

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

=====

UNIVERSITE CONSTANTINE 1  
FACULTE DES SCIENCES EXACTES

=====

DEPARTEMENT DE MATHEMATIQUES

N° d'ordre :..... /.... / 2015

N° de série : ..... / ..... / 2015

THESE PRESENTEE POUR L'OBTENTION

DU

DIPLOME DE DOCTORAT 3 ème CYCLE

EN MATHEMATIQUES

« Analyse bayésienne dans les essais cliniques »

Par

Ahlam LABDAOUI

OPTION

STATISTIQUE APPLIQUEE

Devant le jury :

Mr Z.MOHDEB	Prof. U des Frères Mentouri	Président
M H.MERABET	M.C.U des Frères Mentouri	Rapporteur
Mr N . HAMRI	Prof. C.U de Mila	Examineur
Mr M. BOUZIT	M.C.U Larbi Ben M 'Hidi O.E.B.	Examineur

Soutenu le 07/05/2015

# Table des matières

Remerciements

Résumé

Abstract

Introduction générale 5

1 L'analyse statistique bayésienne 6

**I Théorie de décision 7**

1.1 Introduction . . . . . 8

1.2 Définition des modèles . . . . . 8

1.2.1 Modèle statistique paramétrique . . . . . 8

1.2.2 Modèle statistique bayésien . . . . . 9

1.3 Lois a priori et a posteriori . . . . . 10

1.3.1 Introduction . . . . . 10

1.3.2 La loi a priori . . . . . 11

1.3.3 La loi a posteriori . . . . . 11

1.4 Lois a priori conjuguées . . . . . 12

1.5 Lois a priori non informatives . . . . . 12

1.5.1 La loi a priori de Jeffreys . . . . . 13

1.5.2 L'analyse avec la distribution a priori non informative . . . . . 14

1.6 Incorporation de l'information a priori en pratique . . . . . 16

1.7	Le poids de l'a priori dans la réponse bayésienne . . . . .	17
-----	---	----

**II Applications: Calcul de la puissance et de la taille de l'échantillon pour des données binaires 19**

1.8	Puissance bayésienne prédictive pour un design à deux étapes . . . . .	20
1.8.1	Introduction . . . . .	20
1.8.2	Critères de la taille de l'échantillon bayésienne pour les paramètres binomiale . . . . .	21
1.8.3	La prédiction dans les essais cliniques . . . . .	21
1.8.4	L'approche de la puissance prédictive . . . . .	22
1.8.5	Simulations . . . . .	24
1.8.6	Sommaire des résultats de simulation . . . . .	28
1.8.7	Conclusion . . . . .	28
1.9	La taille de l'échantillon d'une proportion par l'intervalle de confiance dans une approche bayésienne . . . . .	29
1.9.1	Introduction . . . . .	29
1.9.2	Détermination de la taille de l'échantillon pour les données binaires	30
1.9.3	Simulation . . . . .	31
1.9.4	Conclusion . . . . .	32

**2 Les techniques avancées de simulation: les méthodes MCMC 34**

2.1	Introduction . . . . .	34
2.2	Le cadre de l'estimation bayésienne . . . . .	34
2.2.1	Problèmes d'estimation paramétrique . . . . .	35
2.2.2	L'approche bayésienne . . . . .	35
2.2.3	Pourquoi faire de l'estimation bayésienne? . . . . .	35
2.3	Quelques notions sur les chaînes de Markov . . . . .	36
2.4	Modèles de mélanges et modèles de chaînes de Markov cachées . . . . .	38
2.4.1	Modèles de mélanges . . . . .	38
2.4.2	Modèles de chaînes de Markov cachées . . . . .	39

2.5	L'algorithme d'acceptation rejet . . . . .	40
2.6	L'algorithme de Hastings Metropolis . . . . .	41
2.6.1	L'algorithme de Metropolis . . . . .	41
2.6.2	Enoncé de l'algorithme de Hastings Metropolis . . . . .	42
2.6.3	Cas du noyau de transition symétrique . . . . .	42
2.7	L'algorithme du recuit simulé . . . . .	43
2.7.1	Introduction . . . . .	43
2.7.2	Enoncé de l'algorithme du recuit simulé . . . . .	44
2.8	L'échantillonnage de Gibbs . . . . .	44
2.8.1	Algorithme d'échantillonnage de Gibbs . . . . .	44
2.8.2	Exemple: loi gaussienne bidimensionnelle . . . . .	45
2.9	Logiciel de simulation MCMC : WinBUGS . . . . .	47
<b>3</b>	<b>Modèles linéaire et logit dans les essais cliniques avec WinBUGS</b>	<b>48</b>
3.1	Introduction . . . . .	48
3.2	Essais cliniques . . . . .	49
3.2.1	Définition d'un essai clinique . . . . .	49
3.2.2	Les différentes phases d'un essai clinique . . . . .	50
3.3	Modèles linéaire . . . . .	51
3.3.1	Choix de l'a priori des paramètres . . . . .	51
3.3.2	La fonction de vraisemblance . . . . .	52
3.3.3	La densité a posteriori . . . . .	53
3.4	Applications . . . . .	53
3.4.1	Modèle linéaire simple . . . . .	53
3.4.2	Modèle de régression linéaire multiple . . . . .	56
3.4.3	Interprétation des résultats . . . . .	60
3.5	Le modèle logistique . . . . .	60
3.5.1	Modèle logit . . . . .	60
3.5.2	Définition de la loi a posteriori . . . . .	61
3.5.3	Applications . . . . .	61
3.6	Conclusion . . . . .	68

<b>4 Les procédures bayésiennes prédictives dans les designs des essais expérimentaux avec des données binaires</b>	<b>69</b>
4.1 Introduction . . . . .	70
4.2 La méthodologie statistique . . . . .	71
4.2.1 Le design expérimental . . . . .	71
4.2.2 Le design séquentiel . . . . .	79
4.2.3 Comparaison entre les deux prédictions . . . . .	81
4.2.4 Conclusion . . . . .	83

**Conclusion générale**

**Perspectives**

**Bibliographie**

## Remerciements

---

### Remerciements

On peut croire a priori que les remerciements représentent la section <<facile >> d'une thèse. Ce n'est toute fois pas le cas! En effet, on ne veut pas oublier personne sans pour autant remercier le facteur. La réflexion consacrée à la rédaction de cette section peut alors être plus intense que celle menant à une preuve mathématique...

Dieu soit loué, c'est à grâce à son appui que je suis arrivée à ce stade.

Je tiens tout d'abord à remercier et de tout profond du cœur madame MERABET Hayet Maître de conférences. A, ma directrice de thèse pour son soutien scientifique, encouragements et conseils précieux qui m'ont permis de l'aboutissement de ce modeste travail de recherche.

Je remercie grandement Monsieur Pierre Druilhet Professeur des universités, Blaise Pascal-Clermont Ferrand 2, mon responsable de stages de recherche, pour son encadrement, le temps qu'il m'a accordé, ses explications dans le domaine de l'analyse bayésienne et sa disponibilité. .Ainsi toute l'équipe de statistique, université Blaise Pascal en particulier Stéphane Leger pour m'avoir aidée dans le domaine des essais cliniques.

Je remercie aussi Monsieur Rahmani F.L notre chef de département pour son soutenu.

Mes remerciements vont également au Monsieur Mohdeb.Z Professeur à l'université Constantine1 qui a bien voulu de présider le jury. De même, je remercie vivement Monsieur Hamri.N Professeur au centre universitaire de Mila et Monsieur Bouzit.M, Maître de conférences A, à l'université d'O.E.B., qui ont accepté d'examiner ce travail de thèse.

Je remercie également tous les enseignants de l'institut de mathématiques pour leurs aides, orientations et leurs disponibilités.

Je veux remercier aussi ma famille, mes amis pour m'avoir soutenu, encouragé et enduré durant les périodes de stress.

Un grand merci à Khecha Mohamed el Bachir, Benkoussas Abd eldjilil pour leur aide et leurs conseils en informatique.

## **Résumé**

---

### **Résumé**

Ce travail de thèse comporte quatre chapitres :

Le premier parle sur la théorie de décision et le choix de la loi a priori et a posteriori motivé par des applications de calcul de la puissance bayésienne et la taille de l'échantillon avec des données binaire, on a envisagé dans le deuxième chapitre les méthodes MCMC, et on appliqué ces méthodes au modèle de régression linéaire et le modèle logit dans le troisième chapitre. Finalement la prédiction dans le design expérimental et séquentiel est traitée dans le dernier chapitre.

**Mots clés** : Puissance bayésienne, Méthodes MCMC, Modèles de régression, Prédiction.

## **Abstract**

---

### **Abstract**

This thesis has four chapters:

The first talk on the theory of decision and choice of law a priori and a posteriori motivated by computational applications of Bayesian power and sample size with binary data, it was envisaged in the second chapter methods MCMC, and applied these methods to the linear regression model and the logit model in the third chapter. Finally the prediction in the experimental and sequential designs discussed in the last chapter.

**Keywords:** Bayesian Power, MCMC methods, Regression models, Prediction.

## Introduction générale

L'objet de cette thèse est de démontrer que l'approche de l'inférence statistique bayésienne est moderne, parfaitement adaptée aux outils informatiques de simulation et apte à répondre aux problèmes des modélisations les plus avancées dans la conception des designs des essais cliniques.

Cette thèse est structurée en quatre chapitres.

Dans **le premier chapitre**, nous présentons les fondements de l'inférence bayésienne, en insistant sur les spécificités de la modélisation a priori et de la théorie de décision. Puis nous mettons en œuvre explicitement les concepts précédemment introduits en présentant des applications sur le calcul de puissance bayésienne et la taille de l'échantillon.

Nous avons développé dans **le chapitre deux**, succinctement les principales approches des méthodes numériques bayésiennes, à savoir l'algorithme de Metropolis-Hastings et l'échantillonnage de Gibbs. Nous trouvons dans **le chapitre trois** une application par le logiciel Win BUGS d'une étude d'un modèle linéaire simple, multiple et logit, dans le cas de données réelles.

Enfinement **le dernier chapitre** est consacré au traitement de l'approche bayésienne prédictive appliquée à un test statistique fréquentiste. Nous examinons la procédure bayésienne prédictive dans le cadre de la planification expérimentale et des analyses séquentielles pour des essais à deux étapes et pour différentes lois a priori (uniforme, non informative et conjuguée). Nous avons illustré les procédures dans le cas de données binaires successives à l'aide de logiciel R et Matlab.

# Chapitre 1

## L'analyse statistique bayésienne

# Partie I

**Théorie de décision**

## 1.1 Introduction

L'objet principal de la statistique est de faire, à partir d'observations d'un phénomène aléatoire, une inférence au sujet de la loi générant ces observations en vue d'analyser le phénomène ou de prévoir un événement futur. Pour réduire la complexité du phénomène étudié, deux approches statistiques sont utilisées dans la littérature: modélisation non paramétrique et modélisation paramétrique. Dans la modélisation non paramétrique, on considère que l'inférence statistique doit prendre en considération la complexité autant que possible et donc on cherche à estimer la distribution de l'ensemble du phénomène, mettant en œuvre l'estimation des fonctionnelles (densité, régression, etc.). Par opposition, le modèle paramétrique consiste d'observations d'une variable  $X = \{X_1, \dots, X_n\}$ , d'une fonction de densité  $f(x|\theta)$ , où seul le paramètre  $\theta$ , appartenant à un sous-ensemble  $\Theta$  d'un espace vectoriel de dimension finie, n'est pas connu. D'une manière plus générale, une fois le modèle construit, on cherche à établir une inférence sur  $\theta$ , c'est-à-dire à utiliser les observations  $x$  afin d'évaluer  $\theta$ , en vue d'une décision liée à ces paramètres. Notons que la démarche statistique est fondamentalement une démarche d'inversion qui est de remonter des «effets» (observations) aux «causes» (paramètres). Pour certains auteurs, ces deux approches (fréquentiste ou probabiliste, bayésienne) s'opposent alors que d'autres voient en elles une certaine complémentarité. Cette opposition réside en fait dans la démarche même de ces deux approches.

## 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 loi de densité  $f_i(x_i|\theta)$  sur  $R^p$  dont le paramètre  $\theta_i$ , scalaire ou vectoriel est inconnu.

On voit alors que ce modèle se représente plus simplement sous la forme  $x \sim$

$f(x_1|\theta_1)$ , 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

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

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

**Définition 1.2.1** *Un modèle statistique paramétrique sur  $(\Omega, F, P)$  consiste en l'observation d'une variable aléatoire  $X$ , de loi  $f(x|\theta)$ , où seul le paramètre  $\theta$ , appartenant à un sous-ensemble  $\Theta$  d'un espace vectoriel de dimension finie, n'est pas connu.*

## 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  $\theta$ , les méthodes statistiques permettent de conduire, à partir de ces observations, une inférence sur  $\theta$ , 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,

$$l(\theta|x) = f(x|\theta) \tag{1-1}$$

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 utilise 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

$$\pi(\theta|x) = \frac{f(x|\theta) \pi(\theta)}{\int f(x|\theta) \pi(\theta) d\theta} \quad (1-2)$$

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 1.2.2** *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 Lois a priori et a posteriori

### 1.3.1 Introduction

L'aspect de l'analyse bayésienne qui est à la fois le plus délicat, le plus critiquable 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 lois a priori lorsque l'information a priori est trop coûteuse à obtenir ou peu fiable.

### 1.3.2 La loi a priori

Le choix des lois a priori est une étape fondamentale dans l'analyse bayésienne. Ce choix peut avoir différentes motivations. Les stratégies sont diverses. Elles peuvent se baser sur des expériences du passé ou sur une intuition, une idée que le praticien a du phénomène aléatoire qu'il est en train de suivre. Elles peuvent être également motivées par des aspects calculabilité. Enfin, ces stratégies peuvent également tenir compte du fait qu'on ne sait rien par le truchement des lois non informatives.

### 1.3.3 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, on a:

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

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:

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

**Exemple 1.3.1** calcule de la distribution a posteriori pour  $x \sim B(n, p)$  ,  $p \sim \text{Beta}(\alpha, \beta)$

(Avec  $\alpha = \beta = 1$ , dans le cas particulier de Bayes)

$$f(x|p) = \binom{n}{x} p^x (1-p)^{n-x} \quad ; \quad x = 0, 1, \dots, n$$
$$\pi(p) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1} \quad ; \quad 0 \leq p \leq 1$$

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.

## 1.4 Lois a priori conjuguées

Une des difficultés de l'approche bayésienne est le calcul de la loi a posteriori. Ce calcul est facilité lorsque loi a priori et loi a posteriori ont la même forme. Dans ce cas, on parle de loi a priori conjuguée.

Une famille conjuguée est une famille de lois liée au processus étudié et qui est particulièrement intéressante dans des modèles paramétriques ou 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 1.4.1** Une famille  $F$  de lois, sur  $\Theta$  est dite conjuguée (ou ferme par échantillonnage) si, pour tout  $\pi \in F$ , la loi a posteriori  $\pi(\theta|x)$  appartient également à  $F$ .

Dans ce cas, le praticien induit directement la forme de son estimateur dès qu'il a choisi sa loi a priori.

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

Dans la section 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.

## 1.5 Lois a priori non informatives

Une loi non informative est une loi qui porte une information sur le paramètre à estimer dont le poids dans l'inférence est réduit. Certains auteurs la définissent également

comme une loi a priori qui ne contient aucune information sur  $\theta$  ou encore comme une loi qui ne donne pas davantage de poids à telle ou telle valeur du paramètre. Par exemple, supposons  $\Theta$  un ensemble fini de taille  $q$ , une loi a priori non informative pourra être une loi de la forme :

$$P(\theta_i) = 1/q$$

On a équiprobabilité, les valeurs possibles de  $\theta$  se voient attribuer le même poids.

### 1.5.1 La loi a priori de Jeffreys

Une méthode proposée par Jeffreys (1961) permet de fabriquer des lois a priori non informative. Cette méthode utilise l'information de Fischer :  $I(\theta)$ . L'argument pourrait être le suivant.  $I(\theta)$  représente une mesure de la quantité d'information sur  $\theta$  contenue dans l'observation. Plus  $I(\theta)$  est grande, plus l'observation apporte de l'information. Il semble alors naturel de favoriser (au sens rendre plus probable suivant  $\pi(\theta)$ , les valeurs de  $\theta$  pour lesquels  $I(\theta)$  est grande; ce qui minimise l'influence de la loi a priori au profit de l'observation.

Le choix de ce type de loi conduit ainsi souvent à des estimateurs classiques du type maximum de vraisemblance. La règle de Jeffreys consiste donc à considérer des lois a priori de la forme :

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

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 près quand  $\pi^*$  est propre.

**Exemple 1.5.1** 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  $I(\theta)$  a les éléments suivants:

$$I_{ij}(\theta) = -E_{\theta} \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(x|\theta) \right] \quad i, j = 1, \dots, k \quad (1-3)$$

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

## 1.5.2 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:

$$\pi(\mu, \sigma^2) \propto \frac{1}{\sigma^2} \quad (1-4)$$

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

$$f(x_1, \dots, x_n|\mu, \sigma^2) = \prod_{i=1}^{i=n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2\sigma^2} (x_i - \mu)^2 \right] \quad (1-5)$$

la distribution a posteriori jointe de  $(\mu, \sigma^2)$  est donnée par :

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

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 marginal 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

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

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

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  $\frac{s^2}{n}$ .

Supposant 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

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  $\frac{(n-1)s^2}{2}$ .

## 1.6 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 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  $\theta \sim \text{Beta}(\lambda\theta^*, \lambda(1 - \theta^*))$ . 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  $\theta \sim \text{Beta}(0.2\lambda, 0.8\lambda)$ . 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  $I^* = [0.8, 1.2]$ .

## 1.7 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 bayésienne suivant:  $x_i|\theta \sim \text{Bernoulli}(\theta)$  et  $\theta \sim \text{Beta}(a, b)$  et  $\theta|x \sim \text{Beta}(a + x, b + n - x)$ , 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 celui de  $E(\theta)$  et  $\bar{x}$ , affectés respectivement des coefficients  $\frac{a+b}{a+b+n}$  et  $\frac{n}{a+b+n}$ .

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

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

Si  $\mathbf{a+b<n}$  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  $\mathbf{a+b} \rightarrow 0$  et  $\mathbf{a+b} \rightarrow \infty$ . Dans le premier cas, le poids de l'a priori est nul, et  $E(\theta|x) \rightarrow \bar{x}$  qui est la réponse classique, dans le second cas, le poids des données est nul, et  $E(\theta|x) \rightarrow E(\theta)$  qui ne dépend plus de  $x$ .

# Partie II

**Applications: Calcul de la puissance et de la taille de l'échantillon pour des données binaires**

## 1.8 Puissance bayésienne prédictive pour un design à deux étapes

### 1.8.1 Introduction

L'approche bayésienne consiste à synthétiser les données et le jugement afin de parvenir à des conclusions sur les quantités inconnues et faire des prédictions. Méthodes bayésiennes sont devenues de plus en plus populaires ces dernières années, notamment dans la recherche médicale, et bien qu'il existe un certain nombre de livres sur l'analyse bayésienne, peu d'essais cliniques de couverture et les applications bio-statistiques en détail. Bayésienne approches aux essais cliniques et d'évaluation de Santé fournit une vue d'ensemble de ce domaine qui évolue rapidement, y compris les idées de base bayésiennes, les distributions antérieures, les essais cliniques, les études d'observation, synthèse de données et l'analyse coût-efficacité. Couvre un large éventail de sujets essentiels, le renforcement des bases à des techniques plus avancées telles que la puissance prédictive.

(Spiegelhalter et al. 2004, voir la section 6.5.5). Donner une explication simple et l'exemple de la notion de distribution de puissance prédictive. Si l'incertitude sur les quantités qui entrent dans le calcul de l'énergie conventionnelle (par exemple des valeurs moyennes et écart-type) est exprimée sous la forme de distributions antérieures, alors le pouvoir prédictif peut être considéré simplement comme la distribution qui est directement induite ou implicite par ces a priori : en fonction de la puissance conditionnelle et les priors. Spiegelhalter al donner un exemple de comment cela peut être obtenu par simulation Monte Carlo dans un cas simplifié (LW Huson. 2009). Cette idée de base peut être étendue pour inclure à la fois la croyance préalable sur les quantités incertaines et également des renseignements obtenus à partir des données historiques, il devient alors une procédure entièrement bayésienne. La répartition de la puissance prédictive fournit un résumé complet de ce qu'est un bayésien considérerait comme étant raisonnables de croire à la puissance de l'étude. Rubin et Stern (1998) fournissent une discussion théorique plus détaillée de l'utilisation et de l'interprétation des distributions de puissance prédictive.

Il est naturel de se demander dans le milieu d'un essai quelle est la probabilité que l'essai se tire une conclusion ou d'une autre, ou même d'atteindre une conclusion quelconque. Probabilités prédictives fournissent un mécanisme pour résoudre ce problème. Les probabilités prédictives sont plus faciles à comprendre dans le contexte des résultats binaires. Cette section se concentre sur les résultats binaires, bien que les mêmes principes s'appliquent plus largement à d'autres résultats, tels que le temps de l'événement, ( Satoshi Teramukai, un, Takashi Daimonb et Sarah Zoharc, 2012).

### 1.8.2 Critères de la taille de l'échantillon bayésienne pour les paramètres binomiale

Détermination de la taille de l'échantillon pour l'estimation précise d'un paramètre binomial est sans doute la situation de conception la plus commune rencontrée par les statisticiens. Les méthodes de taille de l'échantillon bayésienne utilisent l'information préalable sur le paramètre binôme plutôt qu'une estimation ponctuelle, et entièrement compte de l'incertitude dans les données prédites (E. Cyr mLAN 2008). Soit  $p$  le paramètre binomial être estimée sur la base d'une taille de  $n$  de l'échantillon, nous supposons que le modèle avant-vraisemblance suivante:  $p \sim \text{Beta}(a, b)$ ,  $a, b > 0$ ,  $\text{Beta}(a, b)$  indique une beta distribution de paramètres  $a$  et  $b$ :  $\pi(p) = \frac{1}{B(a, b)} p^{a-1} (1-p)^{b-1}$ , et  $x|p \sim \text{Bin}(n, p)$ ,  $n \geq 2$ ,  $\text{Bin}(n, p)$  représente la distribution binomiale, de paramètres  $n$  et  $p$ . En conséquence, la distribution prédictive marginale de  $x$  est la bêta-binomiale avec

$$p(x|n, a, b) = \frac{B(a+x, b+n-x)}{B(a, b)}$$

Pour un point de données de l'échantillon  $x$  donné, la distribution a posteriori de  $p$ ,  $\pi(p|x, n, a, b)$  et  $B(a+x, b+n-x)$

$$\pi(p|x) = \frac{p^{a+x-1} (1-p)^{b+n-x-1}}{B(a+x, b+n-x)} = \text{Beta}(a+x, b+n-x) \quad (1-6)$$

### 1.8.3 La prédiction dans les essais cliniques

Une des grandes forces de l'approche bayésienne est la facilité de faire des prévisions concernant les événements d'intérêt. Des exemples de prédictions utiles: La probabilité de succès

ou futilité si l'essai se poursuit, la probabilité prédictive de la réussite du procès en vertu d'une conception donnée, la probabilité qu'un patient donné dans l'essai de répondre à un traitement donné, la probabilité que l'avenir du patient répondra à un traitement donné. Nous pourrions évaluer une décision de conception ou d'arrêt en fonction de la probabilité prédictive du nombre total des patients (à l'intérieur et à l'extérieur d'essai) qui obtiennent le traitement supérieur. Nous équilibrons les besoins des patients actuels et futures prévisions fournit la discipline et l'importance pragmatique à la recherche empirique. (Gary Koop, Dale J. Poirier et Justin L 2007) et (BERRY S. M, CARLIN BP, LEE JJ, MULLER P. 2011 ) Supposons une divinité vous a dit les valeurs de tous les paramètres inconnus dans votre modèle afin que l'estimation et tests d'hypothèses sont devenues caduques. Que feriez-vous avec votre modèle? La réponse évidente est d'utiliser pour ce qu'il est censé faire: Faites ex déclarations de probabilité ante sur observables futures. Supposons que l'ensemble de l'information consiste en l'union des données passées  $Y = y$ , ce qui donne la fonction de vraisemblance paramétrique  $L(\theta)$ , et d'autres informations sous la forme d'une distribution a priori  $p(\theta)$ . La distribution d'échantillonnage d'un  $Y_*$  out-of-échantillon (peut-être un vecteur) donnée  $Y = y$  et  $\theta$  serait une distribution prédictive acceptable si  $\theta$  étaient connus, mais sans la connaissance de  $\theta$  celui-ci ne peut pas être utilisé. Ensuite, la distribution de probabilité p prédictive bayésienne  $p(y_*|y)$  est

$$\begin{aligned} p(y_*|y) &= \frac{p(y_*, y)}{p(y)} \\ &= \int_{\theta} \frac{p(y_*, y, \theta)}{p(y)} d\theta \\ &= \int_{\theta} p(y_*|y, \theta) \left[ \frac{p(\theta) p(y|\theta)}{p(y)} \right] d\theta \\ &= \int_{\theta} p(y_*|y, \theta) p(\theta|y) d\theta \\ &= E_{\theta|y} [p(y_*|y, \theta)] \end{aligned}$$

#### **1.8.4 L'approche de la puissance prédictive**

De nombreux auteurs ont préconisé le calcul de la puissance prédictive, en moyenne conditionnelle pouvoir sur les valeurs de  $p$  dans un calcul bayésien. Nous sommes amenés à une approche bayésienne, mais toujours avec un test fréquentiste à l'esprit. Formellement, la prédiction utilise la distribution a posteriori de  $p$ , donné un avant et les données disponibles

à l'analyse intérimaire. Pour l'inférence sur une proportion, les calculs sont particulièrement simple si nous choisissons une distribution a priori conjugué Beta (Lecoutre et al, 1995).

a priori:  $p \sim \text{Beta}(a, b)$

a posteriori :  $p|x \sim \text{Beta}(a + x, b + n - x)$

prédictive:  $y |x \sim \text{Binomial} - \text{Beta}(a + x, b + n - x; n)$

Supposons que  $X \sim B(n, p)$  and  $p \sim \text{Beta}(a, b)$ . Lors de la première étape avec  $n_1$  sujets, la distribution a posteriori de  $p$  donnée  $X$  est donnée par:

$$\pi_1(p|x) = \text{Beta}(a + x, b + n_1 - x) \tag{1-7}$$

De même, à la fin de la deuxième étape avec  $n$  sujets, la distribution a posteriori de  $p$  donnée  $X$  est:

$$(p|x) = \text{Beta}(a + x, b + n - x) \tag{1-8}$$

pour tester l'hypothèse null :  $H_0: \theta \leq 0$  contre une hypothèse alternative  $H_1: \theta > 0$ , la signification bayésienne

$$P_\beta = p(\theta < 0 | data) < \alpha_B$$

En conséquence, le point d'arrêt de l'essai de coupure est choisie de manière à ce que la puissance bayésien à la première étape est  $(1 - \beta_1)$ , au niveau de signification bayésienne  $\alpha_{B_1}$ . De même, le  $n$  est choisi de telle sorte que la puissance bayésien à la deuxième étape est  $(1 - \beta)$  au niveau de signification de Bayes  $\alpha_\beta$  ( BERRY S. M, CARLIN BP, LEE JJ, MULLER P. 2011 ). Basé sur (1-7) et (1-8), la puissance conditionnelle et la puissance prédictive peuvent être dérivées. Compte tenu de  $X$  sur les  $n_1$  patients qui répondent à la première étape, la probabilité ( ou puissance conditionnelle ) d'avoir des réponses supplémentaires au moins  $y$  sur les  $n_2$  patients supplémentaires  $n_2$  à la deuxième étape est donné par:

$$p(y|x, n_1, n_2) = \sum_{i=y}^{n_2} \binom{n_2}{i} \left(\frac{x}{n_1}\right)^i \left(1 - \frac{x}{n_1}\right)^{n_2-i} \tag{1-9}$$

La puissance prédictive ou la probabilité prédictive d'avoir au moins  $y$  intervenants sur  $m$  patients supplémentaires, étant donné le taux de réponse observé de  $x/n_1$  à la première étape, est donnée par:

$$\begin{aligned}
 p(y|x, n_1, n_2) &= \int_0^1 p(Y \succeq y|p, n_2)\pi(p|x) dp & (1-10) \\
 &= \int_0^1 \sum_{i=y}^{n_2} \binom{n_2}{i} p^i (1-p)^{n_2-i} \frac{p^{a+x+1} (1-p)^{b+n_1-x-1}}{\text{Beta}(a+x, b+n_1-x)}
 \end{aligned}$$

$$p(y|x, n_1, n_2) = \sum_{i=y}^{n_2} \binom{n_2}{i} \frac{B(a+x+i, b+n_1-x-i)}{B(a+x, b+n_1-x)} \quad (1-11)$$

### 1.8.5 Simulations

Dans cette section, nous calculons la formule (1.11) avec le logiciel Matlab, nous considérons une étude de phase II, une hypotension comparant un traitement d'essai avec un agent de contrôle actif. En supposant que la taille de l'échantillon pour la première étape,  $n_1 = 23$ , et la taille de l'échantillon pour la seconde étape,  $n_2 = 20$ , avec un temps donné,  $x = 16$  réponses et l'effet du traitement est estimé avec une distribution de bêta ( $a, b$ ).

#### Résultats de simulations

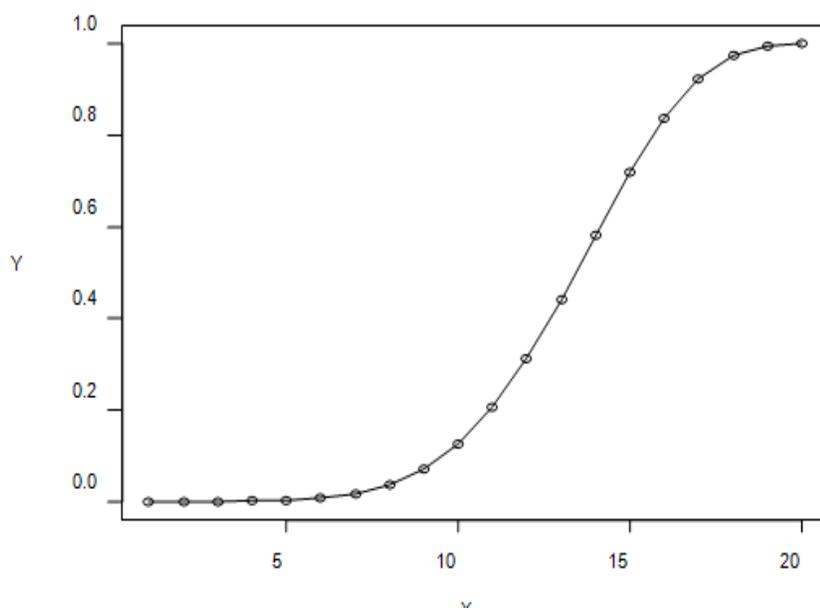
**Calcul de puissance prédictive avec une a priori uniforme**  $x = 16, n_1 = 23, n_2 = 20, a = b = 1$

trials	Y	$p(y/x, n_1, n_2)$
1	1	0.0000006
2	2	0.0000045
3	3	0.0002201
4	4	0.0008388
5	5	0.0026428
6	6	0.0071527
7	7	0.0170525
8	8	0.0364397
9	9	0.0706728
10	10	0.1255437
11	11	0.2056582
12	12	0.3122622
13	13	0.4412050
14	14	0.5820869
15	15	0.7195876
16	16	0.8372853
17	17	0.9230035
18	18	0.9733817
19	19	0.9949055
20	20	0.9999995

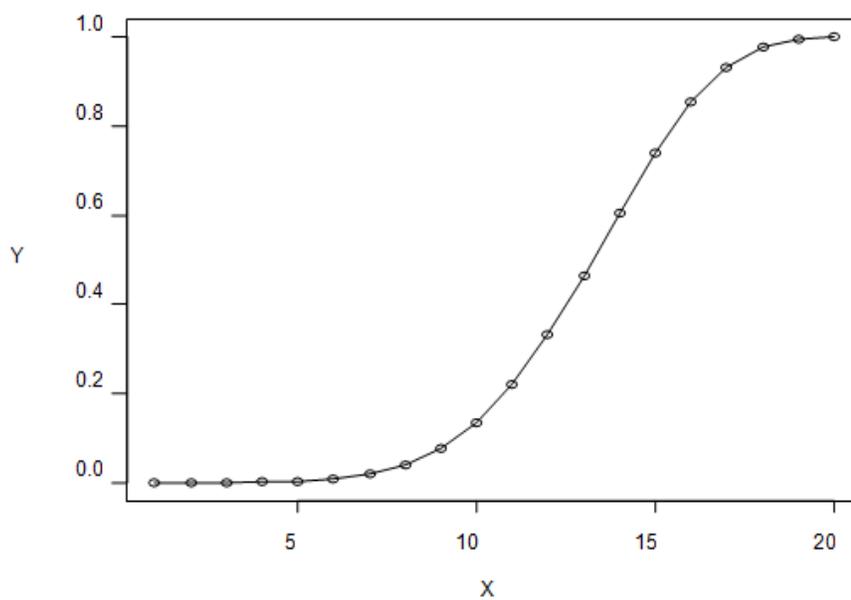
**Calcul de puissance prédictive avec une a priori non informative au sens de Jeffreys**

$$x = 16, n_1 = 23, n_2 = 20, a = b = 0.5$$

trials	Y	$p(y/x, n_1, n_2)$
1	1	0.0000006
2	2	0.0000048
3	3	0.0002385
4	4	0.0009130
5	5	0.0028836
6	6	0.0078103
7	7	0.0186020
8	8	0.0396457
9	9	0.0765646
10	10	0.1352247
11	11	0.2199210
12	12	0.3310849
13	13	0.4633414
14	14	0.6050448
15	15	0.7402081
16	16	0.8528441
17	17	0.9323519
18	18	0.9774064
19	19	0.9958497
20	20	0.9999994



Figure(1, 2) : puissance prédictive avec une *a priori* uniforme,  $a = b = 1$



Figure(1, 3) : puissance prédictive avec une *a priori* non informative au sens de Jeffreys,  $a = b = 0.5$

## 1.8.6 Sommaire des résultats de simulation

Le processus bayésien, au lieu de produire une estimation ponctuelle de la puissance pour le nouvel essai clinique, produit une distribution prédictive (ou postérieur) qui résume l'incertitude sur la puissance d'un nouveau procès prévu, et indique les valeurs plausibles pour cette puissance. La distribution de la puissance prédictive dérivée en utilisant les données binaires de Bayes est résumée sur la figure 1-2 et la figure 1-3.

D'après la première table avec une loi a priori uniforme,  $beta(1, 1)$  et  $x = 16, n_1 = 23, n_2 = 20$  la probabilité de prédiction ou de la puissance prédictive

dans le dernier essai,  $y = 20, p(y|x) = 0.9999995$ .

Dans le deuxième cas avec une loi a priori non informative, a priori de Jeffreys,  $beta(0.5, 0.5)$  et  $x = 16, n_1 = 23, n_2 = 20$ , la puissance prédictive

,  $y = 20, p(y|x) = 0.9999994$ .

D'après les deux tableaux on peut conclure que: la puissance prédictive est élevée et converge vers 1.

## 1.8.7 Conclusion

L'utilisation de l'analyse bayésienne de puissance prédictive est parfois considérée comme un curieux mélange de philosophies bayésiennes et fréquentistes. La " quantité incertaine " qui est estimée la puissance de l'étude, et la puissance d'une étude est un concept fermement fréquentiste. L'analyse bayésienne nous dit, en effet, ce degré de croyance pour attacher à un plage de valeurs de puissance estimée. L'analyse de la puissance prédictive semble encore peu utilisée dans la pratique, et l'une des raisons pour cela peut être que d'une analyse bayésienne classique doit soit utiliser des techniques d'analyse, qui dans la plupart des cas, pour des raisons de traçabilité, restreint sévèrement les hypothèses de distribution qui peuvent être faites ou doit faire usage de logiciels spécialistes bayésienne de tel que le R, BUGS Win, Matlab, ...

Le principal avantage de la bayésienne des données binaires avec une a priori bêta ( $a, b$ ), la technique est que ce calcul est simple à réaliser et peut facilement être programmé dans la plupart des paquets et avec la plupart des langages de programmation. Cela en fait une option attrayante pour les calculs couramment pratiqués comme les analyses de puissance (Chow SC, J. 2008).

Les résultats présentés ici montrent que, dans cet exemple, au moins, la répartition de la puissance de prédiction à partir des données binaires de Bayes est très similaire à celle produite à partir d'une analyse bayésienne plus classique en utilisant R.

Analyse de la puissance prédictive n'est pas la seule approche bayésienne pour la détermination de la taille de l'échantillon et de puissance. Un résumé pratique de certaines des méthodes alternatives de Bayes est donnée par (Chow et al 2008). Une description plus détaillée d'autres méthodes dont certains sont très différents de celui présenté ici sont également donnés par (Grouin et al 2007) et (Pezeshk 2003).

## **1.9 La taille de l'échantillon d'une proportion par l'intervalle de confiance dans une approche bayésienne**

### **1.9.1 Introduction**

La détermination de la taille de l'échantillon est l'un des problèmes les plus fréquemment utilisés et une vaste littérature disponible sur le sujet, y compris les deux fréquentiste (Latchine 1981, Lemeshow et al., 1990) et les approches bayésiennes (Adcock 1988, Pham - Gia et Turkkkan 1992, Joseph, Wolfson et du Berger 1995, Adcock 1997). Dans de nombreuses situations d'échantillonnage rencontrées dans la pratique, cependant, la méthodologie spécifique a échoué. Par exemple, tandis que les formules standards sont disponibles pour le calcul de la taille d'échantillon requise pour les paramètres binomiaux. Prenons l'exemple suivant, qui a motivé le développement de notre méthodologie. Estimation précise de la prévalence du tabagisme actif et passif à prendre des décisions éclairées sur l'impact des bébés de poids de tabac. Une étude épidémiologique menée chez les femmes dans la salle de livraison de 2 Juin 2003 jusqu'au 1er Juin 2004 sont fournies pour estimer la prévalence de ces derniers chez les femmes enceintes de la province de l'Auvergne en France. Supposons que l'on veut déterminer la taille de l'échantillon nécessaire que la puissance est de 90 % avec une précision de mesure de,  $l = 1\%$ , et le facteur d'effet de la fréquence la plus basse que l'alcool de la mère qui est prévu pour  $p = 4\%$ . la formule de la taille de l'échantillon connu sur la base

de l'approximation normale de la distribution binomiale peut être utilisée . Cela donne

$$n = \left( \frac{z_{\alpha/2}}{l} \right)^2 p(1 - p) \quad (1-12)$$

Dans notre étude, la taille de l'échantillon nécessaire pour prendre une décision ,  $n = 1150$  , avec  $z_{\alpha/2} = 1.64$ .

Les méthodes de la taille de l'échantillon bayésienne utilisent l'information préalable sur le paramètre plutôt que d'une estimation ponctuelle , et entièrement compte de l'incertitude dans les données prédites, donc offre une alternative intéressante aux formules fréquentiste.

Les méthodes bayésiennes peuvent éviter le recours à des approximations normales en utilisant exactement la plus forte densité postérieure ( HPD ) intervalles plutôt que moins efficaces intervalles égaux à queue. Cependant, en utilisant des méthodes bayésiennes n'est pas sans défis , à la fois numérique et en choisissant parmi les nombreux critères bayésiens qui ont été publiés dans la littérature . Dans ce papier articulaire nous discuter d'une solution bayésienne critères de taille de l'échantillon pour l'intervalle seront intégrés avec les méthodes d'estimation de la prévalence de partir d'une distribution a priori  $p$  , nous allons calculer la taille de l'échantillon plus petit pour lequel une  $(1 - \alpha)$  postérieure crédible intervalle de la prévalence aura une largeur totale de  $l$  avec une forte probabilité .

### **1.9.2 Détermination de la taille de l'échantillon pour les données binaires**

L'inférence et la décision des problèmes statistiques concernant les proportions peuvent être traitées en utilisant une analyse bayésienne. Dans l'approche bayésienne pour les statistiques, une tentative est faite d'utiliser toutes les informations disponibles (tous deux de l'échantillon et l'information a priori ) afin de réduire le montant de l'incertitude inhérente à un problème d'inférence ou la prise de décision. Comme de nouveaux renseignements, il est combiné avec toute information précédente pour former la base de méthodes statistiques .

#### **L'intervalle de confiance pour les proportions**

Un intervalle de confiance pour une proportion de la population  $p$  , basée sur un échantillon de taille  $n$  de cette population , avec l'échantillon de proportion  $\hat{p}$  est:

$$\left[ \hat{p} - U_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + U_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

### Le choix et l'utilisation d'une distribution a priori

La question est: Comment doit-on utiliser l'importance de l'information a priori? L'analyse bayésienne permet une utilisation efficace de l'information a priori à travers la déclaration d'une distribution a priori. La distribution a priori devrait, bien entendu, tenir compte des décideurs une information a priori, qui peut se produire dans une grande variété d'états. La distribution a priori d'un paramètre  $p$  peut être une fonction de densité de fonction de probabilité ou la probabilité à exprimer notre degré de croyance au sujet de la valeur de  $p$ . L'observation a priori d'un échantillon d'une variable aléatoire  $X$  dont la fonction de distribution dépend de fonctions des différentes distributions  $p$ . peut être caractérisé comme «a priori». Dans cette étude, nous avons pris une loi a priori bêta  $(a, b)$  notre paramètre  $p$  :

$$p|X \sim \text{Beta}(a + \sum X_i, b + n - \sum X_i)$$

### La taille de l'échantillon nécessaire avec longueur cible

D'après l'intervalle de confiance d'une proportion, le longueur de l'intervalle égale à:

$$l = \frac{2 * U_{1-\alpha/2}}{\sqrt{n}} * \sqrt{\hat{p} * (1 - \hat{p})}$$

on note

Esperance de longueur  $E(l) = l_0$  où:

$$l_0 = E \left( \sqrt{\hat{p} * (1 - \hat{p})} \right)$$

Donc

$$n = \left( \frac{2 * U_{1-\alpha/2}}{l_0} \right)^2 \left( \frac{\text{Beta}(a + 1/2, b + 1/2)}{\text{Beta}(a, b)} \right)^2 \quad (1-13)$$

### 1.9.3 Simulation

Dans cette section, nous avons utilisés le logiciel R pour faire notre simulation

### Algorithme d'estimation de la taille de l'échantillon

1. donnée une valeur par simulation avons de  $n$
2.  $p\{k\}$  suit une loi  $\text{beta}(a,b)$
3.  $x\{k\}$  suit une loi binomiale  $(n, p\{k\})$
4. donnée une valeur pour le seuil  $\alpha$
5. calcul de longueur
6. calcul de longueur.

### Exemple

$n=100$

$a=10$

$b=10$

$\alpha=0.05$

$K=1000$

pour  $k$  (in  $1:K$ ) faire

$p\{k\}$  suit une loi  $\text{beta}(a,b)$

$x\{k\}$  suit une loi binomiale  $(n, p\{k\})$

$z_{\text{star}} = \text{qnorm}(1-\alpha/2)$

calcul de  $l_0 = 2 * z_{\text{star}} * (\text{sqrt}(p\{k\}(1-p\{k\})/n))$

calcul de moyenne de  $l = 1/K * \sum l_0\{k\}$

$l_0=0.1959$

la taille de l'échantillon avec une longueur cible  $l_0$

lorsque  $l_0 = 0.1959$ ,

$n=153$  patients.

### 1.9.4 Conclusion

Lors de la planification d'une enquête par échantillon, ou dans de nombreuses formes de système d'armes d'essai, une étape est toujours atteinte au cours de laquelle une décision doit être prise au sujet de la taille de l'échantillon. La décision est importante. Un échantillon trop grand implique un gaspillage de ressources, et un échantillon trop petit diminue l'utilité des résultats. La décision ne peut pas toujours être satisfaisante, car souvent nous ne possédons

pas suffisamment d'informations pour être sûr que notre choix de la taille de l'échantillon est la meilleure.

Cette étude analyse la méthode bayésienne pour déterminer le nombre d'échantillons qui sont nécessaires pour produire une taille de l'intervalle de confiance pour un choix dans une étude épidémiologique est d'estimer la prévalence du tabagisme actif et passif en proportion chez les femmes enceintes. Nous comparons la taille de l'échantillon de méthodes bayésiennes avec des méthodes classiques où une distribution a priori bêta est utilisée.

On peut conclure avec nos résultats que les méthodes bayésiennes, tenant compte d'une information externe pertinente, permettent une réduction du nombre de sujet.

# Chapitre 2

## Les techniques avancées de simulation: les méthodes MCMC

Ce chapitre traite des méthodes de Monte Carlo par chaînes de Markov, méthodes que l'on appelle en anglais Monte Carlo Markov Chain, d'où l'acronyme MCMC.

### 2.1 Introduction

Les méthodes de Monte Carlo par Chaîne de Markov (et abrégée en MCMC) sont apparues il y a 50 ans pour la physique statistique, ont 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  $(\theta^{(m)})$  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. La simulation de Monte Carlo vise à envisager l'estimation par simulation « aléatoire » plutôt que par analyse algébrique.

### 2.2 Le cadre de l'estimation bayésienne

Les méthodes MCMC sont surtout utilisées dans le cadre de l'estimation bayésienne

### 2.2.1 Problèmes d'estimation paramétrique

Supposons que l'on dispose d'un certain nombre d'observations qui  $X_0, X_1, \dots, X_{N-1}$  dépendent aléatoirement d'un paramètre inconnu  $\theta$ . On connaît la loi des observations sachant  $\theta$ . L'estimation paramétrique consiste à estimer  $\theta$  au vu des observations  $X_0, X_1, \dots, X_{N-1}$ .

Par exemple, les variables aléatoires  $X_i$  sont indépendantes et identiquement distribuées selon la loi gaussienne de moyenne  $\theta$  et de variance 1, loi notée  $N(\theta, 1)$ . La loi forte des grands nombres dit que l'estimateur de la moyenne empirique

$$\hat{\theta}_N = \frac{1}{N} \sum_{i=0, N-1} X_i \quad (2.1)$$

est un "bon estimateur" de la vraie moyenne,  $\theta$ , puisque  $\hat{\theta}_N$  converge presque sûrement vers  $\theta$  si  $\theta = E(X)$  est fini.

### 2.2.2 L'approche bayésienne

L'approche bayésienne consiste à considérer le paramètre inconnu  $\theta$  comme étant lui-même une variable aléatoire. La variable aléatoire  $\theta$  suit une certaine distribution a priori, notons le  $\pi\theta$ .

Notons  $X$  les données dont on dispose pour faire l'inférence statistique;  $X$  est généralement multidimensionnel, par exemple  $X = (X_0, X_1, \dots, X_{N-1})$ . Notons  $\pi(x|\theta)$  la densité de la loi de  $X$  sachant  $\theta$ .

La densité de  $\theta$  sachant  $X = x$  (densité de la loi a posteriori) a pour expression:

$$\pi(\theta|x) = \frac{\pi(x|\theta)\pi(\theta)}{\pi(\theta)} \quad (2.2)$$

### 2.2.3 Pourquoi faire de l'estimation bayésienne?

Les avantages de l'estimation bayésienne sont multiples.

- En considérant  $\theta$  comme une variable aléatoire on peut intégrer un certain nombre d'a priori physiques sur  $\theta$ .

Par exemple, dans un problème d'estimation de la moyenne d'une loi gaussienne, si l'on sait que, pour des raisons physiques, cette moyenne ne peut être comprise qu'entre 0 et 1 on peut stipuler que la loi a priori de  $\theta$  est la loi uniforme sur  $[0, 1]$ . Il résulte de l'équation

3.9 que l'estimateur produit ne pourra alors prendre ses valeurs que dans l'intervalle  $[0, 1]$  puisque la densité de  $\theta$  sachant  $x$  sera nulle en dehors de l'intervalle  $[0, 1]$ .

- Il est des applications dans lesquelles l'approche bayésienne s'impose naturellement.

## 2.3 Quelques notions sur les chaines de Markov

**Méthodes MCMC: définitions** Une méthode de Monte Carlo par chaine de Markov est une méthode de simulation qui produit une chaine de Markov  $(X_n)_n$  qui converge en distribution vers la loi objectif  $\pi(\bullet)$ .

L'objectif de cette section est de présenter quelques notions relatives aux chaines de Markov et de préciser quelques propriétés de convergence de ces chaines.

*Généralités*

**Chaine de Markov: définition**

Soit un espace probabilisé  $(\Omega, A, P)$ .

Un processus aléatoire à temps discret  $(X_n)_{n \in N}$  est une chaine de Markov si et seulement si la loi de  $X_n$  sachant  $X_{n-1}, X_{n-2}, \dots, X_0$  est la même que celle de  $X_n$  sachant  $X_{n-1}$ ; autrement dit:

$$P(X_n \in A_n | X_{n-1} \in A_{n-1}, \dots, X_0 \in A_0) = P(X_n \in A_n | X_{n-1} \in A_{n-1}) \quad (2.3)$$

pout tout  $A_0, \dots, A_n$  éléments de la tribu  $A$ .

On peut dire, de façons équivalente, que  $(X_n)_n$  est une chaine de Markov si et seulement si, conditionnellement à  $X_{n-1}$ , la variable aléatoire  $X_n$  est indépendante de  $X_{n-2}, X_{n-3}, \dots, X_0$ .

**Homogénéité**

Si, de plus, la loi de  $X_n$  sachant  $X_{n-1} = x_{n-1}$  ne dépend pas de l'indice  $n$ , la chaine de Markov est homogène:

$$P(X_{n+1} \in B | X_n \in A) = P(X_{m+1} \in B | X_m \in A), \forall m, n \in N \quad (2.4)$$

**Exemple 2.3.1** : *processus autorégressif d'orde 1, AR(1)*

$$X_n = \theta X_{n-1} + \varepsilon_n \quad (2.5)$$

avec  $(\varepsilon_n)_n$  une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d.).  $(X_n)_n$  est une chaîne de Markov homogène.

**Exemple 2.3.2 : marche aléatoire**

$$X_n = X_{n-1} + \varepsilon_n \tag{2.6}$$

avec  $(\varepsilon_n)_n$  une suite indépendante et identiquement distribuée.  $(X_n)_n$  est une chaîne de Markov homogène.

**Noyau de transition**

Soit un espace probabilisable  $(\Omega, A)$ .

Un noyau de transition est une fonction  $P(\bullet, \bullet)$  définie de  $\Omega \times A$  dans  $[0, 1]$  qui vérifie les deux propriétés suivantes:

- (i) pour tout  $x$  dans  $\Omega$ ,  $A \in A \rightarrow P(x, A)$  est une mesure de probabilité.
- (ii) pour tout  $A$  dans  $A$ ,  $x \rightarrow P(x, A)$  est mesurable.

Dans le cas où la mesure  $P(x, \bullet)$  est à densité on appelle aussi noyau de transition la densité  $K(x, y)$  de noyau  $P(x, \bullet)$  :

$$\forall A \in A, P(x, A) = \int_A K(x, y) dy \tag{2.7}$$

**Noyau de transition d'une chaîne de Markov homogène:**

Soit une chaîne de Markov homogène  $(X_n)_n$  ; la loi de  $(X_n)_n$  est caractérisée par la donnée de:

- (i) la distribution initiale de la chaîne (distribution de  $X_0$ )
- (ii) le noyau de transition de la chaîne:

$$P(x, A) = P(X_{n+1} \in A \setminus X_n = x) \tag{2.8}$$

**Cas des chaînes de Markov à valeurs dans un espace discret:**

Dans ce cas,  $\Omega$  est dénombrable, par exemple  $\Omega = \{1, 2, \dots, K\}$  ou encore  $\Omega = N$ . Le noyau de transition  $P(\bullet, \bullet)$  s'écrit alors sous la forme d'une matrice; l'élément  $P(i, j)$  situé sur la ligne  $i$  et sur la colonne  $j$  de la matrice de transition  $P$  est égal à :

$$P(i, j) = P(X_{n+1} = j \setminus X_n = i) \tag{2.9}$$

La matrice  $P$  vérifie les deux propriétés suivantes, qui font d'elle une matrice stochastique:

$$\begin{aligned} P(i, j) &\geq 0, \forall i, j \in \Omega & (2.10) \\ \forall i \in \Omega, \sum_{j \in \Omega} P(i, j) &= 1 \end{aligned}$$

La matrice  $P$  est appelée matrice des probabilités de transition de la chaîne  $(X_n)_n$ .

## 2.4 Modèles de mélanges et modèles de chaînes de Markov cachées

### 2.4.1 Modèles de mélanges

Soit  $X$  une variable aléatoire. On dit que  $X$  est distribuée selon un mélange à  $K$  composantes si il est possible de compléter  $X$  par une variable aléatoire  $Y$  à valeurs dans un ensemble à  $K$  éléments, de telle façon que la loi de  $X$  conditionnellement à  $Y$  soit une loi plus simple que la loi de  $X$ .

Généralement  $X$  est appelée variable observée et  $Y$  est appelée variable cachée, car elle n'est pas observée directement, même si on peut faire une inférence sur  $Y$  au vu de l'observation  $X$ .

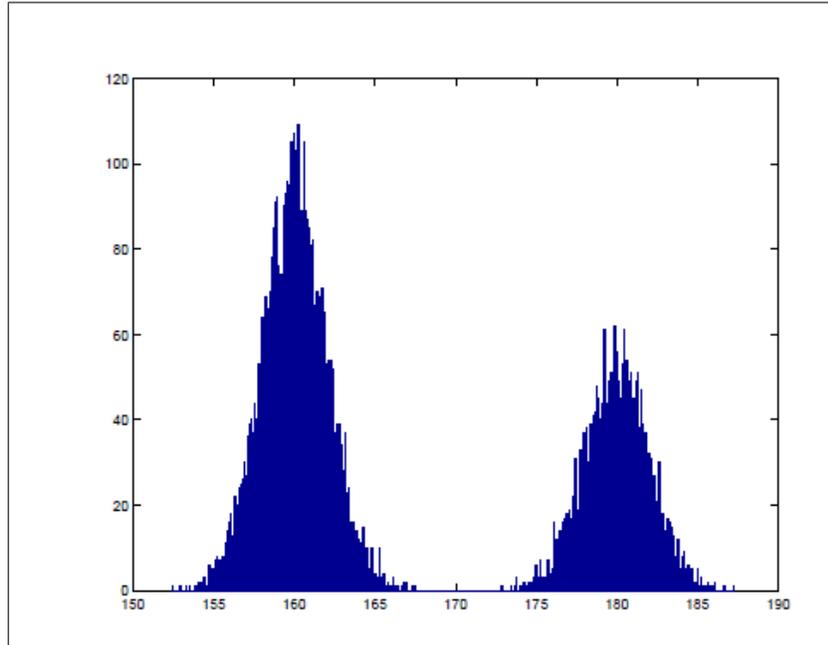
Supposons par exemple que  $X$  représente la taille d'un individu, pris au hasard dans une population où il y a 65% de femmes et 35 d'hommes. Si l'individu est un homme sa taille est distribuée selon une loi gaussienne de moyenne  $m_1 = 180$  (en cm) et de variance  $\sigma_1^2 = 2^2$  (en cm carré). Si l'individu est une femme, sa taille est distribuée selon une loi gaussienne de moyenne  $m_1 = 160$  (en cm) et de variance  $\sigma_1^2 = 2^2$  (en cm carré). Notons  $Y$  le sexe de l'individu:  $Y = F$  si l'individu est une femme, ce qui se produit dans 52% des cas et  $Y = H$  si l'individu est un homme, ce qui se produit dans 48% des cas.

La loi de  $X$  peut être spécifiée de deux façons équivalentes; la loi de  $X$  peut être spécifiée par sa densité de probabilité:

$$p(x) = 0.35 \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2\sigma_1^2} (x - m_1)^2\right) + 0.65 \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2\sigma_2^2} (x - m_2)^2\right) \quad (2.11)$$

ou, de façon équivalente, en utilisant le sexe de l'individu, elle peut être spécifiée par:

$$\begin{aligned}
 P(Y = H) &= 0.35 & P(Y = F) &= 0.35 & (2.12) \\
 [X|Y = H] &\sim N(m_1, \sigma_1^2) & [X|Y = F] &\sim N(m_2, \sigma_2^2)
 \end{aligned}$$

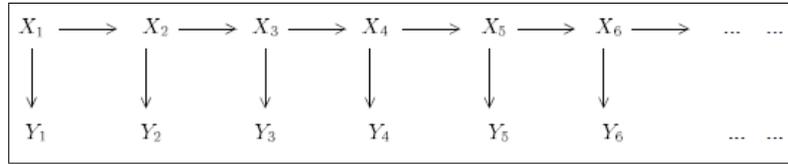


**Figure(2,1)** : représente l'histogramme de la taille d'une population de  $T = 10000$  individus

### 2.4.2 Modèles de chaînes de Markov cachées

Une chaîne de Markov cachée  $(X_t)_{t \in N}$  est une fonction aléatoire d'une chaîne de Markov homogène  $(Y_t)_t$  à valeurs dans un espace fini modèle le plus fini. Dans le modèle le plus simple de chaîne de Markov cachée, le modèle M1-M0, la chaîne de Markov  $(Y_t)_t$  est d'ordre 1, ce qui signifie que  $Y_t$  ne dépend que de  $Y_{t-1}$  et l'observation  $X_t$  ne dépend que de la variable cachée  $Y_t$ . La structure de dépendance de la chaîne de type M1-M0 est résumée sur la figure 3.2.

La loi de la chaîne de Markov cachée est spécifiée par la distribution initiale de la chaîne  $(Y_t)_t$ , par la matrice des probabilités de transition de  $Y_t$ , et par la loi de  $X_t$  conditionnellement à  $Y_t = y_i$ ,  $y_i$  étant un élément de l'espace des états de la chaîne non observée.



**Figure(2, 2)** : représente la structure de dépendance d'une chaîne de type M1-M0.

Considérons la chaîne de Markov cachée M1-M0 suivante.  $(Y_t)_t$  prend ses valeurs dans  $\{1, 2\}$ . La distribution initiale de  $(Y_t)_t$  est donnée par :

$$P(Y_1 = 1) = P(Y_2 = 2) = 0.5 \quad (2.13)$$

La matrice des probabilités de transition de  $(Y_t)_t$  est égale à

$$A = \begin{bmatrix} 0.99 & 0.01 \\ 0.02 & 0.98 \end{bmatrix} \quad (2.14)$$

ce qui signifie que le temps de séjour moyen dans l'état 1 est de  $1/0.01=100$  "coups" et le temps de séjour moyen dans l'état 2 est de  $1/0.02=50$  "coups".

Quand  $Y_t$  est dans l'état 1  $X_t$  est distribuée selon une loi gaussienne de moyenne 1 et de variance 0.01, et quand  $Y_t$  est dans l'état 2  $X_t$  est distribuée selon une loi gaussienne de moyenne 3 et de variance 0.25:

$$\begin{aligned} [X_t \setminus Y_t = 1] &\sim N(m_1 = 1, \sigma_1^2 = 0.01) \\ [X_t \setminus Y_t = 2] &\sim N(m_2 = 3, \sigma_1^2 = 0.25) \end{aligned} \quad (2.15)$$

## 2.5 L'algorithme d'acceptation rejet

L'algorithme d'acceptation rejet s'énonce de la façon suivante:

1. Simuler une variable aléatoire  $Y \sim f(y)$ .  
 $f(y)$  est la densité de probabilité de  $Y$ .
2. Simuler une variable aléatoire  $U \sim U([0, 1])$ .
3. Si  $U \leq \frac{\pi(Y)}{Mf(Y)}$  accepter  $Y$  comme réalisation de la variable aléatoire  $X$ , c'est-à-dire faire

$$X = Y$$

sinon retour en 1.

On termine dès qu'une réalisation  $Y$  a été acceptée.

Le principe de l'algorithme d'acceptation rejet est de "réutiliser" des variables aléatoires  $Y$  simulées selon une loi à densité  $f(y)$  pour produire une variable aléatoire  $X$  distribuée selon  $\pi(x)$ .

On suppose bien sur que l'on sait simuler à moindre cout la loi à densité  $f(\bullet)$  et on va se servir de ce simulateur pour simuler la loi  $\pi(\bullet)$ . la loi  $f(\bullet)$  est appelée la loi instrumentale, tandis que la loi  $\pi(\bullet)$  est appelée la loi objectif.

Dans l'algorithme énoncé ci-dessus  $M$  est une constante qui est telle que, presque surement, on a

$$\pi(Y) \leq Mf(Y)$$

ce qui veut dire que pour tout  $y$  tel que  $f(y) \neq 0$  on a

$$\pi(y) \leq Mf(y)$$

## 2.6 L'algorithme de Hastings Metropolis

### 2.6.1 L'algorithme de Metropolis

L'algorithme de Hasting Metropolis est un autre algorithme utilisé pour simuler une variable aléatoire  $X$  définie par sa densité de probabilité  $\pi(\cdot)$ . Comme l'algorithme d'acceptation rejet présenté dans la section précédente, l'algorithme de Hastings Metropolis réutilise, par un mécanisme d'acceptation rejet, des variables aléatoires  $Y$  produites selon une loi instrumentale.

L'algorithme d'acceptation rejet peut être utilisé pour simuler plusieurs réalisations indépendantes de la loi à densité  $\pi(\bullet)$ . Par contre, les variables aléatoires  $X(1), X(2), \dots, X(i) \dots$  produites par l'algorithme de Hastings Metropolis ne sont pas indépendantes;  $X(2)$  dépend de  $X(1)$ ,  $X(3)$  dépend de  $X(2)$ ,  $X(4)$  dépend de  $X(3)$ ... Le processus  $(X(i), i = 1, 2, \dots)$  est un processus de Markov à temps discret. La chaîne de Markov  $(X(i), i = 1, 2, \dots)$  produite par l'algorithme de Hastings Metropolis converge en distribution vers la loi objectif  $\pi(\bullet)$ .

## 2.6.2 Enoncé de l'algorithme de Hastings Metropolis

On simule une suite de la façon suivante:

1. Sachant  $X_t = x_t$  générer  $Y_t \sim q(y|x_t)$ .

La valeur obtenue est notée  $y_t$ .

2. Calculer le ratio de Hastings Metropolis:

$$\alpha(x_t, y_t) = \min \left\{ 1, \frac{\pi(y_t) q(x_t|y_t)}{\pi(x_t) q(y_t|x_t)} \right\}$$

3. Simuler  $U_t \sim U[0, 1]$ . La valeur obtenue est notée  $u_t$ .

Si  $u_t < \alpha(x_t|y_t)$  alors  $X_{t+1} = y_t$  sinon  $X_{t+1} = x_t$ .

Sous certaines conditions sur le noyau de transition  $q(\bullet, \bullet)$  la variable aléatoire  $X(t)$  converge en distribution vers la loi objectif  $\pi(\bullet)$ .

Le principe d l'algorithme de Hastings Metropolis est de perturber de façon stochastique la valeur courante  $x_t$ , et d'accepter aléatoirement le résultat  $y_t$  de cette perturbation aléatoire.

## 2.6.3 Cas du noyau de transition symétrique

Par définition le noyau de transition  $q(\bullet, \bullet)$  est symétrique quand :

$$q(y|x) = q(x|y)$$

Cette propriété est en particulier vérifiée quand le noyau de transition est une marche aléatoire symétrique:

$$y_t = x_t + z_t \quad \text{avec} \quad z_t \sim \Phi(\bullet) \quad \text{et} \quad \Phi(z) = \Phi(-z)$$

La perturbation aléatoire  $x_t \rightarrow y_t$  est centrée sur  $x_t$  et symétrique autour de  $x_t$ . c'est par exemple la marche aléatoire à accroissement gaussien centré

$$\Phi(z) = \frac{1}{2\pi\sigma} \exp\left(-\frac{z^2}{2\sigma^2}\right)$$

ou encore la marche aléatoire +1 ou -1 équiprobables:

$$\Phi(z) = \frac{1}{2}\delta_1(z) + \frac{1}{2}\delta_{-1}(z)$$

Dans le cas d'une marche aléatoire symétrique le noyau de transition pour expression:

$$q(y|x) = \Phi(|y - x|)$$

et il est clair que ce noyau de transition vérifie la condition de symétrie.

Dans le cas d'un noyau de transition symétrique le ratio de Hasting-Metropolis a une expression simple:

$$\alpha(x_t, y_t) = \min \left\{ 1, \frac{\pi(y_t)}{\pi(x_t)} \right\}$$

On tombe alors sur l'algorithme de Metropolis (1953).

## 2.7 L'algorithme du recuit simulé

### 2.7.1 Introduction

L'algorithme de recuit simulé est un algorithme d'optimisation dérivé de l'algorithme de Hastings Metropolis. L'objectif de l'algorithme du recuit simulé est de minimiser une fonction  $U(x)$  /

$$x^* = \underset{x}{\text{Arg min}} U(x)$$

L'algorithme du recuit simulé

- (i) ne nécessite aucune hypothèse sur la régularité de la fonction à optimiser
- (ii) a la possibilité de sortir des bassins d'attraction des extrema locaux de  $U(x)$  sous des conditions qui portent sur la fonction de température  $T_i$  seulement

Le principe de l'algorithme du recuit simulé est de se donner pour distribution objectif la distribution

$$\pi(x) \propto \exp(-U(x))$$

et de simuler la distribution  $\pi(x)$  par une méthode de Hastings Metropolis. Le fait que la distribution objectif soit définie à un facteur multiplicatif près ne pose pas de problème quand elle est simulée par Hastings Metropolis.

## 2.7.2 Énoncé de l'algorithme du recuit simulé

L'algorithme du recuit simulé s'énonce comme suit:

Sachant  $x_i$ , générer  $x_{i+1}$  de la façon suivante:

1. Générer  $Y \sim q(y|x_i)$ .
2. Calculer la probabilité d'acceptation

$$\alpha_i(x_i, y) = \min \left\{ 1, \frac{\exp(-U(y|T_i)) q(x_i|y)}{\exp(-U(x_i|T_i)) q(y|x_i)} \right\}$$

3.  $U \sim U[0, 1]$ .

$$\begin{cases} y \text{ si } U < \alpha_i(x_i, y) \\ x_i \text{ si } U > \alpha_i(x_i, y) \end{cases}$$

4. Faire décroître la température  $T_i \rightarrow T_{i+1}$ .

## 2.8 L'échantillonnage de Gibbs

### 2.8.1 Algorithme d'échantillonnage de Gibbs

L'algorithme d'échantillonnage de Gibbs est utilisé pour simuler une distribution  $\pi(x)$  quand

1.  $x$  admet une décomposition de la forme

$$x = (x_1, x_2, \dots, x_N)$$

2. et les lois conditionnelles

$$p_i(x_i|x_{-i})$$

sont simulables aisément, en notant  $x_{-i} = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, x_N)$  et  $p_i(x_i|x_{-i})$  la densité de la loi de  $x_i$  sachant  $x_{-i}$ .

Le principe de l'algorithme de Hastings Metropolis est de réactualiser "composante par composante". Un passage de l'échantillonneur de Gibbs se décompose de la façon suivante: sachant  $x^t = (x_1^t, x_2^t, \dots, x_N^t)$ , générer  $x^{t+1}$  comme suit:

$$\begin{aligned}
 x_1^{t+1} &\leftarrow p_1(x_1 \setminus x_2^t, \dots, x_N^t) \\
 x_2^{t+1} &\leftarrow p_2(x_2 \setminus x_1^{t+1}, x_3^t, \dots, x_N^t) \\
 &\vdots \\
 x_i^{t+1} &\leftarrow p_i(x_i \setminus x_1^{t+1}, \dots, x_{i-1}^{t+1}, x_{i+1}^t, \dots, x_N^t) \\
 &\vdots \\
 x_N^{t+1} &\leftarrow p_N(x_N \setminus x_1^{t+1}, \dots, x_{N-1}^{t+1})
 \end{aligned}$$

Sous certaines conditions, la chaîne de Markov  $(x^t)_{t=1,2,\dots}$  qui est produite par l'algorithme d'échantillonnage de Gibbs converge, en distribution, vers la loi jointe  $p(x)$  quand  $t$ , le nombre d'itérations de l'algorithme, converge vers l'infini.

### 2.8.2 Exemple: loi gaussienne bidimensionnelle

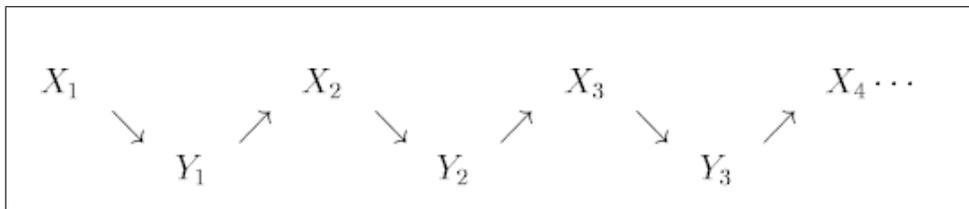
Donnons-nous pour loi objectif la loi gaussienne bidimensionnelle

$$(X, Y) \sim N \left( (0, 0), \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

La loi de  $X$  sachant  $Y=y$  et la loi de  $Y$  sachant  $X=x$  sont des gaussiennes:

$$\begin{aligned}
 [X|Y = y] &\sim N(\rho y, 1 - \rho^2) \\
 [Y|X = x] &\sim N(\rho x, 1 - \rho^2)
 \end{aligned}$$

Le principe de l'algorithme de Hastings Metropolis est, successivement, de simuler  $Y$  sachant  $X=x$  puis  $X$  sachant  $Y=y$  et ainsi de suite, en utilisant les lois conditionnelles.



Partant, par exemple, de  $(X, Y) = (10, 10)$ , on obtient un échantillon dont la distribution est représentée sur la figure 2-3. La droite de régression de  $Y$  sur  $X$  de pente  $\rho$  est en particulier bien visible sur la figure 2-3.

$(X_t)_t$  est une chaîne de Markov; le couple  $((X_t, Y_t))_t$  est également une chaîne de Markov puisque  $Y_{t+1}$  ne dépend que de  $X_{t+1}$  qui elle même ne dépend que de  $Y_t$ ...

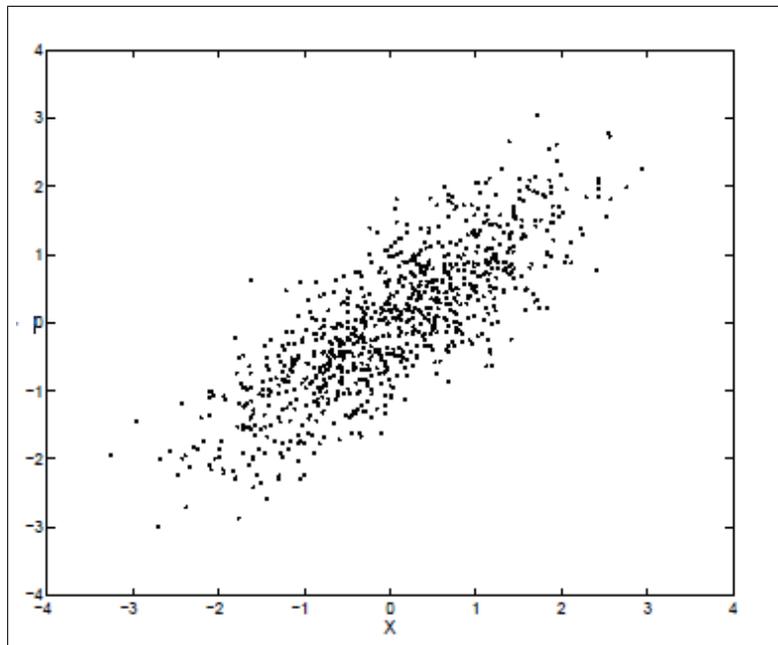
On peut montrer que, partant de  $X_0 = x_0$ , la loi  $X_t$  de est une loi gaussienne de moyenne  $\rho^{2t} x_0$  et de variance  $1 - \rho^{4t}$

$$[X_t | X_0 = x_0] \sim N(\rho^{2t} x_0, 1 - \rho^{4t})$$

Comme le paramètre  $\rho$  est plus petit que 1 en valeur absolue,  $X_t$  converge en distribution vers la loi  $N(0, 1)$  :

$$X_t \xrightarrow{t \rightarrow +\infty} N(0, 1)$$

De même partant de  $X_0 = x_0$  le couple  $(X_t, Y_t)$  a pour distribution:



**Figure(2,4) :** *Uéchantillonnage de Gibbs. Itérations 100 à 1000..*

$$[(X_t, Y_t) \setminus X_0 = x_0] \sim N \left( \begin{pmatrix} \rho^{2t} x_0 \\ \rho^{2t+1} x_0 \end{pmatrix}, \begin{pmatrix} 1 - \rho^{4t} & \rho(1 - \rho^{4t}) \\ \rho(1 - \rho^{4t}) & (1 - \rho^{4t+2}) \end{pmatrix} \right)$$

$$\text{qui tend vers } N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right) \text{ quand } t \rightarrow +\infty$$
$$(X_t, Y_t) \xrightarrow{t \rightarrow +\infty} N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

## 2.9 Logiciel de simulation MCMC : WinBUGS

Les initials BUGS signifient Bayesian Inference Using Gibbs Sampling. Ce logiciel de programmation a été développé par Spiegelhalter, Thomas, Best et Gilks (Spiegelhalter, Thomas, Best et Gilks, 1996a) à l'unité MRC de Bio-statistique de Cambridge. Il vise à tirer avantage des probabilités de l'échantillonnage de Gibbs dans l'inférence bayésienne sur des systèmes complexes. BUGS repose sur un langage proche de C ou de S-Plus et comprend des déclarations sur le modèle, les données et les spécifications a priori, en autorisant plusieurs niveaux dans la modélisation a priori. La plupart des lois standards sont reconnues par BUGS, qui autorise de plus un grand nombre de transformations. BUGS comprend également des commandes comme `compile`, `data`, `update`, `stat`. La sortie de BUGS est une table des valeurs simulées pour l'ensemble des paramètres. Une restriction importante de ce logiciel est d'imposer l'utilisation de lois a priori conjuguées ou log-concaves pour permettre l'emploi de l'échantillonnage de Gibbs et de méthodes de simulation automatiques. Les autres lois peuvent être utilisées par une discrétisation de leur support. Un module de contrôle de convergence des algorithmes MCMC, nommé CODA, a été construit par Best, Cowles et Vines (Best, Cowles et Vines, 1995). Initialement prévu pour traiter les sorties de BUGS, ce logiciel, fonctionnant sous S-Plus, permet d'analyser les résultats d'algorithmes de Gibbs et de Metropolis-Hastings. Les techniques de contrôle sont principalement celles décrites dans Cowles et Carlin (Cowles et Carlin, 1996).

# Chapitre 3

## Modèles linéaire et logit dans les essais cliniques avec WinBUGS

### 3.1 Introduction

La régression est le domaine le plus grand des statistiques théoriques et appliquées. C'est la méthode la plus pratique de l'économétrie. Le modèle de cadre de régression est définie par une variable à prédire notée  $y$ , et une variable (régression simple) et multi variable (régression multiple) variables prédictives (ou indépendante) connue.

L'inférence bayésienne pour les modèles linéaires et les paramètres de régression logistique suit les trois étapes suivantes :

1. Former une distribution a priori de tous les paramètres inconnus.
2. Notez la fonction de vraisemblance des données.
3. Utilisez le théorème de Bayes pour trouver la distribution a posteriori de tous les paramètres.

Dans ce chapitre nous avons appliquées ces modèles dans les essais cliniques à l'aide de logiciel Win BUGS.

## **3.2 Essais cliniques**

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 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 précliniques). 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 structure de recherche agréées publique ou privées.

Pour une classe de loi a priori conjuguée nous donnons les distributions a posteriori, 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.

### **3.2.1 Définition d'un essai clinique**

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.).

### **3.2.2 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 ont 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 européennes de santé ( E M E A ) pour l'obtention de l'autorisation de commercialisation appelée AMM (Autorisation de Mise sur le Marché).

### **Phase VI**

Les essais de **phase VI** 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.

## **3.3 Modèles linéaire**

La régression est pour un type de problème où deux variables continues quantitatives  $X$  et  $Y$  ont un rôle asymétrique d'une variable  $Y$  dépend de la variable  $X$ .

La liaison entre la variable dépendante  $Y$  et la variable indépendante  $X$  peut être modélisée comme une fonction de  $Y = \alpha + \beta \times X + \varepsilon$

$Y$ : variable dépendante (expliquée)

$X$ : variable indépendante (explicative)

$\alpha$ : interception (valeur de  $Y$  pour  $x = 0$ )

$\beta$ : pente (variation moyenne de la valeur de  $Y$  pour une augmentation d'une unité de  $\alpha$  et  $\beta$ ).

$\varepsilon$ : L'erreur résiduel qui suit une loi normale de moyenne 0 et de variance  $\sigma^2$ .

### **3.3.1 Choix de l'a priori des paramètres**

Nous avons besoin d'une distribution a priori sur  $\alpha$ ,  $\beta$ , et  $\sigma$ . Nous allons préciser ces trois priors comme indépendants, le choix est :

$$\alpha \sim \text{uniforme} [ - \infty , + \infty ]$$

$$\beta \sim \text{uniforme} [ - \infty , + \infty ]$$

$$\log(\sigma) \sim \text{uniforme} [ - \infty , + \infty ]$$

Avec ce choix, nous pouvons rapprocher les résultats d'une analyse de régression fréquentiste .mais, on peut prendre autre choix est considéré comme le plus pratique pour le logiciel

WinBUGS :

$$\alpha \sim \text{Normal} ( 0, 0.0001 )$$

$$\beta \sim \text{Normal} ( 0, 0.0001 )$$

$$\sigma \sim \text{uniforme} [ 0, 100 ]$$

Ils sont très diffus (typiquement, dépend de l'échelle des données) , de sorte rapproche de la première série de lois a priori. Bien entendu, si l'information a priori réel est disponible, il peut être incorporé.

- La nécessité pour le log en première série de lois a priori vient du fait la variance doit être positive. L'a priori de  $\sigma$  est équivalent à une densité qui est proportionnelle à une  $\sigma^2$  .

- Nous spécifions une loi a priori non informative de ces trois paramètres. Bien sûr, nous pouvons également inclure des informations a priori lorsqu'ils sont disponibles.

- Première série de lois a priori est en fait " impropre " parce que leurs densités ne s'intègrent pas à un, depuis l'aire sous ces courbes infini ! En général, cela doit être évité car il peut parfois causer des problèmes avec les distributions a posteriori. Ce n'est pas un de ces cas de problème.

### 3.3.2 La fonction de vraisemblance

Comme c'est souvent le cas, la fonction de probabilité utilisée dans une analyse Bayésienne est la même que celui utilisé pour l'analyse fréquentiste.

1. Nous avons le résidus suit une loi normale :  $\varepsilon \sim N(0, \sigma^2)$ .

2. La moyenne de la ligne de régression, étant donné que nous savons  $\alpha$  et  $\beta$  est

$$y = \alpha + \beta \times x.$$

Mettre cela ensemble, nous avons  $y \sim (\alpha + \beta \times x, \sigma^2)$ .

Ainsi, pour un seul patient avec  $x_i$  valeur observée, on a  $y \sim (\alpha + \beta * x_i, \sigma^2)$  .

Ainsi, pour un même patient, la fonction de vraisemblance est:

$$f(y_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(y_i - (\alpha + \beta * x_i))^2}{\sigma^2} \right\} \quad (3-1)$$

Donc, pour un groupe de n patients chacun contribuant données  $(x_i, y_i)$ , la fonction de probabilité est donnée par

$$\prod_{i=1}^n f(y_i) = f(y_1) \times f(y_2) \times f(y_3) \cdot \times .. \times f(y_n)$$

### 3.3.3 La densité a posteriori

La théorème de Bayes dit maintenant de multiplier la fonction de vraisemblance avec l'a priori. Pour la première série a priori, l'a priori conjugué est tout simplement  $1 \times 1 \times \frac{1}{\sigma^2}$ . Pour le second, il s'agit de deux distributions normales multiplié par l'uniforme. Etant donné que l'uniforme est égal à 1 partout, ce qui a réduit à un produit de deux distributions normales.

- Pour la première série a priori, la distribution a posteriori est tout simplement :

$$\prod_{i=1}^n f(y_i) \times \frac{1}{\sigma^2} \quad (3-2)$$

Il s'agit d'un postérieur tridimensionnelle impliquant  $\alpha$ ,  $\beta$ , et  $\sigma^2$ . En intégrant cette densité a posteriori, nous pouvons obtenir les densités marginaux pour chacune des  $\alpha$ ,  $\beta$ , et  $\sigma^2$ .

$$\alpha \sim t_{n-2}$$

$$\beta \sim t_{n-2}$$

$$\sigma^2 \sim \text{Inverse chi deux}$$

## 3.4 Applications

### 3.4.1 Modèle linéaire simple

Le tableau 1 donne les données réelles d'une étude croisée comparant un nouveau laxatif contre un laxatif standard, bisacodyl. Les jours sont utilisés comme critère d'évaluation primaire. Le tableau montre que le nouveau médicament est plus efficace que le bisacodyl.

n° patient	Y-variables	X-variables
1	24	8
2	30	13
3	25	15
4	35	10
5	39	9
6	30	10
7	27	8
8	14	5
9	39	13
10	42	15
11	41	11
12	38	11
13	39	12
14	37	10
15	47	18
16	30	13
17	36	12
18	12	4
19	26	10
20	20	8
21	43	16
22	31	15
23	40	14
24	31	7
25	36	12
26	21	6
27	44	19
28	11	5
29	27	8
30	24	9
31	40	15
32	32	7
33	10	6
34	37	14
35	19	7

Modèle avec le logiciel WinBugs

Y-variables: nouveau treatment.

X-variables: bisacodyl

$Y \sim N(\mu_i, \tau)$

$\mu_i = \alpha + \beta \times X_i$

modèle

```
{
for(i in 1 : 35) {
y[i] ~dnorm(mu[i], tau)
mu[i] <- alpha + beta * X[i]
}
alpha ~dnorm(0, 1.0E-6)
beta ~dnorm(0, 1.0E-6)
```

```

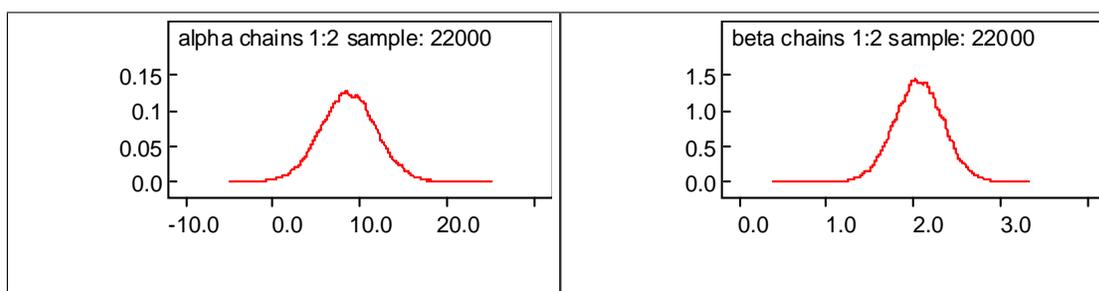
tau ~ dgamma(1.0E-3, 1.0E-3)
sigma <- 1/sqrt(tau)
}

```

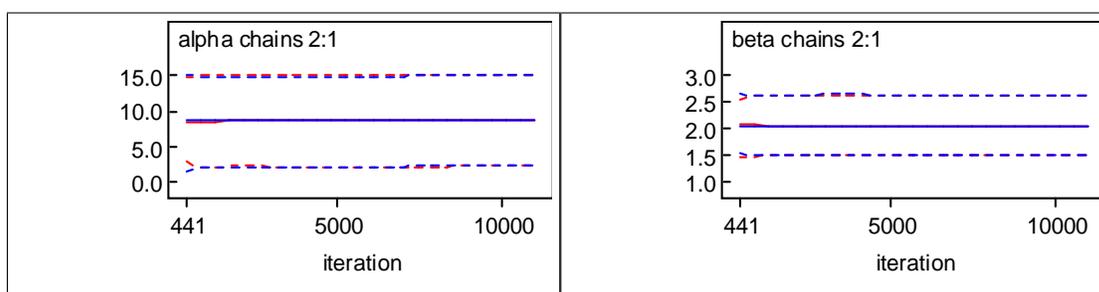
Nous procédons ensuite à estimer, cette fois sur deux chaînes, avec 110 000 itérations (1000) suffisamment chacun, en gardant une itération de 150. Les paramètres de la ligne est estimé,  $\alpha = 8,669$  avec un écart type de 3,236 et  $\beta = 2.062$  avec un écart type de 0,2854 (voir A.Labdaoui & H.merabet 2013). Les sorties de WinBUGS sont:

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
alpha	8.669	3.236	0.02289	2.308	8.658	15.03	1	22000
beta	2.062	0.2854	0.002044	1.505	2.061	2.624	1	22000
tau	0.02641	0.006571	4.7E-5	0.01533	0.02583	0.04088	1	22000

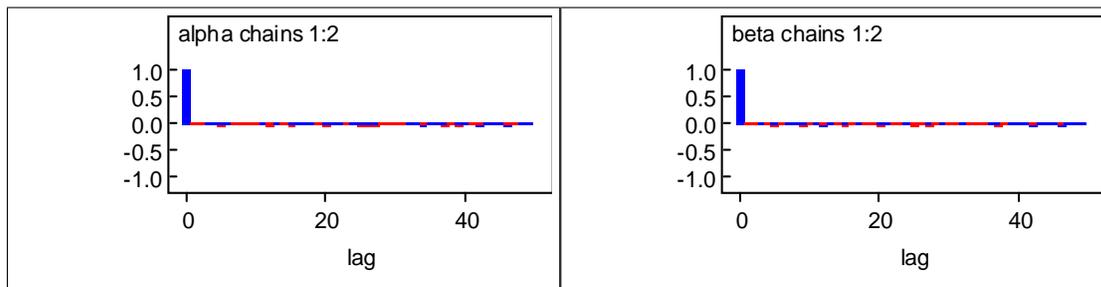
Nous présenter maintenant les graphiques des paramètres alpha et bêta, de la densité du noyau en figure.3.1, figure.3.2 quantiles et la fonction d'auto-corrélation sur la figure.3.3



Figure(3, 1) : la densité du noyau



Figure(3, 2) : *quantiles*



Figure(3, 3) : *fonction d'auto-corrélation*

### 3.4.2 Modèle de régression linéaire multiple

Nous pouvons être intéressés de savoir si l'âge est un facteur indépendant de l'effet de

le nouveau laxatif. Cet objectif d'une équation de régression simple doit être étendu comme suit  $Y = \alpha + \beta_1 \times X_1 + \beta_2 \times X_2$ , deux coefficients de régression partielle sont appelés. Tout comme une régression linéaire simple, régression linéaire multiple peut nous donner la meilleure solution pour les données indiquées, mais il est difficile de voir les corrélations dans une figure. Le tableau 2 donne les données du tableau 1 prolongée par la variable âge.

Patient no	Y-variables	X <sub>1</sub> -variables	X <sub>2</sub> -variables
1	24	8	25
2	30	13	30
3	25	15	25
4	35	10	31
5	39	9	36
6	30	10	33
7	27	8	22
8	14	5	18
9	39	13	14
10	42	15	30
11	41	11	36
12	38	11	30
13	39	12	27
14	37	10	38
15	47	18	40
16	30	13	31
17	36	12	25
18	12	4	24
19	26	10	27
20	20	8	20
21	43	16	35
22	31	15	29
23	40	14	32
24	31	7	30
25	36	12	40
26	21	6	31
27	44	19	41
28	11	5	26
29	27	8	24
30	24	9	30
31	40	15	20
32	32	7	31
33	10	6	29
34	37	14	43
35	19	7	30

Modèle avec le logiciel WinBugs

Y-variables: nouveau treatment.

X1-variables: bisacodyl.

X2-variables: age.

$Y \sim N(mu, tau)$

$mu = \alpha + \beta_1 \times X_1 + \beta_2 \times X_2$

modele

```

{
  for(i in 1 : 35) {
    y[i] ~dnorm(mu[i], tau)
    mu[i] <- alpha + beta1* X1[i] + beta2* X2[i]
  }
  alpha ~dnorm(0, 1.0E-6)

```

```

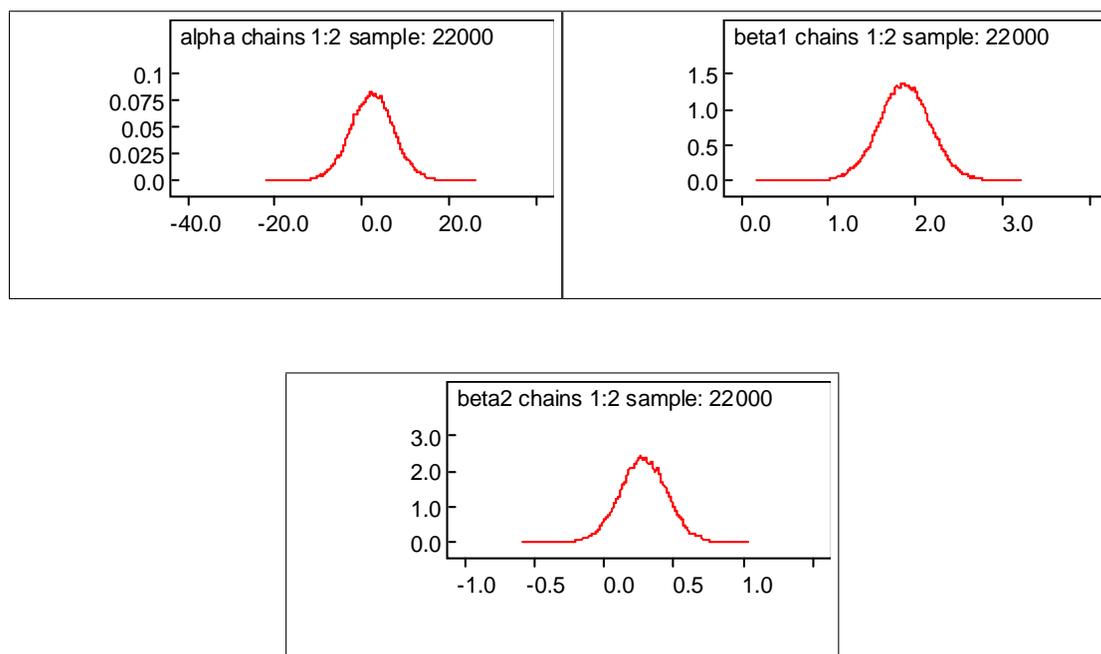
beta1 ~dnorm(0, 1.0E-6)
beta2 ~dnorm(0, 1.0E-6)
tau ~dgamma(1.0E-3, 1.0E-3)
sigma <- 1/sqrt(tau)
}

```

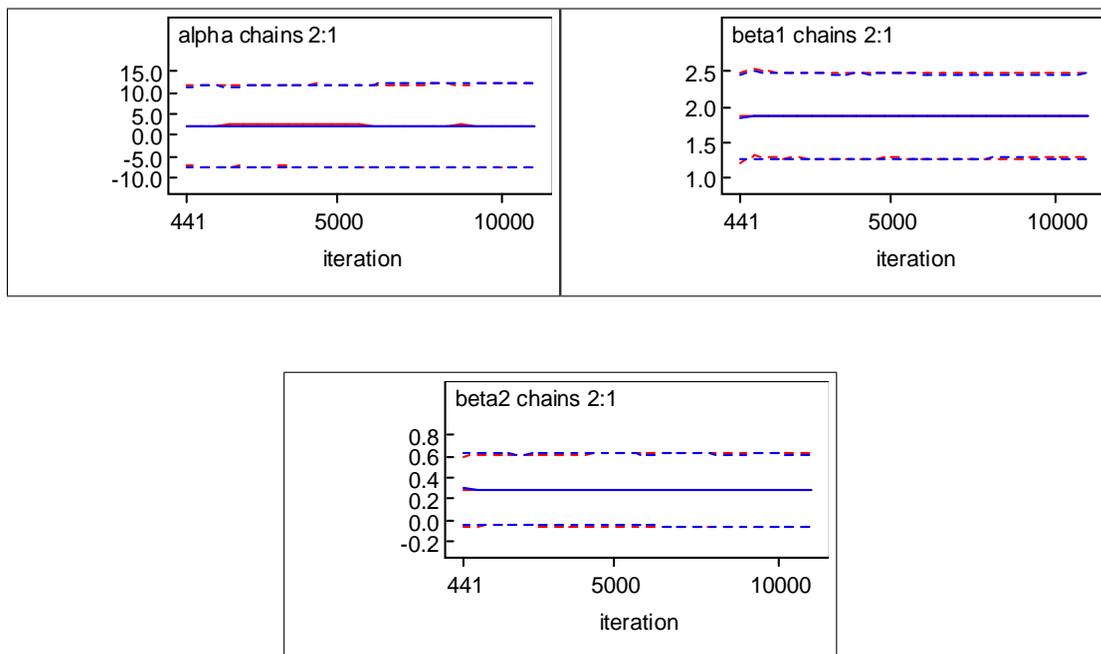
Nous procédons ensuite à estimer, cette fois sur deux canaux, avec 110 000 itérations (1000) suffisamment chacun, en gardant une itération de 150. Les paramètres de la ligne est estimé,  $\alpha = 2,332$  avec un écart type de 4,985 et  $\beta_1 = 1.876$  avec un écart type de 0,3003,  $\beta_2 = 0,282$  avec un écart type de 0,171. Les sorties de WinBUGS sont:

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
alpha	2.332	4.985	0.03595	-7.53	2.353	12.14	1	22000
beta1	1.876	0.3003	0.002056	1.282	1.875	2.472	1	22000
beta2	0.2827	0.1719	0.001191	-0.05832	0.2827	0.6232	1	22000
tau	0.02786	0.006994	5.042E-5	0.01588	0.02728	0.04326	1	22000

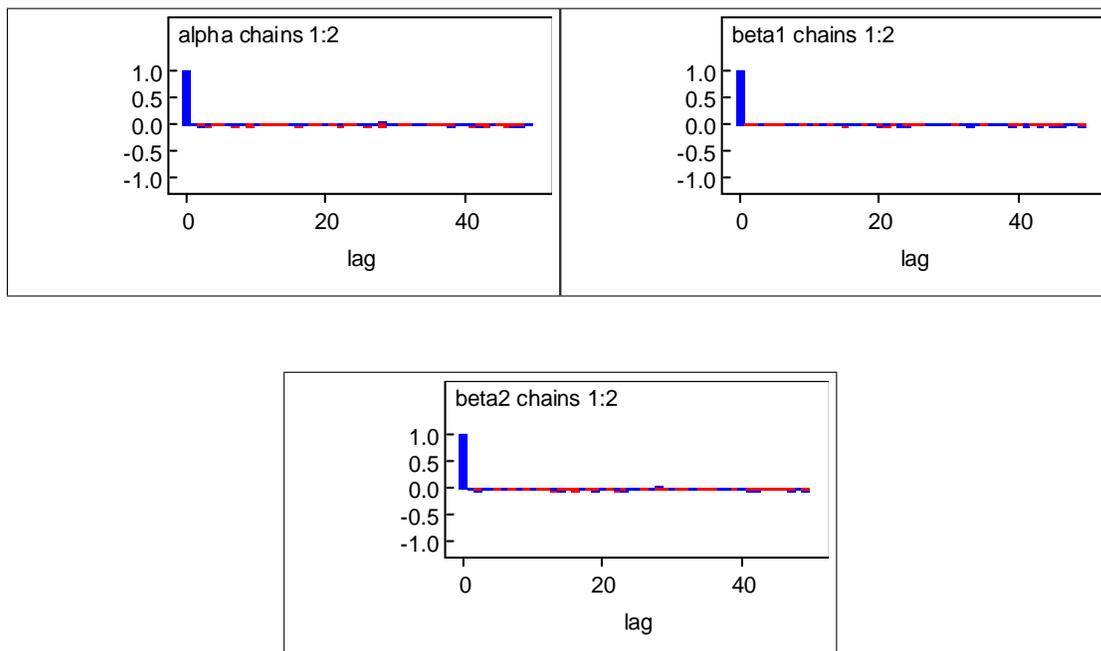
Nous présenter maintenant graphiquement les paramètres alpha et bêta (1), bêta (2) de noyau de densité dans la figure.3.4, quantiles, figure3.5 et le graphique de corrélation automatique en figure3.6:



Figure(3,4) : la densité Kernel



Figure(3, 5) : quantiles



Figure(3, 6) : représente fonction d'autocorrelation

### 3.4.3 Interprétation des résultats

\* Dans le modèle de régression linéaire la droite de régression est  $Y = 8.646 + 2.065 X$  .

La pente est 2.065 et dirigé l'original est 8,646, et si nous avons  $x = 1 \rightarrow y = 10$  donc le nouveau traitement est meilleur que le traitement standard.

\* La droite de régression dans le modèle de régression linéaire multiple est  $Y = 2.332 + 0.282 X_1 + 1.876 X_2$

Nous ajoutons le facteur l'âge ou non le nouveau traitement est la meilleure.

## 3.5 Le modèle logistique

Dans le cadre bayésien, il n'existe pas de différence fondamentale entre l'observation et le paramètre d'un modèle statistique, tous deux étant considérés comme quantités variables, Donc si on note  $x$  la donnée, de loi d'échantillonnage  $f(x|\theta)$ , et  $\theta$  le paramètre du modèle considérés (plus éventuellement, les variables latentes), de loi a priori une inférence formelle requiert la mise à jour de la distribution conditionnelle  $f(\theta|x)$  du paramètre.

### 3.5.1 Modèle logit

Un modèle standard de régression qualitative et le modèle de régression logistique ou modèle logit, où la loi de  $y$  conditionnelle aux variables explicatives  $z \in \mathbb{R}^p$  est :

$$p(y = 1) = 1 - p(y = 0) = \frac{\exp(z^t \gamma)}{1 + \exp(z^t \gamma)} \quad (3-3)$$

Considérons le cas particulier où  $z = (1, x)$  et  $\gamma = (\alpha, \beta)$  : les variables aléatoires  $y_i$  à valeurs dans  $\{0,1\}$  sont associées à des variables explicatives est sont modélisées suivant une loi de Bernoulli de probabilité conditionnelle

$$y_i | x_i \sim B \left( \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)} \right) \quad (3-4)$$

Supposons que nos paramètres suivent a priori une loi impropre  $(\alpha, \beta) = 1$ . La vraisemblance de notre modèle, pour un échantillon  $(y_1, x_1), \dots, (y_n, x_n)$ , est égale à

$$f(y_1, \dots, y_n | x_1, \dots, x_n, \alpha, \beta) = \prod_{i=1}^n \frac{\exp\{(\alpha + \beta x_i) y_i\}}{1 + \exp(\alpha + \beta x_i)} \quad (3-5)$$

La loi a posteriori de  $(\alpha, \beta)$ , se déduit alors par application formelle du Théorème de Bayes

$$\prod_{i=1}^n \frac{\exp(\alpha + \beta x_i) y_i}{1 + \exp(\alpha + \beta x_i)} = \frac{\exp\left\{\sum_{i=1}^n (\alpha + \beta x_i) y_i\right\}}{\prod_{i=1}^n 1 + \exp(\alpha + \beta x_i)} \quad (3-6)$$

### 3.5.2 Définition de la loi a posteriori

L'utilisation de lois a priori impropres, c'est-à-dire de mesure  $\sigma$  finie de masse infinie sur l'espace des paramètres, implique que la dérivation des lois a posteriori par la relation de proportionnalité

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

n'est pas nécessairement acceptable pour mettre en œuvre un algorithme de Metropolis-Hastings sur car la loi correspondante peut ne pas exciter, c'est-à-dire, n'est pas forcément intégrable. On est confronté à cette difficulté pour l'échantillonnage de Gibbs, par exemple, qui contrairement à l'algorithme de Metropolis-Hastings, fonctionne avec des lois conditionnelles extraites de , elle-même représentée par la relation de proportionnalité ci-dessus. Il peut arriver que ces lois soient clairement définies et simulables, mais qu'elles ne correspondent pas à une loi jointe c'est-à-dire que n'est pas intégrable.

### 3.5.3 Applications

Dans cette section nous allons appliquer la méthode MCMC pour le modèle logit et pour cela nous utilisons le logiciel WinBUGS. Notre étude est basée sur la comparaison entre une crème antiseptique et un Placebo, tel que le critère de jugement est la guérison d'une infection (Agrist 2002). Nous cherchons à estimer l'effet de la crème par rapport au placebo, la réponse de 8 centres que nous avons considérés sont présenté dans le tableau 3.3 :

Centre	traitement	Réponse	
		Succès	Echec
1	Crème	11	25
	Placebo	10	27
2	Crème	16	4
	Placebo	22	10
3	Crème	14	5
	Placebo	7	12
4	Crème	2	14
	Placebo	1	16
5	Crème	6	11
	Placebo	0	12
6	Crème	1	10
	Placebo	0	10
7	Crème	1	4
	Placebo	1	8
8	Crème	4	2
	Placebo	6	1

### Modèle 1

Certains centres peuvent avoir une plus grande probabilité de guérison, quel que soit le traitement employé.(voir A.Labdaoui & H.Merabet 2012).

Soient  $r_i^t$  le nombre de succès dans le groupe T (Placebo ou Crème) dans le centre  $i$ , on peut écrire :

$$r_i^T \rightarrow \text{Binomiale}(p_i^T, n_i^T)$$

$$\text{logit}(p_i^p) = \alpha - \beta/2 + u_i$$

$$\text{logit}(p_i^c) = \alpha + \beta/2 + u_i$$

$$u_i \rightarrow \text{Normal}(0, \sigma_u^2)$$

Dans ce modèle, l'OR de la crème versus Placebo est constant et égal à  $e^\beta$ , et l'hétérogénéité (dans la probabilité de succès) entre les centres est mesuré par  $\sigma_u^2$ .

Le modèle de WinBUGS va s'écrire :

```

Modèle
{
for(i in 1 : 8) {
rp[i] ~ dbin(pp[i], np[i]) # Vraisemblance pour Placebo.
rc[i] ~ dbin(pc[i], nc[i]) # Vraisemblance pour Crème.
logit(pp[i]) <- alpha - beta / 2 + u[i] # Modèle Placebo.
logit(pc[i]) <- alpha + beta / 2 + u[i] # Modèle crème.
u[i] ~ dnorm(0.0, tau) # Effet aléatoire.
}
alpha ~ dnorm(0.0, 1.0E-6) # A priori vague pour alpha.
beta ~ dnorm(0.0, 1.0E-6) # A priori vague pour beta.
tau ~ dgamma(0.1, 0.1) # A priori vague pour la précision.
sigma <- 1/ sqrt(tau)
OR <- exp(beta)
}

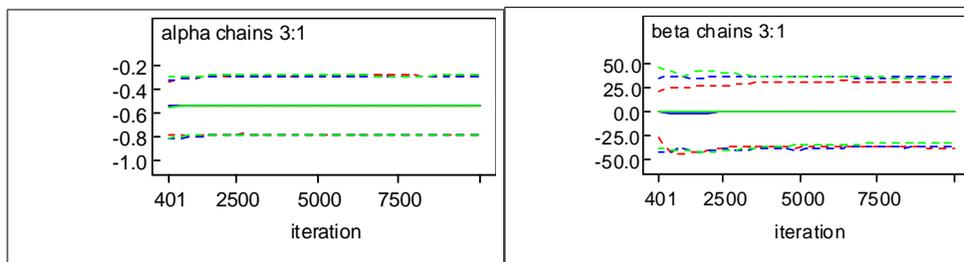
```

Nous procédons ensuite à l'estimation, cette fois sur 3 chaînes, avec 100 000 itérations (1 000 suffisaient) chacune, en conservant une itération sur 150. L'effet (suppose homogène) de la crème est estimé à 0,757, avec un écart-type de 0,304.

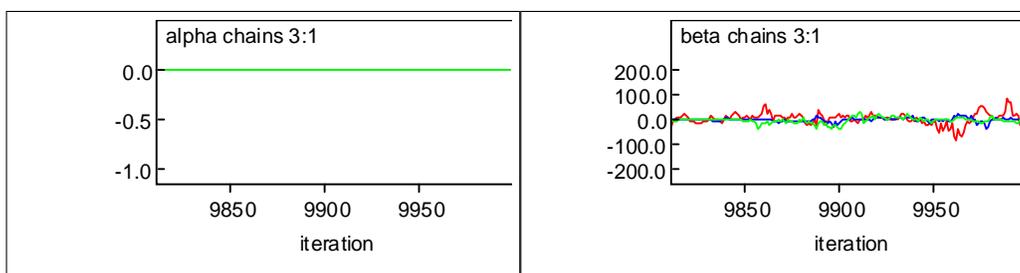
Les sorties de WinBUGS sont les suivantes :

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
OR	2.235	0.7052	0.009858	1.191	2.123	3.88	1	3000
alpha	-0.8362	0.757	0.05594	-2.37	-0.859	0.8177	1	3000
beta	0.7575	0.3043	0.004185	0.1745	0.7528	1.356	1	3000
tau	0.4806	0.7091	0.01471	0.0881	0.3918	1.323	1	3000
u[1]	-0.1132	0.7887	0.05607	-1.787	-0.09136	1.514	1	3000
u[2]	1.893	0.8025	0.05561	0.2114	1.89	3.542	1	3000
u[3]	1.01	0.8064	0.05598	-0.7026	1.008	2.671	1	3000
u[4]	-1.427	0.9	0.05286	-3.341	-1.382	0.2316	1	3000
u[5]	-0.6119	0.8564	0.05491	-2.436	-0.5866	1.111	1	3000
u[6]	-1.871	1.044	0.04911	-4.184	-1.764	-0.05686	1	3000
u[7]	-0.849	0.9642	0.05094	-2.914	-0.7977	0.9901	1	3000
u[8]	1.856	0.9282	0.05278	0.1017	1.806	3.815	1	3000

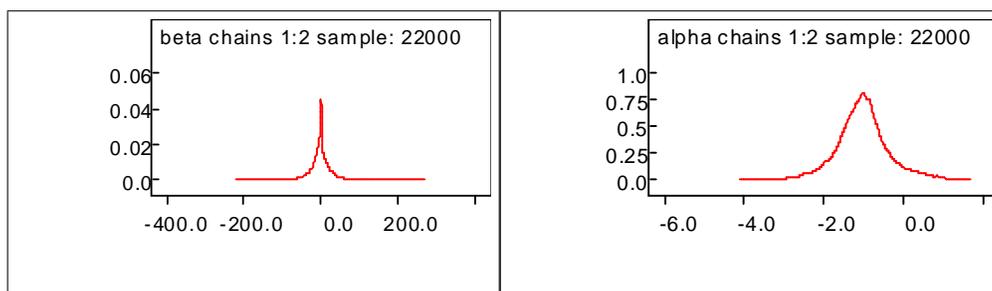
Nous présentons maintenant la représentation graphique pour les paramètres alpha et beta:



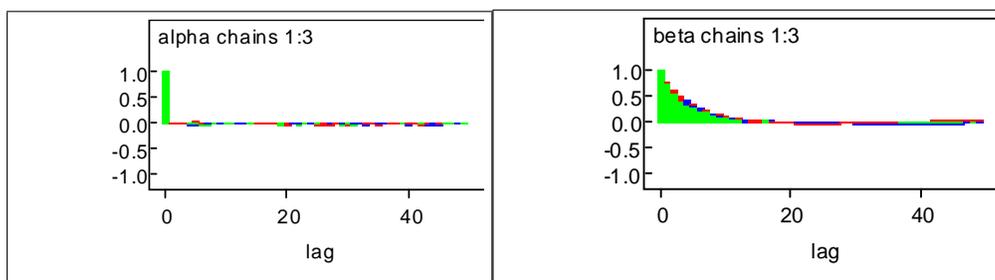
Figure(3,7) : les quantiles



Figure(3,8) : les traces dynamiques



Figure(3,9) : la densité de Kernel



Figure(3, 10) : fonction d'auto-corrélation

## Modèle2

Certains centres peuvent avoir une plus grande probabilité de guérison, quel que soit le traitement employé. De plus, l'effet du traitement peut être hétérogène entre les centres. On complexifie le modèle précédent pour ne pas avoir à faire l'hypothèse d'homogénéité de l'effet de la crème en injectant un effet aléatoire sur  $\beta$  on obtient alors:

$$r_i^T \rightarrow \text{Binomiale}(p_i^T, n_i^T)$$

$$\text{logit}(p_i^p) = \alpha - (\beta + b_i)/2 + u_i$$

$$\text{logit}(p_i^c) = \alpha + (\beta + b_i)/2 + u_i$$

$$u_i \rightarrow \text{Normal}(0, \sigma_u^2)$$

$$b_i \rightarrow \text{Normal}(0, \sigma_b^2)$$

Dans ce modèle, l'OR de la crème versus Placebo est constant et égal à  $e^\beta$ , et l'hétérogénéité (dans la probabilité de succès) entre les centres est mesuré par  $\sigma_u^2$ .

Le modèle de WinBUGS va s'écrire :

modèle

{

for(i in 1 : 8) {

rp[i] ~ dbin(pp[i], np[i]) # Vraisemblance pour Placebo.

rc[i] ~ dbin(pc[i], nc[i]) # Vraisemblance pour Crème.

logit(pp[i]) <- alpha - (beta+b[i]) / 2 + u[i] # Modèle Placebo.

logit(pc[i]) <- alpha + (beta+b[i]) / 2 + u[i] # Modèle crème.

```

u[i] ~dnorm(0.0, tau) # Effet aléatoire sur constante.
b[i] ~dnorm(0.0, taub) # Effet aléatoire sur coefficient.
}
alpha ~dnorm(0.0, 1.0E-6) # A priori vague pour alpha.
beta ~dnorm(0.0, 1.0E-6) # A priori vague pour beta.
tau ~dgamma(0.001, 0.001) # A priori vague pour la précision.
sigma <- 1/ sqrt(tau)
taub ~dgamma(0.001, 0.001)
sigmab <- 1/ sqrt(taub)
OR <- exp(beta)
}

```

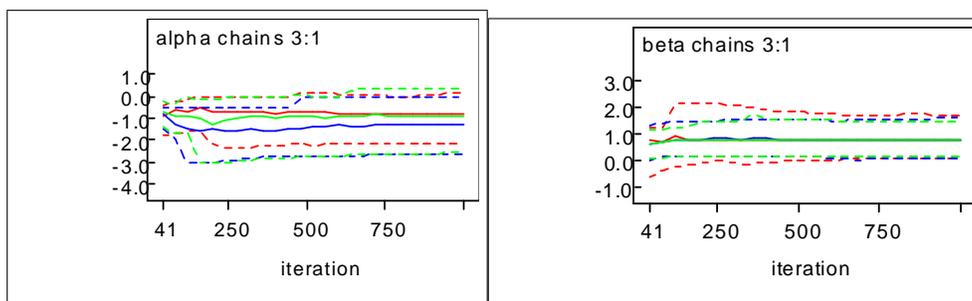
Nous procédons ensuite à l'estimation, avec les mêmes paramètres que la précédente.

L'effet (suppose homogène) de la crème est estime à 0,809, avec un écart-type de 0,388.

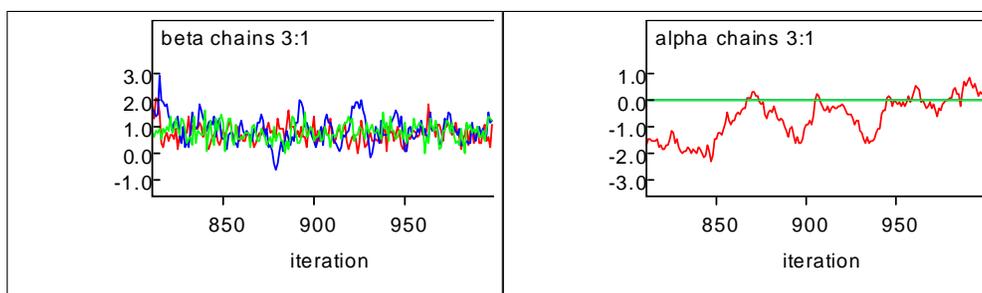
Les sorties de WinBUGS sont les suivantes :

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
OR	2.439	1.194	0.04907	1.114	2.205	5.073	1	3000
alpha	-1.019	0.6656	0.05523	-2.412	-0.9845	0.206	1	3000
b[1]	-0.1855	0.4092	0.02053	-1.343	-0.06496	0.3428	1	3000
b[2]	-0.06375	0.3686	0.01251	-0.9226	-0.01724	0.6018	1	3000
b[3]	0.1396	0.3834	0.01526	-0.4402	0.04621	1.134	1	3000
b[4]	-0.01424	0.3951	0.008387	-0.903	-3.466E-40	0.8319	1	3000
b[5]	0.2099	0.5129	0.02595	-0.3895	0.05254	1.659	1	3000
b[6]	0.04355	0.4719	0.01307	-0.8845	0.008931	1.142	1	3000
b[7]	-0.003306	0.8799	0.4091	0.009099	-0.9549	1.391E-4	1	3000
b[8]	-0.1955	0.5276	0.02515	-1.701	-0.04612	0.4741	1	3000
beta	0.8093	0.3881	0.01441	0.1083	0.7905	1.624	1	3000
u[1]	0.08317	0.7056	0.05601	-1.237	0.05257	1.577	1	3000
u[2]	2.073	0.7483	0.0573	0.6281	2.052	3.593	1	3000
u[3]	1.187	0.7324	0.055	-0.1551	1.16	2.706	1	3000
u[4]	-1.29	0.8366	0.05211	-2.918	-1.279	0.2839	1	3000
u[5]	-0.5029	0.7608	0.05188	-1.987	-0.5165	1.03	1	3000
u[6]	-1.75	0.9604	0.04458	-3.814	-1.691	0.0156	1	3000
u[7]	-0.7091	0.8792	0.04721	-2.478	-0.7052	1.019	1	3000
u[8]	2.026	0.8818	0.05519	0.3233	2.009	3.814	1	3000

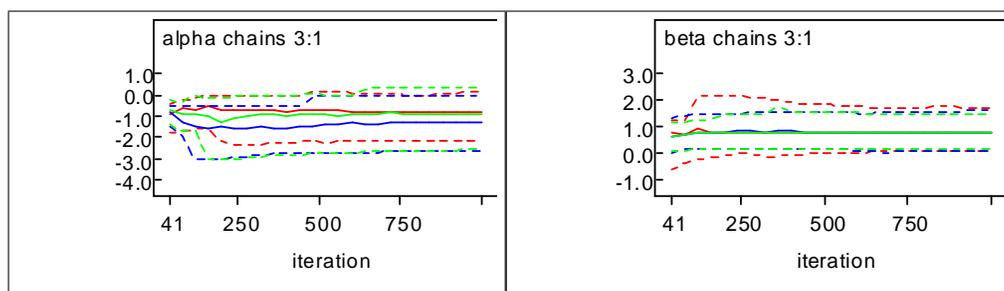
Nous présentent maintenant la représentation graphique pour les paramètres alpha et beta:



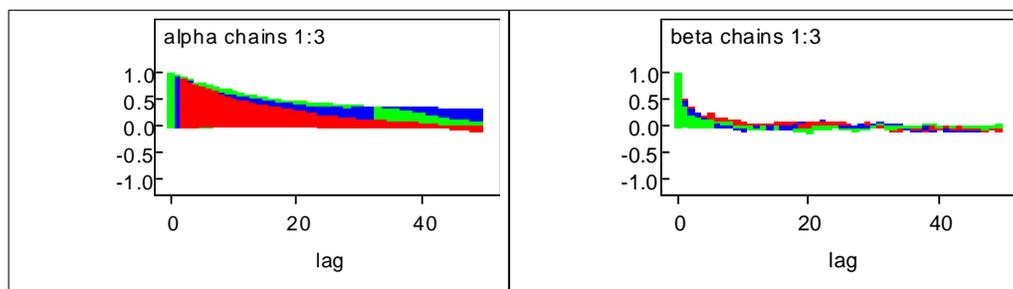
Figure(3, 11) : les quantiles



Figure(3, 12) : les traces dynamiques



Figure(3, 13) : la densité de Kernel



Figure(3, 14) : fonction d'auto-corrélation

### Interprétation des résultats

Nous remarquons après la comparaison entre les deux modèles que le paramètre beta et son écart type sont presque égaux, c'est-à-dire l'ajout de l'effet aléatoire sur le coefficient beta il change rien, ensuite l'OR dans les deux modèles (modèle(1) : OR=2.235, modèle(2) : OR=2.439) est supérieur à 1 et l'intervalle de crédibilité compris entre 1.191 et 3.88 pour le modèle (1) et 1.114 et 5.07 pour le modèle (2) au niveau de 97.5% pour ces résultats on dit que notre crème antiseptique est efficace dans les deux modèles(1) et( 2).

## 3.6 Conclusion

Un des mérites de notre travail est d'avoir montré à l'aide de données expérimentales d'essais cliniques qu'un peut modéliser de façon naturelle et en tirer les inférences adéquates, à savoir estimer les paramètres dans les modèles linéaires et logit avec et sans effet aléatoire, à l'aide des méthodes de Monte Carlo par chaîne de Markov (MCMC) d'autant plus que les performances des ordinateurs, ont rendu faisables des procédés de simulations efficaces et la disponibilité des programmes informatiques a facilité le calcul des probabilités a posteriori, qui étaient jusque là d'une complexité décourageante.

# Chapitre 4

## Les procédures bayésiennes prédictives dans les designs des essais expérimentaux avec des données binaires

### Rappel sur la notion de p-valeur

L'objectif principal de l'indice p-valeur est de mesurer la cohérence des données avec une forte hypothèse nulle. La définition de p-valeur est la probabilité, sous l'hypothèse nulle, que la statistique de test prendrait une valeur au moins aussi extrême que celui de fait observé.

**Définition 4.0.1** Soit  $H_0$  l'hypothèse nulle,  $T$  la statistique de test et  $F_0$  sa fonction de répartition sous l'hypothèse  $H_0$ . On suppose que  $F_0$  est continue.

1. Pour un test bilatéral (rejet des valeurs trop écartées) la p-valeur d'une valeur  $t$  prise par  $T$  est :

$$p(t) = \begin{cases} 2F_0(t) & \text{si } F_0(t) < 0.5, \\ 2(1-F_0(t)) & \text{si } F_0(t) \geq 0.5. \end{cases}$$

2. Pour un test unilatéral à droite (rejet des valeurs trop grandes) la p-valeur d'une valeur  $t$  prise par  $T$  est :

$$p(t) = 1 - F_0(t).$$

3. Pour un test unilatéral à gauche (rejet des valeurs trop petites) la p-valeur d'une valeur  $t$  prise par  $T$  est :

$$p(t) = F_0(t).$$

## 4.1 Introduction

Une des grandes forces du paradigme bayésien est la facilité avec laquelle on peut faire des prédictions sur les observations futures. L'idée prédictive est centrale dans les recherches expérimentales. En outre les probabilités prédictives bayésienne sont des outils efficaces pour le design et le suivi des expériences. Les procédures bayésienne prédictive donnent la recherche appliquée d'une méthode très intéressante pour évaluer les chances qu'une expérience finira pour montrant un résultat probant, ou au contraire un résultat non concluant. Les études expérimentales comme son nom l'indique, ce sont des études dans lesquelles les participants subissent une sorte d'intervention afin d'évaluer son effet. Une intervention pourrait inclure une intervention médicale ou chirurgicale, un nouveau médicament ou une intervention visant à modifier le mode de vie. Parce qu'ils sont la conception la plus méthodologiquement rigoureuse, les expériences sont le choix par défaut pour fournir des preuves pour les meilleures pratiques dans la gestion des patients, de sorte que cette discussion va commencer avec eux.

Dans sa forme la plus simple, une étude expérimentale pour tester l'effet d'un traitement exige que le chercheur déclare formellement l'hypothèse à tester où il doit sélectionner les personnes éligibles pour le traitement, dans lequel l'échantillon est divisé en deux groupes et chaque groupe (le groupe expérimental ou l'intervention) est donnée l'intervention tandis que l'autre (le groupe de contrôle) n'est pas, et les résultats d'intérêt sont comptabilisés sur la durée, et les résultats comparés entre les deux groupes.

Dans ce chapitre, nous considérons les procédures prédictive bayésienne dans le design expérimentale pour cela, nous définissons une hypothèse d'indices de satisfaction et l'anticipation de la satisfaction liée à un test comme une fonction décroissante de la p-valeur, la satisfaction est plus élevée que l'hypothèse nulle est rejetée plus large, ce est-à-dire, la valeur  $p$  est petite. Nous considérons le cas d'une procédure en deux étapes, qui se fait souvent dans

le cas des essais expérimentaux où ces indices de satisfaction sont intéressants des protocoles et quand l'inférence concerne à un effet évalué à partir de l'échantillon avenir. Nous avons traité nos applications à la planification expérimentale et design séquentiels avec des résultats binaires. Les calculs et les résultats de simulation concernent un problème d'inférence sont donnés par le logiciel: Matlab et R.

## 4.2 La méthodologie statistique

### 4.2.1 Le design expérimental

#### La prédiction bayésienne de design

Nous utilisons le cadre bayésien comme un outil pour le design des essais cliniques avec des propriétés fréquentiste souhaitables. Nous tirons un design efficace et flexible. La méthodologie statistique a déjà été utilisée. Rappelons que, le modèle bayésien a été introduit dans le cadre de (B. Lecoutre, G. Derzko, J. M. Grouin (1995) ) et après améliorée dans (Merabet, H. and Raoult J. P (1995), H .Merabet (2004 A), H. Merabet (2004 B)).

Nous spécifions le contexte expérimental en choisissant  $(P_\theta)_{\theta \in \Theta}$  une famille d'observations de probabilité sur un  $\Omega$  de l'espace et où  $\Theta$  est l'espace des paramètres inconnus et de tester l'hypothèse nulle  $\Theta_0$  contre l'hypothèse alternative  $\Theta_1$ . Dans le design asymétrique classique problème de test, une telle situation est généralement dans l'expérimentateur, un désir de mettre en évidence un résultat significatif, ce est-à-dire, pour conclure au rejet de l'hypothèse nulle

- étape 1: Si  $P_\theta < \theta_L$ , arrêter l'essais et de rejeter l'hypothèse alternative;
- étape 2: Si  $P_\theta < \theta_U$ , arrêter l'essais et de rejeter l'hypothèse nulle;
- étape 3: Sinon passez à l'étape suivante jusqu'à atteindre le maximum de nombre de patients.

Typiquement, nous choisissons  $\theta_L$  comme un petit nombre positif et  $\theta_U$  comme un grand nombre positif, à la fois entre 0 et 1 (inclus).  $P_\theta < \theta_L$  indique qu'il est peu probable que le taux de réponse sera plus grand que  $\theta_0$ . à la fin de l'essai compte tenu de l'information actuelle. Lorsque cela se produit, nous pouvons aussi bien arrêter l'essai et de rejeter l'hypothèse alternative à ce point. D'autre part, lorsque,  $P_\theta < \theta_L$  les données actuelles suggèrent que, si la même tendance se poursuit, nous aurons une forte probabilité de conclure que le traitement

est efficace à la fin de l'étude. Ce résultat, puis, fournit des preuves pour arrêter l'essai prématurément en raison d'efficacité. En choisissant  $\theta_L > 0$  et  $\theta_U < 1$ , le procès peut mettre fin prématurément en raison soit de la futilité ou efficacité. Pour les essais de phase IIA, nous préférons souvent choisir  $\theta_L > 0$  et  $\theta_U = 1$  pour permettre l'arrêt précoce de raison de futilité, mais pas en raison de l'efficacité.

### **Indice de satisfaction**

S'il est adopté à un test déterministe de procédure, par rapport à un niveau  $\alpha$ , conduisant à partitionner  $\Omega$  dans une région de non rejet  $\Omega_0^{(\alpha)}$  et une région de rejet  $\Omega_1^{(\alpha)}$ , un indice particulièrement simple de satisfaction est la fonction de l'indicateur  $\Omega_0^{(\alpha)}$ . Il est satisfait si le résultat est significatif à  $\alpha$ , si insatisfait. Mais très souvent, les utilisateurs veulent plutôt faire face à un résultat qui semble susceptible de conduire à leur rejet de  $\Theta_0$ , savent ce que son degré de signification; c'est-à-dire, savent quelle mesure les résultats apparaissent significative.

En utilisant le fait que toute technique de test raisonnable conduit à une famille régions de non rejet  $\Omega_0^{(\alpha)}$  dans le sens de l'inclusion diminuant lorsque  $\alpha$  augmente, c'est-à-dire, lorsque nos précautions d'atténuer, puis utilisez un nouvel indice de satisfaction, un peu moins rude que la précédente, notée  $\Phi^{(\alpha)}$ , et défini en fonction de la *p*-valeur, le plus simple à zéro sur la région de non rejet  $\Omega_0^{(\alpha)}$  et égale à  $(1 - p)$ , ou plus généralement  $(1 - p)^l$  avec  $l > 0$  autrement. En d'autres termes, elle offre un indice de satisfaction qui est égale à zéro si elle ne est pas détecté par ailleurs significative et une fonction décroissante de la valeur *p*, et par conséquent, plus *p* est faible l'expérimentateur estime que le résultat est significatif (G. Koop, D.J. Poirier, and J. L. Tobias (2007)) rappelons que la *p*-valeur est considérée comme une mesure de crédibilité attacher à l'hypothèse nulle que les praticiens utilisent souvent pour répondre à plusieurs inconvénients critiques de l'approche de Neymann-Pearson. La valeur de cet indice de satisfaction est une famille élargie d'indices dans le concept de prédire la satisfaction d'un avenir échantillon d'un premier échantillon.

### **La prediction de satisfaction**

Les contextes expérimentaux que nous avons mentionnés dans l'introduction se prêtent souvent à une analyse en plusieurs phases, et nous nous limitons à deux phases et la situation, ce qui correspond aux exigences des essais expérimentaux, où la première phase est celle indicatif

est destiné uniquement d'examiner si de reprendre les essais pour une deuxième phase, menée indépendamment de la première et dont seul le résultat de test basé sur la conclusion de ce qui est le but ultime de l'étude. Nous notons ici  $\Omega, \Omega'$  et  $\Omega''$  ensembles résultats complets, les résultats de la première phase et les résultats de la deuxième phase ( $\Omega = \Omega' + \Omega''$ ). C'est dans ce contexte que s'est proposé d'introduire un modèle bayésien avec une distribution a priori sur  $\Theta$  et de la famille des probabilités  $(P_\theta)_{\theta \in \Theta}$  sur  $\Omega$ . Il voit dans ce modèle la probabilité de  $\Omega''$ , influencé par le résultat de la première phase  $\omega'$ , que nous désignons par  $P_{\Omega'}^{\omega'}$  rappelons que, selon la terminologie bayésienne habituelle, le terme de probabilité prédictive, la probabilité  $P_\Omega$  sur l'espace des résultats complets, qui est utilisé ici est la probabilité sur  $\Omega''$  qui est déduite par conditionnement par  $\omega'$ . Nous trouvons comme une prédiction sur la vue de la première phase d'un résultat significatif dans la deuxième phase, la valeur  $P_{\Omega'}^{\omega'}(\Omega_1''^{(\alpha)})$ , où  $\Omega_1''^{(\alpha)}$  est la région de rejet du test classique faite sur la base des résultats de la deuxième phase. C'est dans ce sens que voici pratique, les deux statistiques classiques et statistique bayésienne. Nous proposons ici, associer plus généralement avec un indice de satisfaction sur le deuxième la prévision de l'indice de phase est l'espérance mathématique par rapport à  $P_{\Omega'}^{\omega'}$ , la satisfaction fournies par examiner la deuxième phase de l'expérience et de la prédite en utilisant le premier.

Considérons un essai clinique de phase II à un seul bras dans laquelle tous les patients sont affectés à un traitement expérimental  $E$ . Le niveau de dose de  $E$  est déterminé par un essai de phase I précédente et, par conséquent, est considéré comme fixé. Les patients sont recrutés, traités et évalués séquentiellement par groupe, mais la taille d'un groupe peuvent varier en raison de la taille du groupe après le premier regard provisoire est déterminé sur la base des observations les plus au cours de l'essai. Le nombre total de regards intermédiaires peut également varier d'un essai à l'autre, selon les observations et la fonction de perte, mais nous supposons une limite supérieure sur le nombre de patients qui pourraient être courus à l'essai de phase II. L'espace de décision pour l'ensemble de l'essai clinique de phase II est définie comme  $D = (d_1, d_2, d_3)$ .

A chaque évaluation intermédiaire, trois décisions possibles pourraient être apportées:

- $d_1$ : Arrêter l'essai de phase II et de recommander aucune autre étude de  $E$ .
- $d_2$ : Arrêter l'essai de phase II et de recommander  $E$  pour un essai de phase III ultérieure.
- $d_3$ : Continuer à recruter des patients pour l'essai de phase II. Il est démontré que la valeur élémentaire  $\omega'$  un indice de prédiction peut également être obtenu que l'attente

à l'égard de la distribution a posteriori sur la base  $\omega'$ , la valeur moyenne de l'indice de satisfaction liée à la loi échantillonnage de la deuxième phase. Le problème qui se pose est celui du calcul de cette espérance dans des situations des tests, pour un choix de la distribution préalable. Plusieurs modèles sont considérés pour illustrer la procédure bayésienne prédictive proposée.

### **L'inférence statistique pour le design**

Nous définissons les indices de satisfaction et l'anticipation de la satisfaction liée à un test d'hypothèse décroissante en fonction de la p-valeur, la satisfaction est plus élevée que l'hypothèse nulle est rejetée de façon plus générale, c'est-à-dire que p-valeur est petite. Nous considérons le cas d'une procédure en deux étapes, qui se fait souvent dans le cas des essais cliniques où les protocoles de ces indices de satisfaction sont intéressants et quand l'inférence concerne un effet évalué à partir de l'échantillon avenir seulement.

étant fixé  $\alpha$ , un test de niveau  $\alpha$  défini par la première région de l'indice de satisfaction critique  $\Omega_1^{(\alpha)}$ , étudié dans (Merabet, H, 2004) est défini par:

$$\phi(\omega'') = 1_{\Omega_1^{(\alpha)}}(\omega''); \quad (5-1)$$

Le défaut de l'indice fruste ci-dessus est qu'il exprime une satisfaction en "tout ou rien". Il est plus intéressant de prendre en compte jusqu'à quel niveau le résultat apparaîtra toujours significatif. On utilise donc un nouvel indice de satisfaction défini par :

$$\begin{aligned} \phi(\omega'') &= 0 \quad \text{si } p(\omega'') \geq 1 - \alpha \\ &= L(p(\omega'')) \quad \text{else} \end{aligned}$$

où L est une fonction décroissante. On peut généraliser cette procédure à une famille d'indices limités défini par:

$$L(p) = (1 - p)^l \quad \text{where } l \geq 0.$$

Il est préférable de choisir des indices limités en raison de leur faciliter de l'interprétation. Dans le cas où ( $l = 1$ ),  $1 - \Phi(\omega'')$  est la p-valeur et dans le cas où  $l = 0$ , on retrouve la fonction d'indicateur de la zone critique. Par la suite, on choisit  $l = 1$ ,  $L(p) = (1 - p)$  donc

$$\begin{aligned} \phi(\omega'') &= 0 \quad \text{si } p(\omega'') \geq 1 - \alpha \\ &= (1 - p) \quad \text{else} \end{aligned}$$

et évidemment la prédiction est donnée par :

$$\begin{aligned} \pi(\omega') &= \int_{\Omega_1^{\omega'(\alpha)}} \phi(\omega'') P_{\Omega''}^{\omega'}(d\omega'') & (5-2) \\ &= \int_{\Theta} \left( \int_{\Omega_1^{\omega'(\alpha)}} \phi(\omega'') P_{\Omega''}^{\theta}(d\omega'') \right) P_{\Theta}^{\omega'}(d\theta) \end{aligned}$$

On remarque que  $\int_{\Omega_1^{\omega'(\alpha)}} \phi(\omega'') P_{\Omega''}^{\theta}(d\omega'')$  généralise la puissance du test dans la logique de l'indice de satisfaction proposé.

### Application

On se propose de calculer la prédiction de satisfaction dans plusieurs modèles lorsque la loi du paramètre inconnu  $\theta$  est une loi a priori conjuguée ou bien non informative et nous envisageons une approche prédictive qui s'applique tout autant à des conclusions issues de procédures fréquentiste que bayésiennes. L'aspect séquentiel du traitement qui va être adopté est un élément particulièrement innovateur par rapport à la technologie existante, il permet d'alléger des études multi phases plus ambitieuses que l'existantes, de plus autorisée l'analyse statistique, un patient à la fois, est éthique puisque cela permet un arrêt de l'expérience plus bref et moins tardif.

Plusieurs designs de phase II d'essais cliniques sont proposées dans la littérature sont effectués en deux étapes. Soit  $X_i$  la variable de réponse dichotomique, qui prend la valeur 1 si les cliniciens classent le patient  $i$  comme réponse au traitement  $t$  et 0 sinon. Dans un design typique à deux étages, les patients  $T$  sont comptabilisés et traités au première étape et les patients supplémentaires  $T'$  dans la deuxième étape. On note  $\theta$  la probabilité qu'une personne souffrant d'une maladie est traitée avec le traitement  $t$ , ne peut être commercialisé que si  $\theta \geq \theta_0$ . D'un point de vue statistique on peut formuler le problème à l'aide du test suivant:

$$H_0 : \theta \leq \theta_0$$

On se place dans le cadre du modèle d'échantillonnage où on suppose que les  $x_i$  sont des réalisations de variables aléatoires  $X_i$  indépendantes et de même loi de Bernoulli de paramètre  $\theta$ , toujours pour des raisons d'exhaustivité on prend  $\omega' = \sum_{i=1}^T x_i$ . Si on choisit comme loi a priori pour  $\theta$  une loi beta  $B(\underline{\alpha}, \underline{\delta})$ , on sait alors que la loi a posteriori de  $\theta/\omega'$  est encore une

loi beta  $B(a, \bar{b})$  avec  $\bar{\alpha} = a + \omega'$  et  $\bar{\delta} = b + T - \omega'$ . L'indice de satisfaction pour une observation  $\omega'' = 0$  ou 1 est donc:

$$\begin{aligned}\phi(\omega'') &= 0 \quad \text{si } \omega'' < q_0 \\ &= \sum_{s=0}^{\omega''-1} \theta_0^s (1-\theta_0)^{1-s} \quad \text{si } \omega'' \geq q_0,\end{aligned}$$

avec

$$q_0 = \inf \left\{ u, \sum_{s=1}^u \theta_0^s (1-\theta_0)^{1-s} \leq \alpha \right\}$$

D'autre part la densité prédictive de  $\omega'' | \omega'$  est une la beta-binomiale:

$$\begin{aligned}f(\omega'' | \omega') &= \int_0^1 f(\omega'' / \theta) f(\theta / \omega') d\theta & (5-3) \\ &= \int_0^1 \theta^{\omega''} (1-\theta)^{(1-\omega'')} [B(\bar{\alpha}, \bar{\delta})]^{-1} \theta^{\bar{\alpha}-1} (1-\theta)^{\bar{\delta}-1} d\theta \\ &= \left[ \frac{\Gamma(\bar{\alpha} + \omega'') \Gamma(\bar{\delta} + 1 - \omega'')}{\Gamma(\bar{\alpha} + \bar{\delta} + 1)} \right] \frac{\Gamma(\bar{\alpha} + \bar{\delta})}{\Gamma(\bar{\alpha}) \Gamma(\bar{\delta})}\end{aligned}$$

La prédiction de satisfaction est donnée par:

$$\pi(\omega') = \sum_{\omega''=q_0}^1 \sum_{s=0}^{\omega''-1} \theta_0^s (1-\theta_0)^{1-s} \left[ \frac{\Gamma(\bar{\alpha} + \omega'') \Gamma(\bar{\delta} + 1 - \omega'')}{\Gamma(\bar{\alpha} + \bar{\delta} + 1)} \right] \frac{\Gamma(\bar{\alpha} + \bar{\delta})}{\Gamma(\bar{\alpha}) \Gamma(\bar{\delta})} \quad (4.3)$$

Remarquons qu'on peut généraliser à un design d'expérience où  $\omega' = \sum_{i=1}^{T'} x_i$  et on retrouve le modèle Binomial.

L'indice de satisfaction est donc:

$$\begin{aligned}\phi(\omega'') &= 0 \quad \text{si } \omega'' < q_0 \\ &= \sum_{s=0}^{\omega''-1} \binom{T'}{s} \theta_0^s (1-\theta_0)^{T'-s} \quad \text{si } \omega'' \geq q_0,\end{aligned}$$

$$q_0 = \inf \left\{ u, \sum_{t=u}^{T'} C_{T'}^t \theta_0^t (1-\theta_0)^{T'-t} \leq \alpha \right\}$$

La densité prédictive est donc:

$$\begin{aligned}
 f(\omega''|\omega') &= \int_0^1 f(\omega''/\theta) f(\theta/\omega') d\theta & (5-4) \\
 &= \int_0^1 \binom{T'}{\omega''} \theta^{\omega''} (1-\theta)^{(T'-\omega'')} [Beta(\bar{\alpha}, \bar{\delta})]^{-1} \theta^{\bar{\alpha}-1} (1-\theta)^{\bar{\delta}-1} d\theta \\
 &= \binom{T'}{\omega''} \left[ \frac{Beta(\bar{\alpha} + \omega'', \bar{\delta} + T' - \omega'')}{Beta(\bar{\alpha} + \bar{\delta})} \right]
 \end{aligned}$$

On en déduit l'indice de prédiction qui est différent de la puissance prédictive donnée par

$$\pi(\omega') = \sum_{\omega''=q_0}^{T'} \sum_{s=0}^{\omega''-1} \binom{T'}{s} \theta_0^s (1-\theta_0)^{T'-s} \left[ \binom{T'}{\omega''} \frac{Beta(\bar{\alpha} + \omega'', \bar{\delta} + T' - \omega'')}{Beta(\bar{\alpha} + \bar{\delta})} \right] \frac{\Gamma(\bar{\alpha} + \bar{\delta})}{\Gamma(\bar{\alpha}) \Gamma(\bar{\delta})} \quad (5-5)$$

*Résultat de simulation*

Le but des essais cliniques exploratoires, tels que les essais de phase II et des études de preuve de concept, est de déterminer si une nouvelle intervention est prometteur pour d'autres confirmation des tests dans les essais cliniques, tels que la phase III des essais contrôlés randomisés. L'essai clinique est une étude prospective pour évaluer l'effet des interventions chez les humains dans des conditions prédéterminées, est une partie intégrante de la norme et de la médecine moderne. De nombreuses approches adaptatives et séquentielles ont été proposées pour une utilisation dans les essais cliniques pour permettre des adaptations ou des modifications à des aspects d'un essai après l'ouverture sans porter atteinte à la validité et l'intégrité d'essai. En toute rigueur, la règle de Jeffreys donne différentes a priori pour les différents modèles, car il est basé sur les informations de Fisher, showd que Jeffreys offre de nouvelles perspectives pour le développement des procédures bayésiennes avec des bonnes propriétés fréquentiste dans les procédures de test d'hypothèse.

Supposons que deux modalités d'imagerie (par exemple, CT vs.MRI) pour le diagnostic de cancer du poumon sont à comparer sur la base de la précision du test (sensibilité, spécificité, et l'aire sous la courbe ROC).(H. Merabet, A. Labdaoui, 2015).

Suppose que  $T=T'=20$  sont les tailles d'échantillon des deux groupes et la probabilité a priori de l'hypothèse nulle est  $\theta_0 = 0,6$ . La probabilité prédictive  $\pi(\omega')$  au niveau de chaque point est calculé par simulation dans le tableau 5.1, avec  $\pi_1(\omega')$  représente l'a priori uniforme Beta (1,1),  $\pi_2(\omega')$  représente l'a priori non-informative Beta (0.5,0.5),  $\pi_3(\omega')$  est les donnée réel avec une a priori vague Beta (0.6,0.4).

Tableau 5.1. La prédiction dans le design expérimentale

scenarios	$\omega'$	$\pi_1(\omega')$	$\pi_2(\omega')$	$\pi_3(\omega')$
1	0	0.000000004	0.000000001	0.000000002
2	1	0.000000006	0.000000003	0.000000004
3	2	0.000000005	0.000000003	0.000000003
4	3	0.000000003	0.000000002	0.000100
5	4	0.000108	0.000008	0.000368
6	5	0.000376	0.0003	0.000376
7	6	0.001126	0.0010	0.001150
8	7	0.002990	0.0028	0.003157
9	8	0.007164	0.0071	0.007760
10	9	0.015690	0.0160	0.017329
11	10	0.031708	0.0331	0.035516
12	11	0.059535	0.0633	0.067316
13	12	0.104377	0.1124	0.118618
14	13	0.171459	0.1861	0.195014
15	14	0.264466	0.2880	0.299757
16	15	0.383432	0.4170	0.431174
17	16	0.522629	0.5650	0.580360
18	17	0.669467	0.7161	0.730649
19	18	0.805738	0.8490	0.860491
20	19	0.912228	0.9437	0.950618
21	20	0.976381	0.9914	0.993835

Nous notons dans le tableau 5.1 avec:

- l'a priori uniforme Beta (1,1), où  $\omega'$  est compris dans  $[0, 17]$ , le résultat de  $\theta_0 = 0.6$  varie entre 0.000000004 et 0.6694. Par conséquent, nous concluons  $H_0$  pour  $\omega' < 17$ . d'autre part, lorsque  $\omega'$  est compris dans  $[18, 20]$  le résultat  $\pi_1(\omega')$  varie entre 0.8057 et 0.9763.

- l'a priori non-informative Beta (0.5,0.5), où  $\omega'$  est compris dans  $[0, 17]$ , le résultat de  $\theta_0 = 0.6$  varie entre 0.000000007 et 0.7161. Par conséquent, nous concluons  $H_0$  pour  $\omega' < 17$ . d'autre part, lorsque  $\omega'$  est compris dans  $[18, 20]$  le résultat  $\pi_2(\omega')$  varie entre 0.8490 et 0.9914.

- les données réel avec une a priori Beta (0.6,0.4), où  $\omega'$  est compris dans  $[0, 17]$ , le résultat de  $\theta_0 = 0.6$  varie entre 0.000000002 et 0.7306. Par conséquent, nous concluons  $H_0$  pour  $\omega' < 17$ . d'autre part, lorsque  $\omega'$  est compris dans  $[18, 20]$  le résultat  $\pi_3(\omega')$  varie entre 0.8604 et 0.9938.

## 4.2.2 Le design séquentiel

Le design des méthodes statistiques séquentielles a été motivé par la nécessité d'obtenir bénéfices cliniques sous certaines contraintes économiques ou éthiques. Pour un essai avec un résultat positif, l'arrêt de l'essai signifie qu'un nouveau produit peut être exploité plus tôt. Si un résultat négatif est indiqué, l'arrêt de l'essai garantit que les ressources ne sont pas gaspillées.

Nous proposons dans les designs séquentiels, une satisfaction faite par la première et la seconde phase de l'expérience  $(\omega', \omega'')$  dans le cas d'une étude de test classique, et il est prédit en utilisant la différence de la première phase de travail précédente où seul le résultat de la deuxième phase  $\omega''$  est d'établir la conclusion formelle de l'étude (5.2.1), avec le modèle binomial précédent Nous utilisons un test d'habitude sur les résultats  $z$  de la première et la seconde phase définie par :  $z = \omega', \omega''$ .

Les méthodes séquentielles de groupe sont utilisés régulièrement pour surveiller les essais cliniques et de fournir l'arrêt de l'essai lorsqu'il existe des preuves d'un effet de traitement, une absence d'effet ou de préoccupations au sujet de la sécurité des patients. Dans de nombreuses études, la réponse clinique d'intérêt est mesuré un certain temps après le début du traitement et il ya des sujets à chaque analyse

intermédiaire qui ont été traités mais ne sont pas encore répondu (N. A Azadi, P. Fearnhead, G Ridall and J. H. Blok, (2014). L'indice de satisfaction pour l'observation  $z$  est:

$$\begin{aligned}\phi(z) &= 0 \quad \text{si } z < q_0 \\ &= \sum_{t=0}^{z-1} C_{T+T'}^t \theta_0^t (1-\theta)^{T+T'-t} \quad \text{si } z \geq q_0,\end{aligned}$$

avec

$$q_0 = \inf \left\{ u, \sum_{t=U}^{T+T'} C_{T+T'}^t \theta_0^t (1-\theta)^{T+T'-t} \leq \alpha \right\}$$

D'autre part la densité prédictive de  $z/\omega'$  est une beta-binomiale:

$$f(z|\omega') = \int_0^1 f(z/\theta) f(\theta/\omega') d\theta \quad (5-6)$$

$$= \int_0^1 C_{T+T'}^z \theta^z (1-\theta)^{(T+T'-z)} \frac{\theta^{a+\omega'-1} (1-\theta)^{T+b-\omega'-1}}{B(a+\omega', T+b-\omega')} d\theta \quad (5-7)$$

$$= C_{T+T'}^z \frac{Beta(a+\omega'+z, 2T+T'+b-z-\omega')}{Beta(a+\omega', T+b-\omega')} \quad (5-8)$$

A partir de l'observation de la réponse de la première étape  $\omega'$ , La prédiction  $\Pi(\omega')$  est-

$$\Pi(\omega') = \sum_{z=q_0}^{T+T'} \sum_{t=0}^{z-1} C_{T+T'}^t \theta_0^t (1-\theta)^{T+T'-t} \left[ C_{T'}^{z-\omega'} \frac{\text{Beta}(a+z, T+T'+b-z-\omega')}{\text{Beta}(a+\omega', T+b-\omega')} \right] \quad (5-9)$$

*Résultat de simulation*

Le design séquentiel de groupe est probablement l'un des modèles les plus couramment utilisés sur les essais cliniques dans la recherche clinique et de développement, les principales raisons de la réalisation d'analyses de données intermédiaires à payer. Supposons que sont la taille des échantillons des deux groupes et la probabilité a priori des hypothèses nulles est  $\theta_0=0,6$ . La probabilité prédictive  $\Pi(\omega')$  au niveau de chaque point est calculé par simulation dans le tableau 5.2, avec  $\Pi_1(\omega')$  représente l'a priori uniforme Beta (1,1),  $\Pi_2(\omega')$  représente l'a priori non-informative Beta (0.5,0.5),  $\Pi_3(\omega')$  est les donnée réel avec une a priori vague Beta (0.6,0.4).

Tableau 5.2. La prédiction dans le design séquentiel

Scenarios	$\omega'$	$\Pi_1(\omega')$	$\Pi_2(\omega')$	$\Pi_3(\omega')$
1	0	0.0000001	0.00000007	0.00000005
2	1	0.000002	0.0000009	0.000001
3	2	0.00014	0.000008	0.00010
4	3	0.00065	0.00047	0.00055
5	4	0.0023	0.0019	0.0022
6	5	0.0070	0.0060	0.0067
7	6	0.0174	0.0159	0.0173
8	7	0.0382	0.0362	0.0390
9	8	0.0746	0.0728	0.0776
10	9	0.1314	0.1311	0.1383
11	10	0.2115	0.2144	0.2241
12	11	0.3141	0.3216	0.3335
13	12	0.4337	0.4466	0.4597
14	13	0.5611	0.5788	0.5919
15	14	0.6846	0.7051	0.7170
16	15	0.7932	0.8136	0.8231
17	16	0.8789	0.8964	0.9032
18	17	0.9387	0.9515	0.9555
19	18	0.9746	0.9821	0.9841
20	19	0.9922	0.9956	0.9963
21	20	0.9986	0.9995	0.9997

Nous notons dans le tableau 5.2 avec:

- l'a priori uniforme Beta (1,1), où  $\omega'$  est compris dans  $[0, 15]$ , le résultat de  $\theta_0 = 0.6$  varie entre 0.0000001 et 0.7932. Par conséquent, nous concluons  $H_0$  pour  $\omega' < 15$ . d'autre part, lorsque  $\omega'$  est compris dans  $[16, 20]$  le résultat  $\Pi_1(\omega')$  varie entre 0.8789 et 0.9986.

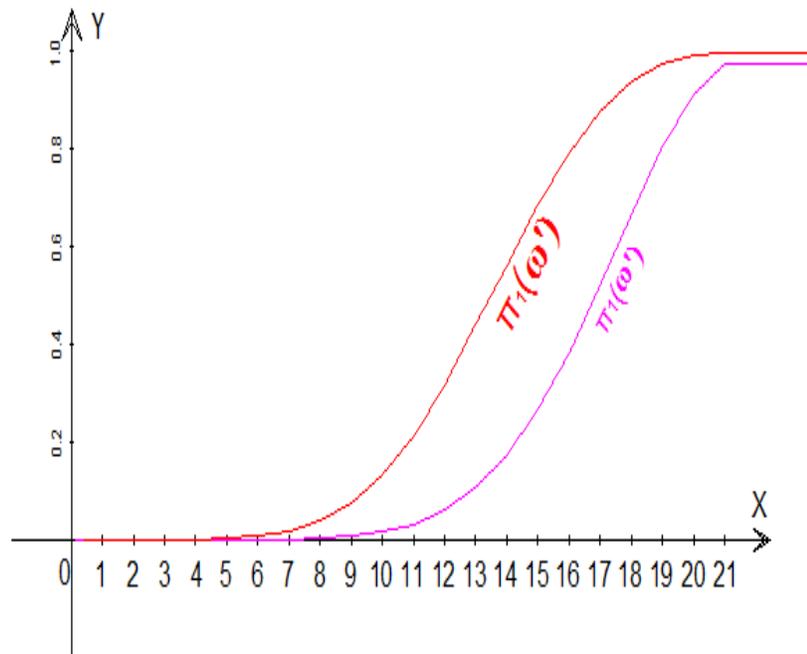
- l'a priori non-informative Beta (0.5,0.5), où  $\omega'$  est compris dans  $[0, 15]$ , le résultat de  $\theta_0 = 0.6$  varie entre 0.00000007 et 0.8136. Par conséquent, nous concluons  $H_0$  pour  $\omega' < 15$ . d'autre part, lorsque  $\omega'$  est compris dans  $[16, 20]$  le résultat  $\Pi_2(\omega')$  varie entre 0.8964 et 0.9995.

- les données réel avec une a priori Beta (0.6,0.4), où  $\omega'$  est compris dans  $[0, 15]$ , le résultat de  $\theta_0 = 0.6$  varie entre 0.00000005 et 0.8231. Par conséquent, nous concluons  $H_0$  pour  $\omega' < 15$ . d'autre part, lorsque  $\omega'$  est compris dans  $[16, 20]$  le résultat  $\Pi_3(\omega')$  varie entre 0.9032 et 0.9997.

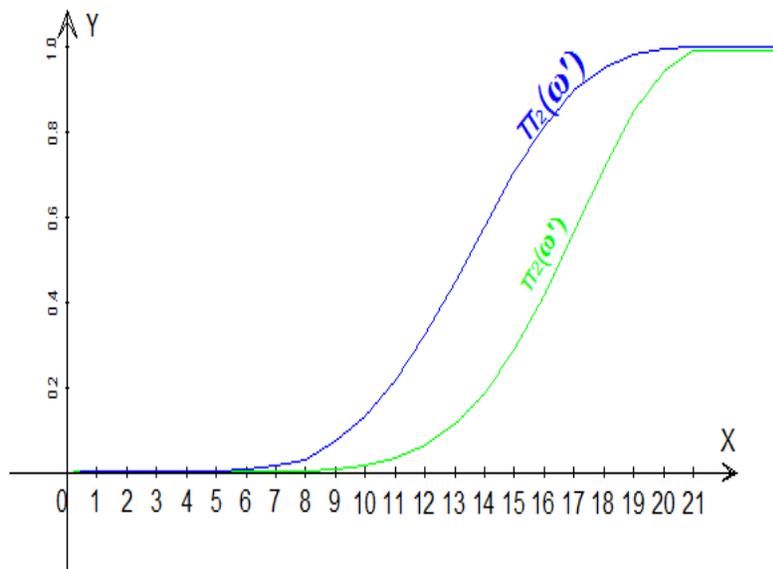
### 4.2.3 Comparaison entre les deux prédictions

Si on note par  $\pi_1(\omega')$ ,  $\pi_2(\omega')$  et  $\pi_3(\omega')$  l'indice de prévision dans le cas de design expérimentale, et  $\Pi_1(\omega')$ ,  $\Pi_2(\omega')$  et  $\Pi_3(\omega')$  l'indice de prévision dans le cas de design séquentielle. la comparaison entre les deux cas elle est présentée par les courbes dans les figures 5-1, 5-2 et 5-3.

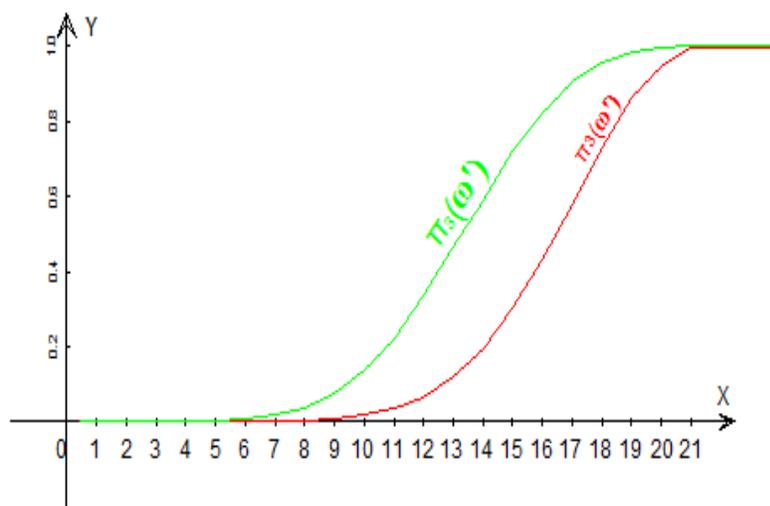
Nous pouvons conclure que la prédiction dans le design séquentielle est meilleure que dans le design expérimental.



Figure(5, 1) : Beta (1, 1)



Figure(5, 2) : Beta (0.5, 0.5)



Figure(5, 2) : Beta (0.6, 0.4)

#### 4.2.4 Conclusion

Le but de notre travail était de proposer le cadre bayésien des modèles expérimentaux et le design séquentiel dans les essais expérimentaux, considéré comme la procédure basée sur la satisfaction de concept design de l'indice de prédiction, nous croyons que nous pouvons dire que l'approche bayésienne prédictive proposée peut être utilisée pour prédire sur la base des états des résultats fréquentiste, nous pensons que l'approche fréquentiste jette un éclairage différent sur les données et ne doit pas être exclu de poursuivre notre objectif est de développer un design efficace et flexible qui possède des propriétés statistiques souhaitables. Nous avons amélioré la méthodologie dans la conception des essais cliniques en offrant une prédiction indices dans un cadre bayésien, comme c'est toujours le cas dans le protocole de l'essai expérimental et nous illustrons nos résultats en utilisant le modèle binomial. En bref, c'est un travail extrêmement utile pour des essais expérimentaux des statisticiens qui souhaitent se tenir au courant avec les approches innovantes qui sont développées au milieu des controverses quant à leurs avantages. Nous croyons qu'il fournit une contribution précieuse à la zone de design des essais expérimentaux séquentiels.

## Conclusion générale

---

### Conclusion générale

Notre travail porte sur l'étude de l'analyse Bayésienne dans les essais cliniques. Dans le domaine de l'essai clinique, l'approche Bayésienne apparaît bien adaptée à l'accumulation progressive d'informations sur l'efficacité et la toxicité de traitements. L'un des éléments permettant de distinguer les statistiques fréquentiste et Bayésiennes est la direction de l'inférence. L'approche fréquentiste est déductive. Elle part de l'hypothèse et la conforte aux données. L'approche Bayésienne est inductive. Elle part des données pour estimer la distribution du paramètre inconnu à partir des données analysées. On a consacré à une synthèse de la théorie sur l'inférence bayésienne et le poids de l'a priori qui résume la connaissance disponible sur le paramètre d'intérêt dans la réponse bayésienne. On a étudié aussi l'utilisation des méthodes Bayésiennes dans des essais de phase II, où l'on souhaite soit éliminer précocement un traitement inefficace à partir des prédictions (futilité), détecter au plus tôt le bénéfice du traitement (efficacité) ou bien déterminer la taille de l'échantillon c'est à dire le nombre de sujets à inclure au cours d'un essai.

Les progrès phénoménaux effectués en termes de puissance par l'informatique de ces dix dernières années, tout comme l'introduction pour la modélisation des fonctions a priori des méthodes de Monte Carlo par Chaîne de Markov (MCMC), ont apporté les éléments manquant à une exploitation riche et efficace de l'approche bayésienne, on a bien étudié ces notions. Une application par Win BUGS au modèle linéaire et Logit a été envisagée.

. On a terminé notre travail par la présentation de l'indice de satisfaction et le calcul des probabilités prédictives dans un modèle de données binaires avec différentes loi a priori (uniforme, non informative) dans le design expérimental et le design séquentiel et comparé les résultats des deux designs et on a conclu que les méthodes séquentielles sont meilleures.

## Perspectives

---

### Perspectives

Nous envisageons la prédiction toujours dans le cas séquentiel pour un test bayésien de modèle binomial en comparant différents designs à l'aide de plusieurs choix d'indice de satisfaction à savoir par exemple:

$$E(\theta | z)$$

et

$$((1-p(z))/\alpha)$$

# Bibliographie

- [1] A. Agresti, *Categorical Data Analysis (Volume 359)*, de Wiley Series in Probability and Statistics, 2002.
- [2] A. Altaleb and C.P. Robert, Analyse bayésienne du modèle logit: Algorithme par tranches ou metropolis-Hastings? *Revue de Statistique Appliquée*, Tome, 49(4) (2001), 53-70.
- [3] Adcock, C. J. A Bayesian Approach to Calculating Sample Sizes. *The Statistician: Journal of the Institute of Statisticians*, 37: 433–439. 1988.
- [4] Adcock, C.J. Sample Size Determination: A Review. *The Statistician: Journal of the Institute of Statisticians*, 46: 261–283. 1997.
- [5] A. Labdaoui, H. Merabet, Bayesian prediction in clinical trials design, *European Scientific Journal*, edition vol.9, No.36, December 2013.
- [6] Berry. S. M, Carlin. B.P., Lee J.J, Muller P, *Bayesian Adaptive Methods for Clinical Trials*. Chapman & Hall/CRC biostatistics series, 2011.
- [7] Best, N.G., Cowles, M.K. et Vines, S.K. 1995, CODA :Convergence Diagnosis and Output Analysis software for Gibbs sampling output. Medical Research Council/ Bioslalislics Unil, Cambridge University. [bsu.cam.ac.uk/bugs/2005](http://bsu.cam.ac.uk/bugs/2005).
- [8] Chib. S, *The Gibbs Sapmpling Algorotms*, <http://fedc.wiwi.Huberlin.de/xplore/ebooks/html/csa/node28.html>, 2010.
- [9] Christian P. Robert CEREMADE, Université Paris Dauphine et CREST, INSEE, Paris, *Le choix Bayésien: Principes et pratique*, Springer-Verlag France-Paris, 2006.
- [10] Chow S.C., J. Shao and H. Wang. *Sample Size Calculations in Clinical Research (see Chapter 13)*. Chapman & Hall/CRC, 2008.

## *Bibliographie*

---

- [11] Cowles, M.K. et Carlin, B.P, Markov Chain Monte-Carlo convergence diagnostics: A comparative study. *Journal of the American Statistical Association*, 91, 883-904. 1996.
- [12] Cyr E. M'lan, Lawrence Joseph and David B. Wolfson, Bayesian Sample Size Determination for Binomial Proportions, *Bayesian Analysis*, 3, Number 2, pp. 269–296, 2008.
- [13] Éric Parent, Jacques Bernier, *Le raisonnement bayésien Modélisation et inférence*, Springer-Verlag France, Paris, 2007.
- [14] Gary Koop, Dale J. Poirier, and Justin L. Tobias, *Bayesian econometric methods*, 2007.
- [15] Grouin J.M., M. Coste, P. Bunouf and B. Lecoutre, Bayesian sample size determination in non-sequential clinical trials: Statistical aspects and some regulatory considerations. *Statistics in Medicine*, 26:4914–4924, 2007.
- [16] Guosheng Yin, Phase II with Bayesian adaptive randomization and predictive probability. *Appl. Statist.* 61, Parts 2, pp. 219-235. 2012.
- [17] Jean-Jacques Boreux, Eric Parent, Jacques Bernier, *Pratique du calcul bayésien*, Springer-Verlag France, Paris, 2010.
- [18] Jean-Michel Marin Christian P. Robert, *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*, Springer Science Business Media, LLC, 2007.
- [19] JM. Grouin., *Procédures bayésiennes prédictives pour les essais expérimentaux*, thèse de Doctorat de l'Université René Descartes, Paris, 1994.
- [20] Jérôme Dupuis (LSP-UPS), *Statistique bayésienne et algorithme MCMC*, .2007.
- [21] Joseph, L., Wolfson, D. B., and Berger, R. d. "Sample Size Calculations for Binomial Proportions Via Highest Posterior Density Intervals." *The Statistician: Journal of the Institute of Statisticians*, 44: 143–154. 1995.
- [22] Labdaoui. A, Merabet. H, *Statistical Bayesian Analysis of Experimental Data*, *Int. J. Pure Appl. Sci. Technol.*, 16(1), 2013, pp. 7-19.
- [23] Labdaoui. Ahlam, Merabet. Hayet, *Régression logistique dans les essais cliniques par MCMC*, *Rev. Ivoir. Sci. Technol.*, 19, 2012, 59 – 71.

## *Bibliographie*

---

- [24] Lyle D. Broemeling, Bayesian Biostatistics and Diagnostic Medicine, Taylor & Francis Group, LLC, 2007.
- [25] Martin, G. Notes de cours, MAT2717, 2009.
- [26] Mazet, V, Introduction aux méthodes de Monte Carlo par chaînes de Markov. [http://lsiit-miv.u-strasbg.fr/lsiit/perso/mazet/publis/mazet\\_mtde03.pdf](http://lsiit-miv.u-strasbg.fr/lsiit/perso/mazet/publis/mazet_mtde03.pdf), 2010.
- [27] Merabet, H, Index and prevision of satisfaction in exponential models for clinical trials. STATISTICA, anno LXIV, n.3, 441-453, 2004.
- [28] Merabet.H, Labdaoui.A, Bayesian prediction using two stages design in experimental trials , Revue « Nature & Technologie ». A- Sciences fondamentales et Engineering, n° 12/ Janvier 2015. Pages 11 à 23.
- [29] O'hagan. A, Stevens. J.W, Campbell. M.J., Assurance in clinical trial design,pharmaceut;Statist,4,187-201, 2005.
- [30] Pezeshk, H. Bayesian techniques for sample size determination in clinical trials: a short review. Statistical Methods in Medical Research 12, 489–504. 2003.
- [31] Pham-Gia, T. and Turkkan, N. Sample size determination in Bayesian analysis. The Statistician 41, 389–397, 1993.
- [32] Robert, C., Casella, G, Monte Carlo Statistical Methods, 2nd Edition. Springer texts in statistics, 2004.
- [33] Robert, C.P. and Casella, G. Monte Carlo Statistical Methods . New York: Springer Verlag , 1999.
- [34] Robert, C.P. L'analyse statistique bayésienne Economica , Paris, 1992.
- [35] Satoshi Teramukai,a, Takashi Daimonb and Sarah Zoharc, A Bayesian predictive sample size selection design for single-arm exploratory clinical trials, statistics in medicine, 2012.
- [36] Shein Chung Chow Mark chang, Adaptive design methods in clinical trials, Chapman & Hall/CRC biostatistics series, 2007.
- [37] S. Michael, G .Hamada Alyson, C. Wilson. Shane Reese Harry F. Martz, Bayesian Reliability, Springer Series in Statistics, 2008.

## *Bibliographie*

---

- [38] Spiegelhalter D., K.R. Abrams and J.P. Myles. Bayesian Approaches to Clinical Trials and Health-Care Evaluation (see Section 6.5.5). Wiley, Chichester , 2004.
- [39] Spiegelhalter, D.J., Thomas, A., Best, N.G. et Gilks, W.R. (1996a) BUGS :Bayesian Inference Using Gibbs Sampling, version 1.4.1. Medical Research Council Biostatistics Unit, Cambridge University. <http://www.mrc-bsu.cam.ac.uk/Jbugs/>(2005).
- [40] Steyerberg, E.W. Clinical Prediction Models A Practical Approach to Development, Validation, and Updating, XXVIII, 500p, Hardcover, 2009.
- [41] Ton J Cleophas, Aeilko H Zwinderman, Toine F Cleophas, Statistics Applied to Clinical Trials, 1 vol. (XIX-559 p.): ill.; 25 cm, 2006.
- [42] Walsh, B, Markov Chains Monte Carlo and Gibbs Sampling, Lecture Notes for EEB581. [http://membres-timc.imag.fr/Olivier.Francois/mcmc\\_gibbs\\_sampling.pdf](http://membres-timc.imag.fr/Olivier.Francois/mcmc_gibbs_sampling.pdf), 2010.

---

## ملخص

تتكون هذه الأطروحة من أربعة ابواب:

الفصل الأول يتحدث على نظرية القرار واختيار القانون بداهة واستدلال بدافع من التطبيقات الحسابية السلطة النظرية الافتراضية وحجم العينة مع البيانات الثنائية، مرورا بطرائق مجمجي الفصل الثاني ، وتطبيق هذه الأساليب لنموذج الانحدار الخطي ونموذج اللوغارتميين في الفصل الثالث. وأخيرا ناقشنا التنبؤ في التصميم التجريبي ومتابعة في الفصل الأخير

الكلمات المفتاحية: الطاقة النظرية الافتراضية، أساليب مجمج، نماذج الانحدار، التنبؤ

---