien de gérer ses doutes relatifs à son modèle statistique.
la procédure de décision Bayésienne est fondé sur la probabilité a posteriori de l'hypothèse nulle ou de manière presque équivalente sur le facteur de Bayes.

#### 1.7.3.2 Facteur de Bayes

Le facteur de Bayes est un critère Bayésien de sélection de modèle, comme il est un outil pour comparer la crédibilité de deux hypothèses.

- **Cas de tests d'hypothèses**
  Supposons que nous avons deux hypothèses :

$$H_0 : \theta \in \Theta_0$$
$$H_1 : \theta \in \Theta_1$$

  En définissant les probabilités a priori :

$$\pi_0 = P(H_0) = P(\theta \in \Theta_0)$$
$$\pi_1 = P(H_1) = P(\theta \in \Theta_1)$$

  avec $\pi_0 + \pi_1 = 1$, on considère le rapport a priori (prior odds) $\frac{\pi_0}{\pi_1}$.
  de même façon, les probabilités a posteriori :

$$p_0 = P(H_0|x) = P(\theta \in \Theta_0|x)$$
$$p_1 = P(H_1|x) = P(\theta \in \Theta_1|x)$$

  avec $p_0 + p_1 = 1$, on définit le rapport a posteriori (posterior odds ) $\frac{p_0}{p_1}$.
  Nous devons choisir une des deux hypothèses dans un concept Bayésien. Pour le faire, il suffit de comparer les probabilités a posteriori des deux hypothèses. la règle de Bayes consiste à choisir l'hypothèse de plus grande probabilité a posteriori.
  Le facteur de Bayes en faveur de $H_0$ relativement à $H_1$ est le rapport des deux odds :

$$BF = \frac{posterior\ odds}{prior\ odds} = \frac{p_0\pi_1}{p_1p_0}$$

  **Interprétation du facteur de Bayes**
  Kass et Raftery (1995) proposent un guide d'interprétation des valeurs du facteur de Bayes qui peut servir d'aide à la décision pour écarter ou non une hypothèse jugée moins crédible.

| BF | $H_0$ |
|---|---|
| de 1 à 3.2 | rejetée |
| de 3.2 à 10 | fortement rejetée |
| de 10 à 100 | très fortement rejetée |
| $> 100$ | rejetée de façon décisive |

**tab.1.1.2.3** : Guide d'interprétation du facteur de Bayes.
Kass et Raftery (1995).

- **Cas de sélection de modèles**
  Lorsqu'on est amené à faire une sélection entre deux modèles $M_1 = \{f_1(x|\theta_1), \pi(\theta_1)\}$ et $M2 = \{f_2(x|\theta_2), \pi(\theta_2)\}$ par l'approche Bayésienne, on affecte à chacun des deux modèles les probabilités a priori $P(M_1)$ et $P(M_2)$
  avec :

$$P(M_1) + P(M_2) = 1$$

On considère le rapport a priori ou quotient d'enjeux a priori (the odds prior)

$$\frac{P(M_1)}{P(M_2)} = \frac{P(M_1)}{1 - P(M_1)}$$

Du théorème de Bayes, on obtient les probabilités a posteriori :

$$P(M_1|X) = \frac{P(X|M_1)P(M_1)}{f(X)} \quad et \quad P(M_2|X) = \frac{P(X|M_2)P(M_2)}{f(X)}$$

Où $X = (X_1, ..., X_n)$ est le $n$-échantillon, et $f(X) = P(X|M_1)P(M_1)+P(X|M_2)P(M_2))$, représente la probabilité marginale des données ($f(X) \neq 0$)
avec :

$$P(M_1|X) + P(M_2|X) = 1$$

conduit au rapport a posteriori ou quotient d'enjeux a posteriori (the odds posterior)

$$\frac{P(M_1|X)}{P(M_2|X)} = \frac{P(M_1|X)}{1 - P(M_1|X)}$$

On peut écrire :

$$\frac{P(M_1|X)}{P(M_2|X)} = \frac{P(X|M_1)P(M_1)}{P(X|M_2)P(M_2)} \tag{1.3}$$

Où

$$BF = \frac{P(X|M_1)}{P(X|M_2)} = \frac{\int_{\Theta_1} f_1(x|\theta_1)\pi(\theta_1)d\theta_1}{\int_{\Theta_2} f_2(x|\theta_2)\pi(\theta_2)d\theta_2}$$

est le **facteur de Bayes** en faveur du modèle $M_1$ contre le modèle $M_2$.
On peut alors exprimer la relation (1.1) comme suite :

posterior odds=BF$\times$ prior odds

**Remarque :** Si on accorde le même poids a priori pour les deux hypothèses $M_1$ et $M_2$, c'est-à-dire, $M_1 = M_2 = \frac{1}{2}$, alors le facteur de Bayes sera le odds posterior :

$$BF = \frac{P(M_1|X)}{P(M_2|X)}$$

et c'est ce qui est utilisé souvent en pratique.

**Exemple :** (Quotient d'enjeux a posteriori)
le quotient d'enjeux a posteriori est un rapport largement utilisé dans les études médicales. C'est un cas particulier du facteur de Bayes. Il correspond au rapport des probabilités a posteriori des deux modèles (odds posterior) qui s'écrit :

$$K = \frac{P(M_1|X)}{P(M_2|X)}$$

***Interprétation*** :

- Si $K > 1$ alors on choisit le modèle $M_1$ ;
- Si $K < 1$ alors le choix sera porté sur le modèle $M_2$ ;
- $K = 1$ situation d'indécision.

### 1.7.4 Estimation par intervalles

#### 1.7.4.1 Intervalle de crédibilité

L'intervalle de crédibilité est l'équivalent Bayésien de l'intervalle de confiance de la statistique classique, en effet, une approche fréquentiste conduit à un intervalle de confiance tandis qu'une approche Bayésienne aboutit à un intervalle de crédibilité. Toutefois, l'interprétation de ces deux derniers est légèrement différente : les intervalles de crédibilités Bayésiens traitent leurs limites comme fixes et le paramètre estimé comme variable aléatoire, les intervalles de confiances fréquentistes traitent leurs limites comme variables aléatoires et le paramètre estimé comme fixe. En outre, les intervalles de crédibilités bayésiens utilisent la connaissance a priori, alors que les intervalles de confiance fréquentistes ne le font pas.
Par définition, un intervalle de crédibilité au niveau $\alpha$ est un intervalle tel que la probabilité que le paramètres d'intérêt $\theta$ appartient a cet intervalle selon la distribution à posteriori est de $1 - \alpha$.

**Définition 1.1.3.** Soit $\alpha \in ]0, 1[$ fixé. Un intervalle $I$ pour lequel on a :

$$P(\theta \in I|x) = \int_I \pi(\theta|x)d\theta = 1 - \alpha$$

est appelé intervalle de confiance a posteriori, de niveau de confiance $1 - \alpha$.
Il y a plusieurs manières de construire cet intervalle ; on peut prendre par exemple celui donné par $\alpha/2$ et $(1 - \alpha)/2$ quantiles.
Si on s'intéresse à plusieurs paramètres simultanément, on parlera de **région de crédibilité**.

#### 1.7.4.2 Région de crédibilité

**Définition.** ( Région $\alpha$-crédible)

Une région $C$ de $\theta$ est dite $\alpha$-crédible si et seulement si

$$P^{\pi}(\theta \in C|X) > 1 - \alpha$$

. Notons que le paradigme Bayésien permet une nouvelle fois de s'affranchir d'un inconvénient de l'approche fréquentiste. En effet, au sens fréquentiste, une région de confiance $C$ est définie par

$$\forall\theta, P_\theta(\theta \in C) \geqslant 1 - \alpha$$

et correspond à l'interprétation suivante ; en refaisant l'expérience un grand nombre de fois, la probabilité que $\theta$ soit dans $C$ est plus grand que $1 - \alpha$.
Une région de confiance n'a donc de sens que pour un très grand nombre d'expériences

tandis que la définition Bayésienne exprime que la probabilité que $\theta$ soit dans $C$ au vue des celles déjà réalisées est plus grande que $1-\alpha$. Il n'y a donc pas besoin ici d'avoir recours à un nombre infini d'expériences pour définir une région $\alpha$-crédible, seule compte l'expérience effectivement réalisée.

Il y a une infinité de régions $\alpha$-crédibles, il est donc logique de s'intéresser à la région qui a le volume minimal. Le volume étant défini par $vol(C) = \int_C dv(\theta)$, si $\pi(\theta|x)$ est absolument continue par rapport à une mesure de référence $v$.

**Définition.** (Région $HPD$ (highest posteriori density))

$C_\alpha^\pi$ est une région $HPD$ si et seulement si :

$$C_\alpha^\pi = \{(\theta, \pi(\theta|x) \geqslant h_\alpha\}$$

où $h_\alpha$ est défini par :

$$h_\alpha = \sup\{h, P^\pi(\{\theta, \pi(\theta|x) \geqslant h\}|x) \geqslant 1-\theta\}$$

$C_\alpha^\pi$ est parmi les régions qui ont une probabilité supérieure à $1-\alpha$ de contenir $\theta$ (et qui sont donc $\alpha$-crédibles) et sur lesquelles la densité a posteriori ne descend pas sous un certain niveau (restant au dessus de la valeur la plus élevée possible).

> **Théorème.** (Régions $HPD$)
> $C_\alpha^\pi$ est parmi les régions $\alpha$-crédibles de volume minimal si et seulement si elle est $HPD$.
> (La démonstration de ce théorème est détaillé dans : Robert, 2006).

### 1.7.5 Lien entre intervalles de confiances et tests d'hypothèses

La présentation de l'intervalle de confiance (crédibilité) a permis de voir que ses procédures et celles utilisées pour la construction d'un test d'hypothèses sont très voisines, il peut même y avoir une sorte d'équivalence entre celles-ci. Pour qu'une valeur estimée de $\theta$ soit acceptée il faut et il suffit qu'elle soit dans l'intervalle de crédibilité. Il y a donc une équivalence pour $\theta$ entre le fait de prendre une valeur acceptée dans le test de niveau $\alpha$ et le fait d'être située dans l'intervalle de confiance de niveau $1-\alpha$. On peut donc voir l'intervalle de confiance comme l'ensemble des valeurs acceptées par le test.

Il est généralement plus facile d'élaborer une procédure de test, et à partir de celle-ci, un intervalle de confiance peut être construit. Pour approfondir la théorie des intervalles de confiance (et, plus généralement, la théorie de l'estimation des tests), on pourra consulter le livre de Cox et Hinkley (1979) ou celui de Shao (1999).

## 1.8 Conclusion

Concluons notre premier chapitre en précisant en quoi l'approche Bayésienne présente divers avantages, mais aussi quelques inconvénients, par rapport à l'approche fréquentiste.

En effet comparé a l'approche statistique classique, l'approche Bayésienne présente de nombreux avantages :

Elle fait un couplage cohérent entre la vision du modélisateur et l'évidence des résultats expérimentaux ; elle donne la possibilité de combiner les informations objectives apportées par les donnée, avec des informations extérieures a l'échantillon observé (Bernardo et Smith, 1994), (Box et Tiano, 1973), (Berry et al, 1996), Cette possibilité est très adaptée à la pratique technique. Dans de nombreux cas, la formalisation statistique n'est qu'une approximation de la réalité avec des données peu représentatives ou incohérentes, mais le statisticien dispose aussi de l'avis technique des experts qui, sur la base de leur expériences et savoir-faire, sont capable de fournir des informations complémentaire de grande utilité et qu'il serai dommage de ne pas prendre en compte (Bernier et al,2000).

Les résultats de l'inférence Bayésienne sont plus riches que les estimateurs fournis par l'inférence classique (Beger, 1985).

Les techniques Bayésienne permettent d'obtenir la loi jointe des paramètres du modèle donc de prendre en compte simultanément l'effet de l'incertitude globale sur l'ensemble des paramètres inconnus sur les prévisions futures du comportement (Krzysztofowicz, 1983).

Un autre avantage de l'approche Bayésienne que l'on détaillera pas dans notre travaille, est, la possibilité d'échantillonner à nouveau les distributions a posteriori des paramètres d'un modèle afin de produire des intervalles qu'on appelles "intervalles de prédictibilité". De tels intervalles tiennent compte de toutes les incertitudes d'un modèle Par exemple, un intervalle de prédictibilité est utilisé pour comparer différents modèles de croissance.

Il existe aussi d'autres avantages lié à l'utilisation e l'approche statistique Bayésien comme la création de variables dérivée, la possibilité de solutionner des modèles complexes dont certains ne peuvent l'être par la statistique classique et finalement la production de résultats non biaisés même avec des échantillon de faibles nombres (Kéry, 2010), (Kéry et Schaub, 2012).

Malheureusement, il y a quelques inconvénients. Comme nous l'avons déjà mentionné, la distribution a posteriori contient toute l'information à propos des paramètres du modèle. De plus, c'est à partir de cette distribution que se base notre inférence. Si cette distribution a posteriori est connue, nous n'avons pas besoin d'échantillonner l'a posteriori pour calculer nos paramètres. Par contre, si l'a posteriori n'est pas une distribution connue, nous devons pouvoir en générer un échantillon en espérant que les valeurs obtenues de nos paramètres ne soient pas trop éloignées de ce qu'elles seraient si cet a posteriori était connu. Afin de générer un tel échantillon, il faut dès lors recourir à des méthodes numériques. Mais ça ne reste toujours pas un problème de taille vue les nouvelles techniques de simulation telles que les méthodes MCMC.

# Chapitre 2

# Robustesse Bayésienne

## 2.1 Introduction

En statistique, il arrive fréquemment qu'un échantillon de données contienne des incertitudes, des valeurs aberrantes (des observations non représentatives). Conséquemment, l'inférence statistique peut être contaminée par leur présence, menant ainsi à des résultats qui ne sont pas en accord avec la majorité des observations.
Dans la mise en œuvre d'une analyse Bayésienne, le statisticien s'est intéressé comme une première étape à proposer un modèle qui explique le comportement des observations, une loi a priori qui génère le paramètre d'intérêt et une fonction de perte qui est utilisée pour évaluer le risque. Etant donné ces trois éléments, le Bayésien cherche à employer des méthodes optimales et à proposé des modèles robustes.
Les analyses de la robustesse Bayésienne (Berger et al, 2000), également appelées analyses de sensibilités Bayésienne à l'incertitude sur les détailles précis de l'analyse, permettent de quantifier l'erreur introduite par l'imprécision des modèles sur les résultats obtenus lors de l'estimation, et éventuellement d'agir en conséquence pour restreindre la classe des résultats.
L'analyse Bayésienne robuste est l'étude de la sensibilité des réponses Bayésienne à des données incertaines. Ces données incertaines sont généralement le modèle, la loi a priori ou la fonction de perte quand il s'agit d'un problème de décision, mais, parfois il peut s'agir d'une combinaison de ces trois derniers.

## 2.2 Les différentes approches de la robustesse Bayésienne

Il existe dans la littérature Bayésienne plusieurs méthodes alternatives dites robustes. Cette robustesse peut être perçu sous trois angles différents, trois approches principales :

La première est *l'approche informelle* : l'approche informelle consiste à considérer un ensemble de loi a priori et comparer les moyennes a posteriori correspondantes, cette approche est intéressante en raison de sa simplicité. Mais, il est parfois facile de prendre des lois a priori incompatibles avec les connaissances a priori disponibles, ce qui mènerait à des moyennes a posteriori très différentes.

La deuxième est appelée *robustesse globale* : l'approche globale fonctionne idéalement à l'approche précédente, elle consiste à considérer une classe de lois a priori compa-

tibles avec les informations a priori disponibles, et évaluer par la suite la différence entre le sup et l'inf des moyennes a posteriori autours de la classe. Cette approche est très populaire elle même, mais les calculs ne sont pas toujours faciles du fait qu'elle exige l'évaluation du sup et de l'inf des moyennes a posteriori, (pour plus de détails sur cette approche, voir Moreno,2000).

La dernière est dite *robustesse locale* : la robustesse locale s'est intéressée au taux de changements dans l'inférence par rapport aux changements dans la loi a priori utilisant différentes techniques. Cette dernière est décrite par Gustafson (2000) et Sivaganesan (2000).
Les mesures de sensibilité (robustesse) locale sont généralement plus faciles à calculer que les mesures globales, mais leur interprétation n'est pas toujours claire.

### 2.2.1 Robustesse par rapport à la loi a priori ou approche informelle (prior robustness)

La robustesse Bayésienne dans ce cas, consiste à construire une classe de lois a priori $\Gamma$, et étudier par la suite les changements effectués sur les quantités a posteriori autours de cette classe. La robustesse est réalisée s'il n'y a pas un grand changement entre les moyennes a posteriori sous les lois a priori, c'est-à-dire que le choix des lois a priori n'a pas d'influence.

**Exemple :** (Ghosh et al, 2006 )
Nous allons énoncer un exemple qui montre combien il est important d'introduire la notion de la sensibilité au choix de la loi a priori.
supposons qu'on observe une variable aléatoire $X$ qui suit la loi de Poisson($\theta$), et supposons qu'il est connu a priori que $\theta$ a une distribution continue avec une médiane égale à 2 et un quantile d'ordre 3 égale à 4. i.e. $P^{\pi}(\theta \leq 2) = 0.5$ et $P^{\pi}(\theta \leq 4) = 0.25$.
Si ces informations sont les seules connaissances disponibles sur le paramètre $\theta$, les trois distributions suivantes peuvent être considérées comme des lois a priori de $\theta$ :

**1.** $\pi_1 : \theta \sim exponentielle\big(\log(2)\big)$ ;

**2.** $\pi_2 : \log(\theta) \sim Normale\big(\log(2), (\log(2)/0.675)^2\big)$ ;

**3.** $\pi_3 : \log(\theta) \sim Cauchy\big(\log(2), \log(2)\big)$.

Pour voir l'influence du choix de la loi a priori, on examine les moyennes a posteriori sous les trois différentes lois a priori :

Pour (1), on a : $\theta|x \sim Gamma(\log(2) + 1, x + 1)$ donc la moyenne a posteriori correspondante est

$$E^{\pi_1}(\theta|x) = \frac{\log(2) + 1}{x + 1}$$

.
Pour (2), on pose $m = \log(\theta)$ et $\sigma = \log(2)/0.675$, on obtient

$$\begin{aligned} E^{\pi_2}(\theta|x) &= E^{\pi_2}(\exp(m)|x) \\ &= \frac{\int_{-\infty}^{+\infty} \exp(-e^m) \exp(m(x+1)) \exp\big(-(m - \log(2))^2/(2\sigma^2)\big) dm}{\int_{-\infty}^{+\infty} \exp(-e^m) \exp(mx) \exp\big(-(m - \log(2))^2/(2\sigma^2)\big) dm} \end{aligned}$$

Pour (3), on pose aussi $m = \log(\theta)$, on obtient

$$E^{\pi_3}(\theta|x) = E^{\pi_3}(\exp(m)|x)$$
$$= \frac{\int_{-\infty}^{+\infty} \exp(-e^m)\exp(m(x+1))\left[1+\big(\big(m-\log(2)\big)/\big(\log(2)\big)\big)^2\right]^{-1} dm}{\int_{-\infty}^{+\infty} \exp(-e^m)\exp(mx)\left[1+\big(\big(m-\log(2)\big)/\big(\log(2)\big)\big)^2\right]^{-1} dm}$$

Pour des valeurs différentes de $x$, on calcule les moyenne a posteriori sous les trois différentes loi a priori, les résultats sont donnée par la table ci-dessous.

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 10 | 15 | 20 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\pi_1$ | 0.749 | 1.485 | 2.228 | 2.974 | 3.713 | 4.456 | 8.169 | 11.882 | 15.595 | 37.874 |
| $\pi_2$ | 0.950 | 1.480 | 2.106 | 2.806 | 3.559 | 4.353 | 8.660 | 13.241 | 17.945 | 47.017 |
| $\pi_3$ | 0.761 | 1.562 | 2.094 | 2.633 | 3.250 | 3.980 | 8.867 | 14.067 | 19.178 | 49.402 |

**Tab.2.1** : Les moyenne a posteriori sous $\pi_1$, $\pi_2$ et $\pi_3$.

Pour un $x$ petit ou modéré ($x \leqslant 10$), il n'y a pas un grand changement entre les moyennes a posteriori sous le trois lois a priori, et donc le choix d'une loi a priori entre les trois n'a pas d'influence, la robustesse est réalisée. Par contre pour les grandes valeurs de $x$, le choix de la loi a priori est très important et a influencé les moyennes a posteriori, il n'y a donc pas de robustesse dans ce cas.
La conclusion de cette analyse serait que la robustesse est vraisemblablement obtenue pour des valeurs plus petites de $x$, mais pas pour des valeurs plus importantes.
Il est clair maintenant qu'il y a des situations où le choix d'une loi a priori parmi d'autres dans une classe peut avoir une influence sur les quantités a posteriori d'intérêt.

### Quelques classes de lois a priori

Pour la mise en œuvre d'une analyse Bayésienne robuste par rapport a une loi a priori $\pi$, on est amené à construire une classe $\Gamma$ de lois a priori de sorte que cette dernière modélise au mieux l'incertitude sur notre loi a priori, selon berger (1990), cette incertitude portant sur la loi a priori peut se présenter par la classe $\Gamma$ de lois a priori, à laquelle $\pi$ est supposée appartenir.
Dans ce qui suit, nous allons donner quelques classes de lois a priori qui sont parmi les classes les plus populaires.

- **Classe des lois conjuguées :**
  Ces classes sont basées sur les lois a priori conjuguées qu'on a déjà vue dans le premier chapitre. Elles sont parmi les classes les plus faciles à utiliser dans la pratique, et elles sont typiquement choisies pour des raisons pratiques parce qu'elle fournissent en général des bornes explicites pour les quantités d'intérêt. Par exemple, si $X \sim N(\mu, \sigma^2)$ tels que : $\mu_1 \leqslant \mu \leqslant \mu_2$ et $\sigma_1^2 \leqslant \sigma^2 \leqslant \sigma_2^2$ , on peut considérer la classe :
  $$\Gamma_c = N(\mu, \sigma^2) \ : \ \mu_1 \leqslant \mu \leqslant \mu_2 \ et \ \sigma_1^2 \leqslant \sigma^2 \leqslant \sigma_2^2$$

pour quelques valeurs spécifiées de $\mu_1$, $\mu_2$, $\sigma_1^2$ et $\sigma_2^2$.
L'avantage de ces classes est que les quantités a posteriori peuvent être facilement calculées (pour les lois conjuguées naturelles), ce qui facilite la minimisation et la maximisation des quantités d'intérêt.
Ces classes sont connues aussi par les classes paramétriques et en général elles sont données par :

$$\Gamma_P = \{P : p(\theta, \omega),\ \omega \in \Omega\}$$

- **Classe d'$\epsilon$-contamination :**
  Étant donnée une loi a priori $\pi_0$ (souvent conjuguée), une classe naturelle de lois a priori pour étudier la sensibilité de la loi a priori est la classe de contamination qui est définit comme suit :

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

  où $\epsilon$ porte l'incertitude sur $\pi_0$ et $\mathcal{Q}$ est la classe des contaminations qui a été considéré par Huber (1973) dans la robustesse classique. Pour la robustesse Bayésienne, la classe d'$\epsilon$-contamination est construite à partir des lois a priori qui ressemblent à la loi a priori de référence $\pi_0$.
  Le problème majeur lié à l'utilisation de telles classes est la détermination difficile de $\epsilon$ et de $\mathcal{Q}$.

- **Classes des moments généralisés :**
  Les classes des moments généralisés ont été considérées pour la première fois par Betro et al. (1994) et Goutis (1994), celles données par :

$$\Gamma_{GM} = \left\{\pi : \int_\Theta H_i \pi(\theta) d\theta \leqslant \alpha_i,\ i = 1, ..., n\right\}$$

  Où les $H_i$ sont des fonctions $\pi$-intégrables et les $\alpha_i$ sont des nombres réels fixés.
  La classe des moments généralisés permet d'autres choix des fonctions $H_i$ .
  Betro et al (1994) ont considéré la classe définit par des bornes sur les probabilités marginales pour des ensembles données $K_i$ , dont ils ont pris :

$$H_i(\theta) = \int_{K_i} f(x|\theta) dx$$

  et en appliquant le théorème de Fubini, ils ont trouvé que :

$$\int_\Theta H_i(\theta)\pi(\theta) d\theta = \int_{K_i} m_\pi(x) dx$$

  où $m_\pi(x) = \int_\Theta f(x|\theta)\pi(\theta) d\theta$ est la densité marginale de $x$ sous $\pi$.
  Un avantage pour ce choix est qu'à partir de l'information sur $\theta$ on peut obtenir des informations sur $X$.
  D'autres classes populaires sont données et détaillées dans Ghosh et al (2006).

### 2.2.2 Robustesse par rapport à la fonction de perte (loss robustness)

En général, un groupe de décideurs pourrait avoir différentes idées, différentes évaluations des conséquences de leur actions, ce qui entraîne différentes fonctions de pertes. Dans ce cas, il peut être nécessaire d'évaluer la robustesse des méthodes Bayésiennes au choix de la fonction de perte exactement de la même manière qu'au choix de la loi a priori : si une classe de fonctions de pertes est disponible, les changement effectués sur les quantités a posteriori peuvent être examinés.

**Exemple :** Soit $X$ une variable aléatoire d'une loi de poisson de paramètre $\theta$,
$X \sim Poisson(\theta)$, et que la loi a priori $\pi(\theta)$ est exponentielle de moyenne 1.ie.$\pi(\theta) = e^{\theta}$.
Supposons aussi que $x = 0$ est observée.Après calcul,la distribution a posteriori de $\theta$ sera une exponentielle de moyenne 1/2. Ainsi, l'estimateur Bayésien de $\theta$ est égal à 1/2 sous la perte quadratique, qui est la moyenne a posteriori. Et est égal à 0.3465 sous la perte absolue, qui est la médiane a posteriori.
Il est clair que ces deux estimateurs sont différents, et cette différence peut avoir un certain impact selon l'utilisation de ces estimateurs.

#### Quelques classes de fonctions de perte

Nous ne présentons que les classes de fonctions de pertes les plus utilisées dans les études de la robustesse Bayésienne.

- **Classes d'$\epsilon$-contamination**
  Comme nous avons déjà vu, les classes d'$\epsilon$-contamination sont très populaires pour définir les classes de lois a priori. On peut aussi définir un voisinage d'une fonction de perte $L_0$ comme suit :

  $$\mathcal{L}^{\epsilon} = \{L : L(c) = (1-\epsilon)L_0(c) + \epsilon M(c) : M \in \mathcal{W}\}$$

  où $\epsilon$ représente l'imprécision sur $L_0$, $c$ est une conséquence de l'ensemble des conséquences $C$ et $\mathcal{W}$ est la classe des fonctions de perte qui contient aussi $L_0$.

- **Classes partiellement connues**
  Sur la base des classes des quantiles des lois a priori, Martin et al (1998) ont considéré une partition finie $C_1, ..., C_n$ de l'ensemble des conséquences $C$ et ont donné les bornes supérieurs et inférieurs des fonctions de pertes pour chaque élément de la partition.

  $$\mathcal{L}_k = \{L : v_{i-1} \leqslant L(c) \leqslant v_i,\ \forall c \in C,\ i = 1, ..., n\}$$

  où quelques ensembles $C_i$ pourraient être vides.

- **Classes paramétriques**
  la classe paramétrique des fonctions de perte la plus populaire est définie dans

Varian (1974) par :

$$\mathcal{L}_\gamma = \{L_\gamma : L_\gamma(c) = e^{\gamma c} - \gamma c - 1, \gamma_L \leqslant \gamma \leqslant \gamma_\mu,\ c = a - \theta\}$$

autre exemple est

$$\mathcal{L}_k = \{L_k : L_k(c) = -e^{-k(c)},\ k > 0,\ c = a - \theta\}$$

D'autres exemples peuvent être trouvés dans Bell(1995). En général, les classes paramétriques sont définies comme suit :

$$\mathcal{L}_w = \{L = L_w, w \in \Omega\}$$

## 2.3 Robustesse conjointe par rapport à la fonction de perte et à la loi a priori (loss and prior robustness)

La recherche sur la sensibilité conjointement par rapport à la loi a priori et à la fonction de perte n'est pas très abondante. En effet, il est possible que le problème soit robuste seulement par rapport à la loi a priori ou seulement par rapport à la fonction de perte, mais plutôt sensible lorsque les deux éléments sont considérés conjointement. Un aperçu de ces analyses générales de sensibilité est donné dans Rios Insua et al (2000). Basu et DasGupta (1995) considèrent conjointement les lois a priori dans une classe de distributions et un nombre fini de fonctions de perte. Martin et Rios Insua (1996) ont utilisé des dérivées de Fréchet pour étudier la sensibilité locale pour les petites perturbations à la fois dans la distribution a priori et dans la fonction de perte.

## 2.4 Autres approches

### 2.4.1 L'analyse Bayésienne hiérarchique

Un des problèmes de l'approche Bayésienne noté par les fréquentistes est l'incertitude concernant la loi a priori. Une approche qui essaie de rectifier ce problème est l'analyse Bayésienne hiérarchique qui met des mesures a priori sur les paramètres de la loi a priori $\pi(\theta)$.

**Définition :** On appelle modèle Bayésien hiérarchique de niveau $n$, un modèle statistique Bayésien avec fonction de vraisemblance, $f(x|\theta)$, la densité a priori $\pi(\theta)$ est décomposée en $n$ densités conditionnelles et une densité marginale, c'est-à-dire :

$$\begin{aligned}
&x|\theta \sim f(x|\theta)\\
&\theta|\theta_1 \sim \pi_1(\theta|\theta_1)\\
&\theta_1|\theta_2 \sim \pi_2(\theta_1|\theta_2)\\
&\vdots\\
&\theta_{n-1}|\theta_n \sim \pi_n(\theta_{n-1}|\theta_n)
\end{aligned}$$

$$\theta_n \sim \pi_{n+1}(\theta_n)$$

Les $\theta_1$, $\theta_2$, ..., $\theta_n$ sont appelés hyperparamètres de niveau $i, i = \overline{1,n}$ et $\pi_{n+1}(\theta_n)$ une loi marginale. Par conséquent

$$\pi(\theta) = \int_{\Theta_1 \times ... \times \Theta_n} \pi_1(\theta|\theta_1)...\pi_n(\theta_{n-1}|\theta_n)\pi_{n+1}(\theta_n)d\theta_1...d\theta_n$$

Habituellement, on se limite au cas où $n = 1$.
Dans ce cas on a, la loi a posteriori de $\theta$ est donnée par :

$$\pi(\theta|x) = \int \pi(\theta|\theta_1, x)\pi(\theta_1|x)d(\theta_1),$$
$$\pi(\theta|\theta_1, x) = \frac{f(x|\theta)\pi_1(\theta|\theta_1)}{f_1(x|\theta_1)},$$
$$f_1(x|\theta_1) = \int f(x|\theta)\pi_1(\theta|\theta_1)d\theta_1,$$
$$\pi(\theta_1|x) = \frac{f_1(x|\theta_1)\pi_2(\theta_1)}{f(x)},$$
$$f(x) = \int f_1(x|\theta_1)\pi_2(\theta_1)d\theta_1.$$

Un aspect positif de l'analyse Bayésienne hiérarchique est qu'elle augmente la robustesse de l'analyse Bayésienne classique d'un point de vue fréquentiste, puisqu'elle réduit l'arbitraire sur le choix de l'hyperparamètre et établit une moyenne de réponse Bayésienne conjuguées.
La décomposition d'une loi a priori $\pi$ en plusieurs composantes $\pi_1$, $\pi_2$, ..., $\pi_n$ permet parfois des approximations plus aisées de certaines quantités a posteriori.
Un inconvénient des modèles hiérarchiques est qu'ils ne permettent en générale pas un calcul explicite des estimateurs de Bayes, même lorsque les niveau successifs sont conjuguées, et il faut donc avoir recours à des techniques numériques d'approximation.
On site par exemple deux techniques générales pour construire des estimateurs Bayésien hiérarchique :
La méthodes de Monte-Carlo (Robert, 2001) ;
La méthode de Monte-carlo avec fonction d'importance (Robert, 2001).

### 2.4.2 L'analyse Bayésienne empirique

L'estimation Bayésienne empirique peut être vue comme un complément du modèle hiérarchique : l'estimation Bayésienne empirique consiste à estimer la densité a posteriori quand les hyperparamètres sont inconnus.
Cette méthode est considérée comme une autre façon de résoudre la difficulté de l'incertitude dans la mesure a priori, elle est utilisée pour l'estimation des paramètres d'une loi quand les observations sont des variables aléatoires indépendantes identiquement distribuées qui suivent cette loi. On mentionne que cette technique n'est pas une méthode Bayésienne pure car elle fait appel à des approximations fréquentistes dans le cas où l'information a priori n'est pas disponible ou est insuffisante ; cette technique peut être vue

comme une bonne combinaison des méthodes Bayésiennes et fréquentistes.
L'analyse Bayésienne empirique repose sur une modélisation a priori conjuguée, en estimant les hyperparamètres à partir des observations et en utilisant ensuite cette loi a priori estimée comme loi a priori normale pour l'inférence. Lorsque l'analyse Bayésienne hiérarchique est trop compliquée à mettre en œuvre, l'analyse Bayésienne empirique se présente comme une méthode intéressante. Cette dernière est considérée comme une méthode duale de l'analyse Bayésienne hiérarchique présentée, et constitue dans certains cas une approximation acceptable lorsque la modélisation Bayésienne réelle est trop compliquée ou trop chère.
pour plus de détailles sur l'approche Bayésienne empirique paramétrique, voir (Morris, 1983).

Nous citons quelques autres façons de résoudre la difficulté de l'incertitude dans la mesure a priori :

L'utilisation d'une approche Bayésienne empirique de Robbins (1955) qui est essentiellement non paramétrique et que l'on appelle estimation bayésienne empirique non paramétrique.

La régression linéaire paramétrique où l'objectif est d'obtenir une inférence a posteriori robuste en présence de valeurs aberrantes. Box et Tiao (1968) furent les premier auteurs à proposer une alternative robuste et ce, en modélisant les erreurs par un mélange de deux lois normales, (pour approfondir dans la méthode, voir : Gagnon, 2012).

Marin (2000) considère une version robuste du modèle linéaire dynamique qui est moins sensible aux valeurs aberrantes. Ceci est obtenu en modélisant à la fois les erreurs de l'échantillonnage et les paramètres en tant que distributions de puissance exponentielles multivariées, au lieu de distributions normales.

Une autre possibilité d'améliorer la robustesse est offerte par l'approche Bayésienne non paramétrique ; Voir Berger (1994). Les modèles paramétriques peuvent être incorporés dans des modèles non paramétriques, ce qui permettra un gain de robustesse. Un exemple est discuté dans MacEachern et Muller (2000), où les mélanges du processus de Dirichlet sont considérés et un système MCMC efficace est présenté.

## 2.5 Les mesures de la sensibilité Bayésienne

Afin d'examiner la robustesse des procédures de l'inférence sous une classe $\Gamma$ de loi a priori, des mesures de sensibilités sont nécessaires. Dans les années récentes, deux types de ces mesures ont été étudiés. Les mesures globales de sensibilité ; et les mesures locales de sensibilité.

### 2.5.1 Les mesures globales de sensibilité

La mesure globale de sensibilité Bayésienne, s'intéresse au calcul de l'étendue (the range) de la variation d'une quantité a posteriori $T(h, \pi)$ quand la loi a priori $\pi$ varie dans

une classe $\Gamma$ de distributions de probabilités.
L'étendue que nous appellerons le range dans toute la suite est donné comme suit :

$$r = \sup_{\pi \in \Gamma} T(h, \pi) - \inf_{\pi \in \Gamma} T(h, \pi) \tag{2.1}$$

où $T(h, \pi)$, comme expliqué dans Berger (1990), peut être en général, dans un sens Bayésienne, une des trois catégories suivantes :

1. Des fonctions linéaires de la loi a priori : $T(h, \pi) = \int_\Theta h(\theta)\pi(d\theta)$, où $h$ est une fonction donnée.
   Si $h$ est la fonction de vraisemblance $f(x|\theta)$, on obtient la densité marginale des données : $m_\pi(x) = \int_\Theta f(x|\theta)\pi(d\theta)$.
2. Les rapports des fonctions linéaires de la loi a priori :

   $$T(h, \pi) = E_\pi(h(\theta)|x) = \frac{1}{m_\pi(x)} \int_\Theta h(\theta) f(x|\theta)\pi(d\theta)$$

   pour quelques fonctions $h$ données. Si par exemple, on prend $h(\theta) = \theta$, $T(h, \theta)$ est la moyenne a posteriori.
3. Les rapports des fonctions non linéaires :$T(h, \pi) = \dfrac{1}{m_\pi(x)} \int_\Theta h(\theta, \phi(\theta)) l(\theta)\pi(d\theta)$ pour quelques fonctions $h$. Pour $h(\theta, \phi(\theta)) = (\theta - \mu(\pi))^2$ avec $\mu(\pi)$ est la moyenne a posteriori, on obtient $T(h, \pi) =$ la variance a posteriori.

   Les valeurs extrêmes des fonctions linéaires de la loi a priori quand cette dernière varie dans une classe $\Gamma$ sont faciles à calculer si les points extrêmes de $\Gamma$ peuvent être identifiés.

(Plusieurs exemples sur ces différentes catégories sont fournis dans Ghosh et al (2006)).

Si le range de la quantité a posteriori (qui est en général un nombre) est petit, alors on peut assurer la robustesse et n'importe quelle loi a priori dans la classe $\Gamma$ peut être choisi. Si ce nombre est grand, et que de nouvelles données peuvent être considérées et/ou une autre classe $\Gamma_1$ rétrécit de la classe $\Gamma$, alors les mesures de la robustesse seront recalculer en s'arrêtant lorsque le nombre deviendra petit, enfin, Si ce nombre est grand et que la classe ne peut pas être modifiée, alors il faut choisir une loi a priori dans la classe $\Gamma$ mais en tenant compte de l'influence que peut avoir ce choix sur les quantités d'intérêt.

**Autre mesure globale de sensibilité :**
Comme nous venons de le voir plus haut, l'interprétation du range n'est pas toujours clair et dans ce cas, aucun critère objectif indiquant que la robustesse est atteinte ne peut être trouver. Et c'est ce qui a poussé les chercheurs à l'idée de mesurer la sensibilité par une quantité a posteriori appropriée afin de se ramener à une évaluation plus facile. Ruggeri et sivaganesan (2000) ont introduit ce qu'ils ont appellé la sensibilité relative.
Pour une quantité d'intérêt $h(\theta)$, la sensibilité relative est définie par :

$$R_\pi = \frac{\big(T(h, \pi) - T(h, \pi_0)\big)^2}{V^\pi} \tag{2.2}$$

Où $T(h, \pi) = E^\pi(h(\theta)|x)$, $T(h, \pi_0) = E^{\pi_0}(h(\theta)|x)$ et $V^\pi$ est la variance a posteriori de $h(\theta)$ relativement à la loi a priori $\pi$. L'idée dans cette considération est que la variance a

posteriori mesure l'exactitude dans l'estimation de $h(\theta)$, par conséquent, si la distance au carrés entre $(T(h,\pi)$ et $T(h,\pi_0)$ relativement à $V^\pi$ n'est pas trop grande, la robustesse peut être prévue.
L'exemple suivant qui est essentiellement de Ruggeri et Sivaganesan (2000) illustre cette idée.

**Exemple**.( Ruggeri et Sivaganesan (2000) )

Soit $X$ une variable aléatoire de distribution de probabilité $N(\theta,1)$. Et soit $\pi_0 = N(0,2)$ une loi a priori pour $\theta$. on veut évaluer la sensibilité des inférences a posteriori de $h(\theta)=\theta$ sous la classe $\Gamma$ de toutes les distributions a priori $N(0,\tau^2)$ avec $0 \leq \tau^2 \leq 10$.
La distribution a posteriori de $\theta$ donnant $x$ sous la loi a priori $N(0,\tau^2)$ est une loi normale de moyenne $\tau^2 x/(\tau^2+1)$ et de variance $\tau^2/(\tau^2+1)$. D'où :

$$T(h(\theta),\pi) - T(h(\theta),\pi_0) = \left(\frac{\tau^2}{\tau^2+1} - \frac{2}{3}\right) x$$

et

$$R_\pi(x) = \frac{(\tau^2-2)x^2}{9\tau^2(\tau^2+1)}$$

Le range de $T(h,\pi) - T(h,\pi_0)$ est $8x/33$ et $\sup_\tau R_\pi(x) = 6.4x^2/99$. Ainsi, la robustesse peut être atteinte pour $0 \leq x \leq 4$ et bien sûr pas pour $x = 10$.

### 2.5.2 Les mesures locales de sensibilité

Les mesures globales de sensibilité peuvent être très complexes à calculer à moins que la classe $\Gamma$ des lois a priori possibles soit une classe paramétrique simple ou une classe dont les points extrêmes sont faciles à utiliser, cette approche globale peut devenir impossible à appliquer pour des modèles très compliqués. Une alternative, qui a attiré beaucoup d'attention ces dernière années est celle d'essayer d'étudier les effets de petites perturbations sur le loi a priori. C'est ce qui a été appelé la sensibilité locale. Dans cette approche, soit la sensibilité de l'ensemble de la distribution a posteriori est étudiée, soit celle d'une certaine quantité a posteriori spécifiée.
Cette approche ne sera pas étudier dans ce présent travail, et pour plus de détailles voir (Ghosh et al, 2006).

## 2.6 Discussion

Il est important de se rendre compte que la robustesse de la loi a posteriori est l'objectif idéal de l'étude de la robustesse Bayésienne. Si cela est atteint, alors, on considère que le problème de sensibilité est résolu. En outre, si la robustesse de la loi a posteriori n'est pas présente, un bayésien tentera un raffinement supplémentaire de la classe $\Gamma$, ou, si c'est possible il cherchera à obtenir plus de données. Malheureusement, les situations où la robustesse de cette loi a posteriori est tout simplement inaccessible sont courantes :

- Soit les délais de temps accordés pour l'étude sont limités, alors un raffinement supplémentaire de $\Gamma$ est impossible ;

- Soit on ne peut pas obtenir plus de données ;
- Ou alors l'analyse Bayésienne est techniquement trop difficile à mettre en œuvre pour une variété convaincantes de lois a priori crédibles (comme dans de nombreux problème non-paramétrique).

Dans des situations pareilles, il peut être raisonnable de dire qu'il n'y a pas de réponse claire au problème, du moins dans les situations où $\Gamma$ est clairement défini et que les différentes lois a priori dans $\Gamma$ donnent de différentes réponses sensibles.
Des alternatives à prendre quand la robustesse de la loi a posteriori ne peux pas être obtenue sont données dans (Berger, 1984).

## 2.7 Conclusion

Dans sa définition, La robustesse est la qualité d'une méthode capable de donner des résultats d'une exactitude et d'une précision acceptables dans des conditions diverses. L'étude de la robustesse Bayésienne fait un objet de recherches pour touts ceux qui sont dans le domaine ; en effet, ces études sur la robustesse ont récemment suscité un intérêt considérable chez les Bayésiens.
Dans ce précédent chapitre, on a vu que la robustesse Bayésienne peut être établie en considérant une classe de lois a priori pour les paramètres du modèle et évaluer les changements sur les quantités a posteriori, cela permet de meilleures performances de l'estimation Bayésienne.

# Chapitre 3

# Application

Application à l'estimation Bayésienne robuste dans un modèle Gaussien sous une fonction de perte asymétrique (LINEX)

## 3.1 Introduction

L'analyse Bayésienne robuste ou analyse de sensibilité Bayésienne vise à quantifier et à interpréter l'incertitude induite par la connaissance partielle de l'un (ou plus) des trois éléments de l'analyse Bayésienne (l'information a priori, la vraisemblance et la fonction de perte). L'objectif de cette analyse est d'aboutir à une décision optimale sous une fonction de perte et une distribution a priori spécifiées sur l'espace des paramètres. L'étude de cette sensibilité se concentre principalement sur le calcul des changements apporté sur les quantités a posteriori lorsque la loi a priori $\pi$ varie dans une classe $\Gamma$ ; l'utilisation d'une classe $\Gamma$ de lois a priori reste la manière la plus efficace pour étudier la robustesse des méthodes Bayésienne, d'une part, parce qu'elle remédie à la critique sur le choix d'une loi a priori unique (en raison de sa partialité et son arbitraire), et d'autre part, les changements dans la vraisemblance sont rarement abordés en raison de la complexité mathématique et le fait que le choix du modèle statistique contrairement à celui de la loi a priori est considéré comme plus objectif.

Dans ce présent chapitre, l'article ***Robust Bayesian estimation in a normal model whith asymmetric loss function*** de Boratyńska et Drozdowicz (1999) est étudié.
Dans cet article, les auteurs estiment un paramètre inconnu $\theta$ en considérant une fonction de perte asymétrique introduite par Varian (1975) (LINEX) définit par :

$$L(\theta, d) = \exp\big(a(\theta - d)\big) - a(\theta - d) - 1, \qquad a \neq 0$$

Ils supposent une certaine incertitude sur la loi a priori et introduisent deux classes de lois a priori, puis construisent les estimateurs les plus robustes et les estimateurs conditionnels $\Gamma$-minimax, puis, donnent les conditions où ces estimateurs peuvent coïncider.

## 3.2 Présentation du modèle

soient $X_1, ..., X_n$ des variables aléatoires *iid* de loi normale $N(\theta, b^2)$ où la moyenne $\theta$ est inconnue et la variance $b^2$ est connue. Pour simplifier les calculs qui suivront, les

auteurs de l'article définissent les variables suivantes :

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \lambda = \lambda(\sigma) = \left(\frac{1}{\sigma^2} + \frac{n}{b^2}\right)^{-1},$$

$$v_n = \frac{n(\overline{X} - \mu_0)}{b^2}, \qquad m = m(\mu) = \mu\left(1 - \frac{n}{b^2}\left(\frac{1}{\sigma^2} + \frac{n}{b^2}\right)^{-1}\right),$$

$$w_n = \left(\frac{a}{2} + \frac{n\overline{X}}{b^2}\right)\left(\frac{1}{\sigma_0^2} + \frac{n}{b^2}\right).$$

## 3.3 Estimation Bayésienne du paramètre $\theta$

Dans ce qui suit, on cherche à estimer le paramètre inconnu $\theta$ en considérant la fonction de coût LINEX (LINear EXponential) :

**Dans le cas où $\pi$ est spécifiée :**

Dans ce cas, ils considère que la loi a priori du paramètre inconnu $\theta$ est donnée par :

$$\pi_{\mu_0,\sigma_0} = N(\mu_0, \sigma_0^2)$$

où la moyenne $\mu_0$ et la variance $\sigma_0$ sont connus, Dans ce cas, et d'après la table des lois conjuguées naturelles, la loi a posteriori $\pi(\theta|x)$ sera donnée par :

$$\pi(\theta|x) = N(\mu_0 + v_0\lambda_0, \lambda_0) = N(m_0 + w_n - \frac{a\lambda_0}{2}, \lambda_0)$$

où $\lambda_0 = \lambda(\sigma_0)$ et $m_0 = m(\mu_0)$.

Par définition, l'estimateur de Bayes associé au coût $L$ , et à la distribution a priori $\pi_{\mu_0,\sigma_0}$, est toute décision $\widehat{\theta}$ qui minimise le risque a posteriori $\rho(\pi, \widehat{\theta})$, c'est-à-dire :

$$\begin{aligned}
&\frac{\partial \rho(\pi, \widehat{\theta})}{\partial \widehat{\theta}} = 0 \\
&\Rightarrow \frac{\partial E^{\pi}[L(\theta, \widehat{\theta})]}{\partial \widehat{\theta}} = 0 \\
&\Rightarrow \frac{\partial}{\partial \widehat{\theta}}\Big(E^{\pi}\big(\exp\{a(\theta - \widehat{\theta})\} - a(\theta - \widehat{\theta}) - 1\big)\Big) = 0 \\
&\Rightarrow \frac{\partial}{\partial \widehat{\theta}}\Big(e^{-a\widehat{\theta}}E(e^{a\theta}) - aE(\theta) + a\widehat{\theta} - 1\Big) = 0 \\
&\Rightarrow -ae^{-a\widehat{\theta}}E(e^{a\theta}) + a = 0 \\
&\Rightarrow \widehat{\theta}^{Bay}_{\mu_0,\sigma_0} = \frac{1}{a}\ln E(e^{a\theta})
\end{aligned}$$

où $E(e^{a\theta}) = \exp\{a\mu_0 + (a^2/2 + av_n)\lambda_0\} = \exp(am + aw_n)$.

**Dans le cas où $\pi$ n'est pas spécifiée :**

Dans ce cas, les auteurs supposent que la distribution a priori n'est pas spécifiée et considèrent deux classes $\Gamma_{\mu_0}$ et $\Gamma^*_{\sigma_0}$ de lois a priori qui expriment respectivement l'incertitude de la variance $\sigma^2$ et de la moyenne $\mu$ de la loi a priori obtenue.

$$\Gamma_{\mu_0} = \{\pi_{\mu_0,\sigma} = \pi_{\mu_0,\sigma} = N(\mu_0, \sigma^2), \quad \sigma \in (\sigma_1, \sigma_1)\}$$

où $\sigma_1 < \sigma_2$ sont fixés et $\sigma_0 \in (\sigma_1, \sigma_2)$. Et

$$\Gamma^*_{\sigma_0} = \{\pi_{\mu,\sigma_0} = \pi_{\mu,\sigma_0} = N(\mu, \sigma_0^2), \quad \mu \in (\mu_1, \mu_1)\}$$

où $\mu_1 < \mu_2$ sont fixés et $\mu_0 \in (\mu_1, \mu_2)$.

En notant $R_x(\mu, \sigma, \widehat{\theta})$ le risque a posteriori de l'estimateur $\widehat{\theta}$ quand la loi a priori est la loi $N(\mu, \sigma^2)$, ce dernier peut être exprimé par les deux formules qui suivent selon l'incertitude :

**1.** *L'incertitude est sur $\sigma^2$ :*

Dans ce cas, on aura comme loi a posteriori la loi $N(\mu_0 + v_n\lambda, \lambda)$, et le risque a posteriori sera donné en fonction de $\lambda$ par :

$$\begin{aligned} R(\mu_0, \sigma, \widehat{\theta}) = \varrho_{\mu_0}(\lambda, \widehat{\theta}) &= E\big(L(\theta, \widehat{\theta})\big) \\ &= E\big(\exp\{-a\widehat{\theta} + a\theta\} - a\theta + a\widehat{\theta} - 1\big) \\ &= e^{-a\widehat{\theta}}E(e^{a\theta}) - aE(\theta) + a\widehat{\theta} - 1 \end{aligned}$$

où $E(\theta) = \mu_0 + v_n\lambda$ et $E(e^{a\theta}) = \exp\big(a\mu_0 + (a^2/2 + av_n)\lambda\big)$, donc :

$$\begin{aligned} \varrho_{\mu_0}(\lambda, \widehat{\theta}) &= e^{-a\widehat{\theta}}e^{a\mu_0 + (a^2/2 + av_n)\lambda} + a\widehat{\theta} - a\mu_0 - av_n\lambda - 1 \\ &= \exp\big(-a\widehat{\theta} + a\mu_0 + (a^2/2 + av_n)\lambda\big) - a(\mu_0 + v_n\lambda) + a\widehat{\theta} - 1 \end{aligned}$$

$\lambda$ est une fonction de $\sigma$ ; donc si $\sigma \in [\sigma_1, \sigma_2]$ alors $\lambda \in [\lambda_1, \lambda_2]$ où $\lambda_i = \lambda(\sigma_i), i = 1, 2$

**2.** *L'incertitude est sur $\mu$ :*

Dans ce cas, on aura comme loi a posteriori la loi $N(\mu + v_n\lambda_0, \lambda_0)$, et le risque a posteriori sera donné en fonction de $m$ par :

$$\begin{aligned} R(\mu, \sigma_0, \widehat{\theta}) = \varrho^*_{\sigma_0}(m, \widehat{\theta}) &= E\big(L(\theta, \widehat{\theta})\big) \\ &= E\big(\exp\{-a\widehat{\theta} + a\theta\} - a\theta + a\widehat{\theta} - 1\big) \\ &= e^{-a\widehat{\theta}}E(e^{a\theta}) - aE(\theta) + a\widehat{\theta} - 1 \end{aligned}$$

où $E(\theta) = \mu + v_n\lambda_0 = m + w_n - a\lambda_0/2$,
et $E(e^{a\theta}) = \exp\big(a\mu + (a^2/2 + av_n)\lambda_0\big) = \exp(am + aw_n)$, donc :

$$\begin{aligned} \varrho^*_{\sigma_0}(m, \widehat{\theta}) &= e^{-a\widehat{\theta}}e^{a\mu + (a^2/2 + av_n)\lambda_0} - a(m + w_n - a\lambda_0/2) + a\widehat{\theta} - 1 \\ &= \exp(-a\widehat{\theta} + am + aw_n) - a(m + w_n) + a^2\lambda_0/2 + a\widehat{\theta} - 1 \end{aligned}$$

$m$ est une fonction de $\mu$ ; donc si $\mu \in [\mu_1, \mu_2]$ alors $m \in [m_1, m_2]$ où $m_i = m(\mu_i), i = 1, 2$

## 3.4 Le range du risque a posteriori de l'éstimateur de Bayes

Dans ce qui suit, les auteurs cherchent à mesurer l'oscillation du risque a posteriori de l'estimateur de Bayes $\widehat{\theta}^{Bay}_{\mu_0,\sigma_0}$ lorsque la loi a priori $\pi$ traverse la classe $\Gamma$.
Ils considère la loi $\pi_{\mu_0,\sigma_0}$, il est évident que cette loi appartient à $\Gamma_{\mu_0}$ et aussi à $\Gamma^*_{\sigma_0}$.

On commence par calculer le range du risque a posteriori de l'estimateur de Bayes sous la loi a priori $\pi_{\mu_0,\sigma}$ arbitraire dans la classe $\Gamma_{\mu_0}$, ce risque est donné par :

$$\varrho_{\mu_0}(\lambda, \widehat{\theta}^{Bay}_{\mu_0,\sigma_0}) = \exp\big((a^2/2 + av_n)(\lambda - \lambda_0)\big) - av_n(\lambda - \lambda_0) + a^2\lambda_0/2 - 1$$

Dans ce cas, le range sera donné par :

$$\begin{aligned} r_{\mu_0}(\widehat{\theta}^{Bay}_{\mu_0,\sigma_0}) &= \sup_{\lambda\in(\lambda_1,\lambda_2)} \varrho_{\mu_0}(\lambda, \widehat{\theta}^{Bay}_{\mu_0,\sigma_0}) - \inf_{\lambda\in(\lambda_1,\lambda_2)} \varrho_{\mu_0}(\lambda, \widehat{\theta}^{Bay}_{\mu_0,\sigma_0}) \\ &= \sup_{\lambda\in(\lambda_1,\lambda_2)} f(\lambda) - \inf_{\lambda\in(\lambda_1,\lambda_2)} f(\lambda) \end{aligned}$$

où $f(\lambda) = \varrho_{\mu_0}(\lambda, \widehat{\theta}^{Bay}_{\mu_0,\sigma_0})$
Maintenant, pour pouvoir calculer $r_{\mu_0}(\widehat{\theta})$, on cherche le $\widehat{\lambda}$ qui minimise $f(\lambda)$ c'est-à-dire résoudre l'équation :

$$\frac{\partial f(\lambda)}{\partial \lambda} = 0$$

Après calcul, la solution de l'équation ci-dessus sera donnée par :

$$\widehat{\lambda} = \lambda_0 + (a^2/2 + av_n)^{-1} \ln \frac{v_n}{a/2 + v_n}$$

d'où :

$$r_{\mu_0}(\widehat{\theta}^{Bay}_{\mu_0,\sigma_0}) = \begin{cases} f(\lambda_2) - f(\lambda_1) & si - a/2 \le v_n < 0 \ et \ a > 0, \\ & ou \ 0 < v_n \le -a/2, \ ou \ \widehat{\lambda} < \lambda_1 \\ f(\lambda_2) - f(\widehat{\lambda}) & sinon \end{cases}$$

$$= \begin{cases} e^{z(\lambda_1-\lambda_0)}(e^{z\delta} - 1) - av_n\delta & si - a/2 \le v_n < 0 \ et \ a > 0, \\ & ou \ 0 < v_n \le -a/2, \ ou \ \widehat{\lambda} < \lambda_1 \\ a^2\delta/2 & si \ \ v_n = -a/2 \\ e^{z(\lambda_1-\lambda_0)} + av_n(\widehat{\lambda} - \lambda_2 - 1/z) & sinon \end{cases}$$

avec : $z = a^2/2 + av_n$ et $\delta = \lambda_2 - \lambda_1$.

Maintenant, on calcule le range du risque a posteriori de l'estimateur de Bayes sous la loi a priori $\pi_{\mu,\sigma_0}$ arbitraire dans la classe $\Gamma^*_{\sigma_0}$, ce risque est donné par :

$$\varrho_{\sigma_0}(m, \widehat{\theta}^{Bay}_{\mu_0,\sigma_0}) = \exp\big(-a(m_0 - m)\big) + a(m_0 - m) + a^2\lambda_0/2 - 1$$

Dans ce cas, le range sera donné par :

$$r^*_{\sigma_0}(\widehat{\theta}^{Bay}_{\mu_0,\sigma_0}) = \sup_{m\in(m_1,m_2)} \varrho^*_{\sigma_0}(m, \widehat{\theta}^{Bay}_{\mu_0,\sigma_0}) - \inf_{m\in(m_1,m_2)} \varrho^*_{\sigma_0}(m, \widehat{\theta}^{Bay}_{\mu_0,\sigma_0})$$

$$= \begin{cases} e^{-a(m_0-m_2)} + a(m_0-m_2) - a & si \;\; m_0 \leq \widehat{m} \\ e^{-a(m_0-m_1)} + a(m_0-m_1) - a & si \;\; m_0 > \widehat{m} \end{cases}$$

où :

$$\widehat{m} = m_1 + \frac{1}{a}\ln\frac{\exp(am_2 - am_1) - 1}{a(m_2 - m_1)}$$

## 3.5 Les estimateurs les plus stables et les estimateurs Γ-minimax conditionnels

L'objectif de la recherche en robustesse Bayésienne, sont des décisions optimales sous une fonction de perte spécifiée et une classe Γ de distributions a priori. Dans ce qui suit, deux concepts d'optimalité correspondant à une analyse Bayésienne robuste sont considérés : la stabilité et la Γ-minimaxité conditionnel. Les conditions où ces deux types coïncident sont énoncés.
Le concept d'action conditionnel Γ-minimax a été considéré et étayé dans DasCupta et Studden (1989), Betro et Ruggeri (1992), et celui de stabilité a été développé dans Meczarski et Zieliński (1991).

La notion d'estimateurs les plus stables est la suivante : on doit trouver les estimateurs $\widehat{\theta}_{\mu_0}$ et $\widehat{\theta}^*_{\sigma_0}$ qui satisfont

$$\inf_{\widehat{\theta}} r_{\mu_0}(\widehat{\theta}) = r_{\mu_0}(\widehat{\theta}_{\mu_0}) \quad et \quad \inf_{\widehat{\theta}} r^*_{\sigma_0}(\widehat{\theta}) = r^*_{\sigma_0}(\widehat{\theta}^*_{\sigma_0})$$

Et celle d'estimateurs conditionnels Γ-minimax est : de trouver $\tilde{\theta}_{\mu_0}$ et $\tilde{\theta}^*_{\sigma_0}$ qui satisfont

$$\inf_{\widehat{\theta}} \sup_{\sigma\in(\sigma_1,\sigma_2)} R_x(\mu_0, \sigma, \widehat{\theta}) = \sup_{\sigma\in(\sigma_1,\sigma_2)} R_x(\mu_0, \sigma, \tilde{\theta}_{\mu_0})$$

et

$$\inf_{\widehat{\theta}} \sup_{\mu\in(\mu_1,\mu_2)} R_x(\mu, \sigma_0, \widehat{\theta}) = \sup_{\mu\in(\mu_1,\mu_2)} R_x(\mu, \sigma_0, \tilde{\theta}^*_{\sigma_0})$$

Dans cette section, on présentera le dernier problème que traite l'article : trouver les estimateurs cités au dessus, puis, faire une comparaison entre eux.
Pour trouver ces estimateurs, les auteurs vérifie les conditions du théorème de Meczarski (1993) :

**Théorème 3.5.1.** *(Meczarski (1993)).*
*Soit* $\Gamma = \{\pi_\alpha : \alpha \in [\alpha_1, \alpha_2]\}$ *une classe de distributions a priori, où* $\alpha$ *est un paramètre réel.*
*Soit* $\varrho(\alpha, d)$ *le risque a posteriori d'une décision d basée sur une observation x lorsque la loi a priori est* $\pi_\alpha$*. On suppose que la fonction* $\varrho(\alpha, d)$ *satisfait les conditions suivantes :*

1. *$\varrho(\alpha, .)$ est une fonction strictement convexe $\forall\alpha$ ;*
2. *$\forall d$, le plus petit point $\alpha_{min}(d)$ de $\varrho(., d)$ est unique, et $\alpha$ est une fonction strictement monotone de $d$*
3. *Pour tout $\overline{\alpha}$ et $\overline{d}$ tel que $\alpha_{min}(\overline{d}) = \overline{\alpha}$ nous avons :*

$$\forall d_1 < d_2 \leq \overline{d} \quad \frac{\varrho(\overline{\alpha}, d_2) - \varrho(\overline{\alpha}, d_1)}{d_2 - d_1} < \frac{\varrho(\alpha_{min}(d_2), d_2) - \varrho(\alpha_{min}(d_1), d_1)}{d_2 - d_1}$$

$$\forall d_2 > d_1 \geq \overline{d} \quad \frac{\varrho(\overline{\alpha}, d_2) - \varrho(\overline{\alpha}, d_1)}{d_2 - d_1} > \frac{\varrho(\alpha_{min}(d_2), d_2) - \varrho(\alpha_{min}(d_1), d_1}{d_2 - d_1};$$

4. *La fonction $\varrho(\alpha_1, d) - \varrho(\alpha_2, d)$ est une fonction monotone de $d$*

*alors :*

*(i) Si il existe un $\widehat{d}$ tel que :*

$$\sup_{\alpha\in[\alpha_1,\alpha_2]} \varrho(\alpha_1, \widehat{d}) = \varrho(\alpha_2, \widehat{d})$$

*alors $\widehat{d}$ est l'estimateur le plus stable ;*

*(ii) Si le $\widehat{d}$ qui satisfait (i) appartient à*

$$\mathcal{L}_\Gamma = \{d : \forall x \in \mathcal{X} \quad \exists\alpha \in [\alpha_1, \alpha_2] d(x) = d_\alpha^{Bay}(x)\}$$

*alors $\widehat{d}$ est un estimateur conditionnel $\Gamma$-minimax.*

Après avoir soumis leurs résultats au conditions du théorème de Meczarski, les auteurs prouvent ces résultats puis énoncent les deux théorèmes qui suivent où ils donnent les estimateurs les plus stables et les estimateurs conditionnels Γ-minimax.

**Théorème 3.5.2.** *(Boratyńska et Drozdowicz, 1999)*

*Si la classe de loi a priori est la classe $\Gamma^*_{\sigma_0}$ alors :*

$$\widehat{\theta}^*_{\sigma_0} = \widehat{\theta}^{Bay}_{\mu_1,\sigma_0} + \frac{1}{a} \ln \frac{\exp\big(a(m_2 - m_1) - 1\big)}{a(m_2 - m_1)}$$

*et $\tilde{\theta}^*_{\sigma_0} = \widehat{\theta}^*_{\sigma_0}$ pour toute valeur de l'observation $x$ de $X$.*

**Théorème 3.5.3.** *(Boratyńska et Drozdowicz, 1999)*

*Si la classe de lois a priori est la classe $\Gamma_{\mu_0}$, alors l'estimateur le plus stable $\widehat{\theta}_{\mu_0}$ existe si et seulement si la valeur de $X$ satisfait :*

$$v_n(v_n + a/2) > 0 \qquad ou \qquad v_n = -a/2$$

*Pour $v_n(v_n + a/2) > 0$,*

$$\widehat{\theta}_{\mu_0} = \widehat{\theta}^{Bay}_{\mu_0,\sigma_1} + \frac{1}{a} \ln \frac{e^{(\lambda_2-\lambda_1)(a^2/2+av_n)} - 1}{av_n(\lambda_2 - \lambda_1)}$$

*Pour $v_n = -a/2$ le range a posteriori ne dépend pas de la valeur de $\widehat{\theta}$.*
*L'estimateur conditionnel $\Gamma$-minimax est*

$$\tilde{\theta}_{\mu_0} = \begin{cases} \widehat{\theta}_{\mu_0} & si \quad v_n(v_n + a/2) > 0 \quad et \quad \exp\big((\lambda_1 - \lambda_2)(a^2/2 + av_n\big) + av_n(\lambda_2 - \lambda_1)) \\ \\ \widehat{\theta}^{Bay}_{\mu_0,\sigma_2} & sinon \end{cases}$$

*L'estimateur le plus stable dans la classe $\mathcal{L} = \{\widehat{\theta} : \forall x \exists \sigma \in [\sigma_1, \sigma_2] \widehat{\theta}(x) = \widehat{\theta}(x)\}$ est égal à l'estimateur conditionnel $\Gamma$-minimax dans la classe de touts les estimateurs.*
*Les auteurs de ces deux théorèmes ont détailler les preuves de ces derniers dans leurs article.*

## 3.6 Simulation des résultats

Dans cette section, nous allons faire une simulation (avec le langage R) des mesures globales de sensibilité où on étudiera les changements apporté sur le risque a posteriori quand la la loi a priori varie dans la classe $\Gamma_{\mu_0}$ et $\Gamma^*_{\sigma_0}$ respectivement.

On a : $X_1, ..., X_n \sim N(\theta, b^2)$ où la moyenne $\theta$ est inconnue et la variance $b^2$ est connue, $\pi_{\mu_0,\sigma_0}(\mu_0, \sigma_0^2)$ est la loi a priori de $\theta$, où $\mu_0, \sigma_0$ sont fixés. dans cette partie, on s'intéresse a vérifier la robustesse globale (calculer le range) de l'action Bayésienne lorsqu'on considère la fonction de perte LINEX et les deux classes de loi a priori $\Gamma_{\mu_0}$ et $\Gamma^*_{\sigma_0}$ vues plus haut.

On considère que : $n = 10$, $\mu_0 = 0$, $\sigma_0^2 = 1$, $b^2 = 1$, $\sigma_1 = 0$, $\sigma_2 = 10$, $\mu_1 = 0$, $\mu_2 = 10$, et $a$ est fixé à 1.2 puis à 0.5.

On commence par considérer la classe $\Gamma_{\mu_0}$ où $\mu = \mu_0$ fixé et $\sigma^2$ est inconnu, on calculera la valeur du risque a posteriori $\varrho_{\mu_0}(\lambda, \widehat{\theta}) = \exp\big(-a\widehat{\theta} + a\mu_0 + (a^2/2 + av_n)\lambda\big) - a(\mu_0 + v_n\lambda) + a\widehat{\theta} - 1$ en faisant varier $\sigma$ dans l'intervalle $(\sigma_1, \sigma_2)$ puis on calculera le range des quantités trouvées.

| $\sigma$ | $\varrho_{\mu_0}(\lambda, \widehat{\theta})$ |
|---|---|
| 0 | 0.00315833 |
| 1 | 0.06545455 |
| 2 | 0.07026034 |
| 3 | 0.07123251 |
| 4 | 0.07157943 |
| 5 | 0.0717412 |
| 6 | 0.0718294 |
| 7 | 0.07188269 |
| 8 | 0.07191732 |
| 9 | 0.07194108 |
| 10 | 0.07195808 |
| $r_{\mu_0}(\widehat{\theta}) = 0.06879975$ | |

**Tab 3.1** : Range du risque
a posteriori avec $\sigma$ qui varie et $a$ fixé à 1.2.

| $\sigma$ | $\varrho_{\mu_0}(\lambda, \widehat{\theta})$ |
|---|---|
| 0 | 0.001199937 |
| 1 | 0.01136364 |
| 2 | 0.01220144 |
| 3 | 0.01237175 |
| 4 | 0.01243259 |
| 5 | 0.01246098 |
| 6 | 0.01247645 |
| 7 | 0.01248581 |
| 8 | 0.01249189 |
| 9 | 0.01249606 |
| 10 | 0.01249904 |
| $r_{\mu_0}(\widehat{\theta}) = 0.0112991$ | |

**Tab 3.2** : Range du risque
a posteriori avec $\sigma$ qui varie et $a$ fixé à 0.5.

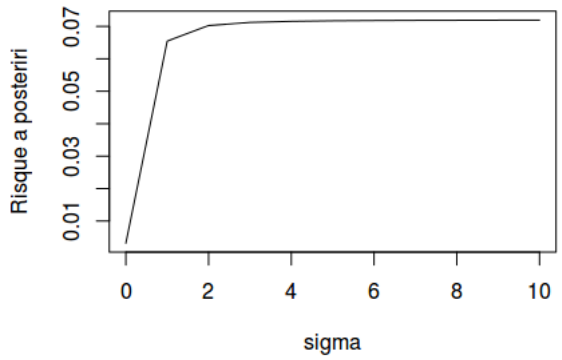


**Figure 3.1** : Variation de $\varrho_{\mu_0}(\lambda, \widehat{\theta})$ selon $\sigma$, $a = 1.2$

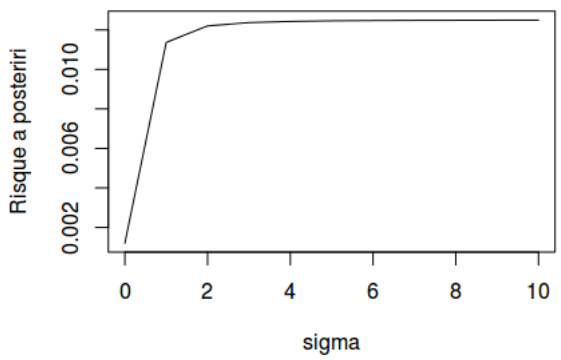


**Figure 3.2** : Variation de $\varrho_{\mu_0}(\lambda, \widehat{\theta})$ selon $\sigma$, $a = 0.5$

Maintenant, on considère la classe $\Gamma^*_{\sigma_0}$ où c'est *sigma* qui est fixé à $\sigma_0$ et $\mu$ varie dans $(\mu_1, \mu_2)$, on calculera la valeur de $\varrho^*_{\sigma_0}(m, \widehat{\theta}) = \exp(-a\widehat{\theta}+am+aw_n)-a(m+w_n)+a^2\lambda_0/2+a\widehat{\theta}-1$, puis le range des quantités trouvées.

| $\mu$ | $\varrho^*_{\sigma_0}(m, \widehat{\theta})$ |
|---|---|
| 0 | 0.06545455 |
| 1 | 0.07162737 |
| 2 | 0.09108592 |
| 3 | 0.1253616 |
| 4 | 0.1761622 |
| 5 | 0.2453925 |
| 6 | 0.3351767 |
| 7 | 0.44788441 |
| 8 | 0.5861568 |
| 9 | 0.7529415 |
| 10 | 0.9515246 |
| $r_{\sigma_0}(\widehat{\theta}) = 0.8860701$ | |

**Tab 3.3** : Range du risque
a posteriori avec $\mu$ qui varie, $a = 1.2$

| $\mu$ | $\varrho^*_{\sigma_0}(m, \widehat{\theta})$ |
|---|---|
| 0 | 0.01149086 |
| 1 | 0.01328549 |
| 2 | 0.01727738 |
| 3 | 0.02356869 |
| 4 | 0.03226636 |
| 5 | 0.04348229 |
| 6 | 0.0573336 |
| 7 | 0.07394264 |
| 8 | 0.09343825 |
| 9 | 0.1159541 |
| 10 | 0.1416307 |
| $r_{\sigma_0}(\widehat{\theta}) = 0.13013984$ | |

**Tab 3.4** : Range du risque
a posteriori avec $\mu$ qui varie, $a = 0.5$

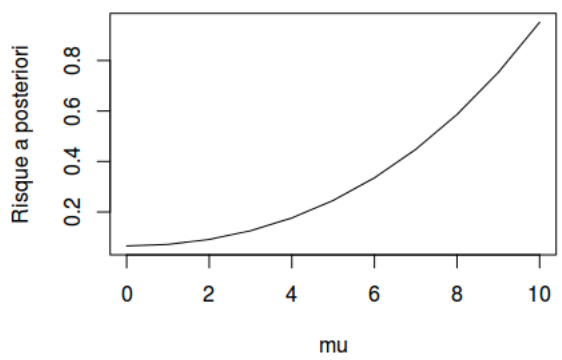


**Figure 3.3** : Variation de $\varrho^*_{\sigma_0}(m, \widehat{\theta})$ selon $\mu$, $a = 1.2$

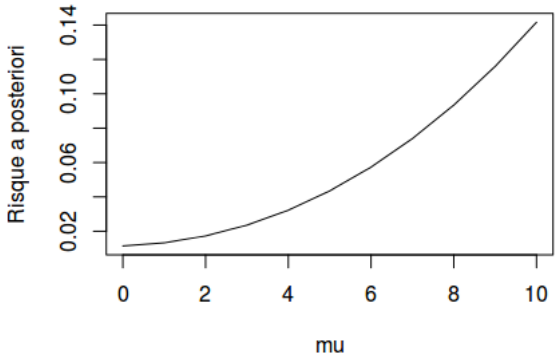


**Figure 3.4** : Variation de $\varrho^*_{\sigma_0}(m, \widehat{\theta})$ selon $\mu$, $a = 0.5$

### 3.6.1 Discussion des résultats

Arrivé au terme de cette simulation, plusieurs remarques et conclusions peuvent être donner :
La première remarque est que la robustesse change quand le paramètre de la fonction de perte change, en effet, pour un $a$ petit ($a = 0.5 < 1$) les valeurs calculées du risque a posteriori sont plus petites que celles calculées pour un $a$ plus grand ($a = 1.2 > 1$), ce qui signifie qu'on a plus de robustesse pour les petites valeurs de $a$.
On remarque aussi que les deux fonctions de risque $\varrho_{\mu_0}(\lambda, \widehat{\theta})$ et $\varrho_{\sigma_0}(m, \widehat{\theta})$ sont toutes les deux monotones et croissantes quand $\sigma$ et $\mu$ croissent respectivement. Mais ces dernières croissent différemment : contrairement à la fonction $\varrho^*_{\sigma_0}(m, \widehat{\theta})$, les changements calculés pour la fonction de risque $\varrho_{\mu_0}(\lambda, \widehat{\theta})$ sont petits et cette dernière finie par se stabiliser pour une certaine valeur de $\sigma$, par exemple, si on prend $\sigma = 100$ la valeur du risque sera 0.0720296 , cette valeur est très proche de celle calculée pour $\sigma = 10$, et le range trouvé pour celle ci est significativement plus petit que celui trouvé pour l'autre quantité a posteriori ($\varrho^*_{\sigma_0}(m, \widehat{\theta})$), ce qui signifie que la robustesse globale est réalisée pour cette classe.
On déduit donc que le choix de la classe $\Gamma$ sera porter sur la classe $\Gamma_{\mu_0}$ et que n'importe quelle loi a priori dans cette classe peut être choisi.

# Conclusion générale

La réalisation de ce mémoire qui a constituer une bonne expérience de recherche, m'a été très bénéfique, il m'a permit d'enrichir et d'approfondir mes connaissances dans une vaste approche de la statistique et qui est la l'approche statistique Bayésienne, ça m'a permis de connaitre les principaux concepts de cette analyse particulière ; comment construire une loi a priori sur le paramètre d'intérêt et quelle démarche a suivre pour mener une inférence Bayésienne robuste sur ce paramètre.

Dans ce travail, nous avons exposé les différentes étapes à suivre : du choix d'une loi a priori sur le paramètre $\theta$, de sa ré-initialisation afin d'avoir une loi a posteriori pour enfin pouvoir mener notre inférence sur ce paramètre. Nous avons aussi vue que le choix de la loi a priori été l'étape fondamentale de cette inférence et qu'un mauvais choix de cette loi pouvais donné de faux résultats et c'est pour cela qu'on s'est diriger vers l'étude de la robustesse des choix Bayésiens afin de mener une inférence plus robuste. Au finale nous avons pris l'article "Robust Bayesian estimation in a normal model with asymmetric loss function" qui illustre une étude Bayésienne robuste comme sujet d'étude, une étape très enrichissante qui m'a permis d'apprendre à bien lire et comprendre un travail de recherche.

# Bibliographie


[1] Bayes.T. Studies in the History of Probabiliy and Statistics, IX. Thomas Bayes's Essay towards solving a problem in the Doctrine of chance,(1763), in Biometrika, 45, p.296-315, (1958).

[2] Basu.S., et DasGupta.A., Robust Bayesian analysis with distribution bands. Statist. Decisions, 13, 333-349, (1995).

[3] Belkacem.N. Modèle d'incertitude appliqués aux problème de managment de l'eau, Mémoire de Magister (Ecole Doctorale) en Statistique, U.M.M.T.O, (2012).

[4] Berger.J.O., D. Rios Insua and F.Ruggeri. Bayesian Robustness.

[5] Berger.J.O, The robust Bayesian viwpoint (with discussion). In robustess of baesian analysis (J.Kadane, ed). Amesterdam : North-Holland, (1984).

[6] Berger.J.O., Statistical Decision Theory and Bayesian Analysis. (2nd Edition). New York : Springer-Verlag, (1985).

[7] Berger.J.O., Robust Bayesian Analysis : Sensitivity to the prior. J. Statist. Plan. Inference,25,303-328, (1990).

[8] Berger.J.O., An overview of Robust Bayesian Analysis. TEST, 3, 5-5, (1994).

[9] Begrer.J.O., Rios Insuna.D. et Ruggeri.F., Bayesian Robustness. In D.Rios Insuna et F.Ruggeri, editors, Robust bayesian analysis, page.1-30. Springer Verlag, (2000).

[10] Bernardo.J., Smith.A., Bayesian theory. John Wiley, New York, (1994).

[11] Berry.D.A., Statistics. A Bayesian perspective, Duxburg Press, (1996).

[12] Bessière.P., Dedieu.E., Lebeltel.O., Mazer.E., and Mekhnacha.K., Interprétation ou description (i) : Proposition pour une théorie probabiliste des systèmes cognitifs sensi-moteurs. Intellectica, 26-27 :257–311, (1998a).

[13] Bessière.P., Dedieu.E., Lebeltel.O., Mazer.E., and Mekhnacha.K., Interprétation ou description (ii) : Fondements mathématiques de l'approche f+d. Intellectica, 26-27 :313–336, (1998b).

[14] Betro.B. et Ruggeri.F., Conditional Γ-minimax actions under convex losses, Comm. Statist. Theory Methods 21 (1992), 1051–1066.

[15] Boratyńska.A., et Drozdowicz.M., Robust Bayesian estimation in a normal model with asymmetric loss function, Institute of Applied Mathematics, University of Warsaw, Banacha 2, 02-097 Warszawa, Poland ,Wojciechowskiego 22, 02-495 Warszawa, Poland, (1999).

[16] Box.G.E.P., et G.C.Tiao., A bayesian approach to some outlier problems. Biometrika, vol.55, no.1, p. 119-129, 1968.

[17] Chan, Y., Anderson, C., and Hadly, E. Bayesian estimation of the timing and severity of a population bottleneck from ancient dna. PLoS Genetics, 2, (2006).

[18] Chatterjee.N.D., Kruger.R., Haller.G., and Olbricht.W., The Bayesian approach to an internally consistent thermodynamic database : theory, database, and generation of phase diagrams. Computer., 133 :149–168, (1998).

[19] Cousins.R.D., Why isn't every physicist a Bayesian ? Amer.J.Phys., 63(5) :398–410, (1995).

[20] Cox.D.R., Hinkley.D.V., Theoretical Statistics. Chapmann and Hall, London, (1979).

[21] DasCupta.A., Studden.W.J., Frequentist behavior of robust Bayes estimates of normal means, Statist. Decisions 7 (1989), 333–361.

[22] Dey.D., Lou.K., and Bose.S., A Bayesian approach to loss robustness. Statistics and Decisions, 16, 65-87, (1998).

[23] Dey.D. and Micheas.A. Ranges of posterior expected losses and robust actions. In Robust Bayesian Analysis, (D. Rios Insua and F. Ruggeri, eds.). New York : Springer-Verlag, (2000).

[24] Demortier.L., Bayesian reference analysis. In Lyons, L. and Ünel, M. K., editors, Statistical problems in particle physics, astrophysics and cosmology, pages 11–+. Imp. Coll. Press, London, (2006).

[25] Fisher.R., Statistical Methods and Scientific inference. Oliver and Boyd, Edinburgh, (1956).

[26] Gagnon.H., Inférence robuste sur les paramètres d'une régression linéaire Bayésienne, Mémoire présenté comme exigence partielle de la maîtrise des mathématiques, Université du Quebec à Montréal, 2012.

[27] Garthwaite.P., et O'Hogan.A., Quantifying expert opinion in the water industry : an experimental study, The Statistician, vol.49, pp.455-477, (2000).

[28] Gelman.A., Carlin.J. B., Stern.H.S., and Rubin.D. B., Bayesian data analysis. Texts in Statistical Science Series. Chapman and Hall/CRC, Boca Raton, FL, second edition, (2004).

[29] Ghouil.D., Aspects de la robustesse et inférence Bayésienne des Modèles AR(1), Thèse de Magistère en Statistique, U.M.M.T.O., (20..).

[30] Ghosh Jayanta K., Delampady Mohan., Samanta Tapas., An Introduction to Bayesian Analysis Theory and Methods, Springer Texts in Statistics, Indian Statistical Institute, (2006).

[31] Gustafson.P., Local robustness in Bayesian analysis. In Robust Bayesian Analysis, (D. Rios Insua and F. Ruggeri, eds.). New York : Springer-Verlag, (2000).

[32] Jaynes.E.T., Information theory and statistical mechanics. Phys. Rev. (2), 106 :620–630, (1957).

[33] Jaynes.E.T., Probability theory. Cambridge University Press, Cambridge. The logic of science, Edited and with a foreword by G. Larry Bretthorst, (2003).

[34] Jeffreys.H., An invariant form for the prior probability in estimation problems. Proc.Roy.London.Ser.A., 186 : 453-461, (1946)

[35] Jeffreys.H., Theory of probability. Third edition. Calender Press. Oxford, (1961).

[36] Kadane.J.B., et Wolfson.L.J., Experiences in elicitation, The Statistician, vol.47, pp.1-20, (1998).

[37] Kass.R.E., et Wasserman.L., The selection of prior distributions by formal rules. Journal of American Statistical Association, (91)(435) : 1343-1370, (1996).

[38] Kording, K. P. Bayesian integration in sensorimotor learning. Nature, 15 (427) :244–7, (2004).

[39] Laplace.P.S., Sur les naissances, les mariages et les morts. Histoire de l'Academic Royale des sciences, (1786).

[40] Laplace.P.S., Théorie analytique des probabilités, vol.II. Éditions Jacques Gabay. Paris. Reprint of the 1820 third edition (book II) and of the 1816, 1818, 1820 and 1825 originals (Supplements), (1995).

[41] Lebeltel, O., Bessière, P., Diard, J., and Mazer, E. Bayesian robots programming. Autonomous Robots, 16(1) :49-79, (2003).

[42] Lindley.D.V., The 1988 Wald Memorial lectures : the present position in Bayesian statistics.
Statist.Sci., 5(1) : 44-89. With comments and a rejoinder by the author, (1990).

[43] MacEachern.S. and Muller.P. Efficient MCMC schemes for robust model extensions using encompassing Dirichlet process mixture models. In Robust Bayesian Analysis, (D. Rios Insua and F.Ruggeri, eds.). New York : Springer-Verlag, (2000).

[44] Marin.J.M., A robust version of the dynamic linear model with an economic application. In Robust Bayesian Analysis, (D. Rios Insua and F. Ruggeri, eds.). New York : Springer-Verlag, (2000).

[45] Martin.J., and Rios Insua D. Local sensitivity analysis in Bayesian Decision Theory (with discussion). In Bayesian Robustness,IMS Lectures-Notes Monograph Series, (J.O. Berger,B. Betro, E. Moreno, L.R. Pericchi, F. Ruggeri, G. Salinetti and L. Wasserman eds.), Vol 29, 119-135. Hayward : IMS, (1996).

[46] Martin.J., Rios Insua.D. et Ruggeri.F., Issues in Bayesian loss robustness. Sankhya, A, 60, 405-417, (1998).

[47] Martin.J., Rios Insua.D. et Ruggeri.F., Issues in Bayesian loss robustness. Sankhya, A, 60, 405-417, (1998). Moreno.E., Global Bayesian robustness for some classes of prior distributions. In Robust Bayesian Analysis, (D.Rios Insua and F.Ruggeri, eds). New York : Springer-Verlag, (2000).

[48] Morris.C. N., Parametric Empirical Bayes Inference : Theory and Applications. Journal of the American Statistical Association , Vol.78, N.381, p.47-65, (1983).

[49] Meczarski. M., et Zieliński.R., Stability of the Bayesian estimator of the Poisson mean under the inexactly specified gamma prior , Statist. Probab. Lett. 12 (1991), 329–333.

[50] Meczarski. M., Stability and conditional Γ-minimaxity in Bayesian inference, Appl. Math. (Warsaw) 22 (1993), 117–122.

[51] Parent.E. et Prevost.E., Inférence Bayésienne de la taille d'une population de saumons par utilisation de sources multiples d'informations, Revue de Statistique Appliquée, n.3, pp.5-38, (2003).

[52] Pohorille, A. and Darve, E. A Bayesian approach to calculating free energies in chemical and biological systems. In Bayesian inference and maximum entropy method in science and engineering, volume 872, pages 23-30. (2006)

[53] Rios Insua.D., Ruggeri.F., et Martin.J. Bayesian Sensitivity Analysis : A review. To appear in Handbook on sensitivity analysis, (A. Saltelli et al. eds.). New York : Wiley, (2000).

[54] Robbins.H., An empirical Bayes Approach to Statistics. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol.1, N.1, p.157-1 63, (1955).

[55] Robert.C.P., The Bayesian Choice. New York : Springer, 2001.

[56] Robert.C.P, The Bayesian choice, From Decision-Theoric Fondations to Computational Implementation. Springer-Verlag, New York, 2 édition, (2007).

[57] Shao.J., Mathematical Statistics. Springer-Verlag, New York, (1999).

[58] Sivaganesan.S., Global and local robustness approaches : uses and limitations. In Robust Bayesian Analysis, (D. Rios Insua and F.Ruggeri, eds.). New York : Springer-Verlag, (2000).

[59] Smyth, G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol., 3 :Art. 3, 29 pp. (electronic), (2004).

[60] Varian.H.R., A Bayesian approache to real estate assessment. In S.Fienberg, and A. Zellner, eds. Studies in Bayesian Econometrics and Statistics in Honor of L. J. Savage. North-Holland, Amesterdam, The Netherlands, (1974).

[61] Varian.H.R., A Bayesian approach to real estate assessment, in Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage, S.E. Fienberg and A.Zellner, Eds., pp. 195-208, North-Holland, Amsterdam, The Netherlands, (1975).

[62] Vines, K. S., Evilia, R. F., and Whittenburg, S. L. Bayesian analysis investigation of chemical exchange above and below the coalescence point. Journal of physical chemistry, 97 :4941–4944. (1993)

[63] Wilkinson.G., In discussion of Godambe, U.P and Thompson, Marye. Bayes, Fiducial and Frequency aspects of statistical inference in regression analysis in survey-sampling.J.Roy.Statist.Soc.Ser.B, 33 : 361-390, (1971).

[64] Wilkinson, D. J. J. Bayesian methods in bioinformatics and computational systems biology. Brief bioinform, (2007).

## Résumé :

Dans ce travail , nous avons donner un aperçu général sur l'approche statistique Bayésienne, ces principaux concepts et la démarche à suivre pour la mise en place d'une inférence sur un paramètre $\theta$, puis nous avons étudier la robustesse des choix Bayésiens afin de mener une inférence robuste.
Enfin, nous avons présenter l'article intitulé *Robust Bayesian estimation in a normal model with asymmetric loss function* où une analyse Bayésienne robuste est menée dans un modèle normale avec la fonction de perte LINEX, nous avons simuler avec le langage R les résultats de ce dernier.



**mots clés :**

Approche Bayésienne, analyse Bayésienne robuste, classes de lois a priori, classes de fonctions de perte, modèle normale, estimation, fonction de perte LINEX.