



Faculté des Sciences Exacte et Informatique
Département de Mathématiques

Mémoire de fin d'études

Présenté pour l'obtention du diplôme de

Master

Spécialité : Mathématiques.

Option : Probabilités et Statistique

Thème

Propriétés asymptotiques des facteurs de Bayes

Présenté par :

- KHALFI Rima
- MESBOUT Souad

Devant le jury :

Président	: BOUDJERDA Khawla	M.A.B Université de Jijel
Encadreur	: ABDI Zeyneb	M.A.A Université de Jijel
Examineur	: GHOUIL Djoweda	M.A.A Université de Jijel
Examineur	: YAKOBI Fatima	M.A.A Université de Jijel

Remerciement

*Tout d'abord, nous remercions dieu le
tout puissant pour son aide et pour nous avoir
guidé pour mener à bien ce travail.*

*Nous voudrions remercier chaleureusement
Notre directrice de mémoire Madame ABDI,
Université de Mohamed Sadik Ben Yahia de
Jijel, d'avoir encadré ce travail avec beaucoup de
compétence et de disponibilité.*

*Nous tenons aussi à remercier les membres de jury
Qui acceptent à examiner notre mémoire de fin d'étude.*

*Pour finir, nous souhaitons exprimer nos remerciements à :
Nos familles surtout nos parents pour leurs intérêts envers notre
travail et pour leurs aide; Nos amies, pour leurs soutien moral
et leurs présence dans les bons et les mauvais moments.*

Merci à tous et à toutes

Rjima- Souad

Table des matières

Introduction générale	3
1 L'inférence bayésienne	5
1.1 Introduction	5
1.2 Principe de l'analyse bayésienne	5
1.2.1 Modèle statistique	6
1.3 Loi a priori non informative	9
1.4 Loi conjuguée	10
1.4.1 La famille des lois exponentielles.	10
1.5 Les tests bayésiens	12
1.5.1 Une première approche à la théorie des tests	12
1.5.2 Une deuxième approche décisionnelle	15
1.6 Le facteur de Bayes	17
2 Asymptotiquement du facteur de Bayes	19
2.1 Introduction	19
2.2 Le facteur de Bayes pour les tests	19
2.2.1 Cas pour tester les paramètres	19
2.2.2 Cas pour sélectionner les modèles	20
2.3 Évaluation d'hypothèses informatives à l'aide du facteur de Bayes	22
2.3.1 Définir les hypothèses à évaluer	22
2.3.2 Calcul du facteur de Bayes	23
2.4 Variations autour du facteur de Bayes	24
2.5 Asymptotiquement du facteur de Bayes	26

3 Application	29
3.1 La sélection de modèle pour la validation	30
3.1.1 Application du facteur de Bayes pour tester les modèles	31
3.2 Application du facteur de Bayes pour tester les paramètres	36
Conclusion	38
Annex	42
Bibliographie	52

Introduction générale

La statistique bayésienne est une approche de la statistique inférentielle qui propose une réponse unitaire et globale dans le cadre paramétrique comme dans le cadre non-paramétrique.

L'objet de la statistique bayésienne est précisément de proposer une procédure qui réponde au problème. Rappelons que la statistique classique, quant à elle, fait reposer l'estimation de θ uniquement sur les observations (c-à-d sur les x_i); au cas particulier l'estimation retenue serait la moyenne des x_i .

Différents livres ont été écrits sur l'histoire et le développement de cette discipline, notamment **Stigler (1986)**, **Dale (1991)**, **Lad et Hald (1998)**, elle se base sur le même cadre de pensée de la statistique fréquentiste mais le paramètre θ n'est plus considéré comme étant totalement inconnu; il est devenu une variable aléatoire à laquelle on associe une loi dite "loi a priori", on utilise pour construire une loi dite "loi a posteriori" un célèbre théorème "théorème de Bayes", il est apparu pour la première fois en **1961** dans le cadre de l'exemple binomial exposé par le révérend **Thomas Bayes** devant la "Royal Society", et publié de façon posthume par son ami **R.Price** en **1763**. **Pierre Simon Laplace** redécouvrit ensuite cette formule dans une plus grande généralité en **1773**. Donc l'analyse statistique bayésienne vise à exploiter le plus efficacement possible l'information apportée par X sur le paramètre θ .

Notre but dans ce mémoire est de calculer le facteur de Bayes qui est un critère bayésien de sélection de modèle et de paramètre à partir des mesures disponibles afin de discriminer laquelle des deux hypothèses statistiques (l'hypothèse nulle et alternative) est la plus probable, avec une attention particulière sera portée au choix des lois a priori pour lesquelles nous proposons plusieurs techniques de construction avec le calcul est linéaire par rapport au paramètre, notre recherche se fait dans le cadre bayésien où nous utilisons les propriétés asymptotiques de facteur du Bayes.

Ce mémoire réparti en trois chapitres :

Le premier chapitre est un rappel sur la statistique bayésienne, on donne quelques notions de base sur les lois a priori, lois a posteriori, et n'oublions par le théorème de Bayes, et la notion du facteur de Bayes.

Le deuxième chapitre est consacré à l'asymptotiquement du facteur de Bayes, première-

rement on distingue le facteur de Bayes pour tester les paramètres puis pour tester les modèles avec une variation autour du facteur de Bayes, et quelques propriétés asymptotiques du facteur de Bayes.

Dans le troisième chapitre, nous avons illustré notre point de vue par des applications numériques, en calculant le facteur de Bayes pour tester les modèles avec un bon choix des lois a priori, et dans le cas linéaire par rapport au paramètre θ , de la forme :

$$f_{\theta}(x) = h(x)^T \cdot \theta$$

Avec $h(x)$ est un vecteur de fonction de régression.

Nous proposons une définition d'un modèle statistique où l'on teste $b = 0$ contre $b \neq 0$ pour les deux modèles M_0 et M_1 où

$$M_0 : y_i = f_{\theta_0}(x_i) + \varepsilon_i$$

$$M_1 : y_i = f_{\theta_1}(x_i) + b(x_i) + \varepsilon_i$$

Et prendre une décision dans le cadre de la statistique bayésienne.

Chapitre 1

L'inférence bayésienne

1.1 Introduction

L'inférence bayésienne peut être présentée comme une généralisation de l'inférence classique. Les paramètres ne sont plus des valeurs inconnues fixées, mais des variables aléatoires. Nous considérons que l'objet principal de la statistique est de tirer, au vu d'observations d'un phénomène aléatoire, soit de prévoir un événement futur, en nous attachant tout particulièrement aux aspects décisionnels de l'inférence bayésienne. Dans ce chapitre, nous allons donner quelques notions de base qui concernent la statistique bayésienne (la loi a priori, la loi a posteriori,...)

1.2 Principe de l'analyse bayésienne

Soit $X = (x_1, \dots, x_n)$ un échantillon de variable aléatoire i.i.d. de densité $f(x_i/\theta)$, $i = 1, \dots, n$ qui dépend d'un paramètre θ inconnu. La vraisemblance des observations x_1, \dots, x_n est notée $f(x/\theta)$ telle que :

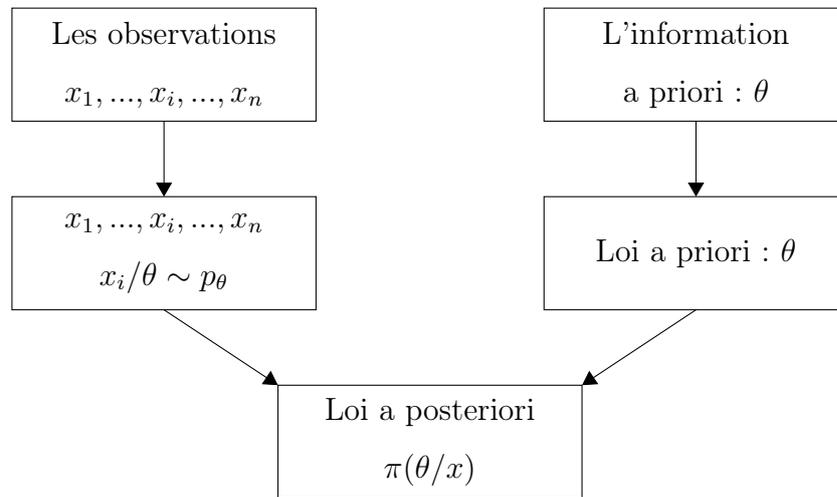
$$f(x/\theta) = \prod_{i=1}^n f(x_i/\theta)$$

Dans l'inférence bayésienne, on considère θ un paramètre aléatoire et on a associée à l'information tirée de l'échantillon, une information provenant d'une autre source, cette information additionnelle sur θ est résumée par une loi de probabilité $\pi(\cdot)$ dite "loi a priori" du paramètre θ .

L'inférence bayésienne est alors fondée sur une loi dite "loi a posteriori" de θ qui est une

fonction de vraisemblance du modèle et de la loi a priori de θ .

Le schéma suivant résume la démarche bayésienne dans le cadre de la statistique paramétrique inférentielle :



1.2.1 Modèle statistique

Le modèle statistique est un objet mathématique associé à l'observation des données issues d'un phénomène aléatoire .

Définition 1.1 *mathématiquement le modèle statistique (la structure statistique) associé à l'expérience est le triplet (χ, A, P)*

où :

χ : est l'ensemble des résultats (l'espace des observations).

A : est la tribu des évènements associés.

P : est famille de loi des probabilités possible définis sur A .

L'intérêt de ce modèle

Est qu'il permet de traiter toutes les observations dans le même formalisme.

Le modèle est discret si χ est fini $A = P(\chi)$.

On dit que le modèle est continue si $\chi \in \mathbb{R}^d, A = B(\chi)$.

Modèle statistique paramétrique

Définition 1.2 *Un modèle statistique sur (χ, A, P) consiste à l'observation d'un variable aléatoire X de loi $f(x_i/\theta), i = 1..n$, où seulement le paramètre θ est inconnu et appartient à un espace de dimension fini .*

modèle statistique bayésien

Comparée à la modélisation probabiliste l'analyse statistique se ramène fondamentalement à une inversion, car elle se propose de remonter, même imparfaitement, des effets (observations) aux causes.

En d'autre termes, quand nous observons un phénomène aléatoire contrôlé par le paramètre θ , les méthodes statistiques permettent de conduire à partir de ces observations à une inférence, alors que la modélisation probabiliste caractérise le comportement des observations conditionnellement à θ . Une telle inversion est flagrante dans la notion même de vraisemblance, puisque formellement elle ne fait que réécrire la densité dans le bon ordre, $l(x/\theta) = \ln(f(x/\theta))$. Comme une fonction de θ inconnu dépendant de x observé. une description générale de l'inversion des probabilités est donnée par le théorème de bayes :

Théorème 1.1 *Soient A et E des évènements tels que*

$$P(E) \neq 0$$

$P(A/E)$ et $P(E/A)$ sont reliées par :

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

Où A^c désigne le complémentaire de A .

Bayes donne en réalité une version continue de ces résultats, à savoir, pour deux variables aléatoires X et Y , de distribution conditionnelle $f(x/y)$ et marginale $g(y)$ alors :

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

Bayes est utilisé la version continue de ce théorème, qui énonce pour la loi a priori $\pi(\theta)$ et la fonction de vraisemblance $f(x/\theta)$ alors :

Définition 1.3 *Un modèle statistique bayésien est constitué d'un modèle statistique paramétrique $f(x/\theta)$, et d'une distribution a priori pour le paramètre θ , notée $\pi(\theta)$*

La loi a priori : La loi a priori de densité $\pi(\theta)$ est la loi marginale du paramètre θ , ou en d'autre terme, la loi du paramètre avant que x soit observé.

Elle résume, donc l'information sur θ disponible a priori (avant observation).

La loi a posteriori : C'est la loi conditionnelle de θ sachant x , sa densité notée $\pi(\theta/x)$, en vertu de la formule de Bayes on a :

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

où encore

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

Cette loi a posteriori peut s'interpréter comme une combinaison de l'information a priori disponible sur θ avec celle apportée par les observations.

La loi marginale de X : Sa densité est notée $m(x)$ et on a :

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

La loi du couple (θ, X) : Sa densité est notée $\varphi(\theta, x)$, est donnée par :

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

Exemple 1.1 *Soit x_1, \dots, x_n n -échantillon*

On considère le modèle suivant :

$$\begin{aligned} x_i &\sim \text{Bernoulli}(\theta) \quad \text{et} \quad \theta \sim \text{Beta}(a, b) \\ f(x/\theta) &= \prod_{i=1}^n P(X = x_i/\theta) = \theta^s (1 - \theta)^{n-s}, \quad s = \sum_{i=1}^n x_i \\ \pi(\theta) &= \frac{1}{\text{Beta}(a, b)} \theta^{a-1} (1 - \theta)^{b-1}, \quad 0 \leq \theta \leq 1 \end{aligned}$$

Alors :

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

$$\begin{aligned} \int_{\Theta} f(x/\theta)\pi(\theta)d\theta &= \int_{\Theta} \frac{1}{Beta(a,b)} \theta^{s+a-1}(1-\theta)^{n+b-s-1}d\theta \\ &= \frac{1}{Beta(a,b)} \int_{\Theta} \theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta \\ &= \frac{Beta(\alpha,\beta)}{Beta(a,b)} \end{aligned}$$

Où

$$\alpha = s + a$$

Et

$$\beta = n + b - s$$

$$f(x/\theta)\pi(\theta) = \frac{1}{Beta(a,b)} \theta^{s+a-1}(1-\theta)^{n+b-s-1}, 0 \leq \theta \leq 1$$

Par conséquent :

$$\theta/x \sim Beta(a + \sum_{i=1}^n x_i, b + n - \sum_{i=1}^n x_i)$$

1.3 Loi a priori non informative

Les lois non informatives représentent une ignorance sur le problème en main, mais ne signifient pas que l'on sache rien sur la distribution statistique du paramètre.

En effet on connaît au moins son domaine de variation, c-à-d l'ensemble des états de la nature Θ , et le rôle de chaque composant du paramètre sur les observations (paramètre de localisation, d'échelle, etc)

Ces lois doivent être donc particulièrement construits à partir de la distribution d'échantillonnage.

A cet égard, les lois a priori non informatives peuvent être considéré comme des lois référence aux quelles chacun pourrait avoir recours quand informations a priori sur θ est absente.

1.4 Loi conjuguée

Une des difficultés de l'approche bayésienne est le calcul de la loi a posteriori. Ce calcul sera facilité lorsque la loi a priori et la loi a posteriori sont de la même famille.

Dans ce cas, on parle de loi a priori conjuguée.

Définition 1.4 Une famille F de loi a priori sur θ est dite conjuguée si pour tout $\pi \in F$, la loi a posteriori $\pi(\theta/x)$ appartient également à F .

Une loi conjuguée peut être déterminé en considérant la forme de la vraisemblance $f(x/\theta)$ et en prenant une loi a priori de la même forme que cette dernière comme une fonction du paramètre. Les lois a priori conjuguées obtenues par ce procédé sont dites naturelles

Dans ce qui suit, nous allons aborder une des très importantes familles de loi a priori :

1.4.1 La famille des lois exponentielles.

Définition 1.5 Soit Θ l'espace des paramètres, χ l'espace des observations, C et h deux fonctions dans \mathbb{R}^+ respectivement de Θ et χ ; et soient R et T deux fonctions de Θ et χ dans \mathbb{R}^k .

1. La famille des distributions de densité

$$f(x/\theta) = h(x)\exp\{R(\theta)T(x) - C(\theta)\}$$

est dite famille exponentielle de dimension k .

2. En reparamétrisant $R(\theta) = \theta$, donc le modèle exponentielle est dit canonique c'est à dire

$$f(x/\theta) = h(x)\exp\{\theta T(x) - C(\theta)\}$$

3. Dans le cas particulier où $\Theta \subset \mathbb{R}^k$, $\chi \subset \mathbb{R}^k$ et

$$f(x/\theta) = h(x)\exp\{\theta x - C(\theta)\}$$

est dite famille exponentielle naturelle.

Exemple 1.2 *Considérons le modèle :*

$$P(X = x/\theta, \beta) = \frac{e^{(\theta-\beta)x}}{1 + e^{(\theta-\beta)}} \quad x \in \{0, 1\}$$

Il s'agit d'une loi logistique. Elle appartient à la famille exponentielle car

$$\begin{aligned} f(X/\theta, \beta) &= P(X = x/\theta, \beta) = \exp\left\{\log\left(\frac{e^{(\theta-\beta)x}}{1 + e^{(\theta-\beta)}}\right)\right\} \\ &= \exp\{(\theta - \beta)x - \log(1 + e^{\theta-\beta})\} \end{aligned}$$

et la représentation : $h(x) = 1$, $R(\theta, \beta) = (\theta, \beta)$, $T(x) = (x, -x)'$ et $C(\theta, \beta) = \log(1 + e^{\theta-\beta})$

Le tableau suivant présente quelques lois a priori conjuguées pour quelques familles exponentielles usuelles :

$f(x/\theta)$	$\pi(\theta)$	$\pi(\theta/x)$
Normale $N(\theta, \sigma^2)$	Normale $N(\mu, \tau^2)$	Normale $N(\rho(\sigma^2\mu + \tau^2x), \rho\sigma^2\tau^2)$ avec $\rho = \frac{1}{\sigma^2 + \tau^2}$
Poisson $P(\theta)$	Gamma $G(\alpha, \beta)$	Gamma $G(\alpha + x, \beta + 1)$
Gamma $G(\nu, \theta)$	Gamma $G(\alpha, \beta)$	Gamma $G(\alpha + \nu, \beta + x)$
<i>Binomiale</i> $B(n, \theta)$	<i>Beta</i> $Beta(\alpha, \beta)$	<i>Beta</i> $Beta(\alpha + x, \beta + n - x)$
Normale $N(\mu, \frac{1}{\theta})$	Gamma $G(\alpha, \beta)$	Gamma $G(\alpha + 0.5, \beta + \frac{(\mu-x)^2}{2})$
Binomiale Négative $Neg(m, \theta)$	Beta $Beta(\alpha, \beta)$	Beta $Beta(\alpha + m, \beta + x)$
Multinomiale $M_k(\theta_1, \dots, \theta_k)$	Dirichet $D(\alpha_1, \dots, \alpha_k)$	Dirichet $D(\alpha_1 + x_1, \dots, \alpha_k + x_k)$

TABLE 1.1 – Loi a priori conjuguée.

1.5 Les tests bayésiens

1.5.1 Une première approche à la théorie des tests

Soit

$$f(x/\theta), \quad \theta \in \Theta$$

Étant donné un sous ensemble d'intérêt Θ_0 de Θ , éventuellement réduit à un point, on cherche à déterminer le vrai paramètre appartenant à Θ_0 , i.e. à tester l'hypothèse :

$$H_0 : \theta \in \Theta_0$$

Sauvant appelée hypothèse nulle. Dans l'approche de Neyman-Pearson, ce problème est envisagé avec un espace de décision D , restreint à { oui, non } ou de manière équivalente à {1,0}, ce qui revient à estimer $I_{\Theta_0}(\theta)$, uniquement par des estimateurs à valeurs dans {0,1}. Il est parfois nécessaire de définir l'hypothèse alternative par rapport à laquelle on va tester H_0 contre :

$$H_1 : \theta \in \Theta_1$$

* Donc un test se formalise de la manière suivante :

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1 \end{cases}$$

De tel sorte que :

$$\Theta_0 \cap \Theta_1 = \emptyset \quad \text{et} \quad \Theta_0 \cup \Theta_1 \subset \Theta$$

* pour un modèle paramétrique :

$$X \in (\mathcal{X}, B, \{P_\theta, \theta \in \Theta\}), \Theta_0 \cup \Theta_1 \subset \Theta$$

Le cas propre correspond au cas où la loi a priori est telle que si θ suit la loi π alors :

$$\pi(\Theta_0) + \pi(\Theta_1) = 1$$

Cette dernière condition peut s'exprimer aussi sous la forme suivante :

$$\pi(\theta/\Theta_0 \cup \Theta_1) = 0$$

Ainsi puisque tout test est en fait un estimateur de $I_{\Theta_0}(\theta)$, il suffit de disposer d'une fonction de coût $L(\theta, \delta)$ pour proposer des estimateurs de Bayes associés.

Approche par la fonction de coût de type 0-1

Dans cette partie, l'espace des décisions est $D = \{0, 1\}$, la décision i correspondant à accepter H_i , $i = 0, 1$.

A ces décisions vus être associées des coûts $L(\theta, \delta_0)$ et $L(\theta, \delta_1)$, la fonction de coût est alors la suivante :

$$L(\theta, \delta) = \begin{cases} 1 & \text{si } \delta = I_{\Theta_0}(\theta) \\ 0 & \text{sinon} \end{cases}$$

par définition, Les décisions bayésiennes sont celle qui minimise le risque a posteriori $\rho(\pi, \delta/x)$

Calculons le risque a posteriori de chaque décision :

$$\begin{aligned} \rho(\pi, \delta_i/x) &= E^\pi(L(\theta, \delta_i(x))) \\ &= \int_{\theta} L(\theta, \delta_i) \pi(\theta/x) d\theta \\ &= \int_{\theta_i} \pi(\theta/x) d\theta \\ &= P^\pi(\theta \in \Theta_i/x), \quad i = 0, 1 \end{aligned}$$

Autrement dit, si

$$P^\pi(\theta \in \Theta_0/x) < P^\pi(\theta \in \Theta_1/x)$$

Alors on choisit de rejeter H_0 , un test bayésien consiste à choisir l'hypothèse ayant la plus grande probabilité a posteriori.

On peut aussi envisager un coût plus générale qui réalise différemment les mauvaises décisions suivant que l'hypothèse nulle est vraie ou fausse :

$$L(\theta, \delta) = \begin{cases} 0 & \text{si } \delta = I_{\Theta_0}(\theta) \\ a_0 & \text{si } \theta \in \Theta_0 \text{ et } \delta = 0 \\ a_1 & \text{si } \theta \in \Theta_1 \text{ et } \delta = 1 \end{cases}$$

On a alors la proposition suivante :

Proposition 1.1 *Sous le coût plus générale, l'estimateur bayésien associé à la loi a priori π est :*

$$\delta^\pi(x) = \begin{cases} 1 & \text{si } P^\pi(\theta \in \Theta_0/x) \geq \frac{a_1}{a_1+a_0} \\ 0 & \text{sinon} \end{cases}$$

Preuve 1.1 *Il suffit d'écrire le risque a posteriori*

$$\begin{aligned} \rho(\pi, \delta_i/x) &= E^\pi(L(\theta, \delta(x))/x) \\ &= \int_{\theta} L(\theta, \delta)\pi(\theta/x)d\theta \\ &= \int_{\theta_0, \delta=1} a_0\pi(\theta/x)d\theta + \int_{\theta_1, \delta=0} a_1\pi(\theta/x)d\theta \\ &= a_0P^\pi(\theta \in \Theta_0/x) + a_1P^\pi(\theta \in \Theta_1/x) \end{aligned}$$

On accepte H_0 si

$$a_1P^\pi(\theta \in \Theta_1/x) < a_0P^\pi(\theta \in \Theta_0/x)$$

I.e

$$\frac{P^\pi(\theta \in \Theta_1/x)}{P^\pi(\theta \in \Theta_0/x)} < \frac{a_0}{a_1}$$

Or

$$P^\pi(\theta \in \Theta_0/x) + P^\pi(\theta \in \Theta_1/x) = 1$$

Alors

$$\frac{P^\pi(\theta \in \Theta_1/x) + P^\pi(\theta \in \Theta_0/x) - P^\pi(\theta \in \Theta_0/x)}{P^\pi(\theta \in \Theta_0/x)} < \frac{a_0}{a_1}$$

D'où

$$\frac{1}{P^\pi(\theta \in \Theta_0/x)} - 1 < \frac{a_0}{a_1}$$

Donc

$$P^\pi(\theta \in \Theta_0/x) > \frac{a_1}{a_1 + a_0}$$

C'est la région d'acceptation de H_0

Exemple 1.3 *Soit $x \sim N(\theta, \sigma^2)$ et $\theta \sim N(\mu, \tau^2)$ alors $\pi(\theta/x)$ est la loi normale $N(\mu, \omega^2)$ avec*

$$\mu(x) = \frac{\sigma^2\mu + \tau^2x}{\sigma^2 + \tau^2} \quad \text{et} \quad \omega^2 = \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}$$

Pour tester

$$H_0 : \theta < 0$$

Nous calculons

$$P^\pi(\theta < 0/x) = P^\pi\left(\frac{\theta - \mu(x)}{\omega} < \frac{-\mu(x)}{\omega}\right) = \Phi\left(\frac{-\mu(x)}{\omega}\right)$$

Si Z_{a_0, a_1} est le quantile $\frac{a_1}{a_1 + a_0}$, donc s'il satisfait

$$\Phi(Z_{a_0, a_1}) = \frac{a_1}{a_1 + a_0}$$

On accepte H_0 lorsque

$$-\frac{\mu(x)}{\omega} > Z_{a_0, a_1}$$

I.e.

$$-\mu(x) > \omega Z_{a_0, a_1}$$

Alors l'hypothèse nulle sera acceptée lorsque :

$$x < -\frac{\sigma^2}{\tau^2}\mu - \left(1 + \frac{\sigma^2}{\tau^2}\right)\omega Z_{a_0, a_1}$$

1.5.2 Une deuxième approche décisionnelle

Dans cette section nous construisons une hypothèse alternative à l'approche de Neyman-Pearson pour justifier les probabilités a posteriori, le problème de test formalisé par Neyman et Pearson peut s'exprimer comme l'estimation de la fonction indicatrice $\mathbb{1}_{\Theta_0}(\theta)$ sous le coût 0 – 1, de façon équivalente, le coût en erreur absolue :

$$L_1(\theta, \delta) = |\delta - \mathbb{1}_{\Theta_0}(\theta)|$$

En effet, si les estimateurs δ ne prennent que les valeurs 0 et 1, il existe de nombreuses manières d'écrire le coût en erreur absolue.

Nous nous tournons alors vers une théorie moins restrictive pour laquelle les estimateurs prennent leurs valeurs dans $D = \{0, 1\}$ et peuvent être considérés comme des indicateurs de degré de certitude contre ou faveur H_0 .

Parallèlement à Shaafsma et al(1989) Hwang et al (1992) examinent cette approche des problèmes de test, pour laquelle les estimateurs de $\mathbb{1}_{\Theta_0}(\theta)$ appartiennent à $\{0, 1\}$ est levée, le choix de coût devient plus important est trop semblable à la fonction de coût 0 – 1, car

elle fournit les mêmes procédures de Bayes :

$$\delta^\pi(x) = \begin{cases} 1 & \text{si } P^\pi(\theta \in \Theta_0/x) \geq P^\pi(\theta \in \Theta_1/x) \\ 0 & \text{sinon} \end{cases}$$

En revanche, les coût quadratiques

$$L_2(\theta, \delta) = (\delta - \mathbf{1}_{\Theta_0}(\theta))^2$$

Mènent à des estimateurs plus adaptatifs.

Proposition 1.2 *Sous le coût quadratique, l'estimateur de Bayes associée à π est la probabilité a posteriori*

$$\delta^\pi(x) = P^\pi(\theta \in \Theta_0/x)$$

Preuve 1.2 : *l'estimateur bayésien comme il est défini précédemment est la règle de décision qui minimise le risque a posteriori $\rho(\pi, \delta/x)$.*

On pose :

$$f(\delta, x) = \rho(\pi, \delta/x)$$

D'où

$$\begin{aligned} f(\delta, x) &= \rho(\pi, \delta/x) = E^\pi(L(\delta(x), \theta)/x) \\ &= E^\pi((\delta - \mathbf{1}_{\Theta_0}(\theta))^2/x) \\ &= E^\pi((\delta^2 - 2\delta\mathbf{1}_{\Theta_0}(\theta) + \mathbf{1}_{\Theta_0}^2(\theta))/x) \\ &= E^\pi(\delta^2/x) - 2\delta E^\pi(\mathbf{1}_{\Theta_0}(\theta)/x) + E^\pi(\mathbf{1}_{\Theta_0}^2(\theta)/x) \\ &= \delta^2 - 2\delta E^\pi(\mathbf{1}_{\Theta_0}(\theta)/x) + E^\pi(\mathbf{1}_{\Theta_0}^2(\theta)/x) \end{aligned}$$

La décision δ qui minimise

$$f(\delta, x)$$

Est celle qui vérifiée

$$\frac{d}{d\delta} f(\delta, x) = 0$$

I.e.

$$\begin{aligned} 2\delta - 2E^\pi(\mathbf{1}_{\Theta_0}(\theta)/x) &= 0 \\ 2\delta &= 2E^\pi(\mathbf{1}_{\Theta_0}(\theta)/x) \\ \delta &= E^\pi(\mathbf{1}_{\Theta_0}(\theta)/x) \end{aligned}$$

Or

$$E^\pi(\mathbf{1}_{\Theta_0}(\theta)/x) = \int_{\Theta_0} \pi(\theta/x) d\theta = P^\pi(\theta \in \Theta_0/x)$$

donc l'estimateur de Bayes sous le coût quadratique est :

$$\delta^{(\pi)}(x) = E^\pi(\mathbf{1}_{\Theta_0}(\theta)/x)$$

1.6 Le 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.

Le facteur de Bayes est le rapport des probabilités a posteriori des hypothèses nulle et alternative sur le rapport des probabilités a priori de ces même hypothèses.

$$\begin{aligned} B_{01}(x) &= \frac{P(H_0/x)P(H_1)}{P(H_1/x)P(H_0)} \\ &= \frac{P(\theta \in \Theta_0/x)P(\theta \in \Theta_1)}{P(\theta \in \Theta_1/x)P(\theta \in \Theta_0)} \end{aligned}$$

Et on a

$$P(H_1) = 1 - P(H_0)$$

Et

$$P(H_1/x) = 1 - P(H_0/x)$$

Interprétation du facteur de Bayes

Dans la pratique on compare B_{01} avec 1

- Si $B_{01} > 1$ on accepte l'hypothèse H_0 .
- Si $B_{01} < 1$ on accepte l'hypothèse H_1

Exemple 1.4 Pour $x \sim B(n, p)$ et $p \sim U_{[0,1]}$, et on veut tester

$$H_0 : p > \frac{1}{2} \quad \text{contre} \quad H_1 : p \leq \frac{1}{2}$$

$$\begin{aligned} P(H_0/x) &= P(p > \frac{1}{2}/x) \\ &= \int_{\frac{1}{2}}^1 \pi(p/x) dp \\ &= \frac{\int_{\frac{1}{2}}^1 p^x (1-p)^{n-x} dp}{\int_0^1 p^x (1-p)^{n-x} dp} \end{aligned}$$

et

$$\begin{aligned} P(H_1/x) &= P(p \leq \frac{1}{2}/x) \\ &= \int_0^{\frac{1}{2}} \pi(p/x) dp \\ &= \frac{\int_0^{\frac{1}{2}} p^x (1-p)^{n-x} dp}{\int_0^1 p^x (1-p)^{n-x} dp} \end{aligned}$$

pour $x = 10$ et $n = 15$

Donc $P(p \leq \frac{1}{2}/x) = 15 * 10^{-4}$, et $P(p > \frac{1}{2}/x) = 0.0283$

Le facteur de Bayes associé est alors :

$$\begin{aligned} B_{01}(x) &= \frac{P(p > \frac{1}{2}/x)}{P(p \leq \frac{1}{2}/x)} \frac{\pi([\frac{1}{2}, 1])}{\pi([0, \frac{1}{2}])} \\ &= \frac{\int_{\frac{1}{2}}^1 p^x (1-p)^{n-x} dp}{\int_0^{\frac{1}{2}} p^x (1-p)^{n-x} dp} \\ &= 18.87 \quad \text{pour } x = 10, \text{ et } n = 15 \end{aligned}$$

$B_{01}(x) = 18.87 > 1$, donc on accepte H_0 .

Chapitre 2

Asymptotiquement du facteur de Bayes

2.1 Introduction

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

On va voir le facteur de Bayes pour tester les paramètres et les modèles avec quelques propriétés asymptotique du facteur de Bayes.

2.2 Le facteur de Bayes pour les tests

2.2.1 Cas pour tester les paramètres

On considère le test d'hypothèse suivant :

$$H_0 : \theta \in \Theta_0 \quad \text{contre} \quad H_1 : \theta \in \Theta_1$$

La loi a priori est donnée par le mélange de lois

$$\pi(\theta) = P(H_0)\pi_0(\theta) + (1 - P(H_0))\pi_1(\theta)$$

où $\pi_i(\theta) = P(\theta/H_i)$ est la probabilité a priori de θ sous l'hypothèses H_i . La fonction de vraisemblance sous H_i s'écrit :

$$L(X/H_i) = \int_{\theta \in \Theta_i} L(X/H_i, \theta)\pi(\theta)d\theta$$

Le facteur de Bayes de l'hypothèse H_0 en faveur de l'hypothèse H_1 s'exprime par

$$B_{01} = \frac{L(X/H_0)}{L(X/H_1)}$$

Remarque 2.1 1. Lorsque les deux hypothèses sont équiprobables le facteur de Bayes correspond au rapport des probabilités a posteriori de H_1

$$B_{01} = \frac{P(H_0/X)}{P(H_1/X)}$$

2. Dans le cas d'un test hypothèse simple contre une hypothèse composée suivant :

$$H_0 : \theta = \theta_0 \quad \text{contre} \quad H_1 : \theta \neq \theta_0$$

où $P(H_0) = P(H_1)$ et $\pi_0(\theta)$ est la masse de Dirac ou point θ_0

Le facteur de Bayes s'écrit

$$B_{01} = \frac{L(X/\theta_0)}{\int_{\theta \neq \theta_0} L(X/\theta) \pi_1(\theta) d\theta}$$

3. Le facteur de Bayes est une quantité non borné il peut prendre des valeurs entre 0 et $+\infty$

2.2.2 Cas pour sélectionner les modèles

Lorsqu'on est appelé à faire une sélection entre deux modèles M_0 et M_1 par l'approche bayésienne, on affecte à chacun des deux modèles des probabilité a priori $P(M_i)$, $i = 0, 1$ vérifiant

$$P(M_0) + P(M_1) = 1$$

Du théorème de Bayes, on peut alors exprime les probabilités a posteriori

$$P(M_i/X) = \frac{P(X/M_i)P(M_i)}{m(X)} \quad i = 0, 1$$

Ou $X = (X_1, \dots, X_n)$ est le n échantillon, et $m(X) = P(X/M_0)P(M_0) + P(X/M_1)P(M_1)$ représente la probabilité marginale des données (qu'on suppose non nulle) avec

$$P(M_1/X) = 1 - P(M_0/X)$$

on peut écrire

$$\frac{P(M_0/X)}{P(M_1/X)} = \frac{P(X/M_0)}{P(X/M_1)} \times \frac{P(M_0)}{P(M_1)} \tag{2.1}$$

où

$$B_{01} = \frac{P(X/M_0)}{P(X/M_1)}$$

représente le facteur de Bayes en faveur du modèle M_0 contre le modèle M_1 .

Le rapport $\frac{P(M_0/X)}{P(M_1/X)}$ est dit quotient d'enjeux a posteriori en faveur du modèle M_0 contre le modèle M_1 .

Dans la littérature anglaise il est dit "posterior odds ratio "

On peut alors exprimer la relation (2.1) comme suite :

quotient d'enjeux a posteriori = Facteur de bayes \times quotient d'enjeux a priori

Remarque 2.2 Si on accord le même poids a priori pour les deux hypothèses i.e :

$$P(M_0) = P(M_1) = \frac{1}{2}$$

le facteur de Bayes s'écrit :

$$B_{01} = \frac{P(M_0/X)}{P(M_1/X)}$$

Dans cette situation, il correspond exactement au quotient d'enjeux a posteriori

Exemple 2.1 On considère deux modèles :

$$M_0 : NB(x_i/1, \theta_0)$$

$$M_1 : P(x_i/\theta_1), i = 1, \dots, n$$

Dans ce cas le facteur de Bayes est donné par :

$$\begin{aligned} B_{01}(x) &= \frac{\prod_{i=1}^n NB(x_i/1, \theta_0)}{\prod_{i=1}^n p(x_i/\theta_1)} \\ &= \frac{\theta_0^n (1 - \theta_0)^{n\bar{x}}}{\theta_1^{n\bar{x}} \{\prod x_i!\}^{-1} \exp\{-n\theta_1\}} \end{aligned}$$

On suppose pour l'illustration que $\theta_0 = \frac{1}{3}$, $\theta_1 = 2$

si $n = 2$, $x_0 = x_1 = 0$

Donc $B_{01}(x) = \frac{e^4}{9} \simeq 6.1 > 1$; indiquant une légère augmentation de la plausibilité de M_0

Si $n = 2$, $x_0 = x_1 = 2$

Donc $B_{01}(x) = \frac{4e^4}{729} \simeq 0.3 < 1$; indiquant une légère augmentation de la plausibilité de M_1

2.3 Évaluation d'hypothèses informatives à l'aide du facteur de Bayes

2.3.1 Définir les hypothèses à évaluer

L'utilisation du facteur de Bayes nécessite l'emploi de modèles à comparer (Kass et Raftery, 1995). Dans le cadre de cette étude, nous parlerons plutôt d'hypothèses (informatives) à évaluer, afin de rester cohérents avec la terminologie employée par Hoijtink (2012) et Hoijtink, Klugkist et Boeleu (2008). Avec un test d'hypothèse classique, nous poserions l'hypothèse nulle (notée H_0) selon laquelle un ensemble de statistique sont égales, par exemple des probabilités notées de $j = 1$ à J

$$H_0 : P_1 = P_2 = \dots = P_J$$

Dans ce cas-ci, l'hypothèse H_0 est considérée comme une hypothèse informative, car elle donne une information très précise sur le lien existant entre les probabilités P_j : elle sont toutes égales. A l'opposé, l'hypothèse alternative classique (notée H_1) est plutôt une hypothèse informative non contrainte, car les paramètres P_j sont tous libres :

$$H_1 : \exists i, j \in \mathbb{N} \text{ tq } : i \neq j \ P_i \neq P_j$$

évidemment, la comparaison de H_0 et H_1 apparaitre une simplicité pour le chercheur qui a des hypothèses précises à évaluer : dans une situation où nous souhaitons tester la présence d'un comportement de réponse au hasard. Dans le cas de répondre à une série de questions ou d'items contenant chacun quatre choix de réponses, nous pourrions poser l'hypothèse suivante :

$$H_{hasard} : 0.25 - \zeta < P_j < 0.25 + \zeta \text{ pour les items } j = 1, \dots, J$$

Donc on va obtenir une probabilité π égale à 0.25 d'obtenir une bonne réponse pour chacun des j items (soit une chance sur quatre). Le symbole ζ permet de borner la probabilité π selon les valeurs désirées. Par exemple, dans une situation où $\zeta = 0.05$, l'hypothèse $H_{hasard} = 0.20 < P_j < 0.30$. L'avantage de cette notation repose donc sur sa grande flexibilité.

Un autre système d'hypothèses d'intérêt consisterait à évaluer si les réponses respectent le principe de monotonie. Cela implique que les réponses doivent être ordonnées en

ordre croissant de difficulté pour tout les items ou questions $j = 1, \dots, J$ d'une épreuve d'évaluation. par exemple, $P_1 > \dots > P_J$ où P_j est la probabilité d'obtenir une bonne réponse à l'item j . Notez les P_j peuvent être estimés à partir de n'importe quel modèle d'intérêt tel que le modèle de Rasch ou à partir de la proportion de bonnes réponses pour chacun des items ou questions. Aussi la monotonie implique que l'ordre de difficulté soit respecté autant pour les items que pour :

$$H_{ordre} : H_1 > \dots > H_J$$

Calculer le facteur de Bayes

Nous l'avons déjà dit : le facteur de Bayes permet de comparer plusieurs hypothèses entre elles. Il existe deux grandes composantes pour calculer le dit facteur : l'adéquation et la complexité, qui sont respectivement générées à partir d'une distribution a priori et d'une distribution a posteriori. Dans le cadre de cet article, nous nous en servons pour évaluer quelle hypothèse est la plus représentative des réponses fournies par chacun des étudiants à une épreuve d'évaluation. Imaginons que nous souhaitons comparer l'hypothèse H_m à l'hypothèse $H_{m'}$ en utilisant les données X . La fonction de probabilité associée à l'hypothèse m peut être définie comme étant égale à :

$$p_m(X) = \int f_m(X/\pi)h_m(\pi)d\pi$$

où π est un paramètre de probabilité, $h_m(\pi)$ est la distribution a priori de ce paramètre et $f_m(X|\pi)$ la fonction de densité des données selon l'hypothèse m . De là, on peut calculer le facteur de Bayes entre ces deux hypothèses en utilisant la forme suivante

$$B_{mm'} = \frac{p_m(X)}{p_{m'}(X)}$$

2.3.2 Calcul du facteur de Bayes

On se place dans le cadre habituel pour $j = 0, 1$, $H_j : \theta \in \Theta_j$ muni de π_j (loi a priori) et $f_j(X/\theta)$ (loi conditionnelle).

Rappelons l'expression du facteur de Bayes :

$$B_{01} = \frac{\int_{\Theta_0} f_0(X/\theta)d\pi_0(\theta)}{\int_{\Theta_1} f_1(X/\theta)d\pi_1(\theta)}$$

Les différentes méthodes de calcul sont les suivantes :

- Formules analytiques de B_{01} (rare, correspond au forme conjuguée).
- Approximation par Bic si : les modèles sont réguliers le max de vraisemblance se calcule sous H_0 et H_1 et de nombreuses observations sont disponibles.

2.4 Variations autour du facteur de Bayes

Nous nous plaçons dans le cas des lois a priori impropres pour définir le facteur de Bayes intrinsèque. L'idée est d'utiliser une partie de l'échantillon de départ pour travailler avec des lois propres.

Définition 2.1 : *Échantillon d'apprentissage (Training sample)*

Si $X^n = (X_1, \dots, X_n)$ sont les données, un sous-échantillon de X^n de taille l s'écrit $X^{(l)} = (X_{i_1}, \dots, X_{i_l})$, $l = i_1, \dots, i_l \subset 1, \dots, n$. Cet échantillon est dit d'apprentissage si et seulement si $\int_{\Theta} f(X^{(l)}(\theta))\pi(\theta)d\theta < +\infty$.

Il est ainsi possible de construire une loi a priori à partir de $X^{(l)}$

Définition 2.2 : *Échantillon d'apprentissage minimal*

$X^{(l)}$ est un échantillon d'apprentissage minimal si pour tout sous-échantillon strict de $X^{(l)}$ l'intégrale précédent sur ce sous-échantillon est impropre.

Le facteur de Bayes intrinsèque est alors défini pour $H_i : \theta \in \Theta_i$, muni de la loi $d\pi_i (i = 0, 1)$

Définition 2.3 : *Facteur de Bayes intrinsèque*

Si $X^{(l)}$ échantillon d'apprentissage pour Θ_0 et Θ_1 est minimal pour la réunion des deux, le facteur de Bayes intrinsèque est défini par :

$$B_{01}^{(l)} = \frac{\int_{\Theta_0} f(X^{(-l)}/\theta)\pi_0(\theta/X^{(l)})dv(\theta)}{\int_{\Theta_1} f(X^{(-l)}/\theta)\pi_1(\theta/X^{(l)})dv(\theta)}$$

En faisant apparaître la loi jointe de $X^{(l)}$ et $X^{(-l)}$, le facteur de Bayes intrinsèque s'écrit :

$$B_{01}^{(l)} = \frac{B_{01}^{\pi_0, \pi_1}(X^n)}{B_{01}^{\pi_0, \pi_1}(X^{(l)})}$$

avec des notations « claires », étant donné que le facteur de Bayes dépend de l'échantillon considéré. Moyenner sur les différents sous-échantillons $X^{(l)}$ permet de s'affranchir de cette propriété.

Pour éviter de moyenner sur les sous-échantillons, O'Hagan a introduit le facteur de Bayes fractionnaire.

Définition 2.4 : *Facteur de Bayes fractionnaire*

Il est défini par :

$$FBF_{01}^{(b)} = B_{01} \times \frac{\int_{\Theta_1} f^b(X/\theta) d\pi_1(\theta)}{\int_{\Theta_0} f^b(X/\theta) d\pi_0(\theta)}$$

où

$$b = \inf \left\{ t > 0 ; \int_{\Theta_1} f^{(t)}(X/\theta) d\pi_1(\theta) < +\infty, \int_{\Theta_0} f^{(t)}(X/\theta) d\pi_0(\theta) < +\infty \right\}$$

Notons que cette méthode présente l'inconvénient d'utiliser deux fois les données. Cependant, tout ce qui a été présenté dans cette section a une justification asymptotique ; en particulier, b défini ci-dessus est très vraisemblablement de la forme $\frac{1}{n}$

Exemple 2.2 *Interprétation du facteur de Bayes fractionnaire comme vraisemblance pénalisée* Considérons le modèle linéaire $M_{[1,n]}$

$$y = X_1\beta_1 + \dots + X_p\beta_p + \epsilon \quad \epsilon \sim N(0, \sigma^2)$$

Le problème consiste à sélectionner les X_j significatifs. A un sous-ensemble $[l]$ donné d'éléments, on associe le modèle $M_{[l]}$ et l'on s'intéresse à $\beta_{[l]} = \{\beta_j, j \in [l]\}$.

Le maximum de vraisemblance est donné par :

$f(X/\hat{\beta}_l, \sigma^2, M_{[l]})$ et on regarde $\text{Argmax}_l f(X/\hat{\beta}_l, \sigma^2, M_{[l]})$ Le maximum est donné par le modèle le plus gros qui inclut toutes les variables proposées dans le modèle de départ. Cependant, il faut garder en tête que β et σ sont considérés comme connus. En fait, c'est la justification de la pénalisation, il faut tenir compte de l'incertitude sur les paramètres qui s'ajoute en gardant toutes les variables. C'est pourquoi, pénaliser permet de ne prendre en compte qu'un nombre limité et raisonnable de variables.

Pour une loi a priori dépendant des variables retenues $\pi_{(l)}$, le critère retenu est de la forme :

$$l_n(\pi_{(l)}) = l_n(\hat{\theta}_{\pi_{(l)}}) - \text{pen}(n, l)$$

$\text{pen}(n, l)$: critère de pénalité

Le rapport suivant permet ainsi de comparer les deux modèles et de prendre en compte l'incertitude sur les paramètres :

$$BF_{ll'} = \frac{\int_{\Theta_l} f_l(X^n/\theta_l) d\pi_l(\theta_l)}{\int_{\Theta_{l'}} f_{l'}(X^n/\theta_{l'}) d\pi_{l'}(\theta_{l'})}$$

2.5 Asymptotiquement du facteur de Bayes

Cette section cherche en outre à établir un lien entre le facteur de Bayes et la vraisemblance pénalisée. Le contexte est désormais usuel : $X^n = (X_1, \dots, X_n)$ et

$X^n \stackrel{iid}{\sim} f(X, \theta)$, $\theta \sim \pi$, le test s'écrit toujours $H_0 : \theta \in \Theta_0$ et $H_1 : \theta \in \Theta_1$ dans le cadre un peu plus restreint $\Theta \in \mathbb{R}^d$.

Si le modèle est régulier et si $\theta \in \Theta_0$ tel que $\pi_1(\theta_0) > 0$ et de surcroît $d\pi_j(\theta) = \pi_j(\theta)d\theta$, alors dans le cas où :

$$d\pi(\theta) = \alpha\pi_0(\theta)\mathbb{1}_{\Theta_0}(\theta) + (1 - \alpha)\pi_1(\theta)\mathbb{1}_{\Theta_1}(\theta);$$

on peut définir une sorte de rapport de vraisemblance par :

$$B_{01} = \frac{\int_{\Theta_0} e^{\log(\theta_n)} d\pi_0(\theta)d\theta}{\int_{\Theta_1} e^{\log(\theta_n)} d\pi_1(\theta)d\theta}$$

où $\log(\theta_n)$ désigne $\log(f(X^n/\theta))$ la valeur de la log-vraisemblance. Les propriétés asymptotiques de B_{01} découlent du maximum du rapport de vraisemblance et sont les suivantes :

- Si $\theta_0 \in \Theta_0$, $B_{01} \xrightarrow[n \rightarrow +\infty]{p_{\theta_0}} +\infty$
- Si $\theta_0 \in \Theta_1$, $B_{01} \xrightarrow[n \rightarrow +\infty]{p_{\theta_0}} 0$

Ainsi le facteur de Bayes distingue la bonne hypothèse, avec une probabilité qui tend exponentiellement rapidement avec n vers 1. Le facteur de Bayes est dit consistant.

Remarquons que l'hypothèse $\pi_1(\theta_0) > 0$ peut paraître illusoire étant donnée la propriété $\Theta_0 \cap \Theta_1 = \emptyset$. En réalité, cela peut être le cas si par exemple Θ_0 est dans l'adhérence de Θ_1 ; en particulier, si Θ_1 est de dimension k , on peut trouver de tels Θ_0 de dimension $K - 1$ Même si cela semble contraignant, il faut se rendre compte que ce cas recouvre un grand nombre de problèmes comme nous le verrons plus loin.

Dans le cadre d'hypothèses « pseudo ponctuelles » de la forme $\Theta_0 = \{\theta_0\}$ ou $\Theta = \{\theta_1, \theta_2\}$ et $\Theta_0 = \{\theta_1 = \theta_{1,0}\}$ et pour $\Theta_1 = \Theta/\Theta_0$, si π_1 est absolument continu par rapport à la mesure de Lebesgue alors $\pi_1(\Theta_0) = 0$ et $\pi(\Theta_1) = \pi(\Theta) = 1$. Comme l'illustre l'exemple suivant, l'hypothèse ainsi formulée est « trop petite ». L'idée est que θ_0 est autant dans Θ_1 que dans Θ_0 au sens où θ_0 appartient à l'adhérence de Θ_1 .

Exemple 2.3 Hypothèse pseudo-ponctuelle

On regarde un modèle que l'on suppose mélangé au sens où il se pourrait que certains éléments soient tirés suivant une loi d'un certain paramètre et d'autres suivant la loi avec

un autre paramètre. La vraisemblance est alors :

$$pf(X/\theta_1) + (1 - p)f(X/\theta_2)$$

Ainsi, $\Theta = (p, \theta_1, \theta_2)$ et on définit : $H_0 : p = 1$ ou $\theta_1 = \theta_2$ Cela correspond à n'avoir qu'un paramètre ou deux : Θ_1 est de dimension 2, Θ_0 est de dimension 1. Le facteur de Bayes converge exponentiellement vers 0 sous H_0 et de façon polynômiale vers l'infini sous H_1 . Le problème est que l'on teste une hypothèse « trop petite » et qu'il est difficile d'accepter H_0 .

Considérons le cadre particulier d'un choix de modèle dans une famille croissante de modèle : $\forall j \Theta_j \subset \Theta_{j+1}$ tels que $\pi_j(\Theta_j) = 1$. L'inclusion doit se comprendre comme le fait que lorsqu'on est dans un modèle on est aussi dans un autre « plus grand ». L'exemple du nombre de variables explicatives dans une régression linéaire est en ce sens parlant. Le but du choix de modèle est d'établir une procédure qui permet de sélectionner le modèle « le plus petit » possible, c'est-à-dire le plus précis. Si θ_0 , la vraie valeur, est contenue dans Θ_j/Θ_{j-1} alors on veut sélectionner Θ_j comme étant le plus petit modèle contenant θ_0 et ce avec une probabilité tendant vers 1, lorsque $n \rightarrow +\infty$.

Théorème 2.1 *Hypothèses emboîtées*

Si $\forall j \theta \in \Theta_j \mapsto f_j(X, \theta)$ est régulière sur Θ_j , et si $\forall j$ la loi a priori, continue sur Θ_j , est telle que $\pi_j(\Theta_j) = 1$ et $\pi_j(\Theta_{j-1}) > 0$ alors en notant $\hat{\theta}_k$ le maximum de vraisemblance sur Θ_k et d_k la dimension de Θ_k :

$$\log \left(\frac{m_{n,h}(X^n)}{m_{n,h-1}(X^n)} \right) = \log(\hat{\theta}_k) - \log(\hat{\theta}_{k-1}) - \left(\frac{d_k - d_{k-1}}{2} \right) \log \left(\frac{n}{2\pi} \right) + O(1)$$

On voit apparaître ici un rapport de vraisemblance pénalisé. Cela correspond au critère BIC. Il est utile de formuler les hypothèses sous la forme $d_1 < d_2 < \dots$ où $\forall j \Theta_j \subset \mathbb{R}^{d_j}$. L'idée à la base des critères BIC est que l'augmentation du nombre de paramètres à estimer augmente l'incertitude et qu'il faut pénaliser le fait qu'on fait comme si les valeurs estimées sont les vraies valeurs. On pénalise donc pour prendre en compte cette incertitude. Comme le montre l'exemple qui suit, le choix de modèle peut être considéré comme un cas particulier de test et donc on peut utiliser le facteur de Bayes.

Exemple 2.4 *Choix de modèle*

Considérons $H_0 : X \sim N(\theta, \sigma^2) \quad \theta \in \mathbb{R}, \sigma > 0$ sous $\pi_0(\theta, \sigma^2)$ et $H_1 : X \sim C(\theta, \sigma^2)$ sous $\pi_1(\theta, \sigma)$

Le facteur de Bayes s'écrit :

$$B_{01} = \frac{\int_{\mathbb{R} \times \mathbb{R}^+} e^{-\frac{(X-\theta)^2}{2\sigma^2}} \frac{\pi_0(\theta, \sigma^2)}{\sqrt{2\theta\sigma^2}} d\theta d\sigma}{\int_{\mathbb{R} \times \mathbb{R}^+} \frac{\pi}{\sigma} \frac{1}{1+\left(\frac{X-\theta}{\sigma}\right)^2} \pi_1(\theta, \sigma) d\theta d\sigma}$$

Remarquons que le rapport de Bayes est une réponse à la fonction de perte 0-1 correspondant à $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} f$, $\{f_\theta \mid \theta \in \Theta\} = \mathcal{F}$, $H_0 : f \in \mathcal{F}$.

Lors d'une simplification de modèle, si on cherche à savoir dans quelle mesure un modèle en approche bien un autre, il faut considérer une distance entre la fonction f à approcher et la famille de modèle \mathcal{F} par laquelle on veut approcher f . On considère une fonction de perte au lieu de la perte 0-1 simple. Pour $\delta \in \{0, 1\}$:

$$L(\theta, \delta) = \begin{cases} (d(f, \mathcal{F}) - \epsilon) \mathbf{1}_{d(f, \mathcal{F}) > \epsilon} & \text{Si } \delta = 0 \\ (\epsilon - d(f, \mathcal{F})) \mathbf{1}_{d(f, \mathcal{F}) < \epsilon} & \text{Si } \delta = 1 \end{cases}$$

Il faut que les lois qu'on regarde (de mesure nulle dans l'espace des lois on le rappelle) ait un critère particulier pour qu'on puisse les différencier des autres lois. Ces lois particulières définissent un ensemble \mathcal{F} .

Pour résoudre notre problème il faut définir une loi a priori sur l'espace de dimension infinie contenant les lois de distribution .

Cette loi π_1 doit être centrée autour de \mathcal{F} . Si l'hypothèse H_0 était vraie alors il existerait θ tel que $F_\theta(X) \sim U_{[0,1]}$, $\forall \theta \in \Theta$ est une variable aléatoire sur $[0, 1]$.

Et puisque si X ne suit pas une loi F_θ pour aucun θ alors $F_\theta(X)$ n'est pas une variable aléatoire à répartition uniforme sur $[0, 1]$, on a un élément qui nous permet d'établir une stratégie de test.

Chapitre 3

Application

La validation des codes de calcul complexes rencontre à l'heure actuelle un intérêt grandissant dans les études industrielles du fait de l'utilisation intensive de la simulation numérique pour prédire des systèmes physiques complexes et pour lesquels l'expérience physique est infaisable ou économiquement très coûteuse à mettre en œuvre. La validation peut être définie comme l'action d'évaluer la capacité d'un code de calcul à prédire une quantité physique d'intérêt. Souvent, elle nécessite une étape préalable de calage du code, formalisée ci-dessous.

Soit $r(x) \in \mathbb{R}$ une quantité physique d'intérêt fonction d'un vecteur x de variables mesurables et soit $f_\theta(x)$ un code de calcul paramétré par un vecteur θ . Plusieurs méthodologies statistiques de validation ont été proposées dans la littérature. Elles reposent toutes sur une quantification exhaustive des incertitudes embarquées à l'intérieur du code, propagées ensuite à la réponse $f_\theta(x)$. Souvent, la principale source d'incertitude provient de la méconnaissance du paramètre θ .

Supposons que nous disposions de n mesures y_i du système physique r en plusieurs configurations x_i . Alors, pour $1 \leq i \leq n$, il est naturel d'introduire l'équation

$$y_i = r(x_i) + \epsilon_i \tag{3.1}$$

Dans lequel $\epsilon_i \sim N(0, \lambda^2)$ est un bruit i.i.d. traduisant l'erreur d'observation. Suivant la méthodologie décrite dans Kennedy et O'Hagan (2001), il n'existe aucun paramétrage permettant au code de prédire exactement $r(x)$. On pose donc,

$$r(x) = f_\theta(x) + b(x) \tag{3.2}$$

Dans lequel b est une fonction d'erreur et θ le paramétrage optimal inconnu. En combinant les équations (3.1) et (3.2), on obtient l'équation

$$y_i = f_\theta(x_i) + b(x_i) + \epsilon_i \quad (3.3)$$

Avec une hypothèse a priori de processus Gaussien pour b :

$$b(x) \sim \mathcal{PG}(0, \sigma^2 \sum_{\Psi}^x)$$

La méthodologie de validation décrite dans Bayarri et al. (2007) combine l'estimation de θ avec la prédiction de r à partir du modèle (3.3). La caractéristique majeure de cette approche est la prise en compte de l'erreur b , sans justification statistique. Par suite, l'incertitude calculée sur $r(x)$ résulte de l'incertitude sur la réponse du code à laquelle vient s'ajouter la fonction d'erreur b estimée conjointement à θ dans (3.3), ce qui peut laisser perplexe quant à la confiance à attribuer au code en termes de représentation du système physique.

Dans la partie d'application, nous proposons une définition alternative où la validation s'écrit comme un problème de sélection d'un modèle statistique où l'on teste $H_0 : b = 0$ contre $H_1 : b \neq 0$. Si $b = 0$, alors le code est validé car il est capable de prédire le comportement du système "sans erreur".

3.1 La sélection de modèle pour la validation

La méthode de validation que nous proposons ici consiste à détecter à partir des mesures disponibles $y = \{y_1, \dots, y_n\}$, s'il existe ou non une erreur systématique b entre le code et la réalité, ceci en amont de l'objectif de prédiction. Deux modèles sont donc en compétition pour prédire r en utilisant f_θ . Le premier suppose qu'il existe une valeur θ_0 du paramètre telle que f_{θ_0} prédit parfaitement r . D'où,

$$M_0 : y_i = f_{\theta_0}(x_i) + \epsilon_i \quad (3.4)$$

Le second suppose qu'il existe une fonction d'erreur telle que

$$M_1 : y_i = f_{\theta_1}(x_i) + b(x_i) + \epsilon_i \quad (3.5)$$

avec $b(x) \sim \mathcal{PG}(0, \sigma^2 \sum_{\Psi}^x)$. Considérons les deux modèles M_0 et M_1 caractérisés par leur vraisemblance respective $f_0 = f_0(y/p_0)$ et $f_1 = f_1(y/p_1)$ où $p_0 = (\theta_0, \lambda^2)$ et

$p_1 = (\theta_1, \sigma^2, \Psi, \lambda^2)$. Pour chacun des modèles M_0 et M_1 ,

$$P(M_i/y) \propto P(M_i) \int_{p_i} f_i(y/p_i) \pi_i(p_i) dp_i \quad (3.6)$$

Avec $\pi_i(p_i)$ désignant la loi a priori des paramètres et $P(M_1) = 1 - P(M_0)$. Le facteur de Bayes (Bernardo et Smith, 1994) s'écrit comme le rapport des vraisemblances marginales de chacun des modèles conditionnellement à y :

$$B_{10}(y) = \frac{P(M_1/y)}{P(M_0/y)} \times \frac{P(M_0)}{P(M_1)} \quad (3.7)$$

$$= \frac{\int_{p_1} f_1(y/p_1) \pi_1(p_1) dp_1}{\int_{p_0} f_0(y/p_0) \pi_0(p_0) dp_0} \quad (3.8)$$

Dans ce travail, on suppose que le modèle est linéaire en θ . En considérant $h(x)$ un vecteur de fonctions de régression, le modèle numérique prend donc la forme suivante :

$$f_\theta(x) = h(x)^T \theta \quad (3.9)$$

Lorsque $B(y) > 1$, le facteur de Bayes indique que les mesures y sont plus probablement issues du modèle M_1 du modèle M_0 et inversement si $B(y) < 1$.

3.1.1 Application du facteur de Bayes pour tester les modèles

Nous adaptons le formalisme de la section précédente aux caractéristiques du cas industriel que nous allons traiter. En plus de l'hypothèse de linéarité du modèle, la variance du bruit d'observation λ^2 est supposée connue.

Maintenant $p_0 = \theta_0$ et $p_1 = (\theta_1, \sigma^2, \Psi)$. Les paramètres θ_0 et θ_1 jouant le même rôle dans les deux modèles, nous choisissons a priori

$$\theta_0 = \theta_1 \sim N(\beta, V)$$

Comme $P(M_0/y)$ est explicite, calculer $B(y)$ revient à calculer $P(M_1/y)$. Lorsque n est grand, l'approximation de la loi a posteriori $\pi(\theta, \sigma^2, \Psi)$ par une loi normale multivariée permet d'approcher $P(M_1/y)$ par son approximation de Laplace. Lorsque le nombre de mesures y est restreint, on préférera utiliser des techniques d'échantillonnage d'importance, voir de lissage par noyau de $\pi(\theta, \sigma^2, \Psi)$ si la dimension de p_1 n'est pas trop élevée. On a la loi a priori de Ψ est $\pi_1(\Psi/c, d)$ et la loi a priori de σ^2 est $\pi_1(\sigma^2/a, b)$.

Le facteur de Bayes (3.8) prend alors la forme suivante :

$$B_{10}(y/a, b, c, d) = \int_{\Psi, \sigma^2} BF(y/\Psi, \sigma^2) \pi_1(\Psi/c, d) \pi_1(\sigma^2/a, b) d\Psi d\sigma^2 \quad (3.10)$$

où $BF(y, \Psi, \sigma^2)$ est appelé facteur de Bayes local, qui est un terme bien adapté pour étudier la variabilité du facteur de Bayes aux lois a priori.

D'après l'application dans langage R pour calculer le facteur de Bayes (voir l'annexe)

Les tableaux suivants représentent les résultats du facteur de Bayes pour les différentes variations des lois a priori de $\Psi(c, d)$ et $\sigma^2(a, b)$

* **supposant BF=1**

On a :

$$B_{10}(y/a, b, c, d) = \int_{\Psi, \sigma^2} \pi_1(\Psi/c, d) \pi_1(\sigma^2/a, b) d\Psi d\sigma^2 \quad (3.11)$$

1. Si a=1, b=4, c=4, d=6

On a les résultats dans la table suivante :

$\Psi(c, d)$ \ / \ $\sigma^2(a, b)$	Beta	Normale	cauchy
Gamma	1.000002	0.9487632	0.4968169
I.Gamma	39908.04	37863.37	19826.95

TABLE 3.1 – Le calcul du facteur de Bayes pour les différentes variations des lois a priori Ψ et σ^2

Interprétation des résultats :

- 1) Pour

$$\Psi \sim Beta(4, 6)$$

$$\sigma^2 \sim Gamma(1, 4)$$

On a : le facteur de Bayes $B_{10}(y) = 1.000002 > 1$. (Voir annexe, programme 1)

Donc le facteur de Bayes indique que les mesures y sont plus probablement issues du modèle M_1 que du modèle M_0 .

- 2) Pour

$$\Psi \sim Beta(4, 6)$$

$$\sigma^2 \sim I.Gamma(1, 4)$$

On a : le facteur de Bayes $B_{10}(y) = 39908.04 > 1$. (Voir annexe, programme 2)

Donc le facteur de Bayes indique que les mesures y sont plus probablement

issues du modèle M_1 que du modèle M_0 .

3) Pour

$$\Psi \sim N(4, 6)$$

$$\sigma^2 \sim \text{Gamma}(1, 4)$$

On a : le facteur de Bayes $B_{10}(y) = 0.9487632 < 1$. (Voir annexe, programme 3)

Donc le facteur de Bayes indique que les mesures y sont plus probablement issues du modèle M_0 que du modèle M_1 .

4) Pour

$$\Psi \sim N(4, 6)$$

$$\sigma^2 \sim I.\text{Gamma}(1, 4)$$

On a : le facteur de Bayes $B_{10}(y) = 37863.37 > 1$. (Voir annexe, programme 4)

Donc le facteur de Bayes indique que les mesures y sont plus probablement issues du modèle M_1 que du modèle M_0 .

5) Pour

$$\Psi \sim \text{Cauchy}(4, 6)$$

$$\sigma^2 \sim \text{Gamma}(1, 4)$$

On a : le facteur de Bayes $B_{10}(y) = 0.4968169 < 1$. (Voir annexe, programme 5)

Donc le facteur de Bayes indique que les mesures y sont plus probablement issues du modèle M_0 que du modèle M_1 .

6) Pour

$$\Psi \sim \text{Cauchy}(4, 6)$$

$$\sigma^2 \sim I.\text{Gamma}(1, 4)$$

On a : le facteur de Bayes $B_{10}(y) = 19826.95 > 1$. (Voir annexe, programme 6)

Donc le facteur de Bayes indique que les mesures y sont plus probablement

issues du modèle M_1 que du modèle M_0 .

2. * Si $a=1$, $b=1$, $c=100$, $d=10000$

On a les résultats dans la table suivante :

$\Psi(c, d)$ $\sigma^2(a, b)$	Beta	Normale	cauchy
Gamma	7.644127	0.3413449	0.496817
I.Gamma	76444.54	3413.57	4968.394

TABLE 3.2 – Le calcul du facteur de Bayes pour les différentes variations des lois a priori Ψ et σ^2

Interprétation des résultats :

1) Pour

$$\Psi \sim \text{Beta}(100, 10000)$$

$$\sigma^2 \sim \text{Gamma}(1, 1)$$

On a : le facteur de Bayes $B_{10}(y) = 7.644127 > 1$. (Voir annexe, programme 7)

Donc le facteur de Bayes indique que les mesures y sont plus probablement issues du modèle M_1 que du modèle M_0 .

2) Pour

$$\Psi \sim \text{Beta}(100, 10000)$$

$$\sigma^2 \sim \text{I.Gamma}(1, 1)$$

On a : le facteur de Bayes $B_{10}(y) = 76444.54 > 1$. (Voir annexe, programme 8)

Donc le facteur de Bayes indique que les mesures y sont plus probablement issues du modèle M_1 que du modèle M_0 .

3) Pour

$$\Psi \sim N(100, 10000)$$

$$\sigma^2 \sim \text{Gamma}(1, 1)$$

On a : le facteur de Bayes $B_{10}(y) = 0.3413449 < 1$. (Voir annexe, programme 9)

Donc le facteur de Bayes indique que les mesures y sont plus probablement issues du modèle M_0 que du modèle M_1 .

4) Pour

$$\Psi \sim N(100, 10000)$$

$$\sigma^2 \sim I.Gamma(1, 1)$$

On a : le facteur de Bayes $B_{10}(y) = 3413.57 > 1$. (Voir annexe, programme 10)

Donc le facteur de Bayes indique que les mesures y sont plus probablement issues du modèle M_1 que du modèle M_0 .

5) Pour

$$\Psi \sim Cauchy(100, 10000)$$

$$\sigma^2 \sim Gamma(1, 1)$$

On a : le facteur de Bayes $B_{10}(y) = 0.496817 < 1$. (Voir annexe, programme 11)

Donc le facteur de Bayes indique que les mesures y sont plus probablement issues du modèle M_0 que du modèle M_1 .

6) Pour

$$\Psi \sim Cauchy(100, 10000)$$

$$\sigma^2 \sim I.Gamma(1, 1)$$

On a : le facteur de Bayes $B_{10}(y) = 4968.394 > 1$. (Voir annexe, programme 12)

Donc le facteur de Bayes indique que les mesures y sont plus probablement issues du modèle M_1 que du modèle M_0 .

• **Un cas particulier :**

Pour

$$\Psi \sim N(0, 1)$$

$$\sigma^2 \sim \text{Gamma}(1, 1)$$

On a : le facteur de Bayes $B_{10}(y) = 0.4999999 < 1$. (Voir annexe, programme 13)

Donc le facteur de Bayes indique que les mesures y sont plus probablement issues du modèle M_0 que du modèle M_1 .

Pour

$$\Psi \sim N(0, 1)$$

$$\sigma^2 \sim I.\text{Gamma}(1, 1)$$

On a : le facteur de Bayes $B_{10}(y) = 5000, 221 > 1$. (Voir annexe, programme 14)

Donc le facteur de Bayes indique que les mesures y sont plus probablement issues du modèle M_1 que du modèle M_0 .

***supposant BF=2**

On a :

$$B_{10}(y/a, b, c, d) = 2 \int_{\Psi, \sigma^2} \pi_1(\Psi/c, d) \pi_1(\sigma^2/a, b) d\Psi d\sigma^2 \quad (3.12)$$

1. Si a=1, b=4, c=4, d=6

On a les résultats dans la table suivante.

$\sigma^2(a, b)$ \ $\Psi(c, d)$	Beta	Normale	cauchy
Gamma	2.000004	1.8975264	0.9936338
I.Gamma	79816.08	75726.74	39653.9

TABLE 3.3 – Le calcul du facteur de Bayes pour les différentes variations des lois a priori Ψ et σ^2

On remarque la même décision après changé la valeur du facteur de Bayes local BF.

3.2 Application du facteur de Bayes pour tester les paramètres

Dans cette partie, on considère le paramètre σ^2 proposé dans la première application (application du facteur de Bayes pour tester les modèles), et on va tester la valeur de σ^2 sous quelques conditions.

Donc on a le test suivant :

$$H_0 : \sigma^2 = \frac{1}{2} \quad \text{contre} \quad H_1 : \sigma^2 \neq \frac{1}{2}$$

Le facteur de Bayes est donné par la forme suivante :

On pose $\sigma^2 = \theta$.

$$B_{01}^{\pi}(x) = \frac{P^{\pi}(\theta \in \Theta_0/x)}{P^{\pi}(\theta \in \Theta_1/x)} / \frac{\pi(\theta \in \Theta_0)}{\pi(\theta \in \Theta_1)} \quad (3.13)$$

1. Avec

$$f(x/\theta) \sim N\left(m, \frac{1}{\theta}\right)$$

$$\theta \sim \text{Gamma}(a, b)$$

Et

$$\theta/x \sim \text{Gamma}\left(a + 0.5, b + \frac{(m - x)^2}{2}\right)$$

Où

$$\theta = \frac{1}{2}, m=0, a=1, b=4, x=3$$

D'après l'application dans langage R le calcul du facteur de Bayes (Voir annexe, programme 15) est :

$$B_{01}^{\pi} = 4.104424 > 1$$

On accepte H_0 donc $\theta \neq \frac{1}{2}$

2. Avec

$$f(x/\theta) \sim N(a, b)$$

$$\theta \sim N(c, d)$$

Et

$$\theta/x \sim N\left(\frac{bc + dx}{b + d}, \frac{bd}{b + d}\right)$$

Où

$$\theta = \frac{1}{2}, a=0, b=1, b=c=1, x=1$$

D'après l'application dans langage R le calcul du facteur de Bayes (Voir annexe, programme 16) est :

$$B_{01}^{\pi} = 0.4226113 < 1$$

On rejete H_0 donc $\theta = \frac{1}{2}$

Conclusion

Dans ce travail, nous avons étudié la théorie mathématique des tests Bayésiens. Et de l'asymptotiquement de facteur de Bayes, nous avons montré leurs intérêts et les conclusions significatives.

Nous avons montré le rôle très important du facteur de Bayes et critiqué ses propriétés et son utilisation dans la validation d'un système physique sur la base d'un modèle statistique reliant les mesures physiques disponibles avec la réponse de calcul.

On propose le problème d'un modèle statistique où l'on teste $H_0 : b = 0$ contre $H_1 : b \neq 0$ s'écrit :

$$M_0 : y_i = f_{\theta_0}(x_i) + \varepsilon_i$$

$$M_1 : y_i = f_{\theta_1}(x_i) + b(x_i) + \varepsilon_i$$

En supposant dans un première temps que le code de calcul est linéaire. Par rapport au paramètre, le facteur de Bayes est calculé à partir de facteur de Bayes locale et une attention particulière portée au choix des différents lois a priori pour les quelles nous prendrons une décisions à l'aide de logiciel R.

Résumé

Dans ce mémoire nous avons discutés certaines propriétés mathématiques de facteur de Bayes dans l'approche Bayésienne, et nous avons vus aussi l'asymptotiquement du facteur de Bayes.

Et en fin ce travail est illustré par quelques applications sous logiciel R, dans notre application le facteur de Bayes est calculé à partir des mesures disponibles afin de discriminer laquelle des deux hypothèses statistiques est la plus probable, avec une attention particulière sera portée au choix des lois a priori pour lesquelles nous proposons plusieurs techniques de construction.

Mots clés :

Approche Bayésienne, lois a priori, Facteur de Bayes, Asymptotiquement du facteur de Bayes.

summary

In this thesis we have discussed some mathematical properties of Bayesian factor in the Bayesian approach, and we have also seen the asymptotically of the Bayesian factor. And in the end this work is illustrated by some applications under R software, in our application the Bayes factor is calculated from the available measures in order to discriminate which of the two statistical hypotheses is most likely, with particular attention to the choice of a priori laws for which we propose several construction techniques.

Keywords :

Bayesian Approach, Laws A priori, Bayes Factor, Asymptotically Bayes Factor.

ملخص

ناقشنا في هذه المذكرة بعض الخصائص الرياضية لعامل بايز في النهج البايزي، ورأينا أيضا عاملا مقربا لعامل بايز وفي النهاية ، يتضح هذا العمل من خلال بعض التطبيقات تحت البرنامج R يتم خلالها حساب عامل بايز من اجل التمييز بين النظريتين الاحصائيتين ، مع ايلاء اهتمام خاص لاختيار القوانين الأولية التي نقترح عليها العديد من تقنيات البناء

الكلمات المفتاحية:

نهج بايزي ، قوانين بديهية، عامل بايز، عامل بايز مقارب

Annex

Dénomination X	Loi de probabilité	Moyenne $E(X)$	Variance $V(X)$
Loi Normale $N(\mu, \sigma^2)$ σ positif	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	μ	σ^2
Loi de Cauchy	$f(x) = \frac{1}{\pi(1+x^2)}$	non définie	non définie
Loi de Gamma $G(a, p)$ $a > 0$ et $p > 0$	$f(x) = \begin{cases} \frac{a^p}{\Gamma(p)} x^{p-1} e^{-ax} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$ où $\Gamma(p) = \int_0^{+\infty} x^{p-1} e^{-x} dx$	$\frac{p}{a}$	$\frac{p}{a^2}$
Loi de Beta $B(p, q)$ $p > 0$ et $q > 0$	$f(x) = \begin{cases} \frac{1}{B(p,q)} x^{p-1} (1-x)^{q-1} & \text{si } 0 < x < 1 \\ 0 & \text{sinon} \end{cases}$ où $B(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx$	$\frac{p}{p+q}$	$\frac{pq}{(p+q)^2(p+q+1)}$
Loi uniforme sur $[a, b]$	$f(x) = \begin{cases} \frac{1}{b-a} & \text{pour } a \leq x \leq b \\ 0 & \text{sinon} \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Loi binomiale $B(n, p)$ $n > 0$; $0 < p < 1$	$P[X = k] = C_n^k p^k (1-p)^{n-k}$ $k = 0, 1, \dots, n$	np	$np(1-p)$

TABLE 3.4 – Lois de probabilité

i.i.d. Identiquement indépendant distribuée.

Le critère d'information Bayésien (BIC) :

BIC est un critère d'information dérivé du critère d'information d'Akaike proposé par Gideon Schwarz en 1978, à la différence, la pénalité dépend de la taille de l'échantillon et pas seulement du nombre du paramètre. Il s'écrit :

$$BIC = -2\ln(L) + k\ln(N)$$

L : la vraisemblance du modèle estimée.

N : le nombre d'observation dans l'échantillon.

k : le nombre de paramètre.

Vraisemblance marginale :

la vraisemblance marginale m est donnée par :

$$m^{-1} = \int \frac{f(\theta)}{L(x/\theta)\pi(\theta)} P(\theta/x) d\theta = E_p \left[\frac{f(\theta)}{L(x/\theta)\pi(\theta)} \right]$$

ou l'espérance est par rapport la densité *inf(a posteriori)* $P(\theta/x)$, $f(\theta)$ est une densité dont les queues de distribution décroissent plus vite que $L(x/\theta)\pi(\theta)$, soit la vraisemblance pondérée par la densité *inf(a priori)*

Le système physique

Définition 3.1 *pour travailler sur un problème de physique, au lieu de traiter le problème sur "l'univers", on se restreint l'espace pour ne le traiter que dans espace plus petit restreint, ne contenant uniquement que ce qui interférera avec ce que l'on étudie, que ce qui rentrera en jeu dans l'expérience. Cet espace plus restreint est nommé système.*

monotonie :

Caractère mono tonique d'un mouvement suivant une seule tendance (croissance, décroissance) sans variation ou alternances.

Processus gaussiens :

un processus est dit gaussien si toutes ses lois fini dimensionnelles $L(x_{t_1}), \dots, x_{t_n}$ sont gaussiennes ($\forall t_1, \dots, t_n \in T$)

Autrement dit $X = (X_t)_t$ est gaussienne si toute combinaison linéaire $a_1 x_{t_1} + \dots + a_n x_{t_n}$ suit un loi gaussienne (pour tout $n \in \mathbb{N}, t_1, \dots, t_n \in T$ et a_1, \dots, a_n)

Masse de Dirac :

Définition 3.2 Soit (X, A) un espace mesurable et $a \in X$. On appelle masse de Dirac en a la mesure notée δ_a définie sur A par :

$$\delta_a = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{sinon.} \end{cases}$$

Code sous R

Si $\Psi = x[1]$; $\sigma^2 = x[2]$

> *library(cubature)*

Programme 1

> $a = 1; b = 4; c = 4; d = 6$

$I = \text{function}(x)\{(x^{(c-1)}) * (1-x)^{(d-1)}\}$

> *integrate(I, lower = 0.001, upper = 1)*

0.001984127

> $h = \text{function}(x)\{x^{(b-1)} * \exp(-x)\}$

> *integrate(h, lower = 0.001, upper = 10000)*

6

> $B = \text{function}(x)\{((x[1]^{(c-1)} * (1-x[1])^{(d-1)})/0.001984127) * ((a^b * x[2]^{(b-1)} * \exp(-a * x[2]))/6)\}$

> *adaptIntegrate(B, lowerLimit = c(0.001, 0.001), upperLimit = c(1, 10000))*

integral

[1]1.000002

Programme 2

> $a = 1; b = 4; c = 4; d = 6$

> $I = \text{function}(x)\{(x^{(c-1)}) * (1-x)^{(d-1)}\}$

> *integrate(I, lower = 0.001, upper = 1)*

0.001984127

> $l = \text{function}(x)\{x^{(a-1)} * \exp(-x)\}$

> *integrate(l, lower = 0.001, upper = 10000)*

0.9990005

> $B = \text{function}(x)\{(x[1]^{(c-1)} * (1-x[1])^{(d-1)})/(0.001984127) * (b^a * (1/x[2])^{(a-1)} * \exp(-b/x[2]))/(0.9990005)\}$

> *adaptIntegrate(B, lowerLimit = c(0.001, 0.001), upperLimit = c(1, 10000))*

integral

[1]39908.04

Programme 3

> a = 1; b = 4; c = 4; d = 6

> h = function(x){x^(b - 1) * exp(-x)}

> integrate(h, lower = 0.001, upper = 10000)

6

> B = function(x)(1/(sqrt(2 * d * pi))) * exp((-1/(2 * d)) * ((x[1] - c)^2)) * (a^b * x[2]^(b - 1) * exp(-a * x[2]))

> adaptIntegrate(B, lowerLimit = c(0, 0.001), upperLimit = c(100, 10000))

integral

[1]0.9487632

Programme 4

> a = 1; b = 4; c = 4; d = 6

> l = function(x){x^(a - 1) * exp(-x)}

> integrate(l, lower = 0.001, upper = 10000)

0.9990005

> B = function(x){(1/(sqrt(2 * d * pi))) * exp((-1/(2 * d)) * ((x[1] - c)^2)) * (b^a * (1/x[2])^(a - 1) * exp(-b/x[2]))/0.9990005}

> adaptIntegrate(B, lowerLimit = c(0, 0.001), upperLimit = c(100, 10000))

integral

[1]37863.37

Programme 5

> a = 1; b = 4

> h = function(x){x^(b - 1) * exp(-x)}

> integrate(h, lower = 0.001, upper = 10000)

6

> B = function(x){(1/(pi * (1 + x[1]^2))) * (a^b * x[2]^(b - 1) * exp(-a * x[2]))/6}

> adaptIntegrate(B, lowerLimit = c(0, 0.001), upperLimit = c(100, 10000))

integral

[1]0.4968169

Programme 6

```

> a = 1; b = 4
> l = function(x){x^(a - 1) * exp(-x)}
> integrate(l, lower = 0.001, upper = 10000)
0.9990005
> B = function(x){(1/(pi*(1+x[1]^2)))*(b^a*(1/x[2])^(a-1)*exp(-b/x[2]))/0.9990005}
> adaptIntegrate(B, lowerLimit = c(0, 0.001), upperLimit = c(100, 10000))
integral
[1]19826.95

```

Programme 7

```

> a = 1; b = 1; c = 100; d = 10000
> I = function(x){(x^(c - 1)) * (1 - x)^(d - 1)}
> integrate(I, lower = 0.001, upper = 1)
7.454345e - 246
> h = function(x){x^(b - 1) * exp(-x)}
> integrate(h, lower = 0.001, upper = 10000)
0.9990005
> B = function(x){((x[1]^(c - 1) * (1 - x[1])^(d - 1))/7.454345e - 246) * ((a^b * x[2]^(b - 1) * exp(-a * x[2]))/0.9990005)}
> adaptIntegrate(B, lowerLimit = c(0.001, 0.001), upperLimit = c(1, 10000))
integral
[1]7.644127

```

Programme 8

```

> a = 1; b = 4; c = 4; d = 6
> I = function(x){(x^(c - 1)) * (1 - x)^(d - 1)}
> integrate(I, lower = 0.001, upper = 1)
0.001984127
> l = function(x){x^(a - 1) * exp(-x)}
> integrate(l, lower = 0.001, upper = 10000)
0.9990005

```

```
> B = function(x){(x[1]^(c-1)*(1-x[1])^(d-1))/(0.001984127)*(b^a*(1/x[2])^(a-1)*exp(-b/x[2]))/(0.9990005)}
```

```
> adaptIntegrate(B, lowerLimit = c(0.001, 0.001), upperLimit = c(1, 10000))
```

```
integral
```

```
[1]39908.04
```

Programme 9

```
> a = 1; b = 4; c = 4; d = 6
```

```
> h = function(x){x^(b-1)*exp(-x)}
```

```
> integrate(h, lower = 0.001, upper = 10000)
```

```
6
```

```
> B = function(x){(1/(sqrt(2*d*pi)))*exp((-1/(2*d))*((x[1]-c)^2))*(a^b*x[2]^(b-1)*exp(-a*x[2]))/6}
```

```
> adaptIntegrate(B, lowerLimit = c(0, 0.001), upperLimit = c(100, 10000))
```

```
integral
```

```
[1]0.9487632
```

Programme 10

```
> a = 1; b = 4; c = 4; d = 6
```

```
> l = function(x){x^(a-1)*exp(-x)}
```

```
> integrate(l, lower = 0.001, upper = 10000)
```

```
0.9990005
```

```
> B = function(x){(1/(sqrt(2*d*pi)))*exp((-1/(2*d))*((x[1]-c)^2))*(b^a*(1/x[2])^(a-1)*exp(-b/x[2]))/0.9990005}
```

```
> adaptIntegrate(B, lowerLimit = c(0, 0.001), upperLimit = c(100, 10000))
```

```
integral
```

```
[1]37863.37
```

Programme 11

```
> a = 1; b = 4
```

```
> h = function(x){x^(b-1)*exp(-x)}
```

```
> integrate(h, lower = 0.001, upper = 10000)
```

```
6
```

```

> B = function(x){(1/(pi * (1 + x[1]^2))) * (a^b * x[2]^(b - 1) * exp(-a * x[2]))/6}
> adaptIntegrate(B, lowerLimit = c(0, 0.001), upperLimit = c(100, 10000))
integral
[1]0.4968169

```

Programme 12

```

> a = 1; b = 4
> l = function(x){x^(a - 1) * exp(-x)}
> integrate(l, lower = 0.001, upper = 10000)
0.9990005
> B = function(x){(1/(pi*(1+x[1]^2)))*(b^a*(1/x[2])^(a-1)*exp(-b/x[2]))/0.9990005}
> adaptIntegrate(B, lowerLimit = c(0, 0.001), upperLimit = c(100, 10000))
integral
[1]19826.95

```

Programme 13

```

> a = 1; b = 1; c = 100; d = 10000
> I = function(x){(x^(c - 1)) * (1 - x)^(d - 1)}
> integrate(I, lower = 0.001, upper = 1)
7.454345e - 246
> h = function(x)x^(b - 1) * exp(-x)
> integrate(h, lower = 0.001, upper = 10000)
0.9990005
> B = function(x){((x[1]^(c-1)*(1-x[1])^(d-1))/7.454345e-246)*((a^b*x[2]^(b-1)
*exp(-a * x[2]))/0.9990005)}
> adaptIntegrate(B, lowerLimit = c(0.001, 0.001), upperLimit = c(1, 10000))
integral
[1]7.644127

```

Programme 14

```

> a = 1; b = 1; c = 0; d = 1
> l = function(x){x^(a - 1) * exp(-x)}

```

```

> integrate(l, lower = 0.001, upper = 10000)
0.9990005
> B = function(x){(1/(sqrt(2 * d * pi))) * exp((-1/(2 * d)) * ((x[1] - c)^2)) * (b^a *
(1/x[2])^(a - 1) * exp(-b/x[2]))/0.9990005}
> adaptIntegrate(B, lowerLimit = c(0, 0.001), upperLimit = c(100, 10000))
integral
[1]5000.221

```

Programme 15

```

> m = 0; a = 1; b = 4; x = 3
> probH0 = pgamma(0.5, a, b)
> probH1 = 1 - probH0
> prior.odds = probH0/probH1
> prior.odds
[1]6.389056
> posterioriH0 = pgamma(0.5, a + 0.5, b + ((m - x)^2)/2)
> posterioriH1 = 1 - posterioriH0
> posteriori.odds = posterioriH0/posterioriH1
> posteriori.odds
[1]26.22339
> BF = posteriori.odds/prior.odds
> BF
[1]4.104424

```

Programme 16

```

> a = 0; b = 1; c = 1; d = 1; x = 1
> probH0 = pnorm(0.5, c, d)
> probH1 = 1 - probH0
> prior.odds = probH0/probH1
> prior.odds
[1]0.4462101
> posterioriH0 = pnorm(0.5, (1/(b + d)) * (b * c + d * x), (b * d)/(b + d))
> posterioriH1 = 1 - posterioriH0

```

```
> posteriori.odds = posterioriH0/posterioriH1
```

```
> posteriori.odds
```

```
[1]0.1885734
```

```
> BF = posteriori.odds/prior.odds
```

```
> BF
```

```
[1]0.4226113
```

Bibliographie

Bibliographie

- [1] **christian P.Robert** :*le choix bayésien (principes et pratique)* Springer Verlag France,Paris,2006,la deuxième édition.
- [2] **Jean-Christophe Breton** :*Master IMA 2ème année*, Université de la Rochelle, Septembre-Décembre 2006.
- [3] **Jerôme Dupuis** (LSP,UPS).(2007). *statistique bayésienne et algorithme MCMC*.
- [4] **Abdi Zeyneb** :*Les tests bayésiens dans les essais cliniques*, Université de Mentouri, Constantine, Département de mathématiques, École Doctorale, Pole de Constantine, Mémoire de magister en mathématiques Option probabilités-statistique.
- [5] **Judith Rousseau** :*Statistique Bayésienne (Notes de Cours)*, Ensae Paris Tech, (2009,2010).
- [6] **Mme Halil Née Khorsi Rachida** :*Inférence Bayésienne en series chronologiques*, Université Mouloud Mammeri, Tiziouzou, Departement de mathématique, Mémoire de magister, soutenu le 10-03-2011.
- [7] **Lauzetto Durand** :*Approches statistiques asymptotique décision et application quant à l'utilisation de l'incertain d'un modèle*, Université de strasbourg UFR Mathématique-Informatique 29 Aôt 2013
- [8] **Handout-08** :*Introduction to probability*, Use R to compute Numerical Integrals 2 Mouvement 2011.
- [9] **TP 5 R** :*Intégration numérique*, L1 Biologie 2013-2014.
- [10] **Gilbert saporta** :*chaire de statistique Appliquée et cerdic CNAM*
- [11] **Article** : **Guillaume Damblin, Merlin Keller, Pierre Barbillon, Alberto Pasanisi, Eric Parent** : *Le facteur de Bayes appliqué à la validation des codes de calcul*,AgroParisTech, 16 rue Claude Bernard 75 005 Paris.

-
- [12] **Rebbet Sarra, Boufermel Hala** : *La régression Bayésienne et Application*, Université Mohamade Seddik Ben Yahia-Jijel, Département de mathématiques, Mémoire de master en mathématiques Option probabilités-statistique, année 2017