

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE  
MINISTERE DE L'ENSEIGNEMENT SUPERIEURE ET DE LA RECHERCHE  
SCIENTIFIQUE**

**UNIVERSITE DE MENTOURI - CONSTANTINE  
FACULTE DES SCIENCES-EXACTES  
DEPARTEMENT DE MATHEMATIQUES  
ECOLE DOCTORALE – POLE DE CONSTANTINE**

N° d'ordre : .....  
N° de série : .....

## **MEMOIRE**

*Présenté pour l'obtention du diplôme de :*  
**MAGISTER EN MATHEMATIQUES**

**OPTION  
PROBABILITES-STATISTIQUES**

**THEME**

# **LES TESTS BAYESIENS DANS LES ESSAIS CLINIQUES**

*Par :*

**ABDI ZEYNEB**

*Devant le jury :*

<i>Président</i>	<b>: N.BENKAFADAR</b>	<b>Prof.Uni Mentouri Constantine</b>
<i>Rapporteur</i>	<b>: H.MERABET</b>	<b>MC.Uni Mentouri Constantine</b>
<i>Examineur</i>	<b>: F.RAHMANI</b>	<b>MC.Uni Mentouri Constantine</b>
<i>Examineur</i>	<b>: D.BOUDAA</b>	<b>MC.Uni Mentouri Constantine</b>

*Soutenu le :*

# Remerciement

*Je voudrais tout d'abord remercier chaleureusement mon directeur de thèse Madame MERABET, maître de conférence, Université de Mentouri de Constantine, d'avoir encadré ce travail de thèse, avec beaucoup de compétence, d'enthousiasme et de disponibilité. La confiance que vous m'avez accordée ainsi que nos nombreuses discussions m'ont permis de progresser et d'acquérir un esprit de synthèse. Je lui exprime ma profonde gratitude pour sa gentillesse, son soutien et sa patience.*

*Je suis également très honoré que Monsieur N.BENKAFADAR, Professeur, Université Mentouri de Constantine soit le président du jury de mon mémoire.*

*Mes remerciements s'adressent également à Monsieur D. BOUDAA maître de conférence à l'université de Constantine et Monsieur F. RAHMANI, Maître de conférence à l'université de Constantine d'avoir accepté de juger ce travail et d'en être les examinateurs.*

*Pour fini, je souhaite exprimer mes remerciements à : Ma famille surtout mes parents pour leurs intérêts envers mon travail et pour leurs aide ; Mon mari, pour son soutien morale, et son amour ; Mon amie Nawal, pour son soutien moral et sa présence dans bons et les mauvais moments.*

*Zeyneb*

# Table des matières

<b>Introduction</b>	<b>iv</b>
<b>1 Introduction à l'approche bayésienne</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Définition des modèles . . . . .	3
1.2.1 Modèle statistique paramétrique . . . . .	3
1.2.2 Modèle statistique bayésien . . . . .	3
1.3 Le problème de choix de l'a priori . . . . .	4
1.4 La loi a priori et a posteriori . . . . .	4
1.4.1 La loi a priori . . . . .	5
1.4.2 La loi a posteriori . . . . .	6
1.5 Lois a priori conjuguées . . . . .	12
1.5.1 Famille exponentielle . . . . .	13
1.5.2 Lois conjuguées des familles exponentielles . . . . .	13
1.6 Lois a priori non informatives . . . . .	14
1.6.1 La loi a priori de Jeffreys . . . . .	14
1.6.2 L'analyse avec la distribution a priori non informative . . . . .	16
1.7 Incorporation de l'information a priori en pratique . . . . .	18
1.8 Le poids de l'a priori dans la réponse bayésienne . . . . .	19
1.9 Etude de la sensibilité de la réponse bayésienne à la loi a priori . . . . .	20
1.10 La prédiction . . . . .	20
1.11 Méthodes usuelles de calcul bayésien . . . . .	23
1.11.1 Méthodes de simulation de Monte Carlo . . . . .	24
1.11.2 Méthode de Monte-Carlo par chaines de Markov . . . . .	26
<b>2 Les tests bayésiens</b>	<b>29</b>
2.1 Introduction . . . . .	29
2.2 L'approche standard . . . . .	30

2.3	La fonction d'utilité . . . . .	30
2.4	Une première approche de la théorie des tests . . . . .	31
2.4.1	Le facteur de Bayes . . . . .	33
2.5	Comparaison avec la théorie classique . . . . .	42
2.5.1	Test UMP et UMPU : . . . . .	42
2.5.2	La p-valeur . . . . .	44
2.6	Une seconde approche décisionnelle . . . . .	45
2.6.1	Equivalence de la problématique . . . . .	45
2.7	Les régions de crédibilités . . . . .	47
2.8	La considération de la fonction de coût . . . . .	48
2.9	Deux optimalités : minimaxité et admissibilité . . . . .	50
2.9.1	La minimaxité . . . . .	50
2.9.2	Le théorème de classe complète sous $L_2$ . . . . .	51
2.9.3	La considération d'admissibilité sous le coût $L_2$ . . . . .	54
2.9.4	L'admissibilité dans le problème bilatéral . . . . .	56
<b>3</b>	<b>Méthodes adaptatives bayésiennes dans les essais cliniques</b> . . . . .	<b>59</b>
3.1	Introduction . . . . .	59
3.2	Les essais cliniques . . . . .	60
3.3	Les différentes phases d'un essai clinique . . . . .	60
3.4	L'adaptativité dans les essais cliniques . . . . .	61
3.5	La comparaison entre l'approche bayésienne et fréquentiste dans les essais cliniques . . . . .	63
3.6	Règle d'arrêt . . . . .	65
3.6.1	Principe de la règle d'arrêt . . . . .	65
3.6.2	Situation conduisant à un arrêt prématuré . . . . .	65
3.7	Les méthodes du modèle d'adaptation dans les essais cliniques . . . . .	66
3.7.1	Quelques notions de base de l'approche bayésienne dans les essais cliniques . . . . .	66
3.8	Design à plusieurs étapes pour un seul bras de l'essai . . . . .	71
3.8.1	L'approche classique pour les deux étapes de conception . . . . .	71
3.8.2	L'approche bayésienne . . . . .	72
3.9	Plans adaptatifs bayésiens optimaux . . . . .	74
3.10	La conformité d'adaptation bayésienne dans les essais cliniques . . . . .	79
3.10.1	La base de l'essai de conformité . . . . .	79
3.11	Le principe des tests diagnostiques . . . . .	80
3.12	La randomisation . . . . .	84
3.12.1	Principe de la randomisation adaptative . . . . .	85

3.12.2 Conclusion . . . . .	86
<b>4 Applications</b>	<b>88</b>
4.1 Le calcul de la probabilité prédictive . . . . .	88
4.2 Le test des paramètres avec le facteur de Bayes . . . . .	91
4.3 La comparaison de la mortalité du groupe de traitement avec le groupe de contrôle correspondant . . . . .	92
4.4 La détermination de la puissance d'un échantillon binomial et la fonction puissance . . . . .	99
4.4.1 La détermination de la fonction puissance pour $H_0$ contre $H_1$ . . . . .	100
4.4.2 La détermination de la puissance pour un échantillon binomial avec l'information a priori . . . . .	101
4.5 Le calcul de la puissance pour deux échantillons binomial . . .	102
4.5.1 La comparaison de deux populations binomial . . . . .	103
4.5.2 Le calcul de la puissance pour deux échantillons binomial	104

# Introduction

L'étude des tests bayésiens a fait l'objet de plusieurs livres et articles récents [Lecoutre (2006), Robert (2007), Bernardo (2010), Shein-Chung Chow (2010)]. L'approche bayésienne offre plus de souplesse à l'analyse des données expérimentales. Dans les domaines expérimentaux, et en particulier dans la recherche médicale, la nécessité de changements dans la présentation des résultats expérimentaux est de plus en plus mise en avant par les "éditeurs" qui demandent aux auteurs de fournir des indicateurs de la taille des effets "effect size" et leurs intervalles d'estimation "interval estimates", en plus ou à la place des tests de signification traditionnels. Les procédures bayésiennes prédictives donnent aux utilisateurs une méthode très séduisante pour répondre à des questions essentielles telles que : "quelle devrait être la taille de l'expérience pour avoir des chances raisonnables de démontrer une conclusion donnée?", "sur la base des données disponibles quelle sont les chances que le résultat final conduise à la même conclusion, ou au contraire ne permette pas de conclure?", alors que les pratiques fréquentistes traditionnelles ne traitent pas ces questions, les probabilités prédictives leur apportent une réponse directe et naturelle.

Ce mémoire est divisé en quatre chapitre, le **premier chapitre** a pour objet d'introduire les notions indispensables à l'inférence bayésienne, aux méthodes prédictives, et au calcul bayésien notamment les méthodes de simulation de Monte-Carlo. Nous avons présentés ces concepts et les avons illustrés par des exemples concrets.

Toute la théorie mathématique des tests est envisagée au **deuxième chapitre**, on a d'abord considéré une première approche bayésienne standard des tests, qui repose sur une évaluation des décisions par des coûts 0-1 et on a proposé par la suite une alternative à l'approche décisionnelle fondée sur des coûts plus adaptés, qui mettent en avant l'évaluation ex post pour des procédures de tests. Par opposition aux procédures de Neymann-Pearson

---

pour lesquelles l'évaluation fonctionne dans un esprit ex ante.

Il apparaît naturel de considérer au **troisième chapitre** des méthodes adaptatives bayésiennes dans les essais cliniques, qui visent à attribuer à plus de sujets le traitement le plus efficace, avec une présentation de principe général des plans adaptatifs qui consistent à attribuer les traitements aux patients, en fonction des observations précédentes. En modifiant les probabilités d'attribution des traitements, l'objectif est de réduire le nombre de patients recevant le (ou les) traitement(s) le (ou les) moins efficace (s) et augmenter au contraire le nombre de patients recevant le (ou les) traitement (s) le (s) plus efficace (s). On fait aussi des tests précisément dans les essais cliniques avec des conclusions et des résultats logiques et légitimes. Puis on assure la conformité d'adaptation bayésienne dans les essais cliniques. Et on termine ce chapitre par le principe et le rôle de la randomisation adaptative.

Le **quatrième chapitre** présente quelques applications numériques, en envisageant le calcul du facteur de Bayes, le calcul de la probabilité prédictive et la puissance prédictive à l'aide des méthodes de Monte Carlo par chaîne de Markov. Nous avons employé le minitab R et compté sur WinBUGS dans les simulations.





# Chapitre 1

## Introduction à l'approche bayésienne

### 1.1 Introduction

L'approche bayésienne peut être présentée comme une généralisation de l'approche 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, une inférence au sujet de la loi générant ces observations ; afin, soit d'analyser un phénomène passé, soit de prévoir un événement futur, en nous attachant tout particulièrement aux aspects décisionnels de l'inférence bayésienne.

L'inférence bayésienne s'accompagne d'une modélisation probabiliste du phénomène observé qui est nécessaire, car elle permet de replacer le phénomène dans une globalité qui autorise des analyses et des généralisations.

Dans ce chapitre, nous allons donner quelques notions de bases qui concernent la statistique bayésienne (la loi a priori, la loi a posteriori, la loi prédictive ...) avec des illustrations.

On note  $\theta$  la probabilité qu'un individu atteint d'une maladie M soit guéri à l'issu d'un traitement T (consistant en l'administration d'un certain médicament). Pour estimer le paramètre  $\theta$  inconnu, on administre à  $n$  patients le traitement T. On note  $x_i \in \{0, 1\}$  l'état du patient  $i$  suite à l'administration du traitement T ;  $x_i = 1$  si le patient  $i$  est guéri et  $x_i = 0$  sinon. On considère par ailleurs que le médicament (associé au traitement T) ne peut être commercialisé que si  $\theta \geq 0.8$ . D'un point de vue statistique on peut formuler

le problème à l'aide du test suivant :

$$H_0 : \theta \leq 0.8 \quad \text{contre} \quad H_1 : \theta \geq 0.8$$

D'un point de vue statistique on se place dans le cadre du modèle d'échantillonnage. Autrement dit on suppose que les  $x_i$  sont des réalisations de variables aléatoires  $X_i$  indépendantes et de même loi (on dit que les  $X_i$  sont *i.i.d.*). Chaque  $X_i$  suit une loi de Bernoulli de paramètre  $\theta$ . Autrement dit on a :  $\Pr(X_i = 1) = \theta$ , et du fait que les variables aléatoires sont *i.i.d.*, la vraisemblance s'écrit :

$$L(\theta; x_1, \dots, x_i, \dots, x_n) = \theta^s (1 - \theta)^{n-s} \quad \text{où} \quad s = \sum_{i=1}^{i=n} x_i$$

où  $s = \sum_{i=1}^{i=n} x_i$  représente le nombre de patients guéris. Par ailleurs, on suppose que l'on dispose d'information sur le paramètre  $\theta$  à estimer (et ce, en plus de l'information déjà apportée par les  $x_i$ ). Cette information (qu'on appellera information a priori) peut, par exemple provenir d'une autre étude portant sur l'efficacité du traitement T et réalisée sur des patients issus, au sens statistique, de la même population (autrement dit la probabilité de guérison de tels patients est égale à  $\theta$ ). Une telle information consiste typiquement alors en une estimation  $\tilde{\theta}$  de  $\theta$  et un intervalle de confiance (observé) I de niveau de confiance 0.95 . Cette information peut être également fournie par un Expert on entend par Expert toute personne compétente sur le sujet et reconnue comme tel par ses pairs. Cette information consiste à nouveau en une estimation de  $\theta$ , notée  $\theta^*$ , et tel que la probabilité  $\Pr(\theta \in J)$  soit élevée (0.90 ou 0.95 en pratique). A noter qu'il conviendra de donner un sens à cette dernière écriture (celle-ci n'ayant pas de sens en statistique classique).

Le problème statistique consiste alors à produire une inférence sur  $\theta$  (estimation ou test) qui repose à la fois sur les observations (c.à.d sur les  $x_i$ ) et sur l'information a priori dont on dispose. L'objet de l'approche 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  $\theta$  uniquement sur les observations (c.à.d sur les  $x_i$ ) ; au cas particulier l'estimation retenue serait la moyenne des  $x_i$ .

## 1.2 Définition des modèles

### 1.2.1 Modèle statistique paramétrique

On considère le modèle paramétrique et on suppose que les observations sur lesquelles l'analyse statistique est fondée,  $x_1, \dots, x_n$ , proviennent de lois de probabilité paramétrées, ie, si  $x_i$  suit une densité  $f_i(x_i/\theta)$  sur  $\mathbb{R}^p$  dont le paramètre  $\theta_i$ , scalaire ou vectoriel est inconnu.

On voit alors que ce modèle se représente plus simplement sous la forme  $x \sim f(x/\theta)$ , où  $x$  est le vecteur des observations et  $\Theta$  l'ensemble des paramètres  $\theta_1, \dots, \theta_n$ , éventuellement tous égaux. Ceci permet de représenter de manière unifiant observations isolées et échantillons  $x_1, \dots, x_n$ , d'une même loi  $f(x_1/\theta_1)$ . Dans le second cas, on a  $x = (x_1, \dots, x_n)$  et

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

Notons à ce point que nous écrivons de manière identique les densités de variables aléatoires continues et discrètes.

### 1.2.2 Modèle statistique bayésien

Comparée à l'approche probabiliste, la démarche statistique est fondamentalement une démarche d'inversion, car elle se propose de remonter, même imparfaitement, des effets (observations) aux causes (paramètres du mécanisme générateur).

En d'autres termes, ayant observé des variables aléatoires générées suivant un mécanisme dirigé par le paramètre, les méthodes statistiques permettent de conduire à partir de ces observations à une inférence, tandis que le modèle probabiliste caractérise le comportement des observations, conditionnellement à  $\theta$ . Une telle inversion est flagrante dans la notion même de vraisemblance, puisque, formellement, elle ne fait que réécrire la densité dans le bon ordre,

$$l(\theta/x) = f(x/\theta)$$

Comme une fonction de  $\theta$  inconnu dépendant de  $x$  observé. De manière générale, l'inversion des probabilités est décrite par le théorème de Bayes : Si  $A$  et  $E$  sont des événements tels que  $P(E) \neq 0$ ,

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

où  $A^c$  désigne le complémentaire de  $A$ .

Bayes utilisait la version continue de ce théorème, qui énonce que, pour une loi (dite a priori)  $\pi$  sur le paramètre  $\theta$  et une observation  $x$  de densité  $f(x/\theta)$ , la distribution de  $\theta$  conditionnellement à  $x$  a pour densité

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

La loi de  $\theta$  conditionnellement à  $x$  est proportionnelle à la loi de  $x$  conditionnellement à  $\theta$  multipliée par la loi marginale de  $\theta$ .

**Définition 1.2.1** *on appelle modèle statistique bayésien sur  $(\Omega, F, P)$  la donnée d'un modèle statistique paramétré  $f(x/\theta)$ , et d'une loi a priori sur les paramètres  $\pi(\theta)$ .*

### 1.3 Le problème de choix de l'a priori

L'aspect de l'analyse bayésienne le plus critiqué et le plus délicat est certainement le choix de la loi a priori des paramètres; en particulier, le recours à des lois usuelles comme la loi normale, gamma, bêta, ne peut pas être défendu comme approche systématique plus aisée.

La détermination de la loi a priori donc est basée sur l'information a priori.

### 1.4 La loi a priori et a posteriori

L'aspect de l'analyse bayésienne, qui est à la fois le plus délicat et le plus critiqué, est certainement celui du choix de la loi a priori des paramètres.

En pratique, il est en effet très rare que l'information a priori disponible soit suffisamment précise pour pouvoir déterminer exactement cette loi, et il faut alors utiliser une approximation dont le choix peut se révéler déterminant dans l'inférence. En particulier, le recours à des lois usuelles et aux autres lois dites conjuguées ne peut être défendu comme approche systématique, car on sacrifie à un traitement mathématique plus aisé la détermination subjective de la loi a priori, donc la prise en compte de l'information a priori. Il peut cependant s'avérer utile de proposer une recherche automatique de loi a priori lorsque l'information a priori est trop coûteuse à obtenir ou peu fiable.

### 1.4.1 La loi a priori

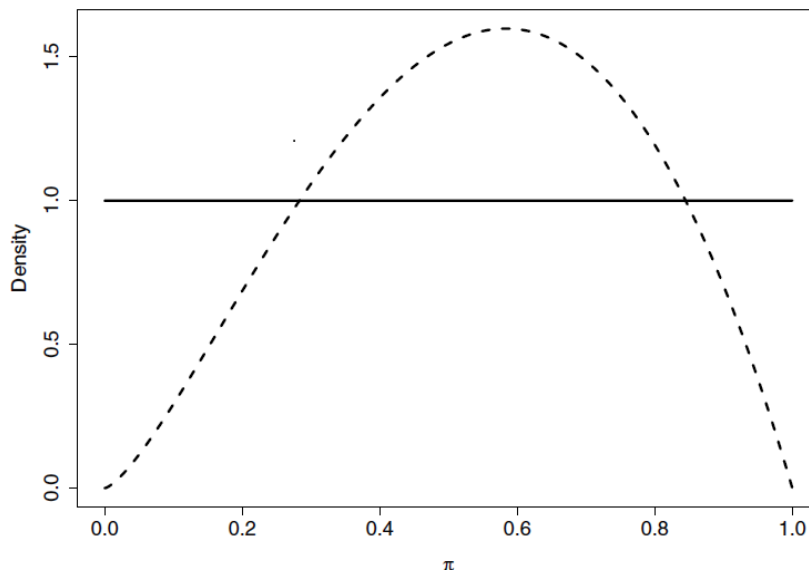
On entend par information a priori sur le paramètre  $\theta$  toute information disponible sur  $\theta$  en dehors de celle apportée par les observations.

L'information a priori sur  $\theta$  est entachée d'incertitude (si ce n'était pas le cas, le paramètre  $\theta$  serait connu avec certitude et on n'aurait pas à l'estimer). Il est naturel de modéliser cette information a priori au travers d'une loi de probabilité appelée **loi a priori**, notée  $\pi(\theta)$ .

**Exemple introductif à la loi a priori :** D'après (**Michael. Hamada et Alyson G, Wilson(2008)**). Pour illustrer le rôle de la distribution a priori, on suppose que l'information sur  $\theta$  est valable. A priori on doit donc supposer que toutes les valeurs de  $\theta$  sont entre 0 et 1. On doit numériser ce type de l'information par l'évaluation que la distribution a priori pour  $\theta$  est uniforme et qui est  $\pi(\theta) = 1, 0 < \theta < 1$ , comme illustré graphiquement dans la **figure(1, 1)**. On peut utiliser cette distribution pour compter les probabilités a priori dans chaque intervalle de  $(0, 1)$ . Pour exemple

$$\pi(\theta < 0.25) = P(\theta > 0.75) = 0.25$$

Et la proportion inconnue peut être **inférieure à 0.25** ou peut être **supérieure à 0.75**. Cette distribution a priori est un exemple de l'a priori de diffuse lorsqu'elle prend la précision de l'information a priori sur la vraie valeur de la proportion.



**Figure(1, 1) :** deux densités a priori pour proportion : la densité uniforme (ligne en gras) qui désigne l'a priori de diffuse, et la densité a priori de  $Beta(2.4, 2)$  (ligne détachée) qui désigne l'information a priori montrant que la vraie proportion est fermée à **0.55** mais non pas arbitrairement fermée à **0** et **1**.

Avec les paramètres  $\alpha = 2.4$  et  $\beta = 2$  la fonction de densité de probabilité est une distribution de beta  $Beta(\alpha, \beta)$  avec  $\alpha = 2.4$  et  $\beta = 2$ , et que :

$$\pi(\theta/\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}, \quad 0 \leq \theta \leq 1, \quad \alpha, \beta > 0$$

On note que l'importance du poids entre 0.2 et 0.8, et la distribution a comme moyenne  $\frac{\alpha}{\alpha + \beta} = 0.545$  et le mode est de  $\frac{\alpha - 1}{\alpha + \beta - 2} = 0.583$ , les distributions quand  $\theta$  appartient aux intervalles  $(0, 0.25)$  et  $(0.75, 1)$  ont pour densités a priori 0.10 et 0.2 respectivement.

Finalement on note que la distribution uniforme est aussi une distribution de Beta, mais avec les paramètres  $\alpha = \beta = 1$ .

### 1.4.2 La loi a posteriori

C'est la loi conditionnelle de  $\theta$  sachant  $x$ , sa densité notée  $\pi(\theta/x)$ , en vertu de la formule de Bayes de (1,1), et on a aussi :

a/ La loi du couple  $(\theta, x)$  : sa densité est notée  $\varphi(\theta, x)$  on a donc

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

b/ La loi marginale de  $x$  : sa densité est notée  $m(x)$ , on a donc

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

L'exemple suivant consiste à calculer la distribution a posteriori à partir des données a priori.

**exemple 1.4.1** si  $x \sim B(n, p)$  ,  $p \sim Beta(\alpha, \beta)$   
(Avec  $\alpha = \beta = 1$ , dans le cas particulier de Bayes)

$$f(x/p) = \binom{n}{x} p^x (1-p)^{n-x} ; x = 0, 1, \dots, n$$

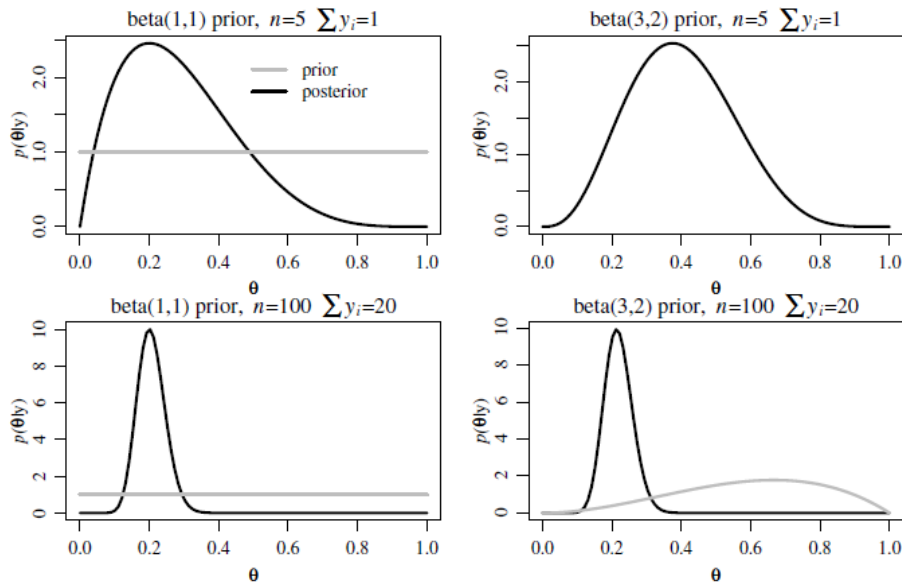
$$\pi(p) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1} ; 0 \leq p \leq 1$$

La distribution a posteriori de  $p$  est :

$$\pi(p/x) = \frac{p^{\alpha+\beta-1} (1-p)^{p+n-x-1}}{\int p^{\alpha+\beta-1} (1-p)^{p+n-x-1} dp}$$

qui est une loi de Beta ( $\alpha + x, \beta + n - x$ ) où nous appliquerons ce résultat dans le quatrième chapitre.

Voici la figure suivante, qui représente la distribution a posteriori de beta sous différentes tailles des échantillons et différentes distributions a priori



**Figure(1, 2)** :la distribution a posteriori de Beta sous deux différences de taille de l'échantillon et deux différentes distributions a priori.

**exemple 1.4.2** Le calcul de la loi a posteriori dans le modèle gaussien lorsque  $\sigma^2$  est connue. On suppose qu'on observe l'observation  $x \sim N(\theta, \sigma^2)$  avec  $\sigma^2$  connue, et que la vraisemblance est

$$f(x/\theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}, \theta \in \mathbb{R} \text{ et } \sigma > 0$$

Si la distribution a priori est spécifique comme  $\pi(\theta) = N(\theta/\mu, \sigma^2)$  avec  $(\mu, \tau^2)$  fixé, donc on obtient

$$\begin{aligned}
 \pi(\theta/x) &= \frac{f(x/\theta) \pi(\theta)}{\int_{\Theta} f(x/\theta) \pi(\theta) d\theta} \\
 &= \frac{\frac{1}{\tau\sqrt{2\pi}} \exp\left(-\frac{(\theta-\mu)^2}{2\tau^2}\right) \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right)}{\int_{-\infty}^{+\infty} \frac{1}{\tau\sqrt{2\pi}} \exp\left(-\frac{(\theta-\mu)^2}{2\tau^2}\right) \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right) d\theta} \\
 &= \frac{\frac{1}{\tau\sigma 2\pi} \exp\left[-\frac{(x-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu)^2}{2\tau^2}\right]}{\int_{-\infty}^{+\infty} \frac{1}{\tau\sigma 2\pi} \exp\left[-\frac{(x-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu)^2}{2\tau^2}\right] d\theta}
 \end{aligned}$$

on pose

$$m(x) = \int_{-\infty}^{+\infty} \frac{1}{\tau\sigma 2\pi} \exp\left[-\frac{(x-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu)^2}{2\tau^2}\right] d\theta$$

et

$$\begin{aligned}
 \frac{(x-\theta)^2}{\sigma^2} - \frac{(\theta-\mu)^2}{\tau^2} &= \frac{x^2 - 2\theta x + \theta^2}{\sigma^2} - \frac{\theta^2 - 2\theta\mu + \mu^2}{\tau^2} \\
 &= \theta^2 \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right) - 2\theta \left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right) + \frac{x^2}{\sigma^2} + \frac{\mu^2}{\tau^2} \\
 &= \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right) \left[\theta - \frac{\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}\right]^2 + \frac{x^2}{\sigma^2} + \frac{\mu^2}{\tau^2} - \frac{\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)^2}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}
 \end{aligned}$$

d'où

$$m(x) = \frac{1}{2\tau\sigma\pi} \int_{-\infty}^{+\infty} \exp\left[-\left(\frac{1}{2\sigma^2} + \frac{1}{2\tau^2}\right) \left[\theta - \frac{\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}\right]^2\right] \times \exp\left[\frac{\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)^2}{2\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)} - \frac{x^2}{\sigma^2} - \frac{\mu^2}{\tau^2}\right] d\theta$$



posons

$$\begin{aligned} t &= \left( \theta - \frac{\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}} \right) \sqrt{\frac{1}{\sigma^2} + \frac{1}{\tau^2}} \\ &= \sqrt{\frac{1}{\sigma^2} + \frac{1}{\tau^2}} d\theta \end{aligned}$$

$$m(x) = \frac{\exp \left[ \frac{\left( \frac{x}{\sigma^2} + \frac{\mu}{\tau^2} \right)^2}{2 \left( \frac{1}{\sigma^2} + \frac{1}{\tau^2} \right)} - \frac{x^2}{\sigma^2} - \frac{\mu^2}{\tau^2} \right]}{2\tau\sigma\pi\sqrt{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}} \int_{-\infty}^{+\infty} \exp -\frac{t^2}{2} dt$$

comme

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp -\frac{t^2}{2} dt = 1$$

on en déduit

$$\begin{aligned} m(x) &= \frac{\exp \left[ \frac{\left( \frac{x}{\sigma^2} + \frac{\mu}{\tau^2} \right)^2}{2 \left( \frac{1}{\sigma^2} + \frac{1}{\tau^2} \right)} - \frac{x^2}{\sigma^2} - \frac{\mu^2}{\tau^2} \right]}{\sigma\tau\sqrt{2\pi}\sqrt{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}} \\ &= \frac{1}{\sqrt{2\pi}} \frac{\exp \left[ \frac{\tau^2\sigma^2 \left( \frac{x}{\sigma^2} + \frac{\mu}{\tau^2} \right)^2}{2(\tau^2 + \sigma^2)} - \frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\tau^2} \right]}{\sqrt{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2 + \tau^2}} \exp \frac{\tau^2\sigma^2}{2(\tau^2 + \sigma^2)} \left( \frac{x^2}{\sigma^2} + \frac{\mu^2}{\tau^2} + 2\frac{x\mu}{\tau^2\sigma^2} \right) - \frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\tau^2} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2 + \tau^2}} \exp \left[ \frac{1}{2(\tau^2 + \sigma^2)} \left( \frac{x^2\tau^2}{\sigma^2} + \frac{\mu^2\sigma^2}{\tau^2} + 2x\mu - \frac{x^2}{\sigma^2} (\tau^2 + \sigma^2) - \frac{\mu^2}{\tau^2} (\tau^2 + \sigma^2) \right) \right] \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2 + \tau^2}} \exp \frac{1}{2(\tau^2 + \sigma^2)} [2x\mu - x^2 - \mu^2] \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2 + \tau^2}} \exp \left[ -\frac{1}{2(\tau^2 + \sigma^2)} (x - \mu)^2 \right] \end{aligned}$$

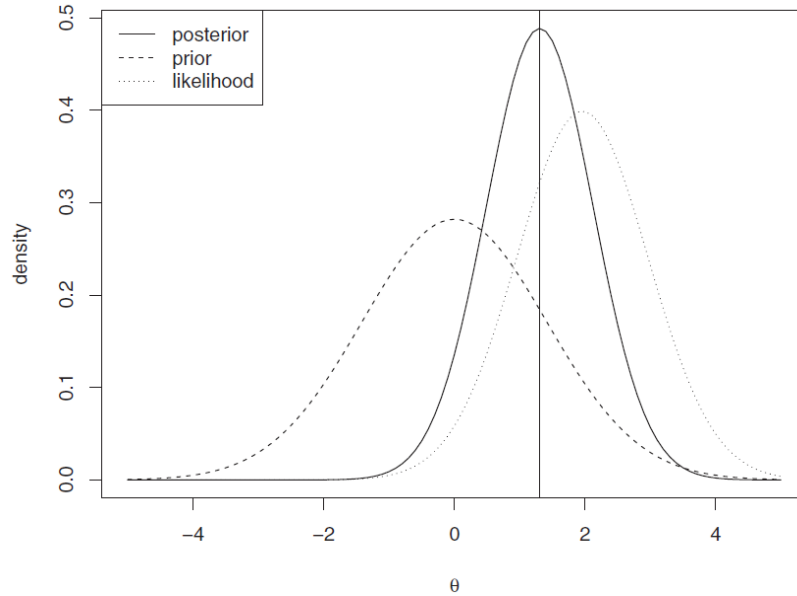
$$x \sim N(\mu, \sigma^2 + \tau^2)$$

On substitue  $m(x)$  dans la formule de  $\pi(\theta/x)$  et on trouve que :

$$\begin{aligned} \pi(\theta/x) &= \frac{\frac{1}{\tau\sigma 2\pi} \exp\left[-\frac{(x-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu)^2}{2\tau^2}\right]}{\frac{1}{\sqrt{2\pi}\sqrt{\sigma^2+\tau^2}} \exp\left[-\frac{1}{2(\tau^2+\sigma^2)}(x-\mu)^2\right]} \\ &= \frac{\sqrt{\tau^2+\sigma^2}}{\tau\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left[\frac{(\theta-\mu)^2}{\tau^2} + \frac{(x-\theta)^2}{\sigma^2} - \frac{(x-\mu)^2}{\tau^2+\sigma^2}\right]\right] \\ &= \frac{\sqrt{\tau^2+\sigma^2}}{\tau\sigma\sqrt{2\pi}} \exp\left[-\frac{(\tau^2+\sigma^2)}{2\sigma^2\tau^2}\left(\theta - \frac{\sigma^2\mu + \tau^2x}{\tau^2+\sigma^2}\right)^2\right] \end{aligned}$$

Donc

$$\theta/x \sim N\left(\frac{\sigma^2\mu + \tau^2x}{\tau^2 + \sigma^2}, \frac{\sigma^2\tau^2}{\tau^2 + \sigma^2}\right)$$



**Figure(1, 3)** : la version graphique de l'estimateur de Bayes, la ligne verticale est la moyenne de l'a posteriori égale à 1.31

L'exemple suivant présente, le calcul de la loi a posteriori lorsque  $\sigma^2$  est inconnue.

**exemple 1.4.3** *Le calcul de la loi a posteriori dans le modèle normale lorsque  $\sigma^2$  est inconnue. On considère l'extension de Normale/modèle normale*

#### 1.4. LA LOI A PRIORI ET A POSTERIORI

---

de l'exemple 3, où  $\sigma^2$  est inconnu, une autre transformation à la précision  $h = \frac{1}{\sigma^2}$ , avec  $h \sim \text{Gamma}(\alpha, \beta)$  et

$$\pi(h) = \frac{\beta^\alpha}{\Gamma(\alpha)} h^{\alpha-1} \exp -h\beta \quad , h > 0$$

lorsque la vraisemblance pour chaque observation  $x_i$  est :

$$f(x_i/\theta, h) = \frac{h^{\frac{1}{2}}}{\sqrt{2\pi}} \exp\left\{-\frac{h}{2}(x_i - \theta)^2\right\}$$

l'a posteriori de  $h$  est proportionnelle au produit de la vraisemblance et l'a priori

$$\begin{aligned} \pi(h/X, \theta) &\propto \left[ \prod_{i=1}^{i=n} f(x_i/\theta, h) \right] \pi(h) \\ &\propto h^{\frac{n}{2}} h^{\alpha-1} \exp\left\{-\frac{h}{2} \sum_{i=1}^{i=n} (x_i - \theta)^2\right\} \times \exp\{-h\beta\} \\ &\propto h^{\frac{n}{2} + \alpha - 1} \exp -h \left[ \beta + \frac{1}{2} \sum_{i=1}^{i=n} (x_i - \theta)^2 \right] \end{aligned}$$

donc

$$h/X, \theta \sim \text{Gamma} \left( \frac{n}{2} + \alpha, \beta + \frac{1}{2} \sum_{i=1}^{i=n} (x_i - \theta)^2 \right)$$

**exemple 1.4.4** le calcul de la loi a posteriori lorsque  $\theta$  et  $\sigma^2$  sont inconnus, on considère la fonction de vraisemblance normale suivante

$$f(x/\theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \theta)^2}{2\sigma^2}\right) \quad , x \in \mathbb{R}, \theta \in \mathbb{R}$$

et en supposant que  $\theta$  et  $\sigma^2$  sont inconnus, de sorte que  $\boldsymbol{\theta} = (\theta, \sigma^2)$ . Et

$$f(\boldsymbol{\theta}) \propto \exp\left(-(\boldsymbol{\theta} - \boldsymbol{a})^2 / b\right)$$

avec l'a priori :

$$\pi_1(\boldsymbol{\theta}) = N(\boldsymbol{\theta}/\boldsymbol{\mu}, \boldsymbol{\tau}^2)$$

et on suppose que :

$$f(\sigma) \propto (\sigma^2)^{-a} \exp -\frac{b}{\sigma^2}$$

En effet, c'est la Gamma inverse : Si  $Y \sim G(\alpha, \beta)$  donc  $X = \frac{1}{Y} \sim \text{Inverse Gamma}(\alpha, \beta) \equiv IG(\alpha, \beta)$  avec la fonction de densité

$$g(x/\alpha, \beta) = \frac{e^{-1/\beta x}}{\Gamma(\alpha) \beta^\alpha x^{\alpha+1}}, \quad x > 0, \alpha > 0, \beta > 0$$

avec l'a priori :

$$\pi_2(\sigma^2) = IG\left(\frac{\sigma^2}{a}, b\right)$$

Et  $a, b$  sont des paramètres, puis on suppose que  $\theta$  et  $\sigma^2$  sont indépendants a priori, donc :

$$\pi(\boldsymbol{\theta}) = \pi_1(\theta) \pi_2(\sigma^2)$$

Donc les distributions a posteriori sont

$$\pi(\theta/x) \propto N\left(\theta / \frac{\sigma^2 \mu + \tau^2 x}{\tau^2 + \sigma^2}, \frac{\sigma^2 \tau^2}{\tau^2 + \sigma^2}\right)$$

et :

$$\pi(\sigma^2/x, \theta) \propto IG\left(\sigma^2/\frac{1}{2} + \alpha, \left[\frac{1}{2}(x - \theta)^2 + \frac{1}{b}\right]^{-1}\right)$$

## 1.5 Lois a priori conjuguées

Une famille conjuguée est une famille de lois liée au processus étudié et qui est particulièrement intéressante dans des modèles paramétriques où des statistiques exhaustives qui peuvent être définies. Il s'agit d'une famille de lois, c-à-d d'un ensemble de lois, qui soit d'une part, assez riche et flexible pour représenter l'information a priori que l'on a sur les paramètres du modèle et, d'autre part, qui se combine remarquablement avec la fonction de vraisemblance dans la formule de Bayes, pour donner a posteriori une loi sur les paramètres du modèle qui appartienne à la même famille.

On verra que le choix de cette famille se déduit à partir du noyau de la vraisemblance considérée comme fonction des paramètres.

**Définition 1.5.1** *une famille  $F$  de lois, sur  $\Theta$  est dite conjuguée si, pour tout  $\pi \in F$ , la loi a posteriori  $\pi(\theta/x)$  appartient également à  $F$ .*

Un **exemple** trivial de famille conjuguée est bien sûr  $F_0$ , ensemble de toutes les lois de probabilité sur  $\Theta$ , qui n'est d'aucune utilité dans le choix d'une loi a priori.

En fait, on essaie de prendre  $F$  aussi petite que possible et paramétrée. Par exemple, on pourrait chercher à définir la famille conjuguée minimale comme l'intersection qui est en général vide. Lorsque  $F$  est paramétrée, le passage des lois a priori aux lois a posteriori se réduit à un changement de paramètres, puisque les lois a posteriori sont alors toujours calculables.

### 1.5.1 Famille exponentielle

Un type particulier de lois de probabilité permet une détermination directe des familles de lois conjuguées. De telles lois sont dites familles exponentielles.

**Définition 1.5.2** *soient  $\mu$ , mesure  $\sigma$  – finie sur  $X$ , et  $\Theta$  l'espace des paramètres. On définit  $C$  et  $H$ , respectivement fonction de  $X$  et  $\Theta$  dans  $\mathbb{R}_+$ , et  $R$  et  $T$ , fonction de  $\Theta$  et  $X$  dans  $\mathbb{R}^k$ . La famille de distributions de densité (par rapport à  $\mu$ )*

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

*est dite famille exponentielle de dimension  $k$ . Dans le cas particulier où  $\Theta \in \mathbb{R}^k$ ,  $X \subset \mathbb{R}^k$  et :*

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

*La famille est dite **naturelle**.*

### 1.5.2 Lois conjuguées des familles exponentielles

Soit

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

appartenant à une famille exponentielle. Alors cette loi admet une famille conjuguée comme :

$$\pi(\theta/\mu, \lambda) = K(\mu, \lambda) \exp\{\theta\mu - \lambda\psi(x)\}$$

Où  $K(\mu, \lambda)$  est la constante de normalisation de la densité. Donc la densité a posteriori est :

$$\begin{aligned}
 \pi(\theta/\mu + x, \lambda + 1) &= \frac{h(x) \exp\{\theta x - \psi(\theta)\} K(\mu, \lambda) \exp\{\theta\mu - \lambda\psi(\theta)\}}{\int h(x) \exp\{\theta x - \psi(\theta)\} K(\mu, \lambda) \exp\{\theta\mu - \lambda\psi(\theta)\} d\theta} \\
 &= \frac{\exp\{\theta x - \psi(\theta)\} \exp\{\theta\mu - \lambda\psi(\theta)\}}{\int \exp\{\theta x - \psi(\theta)\} \exp\{\theta\mu - \lambda\psi(\theta)\} d\theta} \\
 &= K(\theta/\mu + x, \lambda + 1) \exp\{\theta x - \psi(\theta)\} \exp\{\theta\mu - \lambda\psi(\theta)\} \\
 &= K(\theta/\mu + x, \lambda + 1) \exp\{\theta(x + \mu) - \psi(\theta)(1 + \lambda)\}
 \end{aligned}$$

On peut mettre comme exemple de lois conjuguées des familles exponentielles le calcul de la loi a posteriori lorsque  $\sigma^2$  est inconnu.

Dans la partie suivante, quand l'information a priori n'est pas disponible sur le modèle, les lois a priori doivent être construites à partir de la distribution d'échantillonnage. Elles sont appelées les lois non informatives.

## 1.6 Lois a priori non informatives

La section précédente a montré que les lois conjuguées peuvent être utilisées comme approximations des véritables lois a priori par contre lorsqu'aucune information n'est disponible sur le modèle, leur utilisation n'est justifiée que par des considérations analytiques.

Dans de telles situations, il est impossible de bâtir une distribution a priori sur des considérations subjectives. On peut alors chercher à utiliser malgré tout des techniques bayésiennes qui intègrent notre ignorance sur les paramètres du modèle, de telles méthodes sont appelées de manière évidente, **noninformatives**.

### 1.6.1 La loi a priori de Jeffreys

Jeffreys (1946, 1961) propose une approche intrinsèque qui évite effectivement le besoin de prendre en compte une structure d'invariance potentielle, tout en étant souvent compatible lorsque cette structure existe, les lois a

priori non informatives de **Jeffreys** 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 unidimensionnel, sous certaines conditions de régularité, cette information est aussi égale à

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

La loi a priori de Jeffreys est :

$$\pi^*(\theta) \propto I^{\frac{1}{2}}(\theta)$$

Elle est définie à un coefficient de normalisation quand  $\pi^*$  est propre.

**exemple 1.6.1** si  $x \sim B(n, p)$

$$f(x/p) = \binom{n}{x} p^x (1-p)^{n-x}$$

donc

$$I(P) = n \left[ \frac{1}{p} + \frac{1}{1-p} \right] = \frac{n}{p(1-p)}$$

donc la loi de Jeffreys pour ce modèle est :

$$\pi^*(p) \propto [p(1-p)]^{-\frac{1}{2}}$$

et est alors propre, car il s'agit de la distribution Beta  $(\frac{1}{2}, \frac{1}{2})$ .

Dans le cas où  $\theta$  est un paramètre multidimensionnel, on définit la matrice d'information de Fisher, pour  $\theta \in \mathbb{R}^k$ , et  $I(\theta)$  aux éléments suivants :

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

et la loi non informative de Jeffreys est alors définie par :  $\pi^*(\theta) \propto [\det(I(\theta))]^{\frac{1}{2}}$ .

### 1.6.2 L'analyse avec la distribution a priori non informative

La distribution a priori pour le modèle de moyenne et de variance inconnues pour les données de distribution normale est proportionnelle à :

$$\pi(\mu, \sigma^2) \propto \frac{1}{\sigma^2} \quad (1.2)$$

La fonction de densité de probabilité pour la distribution normale de variable aléatoire est :

$$f(x/\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x - \mu)^2\right]$$

où  $\mu$  est la moyenne et  $\sigma^2$  est la variance.

avec

$$f(x_1, \dots, x_n/\mu, \sigma^2) = \prod_{i=1}^{i=n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x_i - \mu)^2\right] \quad (1.3)$$

La fonction de vraisemblance de l'équation (1,3) et la distribution a priori donnée dans l'équation (1,2) conduisent à la distribution a posteriori jointe de  $(\mu, \sigma^2)$  qui est proportionnelle à :

$$\pi(\mu, \sigma^2/X) \propto (\sigma^2)^{-\frac{n}{2}-1} \exp\left[\sum_{i=1}^{i=n} -\frac{1}{2\sigma^2}(x_i - \mu)^2\right]$$

avec  $X = (x_1, \dots, x_n)$ , qui est une fonction de densité propre.

On suppose que  $\mu$  soit le paramètre d'intérêt, pour obtenir la distribution a posteriori marginale de  $\mu$ , on doit intégrer par rapport à  $\sigma^2$  donc :

$$\pi(\mu/X) \propto \int_0^{+\infty} (\sigma^2)^{-\frac{n}{2}-1} \exp\left[\sum_{i=1}^{i=n} -\frac{1}{2\sigma^2}(x_i - \mu)^2\right] d\sigma^2$$



donc on peut écrire l'équation comme suit :

$$\begin{aligned} \pi(\mu/X) &\propto \frac{\Gamma\left(\frac{n}{2}\right)}{\left(\sum_{i=1}^{i=n} -\frac{(x_i-\mu)^2}{2}\right)^{\frac{n}{2}}} \int_0^{+\infty} \frac{\left(\sum_{i=1}^{i=n} -\frac{(x_i-\mu)^2}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} (\sigma^2)^{-\frac{n}{2}-1} \\ &\quad \times \exp\left[\sum_{i=1}^{i=n} -\frac{1}{2\sigma^2} (x_i - \mu)^2\right] d\sigma^2 \\ &= \frac{\Gamma\left(\frac{n}{2}\right)}{\left(\sum_{i=1}^{i=n} -\frac{(x_i-\mu)^2}{2}\right)^{\frac{n}{2}}} \end{aligned}$$

on note que :

$$\sum_{i=1}^{i=n} (x_i - \mu)^2 = n(\bar{x} - \mu)^2 + \sum_{i=1}^{i=n} (x_i - \bar{x})^2$$

donc :

$$\pi(\mu/X) \propto \left[1 + \frac{(\mu - \bar{x})^2}{(n-1)\frac{s^2}{n}}\right]^{-\frac{n}{2}}$$

Où  $\bar{x} = \frac{1}{n} \sum_{i=1}^{i=n} x_i$  et  $s^2 = \frac{\sum_{i=1}^{i=n} (x_i - \bar{x})^2}{(n-1)}$

Donc la distribution a posteriori marginale de  $\mu$  est une distribution de Student de  $n - 1$  degrés de liberté et de moyenne  $\bar{x}$  ( $n > 1$ ), et le paramètre de balance  $\frac{s^2}{n}$ .

Supposons maintenant que le paramètre d'intérêt est  $\sigma^2$ , et que  $\mu$  est le paramètre de nuisance. Pour obtenir la distribution a priori marginale de  $\sigma^2$ ,

nous devons intégrer en dehors de  $\mu$ . Nous pouvons écrire :

$$\begin{aligned}
 \pi(\sigma^2/X) &\propto \int_{-\infty}^{+\infty} (\sigma^2)^{-\frac{n}{2}-1} \exp \left[ \sum_{i=1}^{i=n} -\frac{1}{2\sigma^2} (x_i - \mu)^2 \right] d\mu \\
 &= \int_{-\infty}^{+\infty} (\sigma^2)^{-\frac{n}{2}-1} \exp \left[ -\frac{(\bar{x} - \mu)^2}{2\frac{\sigma^2}{n}} - \frac{(n-1)s^2}{2\sigma^2} \right] d\mu \\
 &\propto (\sigma^2)^{-\frac{n}{2}-1} \exp \left[ -\frac{(n-1)s^2}{2\sigma^2} \right] \sqrt{\frac{2\pi\sigma^2}{n}} \times \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\frac{2\pi\sigma^2}{n}}} \exp \left[ -\frac{(\bar{x} - \mu)^2}{2\frac{\sigma^2}{n}} \right] d\mu \\
 &\propto (\sigma^2)^{-\frac{(n-1)}{2}-1} \exp \left[ -\frac{(n-1)s^2}{2\sigma^2} \right]
 \end{aligned}$$

qui est proportionnel à une densité de probabilité de Gamma inverse des paramètres  $\frac{(n-1)}{2}$  et  $\frac{(n-1)s^2}{2}$ .

## 1.7 Incorporation de l'information a priori en pratique

Nous allons nous limiter au cas d'un paramètre réel  $\theta$  et envisager trois situations selon que  $\theta \in [0, 1]$ ,  $[0, +\infty[$ , ou  $]-\infty, +\infty[$ . Nous supposons par ailleurs que l'information a priori fournie par l'Expert consiste en une estimation  $\theta^*$  de  $\theta$ , et en un intervalle  $I^* = [\alpha^*, \beta^*]$  contenant  $\theta^*$  et tel que la probabilité que  $\theta$  appartienne à  $I^*$  soit élevée (typiquement 0.95).

**Situation No 1 :** le paramètre  $\theta \in [0, 1]$ .

Pour incorporer cette information a priori, il est commode de reparamétriser la loi Beta comme suit : On pose  $a = \lambda\mu$  et  $b = \lambda(1 - \mu)$  où  $\mu = \frac{a}{a+b} = E[\theta]$  et  $\lambda = a+b$ . A noter que  $Var(\theta) = \frac{\mu(1-\mu)}{1+\lambda}$ . Donc à  $\mu$  fixé,  $\lambda$  est inversement proportionnel à la variance et donc s'interprète comme la précision de l'information a priori. Autrement dit, plus  $\lambda$  est grand, et plus  $Var(\theta)$  est petit, et plus l'information a priori est précise ; à l'inverse, plus  $\lambda$  est petit, et plus  $Var(\theta)$  est grand, et moins l'information a priori est précise.

En pratique, on procède comme suit. On choisit comme loi a priori une

## 1.8. LE POIDS DE L'A PRIORI DANS LA RÉPONSE BAYÉSIENNE

loi  $Beta(a, b)$  et on pose :  $E(\theta) = \theta^*$ . Le paramètre  $\lambda$  est lui déterminé par :

$$\Pr(\theta \in I^*) = \int_{I^*} \pi(\theta) d\theta = 0.95$$

Où  $\theta \sim Beta(\lambda\theta^*, \lambda(1 - \theta^*))$ . En pratique  $\lambda$  est déterminé, à l'aide d'un algorithme qu'il convient de programmer. Donnons un exemple. On suppose que :  $\theta^* = 0.2$  et que  $\Pr(\theta \in [0.05, 0.4]) = 0.95$ . On a donc :  $E(\theta) = 0.2$  et  $\theta \sim Beta(0.2\lambda, 0.8\lambda)$ . A l'aide de Matlab, on trouve  $\lambda = 17.5$ ; par conséquent, la loi a priori est une loi Beta de paramètres  $(a, b) = (3.5, 14)$ .

**Situation No 2 :** le paramètre  $\theta \in [0, +\infty[$ .

Dans ce cas on choisit comme loi a priori une loi Gamma qui peut être reparamétriser par sa moyenne et sa variance.

**Situation No 3 :** le paramètre  $\theta \in ]-\infty, +\infty[$ .

Dans ce cas on choisit comme loi a priori, une loi Normale. Incorporer l'information a priori consistant en  $\theta^* = 1$  et  $I^* = [0.8, 1.2]$ .

## 1.8 Le poids de l'a priori dans la réponse bayésienne

D'après (**Jérôme DUPUIS 2007**), on examine cette question sur un exemple pour comprendre comment l'information a priori et l'information contenue dans les observations se combinent l'une à l'autre pour produire la réponse bayésienne.

On se donne le modèle bayésien suivant :  $x_i/\theta \sim Bernoulli(\theta)$  et  $\theta \sim Beta(a, b)$  et  $\theta/x \sim Beta(a + x, b + n - x)$ , il est commode de reparamétriser la loi Beta à l'aide de  $a + b$  et  $\frac{a}{a + b}$  par :

$$\begin{aligned} E(\theta/x) &= \frac{a + x}{a + b + n} \\ &= \frac{a + b}{a + b + n} \cdot \frac{a}{a + b} + \frac{n}{a + b + n} \cdot \frac{x}{n} \\ &= \frac{a + b}{a + b + n} E(\theta) + \frac{n}{a + b + n} \bar{x} \end{aligned}$$

et de travailler avec la formule suivante :

$$E(\theta/x) = \frac{a + b}{a + b + n} E(\theta) + \frac{n}{a + b + n} \bar{x}$$

L'estimation bayésienne de  $\theta$  apparaît donc comme la moyenne pondérée de  $\bar{x}$  (c-à-d de l'estimation de  $\theta$  par maximum de vraisemblance), et de la moyenne a priori  $E(\theta)$ ; le poids de  $\bar{x}$  est la taille  $n$  de l'échantillon, et  $E(\theta)$  et  $\bar{x}$  sont affectés respectivement aux coefficients  $\frac{a+b}{a+b+n}$  et  $\frac{n}{a+b+n}$ .

A noter que  $\mathbf{a}+\mathbf{b}=\mathbf{n}$ , l'estimation bayésienne de  $\theta$  se situe exactement au milieu de l'intervalle  $[E(\theta), \bar{x}]$ .

Si  $\mathbf{a}+\mathbf{b}>\mathbf{n}$  cette estimation est plus proche de  $E(\theta)$  que de  $\bar{x}$ .

Si  $\mathbf{a}+\mathbf{b}<\mathbf{n}$  c'est l'inverse qui se produit.

Pour examiner l'influence de l'a priori sur  $E(\theta/x)$  il est édifiant de s'intéresser aux cas limites  $\mathbf{a}+\mathbf{b}\rightarrow 0$  et  $\mathbf{a}+\mathbf{b}\rightarrow \infty$ . Dans le premier cas, le poids de l'a priori est nul, et  $E(\theta/x) \rightarrow \bar{x}$  est la réponse classique; dans le second cas, le poids des données est nul, et  $E(\theta/x) \rightarrow E(\theta)$  ne dépend plus de  $x$ .

## 1.9 Etude de la sensibilité de la réponse bayésienne à la loi a priori

Le choix de la loi a priori comporte une certaine part d'arbitraire, et ce, à deux niveaux. Le premier niveau réside dans le choix de la famille de probabilités retenue pour la loi a priori (une loi Beta pour un paramètre dans  $[0; 1]$ , une loi Gamma pour un paramètre dans  $[0; +\infty[$ , etc). Le second réside dans le choix des valeurs numériques communiquées par l'Expert (typiquement  $\theta^*$  et  $I^*$ ) pour déterminer les hyperparamètres de la loi a priori. Il convient par conséquent d'examiner dans quelle mesure ces choix affectent l'estimation bayésienne de  $\theta$ . Dans la pratique on se contente d'examiner dans quelle mesure, une petite perturbation des hyperparamètres modifie l'estimation Bayésienne de  $\theta$ . C'est ce qu'on appelle : faire l'étude de sensibilité de la réponse bayésienne à la loi a priori. Disons qu'en pratique on comparera toujours l'estimateur de Bayes obtenu avec la loi a priori informative avec l'estimateur de Bayes correspondant à une loi a priori non informative; et ce, afin d'apprécier le poids de l'a priori dans la réponse Bayésienne informative.

## 1.10 La prédiction

Jusqu'ici, nous nous sommes concentrés sur l'inférence concernant la vraie valeur d'un paramètre. Cependant, dans beaucoup de situations, le but d'une

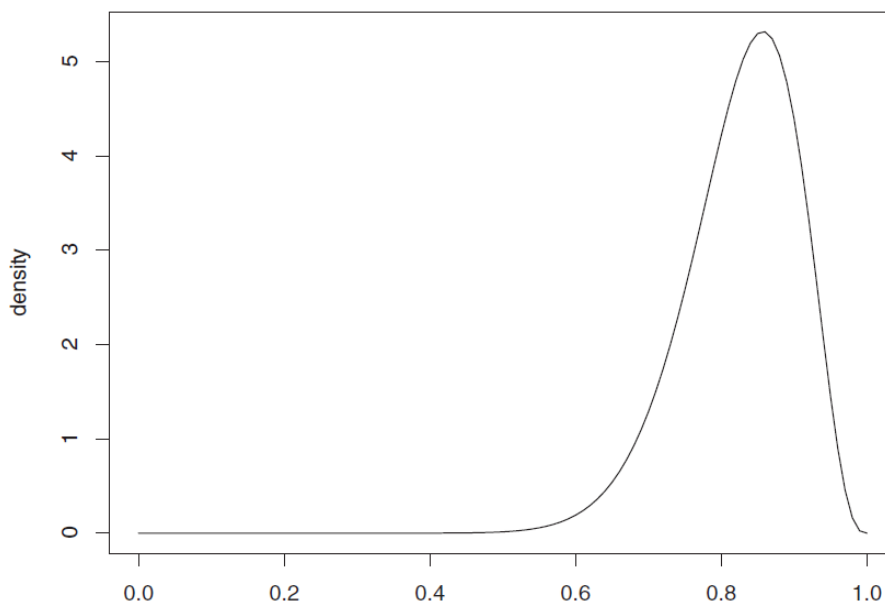
analyse est de prévoir des valeurs d'un futur échantillon.

Cependant, dans la pratique nous ne connaissons pas la valeur de  $\theta$ . Nous savons seulement la distribution a posteriori de  $\theta$ . Dans ce cas, **la probabilité prédictive** de  $y$  est donnée à partir de la distribution a posteriori  $\pi(\theta/x)$  et  $y \sim g(y/\theta, x)$  comme suit :

$$g(y/x) = \int g(y/\theta, x) \pi(\theta/x) d\theta$$

Un avantage de l'approche bayésienne est la capacité de considérer les probabilités non encore observées. Cela est illustré dans l'exemple suivant :

**exemple 1.10.1** *\*une étude publiée par Freireich et Al (1963) à été conçue pour évaluer les effectives de **chimio thérapeutique 6-mercaptopurine**.*



**Figure(1, 4)** : La densité a posteriori de  $\theta$  est  $Beta(19, 4)$ , (6-MP) pour le traitement de la leucémie aigue. Les patients ont été randomisés pour thérapie par paires. Soit  $M$  la proportion de la population de couples dans lesquels la 6 – MP séjour des patients en rémission plus longue que le patient du groupe placebo, (pour distinguer une probabilité  $\mu$  à partir d'une distribution de probabilité sur  $\mu$ , nous appellerons cela une proportion de la population ou une propension) L'hypothèse nulle est  $H_0 : \theta = \frac{1}{2}$  c-à-d pas d'efficacité de

6 – MP, et  $H_1 : \theta \neq \frac{1}{2}$ . Il y avait 21 paires de patients dans l'étude, et 18 d'entre eux favorisaient 6 – MP.

Supposons que la distribution a priori est uniforme sur  $(0, 1)$ . La distribution uniforme de  $(0, 1)$  est également la version de la distribution  $Beta(1, 1)$ . La mise à jour de la distribution  $Beta(a, b)$  après les succès et  $n - y$  échecs est facile, car la version beta est une probabilité a priori conjuguée, ce qui signifie que la distribution a posteriori apparaît comme un membre de la même famille de répartition que la précédente. Pour voir cela dans le cas des bêta-binomiales de la règle de Bayes nous avons :

$$\begin{aligned} \pi(\theta/x) &\propto f(x/\theta) \pi(\theta) \\ &= \binom{n}{x} \theta^x (1 - \theta)^{n-x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1 - \theta)^{b-1} \\ &\propto \theta^{x+a-1} (1 - \theta)^{n-x+b-1} \end{aligned}$$

On note dans cette dernière expression que nous avons réduit tous les termes multiplicatifs qui n'impliquent pas de  $\mu$  dans la constante inconnue de proportionnalité. Mais maintenant, nous reconnaissons cette forme comme étant proportionnelle à une densité de  $beta(x + a, n - x + b)$ . Puisqu'il s'agit de la densité seulement proportionnelle à notre formulaire qui reste intègre à 1, ce doit être la partie a posteriori que nous recherchons. Ainsi, dans notre cas ( $x = 18$  et  $n - x = 3$ ), la distribution a posteriori sous  $H_1$  est  $beta(19, 4)$ , telle que tracée dans la figure (1, 4). Notez que cette distribution est un modèle avec une moyenne de  $\frac{19}{19 + 4} = 0.8261$ .

Dans le contexte prédictif, on suppose qu'il est possible de prendre 5 observations supplémentaires. Comment cela serait-il utile ? Une façon de répondre est d'évaluer la conséquence d'obtenir  $k$  succès au cours des 5 prochaines paires de patients (pour un total de 26 paires) pour  $k = 0, 1, \dots, 5$ , puis de peser ces conséquences par les probabilités de prévisions des valeurs possibles de  $k$ . Pour calculer les probabilités de futures observations, on trouve premièrement ces probabilités en supposant que les paramètres sont connus, et puis on trouve la distribution a posteriori des paramètres. Nous avons ensuite la moyenne des distributions conditionnelles à l'égard de la distribution a posteriori des paramètres. Cela donne la distribution inconditionnelle prédictive d'intérêt.

Pour illustrer ; premièrement, envisager une seule paire de patients, la paire de 22 patients, ayant un résultat binaire variable  $x_{22}$ . Supposons que cette

paire est échangeable avec les 21 premières paires,  $X = (x_1, \dots, x_{21})'$ , pour laquelle  $x = \sum_{i=1}^{i=21} x_i = 18$ . Un nombre de la nouvelle paire est affecté à la 6 – MP, et l'autre est affecté au placebo. La probabilité prédictive que le patient 6 – MP restera dans la remise est plus la moyenne de la distribution a posteriori de  $\mu$ . Pour la distribution Beta(19, 4) considérée séparément, la probabilité prédictive de succès est donnée par la règle du placebo de la succession :  $\frac{19}{(19+4)} = 0.8261$  (Berry, 1996, sect 7.2). Et prouver que la distribution prédictive de  $x_{22}$  est :

$$\begin{aligned} g(x_{22}/X) &= \int g(x_{22}/\theta) \pi(\theta/X) d\theta \\ &= \int \theta^{x_{22}} (1-\theta)^{1-x_{22}} \frac{\Gamma(23)}{\Gamma(19)\Gamma(4)} \theta^{18} (1-\theta)^3 d\theta \end{aligned}$$

une application significative de la probabilité prédictive sera présentée dans le quatrième chapitre.

## 1.11 Méthodes usuelles de calcul bayésien

D'après (Robert (2007)), la statistique a été de tout temps inséparable des moyens de calcul, évoluant avec eux dans une relation duale où leur développement était souvent motivé par des besoins statistiques, mais aussi où leurs limitations empêchant le traitement de phénomènes plus complexes et le développement de techniques différentielles plus ambitieuses. Cette interaction est clairement visible dans le développement de l'informatique de la disponibilité de la puissance de calcul d'une telle magnitude. On a aussi vu éclore de nouvelles théories parmi les quelles on peut citer les méthodes de rééchantillonnage (bootstraps) et les méthodes linéaires généralisées.

Les possibilités ainsi induites par l'informatique autorisent le remplacement de considérations théoriques ou d'expériences effectives par l'appel à la simulation informative, pour plus de détails voir Robert (1986, 2001).

### 1.11.1 Méthodes de simulation de Monte Carlo

#### Motivations

Afin de motiver les méthodes de simulation de Monte Carlo dans un contexte bayésien, nous allons considérer une situation dans laquelle le calcul de l'estimation bayésienne de la grandeur d'intérêt est particulièrement laborieux à la main. Il importe de préciser que l'utilisation des méthodes de simulation de Monte Carlo ne se limite pas à la statistique bayésienne (ni même à la statistique); c'est pourquoi nous allons aussi donner un autre exemple de l'utilisation de ces méthodes, et ce dans un cadre purement probabiliste. Voici le modèle suivant qui consiste à étudier l'efficacité d'un traitement.

**Modèle :** Etude bayésienne de l'efficacité d'un traitement.

$n = 20$  patients atteints d'une même maladie suivent tous le même traitement. L'état de chaque patient est déterminé, régulièrement, lors de visites médicales. On désigne par  $x_{(i,t)}$  l'état du patient  $i$  lors de la  $t$ -ième visite;  $x_{(i,t)} = a$  s'il présente les symptômes de la maladie  $M$ , et  $b$  sinon. Les  $X_{(i,t)}$  sont markoviens ( $i$  étant fixé). La probabilité de transition de  $a$  vers  $a$  est notée  $\alpha$  et celle de  $b$  vers  $b$  est notée  $\beta$ . Le paramètre du modèle est donc  $\theta = (\alpha, \beta)$ . Le suivi du patient  $i$  est noté  $x_i$ . Les  $X_i$  sont supposés indépendants (conditionnellement à  $\theta$ ) et suivent tous la même loi. On suppose que les paramètres  $\alpha$  et  $\beta$  sont a priori indépendants. On adapte des lois uniformes comme lois a priori.

On note  $\omega$  la probabilité qu'un patient soit guéri à l'issue du traitement (on considère que c'est le cas s'il ne présente plus les symptômes de la maladie lors de la quatrième visite). On considère que le traitement est efficace si l'estimation de  $\omega$  dépasse 0.95. Il est facile de se convaincre que le calcul à la main de  $E[\omega/X]$  est particulièrement laborieux, même pour des valeurs modérées du nombre de visites.

#### Le principe de la méthode de Monte Carlo

Dans le problème statistique, l'approximation de l'intégrale

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

doit tirer avantage de la nature particulière de l'intégrale précédente, à savoir le fait que  $\pi$  soit une densité de probabilité (en supposant qu'il s'agisse



d'une loi a priori propre) ou plutôt, que  $f(x/\theta)\pi(\theta)$  soit proportionnel à une densité. Une conséquence naturelle de cette perspective est d'utiliser **la méthode de Monte Carlo**, par exemple, s'il est possible de produire des variables aléatoires  $\theta_1, \dots, \theta_m$  de loi  $\pi(\theta)$ , la moyenne

$$\frac{1}{m} \sum_{i=1}^{i=m} g(\theta_i) f(x/\theta_i)$$

converge presque sûrement vers

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

lorsque  $m$  tend vers  $+\infty$ , selon la loi des Grands Nombres. De la même façon, si un échantillon i.i.d de  $\theta_i$  de  $\pi(\theta/x)$  peut être simulé, la moyenne

$$\frac{1}{m} \sum_{i=1}^{i=m} g(\theta_i)$$

converge vers

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

De plus, si la variance a posteriori  $Var(g(\theta)/x)$  est finie, le théorème central limite s'applique à la moyenne  $\frac{1}{m} \sum_{i=1}^{i=m} g(\theta_i)$ , qui est alors asymptotiquement normale, de variance

$$\frac{Var(g(\theta)/x)}{m}$$

La mise en oeuvre de cette méthode nécessite la production d'une suite i.i.d  $\theta_i$  par ordinateur, reposant sur un générateur pseudo-aléatoire déterministe imitant la génération de  $\pi(\theta)$  ou  $\pi(\theta/x)$  comme suit : Un échantillon iid d'une loi uniforme  $U([0, 1])$  est généré, puis transformé en variables de la loi d'intérêt voir (**Robert et Casella (2004)**).

En réalité, la méthode de Monte Carlo s'applique dans un cadre beaucoup plus général que pour la simulation de  $\pi$ , comme dans le cas ci-dessus. Par exemple, puisque :

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

peut se représenter de plusieurs manières, il n'est pas nécessaire de simuler les lois  $\pi(. / x)$  ou  $\pi$  pour obtenir une bonne approximation de l'intégrale précédente. En effet, si  $h$  est une densité de probabilité telle que  $\text{supp}(h)$  inclut le support de  $g(\theta) f(x/\theta) \pi(\theta)$ , l'intégrale précédente peut aussi être représentée comme une espérance en  $h$ , à savoir

$$\int \frac{g(\theta) f(x/\theta) \pi(\theta)}{h(\theta)} h(\theta) d\theta$$

Cette représentation conduit à **la méthode de Monte Carlo** avec fonction d'importance  $h$  : générer  $\theta_1, \dots, \theta_m$  selon  $h$  et approcher

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

par :

$$\frac{1}{m} \sum_{i=1}^{i=m} g(\theta_i) \omega_i(\theta_i)$$

avec les poids :

$$\omega(\theta_i) = \frac{f(x/\theta_i) \pi(\theta_i)}{h(\theta_i)}$$

De nouveau, par la Loi des Grands Nombres, cette approximation converge presque certainement vers  $\int_{\Theta} g(\theta) f(x/\theta) \pi(\theta) d\theta$ .

Et une approximation de  $E^{\pi} [g(\theta) / x]$  est donnée par :

$$E^{\pi} [g(\theta) / x] = \frac{\sum_{i=1}^{i=m} g(\theta_i) \omega_i(\theta_i)}{\sum_{i=1}^{i=m} \omega_i(\theta_i)}$$

### 1.11.2 Méthode de Monte-Carlo par chaînes de Markov

Nous considérons une méthode de Monte-Carlo plus générale, permettant d'approcher la génération de variables aléatoires d'une loi a posteriori  $\pi(\theta/x)$  lorsque cette loi ne peut être simulée directement.

L'avantage de cette méthode sur les méthodes de Monte Carlo classiques est qu'elle ne nécessite pas la construction précise d'une fonction d'importance puisqu'elle prend en compte les caractéristiques de  $\pi(\theta/x)$ . Cette extension, appelée **Monte Carlo par chaînes de Markov** (et abrégée en MCMC), a des applications presque illimitées, même si ses performances varient largement, selon la complexité du problème. Elle tire son nom de l'idée que, pour produire des approximations acceptables d'intégrales et d'autres fonctions dépendant d'une loi d'intérêt, il suffit de générer une **chaîne de Markov**  $(\theta^{(m)})_m$  de loi limite de la loi d'intérêt. Cette idée d'utiliser le comportement limite d'une chaîne de Markov apparaît à la même époque que la technique de Monte Carlo originelle.

Nous présenterons la technique la plus importante conçue pour créer des chaînes de Markov de loi stationnaire donnée à savoir l'échantillonnage de Gibbs voir (**Casella (2004)**).

Les méthodes de **Metropolis-Hasting** et **Gibbs** sont deux groupes et deux méthodes d'application de la méthode de **Monte Carlo par chaînes de Markov** (MCMC).

### La méthode de Metropolis-Hastings

La méthode de Metropolis-Hastings est historiquement la première des méthodes MCMC. Pour une densité donnée  $\pi(\theta)$  connue un facteur de normalisation près, et une densité conditionnelle  $q(\theta'/\theta)$ , on va construire la chaîne de Markov.

#### Algorithme de Metropolis-Hastings

**Itération 0** : Initialiser avec une valeur arbitraire  $\theta^{(0)}$

**Itération m** : Mettre à jour  $\theta^{(m)}$  par  $\theta^{(m+1)}$  ( $m=1,2,\dots,n$ ), de la façon suivante :

**a/** Générer  $\xi = q(\xi/\theta^{(m)})$

**b/** Poser  $\rho(\theta^{(m)}, \xi) = \frac{\pi(\xi)q(\theta^{(m)}/\xi)}{\pi(\theta^{(m)})q(\xi/\theta^{(m)})} \wedge 1$

**c/** Prendre  $\theta^{(m+1)} = \begin{cases} \xi & \text{avec probabilité } \rho(\theta^{(m)}, \xi) \\ \theta^{(m)} & \text{sinon} \end{cases}$

La loi de densité  $\pi(\theta)$  est souvent appelée loi cible ou loi objet, tandis que la loi de densité  $q(\cdot/\theta)$  est dite loi de proposition. Une propriété intéressante de cet algorithme est d'autoriser un nombre infini de lois de proposition produisant toutes une chaîne de Markov convergeant vers la loi d'intérêt.

### L'échantillonnage de Gibbs

La méthode de Gibbs est essentiellement basée sur les distributions conditionnelles comme les méthodes de modélisation bayésienne, voir (**Robert** (2001)). L'approche de l'échantillonnage de Gibbs est fondé sur la loi  $\pi$ , alors l'échantillonnage de Gibbs tire profit des structures à partir de conditionnellement d'un modèle

$$\pi(\theta/x) = \int \pi_1(\theta/x, \lambda) \pi_2(\lambda/x) d\lambda$$

L'idée est alors de simuler la loi  $\pi_1$  et  $\pi_2$ , pour obtenir  $\pi(\theta/x)$ , la génération de  $\theta$  de  $\pi(\theta/x)$  est équivalente à la génération de  $\lambda$  de  $\pi_2$ , puis de  $\theta$  de  $\pi_1(\theta/x, \lambda)$ .

#### **Algorithme de Gibbs bi varié**

**Initialisation** : commencer par  $\lambda^{(0)}$

**Interaction** : pour  $\lambda^{t-1}$  donné.

**générer** : a/  $\theta^{(t)}$  selon  $\pi_1(\theta/x, \lambda^{(t-1)})$

b/  $\lambda^{(t)}$  selon  $\pi_2(\lambda/x, \theta^{(t)})$

Les applications dans les essais cliniques sont étudiées au quatrième chapitre, étude de procédure de test diagnostique.

# Chapitre 2

## Les tests bayésiens

### 2.1 Introduction

Nous allons présenter les tests bayésiens et pour plus de détails, voir (Robert 1922, Lehmann 1986, Berger 1985), dont nous reprenons ici quelques notions de base sur les tests bayésiens, ainsi que les développements récents de cette théorie qui possède de nombreux avantages dans la statistique appliquée.

Les divergences exhibées entre les diverses approches possibles du problème des tests ont pour but de mettre en évidence l'aspect incomplet et peu naturel de la modélisation classique qui, au contraire de l'approche bayésienne (pour laquelle la notion de probabilité d'une hypothèse a véritablement un sens, a priori comme a posteriori), doit recourir à des concepts artificiels pour fonder ses procédures. De plus, à l'opposé du cadre de l'estimation ponctuelle où les estimateurs optimaux au sens classique sont minimax et admissible et sont des limites des estimateurs de Bayes. Les réponses bayésiennes et fréquentistes sont différentes ; non seulement en nature, mais aussi numériquement, donc peuvent conduire à des conclusions opposées en pratique.

L'approche des tests d'hypothèses est généralement traitée comme un problème des tests, plutôt que l'estimation plus détaillée. Dans ce chapitre, on considère les tests d'hypothèses comme l'estimation de problème (tester  $\mathbf{I}_{\Theta_0}(\theta)$ ), reliée avec la théorie de la décision. Puis on peut arriver à quelques conclusions intéressantes. Dans le cas particulier, la responsabilité de la fonction de coût dans la règle de décision, et quelques propriétés intéressantes de la p-valeur.

## 2.2 L'approche standard

L'hypothèse des tests classiques est construite selon le lemme de Neyman-Pearson [Lehmann (1986)]. Et les résultats dans la règle de décision sont les règles 0 – 1. Ces formes de test bien qu'elles soient optimales dans le sens des fréquentistes strictement mais elles ont des critères pour des directions différentes. Peut être la critique de Neyman-Pearson devient active en pratique, car les formules de la théorie de Neyman-Pearson ne sont pas largement utilisées en pratique.

D'autre part, la p-valeur est implicitement utilisée comme une mesure de la preuve pour les hypothèses, donc Kiefer considère la p-valeur comme une évaluation de la vraisemblance de l'hypothèse nulle.

## 2.3 La fonction d'utilité

D'après (**Robert (2006)**), la notion d'utilité (définie comme l'opposé d'une fonction de coût) est utilisée non seulement en statistique, mais aussi en économie, et dans d'autres domaines comme la théorie des jeux où il est nécessaire d'ordonner les conséquences d'action ou de décisions. Conséquences (ou récompenses) sont des notions génériques qui résument l'ensemble des résultats émanant de l'action du décideur. Dans les cas les plus simples, il peut s'agir d'un gain ou d'un coût financier dû à cette décision.

Dans le cas de l'estimation, l'utilité peut être une mesure de la distance entre l'évaluation et la vraie valeur du paramètre. Les bases axiomatiques de l'utilité ont été attribuées à Von Neumann et Morgenstern (1947) et ont mené à des nombreuses extensions, particulièrement en théorie des jeux. Dans un cadre statistique, cette approche a été considérée par Wald (1950) et Ferguson (1967), voir aussi Chamberlain (2000) pour une connexion avec l'économétrie.

**Définition 2.3.1** *la fonction d'utilité est utilisée en statistique et en particulier dans le domaine économique. Son but principal est d'ordonner les conséquences d'action d'un décideur. Pour cela on considère que  $\mathfrak{R}$  l'espace des récompenses est supposé complètement connu, et d'autre part on suppose qu'il est possible d'ordonner les récompenses. On suppose qu'il existe une relation d'ordre total tel que :  $r_1 \leq r_2$  ou  $r_2 \leq r_1$  ; si  $r_1 \leq r_2$  et  $r_2 \leq r_3$  alors  $r_1 \leq r_3$  permet de choisir la meilleure.*

Nous considérons d'abord dans la section suivante l'approche bayésienne standard des tests, qui repose sur une évaluation des décisions par des coûts 0-1 et comparerons les procédures bayésiennes avec leurs homologues fréquentistes. Nous proposerons ensuite une alternative à l'approche décisionnelle fondée sur des coûts plus adaptés qui mettent en avant l'évaluation ex post pour des procédures de tests, par opposition aux procédures de Neyman-Pearson pour lesquelles l'évaluation fonctionne dans un esprit ex ante.

## 2.4 Une première approche de la théorie des tests

Soit un modèle statistique  $f(x/\theta)$ ,  $\theta \in \Theta$ . Etant donné un sous ensemble d'intérêt  $\Theta_0$  de  $\Theta$ , éventuellement réduit à un point, on cherche à déterminer si  $\theta$  est le vrai paramètre appartenant à  $\Theta_0$ , i.e à tester l'hypothèse

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

souvent appelée hypothèse nulle. Dans l'approche de Neyman-Pearson, ce problème est envisagé avec un espace de décision  $D$ , restreint à  $\{\text{où, non}\}$  ou de manière équivalente à  $\{1, 0\}$ , ce qui revient à estimer  $\mathbf{I}_{\Theta_0}(\theta)$ , 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  $H_0$

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

Lorsque  $\Theta_1 \neq \Theta_0^c$  le complémentaire de  $\Theta_0 \cup \Theta_1$  n'a pas d'intérêt pour le problème. Notons que d'un point de vue bayésien, cela revient à supposer a priori que :

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

Ainsi puisque tout test est en fait un estimateur de  $\mathbf{I}_{\Theta_0}(\theta)$ , il suffit de disposer d'une fonction de coût adéquate  $L(\theta, \varphi)$  pour proposer des estimateurs de bayes associés par exemple la fonction de coût introduite par Neyman et Pearson est le coût 0 - 1

$$L(\theta, \varphi) = \begin{cases} 1 & \text{si } \varphi = \mathbf{I}_{\Theta_0}(\theta) \\ 0 & \text{sinon} \end{cases}$$

où les estimateurs  $\varphi$  sont à valeur dans  $D = \{0, 1\}$ . Dans ce cas, la solution bayésienne est :

$$\varphi^\pi(x) = \begin{cases} 1 & \text{si } P(\theta \in \Theta_0 | x) > P^\pi(\theta \in \Theta_0^c | x) \\ 0 & \text{sinon} \end{cases}$$

Elle se justifie aisément d'un point de vue intuitif. On peut aussi envisager un coût plus général qui réalise différemment les mauvaises décisions suivant que l'hypothèse nulle est vraie ou fautive

$$L(\theta, \varphi) = \begin{cases} 0 & \text{si } \varphi = I_{\Theta_0}(\theta) \\ a_0 & \text{si } \theta \in \Theta_0 \text{ et } \varphi = 0 \\ a_1 & \text{si } \theta \notin \Theta_0 \text{ et } \varphi = 1 \end{cases} \quad (2.1)$$

On a alors la proposition suivante :

**Proposition 2.4.1** *sous le coût (2,1), la solution bayésienne associée à la loi a priori  $\pi$  est :*

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

**Preuve** il suffit d'écrire le coût a posteriori

$$\begin{aligned} L(\pi, \varphi(x)) &= \int_{\Theta} L(\theta, \varphi) \pi(\theta/x) d\theta \\ &= a_0 P^\pi(\theta \in \Theta_0/x) I_{\{0\}}(\varphi) + a_1 P(\theta \notin \Theta_0/x) I_{\{1\}}(\varphi) \end{aligned}$$

L'estimateur de Bayes peut être calculé directement ; par conséquent, l'hypothèse nulle  $H_0$  sera rejetée si la probabilité a posteriori de  $H_0$  est trop petite ; le seuil d'acceptation étant  $\frac{a_0}{(a_0 + a_1)}$  déterminé par le choix du coût. ■

**exemple 2.4.1** *soient  $x \sim B(n, p)$  et  $\Theta_0 = [0, \frac{1}{2}]$   
Sous la loi a posteriori uniforme  $\pi(p) = 1$  on a :*

$$P^\pi \left( p \leq \frac{1}{2}/x \right) = \frac{\int_0^{\frac{1}{2}} p^x (1-p)^{n-x} dp}{B(x+1, n-x+1)}$$

*qui peut donc être calculée et comparée au seuil donné par le coût, suivant les valeurs de  $n$  et de  $x$ .*



## 2.4. UNE PREMIÈRE APPROCHE DE LA THÉORIE DES TESTS

**exemple 2.4.2** soient  $x \sim N(\theta, \sigma^2)$  et  $\theta \sim N(\mu, \sigma^2)$

Alors,  $\pi(\theta/x)$  est la loi  $N(U(x), \rho)$  avec

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

Pour tester  $\theta < 0$ , on considère donc :

$$P^\pi(\theta < 0 \mid x) = P^\pi\left(\frac{\theta - U(x)}{\sqrt{\rho}} < -\frac{U(x)}{\sqrt{\rho}}\right) = \phi\left(-\frac{U(x)}{\sqrt{\rho}}\right)$$

Si  $Z_{a_0, a_1}$  est le fractile d'ordre  $\frac{a_1}{a_0 + a_1}$ , i.e satisfait  $\phi(Z_{a_0, a_1}) = \frac{a_1}{a_0 + a_1}$ , on acceptera l'hypothèse nulle si  $-U(x) > Z_{a_0, a_1} \sqrt{\rho}$ , i.e si :

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

Il est en fait naturel d'un point de vue bayésien de fonder la décision pour ce problème de test sur la probabilité a posteriori que l'hypothèse soit vérifiée.

Aussi le facteur de Bayes donne quelques indications sur le degré de certitude, mais les limites précises séparant une catégorie d'une autre sont conventionnelles et peuvent être changées de façon arbitraire.

### 2.4.1 Le facteur de Bayes

Bien que, d'un point de vue décisionnel, le facteur de Bayes ne soit qu'une transformation bijective de la probabilité a posteriori, il a fini par être considéré comme réponse en théorie des tests bayésiens, sous l'impulsion de Jeffrey (1939).

#### Facteur de Bayes pour tester les modèles

Le problème des tests d'hypothèses a sa propre terminologie conventionnelle qui, dans le cadre que nous adoptons, peut être décrit comme suit, et avec deux modèles alternatives  $M_0$  et  $M_1$  qui sont à l'étude et les deux sont des cas particuliers du modèle prédictif

$$P(x) = \int P(x/\theta) dQ(\theta)$$

Avec la même forme, on prend  $P(x/\theta)$ ,  $\theta \in \Theta$ , paramétrique, mais avec différents choix de  $Q$ , si pour le modèle  $M_i$ ,  $Q_i$  affecte la probabilité à une valeur donnée  $\theta_i$ .

$\omega$  est défini comme étant le modèle  $M_t$  vrai. On peut généraliser la fonction 0 – 1 utilisée auparavant en supposant que :

$$U(m_i, \omega) = -l_{il} \quad \omega = M_j, \quad i = 0, 1, \quad j = 0, 1$$

où

$$U(m_i, \omega) = \begin{cases} 1 & \text{si } \omega = M_j \\ 0 & \text{si } \omega = M_j \end{cases}$$

avec  $l_{00} = l_{11} = 0$  et  $l_{10}, l_{01} > 0$ .

Intuitivement, il y a une perte en choisissant le mauvais modèle, et il n'y a aucune perte dans le choix du bon modèle, compte tenu des données  $x$ , avec l'utilisation :

$$P_i(\omega/x) = \begin{cases} 1 & \text{si } \omega = M_j \\ 0 & \text{sinon} \end{cases}$$

D'après (**Bernardo (2010)**), nous commencerons par le problème des tests d'hypothèse comme un problème du choix du modèle de remplacement des deux hypothèses  $H_0$  et  $H_1$  par deux candidats paramétriques  $M_0$  et  $M_1$ , modèles ayant les vecteurs de paramètres respectifs  $\theta_0$  et  $\theta_1$ . En vertu des densités a priori  $\pi_i(\theta_i)$ ,  $i = 0, 1$ , et les distributions marginales de  $\chi$  sont trouvées en intégrant les paramètres :

$$P(x/M) = \int f(x/\theta_i, M_i) \pi_i(\theta_i) d\theta_i, \quad i = 0, 1$$

**Définition 2.4.1** (*Facteur de Bayes*) : étant donné deux hypothèses  $H_i$  et  $H_j$  correspondant à des hypothèses d'autres modèles  $M_i$  et  $M_j$  pour les données  $x$ , le facteur de Bayes en faveur de  $H_i$  (et contre  $H_j$ ) est donné par :

$$B_{ij}^\pi(x) = \frac{P(x/M_i)}{P(x/M_j)} = \left\{ \frac{P(M_i/x)}{P(M_j/x)} \right\} \Big/ \left\{ \frac{P(M_i)}{P(M_j)} \right\}$$

Intuitivement, le facteur de Bayes fournit une mesure pour savoir si les données  $\chi$  ont augmenté ou diminué les chances de  $H_i$  par rapport à  $H_j$ , ainsi :

## 2.4. UNE PREMIÈRE APPROCHE DE LA THÉORIE DES TESTS

---

$B_{ij} > 1$  signifie donc que  $H_i$  est plus plausible relativement à la lumière de  $x$ .

$B_{ij} < 1$  signifie que la plausibilité relative de  $H_j$  a augmenté.

Si  $\mathbf{i} = \mathbf{0}$ ,  $\mathbf{j} = \mathbf{1}$  :

On peut obtenir la probabilité a posteriori  $P(M_0/x)$  et  $P(M_1/x) = 1 - P(M_0/x)$  pour deux modèles et le facteur de Bayes donné par :

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

avec  $P(M_1/x) \neq 0$ ,  $P(M_1) \neq 0$ ,  $P(M_0) \neq 0$ .

La fonction d'utilité espérée pour  $m_i$  est :

$$\bar{U}(m_i/x) = -[l_{i0}P(M_0/x) + l_{i1}P(M_1/x)]$$

où

$$\bar{U}(m_i/x) = \int U(m_i, \omega) P(\omega/x) d\omega = P(M_1/x), \quad i \in I$$

De sorte que  $\bar{U}(m_0/x) < \bar{U}(m_1/x)$  si et seulement si  $l_{01}P(M_1/x) > l_{10}P(M_0/x)$

Nous choisissons donc  $M_1$  si et seulement si :

$$\frac{P(M_0/x)}{P(M_1/x)} < \frac{l_{01}}{l_{10}}$$

Ce qui suit décrit les formes des tests de Bayes qui se posent dans la comparaison de modèles lorsque ceux-ci sont définis par l'hypothèse paramétrique. L'exemple suivant illustre à la fois les comparaisons des modèles généraux et une instance spécifique de tests d'hypothèses.

**La comparaison de deux modèles** : supposons la comparaisons des deux modèles paramétriques complètement spécifiés **Binomial-négative** et **Poisson** définis pour  $x_1, x_2, \dots, x_n$  sont conditionnellement indépendants par :

$$M_0 : Nb(x_i/1, \theta_0) \quad , \quad M_1 : Pn(x_i/\theta_1) \quad i = 1, \dots, n$$

Dans ce cas le facteur de Bayes est donné par :

$$B_{01}(x) = \frac{\prod_{i=1}^{i=n} Nb(x_i/1, \theta_0)}{\prod_{i=1}^{i=n} Pn(x_i/\theta_1)} = \frac{\theta_1^n (1 - \theta_1)^{n\bar{x}}}{\theta_2^{n\bar{x}} \left\{ \prod_{i=1}^{i=n} x_i! \right\}^{-1} \exp\{-n\theta_2\}}$$

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

Si  $n = 2$ ,  $x_1 = x_2 = 0$ , donc  $B_{01}(x) = \frac{e^4}{9} \simeq 6.1$ ; indiquant une augmentation de la plausibilité de  $M_0$ .

Si  $n = 2$ ,  $x_1 = x_2 = 2$ , donc  $B_{01}(x) = \frac{4e^4}{729} \simeq 0.3$ ; indiquant une légère augmentation de la plausibilité de  $M_1$ .

On suppose maintenant que  $\theta_0, \theta_1$  sont inconnus et ils sont assignés aux distributions a priori suivantes :

$$\pi_0(\theta_0) = Be(\theta_0/\alpha_0, \beta_0) \quad , \quad \pi_1(\theta_1) = Ga(\theta_1/\alpha_1, \beta_1)$$

Il suit que :

$$\begin{aligned} P(x_1, x_2, \dots, x_n/M_0) &= \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} \int_0^1 \theta_0^{n+\alpha_0-1} (1-\theta_0)^{n\bar{x}-\beta_0-1} d\theta_0 \\ &= \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} \frac{\Gamma(\alpha_0 + n)\Gamma(n\bar{x} + \beta_0)}{\Gamma(n + n\bar{x} + \alpha_0 + \beta_0)} \end{aligned}$$

et que :

$$\begin{aligned} P(x_1, x_2, \dots, x_n/M_1) &= \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1) \prod_{i=1}^{i=n} x_i!} \int_0^\infty \theta^{n\bar{x}+\alpha_1-1} \exp\{-(n+\beta_1)\theta\} d\theta_1 \\ &= \frac{\Gamma(n\bar{x} + \alpha_1) \beta_1^{\alpha_1}}{\Gamma(\alpha_1) (n + \beta_1)^{n\bar{x}+\alpha_1}} \frac{1}{\prod_{i=1}^{i=n} x_i!} \end{aligned}$$

noter en outre que :

$$\begin{aligned} E(x_i/M_0) &= \int_0^1 E(x_i/M_0, \theta_0) Be(\theta_0/\alpha_0, \beta_0) d\theta_0 \\ &= \int_0^1 \frac{(1-\theta_0)}{\theta_0} Be(\theta_0/\alpha_0, \beta_0) d\theta_0 \\ &= \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} \frac{\Gamma(\alpha_0 - 1)\Gamma(\beta_0 + 1)}{\Gamma(\alpha_0 + \beta_0)} \\ &= \frac{\beta_0}{\alpha_0} \end{aligned}$$

## 2.4. UNE PREMIÈRE APPROCHE DE LA THÉORIE DES TESTS

---

et

$$\begin{aligned}
 E(x_i/M_1) &= \int_0^{\infty} E(x_i/M_1, \theta_1) Ga(\theta_1/\alpha_1, \beta_1) d\theta_1 \\
 &= \int_0^{\infty} \theta_1 Ga(\theta_1/\alpha_1, \beta_1) d\theta_1 \\
 &= \frac{\alpha_1}{\beta_1}
 \end{aligned}$$

de sorte que la spécification préalable avec  $\alpha_0\alpha_1 = \beta_0\beta_1$ , implique les mêmes moyens pour les deux modèles prédictifs.

	$\alpha_0 = 1, \beta_0 = 2$ $\alpha_1 = 2, \beta_1 = 1$	$\alpha_0 = 30, \beta_0 = 60$ $\alpha_1 = 60, \beta_1 = 30$	$\alpha_0 = 1, \beta_0 = 2$ $\alpha_1 = 4, \beta_1 = 2$
$x_0 = x_1 = 0$	1.5	5.5	0.27
$x_0 = x_1 = 2$	0.29	0.30	0.38

**Table(2, 1)** : dépendance du  $B_{01}(x)$  sur des combinaisons a priori

### Facteur de Bayes pour tester les paramètres

On ne s'intéresse pas seulement aux modèles mais aussi aux paramètres.

**Définition 2.4.2** on appelle facteur de Bayes le rapport

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

**Cas spécial** :  $\Theta_0 = \{\theta_0\}$ ,  $\Theta_1 = \{\theta_1\}$

$$\begin{aligned}
 B_{01}(x) &= \left\{ \frac{P(\theta_0/x)}{P(\theta_1/x)} \right\} \Big/ \left\{ \frac{P(\theta_0)}{P(\theta_1)} \right\} \\
 &= \left\{ \frac{\frac{\pi(\theta_0, x)}{m(x)}}{\frac{\pi(\theta_1, x)}{m(x)}} \right\} \Big/ \left\{ \frac{\pi(\theta_0)}{\pi(\theta_1)} \right\} \\
 &= \left\{ \frac{f(x/\theta_0) \pi(\theta_0)}{f(x/\theta_1) \pi(\theta_1)} \right\} \Big/ \left\{ \frac{\pi(\theta_0)}{\pi(\theta_1)} \right\} \\
 &= \frac{f(x/\theta_0)}{f(x/\theta_1)}
 \end{aligned}$$

A la suite de Jeffrey (1939) et de Good (1952), le facteur de Bayes est désormais un outil à part entière, voir (Kass et Raftery, 1995). En particulier, Jeffrey (1939) a développé une échelle “**absolue**” pour évaluer le degré de certitude en faveur ou au détriment de  $H_0$  apporté par les données, en l’absence d’un cadre décisionnel véritable, l’échelle de Jeffrey est la suivante :

- (i) Si  $0 < \log(B_{10}^\pi) < \frac{1}{2}$ , la certitude que  $H_0$  est faible.
  - (ii) Si  $\frac{1}{2} < \log(B_{10}^\pi) < 1$ , cette certitude est substantielle.
  - (iii) Si  $1 < \log(B_{10}^\pi) < 2$ , elle est forte .
  - (iiii) Si  $\log(B_{10}^\pi) > 2$ , elle est décisive.
- Où  $B_{10}^\pi = \frac{1}{B_{01}^\pi}$ .

Donc le facteur de Baye permet l’utilisation de test bayésien sans avoir recours à une fonction de perte spécifique.

### Insuffisances ou critiques

Le facteur de Bayes n’est en effet défini que si  $P^\pi(\Theta_0) \neq 0$  et  $P^\pi(\Theta_1) \neq 0$ . En d’autres termes, si  $H_0$  et  $H_1$  sont a priori impossible, il est inutile de chercher à modifier ces probabilités en fonction des observations, les probabilités resteront les mêmes. Par conséquent, il est impossible de tester les hypothèses ponctuelles lorsque la loi a priori est continue. Il faut donc adapter le choix des lois a priori au problème de test, considéré de manière a ce que hypothèses nulle et alternative aient des probabilité a priori strictement positives.

Une critique usuelle sur les hypothèses ponctuelles est qu’elles ne sont pas réalistes. Ainsi, il existe cependant des cas où on cherche à mettre en évidence un effet ou l’absence d’effet et ces problèmes se traduisent par des hypothèses ponctuelles.

En fait, on pourrait argumenter qu’au contraire, seules les hypothèses ponctuelles sont importantes car même si une hypothèse unilatérale permet de déterminer si un traitement a une influence positive ou négative, il sera nécessaire de tester ensuite si ce traitement retenu a effectivement une influence.

### Modification de la loi a priori

En générale, le facteur de Bayes dépend effectivement de l’information a priori mais traduit néanmoins une “réponse objective”, d’un point de vue Bayésien ; car il élimine une partie d’influence de l’information a priori sur la

## 2.4. UNE PREMIÈRE APPROCHE DE LA THÉORIE DES TESTS

---

vraisemblance de l'hypothèse nulle et mesure ainsi la part due aux observations si :  $\rho_0 = \pi(\theta \in \Theta_0) \neq 0$ ,  $\rho_1 = \pi(\theta \in \Theta_1) = 1 - \rho_0 \neq 0$ .

Si  $H_0$  et  $H_1$  sont impossibles à réaliser.

j'appelle :  $g_0 \propto \pi(\theta) I_{\Theta_0}(\theta)$ .

$$g_1 \propto \pi(\theta) I_{\Theta_1}(\theta)$$

$$\pi(\theta) = \rho_0 \pi(\theta) + \rho_1 \pi(\theta)$$

L'a priori sur  $\Theta_0$  est la masse de Dirac. Le test à une hypothèse nulle ponctuelle  $\theta = \theta_0$ , et la probabilité a posteriori que  $\theta = \theta_0$  est alors :

$$\pi(\theta_0/x) = \frac{f(x|\theta_0)\rho_0}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta} = \frac{f(x|\theta_0)\rho_0}{f(x|\theta_0)\rho_0 + (1-\rho_0)m_1(x)}$$

avec :

$$m_1(x) = \int_{\theta_1} f(x|\theta)g_1(\theta)d\theta$$

donc :

$$\pi(\theta_0/x) = \left[ 1 + \frac{1-\rho_0}{\rho_0} \cdot \frac{m_1(x)}{f(x|\theta_0)} \right]^{-1}$$

et on a :

$$B_{01}^{\pi}(x) = \frac{f(x|\theta_0)\rho_0}{m_1(x)(1-\rho_0)} \Big/ \frac{\rho_0}{1-\rho_0} = \frac{f(x|\theta_0)}{m_1(x)}$$

donc :

$$\pi(\theta_0/x) = \left[ 1 + \frac{1-\rho_0}{\rho_0} \cdot \frac{1}{B_{01}^{\pi}(x)} \right]^{-1}$$

**exemple 2.4.3** soit à tester  $H_0 = \frac{1}{2}$  contre  $H_1 \neq \frac{1}{2}$ , il vient :

$$\begin{aligned} \pi(\theta_0/x) &= \left[ 1 + \frac{1-\rho_0}{\rho_0} 2^n B(x+1, n-x+1) \right]^{-1} \\ &= \left[ 1 + \frac{1-\rho_0}{\rho_0} \frac{x!(n-x)!}{n!} \cdot 2^n \right]^{-1} \end{aligned}$$

Puisque  $m(x) = \binom{n}{x} B(x+1, n-x+1)$ , par exemple si  $n = 5$ ,  $x = 3$ ,  $\rho_0 = \frac{1}{2}$ , on obtient une probabilité a posteriori de  $(1 + \frac{2}{20} \cdot 2^5)^{-1} = \frac{5}{21}$ , et un facteur de Bayes de  $\frac{5}{16}$ .

Lorsque la taille de l'échantillon augmente, le champ des réponses possibles

s'élargit, ainsi si  $\pi(p)$  est  $B(\frac{1}{2}, \frac{1}{2})$  et  $n = 10$ , on obtient la table(2, 2) pour les probabilités a posteriori, bien que la loi a priori soit très défavorable (puisque favorisant les extrêmes 0 et 1).

$x$	0	1	2	3	4	5
$P(p = \frac{1}{2}/x)$	0.0055	0.0953	0.3737	0.6416	0.7688	0.8025

**Table(2, 2)** : probabilité a posteriori de  $\left\{p = \frac{1}{2}\right\}$  pour  $n=10$

**exemple 2.4.4** pour  $x \sim N(\mu, \sigma^2)$  et  $\mu \sim N(\xi, \tau^2)$ , on considère le test pour  $H_0 : \mu = 0$ , on choisit  $\xi = 0$ . Donc le facteur de Bayes est :

$$\begin{aligned} B_{10}^{\pi}(x) &= \frac{m_{1(x)}}{f(x | \mu = 0)} \\ &= \frac{\sigma}{\sqrt{\sigma^2 + \tau^2}} \cdot \frac{\exp\{-x^2/2(\sigma^2 + \tau^2)\}}{\exp\{-x^2/2\sigma^2\}} \\ &= \sqrt{\frac{\sigma^2}{\sigma^2 + \tau^2}} \cdot \exp\left\{\frac{\tau^2 x^2}{2\sigma^2(\sigma^2 + \tau^2)}\right\} \end{aligned}$$

La probabilité a posteriori s'en déduit immédiatement

$$\pi(\mu = 0/x) = \left[1 + \frac{1 - \rho_0}{\rho_0} \cdot \sqrt{\frac{\sigma^2}{\sigma^2 + \tau^2}} \exp\left(\frac{\tau^2 x^2}{2\sigma^2(\sigma^2 + \tau^2)}\right)\right]^{-1}$$

En considérant le cas où  $\pi_0 = \frac{1}{2}$  et  $\sigma = \tau$ , on obtient la table (2, 3) en fonction de  $Z = \frac{x}{\sigma}$ .

$Z$	0	0.68	1.28	1.96
$p(\mu = 0/x)$	0.586	0.528	0.484	0.244

**Table(2, 3)** : probabilité a posteriori de  $\mu=0$  suivant les valeurs de  $Z$

Dans le cas où  $\tau = 10\sigma$ , reflétant une information plus vague sur  $\mu$ . les probabilités modifiées sont données dans la table (2, 4).

$Z$	0	0.68	1.28	1.96
$p(\mu = 0/x)$	0.786	0.729	0.612	0.366



**Table(2, 4)** : probabilité a posteriori de  $\mu = 0$  pour une loi a priori plus diffuse

L'utilisation de loi non informatives dans le contexte de la théorie des tests est assez limitée. Comme nous l'avons remarqué ci-dessus le contexte des tests n'est pas compatible avec une approche entièrement non informative, puisqu'ils présupposent une division de l'espace des paramètres en deux sous-espaces (au moins). Cette incompatibilité est en fait plus fondamentale car elle empêche l'utilisation de lois impropres dans les tests d'hypothèses bilatérales.

### L'interdiction de l'a priori impropre

La décomposition de la distribution a priori en deux sous a priori apporte une difficulté reliée à l'a priori impropre ; en pratique, il existe une interdiction de l'utilisation dans les tests si on utilise la représentation :

$$\pi(\theta) = P(\theta \in \Theta_0) \pi_0(\theta) + P(\theta \in \Theta_1) \pi_1(\theta)$$

Les poids de  $P(\theta \in \Theta_0)$  et  $P(\theta \in \Theta_1)$  sont significatifs si  $\pi_0$  et  $\pi_1$  sont des densités de probabilité normalisé.

**exemple 2.4.5** si  $x \sim N(\mu, 1)$  ,  $H_0 : \mu = 0$  .

*L'a priori impropre de Jeffrey*

$$\begin{aligned} \pi_1(\mu) &= 1 \cdot \pi(\mu) = \frac{1}{2} \cdot I_0(\mu) + \frac{1}{2} \cdot 1 \\ \pi(\mu = 0/x) &= \frac{f(x/\mu) \pi(\mu)}{\int_{\mathbb{R}} f(x/\mu) \pi(\mu) d\mu} \\ &= \frac{\frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\}}{\frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\{-(x-\theta)^2/2\} d\mu} \\ \pi(\mu = 0/x) &= \frac{\exp\{-x^2/2\}}{\exp\{-x^2/2\} + \int_{-\infty}^{+\infty} \exp\{-(x-\mu)^2/2\} d\theta} \\ &= \frac{1}{1 + \sqrt{2\pi} \exp\{x^2/2\}} \\ &\leq \frac{1}{1 + \sqrt{2\pi}} = 0.285 \end{aligned}$$

Cette probabilité a posteriori de  $H_0$  est majorée par  $\frac{1}{(1 + \sqrt{2\pi})} = 0.285$ , donc ne donne que très peu de chances à l'hypothèse nulle, même dans le cas le plus favorable. A moins de changer l'échelle des comparaisons, donc de modifier le coût on acceptera rarement l'hypothèse nulle.

## 2.5 Comparaison avec la théorie classique

### 2.5.1 Test UMP et UMPU :

L'approche classique en théorie des tests est celle de Neyman-Pearson exposée par exemple dans Lehmann (1986) sous le coût "0 – 1" noté  $L_1$ . On considère la notion d'optimalité suivante.

**Définition 2.5.1** si  $\alpha \in ]0; 1[$  et sous la contrainte

$$\sup_{\theta \in \Theta_0} E_{\theta} [L(\theta, \varphi/x)] = \sup_{\theta \in \Theta_0} P_{\theta}(\varphi(x) = 0) \leq \alpha$$

Un test  $\varphi$  est dit uniformément plus puissant au niveau  $\alpha$  (UMP) s'il minimise le risque  $E_{\theta} [L(\theta, \varphi(x))]$  uniformément sur  $\Theta_1$ . Cette notion d'optimalité est en fait plus faible que la notion d'admissibilité qui sera développée après, car le coût envisagé ici est d'une certaine manière bidimensionnel du fait de la restriction sur le risque de première espèce  $\sup_{\Theta_0} [L(\theta, \varphi)]$ ; cette restriction est en générale nécessaire pour obtenir un test optimal mais :

- (i) elle conduit à une dissymétrie entre hypothèse nulle et alternative qui se transmet dans les tests par un comportement peu naturel.
- (ii) elle nécessite le choix du niveau  $\alpha$  par le décideur indépendamment du coût (puisque travaillant avec le coût  $L_1$ ).
- (iii) elle n'aboutit pas nécessairement à une réduction suffisante de l'ensemble des estimateurs. Pour permettre de sélectionner un estimateur optimal il faut parfois introduire de nouvelles restrictions.

**Proposition 2.5.1** dans le cas simple où l'hypothèse nulle et alternative sont ponctuelles  $H_0 : \theta = \theta_0$  et  $H_1 : \theta = \theta_1$ , le paramètre de Neymann-Pearson établit qu'il existe des tests UMP et qu'ils sont de la forme

$$\varphi(x) = \begin{cases} 1 & \text{si } f(x/\theta_1) < kf(x/\theta_0) \\ 0 & \text{sinon} \end{cases}$$

## 2.5. COMPARAISON AVEC LA THÉORIE CLASSIQUE

---

$k$  étant relié au seuil  $\alpha$  considéré (voir Lehmann, 1986); dans le cas des familles à rapport de vraisemblance monotone pour lesquelles il existe une statistique  $T(x)$  telle que  $\frac{f(x/\hat{\theta})}{f(x/\theta)}$  soit croissante en  $T(x)$  pour  $\hat{\theta} > \theta$ , Kerlin et Rubin (1956) ont montré l'extension suivante (voir Lehmann, 1986, p79).

**Proposition 2.5.2** soit  $f(x/\theta)$  à rapport de vraisemblance monotone en  $T(x)$  pour  $H_0 : \theta \leq \theta_0$  et  $H_1 : \theta > \theta_0$ . Il existe un test UMP de la forme :

$$\varphi(x) = \begin{cases} 1 & \text{si } T(x) < c \\ \gamma & \text{si } T(x) = c \\ 0 & \text{sinon} \end{cases}$$

$\delta$  et  $c$  étant déterminés par :

$$E_{\theta_0} [\varphi(x)] = 1 - \alpha$$

Remarquons d'autre part qu'un cas important de familles à rapport de vraisemblance monotone est fournir pour les familles exponentielles puisque :

$$\frac{f(x/\hat{\theta})}{f(x/\theta)} = \frac{\exp\{\hat{\theta}x - \psi(\hat{\theta})\}}{\exp\{\theta x - \psi(\theta)\}} = \frac{\exp\{(\hat{\theta} - \theta)x\}}{\exp\{\psi(\hat{\theta}) - \psi(\theta)\}}$$

Pfanzagl (1968) a aussi montré que l'existence d'un test UMP pour toutes les tailles d'échantillon et un seuil  $\alpha$  donné; implique que la famille considérée est exponentielle.

**Proposition 2.5.3** soient  $x \sim P(\lambda)$  et  $H_0 : \lambda \leq \lambda_0$  et  $H_1 : \lambda > \lambda_0$ , pour  $m$  observations indépendantes de cette loi; une statistique exhaustive est  $S = \sum_i x_i \sim P(m\lambda)$  et d'après la proposition (2.5.2) un test UMP est donné par

$$\varphi(x) = \begin{cases} 1 & \text{si } S < k \\ \gamma & \text{si } S = k \\ 0 & \text{sinon} \end{cases}$$

Pour

$$E_{\lambda_0} [\varphi(x)] = P_{\lambda_0}(S < k) + \gamma P_{\lambda_0}(S = k) = 1 - \alpha$$

La proposition (2.5.2) souligne un problème majeur de l'approche de Neuman-Pearson à savoir que tous les seuils ne sont pas accessibles à moins de recourir à la randomisation ; en effet puisque l'espace des décisions est réduit à  $D = \{0, 1\}$ , si  $\varphi(x) = \gamma$  ; on répondra  $\varphi(x) = 1$  avec la probabilité  $\delta$ . Une telle approche est évidemment incompatible avec le principe de vraisemblance bien qu'elle n'apparaisse pas dans le cas où  $T(x)$  est continue. D'autre part, la proposition (2.5.2) ne s'applique qu'aux hypothèses unilatérales. Dans le cas d'hypothèse bilatérale on a cependant (Lehman , 1986 , pp.101-103) .

**Proposition 2.5.4** *soient*

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

et  $H_0 : \theta \leq \theta_1$  ou  $\theta \geq \theta_2$ ,  $H_1 : \theta_1 \leq \theta \leq \theta_2$ . Il existe alors un test UMP de la forme

$$\varphi(x) = \begin{cases} 0 & \text{si } c_1 < T(x) < c_2 \\ \gamma_i & \text{si } T(x) = c_i \quad (i = 1, 2) \\ 1 & \text{sinon} \end{cases}$$

avec  $(i = 1, 2)$  et  $E_{\theta_i}[\varphi(x)] = 1 - \alpha$  . Mais il n'existe pas de test UMP pour le cas symétrique i.e ,  $H_0 : \theta_1 \leq \theta \leq \theta_2$  ce qui est choquant et jette un doute sur les prétentions à la cohérence d'une telle approche bayésienne, dans ce cas il faut se restreindre au **tests uniformément plus puissants sans biais** (UMPU) i.e, ce qui vérifient

$$\sup_{\Theta_0} E_{\theta} [L(\theta, \varphi(x))] \leq \inf_{\Theta_1} E_{\theta} [L(\theta, \varphi(x))]$$

Donc renforcent la dissymétrie entre  $H_0$  et  $H_1$ .

## 2.5.2 La p-valeur

Les praticiens ont tenté de répondre à plusieurs critiques et désavantages de l'approche de Neyman-Pearson en faisant disparaître le seuil de signification  $\alpha$  et en proposant une réponse alternative à valeur dans  $[0, 1]$ .

**Définition 2.5.2** *Etant donné un test d'hypothèse, on appelle p-valeur associée à une observation  $x$ , le plus petit seuil pour lequel cette observation conduirait au rejet de l'hypothèse nulle. Si on note  $R_\alpha$  la région de rejet au niveau du test au seuil  $\alpha$  alors :*

$$p(x) = \inf \{ \alpha, x \in R_\alpha \}$$

## 2.6. UNE SECONDE APPROCHE DÉCISIONNELLE

---

La  $p$ -valeur est donc considérée comme une mesure de la crédibilité à attacher à l'hypothèse nulle. La décision de rejeter ou de ne pas rejeter l'hypothèse nulle est alors prise à la lumière du nouveau degré de conviction entretenu à l'égard de l'hypothèse nulle et des conséquences consécutives aux erreurs de première et seconde espèces.

**Calcul de la  $p$ -valeur** : puisque la région critique (i.e la région de rejet de  $H_0$  du test) est de la forme  $\{|x| > k\}$ , la  $p$ -valeur est donnée par :

$$\begin{aligned} p(x) &= \inf_{\{|x| > k_\alpha\}} \alpha = P^x (|X| > |x|) && x \sim N(0, 1) \\ &= 1 - P^x (|X| < |x|) \\ &= 2 [1 - \Phi(x)] \end{aligned}$$

Par conséquent, si  $x = 1.68$ ,  $p(x) = 0.10$  et, si  $x = 1.96$ ,  $p(x) = 0.05$

Les  $p$ -valeurs sont donc des statistiques acceptables d'un point de vue fréquentiste, suivant les directives d'approche conditionnelle, suggérées par Kiefer (1977) et Robinson 1979.

La  $p$ -valeur n'a pas de justification intrinsèque, car sa prétendue "optimalité" découle de celle des procédures de test, dont elle est dérivée. En un sens, la même remarque s'applique aux probabilités a posteriori, car bien qu'elles soient intuitivement justifiables, celles-ci ne sont pas validées par un processus de décision. Dans la section suivante nous construisons une alternative à l'approche de Neyman-Pearson pour justifier les probabilités a posteriori et la  $p$  valeur.

## 2.6 Une seconde approche décisionnelle

### 2.6.1 Equivalence de la problématique

On considère le même problème du test d'hypothèse de la première approche de la théorie des tests (on observe la valeur  $x$  du vecteur  $X$  avec la densité  $f(x/\theta)$  et on désire une conclusion sur les hypothèses

$$H_0 : \theta \in \Theta_0 \text{ contre } H_1 : \theta \in \Theta_0^c \quad (2.2)$$

Où  $\Theta_0$  est un sous ensemble d'espace de paramètre  $\Theta$ . Le test  $H_0 : \theta \in \Theta_0$  contre  $H_1 : \theta \in \Theta_1$  où  $\Theta_1 \neq \Theta_0^c$  et on va estimer la fonction  $I_{\Theta_0}(\theta)$ .

L'exécution de l'estimateur de décision  $\phi(x)$  est évaluée avec le respect de la fonction de coût suivante :

$$L(\theta, \phi) = d(I_{\Theta_0}(\theta) - \phi(x)) \quad (2.3)$$

où la fonction  $d(t)$  est minimale lorsque  $t = 0$ , non décroissante pour  $t > 0$  et croissante pour  $t < 0$ . Comme l'estimateur n'est pas une décision entre  $H_0$  et  $H_1$ . Donc on fait une évaluation de  $H_0$  plutôt que de tirer une conclusion sur  $H_1$ ; on essaie d'estimer  $I_{\Theta_0}(\theta)$  avec  $\phi(x)$ ; lorsqu'on considère le paramètre  $I_{\Theta_0}(\theta)$  comme une mesure de la précision de test, l'estimateur  $\phi(x)$  a une interprétation que les grandes valeurs de soutient  $H_0$  et les petites valeurs de soutient  $H_1$  sont un à peu comme la p-valeur.

Bien que (2, 2) et (2, 3) définissent un problème général d'estimation dans les tests d'hypothèses, on envisage que dans les cas spéciaux on utilise les coûts de la forme :

$$L_k(\theta, \phi) = |I_{\Theta_0} - \phi(x)|^k \quad k = 1, 2 \quad (2.4)$$

Avec la fonction de risque associée :

$$R_k(\theta, \phi) = E_\theta \left( |I_{\Theta_0}(\theta) - \phi(x)|^k \right) \quad k = 1, 2$$

Si  $k = 1$

$$L_1(\theta, \phi) = |\phi - I_{\Theta_0}(\theta)|$$

avec l'estimateur de Bayes associé :

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

Si  $k = 2$

$$L_2(\theta, \phi) = (\phi - I_{\Theta_0}(\theta))^2 \quad (2.5)$$

amène à des estimateurs plus adaptatifs, puisque la proposition suivante, présente l'estimateur de Bayes sous le coût  $L_2$ .

**Proposition 2.6.1** *sous le coût  $L_2$ , l'estimateur de bayes associé à  $\pi$  est la probabilité a posteriori  $\phi^\pi(x) = P^\pi(\theta \in \Theta_0/x)$ .*

**Preuve** on a :

$$P^\pi (\theta \in \Theta_0/x) = E^\pi [I_{\Theta_0} (\theta) /x]$$

$$\begin{aligned} E^\pi [(I_{\Theta_0} (\theta) - \phi (x))^2 /x] &= E^\pi [(I_{\Theta_0} (\theta))^2 - 2\phi (x) I_{\Theta_0} (\theta) + (\phi (x))^2/x] \\ &= E^\pi [(I_{\Theta_0} (\theta))^2/x] - 2\phi (x) E^\pi [I_{\Theta_0} (\theta) /x] + [\phi (x)]^2 \\ &= \phi^2 (x) - 2E^\pi [I_{\Theta_0} (\theta) /x] \phi (x) + E^\pi [(I_{\Theta_0} (\theta))^2/x] \end{aligned}$$

Le minimum de coût a posteriori est effectivement atteint par

$$\phi^\pi (x) = E^\pi [I_{\Theta_0} (\theta) /x] = P^\pi (\theta \in \Theta_0/x)$$

■

En effet , la moyenne a posteriori de  $I_{\Theta_0} (\theta)$  donne exactement la probabilité a posteriori ; le coût  $L_2$  fournit donc une approche décisionnelle justifiée qui redonne le coût “0–1” pour des estimateurs restreints à  $\{0, 1\}$  et surtout conduit à la probabilité a posteriori comme estimateur de Bayes, de tels coût seront dits **propres** (Lindley 1985). Il existe d’autres coûts propres que  $L_2$  ; mais Hwang et Pemantle (1990) ont montré qu’il suffit d’examiner le coût quadratique en terme d’admissibilité et de classe complète ; bien qu’il existe une connexion technique entre l’approche de la théorie de décision avec le coût d’erreur absolue de la théorie de Neyman-Pearson, les réponses sont différentes dans la théorie de Neyman-Pearson où l’objectif est de maximiser la puissance pour le seuil  $\alpha$  fixé, alors qu’ici l’objectif est d’estimer  $\mathbf{I}_{\Theta_0} (\theta)$  en utilisant le coût  $L_2$ .

Dans notre test d’hypothèse nous évaluons  $\phi (x)$  comme un estimateur de  $\mathbf{I}_{\Theta_0} (\theta)$  par l’utilisation de coût (2,4).

Un autre apport important est l’interprétation de l’intervalle de confiance usuel en termes naturels. Dans le cadre bayésien, cet intervalle est habituellement appelé "intervalle de crédibilité", ce qui rend compte explicitement de la différence d’interprétation.

## 2.7 Les régions de crédibilités

Une perspective conduit à la même étude critique des régions usuelles ; à la suite de Stein (1962) et Lindley (1962), Brown (1966) et Joski (1967) ont montré que ces régions  $C_x^0$  ne sont pas optimales au sens qu’elles existent telle que :

$$P_0 (\theta \in C'_x) \geq P_\theta (\theta \in C_x^0) \quad \text{et} \quad Vol (C'_x) \leq Vol (C_x^0)$$

Par conséquent, l'ensemble  $C_x^c$  est préférable puisque pour un volume inférieur ou égale, il donne une plus grande probabilité de contenir la vraie valeur du paramètre.

**Définition 2.7.1** *pour une loi a priori  $\pi$ , un ensemble  $C_x$  est dit région  $\alpha$ -crédible. Si  $P^\pi(\theta \in C_x/x) \geq 1 - \alpha$  cette région est dite **région  $\alpha$ -crédible PHDP** (pour plus haute densité a posteriori) se peut s'écrire sous la forme*

$$\{\theta, \pi(\theta/x) > k_\alpha\} \subset C_x^\pi \subset \{\theta, \pi(\theta/x) \geq k_\alpha\}$$

où  $k_\alpha$  est la plus grande borne telle que :

$$P^\pi(\theta \in C_x^\alpha/x) \geq 1 - \alpha$$

Le fait de se restreindre à la région PHDP est motivé par une recherche de précision maximale, puisque les ensembles PHDP minimisent le volume parmi les régions  $\alpha$ -crédibles.

**exemple 2.7.1** *si  $\theta \sim N(0, \tau^2)$  et  $\theta/x \sim N(\mu(x), \rho^{-2})$  Avec  $\rho^2 = \tau^{-2} + \sigma^{-2}$  et  $\mu(x) = \frac{\tau^2 x}{\tau^2 + \sigma^2}$  alors :*

$$C_\alpha^\pi = [\mu(x) - k_\alpha \rho^{-1}; \mu(x) + k_\alpha \rho^{-1}]$$

où  $k_\alpha$  est le fractile d'ordre  $\frac{\alpha}{2}$  de  $N(0, 1)$ ; en particulier si  $\tau$  tend vers  $+\infty$ ,  $\pi(\theta)$  tend vers la loi uniforme généralisée et donne

$$C_\alpha = [x - k_\alpha \sigma; x + k_\alpha \sigma]$$

qui est un intervalle usuel.

L'intervalle de crédibilité à 95% est :

$$C_{0.05} = [x - 1.96\sigma; x + 1.96\sigma]$$

## 2.8 La considération de la fonction de coût

Pour le problème de test d'hypothèse

$$H_0 : \theta \in \Theta_0 \quad \text{contre} \quad H_1 : \theta \in \Theta_0^c \tag{2.6}$$



## 2.8. LA CONSIDÉRATION DE LA FONCTION DE COÛT

---

nous étudions maintenant les formes raisonnables pour la fonction de coût  $L(\theta, \phi)$  pour évaluer la valeur de l'estimateur  $\phi(x)$  de  $I_{\Theta_0}(\theta)$ , (voir Juinn Tzon Hwang, George Casella, Christian Robert, Martin T, Wells et Roger H. Farrell 1992). Depuis que notre paramètre intéressé eu plus que deux valeurs, la fonction de coût est donnée par :

$$L(\theta, \phi) = \begin{cases} L(1, \phi(x)) & \text{si } \theta \in \Theta_0 \\ L(0, \phi(x)) & \text{si } \theta \notin \Theta_0 \end{cases}$$

Le coût a posteriori attendu pour les fonctions de coût précédentes sachant  $X = x$  est :

$$\begin{aligned} E(L(\theta, \phi(x)) / X = x) &= \int_{\Theta} L(\theta, \phi(x)) \pi(\theta/x) d\theta \\ &= L(1, \phi(x)) P(\theta \in \Theta_0/x) + L(0, \phi(x)) P(\theta \in \Theta_0^c/x) \end{aligned}$$

où

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

Et la distribution a posteriori est :

$$P(\theta \in \Theta_0/x) = \int_{\Theta_0} \pi(\theta/x) d\theta$$

La borne stratégique bayésienne dit la vérité, et le bon estimateur de  $\mathbf{I}_{\Theta_0}(\theta)$  est la bonne évaluation bayésienne de la probabilité de son apparition, donc  $L(\theta, \phi(x))$  est propre si :

$$\min_{\phi(x)} E[L(\theta, \phi(x)) / X = x] = E[L(\theta, P(\theta \in \Theta_0/x)) / X = x]$$

Comme exemple, il y a un risque équivalent entre les tests randomisés et l'estimateur de  $\mathbf{I}_{\Theta_0}(\theta)$ . Ce fait explique en partie pourquoi  $L(\theta, \phi(x)) = |\phi(x) - I_{\Theta_0}(\theta)|$  ce qui conduit à la solution de Bayes 0-1. Cette équivalence

est simple à voir si on écrit le risque de la règle de décision  $\phi(x)$  comme suit :

$$\begin{aligned} R(\theta, \phi(x)) &= \int_{-\infty}^{+\infty} |I_{\Theta_0}(\theta) - \phi(x)| f(x/\theta) dx \\ &= I_{\Theta_0^c}(\theta) \int_{-\infty}^{+\infty} \phi(x) f(x/\theta) dx + I_{\Theta_0}(\theta) \int_{-\infty}^{+\infty} (1 - \phi(x)) f(x/\theta) dx \end{aligned}$$

Il est également possible dans les conditions de régularité d'établir une réciproque, c'est que le coût  $L_1$  est le seul coût en vertu duquel il y a une correspondance entre l'estimateur de  $\mathbf{I}_{\Theta_0}(\theta)$  et les tests randomisés.

Le fait que le coût  $L_1$  ce qui est si étroitement lié au coût 0–1 de Neymann Pearson conduit au les estimateurs qu'ils ne seront pas lisses.

## 2.9 Deux optimalités : minimaxité et admissibilité

Cette section est consacrée à deux notions fondamentales de la théorie de la décision bayésienne à savoir la minimaxité et l'admissibilité. la logique fréquentistes n'est pas assez réductrice pour conduire à un seul estimateur optimal, bien que nous nous intéressions surtout aux aspects bayésiens de la théorie de la décision, Il est nécessaire malgré tout d'étudier ces notions en détail parce qu'elles montrent que les estimateurs de Bayes sont souvent optimaux (minimax et admissibles) et devraient donc être utilisés même lorsque l'information a priori est omise.

### 2.9.1 La minimaxité

Pour démontrer la règle minimax sous le coût  $L_1$  ou  $L_2$  nous déclarons le résultat dans le théorème suivant :

**Théorème 2.9.1** *pour le problème de test d'hypothèse (2,6) avec la densité continue  $f(x/\theta)$  et le coût (2,4), on suppose que  $\Theta_0$  et  $\Theta_1$  ont un point limite commun. Si  $0 \leq \phi \leq 1$  est une décision*

$$\max_{\theta} E_{\theta}(L_k(\theta, \phi)) \geq \max_{\theta} E_{\theta}(L_k(\theta, \phi_0))$$

## 2.9. DEUX OPTIMALITÉS : MINIMAXITÉ ET ADMISSIBILITÉ

---

où  $\phi_0(x) = \frac{1}{2}$  , si  $k > 1$ , donc  $\phi_0$  est unique minimax par conséquent admissible .

**Preuve** la démonstration utilise les applications standards du théorème de convergence bornée et l'inégalité de Liapouov et il est valable pour tout  $k \geq 1$  car si  $\phi$  est dominée par une autre décision  $\phi_0$ , on dit que  $\phi$  est inadmissible; c-à-d :

$$L_k(\theta, \phi) \geq L_k(\theta, \phi_0) , \forall \theta, \exists \theta' : L_k(\theta', \phi) > L_k(\theta', \phi_0)$$

D'autre part, on dit qu'une partie  $\psi_0$  ( $\psi_0 \subset \Psi$ ), est complète si,  $\forall \phi \notin \psi_0$ , il existe  $\phi_0 \in \psi_0$  qui domine  $\phi$  donc :

$$\begin{aligned} |I_{\Theta_0}(\theta) - \phi(x)|^k &\geq |I_{\Theta_0}(\theta) - \phi_0(x)|^k \\ \int_x L_k(\theta, \phi) f(x/\theta) dx &\geq \int_x L_k(\theta, \phi_0) f(x/\theta) dx \end{aligned}$$

donc

$$E_\theta [L_k(\theta, \phi)] \geq E_\theta [L_k(\theta, \phi_0)]$$

où

$$\max_\theta E_\theta [L_k(\theta, \phi)] \geq \max_\theta E_\theta [L_k(\theta, \phi_0)]$$

l'unicité de  $\phi_0$  résulte de la stricte convexité du coût si  $k > 1$ .

Il est intéressant de noter que la p-valeur est un estimateur minimax sous  $L_1$ , donc malgré que la minimaxité ne prouve pas une propriété intéressante pour  $L_2$ , il fournit une propriété optimale de la p-valeur sous  $L_1$ . ■

### 2.9.2 Le théorème de classe complète sous $L_2$

On caractérise la classe complète de règle de décision pour les deux problèmes de test unilatéral et bilatéral

**Unilatéral**  $H_0 : \theta \leq \theta_0$  contre  $H_1 : \theta > \theta_0$

**Bilatéral**  $H_0 : \theta \in [\theta_0, \theta_1]$  contre  $H_1 : \theta \in [\theta_0, \theta_1]^c$ .

Où  $\theta_0$  et  $\theta_1$  sont des spécifiées.

Pour le théorème des classes complètes, on considère la famille d'exponentielle. On observe  $X = x$ , où  $X$  a la densité dans la famille d'exponentielle uni paramètre, depuis les estimateurs qui sont des fonctions de la statistique

$T(x)$  forment une classe complète. On suppose que les fonctions de densités définies sur  $\mathbb{R}$  l'éventail de  $T$  sont :

$$f(t/\theta) = \exp\{\theta t - \psi(\theta)\} \quad (2.7)$$

où

$$\exp \psi(x) = \int_{\mathbb{R}} \exp\{\theta t\} d\mu(t) , \theta \in \Theta \quad (2.8)$$

Car les résultats sont valides dans le problème de cas discret et continu. Les règles de la classe complète sont généralisées essentielles. L'estimateur de Bayes après la modification de l'espace de paramètre pour la troncation qui nous le définit maintenant dans le problème unilatéral. L'intervalle  $[t_1, t_2]$  est un ensemble de troncation pour l'estimateur  $\phi$ .

Si  $t < t_1$  implique  $\phi(t) = 1$  et  $t > t_2$  implique  $\phi(t) = 0$

Dans le problème bilatéral l'intervalle  $[t_1, t_2]$  est un ensemble de troncation pour l'estimateur  $\phi$  si  $t \in [t_1, t_2]^c$  implique  $\phi(t) = 0$ .

**Théorème 2.9.2** dans le problème bilatéral

$$H_0 : \theta \in [\theta_0; \theta_1] \quad \text{contre} \quad H_1 : \theta \in [\theta_0; \theta_1]^c \quad (2.9)$$

Un estimateur  $\phi$  d'ensemble de troncature  $[t_1; t_2]$  est admissible sous le coût  $L_2$ , s'il existe une mesure de probabilité  $\pi_0$  sur  $[\theta_0; \theta_1]$  et une mesure  $\sigma$ -finie  $\pi_1$  sur  $(-\infty; \theta_0] \cup [\theta_1; +\infty)$  pour tout  $t_1 < t < t_2$

$$\int f(t/\theta) \pi_1(d\theta) < \infty \quad (2.10)$$

Et

$$\phi(t) = \frac{\int f(t/\theta) \pi_0(d\theta)}{\int f(t/\theta) \pi_0(d\theta) + \int f(t/\theta) \pi_1(d\theta)} \quad (2.11)$$

Contrairement si  $\phi$  est admissible donc il existe un ensemble de troncation  $[t_1; t_2]$ , une mesure de probabilité  $\pi_0$  sur  $[\theta_0; \theta_1]$  et une mesure  $\sigma$ -finie  $\pi_1$  sur  $(-\infty; \theta_0] \cup [\theta_1; +\infty)$  de telle sorte que (2, 10) et (2, 11) pour tout  $t \in (t_1; t_2)$ .

**Preuve** Premièrement on établit quelques notions préliminaires et les lemmes suivants :

## 2.9. DEUX OPTIMALITÉS : MINIMAXITÉ ET ADMISSIBILITÉ

---

**Lemme 2.9.1** *Pour le coût  $L_2$ , si  $\phi'$  est bon que  $\phi$  et  $\phi$  a un ensemble de troncation  $[t_1; t_2]$ , donc  $\phi(t) = \phi'(t)$  pour  $t < t_1$  et  $t > t_2$ .*

Pour montrer le bon résultat possible, il est nécessaire de modifier l'espace de paramètres. On discute de cela pour le problème bilatéral; soit :

$$I_0 = \{(0, \theta) : \theta_0 \leq \theta \leq \theta_1\}$$

et

$$I_1 = \{(1, \theta) : \theta \leq \theta_0 \text{ ou } \theta \geq \theta_1\}$$

definit

$$R_1(\theta, \phi) = E(1_{[\theta_0, \theta_1]}(\theta) - \phi(x))^2$$

$$R_2[(i, \theta), \phi] = E_\theta(1_{I_0}(0, \theta) - \phi(x))^2, \quad (i, \theta) \in I_0$$

et

$$R[(i, \theta), \phi] = E_\theta(1 - 1_{I_0}(i, \theta) - \phi(x))^2, \quad (i, \theta) \in I_1$$

car  $R_2(\cdot, \phi)$  peut être obtenue comme limite de  $R_1(\cdot, \phi)$ . Le lemme suivant est facile à vérifier.

**Lemme 2.9.2**  *$\phi$  est admissible pour tous risque  $R_1$  si et seulement si  $\phi$  est admissible pour le risque  $R_2$ .*

Aujourd'hui, on retourne à la démonstration du théorème précédent; pour ça, on suppose que  $\phi(x)$  est un estimateur admissible, pour Brown [(1986), proposition 4A, 7 et théorème 4A, 12] il existe une séquence a priori finie  $(u_{0j}, u_{1j})$  concentrée sur les sous ensembles finis tels que les estimateurs de Bayes  $\phi^{u_j}(t)$  convergent vers  $\phi(t)$  presque sûrement. Le cas spécial  $\phi \equiv 0$  est évident, et suppose que  $E(\phi[T(X)]) > 0$ , par le théorème de convergence dominé

$$\lim_{j \rightarrow \infty} E_\theta \phi^{u_j}[T(X)] = E_\theta \phi[T(X)]$$

de sorte que  $j$  soit grand,  $\phi^{u_j}(t) > 0$  avec mesure positive, par conséquent  $u_{0j}([\theta_0; \theta_1]) \succ 0$  et par la renormalisation, on suppose que  $u_{0j}([\theta_0; \theta_1]) = 1$ .

L'ensemble convexe  $C = \{t, \lim_{j \rightarrow \infty} \sup \int \exp(t\theta - \psi(\theta)) u_{1j}(d\theta) < \infty\}$

est un intervalle. Clairement de (2, 11) pour presque tout  $t$  n'étant pas dans

$C'$ ,  $\phi(t) = 0$ ; suppose  $C = [t_1, t_2]$  être la fermeture de  $C'$  et utilise  $C$  comme une ensemble de troncation à partir de (2, 11), pour presque tout  $t_1 < t < t_2$  :

$$0 \prec \lim_{j \rightarrow \infty} \phi^{u_j}(t) = \phi(t)$$

Si nécessaire par le choix de séquence, on peut supposer que  $u_{0j}$  converge vers la mesure de probabilité  $u_0$  faiblement tel que :

$$\lim_{j \rightarrow \infty} \int f(t/\theta) u_{0j}(d\theta) = \int f(t/\theta) u_0(d\theta)$$

Il s'ensuit, pour presque tout  $t_1 < t < t_2$ , que :

$$\lim_{j \rightarrow \infty} \int f(t/\theta) u_{1j}(d\theta)$$

existe et elle est finie, donc si  $t_1 < t_2$ , donc l'argument standard peut être utilisé pour voir qu'il existe une mesure de limitation  $\sigma$ -fini  $u_1$ , tel que : si  $u_{1j} \rightarrow u_1$  et  $t_1 < t < t_2$  donc

$$\lim_{j \rightarrow \infty} \int \exp(\theta t - \psi(\theta)) u_{ij}(d\theta) = \int \exp(\theta t - \psi(\theta)) u_j(d\theta) \quad i = 1, 2$$

Donc pour presque tout  $t_1 < t < t_2$ ,  $\phi(t)$  peut être exprimée comme dans (2, 11) , instituant la deuxième partie du théorème, la première partie du théorème résulte du lemme 2 et l'unicité de l'estimateur de Bayes généralisé comme un minimisant du risque de Bayes généralisé. ■

Comme mentionné précédemment le point essentielle du théorème précédent et la classe complète ont été donnés par l'estimateur de Bayes généralisé ; dans la partie suivante on applique cette stratégie sur la p-valeur.

### 2.9.3 La considération d'admissibilité sous le coût $L_2$

Ici, on retourne à l'exploration du comportement de la p-valeur, et trouver des résultats intéressants sous le coût  $L_2$ .

Dans le problème du test unilatéral, la p-valeur est dans plusieurs cas admissible sous le coût  $L_2$ , et la p-valeur est une mesure raisonnable de la précision. Et dans le cas bilatéral la p-valeur est inadmissible mais difficile à dominer. Malgré que la p-valeur est inadmissible, elle ne peut pas être dominé par aucun bon estimateur de Bayes.

**Les exemples d'admissibilité dans le problème unilatéral**

Dans les deux exemples suivants, l'admissibilité de la p-valeur peut être simple à établir.

**Théorème 2.9.3** pour le problème de test d'hypothèse unilatéral de

$$H_0 : \theta \leq \theta_0 \quad \text{contre} \quad H_1 : \theta > \theta_0$$

Avec la fonction de coût

$$L_2 = (\phi - I_{\Theta_0}(\theta))^2$$

Soit  $X_1, X_2, \dots, X_n$  être (i.i.d)  $n(\theta, 1)$  la p-valeur

$$p(\bar{x}) = p_{\theta_0}(\bar{X} > \bar{x}) = 1 - \Phi(\sqrt{n}(\bar{x} - \theta_0))$$

est admissible, où  $\bar{X}$  est la moyenne de  $X_1, X_2, \dots, X_n$  avec la valeur observée  $\bar{x}$ .

**Preuve** utilisation de la suffisance, on peut supposer  $n = 1$ . Notons que la p-valeur est de Bayes généralisé à l'égard de la mesure de Lebesgue avant [ il y a aussi une limite des estimateurs de Bayes contre la séquence de l'a priori  $n(\theta_0, r)$  lorsque  $r \rightarrow \infty$  ]. En outre, le risque (généralisé) de la p-valeur est finie, par conséquent la p-valeur est admissible.

Nous avons maintenant établi l'admissibilité de la p-valeur pour certaines distributions discrètes en utilisant une méthode similaire, nous résumons ces résultats dans le théorème suivant. ■

**Théorème 2.9.4** pour le problème du test bilatéral de (2,9) avec la fonction de coût  $L_2$ .

a / si  $f(x/\theta)$  est une binomiale des paramètres  $(n, \theta)$ , la p-valeur  $p(x)$  est :

$$p(x) = p_{\theta_0}(X \geq x) = \sum_{k=x}^n \binom{n}{k} \theta_0^k (1 - \theta_0)^{n-k}$$

est admissible.

b / si  $f(x/\theta)$  est une poisson du paramètre  $(\theta)$ , la p-valeur  $p(x)$  est :

$$p(x) = p_{\theta_0}(X \geq x) = \sum_{k=x}^{\infty} e^{-\theta_0} \theta_0^k / k!$$

est admissible.

**Preuve** pour (a) : envisager la généralisation  $\frac{1}{\theta}$  densité a priori qui correspond à l'estimateur de Bayes généralisé  $p(x)$ . Le fait que cet estimateur a de risque de bayes généralisé finie découle du fait que l'estimateur

$$\delta(x) = \begin{cases} 1 & \text{si } x = 0 \\ 0 & \text{sinon} \end{cases}$$

A de risque de Bayes généralisé fini, ainsi  $p(x)$ .

Pour (b) : peut être établie similairement en considérant une nouvelle densité généralisé a priori  $\frac{1}{\theta}$ .

Dans les deux cas de binomial et poisson, la généralisation à un échantillon iid est immédiate, en plus dans le cas de problème du test unilatéral, la p-valeur est admissible comme un estimateur de  $I_{]-\infty; \theta_0[}(\theta)$ . ■

### 2.9.4 L'admissibilité dans le problème bilatéral

Le théorème de la classe complète (**théorème 2.9.2**) nous donne un outil puissant pour explorer l'admissibilité de la p-valeur dans le problème de bilatéral. Le théorème suivant nous donne des conclusions décisives sur l'inadmissibilité de la p-valeur et c'est une corollaire du théorème 2.9.2.

**Théorème 2.9.5** *Pour le problème des tests d'hypothèse bilatéral*

$$H_0 : \theta \in [\theta_0; \theta] \quad \text{contre} \quad H_1 : \theta \in [\theta_0; \theta_1]^c$$

*Avec la fonction de coût  $L_2$ , on suppose que l'estimateur  $\varphi(T(x)) > 0$  est continu, non constant et pour quelque valeur  $x_0$   $\varphi(T(x_0)) = 1$ ; donc  $\varphi$  est inadmissible.*

**Preuve** Si  $\varphi$  était admissible alors presque surement (2, 10) depuis  $\varphi > 0$ .

Pour tout  $x$   $\int f(x/\theta) \pi(d\theta) < \infty$  pour presque tout  $x$ , donc le bilatéral de

(2, 10) est continu en  $x$  et donc l'égalité pour tous  $x$  dans le support de  $X$ , depuis  $f(x/\theta) > 0$ ,  $\varphi(T(x_0)) = 1$  pour tout  $x$ , une contradiction. ■

Le résultat du théorème précédent nous permet aujourd'hui de répondre à la question d'admissibilité de la p-valeur.

Malgré que le théorème 7 est négatif à son évaluation de la p-valeur, aujourd'hui on va voir un cas spécial de test d'hypothèse de point nul sur



## 2.9. DEUX OPTIMALITÉS : MINIMAXITÉ ET ADMISSIBILITÉ

une moyenne normale et on trouve que la p-valeur ne peut être dominée par toute autre procédure appropriée de Bayes. Donc comme la p-valeur est inadmissible pour le test d'hypothèse de point nul, il est difficile d'exposer le bon estimateur et il suffit de considérer le cas d'une observation.

**Théorème 2.9.6** *pour tester l'hypothèse*

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

*Basé sur une observation  $X$  de la densité  $n(\theta, 1)$  et utiliser la fonction de coût  $L_2$ . La p-valeur ne peut être dominé par aucun règle de Bayes.*

**Preuve** En supposant, sans le coût généralisé, que  $\theta = 0$ , l'estimateur de Bayes pour ce problème est de la forme :

$$\phi^\pi(x) = \frac{\pi_0 f(x/0)}{\pi_0 f(x/0) + (1 - \pi_0) \int f(x/\theta) g(\theta) u(d\theta)}$$

où  $f(x/\theta)$  est la densité  $n(\theta, 1)$  et  $\int_{-\infty}^{+\infty} g(\theta) u(d\theta) = 1$ , voir (la section 2 de

Juinn Tzon Hwang, George Casella, Christian robert (1992)), et la p-valeur est donnée par

$$p(x) = p(|X| \geq x) = 2(1 - \phi(x))$$

On considère les 3 cas :

Cas 1 :  $\pi_0 = 1$ , dans ce cas  $\phi^\pi(x) = 1$ , lorsque  $\theta \rightarrow \infty$ ,  $R(\theta, p(x)) \rightarrow 0$  mais  $R(\theta, \phi^\pi(x)) = 1$ , pour  $\theta \neq 0$ , donc  $\phi(x)$  ne peut pas dominer  $p(x)$ .

Cas 2 :  $\pi_0 = 1$ , dans ce cas  $\phi^\pi(x) = 0$ , donc  $R(0, \phi^\pi) = 1 > R(0, p(x)) = \frac{1}{3}$ , et  $\phi^\pi(x)$  ne peut pas dominer  $p(x)$ .

Cas 3 :  $0 < \pi_0 < 1$ , nous montrons que lorsque  $\theta \rightarrow \infty$ ,  $R(\theta, p(x))$  devient plus petit que  $R(\theta, \phi^\pi(x))$ . Premièrement on note que  $|x| > a > 0$ ,  $\phi^\pi(x) > p(x)$ , ceci découle de fait que :

$$\phi^\pi(x) \geq \frac{\pi_0 f(x/\theta_0)}{\pi_0 f(x/\theta_0) + (1 - \pi_0) f(x/\theta)} > p(x) \quad (2.12)$$

pour  $|x|$  suffisamment grand, où  $\hat{\theta} = x$  est l'estimateur de maximum de vraisemblance de  $\theta$ . Pour  $\theta \neq \theta_0$ , la différence de risque est donnée par

$$E_\theta(I_{\{\theta_0\}}(\theta) - \phi^\pi(x))^2 - E(I_{\{\theta_0\}}(\theta) - p(x))^2 = E_\theta(\phi^\pi(x)^2 - p(x)^2)$$

et pour (2,12), par continuité il existe  $\varepsilon > 0$ , tel que  $\phi^\pi(x)^2 - p(x)^2 > \varepsilon$  pour toute  $a < |x| < a + \varepsilon$ , par conséquent

$$E_\theta (\phi^\pi(x)^2 - p(x)^2) \geq \varepsilon P_\theta(a < |x| < a + \varepsilon) - P_\theta(|x| < a)$$

Cette borne inférieure est positive pour  $\theta$  grand

$$\frac{P_\theta(a < |x| < a + \varepsilon)}{P_\theta(|x| < a)} \rightarrow \infty, \text{ lorsque } \theta \rightarrow \infty$$

Par la règle de L'hôpital et par conséquent la différence de risques est strictement positive pour  $\theta$  grand et  $\phi^\pi(x)$  ne peut pas dominer  $p(x)$ . ■

# Chapitre 3

## Méthodes adaptatives bayésiennes dans les essais cliniques

### 3.1 Introduction

La recherche clinique est une activité médicale visant à améliorer la connaissance soit d'une maladie, soit d'une thérapeutique. La recherche clinique concerne l'être humain. En pharmacologie, la recherche clinique est dominée par les études du médicament administré à l'homme dans le cadre des essais cliniques (essais cliniques ou évaluations cliniques).

Chronologiquement, les essais cliniques se déroulent après les études dites de pharmacologie expérimentale, qui se déroulent en laboratoire (stade pré-clinique). Les études chez l'homme obéissent à une technique (méthodologie des essais cliniques), à une législation et à une éthique. Ces études peuvent se dérouler soit en médecine de ville, soit dans les centres hospitaliers, soit dans des structures de recherche agréées publiques ou privées.

Pour une classe de loi a priori conjuguée nous donnons les distributions a posteriori et les intervalles de crédibilité a posteriori des paramètres d'intérêt nécessaires pour montrer la supériorité d'un traitement ou l'équivalence des traitements. Nous explicitons aussi les distributions prédictives qui permettent de choisir les effectifs ou de prendre une décision d'arrêter l'expérience dans le cas des analyses intermédiaires, voir (**Khadidja ELQASYR (2008)**).

## 3.2 Les essais cliniques

Les essais cliniques sont des études médicales statistiques effectuées chez l'homme pour prouver la validité d'un nouveau traitement. La réalisation d'essais cliniques est une étape essentielle du développement de nouveaux médicaments. Un essai clinique permet d'étudier et de s'assurer des propriétés de molécules afin de pouvoir en faire, par la suite, un usage sécurisé et responsable. Ces molécules présenteront, peut être, un nouveau progrès thérapeutique pour des centaines de milliers de patients et pour le corps médical.

Les essais cliniques sont le plus souvent pratiqués en recherche médicale ou biomédicale, mais d'autres disciplines cliniques, telle la psychologie, mènent des travaux de recherche dans le but d'évaluer - ou habituellement de comparer - des interventions. Les essais cliniques sont envisagés sous l'angle de la recherche biomédicale et surtout des essais pharmaceutiques. Les recherches qui réalisent des essais cliniques poursuivent divers objectifs de différentes façons. De tels essais peuvent comprendre des questions indirectement reliées à une thérapie (rapport coût - efficacité, méthodisme des médicaments, *...* *etc.*).

## 3.3 Les différentes phases d'un essai clinique

Ce n'est qu'après les multiples étapes du développement préclinique que les premiers essais thérapeutiques sur l'homme peuvent être réalisés. On parle alors de développement clinique. Les essais cliniques (ou essais thérapeutiques) sont une étape obligatoire et systématique du développement d'un médicament. Ils permettent de préciser l'effet d'un traitement chez l'homme, d'en déterminer l'efficacité ainsi que les éventuels effets indésirables.

On distingue 4 phases dans les essais cliniques.

### Phase I

Lors de la **phase I**, les essais sont généralement réalisés chez le volontaire sain, ces essais ont lieu dans des centres spécialisés qui ont reçu un agrément de la part des autorités de santé. Ces études ont deux objectifs majeurs :

*Premièrement* : il s'agit de s'assurer que les résultats concernant la toxicité obtenus lors du développement préclinique, sont comparables à ceux obtenus chez l'homme. Cela permet de déterminer quelle est la dose maximale du médicament en développement tolérée chez l'homme.

*Deuxièmement* : il s'agit de mesurer via des études de pharmacocinétique, le devenir du médicament au sein de l'organisme en fonction de son mode

d'administration (absorption, diffusion, métabolisme et excrétion).

#### **Phase II**

Les essais de **phase II** ont pour objectif de déterminer la posologie optimale du produit en terme d'efficacité, et de tolérance sur une population limitée et homogène de patients (quelques centaines). Les interactions médicamenteuses ainsi que la pharmacocinétique font parfois l'objet d'études dès cette phase.

#### **Phase III**

Ces essais de plus grande envergure sont conduits sur plusieurs milliers de patients représentatifs de la population de malades à laquelle le traitement est destiné.

Il s'agit d'essais comparatifs au cours desquels le médicament en développement est comparé à un traitement efficace déjà commercialisé ou, dans certains cas, à un placebo, c'est à dire un traitement sans activité pharmacologique.

Cette comparaison se fait le plus souvent, en double insu et avec tirage au sort, c'est-à-dire que les traitements sont attribués de manière aléatoire sans que le patient et le médecin chargé du suivi soient informés de quelle attribution ils font l'objet. Ces essais visent à démontrer l'intérêt thérapeutique du médicament. C'est à l'issue de la **phase III** que les résultats peuvent être soumis aux autorités algériennes de santé pour l'obtention de l'autorisation de commercialisation.

#### **Phase IV**

Les essais de **phase IV** sont réalisés une fois le médicament commercialisé, sur un nombre de patients souvent très important (jusqu'à plusieurs dizaines de milliers de personnes). Ils permettent d'approfondir la connaissance du médicament dans les conditions réelles d'utilisation et d'évaluer à grande échelle sa tolérance.

La pharmacovigilance permet ainsi de détecter des effets indésirables très rares qui n'ont pu être mis en évidence lors des autres phases d'essai.

## 3.4 L'adaptativité dans les essais cliniques

La majeure partie de cette section est consacrée à la détractation de la subtilité de l'approche bayésienne, et la distinction des approches fréquentistes correspondantes. Cependant, nous nous arrêtons ici pour décrire ce que nous entendons par le mot "**adaptative**". Il y a un grand nombre d'inno-

vations récentes en matière d'essais cliniques qui vont sous ce nom à la fois fréquentiste et bayésien mais il ne sera peut être pas surprenant à ce stade que les deux camps affichent le terme plutôt différemment.

Préoccupé comme il se doit de l'erreur de première espèce, les fréquentistes ont parfois appelé à une procédure pour modifier les frontières d'arrêt au fil du temps, tout en protégeant l'ensemble du taux d'erreur de première espèce notée comme "**adaptative**". Plus récemment, plus récemment les fréquentistes et les bayésiens donnent une procédure sur la base des résultats de l'essai. Mais bien sûr il s'agit d'un sérieux changement dans le design expérimental, et donc celui qui doit être reflété dans le calcul d'erreur de première espèce. En particulier, nous n'avons pas besoin de sélectionner la taille de l'échantillon de première instance à l'avance (même si une taille maximale de l'échantillon est souvent donnée). Mais tout cela soulève aussi la question de la connaissance de quelques sortes d'adaptation que nous envisageons dans nos épreuves. Ces deux questions sont liées, puisque la tâche à accomplir dépend sur la phase. Mais il est certainement vrai que lors de la rédaction actuelle, un grand nombre de taille d'échantillon non fixé est en cours d'exécution à travers une variété de phases du processus de réglementation. Dans les études de phase précoce, il semble plus important au point de vue éthique d'être adaptative, car les patients sont souvent très malades, ce qui impose un changement soudain de traitement à la fois plus nécessaire et peut être plus fréquentiste.

L'étude de **phase I** de drogue est généralement sur la sécurité. Ce qui signifie dans ce dernier cas que la dose personnelle que reçoit le patient n'est pas fixée à l'avance, mais plutôt déterminée par les résultats observés chez les patients traités à ce jour, mais se révèle être en **phase II**. Nous avons généralement chercher à établir l'efficacité lorsque tous les patients peuvent encore être protégés contre la toxicité de l'excès et encore contre la fatalité, c-à-d poursuivant un essai qui ne produira vraisemblablement jamais un résultat significatif. Même si tous les patients disponibles pour les mesures sont inscrits, on va discuter ci-dessous à nouveau l'adaptativité et arrêter l'essai plus tôt si l'une des trois conclusions (efficacité, toxicité ou futilité) peut être atteinte au début. Dans la **phase III**, le besoin d'adaptativité doit être réduit mais les traitements éthiques des patients utilisant ses données requièrent de la flexibilité. L'essai particulier doit commencer par des doses multiples pour une situation particulière, et avec l'intention qui consiste en deux phases consécutives : La première détermine la dose appropriée, et la deuxième pour comparer cette efficacité avec les références standards, et alors

il est possible de décider d'éliminer entièrement la **phase III**.

Finalement, l'adaptativité est nécessaire dans les essais cliniques, pour adapter la dose, la fraction de randomisation, la taille totale de l'échantillon...etc.

### 3.5 La comparaison entre l'approche bayésienne et fréquentiste dans les essais cliniques

D'après (**Shein-Chung Chow (2010)**), cette section porte sur certaines des différences entre les approches bayésienne et fréquentiste. La liste des différences donne une vue unilatérale, néanmoins il y a de nombreuses similitudes entre les deux approches. Par exemple, les deux approches reconnaissent la nécessité de contrôles lors de l'évaluation d'un traitement expérimental. Pourtant, avec cet avertissement, voici quelques différences principales entre les deux approches :

**1/ Probabilités des paramètres :** Tous les inconnus ont des distributions de probabilité dans l'approche bayésienne. Mais dans l'approche fréquentiste, les probabilités sont définies seulement sur l'espace de données. En particulier les distributions de probabilité bayésiennes sont appelées distributions a priori et a posteriori. Les premières informations sur les paramètres sont résumées des valeurs fixées avant que les données ne soient recueillies, et que la dernière condition sur les données sont une fois observées.

**2/ Utilisation de toutes les preuves disponibles :** La quantité fondamentale d'inférence dans l'approche bayésienne est la distribution a posteriori des différents paramètres inconnus. Cette distribution dépend de toutes les informations actuellement disponibles sur ces paramètres. En revanche, les mesures fréquentistes sont spécifiques à une expérience particulière. Cette différence rend l'approche bayésienne plus attrayante dans ce sens. Une approche de la combinaison de données est la modélisation hiérarchique. Ceci est particulièrement facile à mettre en oeuvre à partir d'un point de vue bayésien, et conduit à l'emprunt de la résistance estimative à travers des expériences similaires mais indépendantes. L'utilisation de modèles hiérarchiques pour combiner les données des études est une approche bayésienne de méta-analyse.

**3/ Condition sur les résultats effectivement observés :** Les inférences bayésiennes dépendent de la présente étude que par les données réelle-

ment observées, tandis que les mesures fréquentistes impliquent des probabilités des données (calcul par le conditionnement sur des valeurs particulières des paramètres inconnus) qui étaient possibles compte tenu de la conception de l'essai, mais qui n'ont pas été observées réellement.

**4/ Flexibilité :** Les inférences bayésiennes sont flexibles en ce sens qu'elles peuvent être mises à jour continuellement à mesure que les données s'accumulent. Par exemple, la raison pour arrêter un essai affecte des mesures fréquentistes mais pas des inférences bayésiennes. Les mesures fréquentistes exigent une expérience complète qui réalisée en fonction de la conception pré-définie. Certains fréquentistes ne sont pas entravés par de telles restrictions et raisonnablement faites, mais les conclusions résultantes ne sont pas des interprétations déductives claires. Dans une approche Bayésienne, la taille de l'échantillon ne doit pas être choisie à l'avance. Avant un essai, la seule décision nécessaire est de savoir si oui ou non on peut le démarrer.

L'analyse Bayésienne peut toujours être possible dans des situations où les analyses ne sont pas fréquentistes.

**5/ Rôle de la randomisation :** Les essais commandés randomisés sont la norme de la recherche médicale. C'est vrai indépendamment de l'approche statistique. La randomisation réduit au minimum la possibilité de polarisation de choix, et elle tend à équilibrer les groupes de traitement sur les covariables connues et inconnues. Il y a des différences bayésiennes et fréquentistes de la randomisation. Dans ce dernier, la randomisation sert de base à l'inférence, tandis que la base pour l'inférence dans l'approche bayésienne est la probabilité subjective qui n'exige pas la randomisation.

**6/ Probabilité prédictive :** Une approche bayésienne permet de calculer la probabilité prédictive. Les probabilités de futures observations sont possibles dans une approche fréquentiste formelle que par le conditionnement sur des valeurs particulières des paramètres. La moyenne bayésienne de ces probabilités conditionnelles de plus des paramètres inconnus, utilisant le fait que la probabilité sans condition est la valeur prévue des probabilités conditionnelles.

**7/ La prise de décision :** L'approche bayésienne est idéale pour effet. La conception d'un test clinique est un problème de décision de tirer une conclusion d'un essai, telle que recommander une thérapie, pour tirer l'arrêt du développement du dispositif est un problème de décision. Il y a des coûts impliqués dans chaque problème pouvant survenir. Dans l'approche bayésienne, ces coûts peuvent être comme une évaluation pour chaque ordre possible de futures observations. La considération d'une décision particulière



donnera lieu à une parmi une série de futures observations possibles, chacune ayant des coûts et des avantages. Celles-ci peuvent être mis en balance par leurs probabilités prédictives correspondantes.

**8/ La faisabilité des méthodes bayésiennes :** Toutes les procédures fréquentistes traditionnelles peuvent être dérivées comme une extension directe des procédures de base élémentaires utilisé en statistique inférentielle. Et les procédures bayésiennes deviennent faciles à mettre en oeuvre. Nous avons tout particulièrement développé les méthodes bayésiennes dans le cadre de l'analyse de variance, qui est d'une grande importance pour l'analyse des données expérimentales. Des procédures bayésiennes ont été développées dans ce domaine, mais elle sont généralement considérées comme difficiles à mettre en oeuvre et ne sont pas incluses dans les logiciels statistiques habituels.

## 3.6 Règle d'arrêt

D'après (**Robert (2006)**). Au cours d'un essai, l'information s'accumule progressivement au fur et à mesure des inclusions et du suivi des patients, mais c'est seulement au terme de l'essai après avoir recruté l'effectif prévu, que la quantité d'information est suffisante et que les données peuvent être analysées et parfois on doit s'arrêter complètement.

### 3.6.1 Principe de la règle d'arrêt

Si une suite d'expérience  $\zeta_1, \zeta_2, \dots$  admet une règle d'arrêt  $\Upsilon$ , qui indique quand doivent les expériences s'arrêter, l'inférence sur  $\theta$  ne doit dépendre de  $\Upsilon$  qu'à travers l'échantillon résultant.

### 3.6.2 Situation conduisant à un arrêt prématuré

En fait, si l'effet réel du traitement est bien supérieur à l'effet initialement suspecté ou que le risque de base de patients inclus est bien supérieur à celui attendu, il sera possible de mettre en évidence l'effet du traitement avec moins de sujets que l'effectif prévu. Dans les deux cas, il y a eu sous-estimation de l'un ou de ces deux paramètres dans le calcul du nombre de sujets nécessaires et l'effectif initialement calculé est surdimensionné par rapport à la réalité.

Cette situation est imaginable car, assez souvent des données fiables manquent pour faire les hypothèses du calcul du nombre de sujets. Il est

alors envisageable que les valeurs retenues soient éloignées de la réalité. Les mêmes raisons expliquent aussi l'échec de certains essais qui avaient surestimés la taille de l'effet et donc sous estimé l'effectif nécessaire pour le mettre en évidence.

En effet, on doit arrêter l'essai pour les raisons suivantes : **efficacité, toxicité, futilité.**

## 3.7 Les méthodes du modèle d'adaptation dans les essais cliniques

On trouve dans (Shein-Chung, Mark Chang(2006)) que l'approche bayésienne dans les essais cliniques a un grand intérêt dans le développement de la production médicale, car elle offre une information valide.

Dans cette section, notre est l'utilisation des différentes utilités de l'approche bayésienne dans les essais cliniques.

### 3.7.1 Quelques notions de base de l'approche bayésienne dans les essais cliniques

Ici, les notions de base de l'approche bayésienne dans les essais cliniques sont représentées par la règle de Bayes, le calcul du nombre de sujets nécessaire (la taille de l'échantillon) et le calcul du la puissance bayésienne.

#### Calcul du nombre nécessaire de sujets

Ici on considère un exemple qui consiste à déterminer la taille de l'échantillon à partir de la puissance espérée. Cette approximation peut être utilisée pour calculer le nombre de sujets nécessaires.

*exemple 3.7.1 exemple 3.7.2 en considérant une conception de groupe parallèle de deux bras comparant un traitement d'essai et une thérapie standard ou un agent de contrôle actif, sous l'essai de deux bras, il peut être vérifié que la puissance est une fonction de l'effet de  $\varepsilon$*

$$puissance(\varepsilon) = \Phi_0\left(\frac{\sqrt{n_\varepsilon}}{2} - Z_{1-\alpha}\right)$$

### 3.7. LES MÉTHODES DU MODÈLE D'ADAPTATION DANS LES ESSAIS CLINIQUES

---

où  $\Phi_0$  est la fonction de répartition de la distribution normale standard. On suppose que la distribution a priori de  $\varepsilon$  est  $\pi(\varepsilon)$ , donc la puissance espérée est donnée par :

$$P_{esp} = \int \Phi_0 \left( \frac{\sqrt{n_\varepsilon}}{2} - Z_{1-\alpha} \right) \pi(\varepsilon) d\varepsilon \quad (3.1)$$

En pratique, l'intégration numérique est généralement utilisée pour l'évaluation de l'équation (3,1). Pour illustrer l'équation (3,1) on suppose l'unilatérale  $\alpha = 0.025$  (i.e ,  $Z_{1-\alpha} = 1.96$ ), et l'a priori de  $\varepsilon$  est :

$$\pi(\varepsilon) = \begin{cases} \frac{1}{3} & \text{si } \varepsilon = 0.1; 0.25; 0.4 \\ 0 & \text{sinon} \end{cases}$$

Par convention, on utilise la moyenne (médiane) de la taille effective  $\bar{\varepsilon} = 0.25$ . Pour le design d'essais et pour performance de la taille de l'échantillon, on suppose que  $\bar{\varepsilon} = 0.25$  est une vraie taille effective. Pour le design à deux bras balancé avec  $\beta = 0.2$  ou puissance = 80% l'approche classique donne la taille de l'échantillon suivante :

$$n = \frac{4(Z_{1-\alpha} + Z_{1-\beta})^2}{\varepsilon^2} = \frac{4(1.96 + 0.842)^2}{0.25^2} = 502$$

D'autre part, l'approche Bayésienne basée sur la puissance espérée de l'équation (3,1) est donnée :

$$\begin{aligned} P_{esp} &= \frac{1}{3} \left[ \Phi_0 \left( \frac{0.1\sqrt{n}}{2} - Z_{1-\alpha} \right) + \Phi_0 \left( \frac{0.25\sqrt{n}}{2} - Z_{1-\alpha} \right) + \Phi_0 \left( \frac{0.4\sqrt{n}}{2} - Z_{1-\alpha} \right) \right] \\ &= \frac{1}{3} \left[ \Phi_0 \left( \frac{0.1\sqrt{502}}{2} - 1.96 \right) + \Phi_0 \left( \frac{0.25\sqrt{502}}{2} - 1.96 \right) + \Phi_0 \left( \frac{0.4\sqrt{502}}{2} - 1.96 \right) \right] \\ &= \frac{1}{3} [\Phi_0(-0.83973) + \Phi_0(0.84067) + \Phi_0(2.5211)] \\ &= \frac{1}{3} (0.2005 + 0.7997 + 0.9942) \\ &= 0.6648 \\ &= 66\% \end{aligned}$$

Donc la puissance espérée est inférieure à la puissance désirable de 80%. Pour obtenir la même puissance, il est nécessaire d'augmenter la taille de l'échantillon. Si  $\pi(\varepsilon) \sim N\left(\mu, \frac{\sigma^2}{n_0}\right)$  et

$$\begin{aligned}
 P\left(X > \frac{1}{\sqrt{n}}Z_{1-\alpha}\sigma\right) &\iff P\left(\frac{X - \mu}{\sigma\sqrt{\frac{n_0+n}{n_0n}}} > \frac{\frac{\sigma}{\sqrt{n}}Z_{1-\alpha} - \mu}{\sigma\sqrt{\frac{n_0+n}{nn_0}}}\right) \\
 &= 1 - \Phi\left(\frac{\frac{\sigma}{\sqrt{n}}Z_{1-\alpha} - \mu}{\sigma\sqrt{\frac{n_0+n}{n_0n}}}\right) \\
 &= 1 - \Phi\left(\sqrt{\frac{n_0}{n_0+n}}Z_{1-\alpha} - \frac{\sqrt{n}\sqrt{n_0}\mu}{\sigma\sqrt{n+n_0}}\right) \\
 &= \Phi\left(\frac{\sqrt{n_0}\sqrt{n}\mu}{\sigma\sqrt{n_0+n}} - \sqrt{\frac{n_0}{n_0+n}}Z_{1-\alpha}\right) \\
 &= \Phi\left(\sqrt{\frac{n_0}{n_0+n}}\left(\frac{\sqrt{n}\mu}{\sigma} - Z_{1-\alpha}\right)\right) \\
 P_{esp} &= \Phi\left(\sqrt{\frac{n_0}{n+n_0}}\left(\frac{\mu\sqrt{n}}{\sigma} - Z_{1-\sigma}\right)\right)
 \end{aligned}$$

Avec la taille de l'échantillon totale est

$$n(\varepsilon) = \frac{4(Z_{1-\alpha} + Z_{1-\beta})^2}{\varepsilon^2}$$

Et la taille de l'échantillon espérée est :

$$n_{exp} = \int \frac{4(Z_{1-\alpha} + Z_{1-\beta})^2}{\varepsilon^2} \pi(\varepsilon) d\varepsilon$$

- Si

$$\pi(\varepsilon) \sim \frac{1}{b-a} \quad \text{où } a \leq \varepsilon \leq b$$

On a le nombre de sujets nécessaires espéré :

$$n_{esp} = \int_a^b \frac{4(Z_{1-\alpha} - Z_{1-\beta})^2}{\varepsilon^2} \cdot \frac{1}{b-a} d\varepsilon = \frac{4}{ab} (Z_{1-\alpha} + Z_{1-\beta})^2$$

### 3.7. LES MÉTHODES DU MODÈLE D'ADAPTATION DANS LES ESSAIS CLINIQUES

---

Cela donne le rapport de taille d'échantillon :

$$R_n = \frac{n_{esp}}{n} = \frac{\varepsilon^2}{ab}$$

• Si  $\varepsilon = 0.25$ ,  $\alpha = 0.025$ ,  $\beta = 0.8$ ,  $n = 502$ ,  $a = 0.1$ ,  $b = 0.4$  (on note que  $\frac{a+b}{2} = \varepsilon$ ), on a donc

$$R_n = \frac{0.25^2}{0.1 \times 0.4} = 1.56$$

#### La puissance Bayésienne

Pour tester l'hypothèse nulle que  $H_0 : \theta \leq 0$  contre l'hypothèse alternative que,  $H_\alpha : \theta > 0$ , on définit la signification bayésienne :  $P_\beta = P^\pi(\theta < 0/\text{données}) < \alpha_\beta$ .

On note que l'interprétation bayésienne peut être facile à trouver, et elle est basée sur la distribution a posteriori, pour le cas où les données et l'a priori suivent une distribution normale. La distribution a posteriori est donnée par :

$$\pi(\theta/x) = N\left(\frac{n_0\mu + nx}{n_0 + n}, \frac{\sigma^2}{n_0 + n}\right)$$

$$\begin{aligned} P^\pi(\theta < 0/\text{données}) &= P^\pi\left(\frac{\theta - \frac{n_0\mu + nx}{n_0 + n}}{\frac{\sigma}{\sqrt{n_0 + n}}} < \frac{-\frac{n_0\mu + nx}{n_0 + n}}{\frac{\sigma}{\sqrt{n_0 + n}}}/x\right) \\ &= \Phi\left(-\left(\frac{n_0\mu + nx}{n_0 + n}\right) \times \frac{\sqrt{n_0 + n}}{\sigma}\right) \\ &= \Phi\left(-\frac{n_0\mu + nx}{\sigma\sqrt{n_0 + n}}\right) \end{aligned}$$

$Z_{1-\alpha}$  est le quantile de  $N(0, 1)$ .

$H_0$  est acceptée lorsque :

$$-\frac{n_0\mu + nx}{\sigma\sqrt{n_0 + n}} < Z_{1-\alpha} \iff -(n_0\mu + nx) < \sigma Z_{1-\alpha}\sqrt{n_0 + n}$$

Donc la région d'acceptation est :

$$x > \frac{-\sqrt{n_0 + n}\sigma Z_{1-\alpha} - n_0\mu}{n}$$

Et la puissance est donnée par :

$$P \left( \frac{X - \mu}{\sigma \sqrt{\frac{n_0 + n}{n_0 n}}} > \frac{\frac{-\sqrt{n_0 + n} \sigma Z_{1-\alpha} - n_0 \mu}{n} - \mu}{\sigma \sqrt{\frac{n_0 + n}{n_0 n}}} \right) = \Phi \left( \frac{\mu \sqrt{n_0 + n} \sqrt{n_0}}{\sigma \sqrt{n}} - \sqrt{\frac{n_0}{n}} Z_{1-\alpha} \right)$$

L'exemple suivant consiste à déterminer le nombre de sujets nécessaires à partir d'une puissance bayésienne donnée.

**exemple 3.7.3** *pour une bonne illustration, on considère une étude de **phase II**, hypotension comparant un traitement d'essai avec un agent de contrôle actif. Supposant que le point final primaire est la réduction de la tension artérielle systolique (SBP) et que l'effet du traitement est estimé à une distribution normale, i.e  $\theta \sim N\left(\mu, \frac{2\sigma^2}{n_0}\right)$ . Le design vise à construire la puissance Bayésienne à  $(1 - \beta_\beta)$  dans le niveau de signification Bayésienne de  $\alpha_\beta = 0.2$ , pour la taille de l'échantillon. La différence de moyenne d'échantillon est  $\hat{\theta} \sim N\left(\theta, \frac{2\sigma^2}{n}\right)$ , où  $n$  est la taille de l'échantillon pour chaque groupe pour une grande taille d'échantillon. On peut supposer que  $\sigma$  est connu, dans ce cas la taille de l'échantillon  $n$  est la solution de l'équation suivante :*

$$\Phi \left( \frac{\mu \sqrt{n_0 + n} \sqrt{n_0}}{\sigma \sqrt{2n}} - \sqrt{\frac{n_0}{n}} Z_{1-\alpha_\beta} \right) = 1 - \beta_\beta$$

Ceci conduit à :

$$\frac{\mu \sqrt{n_0 + n} \sqrt{n_0}}{\sigma \sqrt{2n}} - \sqrt{\frac{n_0}{n}} Z_{1-\alpha_\beta} = Z_{1-\beta_\beta} \quad (3.2)$$

La dernière équation de (3,2) peut être écrite comme suit :

$$An + B\sqrt{n} + C = 0 \quad (3.3)$$

où

$$\begin{cases} A = Z_{1-\beta_\beta}^2 - \mu^2 n_0 \\ B = 2Z_{1-\beta_\beta} Z_{1-\alpha_\beta} \sqrt{2n_0} \\ C = 2Z_{1-\alpha_\beta}^2 n_0 - \mu^2 n_0^2 \end{cases}$$

La solution de l'équation (3,3) est :

$$n = \left( \frac{-B + \sqrt{B^2 - 4AC}}{2A} \right)^2$$

## 3.8 Design à plusieurs étapes pour un seul bras de l'essai

D'après (Shein-Chung Chow (2006)), il est indésirable d'arrêter une étude tôt où le traitement d'essai n'est pas efficace à la considération éthique, et désirable de déterminer l'étude prématurément lorsque le traitement de test n'est pas efficace à cause des considérations éthiques. Par conséquent, le design à plusieurs étapes pour un seul bras de l'essai est employé pour déterminer si le test traité est promis comme un test final. On distingue deux méthodes :

### 3.8.1 L'approche classique pour les deux étapes de conception

Le concept optimal de l'approche classique permet d'arrêter lorsque les longues conséquences du premier événement ont échoué. On considère les hypothèses suivantes :

$$H_0 : p \leq p_0 \quad \text{contre} \quad H_\alpha : p \geq p_1$$

où  $p_0$  : est la proportion de la réponse indésirable.

$p_1$  : est la proportion de la réponse désirable.

et  $p_1 > p_0$ .

Si la réponse proportionnelle pour le traitement de test est de niveau indésirable, on le rejette comme un traitement non efficace avec une grande probabilité, et si la réponse proportionnelle est de niveau désirable on doit l'accepter.

Soient  $n_1$  et  $n_2$  deux nombres de sujets respectivement dans la première et la deuxième étape ;  $n_1$  patients sont traités dans la première étape, s'il y a au moins que  $r_1$  réponses, donc on arrête l'essai. Sinon il y a addition de  $n_2$  patients recrutés et testés dans la deuxième étape.

• **La décision observée** : si le traitement de test est basé sur la réponse proportionnelle de  $n = n_1 + n_2$  sujets, on rejette  $H_0$  (ou  $H_\alpha$ ) de sens que la dernière étude du test traitée doit être exécutée.

On a proposé de choisir un design à deux étapes optimales qui fait une taille d'échantillon espérée minimale sous l'hypothèse nulle.

$n_{\text{exp}}$  : la taille d'échantillon espérée.

$P_{et}$  : la probabilité d'échantillon espérée et la probabilité de déterminer après la première étape.

$$n_{\text{exp}} = n_1 + (1 - P_{et}) n_2$$

A la fin de la première étape, on doit déterminer et rejeter l'hypothèse nulle si  $r_1$  représente les petites réponses observées.  $p_{et}$  est donné par :  $p_{et} \sim B_c(r_1; n_1, p)$  avec  $B_c(r_1; n_1, p)$  est la distribution binomiale cumulative que  $X \leq r_1$ .

On rejette le traitement de test à la fin de la deuxième étape si  $r$  est la réponse la moins observée. La probabilité de rejeter le traitement de test avec la probabilité de succès  $p$  est donnée par :

$$B_c(r_1; n_1, p) + \sum_{x=r_1+1}^{\min(n_1, r)} B(x; n_1, p) B_c(r-x; n_2, p)$$

Où  $B(x; n_1, p)$  est la probabilité de fonction de densité binomiale pour les valeurs spéciales  $p_0, p_1, \alpha, \beta$ . Le design de deux étapes optimale peut être obtenu comme le design de deux étapes qui satisfait l'erreur construite et minimiser la taille d'échantillon espérés où la réponse proportionnelle est  $p_0$ .

### 3.8.2 L'approche bayésienne

On suppose que  $X \sim B(n, p)$ ;  $p \sim \text{Beta}(a, b)$ . Dans la première étape avec  $n_1$  sujets, la distribution a posteriori de  $p$  sachant  $X$  est donnée par :

$$\pi_1(p/x) = \text{Beta}(a+x, b+n_1-x) \quad (3.4)$$

La même chose à la fin de la deuxième étape avec  $n$  sujets, la distribution a posteriori de  $p$  sachant  $X$  est :

$$\pi(p/x) = \text{Beta}(a+x, b+n-x) \quad (3.5)$$

Par conséquent, le point d'arrêt du l'essai est choisi dans le cas où la puissance Bayésienne dans la première étape est  $(1 - \beta_1)$ . Le niveau de signification Bayésienne est  $\alpha_{\beta_1}$ ,  $n$  est choisi dans le cas où la puissance Bayésienne dans la deuxième étape est  $(1 - \beta)$  avec le niveau de significatif  $\alpha_{\beta}$  basé sur (3,4) et (3,5), la puissance conditionnelle et la puissance prédictive peuvent être dérivées.



### 3.8. DESIGN À PLUSIEURS ÉTAPES POUR UN SEUL BRAS DE L'ESSAI

---

Sachant  $X$  de  $n_1$  patients où la réponse dans la première étape, la probabilité d'obtenir au moins  $y$  réponses hors  $n_2$  patients dans la deuxième étape donnée par :

$$P(y/x, n_1, n_2) = \sum_{i=y}^{n_2} \binom{n_2}{i} \left(\frac{x}{n_1}\right)^i \left(1 - \frac{x}{n_1}\right)^{n_2-i}$$

Donc la distribution d'échantillonnage de  $X$  sachant  $p \in [0; 1]$  est donnée par :

$$P(X = x/p) = \binom{n_1}{x} p^x (1-p)^{n_1-x}$$

Tandi que :

$$P(a < p < b \cap X = x) = \int_a^b \binom{n_1}{x} p^x (1-p)^{n_1-x} dp$$

et

$$P(X = x) = \int_0^1 \binom{n_1}{x} p^x (1-p)^{n_1-x} dp$$

Si  $\mathbf{a}=\mathbf{b}=\mathbf{1}$ , la distribution a posteriori de  $p$  conditionnelle les réponses  $X = x$  de  $n_1$  est une distribution de Beta, i.e :

$$\pi(p/x) = \frac{p^x (1-p)^{n_1-x}}{B(x+1, n_1-x+1)}$$

Et la probabilité prédictive d'obtenir au moins  $y$  réponses d'addition  $n_2$  malades sachant que la réponse observée  $x/n_1$  dans la première étape est

donnée par :

$$\begin{aligned}
 P(y/x, n_1, n_2) &= \int_0^1 P(Y > y/p, n_2) \pi(p/x) dp \\
 &= \int_0^1 \sum_{i=y}^{n_2} \binom{n_2}{i} p^i (1-p)^{n_2-i} \frac{p^x (1-p)^{n_1-x}}{B(x+1, n_1-x+1)} dp \\
 &= \sum_{i=y}^{n_2} \binom{n_2}{i} \frac{\int_0^1 p^{i+x} (1-p)^{n_2+n_1-i-x} dp}{B(x+1, n_1-x+1)} \\
 &= \sum_{i=y}^{n_2} \binom{n_2}{i} \frac{B(x+i+1, n_2+n_1-x-i+1)}{B(x+1, n_1-x+1)}
 \end{aligned}$$

Avec

$$\int_0^1 p^a (1-p)^b dp = B(a+1, b+1)$$

Par conséquent, les conditions et la puissance prédictive peuvent être utilisées pour concevoir le design d'adaptation séquentielle ou d'autres.

### 3.9 Plans adaptatifs bayésiens optimaux

D'après (Shein-Chung Chow 2006), en pratique, la différence d'adaptativité et le choix de l'a priori avec des résultats probables peuvent être conduits à des différents types de modèle d'adaptation. Comment va-t-on choisir une adaptation efficace ?

Parmi ces designs de conceptions cette question devient très intéressante. Dans cette partie, on propose une évaluation pour choisir le Bayésien optimal des plans adaptatifs, qui est un indicateur des résultats de la santé des malades.

La conception bayésienne optimale est le design qui a une grande utilité expérimentale.

Pour l'illustration, on applique l'approche du choix d'un design (conception) bayésien optimal parmi les trois design à deux bras communément considérés dans la recherche pharmaceutique et le développement.

### 3.9. PLANS ADAPTATIFS BAYÉSIENS OPTIMAUX

---

Les trois conceptions designs de l'essai à deux bras communément considérés incluent l'approche classique de deux conceptions distinctes de deux bras (i.e deux bras de la **phase II** de conception suivie d'une phase à deux bras), et deux différentes phases de continuité 2/3 design. On calcule l'évaluation du poids par leurs probabilités a priori pour obtenir une utilité espérée pour le design. On considère 3 scénarios de connaissance préalable dans le tableau suivant :

<i>Scénario</i>	<i>Taille d'effet</i>	<i>Pr ob a priori</i>
1	0	0.2
2	0.1	0.2
3	0.2	0.6

**Table(3, 1)** : connaissance préalable

Pour l'approche classique de deux conceptions distinctes à deux bras, on considère la **phase II** et la **phase III**. Pour étudier la **phase II**, on suppose  $\delta = 0.2$ ; unilatérale  $\alpha = 0.1$  et la **puissance =0.8**, donc la taille de l'échantillon est  $n_1 = 45$ .

Pour la **phase II**, on suppose  $\delta = (0.2)(0) + (0.2)(0.1) + (0.6)(0.2) = 0.14$  (à partir du tableau (2, 1)).

D'autre part, si on considère  $\alpha = 0.025$  (*unilatérale*) et la **puissance = 0.9**, la taille d'échantillon totale recruté est **n = 2144**. Si l'étude de la **phase II** n'a pas de signification statistiques, on ne passe pas directement au design de la **phase III**.

<i>Scénario</i>	<i>Taille d'effet</i>	<i>Pr obabilité a priori</i>	<i>Pr obabilité de continuité de phase III, <math>P_c</math></i>	<i>Phase puissance <math>p_3</math></i>
1	0	0.2	0.1	0.025
2	0.1	0.2	0.4	0.639
3	0.2	0.6	0.9	0.996

**Table(3, 2)** : caractéristiques classiques des essais de **phase II** et **phase III**

La probabilité de continuité à la **phase III** est donnée (a partir de la

CHAPITRE 3. MÉTHODES ADAPTATIVES BAYÉSIENNES DANS  
LES ESSAIS CLINIQUES

---

table (3, 2)) par :

$$\begin{aligned}
 p_c &= \sum_{i=1}^3 p_c(i) \pi(i) \\
 &= 0.2(0.1) + 0.2(0.4) + 0.6(0.9) \\
 &= 0.64
 \end{aligned}$$

Donc la taille de l'échantillon espérée de la **phase III** des essais est :

$$\begin{aligned}
 \tilde{N} &= (1 - p_c) n_1 + p_c n \\
 &= (1 - 0.64)(450) + (0.64)(2144) \\
 &= 1500
 \end{aligned}$$

La puissance espérée totale est donnée par :

$$\begin{aligned}
 \tilde{P} &= \sum_{i=1}^3 p_c(i) \pi(i) p_3(i) \\
 &= (0.2)(0.1)(0.025) + (0.2)(0.4)(0.639) + (0.6)(0.9)(0.996) \\
 &= 0.59
 \end{aligned}$$

Les conceptions distinctes de phase **II** et phase **III** ont une puissance totale de 59% avec une taille d'échantillon expérimentale totale de 1500.

<i>Scénario</i>	<i>Taille d'effet</i>	<i>Pr ob a priori</i> $\pi$	$N_{\text{exp}}$	<i>Puissance</i>
1	0	0.2	1600	0.025
2	0.1	0.2	1712	0.46
3	0.2	0.6	1186	0.98

**Table( 3, 3)** : *caractéristiques de tubes sans soudure phase II/III de conception avec les limites OB*

D'autre part, pour la continuité des conceptions de phase **II/III** on suppose :

- (i)  $\delta = 0.14$ .
- (ii) Unilatérale  $\alpha = 0.025$ .
- (iii) Puissance 0.90.

### 3.9. PLANS ADAPTATIFS BAYÉSIENS OPTIMAUX

---

(iv) Les bornes d'arrêt d'efficacité (de O'Brien-Fleming).

(v) Les bornes d'arrêt symétriques.

On suppose qu'il existe un plan d'analyse lorsque 50% des patients sont enregistrés donc le design a une analyse final et on a :

-La taille d'échantillon de deux analyses est de 1,085 et 2,171 respectivement.

-La taille d'échantillon proportionnelle entre les deux groupes est de 1.

-La taille d'échantillon maximum pour le design est de 2,171.

-Sous l'hypothèse nulle, la taille d'échantillon espérée est de 1,625.

• **La règle de décision est :**

Dans la première étape : on accepte l'hypothèse nulle si  $Z_1 < 0$  et on rejette l'hypothèse nulle si  $Z_1 \geq 2.79$ .

Dans la deuxième étape : on accepte l'hypothèse nulle si  $Z < 1.974$  et on rejette l'hypothèse nulle si  $Z \geq 1.97$

-La probabilité d'arrêter dans la première est étape est donnée par :

\* L'arrêt pour  $H_0$  lorsque  $H_0$  vraie est 0.500.

\* L'arrêt pour  $H_\alpha$  lorsque  $H_0$  vraie est 0.0026 .

\* L'arrêt pour  $H_0$  lorsque  $H_\alpha$  vraie est 0.0105 .

\* L'arrêt pour  $H_\alpha$  lorsque  $H_\alpha$  vraie est 0.3142.

-La moyenne totale de la taille d'échantillon espérée peut être obtenue comme :

$$\begin{aligned} \sum \pi(i) N_{\text{exp}}(i) &= (0.2)(0.025) + (0.2)(0.46) + (0.6)(0.98) \\ &= 1374 \text{ (d'après la caractéristique de la table (2,3))} \end{aligned}$$

-Et la puissance totale est donnée par :

$$\sum \pi(i) N_{\text{exp}}(i) \text{power}(i) = (0.2)(0.025) + (0.2)(0.46) + (0.6)(0.98) = 0.69$$

<i>Scénario<sub>i</sub></i>	<i>Taille d'effet</i>	<i>Pr ob a priori</i> $\pi$	$N_{\text{exp}}$	<i>Puissance</i>
1	0	0.2	1492	0.025
2	0.1	0.2	1856	0.64
3	0.2	0.6	1368	0.996

**Table (3,4) :** Caractéristiques de tubes sans soudure phase **II/III** Pocock avec Boudry

Aujourd'hui, on considère un autre type de la continuité adaptative de phase **II/III** et on essaye d'utiliser l'efficacité des bornes d'arrêt de Pocock,

et les bornes d'arrêt symétriques basées sur les mêmes paramètres de spécification, i.e :  $\alpha = 0.025$  ; **puissance = 0.9** ; **la différence moyenne = 0.14** ; **la dérivation standard = 1** et on a :

-La taille d'échantillon de deux analyses est de 1,274 et 2,549 respectivement.

-La taille d'échantillon proportionnelle entre les deux groupes est de 1.

-La taille d'échantillon maximum pour le design est de 2,549.

-Sous l'hypothèse nulle, la taille d'échantillon espérée est de 1,492.

-Sous l'hypothèse alternative, la taille d'échantillon espérée est de 1,669.

• **La règle de décision est :**

Dans la première étape : on accepte l'hypothèse nulle si p-valeur  $> 0.1867$  (i.e  $Z_1 < 0.89$ ). On rejette l'hypothèse nulle si p-valeur  $\leq 0.0158$  (i.e  $Z_1 > 2.149$ ).

Dans la deuxième étape : on accepte l'hypothèse nulle si p-valeur  $> 0.0158$  (i.e  $Z = 2.149$ ). On rejette l'hypothèse nulle si p-valeur  $< 0.0158$ .

-La probabilité d'arrêter dans la première étape est donnée par :

\*L'arrêt pour  $H_0$  lorsque  $H_0$  vraie est 0.8133.

\*L'arrêt pour  $H_\alpha$  lorsque  $H_0$  vraie est 0.0538.

\*L'arrêt pour  $H_0$  lorsque  $H_\alpha$  vraie est 0.0538.

\*L'arrêt pour  $H_\alpha$  lorsque  $H_\alpha$  vraie est 0.6370.

-La moyenne totale de la taille d'échantillon espérée peut être obtenue comme :

$$\begin{aligned} \sum \pi(i) N_{\text{exp}}(i) &= (0.2)(1492) + (0.2)(1856) + (0.6)(1368) \\ &= 1490 \text{ (d'après les caractéristiques du tableau (3,4))} \end{aligned}$$

et la puissance totale est donnée par :

$$\begin{aligned} \sum \pi(i) N_{\text{exp}}(i) \text{power}(i) &= (0.2)(0.025) + (0.2)(1856) + (0.6)(1368) \\ &= 1490 \text{ (d'après les caractéristiques du tableau (3,4))} \end{aligned}$$

et la puissance totale est donnée par :

$$\begin{aligned} \sum \pi(i) N_{\text{exp}}(i) \text{power}(i) &= (0.2)(0.025) + (0.2)(0.64) + (0.6)(0.996) \\ &= 0.73 \end{aligned}$$

## 3.10 La conformité d'adaptation bayésienne dans les essais cliniques

D'après (Shein-Chung Chow, 2010), on considère les problèmes communs qui se posent dans la conformité Bayésienne de design inflammatoire. Ces problèmes impliquent généralement les caractéristiques d'adaptation de design et leurs effets sur l'erreur de première espèce. Le choix de la distribution a priori pour l'essai de conformité est rarement un problème, comme l'a priori non informative est généralement choisie. Pour démontrer l'idée d'un essai de conformité, on développe un très simple essai, et ensuite on construit sur cette base une conformité pour ajouter des aspects d'adaptation et l'effet de ses fonctions adaptatives. Pour être clair voici l'exemple suivant :

### 3.10.1 La base de l'essai de conformité

Supposons que nous ayons un essai manchot au cours duquel chaque sujet est considéré comme un succès ou un échec. Les informations qui nous renseignent sur un succès ou un échec du sujet sont observées immédiatement, ce peut être le cas dans une étude de la douleur avec la probabilité qu'un objet est de  $p$  succès. Chaque observation est supposée indépendante, par conséquent pour  $n$  observations, le nombre de succès est :

$$X/p \sim \text{Binomial}(n, p)$$

On suppose la régularité qu'a démontré le succès statistique du traitement médical de l'essai et doit montrer que  $p > 0.5$ . Dans un cadre de test d'hypothèse :

$$H_0 : p \leq 0.5 \quad \text{contre} \quad H_1 : p > 0.5$$

La taille de l'échantillon de l'essai dans le cas unilatéral au niveau 0.05 est  $n = 100$ , et ne résulte dans l'essai successif que si  $X \geq 59$ . La probabilité de  $X \geq 59$ , supposant  $p = 0.5$  est 0.0443. Le succès de  $X \geq 58$ , se traduirait par une erreur de première espèce de 0.066. Pour cet exemple nous supposons que la restriction réglementaire est à sens unique d'erreur de première espèce de 0.05. Ainsi l'essai serait considéré comme réussi que si  $X \geq 59$ . 58 succès d'observation seraient considérés comme une étude d'échec. En réalité, ce n'est pas toujours le cas comme critère d'évaluations secondaires et des informations supplémentaires peuvent signifier que les régulateurs permettent le traitement à commercialiser. Pour nos fins, nous nous concentrons sur la

réalisation de signification statistique avec le critère d'évaluation primaire. Supposons plutôt que l'approche bayésienne est utilisée ici pour les analyses prématurées, l'analyse prématurée est à conclure des succès statistiques si la probabilité a posteriori de  $H_1$  est supérieure à 95%. Pour l'analyse bayésienne, nous supposons que la distribution a priori de  $p \sim Beta(1, 1) \doteq Unif(0, 1)$  résulte la distribution a posteriori  $p/X \sim Beta(1 + X, 1 + n - X)$  la probabilité a posteriori pour supériorité.  $\Pr(p > 0.5/X)$  pour  $X = 57, 58, 59$  et  $60$  sont  $0.918, 0.945, 0.964$  et  $0.977$  respectivement ; De plus la règle que le succès statistique est la probabilité a posteriori de supériorité dans  $0.95$  correspond à la règle que  $59$  ou plus de succès implique le succès statistique.

### 3.11 Le principe des tests diagnostiques

Soit un test diagnostique  $T$  (un symptôme, une mesure échographique, ou un dosage sanguin) visant à établir la présence ou l'absence d'une maladie  $M$ . Le test fournit un résultat qualitatif, positif ou négatif. Lors de la mise au point d'un test diagnostique, on classe par ailleurs les sujets en malades et sains en utilisant une méthode de référence pour déterminer ce statut. En croisant les résultats du test avec le statut de référence *malade/sain* des sujets, On établit un tableau permettant d'évaluer les qualités diagnostiques de ce test (table(3,5)). Ce tableau permet de déterminer les nombres de sujets sains ou malades qui, d'après le test, seront correctement ou incorrectement classés en malade ou sain par un test positif ou négatif. En prenant les notations proposées dans la table (3,5), on définit les vrais positifs ( $VP$ ) qui sont les sujets malades ayant un résultat de test positif, les faux positifs ( $FP$ ) qui sont les sujets sains ayant un résultat de test positif, les vrais négatifs ( $VN$ ) qui ont les sujets sains qui sont un résultat de test négatif et les faux négatifs ( $FN$ ) qui sont les sujets malades ayant un résultat de test négatif. A ce stade, c'est l'investigateur qui fixe les nombres  $M$  de sujets malades et  $S$  de sujets sains.

	<i>Malades</i>	<i>Sains</i>	<i>Total</i>
<i>Test positif</i>	<b>VP</b>	<b>FP</b>	<b>P</b>
<i>Test négatif</i>	<b>FN</b>	<b>VN</b>	<b>N</b>
<i>Total</i>	<b>M</b>	<b>S</b>	

**Table(3, 5)** : éléments de test diagnostique : Lecture verticale du test négatif



### 3.11. LE PRINCIPE DES TESTS DIAGNOSTIQUES

---

**VP** = vrais positifs, **FP**=faux positifs, **FN**=faux négatifs, **VN**=vrais négatifs, **P** = nombre de test positifs, **N**= nombre de tests négatifs, **M**=nombre de sujets malades, **S** =nombre de sujets sains.

A partir de ces éléments, on calcule :

$$\begin{aligned} \text{La sensibilité (Se)} &= \frac{VP}{VP + FN} \\ \text{La spécificité (Sp)} &= \frac{VN}{VN + FP} \end{aligned}$$

La *Se* est aussi la probabilité d'avoir un test positif lorsque l'on est malade, que l'on note  $Pr(P/M)$ . De même la *Sp* est la probabilité d'avoir un test négatif lorsque l'on est sain, soit  $Pr(N/S)$ . La *Se* et la *Sp* sont des caractéristiques indépendantes de la prévalence de la maladie dans la population étudiée.

Après cette phase d'élaboration du test, son utilisation en pratique quotidienne vise à déterminer la présence ou l'absence de la maladie chez un sujet donné. Le tableau croisé précédemment décrit doit alors être utilisé dans l'autre sens, c-à-d en lui appliquant une lecture horizontale, ce qui nécessite de modifier un peu la table (3,5) pour obtenir la table (3,6). A partir du résultat du test diagnostique, on cherche donc à déterminer le statut *malade/sain* du sujet.

Avant d'avoir appliqué le test au patient, le médecin ne sait pas s'il présente ou non la maladie. Il sait seulement que la probabilité qu'il présente la maladie n'est pas nulle même si souvent elle est très faible. On peut estimer la probabilité que le patient ait la maladie avant le test par la prévalence de la maladie dans la population générale, notée ici  $Prév(M)$  (voir la table(3,6)). Cette prévalence constitue une connaissance a priori, c-à-d existant avant la réalisation du test diagnostique et en l'absence d'autres informations cliniques sur le patient, c'est la meilleure estimation par rapport que l'on puisse avoir du risque pour le sujet d'être malade. Dans la *table(3,6)* par rapport à la *table(3,5)*, la dernière ligne doit être modifiée et les totaux en colonne des sujets malades et sains ayant servi à concevoir le test doivent être remplacés par les estimations de leurs prévalences respectives,  $Pr(M)$  et  $Pr(S)$ , dans la population dont est issu le patient considéré (population générale,

population hospitalière, consultants en ville, etc.)

	Malades	Sains	Total
<i>Test positif</i>	$Pr(p/M) \times Prév(M)$	$aPr(p/S) \times Prév(S)$	$bp$
<i>Test négatif</i>	$Pr(N/M) \times Prév(M)$	$Pr(N/S) \times Prév(S)$	$N$
<b>Total</b>	$Prév(M)$	$Prév(S)$	

**Table(3, 6)** : éléments du test diagnostique : lecture horizontale

**Prév(M)** =prévalence de sujets malades dans la population.

**Prév(S)** =prévalence de sujets sains dans la population.

**Pr(p/M)** =probabilité d'avoir un test positif lorsque l'on est malade.

**Pr(p/S)** =probabilité d'avoir un test positif lorsque l'on est sain.

**Pr(N/M)** =probabilité d'avoir un test négatif lorsque l'on est malade.

**Pr(N/S)** = probabilité d'avoir un test négatif lorsque l'on est sain.

Après réalisation du test, en supposant que celui-ci soit positif, le médecin ne sait toujours pas si le sujet est porteur de la maladie mais il sait néanmoins que le sujet se trouve dans la première ligne du tableau, celle des résultats positifs. Il reste donc pour le médecin à déterminer si le sujet est dans la première ou la seconde case de cette ligne. La proportion de sujets malades qui ont un test positif est le produit de la prévalence de la maladie  $Prév(M)$  par la sensibilité du test ( $Se = Pr(p/M)$ ). De même, la proportion de sujets sains ayant un test positif est égale au produit de la prévalence des sujets indemnes de la maladie  $Prév(S)$  par la probabilité d'avoir un test positif lorsque l'on n'est pas malade, c-à-d  $Pr(p/S)$ , qui vaut  $1 - Sp$ . Pour connaître la probabilité d'être malade lorsque l'on a un test positif ( $Pr(M/p)$ ), on rapporte la proportion de sujets malades ayant un test positif au nombre total de sujets ayant un résultat positif. Ce rapport correspond à la valeur prédictive positive ( $VPP$ ) et il s'obtient via le théorème de Bayes de la manière suivante :

$$Pr(M/p) = \frac{Pr(p/M) \cdot Prév(M)}{Pr(p/M) \cdot Prév(M) + (1 - Pr(N/S)) \cdot Prév(S)}$$

Dans cette équation, on trouve les éléments suivants :

- $Pr(M/p)$  est la probabilité d'avoir la maladie lorsque le test est positif. C'est aussi la  $VPP$  ;

- $Pr(p/M)$  est la probabilité d'avoir un résultat positif lorsque l'on est malade. C'est la sensibilité du test ;

### 3.11. LE PRINCIPE DES TESTS DIAGNOSTIQUES

---

- $Pr(N/S)$  est la probabilité d'avoir un test négatif lorsque l'on n'est pas malade. C'est la spécificité du test ;
- $Prév(M)$  est la probabilité a priori d'avoir la maladie : c'est la prévalence de la maladie ;
- $Prév(S)$  est la probabilité a priori de ne pas être malade c'est le "complément à 1" de la prévalence.

En reprenant les notations du tableau (3, 6), la **VVP** vaut donc  $\frac{a}{a+b}$ .

**Résumé :** après application du test diagnostique, le médecin attribue au sujet une nouvelle probabilité d'être atteint de la maladie  $M$ ,  $Pr(M/p)$  qui se lit "probabilité d'avoir la maladie sachant que le test est positif". Cette **VVP** est la probabilité a posteriori de la maladie, c-à-d la probabilité que le sujet soit malade, connaissant le résultat du test diagnostique et donc après incorporation de l'information obtenue (le test est positif). La **VVP** combine donc l'information a priori (la prévalence de la maladie) et le résultat du test (le test est positif).

L'obtention du résultat du test permet donc d'établir la probabilité d'une hypothèse (être malade) à partir de l'estimation initiale qu'est la prévalence. Plus précisément, le test transforme une probabilité a priori en une probabilité a posteriori. La prévalence est une connaissance a priori et la **VVP** est une connaissance a posteriori. La prévalence a été "corrigée" en une **VVP** qui mesure la probabilité que le sujet soit porteur de la maladie connaissant le résultat positif du test. On a en fait réalisé une mise à jour de l'information sur la situation du patient grâce au théorème de Bayes. Deux exemples sont donnés dans les tables (3, 7) et (3, 8), concernant le diagnostic d'une sérologie positive pour le VIH chez une femme de 75 ans sans antécédents et chez un toxicomane de 27 ans. Ces exemples illustrent le rôle parfois important que jouent les connaissances a priori (la prévalence de la maladie dans la population concernée).

	Malades	Sains	Total
<b>Test positif</b> $p$	0.97	0.02	$p$
<b>Test négatif</b> $N$	0.03	0.98	$N$
<b>Total</b>	1/500000	994999/500000	

**Table(3, 7) :** illustration chez une femme âgée, prévalence= $\frac{1}{500000}$

Chez cette patiente, la **VVP** vaut :

$$\mathbf{VPP} = \frac{0.97 \cdot 1/500\,000}{0.97 \cdot 1/500\,000 + (1 - 0.98) \cdot 499\,999/500\,000} = 0.00097$$

	Malades	Sains	Total
<b>Test positif</b> $p$	0.97	0.02	$p$
<b>Test négatif</b> $N$	0.03	0.98	$N$
<b>Total</b>	1/10	9/10	

**Table(3, 8)** : illustration chez un toxicomane, prévalence=1/10

Chez ce patient, la **VPP** vaut :

$$\mathbf{VPP} = \frac{0.97 \cdot 0.1}{0.97 \cdot 0.1 + (1 - 0.98) \cdot 0.9} = 0.982$$

Le plus souvent, un test diagnostique compare deux entités seulement : présence ou absence de la maladie. Mais il est tout à fait possible d’imaginer des situations diagnostiques où l’on est amené à comparer trois possibilités (pas de tumeur, tumeur bénigne, tumeur maligne). Le principe du théorème de Bayes peut parfaitement s’appliquer et on peut en principe comparer autant d’hypothèses diagnostiques que l’on veut. On pourrait très bien imaginer l’usage du raisonnement bayésien devant un patient fébrile l’utilisation du raisonnement bayésien pour déterminer laquelle parmi toutes les pathologies provoquant une hyperthermie est la plus probable. Le tableau contiendrait alors deux lignes et quelques dizaines de colonnes. Chaque diagnostique serait assorti d’une valeur prédictive positive donnant la probabilité a posteriori de chacun de ces diagnostiques.

### 3.12 La randomisation

D’après (**L.M.Friedman et al. 2010**), un essai clinique randomisé est fréquemment utilisé dans la recherche de médicaments. Cependant, il est une opposition considérable à l’utilisation de cette conception. Une préoccupation majeure est l’éthique problème de permettre à un évènement aléatoire pour déterminer le traitement d’un patient. On a fait valoir qu’un essai comparatif qui montre de grandes différences entre deux traitements est un mauvais essai car la moitié des patients ont reçu un traitement inférieur. D’autre part, dans un essai prospectif l’assignation aléatoire des traitements évite de nombreux biais potentiels. Plus préoccupant est l’essai au cours duquel un nouveau traitement est comparé à un ancien traitement quand il y a des informations sur l’efficacité de l’ancien traitement par des données historiques.

### 3.12.1 Principe de la randomisation adaptative

L'attribution adaptative de dose dont nous en discutons ici inclut toujours la randomisation. elle est distincte de la tâche adaptative déterministe de dose, comme le jeu de la règle de gagnant. Berry et Eick (1995) donnent une discussion comparative de play-de winner contres les diverses règles de la randomisation, la randomisation adaptative est une solution de la théorie de décision. Ici nous centrons sur des conceptions des résultats d'adaptation, par opposition aux signes de covariable d'adaptative qui cherchent à équilibrer des covariates à travers des traitements.

Supposons qu'il y a deux bras,  $A_1$  et  $A_2$ , soit  $P_{1<2}$  qui désigne la probabilité a posteriori que le bras  $A_2$  domine le bras  $A_1$ .

**exemple 3.12.1** *supposons que le résultat est une réponse d'efficacité binaire, et de laisser désigner  $\theta_1$  et  $\theta_2$  la probabilité de réponse sous chaque bras de traitement. Soit  $Y$  génériquement dénoter les données actuellement disponibles. Ensuite  $P_{1<2} = P(\theta_1 < \theta_2/Y)$ . **Thall et Wathen (2007)** proposent de randomiser les traitements  $A_1$  et  $A_2$  avec probabilité proportionnelle à :*

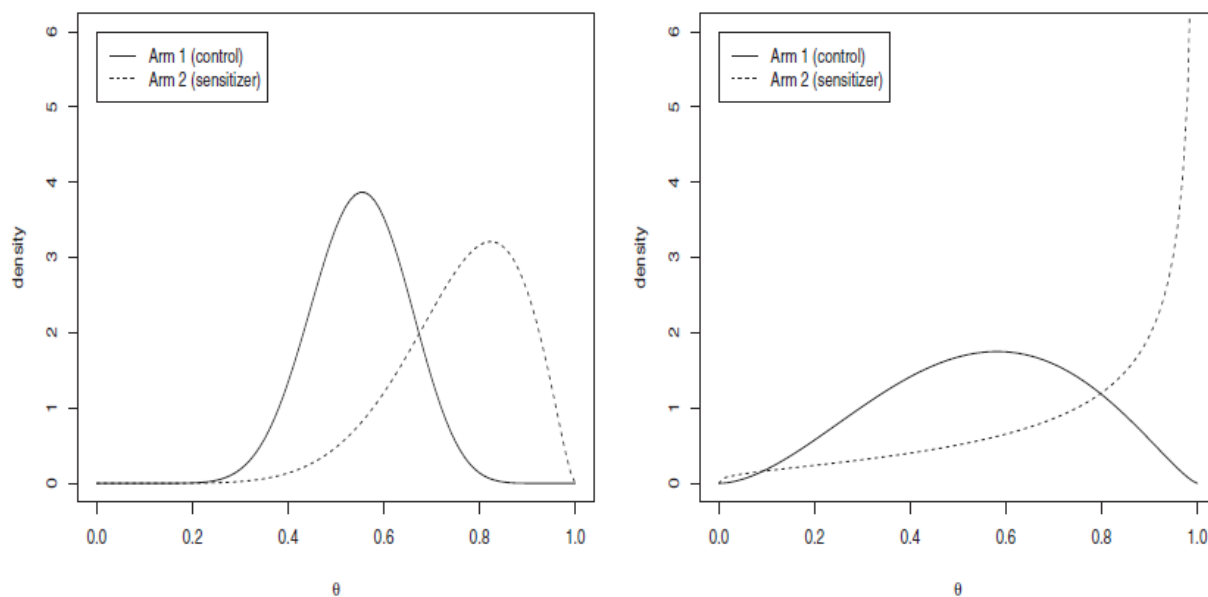
$$r_2(Y) = \{P_{1<2}(Y)\}^c \text{ et } r_1(Y) = \{1 - P_{1<2}(Y)\}^c$$

*En général, pour plus de deux bras, l'utilisation*

$$r_i(Y) \propto \left\{ P\left(\theta_j = \max_k \theta_k/Y\right) \right\}^c$$

**Thall et Wathen (2007)** proposent d'utiliser  $c = \frac{n}{2N}$  où  $N$  est le nombre maximum de patients et  $n$  est le nombre de patients actuellement inscrits. Cette recommandation est fondée sur des preuves empiriques sous les scénarios

rios typiques. (Voir fig (3,1)).



**Figure(3,1)** : la distribution a priori de  $\theta$ , **a gauche** : l'a priori standard et **a droite** : l'a priori conservative. Deux bras 1 a priori a comme moyenne 0.55 et deux bras 2 a priori ont comme moyenne 0.75, l'a priori conservative a déviation standard est égale au double de l'a priori standard.

### 3.12.2 Conclusion

L'assignation aléatoire des patients, soit le groupe de traitement ou de contrôle des essais cliniques est l'un des progrès les plus importants dans l'histoire de la recherche médicale, aucun autre ne donne un niveau comparablement élevé dans les résultats de l'essai. La randomisation garantit que le traitement est sans biais. En particulier, la randomisation permet de tenir compte des changements aux patients dans la qualité des soins et les options de compétitions du traitement au fil du temps, sans randomisation le pronostic du patient peut être corrélé avec la thérapie attribuée. Cependant, le réglage des covariables peut être utiles pour l'analyse des effets du traitement dans un essai non randomisé, l'importance de covariable ne peut pas être connue ou évaluables. Un autre avantage de la randomisation c'est qu'elle sert à équilibrer l'affectation des différents traitements y compris au sein des sous-groupes des patients.

Toute fois, si la randomisation est très importante dans l'inférence fréquentiste, il ne joue aucun rôle dans le calcul de probabilités a posteriori (à l'exception d'un bayésien peut choisir d'actualisation des données provenant d'essais non randomisés par rapport aux l'essai randomisé). Cela à conduit certains à penser que la randomisation bayésienne n'est pas disponible.

# Chapitre 4

## Applications

Dans ce chapitre on va donner quelques applications sur les tests bayésiens et le facteur de Bayes dans les essais cliniques.

### 4.1 Le calcul de la probabilité prédictive

Dans cette application, nous présentons des probabilités prédictives, et nous décrivons leur utilisation dans la conception des tests cliniques de la **phase II**. Un avantage de cette approche est qu'elle imite le processus décisionnel clinique basé sur les données a priori. La probabilité prédictive est obtenue en calculant la probabilité d'une conclusion positive (de rejet de l'hypothèse nulle). Pour prédire une autre échantillon  $y$  conditionnement  $x$ .

**Le calcul de la probabilité prédictive** : D'après (**Shein-Chung Chow (2010)**), pour un essai de **phase II** notre objectif est d'évaluer la proportion  $p$  pour un nouveau médicament en testant l'hypothèse

$$H_0 : p \leq p_0 \text{ contre } H_1 : p \geq p_1$$

Supposons que la distribution a priori de  $p$  est  $\pi(p)$ , suite à une distribution Beta( $a_0, b_0$ ), avec la quantité  $\frac{a_0}{(a_0 + b_0)}$  qui indique la moyenne a priori, depuis les quantités  $a_0$  et  $b_0$  peuvent être considérées comme le nombre de réponses a priori efficaces et non-réponses, respectivement,  $a_0 + b_0$  peut être considéré comme une mesure de précision a priori, la plus importante de cette somme.

Supposons que nous avons fixé un nombre maximum de patients à payer  $N_{\max}$ , et supposons que  $X$  est le nombre de réponses chez les patients parmi



#### 4.1. LE CALCUL DE LA PROBABILITÉ PRÉDICTIVE

---

les patients en cours  $n$ . ( $n \leq N_{\max}$ ) suite à une distribution Binomial( $n, p$ ), par la conjugaison de l'a priori beta et la vraisemblance binomial, la distribution a posteriori de  $p$  est une distribution de Beta, donc :

$$p/x \sim Beta(a_0 + x, b_0 + n - x)$$

L'approche de probabilité prédictive regarde vers l'avenir sur la base des données actuelles observées au projet. Si une conclusion positive à la fin de l'étude est vraisemblable ou non, elle prend alors une décision raisonnable à l'heure actuelle en conséquence.

Soit  $Y$  le nombre de réponses dans le potentiel de futurs patients  $m = N_{\max} - n$ .

Supposons que notre conception est de déclarer l'efficacité, si la probabilité a posteriori de  $p$  qui est supérieur à un certain niveau prédéfini  $p_0$  est supérieure à un seuil  $\theta_T$ . Marginaliser  $p$  de la probabilité binomial, il est bien connu que  $Y$  suit une distribution de beta-binomial, la distribution a posteriori de  $p/(X = x, Y = i)$  est une  $Beta(a_0 + x + i, b_0 + N_{\max} - x - i)$ . La probabilité prédictive (**pp**) du succès du procès peut alors être calculée comme suit : soit  $B_i = \Pr(p > p_0/x, Y = i)$  et  $I_i = I(B_i > \theta_T)$ , on a :

$$\begin{aligned} \mathbf{pp} &= E\{I[\Pr(p > p_0/x, Y) > \theta_T] / x\} \\ &= \int I[\Pr(p > p_0/x, Y) > \theta_T] dp(Y/x) \\ &= \sum_{i=0}^{i=m} \Pr(Y = i/x) \times I[\Pr(p > p_0/x, Y) > \theta_T] \\ &= \sum_{i=0}^{i=m} \Pr(Y = i/x) \times I(B_i > \theta_T) \\ &= \sum_{i=0}^{i=m} \Pr(Y = i/x) \times I_i \end{aligned}$$

La quantité  $B_i$  est la probabilité que la proportion  $p$  est plus grande que les réponses  $p_0$  données dans  $x$  réponses dans  $n$  patients et les données actuelles et  $i$  réponses dans  $m$  futurs patients. La comparaison de  $B_i$  à une valeur seuil  $\theta_T$  qui donne un indicateur  $I_i$  pour considérer l'efficacité du traitement à la fin de l'essai compte tenu des données actuelles et des résultats potentiels des  $Y = i$ .

L'exemple ci-dessous présente une illustration concrète du calcul de **pp** à l'aide des formules précédentes. La somme pondérée des indicateurs de rendement  $I_i$  est la probabilité prédictive de la conclusion d'un résultat positif d'ici la fin du procès sur la base des informations cumulées dans le stade actuel. La **pp** élevée signifie que le traitement est susceptible d'être efficace d'ici la fin de l'étude compte tenu des données actuelles alors qu'un faible **pp** suggère que le traitement ne peut avoir une activité suffisante. Par conséquent **pp** peut être utilisé pour déterminer si l'essai doit être arrêté prématurément en raison d'efficacité / de futilité ou poursuivi parce que les données actuelles ne sont pas encore concluantes. Nous définissons une règle en introduisant deux seuils sur  $p$ . La règle de décision peut être construite comme suit :

**Algorithme (phase II basée sur la conception de **pp**)**

**Step1** : Si **pp**  $< \theta_L$ , arrêter l'essai et rejeter l'hypothèse alternative.

**Step2** : Si **pp**  $> \theta_U$ , arrêter l'essai et rejeter l'hypothèse nulle.

**Step3** : Sinon, passez à l'étape suivante jusqu'à ce qu'il atteigne  $N_{\max}$  patients.

**exemple 4.1.1** *supposons qu'un chercheur envisage de scolariser un maximum de  $N_{\max} = 40$  patients dans l'étude de **phase II** à un moment donné,  $x = 16$  réponses sont observées dans  $n = 23$  patients ce qui est  $p$  ( $p > 60\%$ ) ? l'hypothèse d'une vague distribution est une Beta (0.6, 0.4) sur la proportion  $p$  et soit  $Y$  est le nombre de réponse dans  $m = 17$  futures observations, donc  $Y \sim \text{Beta} - \text{binomial} (17, 16.6, 7.4)$ . A chaque valeur possible de  $Y = i$ , l'a posteriori conditionnelle de  $p$  suit une distribution de beta,  $p/x$  ;  $Y = i \sim \text{Beta} (16.6 + i, 24.4 - i)$ , dans cet exemple, on prend  $\theta_T = 0.90$*

*Comme on peut le voir au tableau(4, 1) lorsque  $Y$  est inclu dans  $[0, 11]$ , le résultat  $p$  (proportion  $> 0.60$ ) varie de 0.0059 à 0.8415. Par conséquent nous concluons  $H_0$  pour  $Y \leq 11$ . D'autre part, lorsque  $Y$  est inclu dans  $[12, 17]$  le résultat  $p$  varie (proportion  $> 0.60$ ) varie de 0.9089 à 0.9990. Dans ce cas, au lieu de décider en faveur de  $H_1$  on considère que la probabilité prédictive est la moyenne pondérée de l'indicateur d'un essai qui se déroule jusqu'à la fin de l'étude que le calcul donne **pp** = 0.5656. si  $\theta_T = 0.10$ , l'essai ne serait pas arrêté à cause d'inutilité, car **pp** est supérieur à  $\theta_L$ . De la même façon si on choisi  $\theta_U = 0.95$ , l'essai doit être arrêter s'il soit efficace ou pas. Donc en se basant sur des données intermédiaires, nous devrions poursuivre l'étude parce que la preuve n'est pas encore suffisante pour tirer une conclusion définitive*

## 4.2. LE TEST DES PARAMÈTRES AVEC LE FACTEUR DE BAYES

---

dans les deux sens.

$\mathbf{y = i}$	$\Pr(y = i/x)$	$\mathbf{B}_i = \Pr(p > 0.60/x, y = i)$	$\mathbf{I}(\mathbf{B}_i > 0.90)$
<b>0</b>	0.0000	0.0059	0
<b>1</b>	0.0000	0.0138	0
<b>2</b>	0.0001	0.0296	0
<b>3</b>	0.0006	0.0581	0
<b>4</b>	0.0021	0.1049	0
<b>5</b>	0.0058	0.1743	0
<b>6</b>	0.0135	0.2679	0
<b>7</b>	0.0276	0.3822	0
<b>8</b>	0.0497	0.5085	0
<b>9</b>	0.0794	0.6349	0
<b>10</b>	0.1129	0.7489	0
<b>11</b>	0.1426	0.8415	0
<b>12</b>	0.1587	0.9089	1
<b>13</b>	0.1532	0.9528	1
<b>14</b>	0.1246	0.9781	1
<b>15</b>	0.0811	0.9910	1
<b>16</b>	0.0381	0.9968	1
<b>17</b>	0.0099	0.9990	1

**Table (4,1)** : le calcul de la probabilité prédictive bayésienne pour  
 $p_0 = 0.60, \theta_T=0.90, N_{\max}=40, x=16, n=23,$  et  $Beta(0.6,0.4)$   
est la distribution a priori de  $p$ .

## 4.2 Le test des paramètres avec le facteur de Bayes

Dans cette application, on fait une étude des tests d'hypothèses dans les essais cliniques avec le facteur de Bayes pour tester les paramètres.

On considère un essai typique où la quantité historique pour la réponse de thérapie est **0.20**, donc l'essai est stoppé si le taux dépasse la valeur historique et la quantité de réponse est définie dans l'étude suivante :

$$H_0 : \theta \leq 0.20 \quad \text{et} \quad H_1 : \theta > 0.20$$

où  $\theta$  est la probabilité de réponse de thérapie, et l'information a priori est disponible avec **50 patients**, **10** d'entre eux ont la réponse à la thérapie, donc le facteur de Bayes est donné par la forme suivante :

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

Avec  $f(x/\theta) \sim B(n, \theta)$

Et  $\theta \sim \text{Beta}(\alpha, \beta)$

Et  $\theta/x \sim \text{Beta}(\alpha + x, n + \beta - x)$

Où  $\theta_0 = 0.20$ ,  $n = 50$ ,  $x = 10$ ,  $\alpha = \beta = 1$ .

Donc on va calculer le facteur de Bayes (**BF**) par les programmes de statement qui sont donnés ici :

```
> alpha = 1; beta = 1
> probH0 = pbeta(0.2, 1, 1)
> probH1 = 1 - probH0
> prior.odds = probH0/probH1
> prior.odds
[1] 0.25
> alpha = 11; beta = 41
> post.odds = pbeta(0.2, 11, 41)
> post.odds = pbeta(0.2, 11, 41)/(1 - pbeta(0.2, 11, 41))
> post.odds
[1] 0.7998702
> BF = post.odds/prior.odds
> BF
[1]3.199481
```

Et on a :  $B_{10}^{\pi} = \frac{1}{B_{01}^{\pi}} = \frac{1}{3.199481} = 0.3125507$

Donc d'après les critiques du facteur de Bayes, la certitude de  $H_0$  est faible.

### 4.3 La comparaison de la mortalité du groupe de traitement avec le groupe de contrôle correspondant

Dans cette partie, on fait une étude de 22 essais avec la simulation de la technique de MCMC, puis avec des représentations graphiques des para-

4.3. LA COMPARAISON DE LA MORTALITÉ DU GROUPE DE  
TRAITEMENT AVEC LE GROUPE DE CONTRÔLE  
CORRESPONDENT

---

mètres.

D'après (Lyle D. Broemeling (2006)), on comprend l'exemple suivant de 22 essais de **Bétabloquant** pour prévenir la mortalité après un infarctus du myocarde, voir la *table* (4, 2).

Essai	Traité	Control
1	3/38	3/39
2	7/114	14/116
3	5/69	11/93
4	102/1533	127/1520
...	...	...
20	32/209	40/218
21	27/391	43/364
22	22/680	39/674

**Table(4, 2)** : étude de mortalité morts/total

Pour chaque essai, la différence de probabilité du succès est donnée pour le contrôle et les groupes de traitements et les probabilités sont transformées à l'échelle, et les paramètres sont données par l'a priori non informative normale. Donc  $ms$  donne des distributions normales indépendantes avec la moyenne  $0.0$  et la précision  $0.00001$ ,  $ds$  sont donnés une distribution normal ( $delta, tau$ ) et  $delta \sim N(0.0, 0.000001)$ .

Il y a 3 niveaux de paramètres impliqués dans cet exemple, la probabilité de succès, les paramètres de logis appelés  $ms$  et  $ds$ , et  $delta \sim N(0.0, 0.000001)$ . D'autre troisième niveau de paramètres est la précision  $tau$  des paramètres de  $delta$ , qui donne une distribution de gamma non informative ( $0.001, 0.001$ ), où le paramètre a une moyenne des paramètres est  $d$ , et  $tau$  est l'inverse de la variance.

L'essentielle de paramètre d'intéressante est  $delta$ , car si  $delta = 0$  il n'existe pas une différence de mortalité entre le contrôle et les groupes de traitements. Dans cet exemple, où la mortalité de traitement et le groupe de contrôle sont comparés sur la base de Meta-analyses de **22 essais cliniques**. L'analyse a posteriori sera exécutée avec l'accent mis sur la distribution a posteriori de  $delta$ , avec les mesures d'effets de le traitement de **Bétabloquants**. Si l'effet est **zéro**, les **Bétabloquants** n'ont pas d'effet sur la comparaison de mortalité du groupe de contrôle. Les analyses sont exécutées avec **50 000** valeurs générées de la distribution a posteriori de  $delta$  et  $sigma = \frac{1}{tau}$ . Le paramètre  $d$  est la moyenne des paramètres de  $delta$

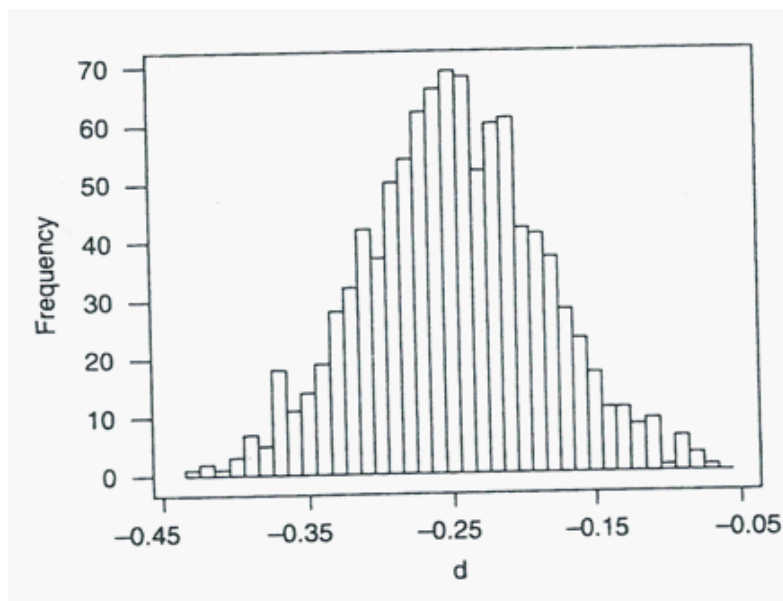
autant que *sigma* est la déviation standard, et les deux paramètres ont été donnés à partir des distributions a priori non informatives. Les caractéristiques de la distribution a posteriori de *delta* et *tau* sont données dans la table suivant.

Paramètre	moyenne	Dev.std	Mediane	Inférieur	Supérieur
<i>delta</i>	-0.2513	0.061	-0.2313	-0.3701	-0.1265
<i>sigma</i>	0.1143	0.0672	0.1020	0.0265	0.2731

Table(4, 3) : la distribution a posteriori de *delta* et *sigma*

Les analyses génèrent aussi le graph de la densité a posteriori (voir **figure(4, 1)**), pour le graph de la distribution a posteriori de *d*. Est-ce qu'il existe des traitements d'effet ?

La masse de la distribution a posteriori est à gauche de zéro où le point **inférieur** est **-0.3701**, et le point **supérieur** est **-0.1265**, donc l'évidence a posteriori suggère que *delta* est différent à **zéro** sur l'échelle, et les **Bétabloquants** a faiblissent la mortalité d'une crise cardiaque.

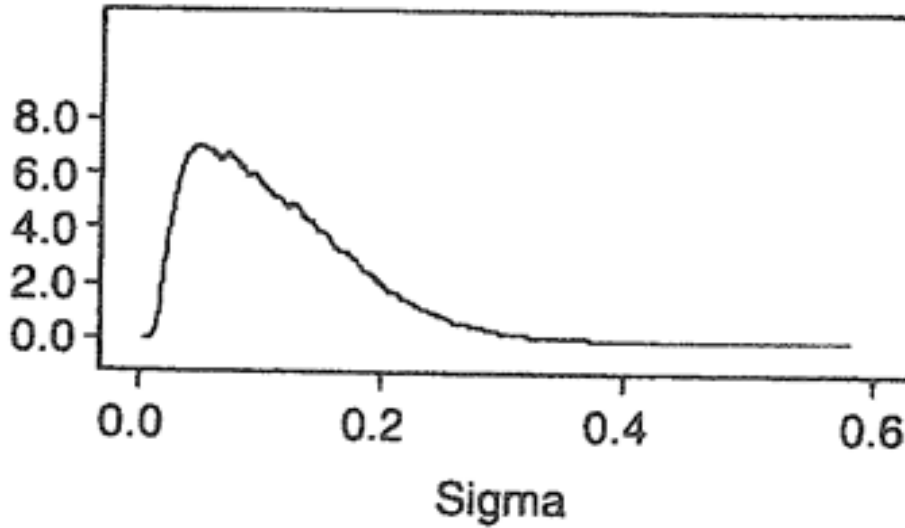


Figure(4, 1) : la distribution a posteriori de *d*

4.3. LA COMPARAISON DE LA MORTALITÉ DU GROUPE DE  
TRAITEMENT AVEC LE GROUPE DE CONTRÔLE  
CORRESPONDENT

---

Et la **figure(4, 2)** présente la distribution a posteriori de *sigma*.



**Figure(4, 2)** : densité a posteriori de **sigma**

***L'exécution de la simulation :***

La vraie puissance de la technique de MCMC est que l'analyse a posteriori est basée sur le modèle hierchique qui a plusieurs niveaux de paramètres. L'exemple précédent (22 essais) avait pour but de fournir une illustration de l'analyse bayésienne a posteriori et les programmes de statement sont donnés ici :

```

model
{
  for(i in 1 : Num) {
    rc[i] ~ dbin(pc[i] , nc[i])
    rt[i] ~ dbin(pt[i] , nt[i])
    logit (pc[i]) < -m[i]
    logit (pt[i]) < -m[i] + d[i]
    m[i] ~ dnorm(0.0 , 1.0E - 5)
    d[i] ~ dnorm(delta , tau)
  }
  delta ~ dnorm(0.0 , 1.0E - 6)
  tau ~ dgamma(0.001 , 0.001)

```

*Data*

```
list(rt = c(3, 7, 5, 102, 28, 4, 98, 60, 25, 138, 64, 45, 9, 57, 25, 33, 28, 8, 6, 32, 27, 22),
,
nt = c(38, 114, 69, 1533, 355, 59, 945, 632, 278, 1916, 873, 263, 291, 858, 154, 207, 251,
151, 174, 209, 391, 680),
rc = c(3, 14, 11, 127, 27, 6, 152, 48, 37, 188, 52, 47, 16, 45, 31, 38, 12, 6, 3, 40, 43, 39),
nc = c(39, 116, 93, 1520, 365, 52, 939, 471, 282, 1921, 583, 266, 293, 883, 147, 213, 122,
154, 134, 218, 364, 674),
Num = 22)
```

*Inits*

```
list(delta = 0, d.new = 0, tau = 1, m = c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0),
d = c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0))
```

C'est la Meta analyse qui combine l'information de 22 essais cliniques qui compare la mortalité du groupe de traitement avec le groupe de contrôle correspondent. Le statement précédent est interprété comme suit :

La **table(4, 1)** indique que dans le premier essai le groupe de traitement compte 38 *malades* dont 3 sont morts tandis que dans le groupe de contrôle il y a 39 *patients* dont 3 morts.

Pour le programme précédent :

*rt* colonne est le nombre de morts dans 22 essais pour le groupe de traitement.

*nt* colonne est le nombre de patients dans 22 essais de groupe de contrôle.

*nc* le nombre de patients dans 22 essais de groupe de contrôle.

*Num* = 22 est le nombre d'essais clinique dans la Meta analyse.

Les valeurs initiales **delta = 0** , **tau = 1** , dans 22 valeurs initiales (i.e, zéros) pour  $m_i$  et  $d_i$ .

Cette partie est une suite de la partie précédente. On suppose que nous voulons estimer le  $\theta$  de l'exemple binomial, où la densité est montrée dans la **Figure(4, 3)**. La distribution a posteriori est une Beta(21, 46) avec les caractéristiques suivantes :

moyenne=0.361, mediane=0.362, écart type=0.55, point inférieur 2 $\frac{1}{2}$ % = 0.254, point supérieur 2 $\frac{1}{2}$ % = 0.473.



### 4.3. LA COMPARAISON DE LA MORTALITÉ DU GROUPE DE TRAITEMENT AVEC LE GROUPE DE CONTRÔLE CORRESPONDENT

---

La moyenne et la médiane sont identiques implicitement par la symétrie de la parcelle de terrain sur la figure(4,3), où les points inférieurs et supérieurs de  $2\frac{1}{2}\%$  déterminent un intervalle de crédibilité à 95% de (0.254, 0.473) de  $\theta$ .

Les inférences à la population normale  $(\mu, \tau)$  sont légèrement plus exigeantes, parce que les deux paramètres sont inconnus. En supposant que la densité a priori est  $\pi(\theta) = \frac{1}{\tau}$ , la distribution a posteriori marginale de la moyenne de population  $\mu$  est une t-distribution avec **n-1** degrés de liberté de moyenne  $\bar{x}$  et de précision  $\frac{n}{s^2}$ . Donc la moyenne et la médiane sont identiques et fournissent un estimateur normal de  $\mu$ .

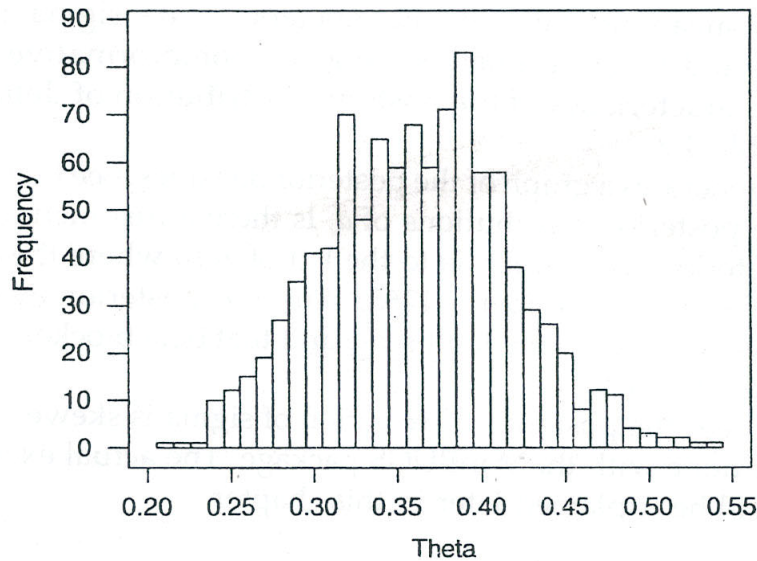
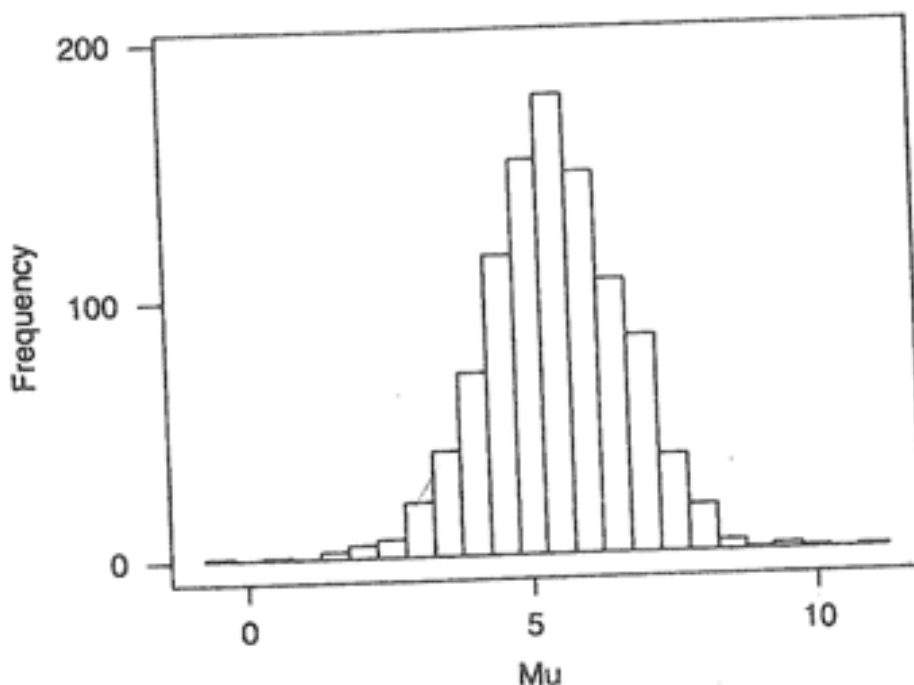


Figure (4,3) : distribution a posteriori de **theta**

**Le calcul de mu et tau** : D'après (Lyle D. Broemeling (2006)) on suppose l'échantillon de dix éléments suivant  $x = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$ , donc l'histogramme de **1000 valeurs** généré de t-distribution (9, 5.5, 1.090)

est donné dans la **figure(4, 4)** par :



**Figure(4, 4)** : distribution a posteriori de  $\mu$

L'intervalle de crédibilité à 95% est :

$$\left(x - t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}, x + t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}\right) = \left(5.5 - 0.822 \frac{3.028}{3.162}, 5.5 + 0.822 \frac{3.028}{3.162}\right) = (4.713, 6.287).$$

Par l'utilisation du même ensemble des données, les instructions de Win-BuGs suivantes sont utilisées pour analyser le problème :

```
model
{for(i in 1 : 10) {X[i] ~ dnorm(mu, tau)}}
mu ~ dnorm (0.0, 0.0001)
tau ~ dgamma (0.0001, 0.0001)
sigma < -{1/tau}
list (X = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10))
list (mu = 0 ,tau = 1)
```

$\mu$  et  $\tau$  sont indépendantes, l'analyse correspondente est donnée dans la

#### 4.4. LA DÉTERMINATION DE LA PUISSANCE D'UN ÉCHANTILLON BINOMIAL ET LA FONCTION PUISSANCE

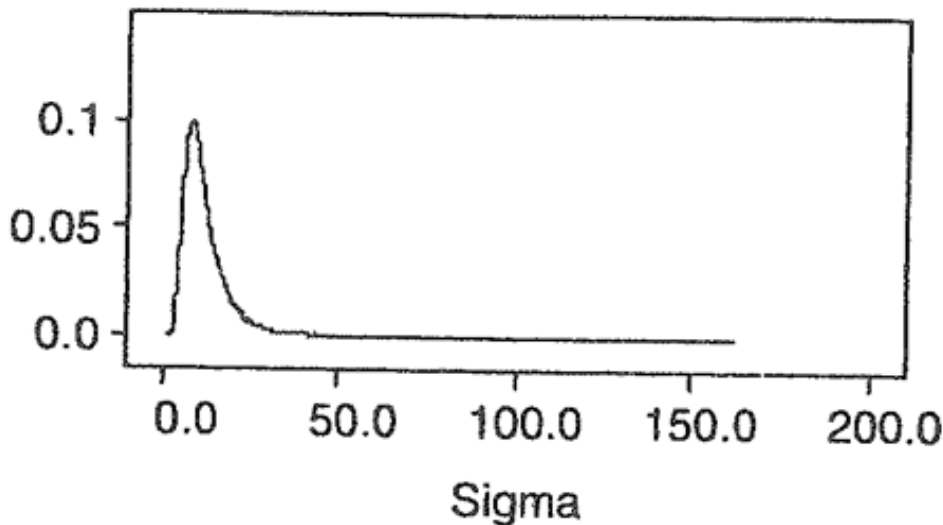
---

table suivante.

Paramètre	Moyenne	Dev.std	Médiane	Inférieur	Supérieur
<b>mu</b>	5.48	1.10	5.49	3.30	7.67
<b>sigma</b>	11.86	7.77	9.92	4.33	30.47

Table(4, 4) : distribution a posteriori de  $\mu$  et  $\sigma = \frac{1}{\sqrt{\tau}}$

La densité a posteriori de  $\mu$  est la même dans la **figure(4, 4)**, la densité a posteriori de **sigma** (voir *figure (4, 5)*) montre l'asymétrie à droite avec la médiane de 9.921, *moyenne* = 11.86 et l'intervalle de crédibilité à 95% est (4.33, 30.47)



Figure(4.5) : densité de *sigma*

#### 4.4 La détermination de la puissance d'un échantillon binomial et la fonction puissance

Cette application présente deux exemples d'application. Le premier consiste à calculer la fonction puissance pour  $H_0$  contre  $H_1$ , et le deuxième consiste à calculer la puissance pour un échantillon binomial avec l'information a priori.

On considère un échantillon aléatoire de la population de Bernoulli avec les paramètres  $n$  et  $\theta$ , où  $n$  est le nombre de malade et  $\theta$  est la probabilité de réponse. Soit  $X$  le nombre des réponses parmi  $n$  patients, et on suppose que l'hypothèse nulle est :

$$H_0 : \theta \leq \theta_0 \text{ contre } H_1 : \theta > \theta_0$$

L'information a priori de  $\theta$  est déterminée pour être  $Beta(\alpha, \beta)$ , donc la distribution a posteriori de  $\theta$  est  $Beta(x + \alpha, n - x + \beta)$  où  $x$  est le nombre observé de réponses parmi  $n$  malades. L'hypothèse nulle est rejetée et on accepte l'alternative lorsque :

$$\Pr[\theta > \theta_0/x, n] > \gamma$$

Où  $\gamma$  prend généralement de "grandes" valeurs 0.90 , 0.95 , 0.99.

L'équation précédente détermine la région critique du test, donc la fonction puissance est :

$$g(\theta) = \Pr_{X/\theta} \{ \Pr[\theta > \theta_0/x, n] > \gamma \}$$

Dans la donnée de la valeur  $\theta$ , la puissance est interprétée comme la simulation suivante :

- 1-choisir  $n$  et  $\theta$  et soit  $I = 0$
- 2-Générer  $X \sim Binomial(\theta, n)$
- 3-Générer  $\theta \sim Beta(x + \alpha, n - x + \beta)$
- 4-Si  $\Pr[\theta > \theta_0/x, n] > \gamma$ , laissez le compteur  $I = I + 1$  sinon laissez  $I = I$
- 5-Répéter (1)-(4)  $M$  fois, où  $M$  est "grand"
- 6-Choisir d'autres  $\theta$  et répété (2)-(5), donc la puissance est estimée comme  $\frac{I}{M}$ .

#### 4.4.1 La détermination de la fonction puissance pour $H_0$ contre $H_1$

On considère un essai typique où la quantité historique pour la réponse de thérapie est 0.20, donc l'essai est stoppé si le taux dépasse la valeur historique. La quantité de réponse est définie dans l'étude suivante :

$$H_0 : \theta \leq 0.20 \text{ contre } H_1 : \theta > 0.20$$

#### 4.4. LA DÉTERMINATION DE LA PUISSANCE D'UN ÉCHANTILLON BINOMIAL ET LA FONCTION PUISSANCE

---

Où  $\theta$  est la quantité de réponse de thérapie. L'hypothèse nulle est rejetée si la probabilité a posteriori de l'hypothèse alternative est supérieure à la valeur seuil  $\gamma$ . Lorsque l'a priori uniforme est appropriée, la courbe de la puissance du scénario précédent est computed (voir l'équation de simulation précédente) et avec les tailles de l'échantillon  $n = 125, 205, \text{ et } 500$ ,  $\gamma = 0.90, 0.95, 0.99$ , et  $M = 1000$ , et la valeurs nulle est  $\theta_0 = 0.20$  (voir la *table* (4, 5) pour la puissance du test pour les valeurs de  $\theta$ ).

On note que la puissance du test à  $\theta = 0.30$  et  $\gamma = 0.95$  est : **0.841, 0.958** et **0.999** pour  $N = 125, 205$  et **500** respectivement ; Le test bayésien se comporte d'une manière raisonnable, pour l'erreur de première espèce conventionnelle de **0.05**, et la taille de l'échantillon  $N = 125$ , peut être suffisante à détecté la différence **0.3** contre **0.2** avec la puissance **0.841** (voir la *table* (4, 5)). On doit premièrement utiliser la simulation précédente puis la programmation en Pass<sup>R</sup>.

$\theta$	$\gamma = 0.90$	$\gamma = 0.95$	$\gamma = 0.99$
<b>0</b>	0, 0, 0	0, 0, 0	0, 0, 0
<b>0.1</b>	0, 0, 0	0, 0, 0	0, 0, 0
<b>0.2</b>	0.107, 0.099, 0.08	0.047, 0.051, 0.05	0.013, 0.013, 0.008
<b>0.3</b>	0.897, 0.97, 1	0.841, 0.958, 0.999	0.615, 0.82, 0.996
<b>0.4</b>	1, 1, 1	1, 1, 1	0.996, 1, 1
<b>0.5</b>	1, 1, 1	1, 1, 1	1, 1, 1
<b>0.6</b>	1, 1, 1	1, 1, 1	1, 1, 1
<b>0.7</b>	1, 1, 1	1, 1, 1	1, 1, 1
<b>0.8</b>	1, 1, 1	1, 1, 1	1, 1, 1
<b>0.9</b>	1, 1, 1	1, 1, 1	1, 1, 1
<b>1.0</b>	1, 1, 1	1, 1, 1	1, 1, 1

**Table(4, 5) : la fonction puissance pour  $H_0$  contre  $H_1$  :125, 205, 500**

#### 4.4.2 La détermination de la puissance pour un échantillon binomial avec l'information a priori

Un échantillon binomial avec l'information a priori disponible avec **50 patients**, **10** parmi eux ont la réponse à la thérapie d'après les hypothèses nulles et alternatives précédentes, on rejette l'hypothèse nulle lorsque :

$$\Pr[\theta > \phi/x, n] > \gamma \tag{4.1}$$

Où  $\theta$  est indépendant de  $\phi \sim \text{Beta}(10, 40)$ . Comme précédent, on utilise la règle de simulation précédente, et la fonction puissance pour la région de confiance de l'équation (4,1) est calculée ( voir la *table(4, 6)*) avec la même taille de l'échantillon et des valeurs seuils comme dans la *table(4, 5)*). La puissance de test dans **0.758**, **0.865** et **0.982** pour  $\theta = 0.4$  , et  $N = 125, 205$  et **500** respectivement est calculée dans la table (4, 6). On voit l'importance de l'information a priori dans les tests d'hypothèses si l'hypothèse est rejetée avec la région critique suivante :

$$\Pr(\theta > 0.2/x, n) > \gamma \tag{4.2}$$

La puissance (voir la *table (4, 5)*) sera supérieure à la puissance correspondante (voir la *table(4, 6)*), déterminée par la région critique de l'équation (4,1). Donc les grandes taille d'échantillons sont nécessaires avec l'approche de l'équation (4,1) pour atteindre la même puissance comme le test donné par l'équation (4,2). Voici la *table(4, 6)* qui donne la puissance pour un échantillon binomial avec l'information a priori.

$\theta$	$\gamma = 0.90$	$\gamma = 0.95$	$\gamma = 0.95$
<b>0</b>	0,0,0	0,0,0	0,0,0
<b>0.1</b>	0,0,0	0,0,0	0,0,0
<b>0.2</b>	0.016 , 0.001, 0.000	0.002,0.000,0.000	0.000,0.000,0.000
<b>0.3</b>	0.629,0.712,0.850	0.362,0.374,0.437	0.004,0.026,0.011
<b>0.4</b>	0.996,0.999,1	0.973,0.998,1	*0.758,0.865,0.982
<b>0.5</b>	1,1,1	1,1,1	0.999,1,1
<b>0.6</b>	1,1,1	1,1,1	1,1,1
<b>0.7</b>	1,1,1	1,1,1	1,1,1
<b>0.8</b>	1,1,1	1,1,1	1,1,1
<b>0.9</b>	1,1,1	1,1,1	1,1,1
<b>1.0</b>	1,1,1	1,1,1	1,1,1

**Table(4, 6)** : puissance pour un échantillon binomial avec l'information a priori

## 4.5 Le calcul de la puissance pour deux échantillons binomial

D'après (**Lyle D. Broemeling (2006)**), cette application resoud un problème commun dans la statistique qui est la comparaison de deux populations binomiales, avec le calcul de la puissance de deux échantillons binomiales.

### 4.5.1 La comparaison de deux populations binomial

La comparaison de deux populations binomiales est implique l'hypothèse

$$H_0 : \theta_1 = \theta_2 \text{ contre } H_1 : \theta_1 \neq \theta_2$$

Où  $\theta_1$  et  $\theta_2$  sont des paramètres de la population de Bernoulli. Les deux paramètres de Bernoulli doivent être sensibles à deux modèles de diagnostique.

On suppose que la probabilité a posteriori de l'hypothèse nulle est  $\pi$ , et on signe à l'indépendance uniforme a priori de deux paramètres de Bernoulli. Il peut être montré que le test bayésien rejette  $H_0$  et accepte  $H_1$  si la probabilité a posteriori  $P$  de l'hypothèse alternative satisfait :

$$P > \gamma$$

où

$$P = D_2/D \tag{4.3}$$

et

$$D = D_1 + D_2$$

avec

$$D_1 = \frac{\left\{ \pi \binom{n_1}{x_1} \binom{n_2}{x_2} \Gamma(x_1 + x_2 + 1) \Gamma(n_1 + n_2 - x_1 - x_2 + 1) \right\}}{\Gamma(n_1 + n_2 + 2)}$$

où  $\Gamma$  est la fonction de gamma.

et

$$D_2 = (1 - \pi) (n_1 + 1)^{-1} (n_2 + 1)^{-1}$$

où  $\pi$  est la probabilité a priori de l'hypothèse nulle,  $X_1$  et  $X_2$  sont les nombres de réponses de deux populations binomiales avec les paramètres  $(\theta_1, n_1)$  et  $(\theta_2, n_2)$  respectivement, et on note la fonction de puissance

$$g(\theta_1, \theta_2) = \Pr_{x_1, x_2 / \theta_1, \theta_2} [ \theta > \gamma / x_1, x_2, n_1, n_2 ], \quad (\theta_1, \theta_2) \in (0, 1) (0, 1)$$

Où  $P$  est donné par l'équation (4,3).

### 4.5.2 Le calcul de la puissance pour deux échantillons binomial

On suppose  $n_1 = n_2 = 20$ , la taille des échantillons de deux groupes et la probabilité a priori de l'hypothèse nulle est **0.5**. La puissance de chaque point  $(\theta_1, \theta_2)$  est calculée par la simulation, avec  $\gamma = \mathbf{0.90}$ . Les valeurs sont données dans la table suivante :

$\theta_1 \backslash \theta_2$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	0.004	0.03	0.13	0.36	0.62	0.84	0.95	0.99	1	1
0.2	0.03	0.01	0.02	0.10	0.28	0.53	0.74	0.91	0.99	1
0.3	0.17	0.02	0.006	0.02	0.10	0.25	0.48	0.76	0.96	1
0.4	0.36	0.09	0.02	0.01	0.02	0.07	0.24	0.54	0.84	0.99
0.5	0.61	0.28	0.10	0.02	0.007	0.01	0.10	0.29	0.64	0.98
0.6	0.82	0.52	0.23	0.08	0.03	0.005	0.02	0.11	0.35	0.88
0.7	0.95	0.77	0.46	0.25	0.11	0.03	0.01	0.04	0.17	0.58
0.8	0.99	0.92	0.76	0.49	0.31	0.13	0.02	0.01	0.04	0.20
0.9	1	0.99	0.94	0.84	0.64	0.35	0.15	0.03	0.006	0.01
1.0	1	1	1	1	0.98	0.87	0.56	0.20	0.01	0.00

Table(4, 7) : puissance de deux échantillons binomiales

Lorsque la puissance est calculée pour les deux échantillons, on a le test bilatérale binomial avec **alpha = 0.013**, et les tailles des échantillons  $n_1 = n_2 = \mathbf{20}$  et  $(\theta_1, \theta_2) = (0.8, 0.2)$ , donc la puissance est **0.922** qui est faible par rapport à la puissance du test de Bayes, Pour  $\gamma$  donnée dans le test de Bayes, on doit donner une puissance égale à la puissance du test conditionnel, pour les valeurs de  $\theta_1$  et  $\theta_2$  considérées. L'erreur maximum de la première espèce du test de Bayes est approximativement à **0.013**.



# Conclusion

Dans ce travail, nous avons étudié la théorie mathématique des tests bayésiens et fait une comparaison avec les tests dans l'approche fréquentiste. Nous avons montré leurs intérêts et les conclusions significatives qu'elles peuvent conduire dans les essais cliniques. De plus nous avons étudié le motif des plans adaptatifs prenant en compte les réponses disponibles comme une solution alternative à la randomisation équilibrée classique. Ces plans sont explicitement intéressants puisqu'ils permettent pour des raisons éthiques d'exposer le moins de sujets possible aux traitements les moins efficaces.

Nous avons montré le rôle très important du facteur de Bayes et critiqué son utilisation.

Nous avons pu voir des exemples précis où à partir d'une étude pilote, les probabilités prédictives portant sur les limites de crédibilité fournissent un résumé utile pour aider au choix de la taille de l'échantillon d'une expérience. On peut également prédire le résultat final pour l'ensemble des données, dans le cas où les données prédictives peuvent aussi être utilisées pour aider à la décision d'interrompre une expérience si la probabilité prédictive apparaît insuffisante.

# Bibliographie

- [1] Jiunn Tzon Hwang, George Casella, Christian Robert, Martin T Wells and Roger H. Farrell. (1992). Estimation of accuracy in testing . The annals of statistics, vol. 20, NO. 1, 490 - 509
- [2] Albert, J. (1996), Bayesian Computation Using Minitab, Belmont , CA : Duxbury press. parfois
- [3] Gorui Bian. (1997). Bayesian Inference in location-scale Distributions with independent Bivariate Priors. vol. 6, NO. 1, pp.137-157
- [4] Sweeting T J (1999) On the construction of Bayes-confidence regions. JR stat Soc Ser B stat methodol 61(4) : 849-861.
- [5] Lecoutre, B. (1999). Two useful distribution for Bayesian predictive procedures under normal models. J. Statist. Planning and inference 77, 93-105.
- [6] Erik Holst. Poul thyregod and Peter-Th . Wilrich. (2001). On conformity testing and the use of tow stage procedures. interessante.
- [7] Bradley Efron and Robert Tibshirani. (2001). Empirical Bayes methods and false discovery rates for microarrays.
- [8] Russell V. Lenth, (2001), Some practical guidelines for effective sample-size determination.
- [9] Lelia Duley and Barbara Farrell. (2002). Clinical trials.
- [10] Biard Olivier, Deschamps Céline, Marlet Sandrine (2002). Méthods bayésiennes.
- [11] Peter Congdon, Applied bayesian Modelling . (2003). John & Sons, Ltd.
- [12] Gelman A, Carlin JB, Stern HS, Rubin DB (2004). Bayesian data analysis , 2nd edn. Texts in statistical science series, Champman & Hall / CRC , Boca Raton, FL.

- [13] Bruce Westerm, Filiz Garip. (2005). Bayesian analysis of comparative survey data.
- [14] Pierre Legendre, (2005). Introduction aux fonctions du langage R.
- [15] Andrew D. Martin. (2005). Bayesian analysis.
- [16] Pierre Bunouf, Bruno Lecoutre . (2006) . Bayesian priors in sequential binomial design.
- [17] Shein-Chung chow Mark chang. (2006). Adaptive design methods in clinical trials.
- [18] Lyle D. Broemeling. (2006). Bayesian biostatistics and diagnostic medicine.
- [19] Bruno Lecoutre. (2006). Former les étudiants et les chercheurs aux méthodes bayésiennes pour l'analyse des données expérimentales.
- [20] Bruno Lecoutre. (2006). Former les étudiants et les chercheurs aux méthodes bayésiennes pour l'analyse des données expérimentales.
- [21] Jayanta K. Ghosh, Mohan Delampady (2006) . An introduction to bayesian analysis theory and methods.
- [22] T.J. Cleophas. A. H. Zwinderman and T. F. Cleophas. (2006). Springer. Statistics applied to clinical trials.
- [23] Sheldon M. Ross, simulation.(2006), Elsevier Inc. All rights reserved.
- [24] Lecoutre M-P., Rovia K , Lecoutre B, Poitevineau J. (2006). People's intuitions about randomness and probability : an empirical study. Statistics Education Research Journal, 5(1), 20-35.
- [25] Robert, C.P, Le choix bayésien : Principe et implémentation. Springer-Verlag, Paris (2006).
- [26] Jérôme Dupuis (LSP-UPS) . (2007). Statistique bayésienne et algorithme MCMC.
- [27] Christophe Dutang & Alaeddine Falleh. (2007). Prediction based on certain uncertainties- a bayesian credit portfolio approach.
- [28] Fabien Campillo .(2007). Modélisation probabiliste et inférence bayésienne.
- [29] Eric Parent, Jacques Bernier. (2007). Le raisonnement bayésien.
- [30] Marin, J-M and Robert, C.P, Bayesian Core : A Practical Approach to computational bayesian statistics, Springer-Verlag, New York (2007).

- [31] Grouin J-M , Cosre M, Bunouf P, Lecoutre B (2007). Bayesian sample size determination in non-sequential clinical trials : Statistical aspects and some regulatory considerations. *Statistics in medicine*, 26,4914-4924.
- [32] Lecoutre B.(2007). Bayesian methods for experimental data analysis. In C.R . Rao, J. Miller & D.C.Rao (Eds), *Handbook of statistics : Epidemiology and Medical statistics (vol27)* , Amsterdam : Elsevier, 775-812.
- [33] Anne Philippe. *Statistique bayésienne, laboratoire de mathématiques Jean Lerary* (2007).
- [34] John Wiley & Sons, Ltd (2007), *Simulation and Monte carlo*.
- [35] Bunouf P, Lecoutre B (2008). On Bayesian estimators in multistage binomial designs. *Journal of statistical planning and inference*, 138, 3915-3926.
- [36] Lecoutre B, Elqasyr K (2008). Adaptative designs for multi-arm clinical trials : The play-the-winner rule revisited. *Communications in statistics-simulation*, 37, 590-601.
- [37] Robert C, Casella G (2008) A history of markov chain monte carlo-subjective recollections from incomplete data. **arxiv : 0808 . 2902 [stat .CO] , arxiv : 0808 . 2902.**
- [38] Michael S . Hamada, Alyson G. Wilson, C.Shane Reese, Harry F.Martz. *Bayesian Reliability* (2008) Springer science +business media. LLC.
- [39] Christian P.Robert. *Licence MI2, (2008-2009). Nouveaux outils informatiques pour la statistique exploratoire.*
- [40] Jim Albert .(2009). *Bayesian computation with R.*
- [41] Jhean-Larose S, Denhière G, Lecoutre B (2009). Etude développementale de l'interprétation de combinaisons conceptuelles nominales "Relation" et "Propriété". *Psychologie Française* ,54,121-135.
- [42] Cucal, J, Marin, J-M, Robert, C.P. and Tettington, D.M. A bayesian reassessment of nearest-neighbour classification. *Journal of the american statistical association*, March1, 2009, 104 (485) (2009).
- [43] Ch.Mélot , MD, PhD, MSCiBiostat. *Biomarqueurs et décision médicale. Réunion interface INSERM-SFAR-SRLF* (2009).
- [44] Jim Albert, *Bayesian computation with R.* Springer science+business media, LLC (2009).
- [45] Lawrence M. Friedman, Curt D.Furberg, David L.Demets. *Fundamentals of clinical trials*, springer science+business media, LLC 2010.

## *BIBLIOGRAPHIE*

---

- [46] Lawrence M. Friedman, Curt D.Furberg, David L.Demets. Fundamentals of clinical trials, springer science+business media, LLC 2010.
- [47] Shein-chung chow, p h.D .Bayesian adaptative methods for clinicals trials (2010).
- [48] José M. Bernardo and Adrian F.M. Smith (2010). Bayesian theory.
- [49] David la veccha and Fabio Trojani. (2010). American statistical association. Journal of the American statistical association.

## **Résumé**

L'objectif de ce mémoire est d'étudier les aspects décisionnels de l'inférence statistique dans les essais cliniques. Nous avons utilisé une modélisation bayésienne parce qu'elle offre plus de souplesse et d'objectivité dans le traitement statistique des données. Des applications numériques et les simulations ont illustré la méthodologie dans différents modèles pour le calcul du facteur de Bayes et probabilités prédictives.

## **Mots clés :**

- Statistique bayésienne
- Test bayésienne
- Essai clinique
- Facteur de Bayes
- P-Valeur

## **ABSTRACT**

The objective of this memory is to study the decisional aspects of the statistical inference in the clinical trials. We used a Bayesian modeling because it offers more flexibility in the statistical processing of data. Numerical applications and simulation illustrated methodology in various models for the calculation of the Bayes factor and probabilities predictive.

### **Key words:**

- Bayesian Statistic.
- Bayesian Test.
- Bayes Factor.
- p-value.

## ملخص:

الهدف من هذه المذكرة هو دراسة عملية صنع القرار من جانب الاستدلال الإحصائي في التجارب السريرية. لذلك استخدمنا نموذج بايز لأنه يوفر المزيد من المرونة والموضوعية في المعالجة الإحصائية للبيانات. وقد أظهرت التطبيقات الرقمية والمحاكاة المنهجية نماذج مختلفة لحساب عامل بايز والاحتمالات التنبؤية.

## الكلمات الدلالية

- إحصائيات بايز.
- اختبار بايز.
- التجارب السريرية.
- ف القيمة.
- عامل بايز.