

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
Ministère de l'enseignement Supérieur et de la Recherche Scientifique
Université Frères Mentouri Constantine 1 - Constantine



Faculté des Sciences Exactes et Informatique

Département de Mathématiques

N° d'ordre : 123/Ds/2019.

N° de série : 06/Math/2019.

Thèse

présentée pour l'obtention du diplôme de

Doctorat Ès-Sciences

Spécialité : Probabilités et Statistique

Option : Statistique Appliquée

Présentée par :

ZOHRA DJERIDI

**Choix de l'a priori en statistique
Bayésienne dans les essais expérimentaux.**

Dirigée par :

HAYET MERABET

Soutenue le 07 Juillet 2019 devant le jury composé de :

Président : *M^r* BESSILA KHALED (M.C.A) Univ. Frères Mentouri, Constantine 1.
Rapporteur : *M^{me}* MERABET HAYET (Prof.) Univ. Frères Mentouri, Constantine 1.
Examineur: *M^r* HAMRI NASR-EDDINE (Prof.) Centre universitaire de Mila.
Examineur: *M^r* BENCHAAABENE ABBES (M.C.A) Univ de Guelma .
Examineur: *M^r* BERKANE ABDELHAK (M.C.A) Univ. Frères Mentouri, Constantine 1.

Dédicace

À MA TRÈS CHÈRE MÈRE

À la plus douce et la plus merveilleuse de toutes les mamans. Tu n'a pas cessé de me soutenir et de m'encourager, ton amour, ta générosité exemplaire et ta présence constante ont fait de moi ce que je suis aujourd'hui. Tes prières ont été pour moi un grand soutien tout au long de mes études.

À MON TRÈS CHER PÈRE

À celui qui m'a aidé à découvrir le 'savoir' le trésor inépuisable. De tous les pères, tu as été le meilleur, tu as su m'entourer d'attention, m'inculquer les valeurs nobles de la vie, m'apprendre le sens du travail, de l'honnêteté et de la responsabilité.

Que ALLAH te rend à moi.

À MON CHER MARI MESSAOUD

Dans tes yeux, j'ai toujours pu lire de la tendresse, Tu m'as toujours soutenu, compris et réconforté tu es et restera toujours ma source d'inspiration. Merci pour ton attention, ta patience et tes encouragements.

À MES ENFANTS MOHAMED AMINE FERDOUS ET ZAKARIA.

À MES CHÈRS FRÈRES ET SOEURS

Salah Addine ; Abdou ; Séddik ; Moufida ; Farès, et surtout à ma chère Bochra, merci pour cette complicité que nous partageons, et pour ta capacité à toujours croire en ta grande soeur.

Plus largement, merci à tous les membres de mes deux familles Djeridi et Boudelal pour tous ces bons moments partagés.

Je désire remercier mes amis de l'Université de MSBY de Jijel et surtout l'équipe de Probabilités et Statistique, qui m'ont soutenue et divertie tout au long de l'accomplissement de ce travail.

Merci en particulier à **Leyla Tebboub** qui m'a aidé à finaliser cette thèse.

Remerciement

Tout d'abord, nous tenons à remercier Allah, le tout puissant et le miséricordieux, de nous avoir donné la santé, la volonté et la patience pour mener à terme ce travail.

*Mes premiers remerciements s'adressent à ma directrice de recherche, M^{me} **MERABET HAYET**, professeur au département de mathématiques de l'Université Constantine 1, pour son indispensable soutien tout au long de la réalisation de ma thèse. J'ai beaucoup apprécié travailler sous sa direction, d'autant plus que ceci m'a permis d'approfondir mes connaissances grâce à ses judicieux conseils et ces encouragements.*

*Je tiens à remercier, Mr **BESSILA KHALED**, d'avoir accepté la présidence du jury de mon travail, qu'il trouve ici toutes mes expressions respectueuses.*

*Je remercie également, Mr **HAMRI NASR-EDDINE**, de m'avoir fait l'honneur de faire partie des membres du jury et d'examiner ce travail. je tiens à vous remercier.*

*Je remercie également, Mr **BENCHAABENE ABBES** de m'avoir fait l'honneur de faire partie des membres du jury et d'examiner ce travail. Je tiens à vous remercier.*

*Je remercie également, Mr **BERKANE ABDELHAK**, de m'avoir fait l'honneur de faire partie des membres du jury et d'examiner ce travail. Je tiens à vous remercier.*

Mes remerciements les plus sincères s'adressent à ma famille pour leur soutien sans faille et pour l'équilibre qu'elle m'a apporté et pour leurs encouragements.

Enfin, je veux remercier toutes les personnes qui ont contribué de loin ou de près à l'avancement de ce travail.

TABLE DES MATIÈRES

| | |
|---|----------|
| Introduction générale | 1 |
| 1 Choix de la loi a priori | 7 |
| 1.1 Introduction | 7 |
| 1.2 Lois a priori subjectives | 8 |
| 1.2.1 Construction axiomatique de lois a priori | 8 |
| 1.2.2 Approximation à la distribution a priori | 10 |
| 1.2.3 Lois a priori d'entropie maximale | 11 |
| 1.2.4 Lois conjuguées | 14 |
| 1.2.5 Loi conjuguée de la famille exponentielle | 16 |
| 1.3 Distribution a priori non informative | 17 |
| 1.3.1 Lois a priori invariantes | 18 |
| 1.3.2 Lois a priori de Jeffreys | 19 |
| 1.3.3 Lois a priori de référence | 20 |
| 1.4 Lois a priori impropres | 20 |
| 1.5 Modèle hiérarchique | 22 |

| | | |
|----------|--|-----------|
| 1.5.1 | Décompositions conditionnelles | 22 |
| 1.6 | Incorporation de l'information a priori en pratique | 23 |
| 1.7 | Le poids de l'a priori dans la statistique bayésienne | 25 |
| 1.8 | Étude de la sensibilité de la réponse bayésienne à la loi a priori | 28 |
| 1.9 | Méthode de Monte Carlo | 29 |
| 1.9.1 | Principe de la méthode | 29 |
| 1.9.2 | Contexte bayésien | 29 |
| 1.9.3 | Méthode de Monte Carlo par Chaîne de Markov (Méthode MCMC) | 30 |
| 1.10 | Exemple | 33 |
| 1.10.1 | Méthode classique | 34 |
| 1.10.2 | Méthodes bayésiennes | 34 |
| 1.10.3 | Interprétation des résultats | 37 |
| 2 | Analyse séquentielle | 39 |
| 2.1 | Introduction | 39 |
| 2.2 | L'a priori séquentielle | 39 |
| 2.2.1 | Exhaustivité | 40 |
| 2.2.2 | Notion de séquentialité | 42 |
| 2.2.3 | Choix de l'a priori initiale | 44 |
| 2.3 | Prédiction séquentielle | 45 |
| 2.3.1 | Le modèle bêta-binomiale | 46 |
| 2.4 | Tests séquentiels | 47 |
| 2.4.1 | Inférence non-décisionnelle | 48 |
| 2.5 | Tests séquentiels dans les essais cliniques à un seul bras | 51 |
| 2.5.1 | Cadre général | 51 |

| | | |
|----------|---|-----------|
| 2.5.2 | Design à une seule étape | 52 |
| 2.5.3 | Design à plusieurs étapes | 54 |
| 2.6 | Règles d'arrêt utilisés dans les essais cliniques | 55 |
| 2.6.1 | Design fréquentiste | 56 |
| 2.6.2 | Design Bayésien | 57 |
| 3 | Choix de l'a priori dans la planification des essais cliniques de phase II | 67 |
| 3.1 | Procédure de test dans un cadre hybride | 68 |
| 3.1.1 | Test de Neymann-Pearson | 68 |
| 3.1.2 | Indice de satisfaction | 69 |
| 3.1.3 | Prédiction de satisfaction | 70 |
| 3.2 | Applications | 72 |
| 3.2.1 | Cas de la loi Binomiale | 72 |
| 3.2.2 | Échantillonnage de Poisson | 73 |
| 3.2.3 | Modèle Gamma | 74 |
| 3.2.4 | Modèle Gaussien | 74 |
| 3.3 | Exemples | 76 |
| 3.3.1 | Cas des résultats binaire | 76 |
| 3.3.2 | Simulation du modèle Gaussien par Monte Carlo | 77 |
| 3.4 | La règle d'arrêt PS | 79 |
| 4 | PS designs pour surveiller les essais cliniques en plusieurs étapes | 83 |
| 4.1 | Apeçu sur les conceptions bayésiennes de phase II | 84 |
| 4.2 | Indice de satisfaction dans l'analyse intermédiaire | 86 |
| 4.2.1 | Indice de Satisfaction pour deux étapes | 86 |
| 4.2.2 | Choix de la distribution a priori dans le cadre séquentiel | 88 |

| | | |
|-------|---|------------|
| 4.3 | Utilisation séquentielle de PS pour les réponses binaires | 89 |
| 4.3.1 | PS-Designs pour les essais en deux étapes | 92 |
| 4.3.2 | Comparaison entre PS-designs et les designs en deux étapes de Simon | 93 |
| 4.4 | Exemple illustratif | 95 |
| 4.5 | Utilité de PS pour futilité | 97 |
| 4.5.1 | Comparaison des différentes métriques bayésienne | 98 |
| 4.5.2 | Discussion des bornes d'arrêt d'Emersson-Fleming | 99 |
| | Conclusion générale | 101 |
| | Annexes | 104 |
| | Résumé | 115 |
| | Abstract | 116 |
| | Bibliographie | 118 |

INTRODUCTION GÉNÉRALE

L'approche bayésienne apporte une grande souplesse à la méthodologie statistique des essais expérimentaux. Un des atouts majeurs du paradigme bayésien est la facilité avec laquelle on peut faire des prédictions sur les observations futures. En particulier, nous nous intéressons à l'utilisation de cette approche dans la prévision dans le contexte d'essais expérimentaux en raison du rôle essentiel que joue la probabilité prédictive dans la conception et la surveillance des essais (fiabilité des systèmes, médecine, biologie, écologie, . . .). (Saville et al., 2014, Hamada et al., 2008).

Les procédures prédictives bayésiennes ont largement contribué à l'inférence et à l'analyse des données. Dans cette perspective, les probabilités prédictives bayésiennes constituent un outil particulièrement utile pour communiquer avec le chercheur. Ils leur donnent une méthode très attrayante pour répondre à des questions essentielles telles que : "Compte tenu des données actuelles, quelle est la probabilité que le résultat final soit en quelque sorte concluant, ou au contraire non concluant?" Cette question est inconditionnelle en ce qu'elle nécessite toutes les valeurs possibles des paramètres. Alors que la méthode fréquentiste traditionnelle ne répond pas à ces questions, les probabilités prédictives leur donnent des réponses directes et naturelles. En particulier, les procédures prédictives peuvent être utilisées pour illustrer les effets de la planification d'une expérience avec un très petit échantillon et pour aider à la décision d'abandonner précocement une expérience.

Les essais cliniques sont des études prospectives visant à évaluer l'effet d'interventions chez l'homme dans des conditions prédéfinies. Ils sont devenus une norme et

une partie intégrante de la médecine moderne. Un essai clinique correctement planifié et exécuté est l'outil le plus déterminant pour évaluer l'effet et l'applicabilité de nouvelles modalités de traitement (Chow et al., 2012; Chow and Chang, 2007; Guosheng et al., 2012; Merabet, 2004).

La méthodologie adaptée au contexte des essais cliniques est caractérisée par de nombreuses contraintes et insatisfactions et fait l'objet d'un développement profond et continu (Chow, 2005; Hoff, 2009; Robert, 2006). L'une des raisons de cet intérêt tient probablement au fait que les autorités de santé publique sont responsables de l'autorisation de mise sur le marché des médicaments et jouent un rôle primordial dans l'élaboration d'une méthodologie rigoureuse d'essais cliniques en vue de tous les acteurs de ce domaine (industries, instituts publics de recherche, hôpitaux et revues scientifiques).

I. ASPECTS BAYÉSIENS GÉNÉRAUX

D'un point de vue pratique, l'analyse bayésienne n'est rien d'autre qu'une méthode d'analyse statistique descriptive parmi les autres. En effet, c'est pour des raisons essentiellement méthodologiques que des extensions d'analyses classiques ou traditionnelles ont été élaborées dans le cadre de la théorie bayésienne, afin d'utiliser tout simplement toutes les ressources mathématiques, qui, alliées aux immenses possibilités ouvertes par le calcul automatique, en font une théorie intéressante. Bien que sophistiquée, cette méthodologie permet de couvrir une bonne partie des domaines d'applications des méthodes d'analyses habituelles : Elle apporte à la fois des compléments appréciables dans la pratique expérimentale et des conclusions plus complètes ; elle permet d'améliorer les procédures existantes parfois mal adaptées à des situations particulières.

L'origine de cette idée remonte à Bayes qui l'a introduite dans son célèbre mémoire de l'année 1763. D'une manière générale, cette analyse apporte, dans le domaine des modèles descriptifs, les idées suivantes :

1. L'analyse statistique permet, dans la planification des expériences, de considérer toutes les informations quantitatives et qualitatives sur l'incertitude dans les modèles ; ici, il s'agit du paramètre θ .
2. Le paramètre θ est considéré, dans l'approche classique, comme une grandeur inconnue (un vecteur dans le cas multidimensionnel), mais certaine. Dans l'approche bayésienne, il peut prendre plusieurs valeurs possibles, avec des probabilités associées, ce qui conduit à trouver des distributions sur l'espace des paramètres. Différents modèles sont adaptés à ces développements.
3. Des informations sur ce paramètre θ avant échantillonnage permettent d'évaluer

une loi de distribution, dite "a priori". L'analyse bayésienne consiste à déduire de cette loi de distribution, grâce au Théorème de Bayes, une distribution dite "a posteriori", en ajustant la valeur du paramètre par un jugement probabiliste de l'incertitude compte tenu des données recueillies dans un échantillon x .

Cette analyse bayésienne intègre, outre le plan d'échantillonnage traditionnel, les résultats d'expériences conçues antérieurement et les exploite de manière optimale lorsque les données expérimentales sont insuffisantes pour appliquer l'analyse fréquentiste.

De plus, cette analyse offre une possibilité supplémentaire : La loi a priori représente toutes les informations pertinentes provenant de données antérieures, mais il est possible d'y incorporer toute connaissance ou opinion, même très conjecturale ou "subjective" que l'on peut avoir sur le problème étudié.

Après l'établissement d'une distribution a priori du paramètre inconnu θ , si des expériences complémentaires sont réalisées, on reprendra la distribution a posteriori de ce paramètre comme nouvelle distribution a priori. On aura ainsi une distribution réactualisée qui incorporera les apports des expériences successives, d'où la méthode dite séquentielle (cf. chapitre 2). La recherche d'une stabilité de forme de ces distributions sera indispensable en pratique, notamment pour réaliser des estimations fonctionnelles, d'où la propriété de conjugaison.

La distribution a priori non-informative vise à représenter un "état d'ignorance" sur le paramètre θ ; la distribution a posteriori correspondante pourra alors être interprétée comme résultant de l'apport propre des données. Le cadre bayésien se présente donc comme une théorie formalisée de l'apprentissage par l'expérience .

II. FORMALISME

Ici, nous considérons une variable aléatoire X à distribution continue admettant $F(x|\theta)$ comme fonction de répartition et $f(x|\theta)$ comme fonction de densité. Supposons connue la forme de cette distribution, tandis que le paramètre θ est inconnu et aléatoire. Nous cherchons dans un premier temps à connaître la distribution p de θ en exprimant une connaissance a priori sur l'espace Θ des valeurs possibles de θ , puis à évaluer ensuite X sous forme d'une distribution marginale.

Supposons que nous disposons comme information a priori d'une série d'observations x_1, x_2, \dots, x_n dans un processus séquentiel, et que toutes ses expériences recueillies soient indépendantes l'une de l'autre.

L'analyse classique ignore d'une certaine manière ces observations antérieures et n'uti-

lise que l'observation x pour estimer une valeur θ_{n+1} à partir de la fonction $F(x_{n+1}|\theta)$ ou $f(x_{n+1}|\theta)$, selon la méthode utilisée et le type d'information disponible. L'approche bayésienne au contraire utilise ces observations d'une manière optimale pour ajuster une distribution a priori notée par $p(\theta)$, puis pour calculer la distribution a posteriori notée par $p(\theta|x_{n+1})$, grâce au théorème de Bayes :

$$p(\theta|x_{n+1}) = \frac{f(x_{n+1}|\theta)p(\theta)}{\int_{\Theta} f(x_{n+1}|\theta)p(\theta)d\theta}$$

Alors, la moyenne bayésienne ou prédictive (a posteriori) $E(\theta|x_{n+1})$ déduite de $p(\theta|x)$ est un estimateur de θ au sens des moindres carrés.

Nous nous plaçons ici dans le cadre paramétrique, et nous nous intéressons plutôt aux distributions $p(\theta)$ et $p(\theta|x)$, afin d'obtenir finalement les distributions prédictives :

$$f(x) = \int_{\Theta} f(x|\theta)p(\theta)d\theta;$$

$$\nu(y|x) = \int_{\Theta} f(y|\theta)p(\theta|x)d\theta$$

pour un échantillon futur y , sur lesquelles est basée l'analyse prévisionnelle. Ceci constitue un outil supplémentaire pour le processus de contrôle, permettant éventuellement de réduire le volume d'essais ou la taille d'échantillons et de traiter le cas de petits échantillons sans perdre de précision ou de fiabilité.

Pour démarrer l'analyse bayésienne, il faut avant tout évaluer la distribution a priori de la variable θ , au lieu d'estimer une valeur certaine dans le cas classique. Nous avons vu que d'autres distributions peuvent être obtenues facilement, en utilisant le théorème de Bayes et la théorie des distributions conditionnelles. Ainsi, la distribution a priori joue un rôle primordial dans cette procédure d'analyse. Différentes techniques sont proposées pour établir une telle distribution et associées avec certaines conditions d'applicabilité en fonction de la nature des données.

Bien entendu, toute étude dépend de la nature des données disponibles. Au vu des résultats obtenus et de la qualité demandée, on choisira un processus de traitement bayésien ou classique. Dans les cas de la vérification d'efficacité d'un traitement dans un essai clinique, les analyses sont effectués progressivement et en utilisant l'indice de satisfaction définit dans un cadre hybride Bayésien-fréquentiste on peut évaluer à quel point nous somme satisfait de l'efficacité de ce traitement.

Dans cette situation, nous avons souvent des informations expérimentales primaires sous forme de phase I, aux quelles nous devons confirmer certains résultats (Merabet, 2004 ;

Labdaoui, 2015 ; Djeridi and Merabet, 2016, Merabet et al., 2017). Formellement, nous considérons la situation suivante : En se basant sur les données du premier échantillon, nous pouvons planifier une expérience (un nouvel échantillon) afin d'avoir de bonnes chances d'obtenir la conclusion voulue si l'expérimentation n'est pas écartée. Nous proposons la procédure basée sur le concept de l'indice de satisfaction qui est une fonction décroissante de la p-valeur, et nous envisageons, compte tenu des données disponibles, de calculer une satisfaction prédite de cet indice en fonction des observations précédentes .

III. RÉSULTATS

L'utilisation d'essais adaptatifs pourrait considérablement améliorer l'efficacité du développement de médicaments ; l'intégration de l'approche bayésienne est un pas de plus dans cette direction (Chang, 2008 ; Chow et al., 2012). Les conceptions d'essais adaptatives bayésiennes constituent un outil puissant pour rationaliser le processus d'apprentissage séquentiel, prédire d'autres résultats et synthétiser les preuves sur différentes ressources. Les résultats obtenus avec les approches bayésiennes sont plus faciles à interpréter et plus utiles à la prise de décision (Jennison et Turnbull, 2000 ; *Everit and Pickles*, 1999). Pour une discussion comparant les conceptions bayésienne et fréquentiste dans le contexte des essais cliniques, voir Berry (1985), Stangl and Berry (1998), Spiegelhalter et al. (1994), et Lee and Liu (2008).

Dans de nombreuses expériences ou enquêtes au cours desquelles les données sont accumulées de manière constante sur une période donnée, il est judicieux de surveiller les résultats au fur et à mesure de leurs apparitions afin de prendre des mesures telles que l'arrêt précoce ou une modification du plan d'étude. Des méthodes statistiques séquentielles ont été développées afin d'obtenir des avantages. Pour un essai avec résultat positif, un arrêt précoce signifie qu'un nouveau produit peut être exploité plus tôt. Dans le cas contraire, un arrêt précoce garantit que les ressources ne sont pas gaspillées ; Gould (1983) a qualifié cette situation "d'abandonner une cause perdue". Les méthodes séquentielles permettent généralement d'économiser la taille d'échantillon, le temps et le coût par rapport aux procédures standard d'échantillon fixe.

Labdaoui (2015) a démontré que l'utilisation de la prédiction de satisfaction donne de bons résultats dans le design séquentiel que dans la planification expérimental à cause d'accumulation progressive d'informations. Dans cet égard, l'objectif principal de ce travail est de fournir une procédure hybride bayésienne-fréquentiste pour les conceptions

à plusieurs étapes afin de tester l'efficacité d'un nouveau traitement. Cette procédure est basée sur le concept d'indice de satisfaction (Merabet et al., 2017) ; où nous envisageons, compte tenu des données disponibles, de calculer une satisfaction prédite de cet indice par l'approche bayésienne, en tenant compte des observations précédentes. Cette valeur prédite généralise la puissance du test dans la logique d'indice de satisfaction proposé. Nombreux auteurs ont préconisé l'utilisation d'indices dans de telles situations comme Lecoutre et al. (1995), Merabet et Labdaoui (2015), Djeridi et Merabet (2016) en raison de la simplicité et de la souplesse dont ils disposent pour mesurer le degré de satisfaction dans le cas d'obtenir un résultat significatif. Cet indice est utilisé comme règle d'arrêt pour la conception d'essais cliniques de phase II.

Dans le cas de futilité pour les essais de phase II à un seul bras avec des données binaires, la condition d'acceptation du traitement est le respect d'une borne inférieure de la prédiction de satisfaction avec une probabilité donnée γ , et un risque de 1er espèce s'il s'agit de l'investigateur (contrôle final de production), ou de 2ème espèce s'il s'agit d'un patient (contrôle de réception). Mais, le minimum des données est bien inférieure au nombre de patient prévu. L'avantage de cette procédure est l'arrêt prématuré de l'essai à cause de futilité/ efficacité , et donc le coût d'essai, à la condition que les rejets de l'hypothèse nulle restent significatifs.

Cette thèse comporte quatre chapitres :

- Le premier chapitre, est consacré à présenter les différentes méthodes de la construction d'une loi a priori, et nous effectuons une étude comparative de paramètres estimés en utilisant deux lois a priori différentes d'un modèle de poisson hiérarchique.
- Dans le deuxième chapitre, nous donnons le principe des méthodes séquentielle notamment le choix de l'a priori initiale ainsi le choix de la règle d'arrêt pour les tests séquentiels fréquentistes et bayésiens.
- Dans le troisième chapitre, nous décrivons la dualité qui existe entre l'approche fréquentiste et l'approche bayésienne. En considérant un cas d'application hybride pour les tests d'hypothèses en choisissant la prédiction de satisfaction comme règle d'arrêt séquentielle dans la planification expérimental des essais cliniques.
- Finalement, Nous présentons le principe de notre conception, en donnant ses points forts par rapport aux autres méthodes et les limites de son utilisation et l'appliquer, rétrospectivement, à un essai clinique sur le cancer du poumon.

CHAPITRE 1

CHOIX DE LA LOI A PRIORI

1.1 Introduction

Le choix de la loi a priori est l'aspect le plus critiqué dans l'analyse statistique bayésienne. Ce choix peut avoir différentes motivations. Car, une fois cette loi a priori est connue, l'inférence peut être conduite d'une façon quasi mécanique en minimisant le coût a posteriori, en calculant les régions de confiance (de plus forte probabilité de la densité a posteriori) ou en intégrant par rapport au paramètres pour obtenir la distribution prédictive.

La loi a priori est la clé de voûte de l'inférence bayésienne et sa détermination est donc l'étape la plus importante dans la mise en oeuvre de cette inférence.

Comment passer des informations a priori à des lois a priori ? est la question fondamentale et légitime dans la mise en oeuvre de toute analyse au sens bayésien.

En pratique, il est rare que l'information a priori soit suffisamment précise pour conduire à une détermination exacte de la loi a priori au sens où plusieurs lois de probabilités soient des lois usuelles (loi normale, loi gamma, etc) dites conjuguées.

En absence d'information a priori on introduira la notion de loi a priori non informative qui permet de rester dans un cadre bayésien, alors même que l'on ne dispose pas d'information a priori.

L'information a priori peut être codée selon une des façons suivantes :

1. Déterminer une loi subjective ;
2. Prendre une loi a priori vague, c'est à dire non informative.

1.2 Lois a priori subjectives

Les critiques contre l'approche bayésienne ont une certaine validité au sens où elles attirent l'attention sur le fait qu'il n'y a pas une façon unique de choisir une loi a priori, et que le choix de cette loi a un impact sur l'inférence résultante. Cet impact peut être négligeable, modéré ou énorme, puisqu'il est toujours possible de choisir une loi a priori qui donnera la réponse qu'on souhaite obtenir. Mais le point essentiel ici est que, premièrement, les lois a priori non fondées fournissent des inférences a posteriori non justifiées et, deuxièmement, le concept d'une loi a priori unique n'a pas de sens, sauf dans des cas très particuliers.

1.2.1 Construction axiomatique de lois a priori

Pour démontrer l'existence d'une loi a priori subjective, DeGroot (1970) avait besoin de définir un ordre des événements :

Supposons que le décideur, le client ou le statisticien soient à déterminer une relation d'ordre sur une σ -algèbre $\mathcal{B}(\Theta)$. Cette relation, notée \preceq , est telle que :

- $B \prec A$ signifie que A est plus vraisemblable que B ,
- $B \preceq A$ que A est au moins vraisemblable que B ,
- $B \sim A$ signifie que A et B sont aussi vraisemblable l'un que l'autre.

Bien entendu, s'il existe une distribution P de probabilité sur $(\Theta, \mathcal{B}(\Theta))$, P induit directement une relation d'ordre sur $\mathcal{B}(\Theta)$.

Considérons, maintenant, les hypothèses sous lesquelles la réciproque peut être établie :

1. (A_1) La relation d'ordre est totale : Pour tout ensemble mesurable A et B , une et seulement une des relations suivantes est satisfaite :

$$A \prec B, B \prec A, A \sim B$$

2. (A_2) Si A_1, A_2, B_1, B_2 sont des ensembles mesurables vérifiant $A_1 \cap A_2 = B_1 \cap B_2 = \emptyset$ et $A_i \preceq B_i$ ($i = 1, 2$), alors $A_1 \cup A_2 \preceq B_1 \cup B_2$. De plus, si $A_1 \prec B_1$, $A_1 \cup A_2 \prec B_1 \cup B_2$.

3. (A_3) Pour tout événement A , $\emptyset \preceq A$ et $\emptyset \prec \Theta$. La condition supplémentaire $\emptyset \prec \Theta$ évite le cas trivial où tous les événements sont équivalents. Il est aussi nécessaire de permettre la comparaison d'une suite infinie d'événements.
4. (A_4) Si $A_1 \supset A_2 \supset \dots$ est une suite décroissante d'ensembles mesurables et B est un événement donné tel que $B \preceq A_i$ pour tout i , alors

$$B \preceq \bigcap_{i=1}^{+\infty} A_i$$

L'hypothèse (A_2) entraîne la transitivité de la relation d'ordre, l'hypothèse (A_3) empêche l'existence d'ensembles mesurables de vraisemblance négative (donc moins vraisemblables, que l'ensemble vide). Tandis que (A_4) assure en quelque sorte, la continuité de l'ordre des préférences et est reliée à la propriété de σ -additivité des mesures de probabilité. Cependant, les axiomes (A_1) – (A_4) sont insuffisants pour obtenir l'existence d'une distribution de probabilité à partir de l'ordre des vraisemblances. En fait, passer d'une échelle de comparaison qualitative à une comparaison quantitative requiert une dernière hypothèse.

5. (A_5) Il existe une variable aléatoire X sur $(\Theta, \mathcal{B}(\Theta))$ de distribution uniforme sur $[0, 1]$, c'est-à-dire, pour tout I_1, I_2 , intervalles de $[0, 1]$,

$$\{X \in I_1\} \preceq \{X \in I_2\}$$

si et seulement si

$$\lambda(I_1) \preceq \lambda(I_2)$$

où λ est la mesure de Lebesgue.

Cette hypothèse supplémentaire permet alors d'établir le résultat d'existence suivant (voir DeGroot, 1970, pour une démonstration).

Théorème 1.1. *Sous les axiomes (A_1) – (A_5), il existe une distribution P telle que $P(A) \leq P(B)$ si et seulement si $A \preceq B$.*

Limites de la construction axiomatique

Selon Robert (2006), les développements précédents sur les fondations axiomatiques de la loi a priori sont plus limités :

- Les hypothèses ci-dessus et le cadre formel correspondant sont plus difficiles à justifier. En fait, le fait qu'un statisticien soit à même d'exprimer la vraisemblance d'un événement

signifie qu'il a, consciemment ou pas, construit un modèle probabiliste sous-jacent et, donc, que la construction précédente est en quelque sorte tautologique.

- L'hypothèse (A_5) est particulièrement forte et peut rarement être vérifiée en pratique.

- Selon le Théorème (1.1), le décideur peut construire une loi a priori à partir de son ordre des vraisemblances. Cependant, il est très vraisemblable, surtout si Θ n'est pas fini, que cet ordre sera grossier, c'est-à-dire que la σ -algèbre $\mathcal{B}(\Theta)$ correspondante ne sera pas la σ -algèbre borélienne usuelle sur Θ , empêchant par là même l'utilisation des distributions classiques sur Θ . Cependant, il est rassurant de pouvoir justifier l'utilisation d'une loi a priori par d'autres raisonnements que ceux de l'approche fréquentiste, supposant la répétabilité des expériences, même si cela est d'un intérêt limité en pratique.

1.2.2 Approximation à la distribution a priori

A moins que le décideur (ou le statisticien) ne soit informé sur le mécanisme (physique, économique, biologique, etc.) sous-jacent de génération du paramètre θ , il est généralement très difficile de proposer une forme exacte ou même paramétré pour la distribution a priori sur θ . En fait, dans la plupart des cas, θ n'a pas de réalité propre (intrinsèque), mais correspond plutôt à une paramétrisation de la loi décrivant le phénomène aléatoire observé.

La loi P est alors un moyen de résumer l'information disponible sur ce phénomène, ainsi que l'incertitude liée à cette information. Ces situations impliquent évidemment des approximations de la vraie distribution a priori (si une vraie loi existe !). Effectivement, les modèles statistiques sont le plus souvent des représentations simplifiées de ces phénomènes aléatoires et, puisqu'il n'existe pas de vrai modèle (mais seulement un modèle plus proche du phénomène pour une distance appropriée) il est conceptuellement difficile de parler de la vraie valeur de θ et, a fortiori, d'une vraie loi a priori.

Précisons tout d'abord que cette démarche n'est pas forcément facile dans la pratique.

La démarche constructive est de représenter l'incertitude de l'expert sur θ par une distribution de probabilité a priori. Comment l'évaluer ?

L'idée générale est de présenter des valeurs ponctuelles de θ à l'expert et de lui demander de donner des chances pour chacune d'elles qu'il lui accorde.

Il est important de comprendre que ces valeurs ponctuelles peuvent être choisi par l'ex-

pert.

Quand aucune information directe n'est disponible sur θ , une alternative est de recourir à la distribution marginale de x

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

Afin d'obtenir de l'information sur P . Plusieurs techniques ont été proposées dans la littérature comme la technique d'entropie maximale.

1.2.3 Lois a priori d'entropie maximale

L'entropie est une grandeur bien connue chez les physiciens comme étant une mesure du désordre. Elle a été développée par Jaynes (1980,1983) pour déterminer la loi a priori sous certaines contraintes qui caractérisent la loi a priori, par exemple des moments, des quantiles etc.

Dans le cadre de la statistique, elle mesure la quantité d'incertitude à la loi de probabilité.

Définition 1.1. : "Dans le cas discret"

Soient Θ l'espace des paramètres discrets et finis, et p une probabilité sur Θ .

L'entropie de p notée $Ent(p)$ est définie par

$$\forall i = \overline{1..n} : \text{telque } p(\theta_i) \geq 0 \text{ et } \sum_{i=1}^n p(\theta_i) = 1$$

$$Ent(p) = - \sum_{\Theta} p(\theta_i) \log p(\theta_i),$$

cette quantité à été introduite par Shannon (1948) comme une mesure d'incertitude en théorie de l'information et en traitement du signal.

- Sous les contraintes de moments qui peuvent s'écrire comme des espérances a priori comme suit

$$E^p[g_k(\theta)] = \sum_{i=1}^n g_k(\theta_i)p(\theta_i) = w_k, \quad \forall k = \overline{1..K}. \quad (1.1)$$

L'objectif est de maximiser $Ent(p)$ sous les contraintes (1.1), donc la loi a priori d'entropie maximale est donnée par

$$p^*(\theta_i) = \frac{\exp\{\sum_{k=1}^K \lambda_k g_k(\theta_i)\}}{\sum_{\Theta} \exp\{\sum_{k=1}^K \lambda_k g_k(\theta_i)\}}$$

où les λ_k sont les multiplicateurs de Lagrange associés, qui se déterminent dans la pratique par un système d'équations à partir des contraintes.

• Sans contraintes sur p , la distribution d'entropie maximale est la distribution uniforme sur Θ .

Exemple 1.1. Soit $\Theta = \mathbb{N}$, supposons $E^p(\theta) = 5$ et $n = 1$.

On a $g_1(\theta) = \theta$, $w_1 = 5$ et $\lambda_1 < 0$

donc

$$\begin{aligned} p^*(\theta) &= \frac{\exp\{\sum_{k=1}^K \lambda_k g_k(\theta_i)\}}{\sum_{\theta \in \mathbb{N}} \exp\{\sum_{k=1}^K \lambda_k g_k(\theta_i)\}} \\ &= \left[\frac{1}{\sum_{\theta=0}^{\infty} e^{\lambda_1 \theta}} \right] e^{\lambda_1 \theta}. \end{aligned}$$

On a

$$\sum_{\theta=0}^n e^{\lambda_1 \theta} = 1 + e^{\lambda_1} + e^{2\lambda_1} + \dots + e^{n\lambda_1} = \frac{1 - e^{(n+1)\lambda_1}}{1 - e^{\lambda_1}},$$

qui est la somme d'une suite géométrique du premier terme 1 et de raison e^{λ_1} .

Donc

$$\begin{aligned} \sum_{\theta=0}^{\infty} e^{\lambda_1 \theta} &= \lim_{n \rightarrow \infty} \sum_{\theta=0}^n e^{\lambda_1 \theta} \\ &= \lim_{n \rightarrow \infty} \frac{1 - e^{(n+1)\lambda_1}}{1 - e^{\lambda_1}} \\ &= \frac{1}{1 - e^{\lambda_1}}, \end{aligned}$$

d'où

$$p^*(\theta) = (1 - e^{\lambda_1}) e^{\lambda_1 \theta}$$

est une loi géométrique de paramètre $(1 - e^{\lambda_1})$.

On détermine, alors, λ_1 par

$$E^p(g_1(\theta)) = E^p(\theta) = \frac{e^{\lambda_1}}{1 - e^{\lambda_1}} = 5,$$

ce qui donne

$$\lambda_1 = \log(5/6).$$

Donc p^* est une loi géométrique de paramètre $(1 - e^{\log(5/6)}) = 1 - \frac{5}{6} = \frac{1}{6}$.

L'extension au cas continu est plus délicate, car elle implique le choix d'une mesure de référence p_0 , qui peut être caractérisée comme la distribution complètement non

informative. Il s'agit, en effet, de l'a priori d'entropie maximale en absence de contrainte. Cette mesure de référence peut être obtenue de plusieurs façons (voir Robert (2006)) et la distribution d'entropie maximale dépend de ce choix. On peut proposer la définition suivante de l'entropie de p par rapport à une mesure p_0 ;

Définition 1.2. : "Dans le cas continu"

Une fois la mesure de référence p_0 choisie, l'entropie de p , qui est aussi la distance de Kullback-Leibler entre p et p_0 , est définie par

$$Ent(p|p_0) = E^{p_0}[\log(\frac{p(\theta)}{p_0(\theta)})] = \int_{\Theta} \log(\frac{p(\theta)}{p_0(\theta)})p_0(d\theta) \quad (1.2)$$

où $p_0(\theta)$ est la loi a priori non informative.

Comme précédemment, si on dispose d'une information a priori du type

$$E^p[g_k(\theta)] = \int_{\Theta} g_k(\theta)p(\theta)d\theta = w_k, \quad \forall k = \overline{1, K}.$$

La loi a priori d'entropie maximale est alors donnée par

$$p^*(\theta_i) = \frac{p_0(\theta) \exp\{\sum_{k=1}^K \lambda_k g_k(\theta)\}}{\int_{\Theta} p_0(\theta) \exp\{\sum_{k=1}^K \lambda_k g_k(\theta)\}d\theta}$$

où les λ_k sont les multiplicateurs de Lagrange.

Notons que les distributions p^* ci-dessus appartiennent formellement à une famille exponentielle (section 1.2.4).

Exemple 1.2. Soit $\Theta = \mathbb{R}$.

Supposons que θ est un paramètre de position. Une loi a priori naturelle non informative est alors $p_0(\theta) = 1$. Si on fixe les valeurs de la moyenne et de la variance de la loi a priori à (μ, σ^2) tel que $E^p(\theta) = \mu$ et $g_1(\theta) = \theta$, $w_1 = \mu$ et $E^p[(\theta - \mu)^2] = \sigma^2$ et $g_2(\theta) = (\theta - \mu)^2$, $w_2 = \sigma^2$.

La loi d'entropie maximale est alors

$$p^*(\theta) = \frac{\exp\{\lambda_1\theta + \lambda_2(\theta - \mu)^2\}}{\int_{\Theta} \exp\{\lambda_1\theta + \lambda_2(\theta - \mu)^2\}d\theta}$$

Calculons λ_1, λ_2 . On a :

$$\begin{aligned} \lambda_1\theta + \lambda_2(\theta - \mu)^2 &= \lambda_1\theta + \lambda_2(\theta^2 - 2\theta\mu + \mu^2) \\ &\propto \lambda_2\theta^2 - 2\theta\lambda_2(\mu - \frac{\lambda_1}{2\lambda_2}) \\ &\propto \lambda_2[\theta - (\mu - \frac{\lambda_1}{2\lambda_2})]^2 \end{aligned}$$

$p^*(\theta)$ est donc la loi normale de paramètres $(\mu - \frac{\lambda_1}{2\lambda_2}, \frac{-1}{2\lambda_2})$.

On cherche, alors, λ_1 et λ_2 tel que :

$$\mu - \frac{\lambda_1}{2\lambda_2} = \mu \text{ et } \sigma^2 = -\frac{1}{2\lambda_2}.$$

Il vient donc $\lambda_1 = 0$ et $\lambda_2 = -\frac{1}{2\sigma^2}$.

D'où $p^*(\theta)$ est la loi normale de paramètres (μ, σ^2) .

Remarque 1.1. *Seidenfeld (1987) et Kass et Wasserman (1996) avancent des critiques sur l'approche par entropie maximale. Parmi ces critiques : La dépendance à p_0 exhibée par (1.2), le manque d'invariance par paramétrisation, en plus les contraintes (1.1) ne sont pas toujours suffisantes pour obtenir une distribution sur θ .*

1.2.4 Lois conjuguées

Une des difficultés de l'approche bayésienne est le calcul de la loi a posteriori. Ce calcul sera facilité lorsque la loi a priori et la loi a posteriori sont de la même famille. Dans ce cas, on parle de lois a priori conjuguées. L'approche a priori conjuguée, introduite par Raiffa et Schlaifer (1961), peut être justifiée partiellement par un raisonnement d'invariance. En fait, quand l'observation de $x \sim f(x|\theta)$ modifie $p(\theta)$ en $p(\theta|x)$, l'information transmise par x sur θ est évidemment limitée ; par conséquent, elle ne devrait pas entraîner une modification de toute la structure de $p(\theta)$, mais simplement de ses paramètres. En d'autres termes, la modification résultant de l'observation de x devrait être de dimension finie (Robert, 2006).

Définition 1.3. *Une famille \mathcal{F} de distributions de probabilité sur Θ est dite conjuguée pour une vraisemblance $f(x|\theta)$ si, pour toute loi a priori $p(\theta) \in \mathcal{F}$, la distribution a posteriori $p(\cdot|x)$ appartient également à \mathcal{F} .*

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 comme une fonction du paramètre θ . Les lois a priori conjuguées obtenues par ce procédé sont dites naturelles.

Dans ce qui suit, nous allons aborder une des plus importantes familles de lois a priori : **la famille des lois exponentielles.**

Définition 1.4. *Soient Θ l'espace des paramètres, χ l'espace des observations, C et h*

deux fonctions dans \mathbb{R}^+ respectivement de Θ et χ et soient R et T deux fonctions de Θ et χ dans \mathbb{R}^k .

1. La famille des distributions de densité

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

est dite famille exponentielle de dimension k .

2. En cas de paramétrisation $R(\theta) \equiv \theta$, donc le modèle exponentielle est dite canonique c'est à dire

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

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

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

est dite famille exponentielle naturelle.

Exemple 1.3.

1) Soit $X \sim \mathcal{B}(n, x)$, alors

$$\begin{aligned} f(x|\theta) = P(X = x|\theta) &= C_n^x \theta^x (1 - \theta)^{n-x} \\ &= C_n^x \exp\left\{x \log\left(\frac{\theta}{1 - \theta}\right) + n \log(1 - \theta)\right\} \end{aligned}$$

donc, on a

$$R(\theta) = \log\left(\frac{\theta}{1 - \theta}\right), T(x) = x, C(x) = -n \log(1 - \theta) \text{ et } h(x) = C_n^x$$

2) Considérons le modèle :

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

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

Car

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

et la représentation : $h(x) = 1, \theta = (\theta, \beta), T(x) = (x, -x)'$

et $C(\theta, \beta) = \log(1 + e^{\theta - \beta})$.

3) Soit $x \sim \mathcal{P}(\lambda)$. Alors

$$\begin{aligned} f(x|\lambda) &= \frac{e^{-\lambda}\lambda^x}{x!} \\ &= \frac{1}{x!} e^{\theta x - e^\theta} \end{aligned}$$

donc $h(x) = \frac{1}{x!}$ et $C(\theta) = e^\theta$ pour le paramètre naturel $\theta = \log \lambda$.

On a le résultat suivant qui donne la forme des lois naturelles conjuguées dans le cas du modèle exponentiel.

1.2.5 Loi conjuguée de la famille exponentielle

Soit la fonction :

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

la loi générique d'une famille exponentielle.

Cette loi admet alors une famille conjuguée, comme le démontre le résultat suivant

Proposition 1.1 (Robert, (2006)). *Soit $f(x|\theta)$ appartenant à une famille exponentielle. Alors une famille de lois a priori conjuguées pour $f(x|\theta)$ est donnée par :*

$$p(\theta|\mu, \lambda) = K(\mu, \lambda) \exp\{\theta\mu - \lambda C(\theta)\}$$

où $K(\mu, \lambda)$ est une constante de normalisation.

Donc la loi a posteriori est de la forme :

$$\begin{aligned} p(\theta|\mu + x, \lambda + 1) &= \frac{h(x) \exp\{\theta x - C(\theta)\} K(\mu, \lambda) \exp\{\theta\mu - \lambda C(\theta)\}}{\int_{\Theta} h(x) \exp\{\theta x - C(\theta)\} K(\mu, \lambda) \exp\{\theta\mu - \lambda C(\theta)\} d\theta} \\ &= \frac{\exp\{\theta x - C(\theta)\} \exp\{\theta\mu - \lambda C(\theta)\}}{\int_{\Theta} \exp\{\theta x - C(\theta)\} \exp\{\theta\mu - \lambda C(\theta)\} d\theta} \\ &\propto \exp\{\theta x - C(\theta)\} \exp\{\theta\mu - \lambda C(\theta)\} \\ &\propto \exp\{(\mu + x)\theta - (\lambda + 1)C(\theta)\}. \end{aligned}$$

Exemple 1.4. (Suite de l'exemple 1.3(2))

En appliquant la proposition, on obtient une loi a priori de la forme :

$$p(\theta, \beta|\mu_1, \mu_2, \lambda) \propto \exp\{[\theta \ \beta][\mu_1 \ \mu_2]' - \lambda C(\theta, \beta)\} = \frac{\exp\{\mu_1\theta + \mu_2\beta\}}{(1 + \exp\{\theta - \beta\})^\lambda}.$$

La loi a posteriori aura la forme suivante :

$$p(\theta, \beta|x) \propto \frac{\exp\{(\mu_1 + x)\theta + (\mu_2 + x)\beta\}}{(1 + \exp\{\theta - \beta\})^{\lambda+1}}.$$

TABLE 1.1 – Lois a priori conjuguées pour quelque familles exponentielles usuelles

| $f(x \theta)$ | $p(\theta)$ | $p(\theta x)$ |
|--|---|--|
| Normale $\mathcal{N}(\theta, \sigma^2)$ | Normale $\mathcal{N}(\mu, \tau^2)$ | $\mathcal{N}(\rho(\sigma^2\mu + \tau^2x), \rho\sigma^2\tau^2)$ $teq\rho^{-1} = \sigma^2 + \tau^2$ |
| Poisson $\mathcal{P}(\theta)$ | Gamma $\mathcal{G}(\alpha, \beta)$ | $\mathcal{G}(\alpha + x, \beta + 1)$ |
| Gamma $\mathcal{G}(\nu, \theta)$ | Gamma $\mathcal{G}(\alpha, \beta)$ | $\mathcal{G}(\alpha + \nu, \beta + x)$ |
| Binomiale $\mathcal{B}(n, \theta)$ | Bêta $\mathcal{Beta}(\alpha, \beta)$ | $\mathcal{Beta}(\alpha + x, \beta + n - x)$ |
| Binomiale négative $\mathcal{Neg}(m, \theta)$ | Bêta $\mathcal{Beta}(\alpha, \beta)$ | $\mathcal{Beta}(\alpha + m, \beta + x)$ |
| Multinomiale $\mathcal{M}_k(\theta_1, \dots, \theta_k)$ | Dirichlet $\mathcal{D}(\alpha_1, \dots, \alpha_k)$ | $\mathcal{D}(\alpha_1 + x_1, \dots, \alpha_k + x_k)$ |
| Normale $\mathcal{N}(\mu, \frac{1}{\theta})$ | Gamma $\mathcal{G}(\alpha, \beta)$ | $\mathcal{G}(\alpha + 0.5, \beta + \frac{(\mu-x)^2}{2})$ |

Le tableau (1.1) représente quelques lois a priori conjuguées pour quelque familles exponentielles.

1.3 Distribution a priori non informative

Ces distributions sont conçues dans le but de faire de l'analyse bayésienne lorsqu'il y a absence d'information a priori sur le paramètre d'intérêt ou dans le cas où il est difficile de traduire en terme de loi a priori l'information disponible sur les paramètres en une loi de probabilité.

Ce choix repose sur l'équiprobabilité des valeurs possibles du paramètre θ sur son domaine de définition.

Dans le cas où l'espace des paramètres est discret et fini, par exemple, $\theta = (\theta_1, \theta_2, \dots, \theta_n)$.

Alors, la distribution a priori non informative est

$$p(\theta_i) = \frac{1}{n}, \quad \forall i = \overline{1, n}.$$

Aussi, dans le cas où nous avons un espace du paramètre continue et borné, comme par exemple $\Theta = [a, b]$, $-\infty < a < b < \infty$, alors la distribution a priori est uniforme

$$p(\theta) = \frac{1}{b-a}; \quad a < \theta < b$$

donc il s'agit d'une distribution non informative pour θ .

Nous décrirons, dans ce qui suit, quelques techniques (les plus importantes) dans la construction de la loi a priori non informative.

1.3.1 Lois a priori invariantes

La loi a priori invariante s'exprime lorsque l'absence de l'information a priori se traduit par une propriété d'invariance par rapport à une famille de transformation de Θ . Deux exemples introductifs sont présentés ci-dessous.

Exemple 1.5. *La famille des lois invariantes par translation, c'est-à-dire, la loi a priori doit vérifier*

$$p(\theta) = p(\theta - \theta_0), \text{ pour tout } \theta_0 \in \Theta.$$

Par conséquent, la loi a priori solution de cette équation est la loi uniforme sur Θ définie par

$$p(\theta) = c$$

où c est une constante et θ paramètre de position.

Exemple 1.6. *La famille des lois invariantes par changement d'échelle ayant la forme*

$$p(\theta) = \frac{1}{c} p\left(\frac{\theta}{c}\right)$$

où c est une constante strictement positive et θ est dit paramètre d'échelle .

La loi a priori solution de cette équation s'écrit alors

$$p(\theta) = \frac{c}{\theta}.$$

L'approche invariante n'est que partiellement satisfaisante, car elle implique la référence à une structure d'invariance, qui peut être parfois choisie de plusieurs manières, ne pas exister, ou être sans intérêt pour le décideur.

1.3.2 Lois a priori de Jeffreys

La lois a priori non informative de Jeffreys est la généralisation de la loi invariante.

En effet Jeffreys (1961), propose une approche plus globale basée sur la matrice d'information de Fisher, qui évite de prendre en compte une structure invariante particulière de la lois des observations.

- Dans le cas unidimensionnel, les lois a priori de Jeffreys sont définies par

$$p(\theta) \propto [I(\theta)]^{1/2}$$

où $I(\theta)$ est la matrice d'information de Fisher

$$I(\theta) = E_{\theta} \left[\left(\frac{\partial \log f(x|\theta)}{\partial \theta} \right)^2 \right] \quad (1.3)$$

sous certaines conditions de régularité l'information de Fisher, est égale à

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

- Dans le cas multidimensionnel, $\theta \in \mathbb{R}^k$, la matrice d'information de Fisher s'obtient par généralisation de (1.3), c'est à dire,

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

et la lois non informative de Jeffreys est alors :

$$p(\theta) \propto |I(\theta)|^{1/2}$$

La technique de Jeffreys fournit une des meilleures techniques pour construire une loi a priori non informative et elle permet bien souvent de retrouver les estimateurs classiques du type maximum de vraisemblance dans des cas unidimensionnels mais l'extension multidimensionnelles peut parfois conduire à des incohérences (paradoxe de marginalisation) cette technique implique une réduction de la dimension de l'espace de paramètres à estimer .

Exemple 1.7. Soit $X \sim \mathcal{N}(\mu, \sigma^2)$ avec $\theta = (\mu, \sigma^2)$ inconnu.

On a

$$f(x|\mu, \sigma^2) \propto \frac{1}{\sigma} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\};$$

alors

$$\log f(x|\mu, \sigma^2) \propto \frac{-1}{2} \log \sigma^2 - \frac{(x - \mu)^2}{2\sigma^2};$$

donc

$$\begin{cases} \frac{\partial}{\partial \mu} \log f(x|\mu, \sigma^2) = \frac{(x-\mu)}{\sigma^2}; \\ \frac{\partial}{\partial \sigma^2} \log f(x|\mu, \sigma^2) = -\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2(\sigma^2)^2}. \end{cases}$$

et

$$\begin{cases} \frac{\partial^2}{\partial \mu^2} \log f(x|\mu, \sigma^2) = -\frac{1}{\sigma^2}; \\ \frac{\partial^2}{\partial (\sigma^2)^2} \log f(x|\mu, \sigma^2) = \frac{1}{2(\sigma^2)^2} - \frac{(x-\mu)^2}{(\sigma^2)^3}; \\ \frac{\partial^2}{\partial \mu \partial \sigma^2} \log f(x|\mu, \sigma^2) = -\frac{(x-\mu)}{2(\sigma^2)^2}. \end{cases}$$

La matrice d'information de Fisher s'obtient en calculant, l'espérance mathématique des dérivées secondes.

On a $E(X - \theta) = 0$ et $E[(X - \theta)^2] = \sigma^2$

Donc

$$I(\mu, \sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^2} \end{pmatrix}$$

et la loi de Jeffreys est

$$p(\theta) \propto \frac{1}{\sigma^2}.$$

1.3.3 Lois a priori de référence

La loi a priori de référence est introduite par Bernardo(1979) suite à la difficulté rencontrée dans l'utilisation de l'approche de Jeffreys dans le cas multidimensionnel.

Quand $X \sim f(X|\theta)$ et $\theta = (\theta_1, \theta_2)$, où θ_1 est le paramètre d'intérêt et θ_2 est le paramètre de nuisance. La stratégie introduite par Bernardo est la suivante :

Pour θ_1 fixé, on détermine la densité conditionnelle $p(\theta_2|\theta_1)$ comme la loi de Jeffreys puis en calculant $p(\theta_1)$ qui est la loi de Jeffreys associée à la loi marginale

$$\tilde{f}(X|\theta_1) = \int f(X|\theta_1, \theta_2)p(\theta_2|\theta_1)d\theta_2.$$

La loi de référence de θ est le produit des deux lois, c'est à dire :

$$p(\theta_1, \theta_2) = p(\theta_2|\theta_1)p(\theta_1).$$

Cette manière de faire peut être généraliser si $\theta = (\theta_1, \dots, \theta_n)$ où l'on a ordonné, sans perte de généralité, les θ_i par intérêt croissant.

1.4 Lois a priori impropres

Lorsque le paramètre θ peut être traité comme une variable aléatoire avec une distribution de probabilité p connue, nous avons vu que le théorème de Bayes est la base

de l'inférence bayésienne, car il donne la distribution a posteriori.

Cependant, dans de nombreux cas, la distribution a priori est déterminée par des critères subjectifs ou théoriques qui conduisent à une mesure σ -finie sur l'espace des paramètres Θ , c'est à dire, une mesure p telle que :

$$\int_{\Theta} p(\theta) d\theta = +\infty.$$

On dit que la distribution a priori est impropre (ou généralisée).

La difficulté pratique dans l'utilisation des lois impropres est de vérifier la condition d'intégrabilité

$$\int_{\Theta} f(x|\theta)p(\theta)d\theta < \infty.$$

L'estimateur de Bayes généralisé de θ qui vérifie la condition d'intégrabilité est la moyenne de la loi a posteriori.

Exemple 1.8. Soit $X|\theta \sim \mathcal{N}(\theta, 1)$ et $p(\theta) = 1$.

On a

$$\int_{\Theta} p(\theta) d\theta = +\infty;$$

d'où la loi a priori est impropre.

Et

$$\int_{\Theta} f(x|\theta)p(\theta)d\theta = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\theta)^2\right\} d\theta = 1 < \infty.$$

Donc l'estimateur de Bayes généralisé est

$$\hat{\theta}_B = E(\theta|x).$$

La loi a posteriori

$$\begin{aligned} p(\theta|x) &= \frac{f(x|\theta)p(\theta)}{\int_{\Theta} f(x|\theta)p(\theta)d\theta} \\ &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(\theta-x)^2\right\}; \end{aligned}$$

donc

$$\theta|X \sim \mathcal{N}(x, 1);$$

alors

$$\hat{\theta}_B = X \in \mathbb{R}.$$

1.5 Modèle hiérarchique

Pour des raisons liées à la modélisation des observations ou à la décomposition de l'information a priori, il peut arriver que le modèle statistique bayésien soit hiérarchique, c'est à dire mettre en jeu plusieurs niveaux de distributions a priori conditionnelles.

Définition 1.5. *Un modèle bayésien hiérarchique est un modèle statistique bayésien $(f(x|\theta), p(\theta))$, dans lequel la loi a priori $p(\theta)$ est décomposée en plusieurs lois conditionnelles*

$$p_1(\theta|\theta_1), p_2(\theta_1|\theta_2), \dots, p_n(\theta_n|\theta_n)$$

et une loi marginale $p_{n+1}(\theta_n)$ telle que

$$p(\theta) = \int_{\Theta_1 \times \dots \times \Theta_n} p_1(\theta|\theta_1) p_2(\theta_1|\theta_2) \dots p_{n+1}(\theta_n) d\theta_1 \dots d\theta_n.$$

Les paramètres $(\theta_i)_{i=1..n}$ sont appelés hyperparamètres.

Remarquons qu'un modèle bayésien hiérarchique

$$x \sim f(x|\theta), \theta \sim p_1(\theta|\theta_1), \theta_1 \sim p_2(\theta_1|\theta_2), \dots, \theta_n \sim p_{n+1}(\theta_n)$$

est un cas particulier du modèle bayésien usuel :

$$x \sim f(x|\theta) \text{ et } \theta \sim p(\theta)$$

avec

$$p(\theta) = \int_{\Theta_1 \times \dots \times \Theta_n} p_1(\theta|\theta_1) p_2(\theta_1|\theta_2) \dots p_{n+1}(\theta_n) d\theta_1 \dots d\theta_n.$$

Cette structure hiérarchiques est trouvée également, dans des modèles statistiques classiques qui sont bénéficié des propriétés générales d'optimalité de l'approche bayésienne avec quelques avantages additionnels liés à la décomposition de la loi a priori.

1.5.1 Décompositions conditionnelles

Une caractéristique particulièrement intéressante des modèles hiérarchiques est que le conditionnement est possible à tous les niveaux et cette liberté, dans la décomposition de la loi a posteriori, compense l'augmentation apparente de complexité de la structure.

Par exemple, si

$$x \sim f(x|\theta), \theta|\theta_1 \sim p_1(\theta|\theta_1), \theta_1 \sim p_2(\theta_1),$$

nous avons le résultat suivant.

Lemme 1.1. *La loi a posteriori de θ est*

$$p(\theta|x) = \int_{\Theta_1} p_1(\theta|\theta_1, x) p(\theta_1|x) d\theta_1,$$

avec

$$p_1(\theta|\theta_1, x) = \frac{f(x|\theta_1) p_1(\theta|\theta_1)}{f_1(x|\theta_1)}$$

où

$$f_1(x|\theta_1) = \int_{\Theta} f(x|\theta) p_1(\theta|\theta_1) d\theta$$

et

$$p(\theta_1|x) = \frac{f_1(x|\theta_1) p_2(\theta_1)}{f(x)}$$

où

$$f(x) = \int_{\Theta_1} f_1(x|\theta_1) p_2(\theta_1) d\theta_1$$

Preuve

En remplaçant $p_1(\theta|\theta_1, x)$ et $p(\theta_1|x)$ par leur expression sous l'intégrale, il vient :

$$\begin{aligned} p(\theta|x) &= \int_{\Theta_1} \frac{f(x|\theta_1) p_1(\theta|\theta_1)}{f_1(x|\theta_1)} \frac{f_1(x|\theta_1) p_2(\theta_1)}{f(x)} d\theta_1 \\ &= \int_{\Theta_1} \frac{f(x|\theta) p_1(\theta|\theta_1) p_2(\theta_1)}{f(x)} d\theta_1 \\ &= \frac{f(x|\theta)}{f(x)} \int_{\Theta_1} p_1(\theta|\theta_1) p_2(\theta_1) d\theta_1 \\ &= \frac{f(x|\theta) p(\theta)}{f(x)}. \end{aligned}$$

1.6 Incorporation de l'information a priori en pratique

On considère le cas d'un paramètre réel θ . On envisage trois situations :

- $\theta \in [0, 1]$,
- $\theta \in [0, +\infty[$,
- $\theta \in]-\infty, +\infty[$,

On suppose que l'information a priori fournie par l'expert consiste en une estimation ponctuelle θ^* de θ et un intervalle $I^* = [\alpha^*, \beta^*]$ contenant θ^* , tel que $P(\theta^* \in I^*)$ soit élevée (typiquement 0.95).

Situation 1 : $\theta \in [0, 1]$

On choisit dans ce cas une loi a priori de type bêta,

$$\theta \sim p(\theta) = \text{Beta}(a, b).$$

On procède à la reparamétrisation comme suit :

$$\begin{cases} a = \lambda\mu \\ b = \lambda(1 - \mu), \text{ où } \mu = E(\theta) \end{cases} \implies \begin{cases} \mu = E(\theta) = \frac{a}{a+b} \\ \lambda = a + b \end{cases}$$

On a

$$\text{Var}(\theta) = \frac{ab}{(a+b)^2(a+b+1)} = \frac{\mu(1-\mu)}{1+\lambda}$$

On obtient qu'à fixé, λ est proportionnel à la variance et donc s'interprète comme la précision de l'information a priori. Autrement dit, plus λ est grand plus $\text{Var}(\lambda)$ est petit et plus l'information a priori est précise. Inversement, plus est petit plus $\text{Var}(\lambda)$ est grand et moins l'information a priori est précise.

En pratique, on pose $E(\theta) = \theta^* = \mu$. D'où,

$$\theta \sim p(\theta) = \text{Beta}(\lambda\theta^*, \lambda(1 - \theta^*)).$$

λ est déterminé tel que :

$$P(\theta^* \in I^*) = \int_{I^*} p(\theta) d\theta = 0.95.$$

Pour $E(\theta) = \theta^* = 0.2$ et $I^* = [0.05, 0.4]$, tel que, $P(\theta \in I^*) = 0.95$, on obtient sur le logiciel R $\lambda = 17.5$.

D'où la loi a priori, $p(\theta) = \text{Beta}(3.5, 14)$.

Situation 2 : $\theta \in [0, +\infty[$

On choisit, dans ce cas, une loi Gamma $G(a, b)$ et on procède à la reparamétrisation comme suit :

$$\begin{cases} a = \lambda\mu \\ b = \lambda \end{cases} \implies \begin{cases} \mu = \frac{a}{b} \\ \lambda = b \end{cases} \text{ avec } \mu = E(\theta)$$

On a $E(\theta) = \mu = \frac{a}{b}$ et $Var(\theta) = \frac{a}{b^2} = \frac{\mu}{\lambda}$.

D'où pour μ fixé, λ est inversement proportionnel à la variance. Autrement dit, λ plus est grand plus $Var(\theta)$ est petit et plus l'information a priori est précise. Inversement, plus λ est petit plus $Var(\theta)$ est grand et moins l'information a priori est précise.

En pratique, on pose $E(\theta) = \theta^* = \mu$. On obtient

$$\theta \sim G(\lambda\theta^*, \lambda)$$

où λ est déterminé tel que

$$P(\theta^* \in I^*) = \int_{I^*} p(\theta) d\theta = 0.95.$$

Situation 3 : $\theta \in \mathbb{R} =]-\infty, +\infty[$

Dans ce cas, on choisit comme loi a priori la loi normale $N(a, b)$ avec $a = E(\theta)$ et $b = Var(\theta)$.

1.7 Le poids de l'a priori dans la statistique bayésienne

Examinons cette question sur un exemple pour comprendre comment l'information a priori et l'information contenue dans les observations se combinent l'une à l'autre pour produire la réponse bayésienne.

On se donne le modèle bayésien suivant : $X_i|\theta \sim \mathcal{B}(\theta)$ et $\theta \sim \mathcal{Beta}(a, b)$.

Il est commode de re-paramétrer la loi Bêta comme suit :

on pose $a = \lambda\mu$ et $b = \lambda(1 - \mu)$ où $\mu = \frac{a}{a+b}$ et $\lambda = a + b$ et de travailler avec $E(\theta|X)$.

Calculons $E(\theta|X)$:

• On a :

$$\begin{aligned} f(x|\theta) &= \prod_{i=1}^n \mathbb{P}(X = x_i|\theta) \\ &= \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} \\ &= \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{\sum_{i=1}^n (1-x_i)} \\ &= \theta^s (1 - \theta)^{n-s} \end{aligned}$$

où $s = \sum_{i=1}^n x_i$.

• Comme $\theta \sim \text{Beta}(a, b)$, on a :

$$p(\theta) = \frac{1}{B(a, b)} \theta^{a-1} (1 - \theta)^{b-1} \mathbb{1}_{[0,1]}(\theta).$$

Alors, la loi a posteriori est :

$$\begin{aligned} p(\theta|x) &= \frac{f(x|\theta)p(\theta)}{\int_{\Theta} f(x|\theta)p(\theta)d\theta} \\ &= \frac{\theta^s (1 - \theta)^{n-s} \frac{1}{B(a,b)} \theta^{a-1} (1 - \theta)^{b-1}}{f(x)} \mathbb{1}_{[0,1]}(\theta) \\ &= \frac{\frac{1}{B(a,b)} \theta^{a+s-1} (1 - \theta)^{n+b-s-1}}{f(x)} \mathbb{1}_{[0,1]}(\theta) \end{aligned}$$

où

$$\begin{aligned} f(x) &= \int_0^1 \frac{1}{B(a, b)} \theta^{a+s-1} (1 - \theta)^{n+b-s-1} d\theta \\ &= \frac{1}{B(a, b)} \int_0^1 \theta^{\alpha-1} (1 - \theta)^{\beta-1} d\theta \\ &= \frac{B(\alpha, \beta)}{B(a, b)}; \end{aligned}$$

et $\alpha = a + s$, $\beta = b + n - s$, d'où :

$$p(\theta|x) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1} \mathbb{1}_{[0,1]}(\theta).$$

Par conséquent :

$$\theta|x \sim \text{Be}(a + \sum_{i=1}^n x_i, b + n - \sum_{i=1}^n x_i).$$

Donc

$$\begin{aligned}
 E(\theta|x) &= \frac{a + \sum_{i=1}^n x_i}{a + b + n} \\
 &= \frac{a + n\bar{x}}{\lambda + n} \\
 &= \frac{\lambda\mu}{\lambda + n} + \frac{n\bar{x}}{\lambda + n} \\
 &= \frac{\lambda}{\lambda + n} E(\theta) + \frac{n}{\lambda + n} \bar{x},
 \end{aligned}$$

L'estimation bayésienne de θ apparaît donc comme une forme linéaire de la moyenne pondérée de \bar{x} (c'est à dire de l'estimation de θ 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 $E(\theta)$ est λ qui s'interprète comme la précision de l'a priori (le dénominateur commun $\lambda + n$ a été sciemment omis, car seul compte en fait le numérateur). Géométriquement, $E[\theta|x]$ est le barycentre des points de coordonnées $E[\theta]$ et \bar{x} , affectés respectivement des coefficients $\frac{\lambda}{\lambda+n}$ et $\frac{n}{\lambda+n}$.

- Si $\lambda = n$, l'estimation bayésienne de θ se situe exactement au milieu de l'intervalle $[E(\theta); \bar{x}]$.
- Si $\lambda > n$ cette estimation est plus proche de $E(\theta)$ que de \bar{x} .
- Si $\lambda < n$ 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 : $\lambda \rightarrow 0$ et $\lambda \rightarrow \infty$ (la taille n de l'échantillon étant fixé, ainsi que $\mu = E(\theta)$).

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 deuxième cas, le poids des données est nul, et $E(\theta|x) \rightarrow E(\theta)$ qui ne dépend plus de x .

Le tableau ci-dessous résume la situation.

| | $\lambda \rightarrow 0$ | $\lambda \rightarrow \infty$ |
|-----------------------|---------------------------|-----------------------------------|
| Loi a priori | Loi de Haldane | Loi concentrée en μ |
| Interprétation | Situation Non informative | Situation Extrêmement informative |
| Estimation Bayésienne | \bar{x} | $E(\theta)$ |

Il est également intéressant de regarder ce que devient $E(\theta|x)$ quand $n \rightarrow \infty$, λ et μ étant fixés; dans ce cas, le poids de l'a priori devient négligeable, et la réponse bayésienne coïncide avec la réponse classique, c'est à dire \bar{x} (l'estimation de θ par maximum de vraisemblance).

1.8 Étude de la sensibilité de la réponse bayésienne à la loi a priori

Le choix de la loi a priori comporte une certaine part d'arbitraire, et ce, à deux niveaux :

Le premier niveau réside dans le choix de la famille de probabilités retenue pour la loi a priori (une loi Bêta pour un paramètre dans $[0; 1]$, une loi Gamma pour un paramètre dans $[0; \infty[$, etc).

Le deuxième réside dans le choix des valeurs numériques communiquées par l'expert (typiquement θ^*) pour déterminer les hyper-paramètres de la loi a priori.

Il convient par conséquent d'examiner dans quelle mesure ces choix affectent l'estimation bayésienne de θ .

Dans la pratique, on se contente d'examiner, dans quelle mesure, une petite perturbation des hyper paramètres modifie l'estimation bayésienne de θ . C'est ce qu'on appelle : faire l'étude de la sensibilité de la réponse bayésienne à la loi a priori.

Pour conclure, disons qu'en pratique, on comparera toujours l'estimateur de Bayes obtenu avec la loi a priori informative, avec l'estimateur de Bayes correspondant à une loi a priori non informative; et ce, afin d'apprécier le poids de l'a priori dans la réponse bayésienne informative.

1.9 Méthode de Monte Carlo

Il existe des méthodes numériques d'intégration et d'optimisation qui peuvent être intéressantes pour résoudre les problèmes d'implémentation pratique de l'analyse bayésienne (voir Fletcher, 1987). Une autre alternative est donnée par les méthodes de simulation. On cite les méthode de simulation de Monte-carlo et les méthodes MCMC, développées ci-dessous.

1.9.1 Principe de la méthode

Soit $\theta \in \mathbb{R}$ de densité $p(\theta)$. On souhaite calculer $E(g(\theta))$ où $g(\theta)$ est un réel.

Le principe de la méthode de simulation de Monte-Carlo est de simuler une suite de v.a. $\theta_1, \theta_2, \dots, \theta_L$ i.i.d. selon la loi de θ puis approcher $E(g(\theta))$ à l'aide de la loi forte des grands nombres par

$$\frac{1}{L} \sum_{i=1}^L g(\theta_i), \text{ lorsque } L \rightarrow \infty.$$

1.9.2 Contexte bayésien

On veut calculer $E(g(\theta|x))$, où g est une fonction d'intérêt.

On a :

$$E(g(\theta|x)) = \int_{\Theta} g(\theta) p(\theta|x) d\theta,$$

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

D'où,

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

Ce qui revient à approcher l'intégrale $\int_{\Theta} g(\theta) f(x|\theta) p(\theta|x) d\theta$ par la méthode de simulation de Monte Carlo.

– On peut simuler $\theta_1, \theta_2, \dots, \theta_L$ selon la fonction d'importance $p(\theta)$. L'intégrale peut alors être approchée par :

$$\frac{1}{L} \sum_{i=1}^L g(\theta_i) f(x|\theta_i), \quad L \rightarrow \infty.$$

– On peut simuler $\theta_1, \theta_2, \dots, \theta_L$ selon la fonction d'importance $f(x|\theta)p(\theta|x)$. L'intégrale peut alors être approchée par :

$$\frac{1}{L} \sum_{i=1}^L g(\theta_i), \quad L \rightarrow \infty.$$

– On peut simuler $\theta_1, \theta_2, \dots, \theta_L$ selon la fonction d'importance $p(\theta|x)$. La quantité $E(g(\theta|x))$ peut alors être approchée, à l'aide de la loi forte des grands nombres, par :

$$\frac{1}{L} \sum_{i=1}^L g(\theta_i), \quad L \rightarrow \infty.$$

De plus, si la variance a posteriori $Var(g(\theta|x))$ est finie le Théorème Central-Limite (TCL) s'applique à la moyenne $E(g(\theta|x)) \simeq \frac{1}{L} \sum_{i=1}^L g(\theta_i)$ qui est asymptotiquement normale de variance $\frac{Var(g(\theta|x))}{L}$. Des régions de confiance peuvent alors être construites à partir de cette approximation. L'ordre de grandeur est de $\frac{1}{\sqrt{L}}$, quelque soit la dimension du problème, contrairement aux méthodes numériques.

1.9.3 Méthode de Monte Carlo par Chaîne de Markov (Méthode MCMC)

C'est une méthode de simulation très puissante. Elle propose une structure universelle de simulation facilement implémentable quelque soit la complexité et la dimension de la densité à simuler. Elle permet d'approcher la génération de v.a. de loi a posteriori $p(\theta|x)$ lorsque celle-ci ne peut pas être simulée directement.

Plusieurs types d'inférences statistiques sont possibles à partir d'échantillons produits par des méthodes MCMC comme l'estimation des moments, de modes ou de densités marginales, l'inférence prédictive, l'analyse de sensibilité,...(Chib and Greenberg, 1995).

Principe de la méthode

Le principe de la méthode MCMC est de générer une chaîne de Markov (CM) : $(\theta^{(l)}, l \geq 0)$, avec $dim\theta^{(l)} = dim\theta$, qui suit une loi asymptotiquement distribuée selon la loi a posteriori de θ , puis d'approcher la moyenne a posteriori à l'aide du théorème ergodique,

$$\frac{1}{L} \sum_{i=1}^L g(\theta^{(l)}) \xrightarrow{L \rightarrow +\infty} E(g(\theta|x))$$

Il existe deux algorithmes pour produire une telle chaîne de markov : "Algorithme de Gibbs" et "Algorithme de Metropolis-Hastings" (MH). L'algorithme de Metropolis-Hastings

est considéré comme l'algorithme de base d'une grande partie des algorithmes MCMC. L'échantillonneur de Gibbs est considéré comme un cas particulier de l'algorithme (MH) mais a des motivations méthodologiques et historiques fondamentalement différentes.

Algorithme de Metropolis-Hastings

Les méthodes (MH) ont l'avantage de ne requérir qu'une connaissance limitée de la densité à simuler en autorisant une grande liberté dans leur implémentation. Cet algorithme a été proposé par Metropolis en 1953 (Metropolis, 1953) pour le traitement de problèmes de physique, puis généralisé par Hastings en 1970 (Hastings, 1970) pour simuler des systèmes complexes. Pour une revue complète concernant cette méthode consulter Chib and Greenberg (1995) et Robert (1996).

Cet algorithme permet de fabriquer une chaîne de Markov dont on s'est donné la loi stationnaire. Sa mise en oeuvre présente l'avantage de ne nécessiter la définition de p qu'à une constante près.

Pour une densité $p(\theta|x)$ connue à une constante près et une loi conditionnelle $q(\cdot|\cdot)$, appelée loi de proposition choisie symétrique au sens où $q(\theta'|\theta) = q(\theta|\theta')$, l'algorithme génère une CM comme suit :

Partant d'un vecteur initial $\theta(0)$ arbitraire, à la $p^{\text{ème}}$ étape, disposant du vecteur $\theta^{(p)} = (\theta_1^{(p)}, \dots, \theta_m^{(p)})$.

– Générer $\tilde{\theta}$ selon la loi $q(\tilde{\theta}|\theta^{(p-1)})$ (θ est choisi comme paramètre de position),

– Calculer la probabilité d'acceptation $\alpha = \min \left\{ 1, \frac{p(\theta|x)q(\theta^{(p-1)}|\tilde{\theta})}{p(\theta^{(p-1)}|x)q(\tilde{\theta}|\theta^{(p-1)})} \right\}$.

– Prendre

$$\theta^{(p)} = \begin{cases} \tilde{\theta}, & \text{avec probabilité } \alpha \\ \theta^{(p-1)}, & \text{avec probabilité } 1 - \alpha \end{cases}$$

Remarque 1.2. On peut autoriser un nombre infini de lois de proposition, produisant toutes une CM convergentes vers la loi d'intérêt.

Échantillonnage de Gibbs

L'échantillonneur de Gibbs introduit par Geman et al. (1984) dans le cadre de la restauration d'images et a été ensuite généralisé par Tanner et al. (1987) et par Gelfand

et al. (1990) en statistique appliquée. Le principe repose sur une décomposition d'une loi $f(x)$ suivant des lois conditionnelles.

Considérons le cas élémentaire $f(x, y)$. On suppose que $f(x|y)$ et $f(y|x)$ sont disponibles. On peut alors générer une séquence de Gibbs de la manière suivante : Partant d'une valeur x_0 , on génère y_0 suivant $p(\cdot|x_0)$, puis x_1 suivant $p(\cdot|y_0)$, puis y_1 suivant $p(\cdot|x_1)$ et ainsi de suite.

Après M itérations de ce schéma, il vient une séquence $(x_0, y_0, x_1, y_1, \dots, x_M, y_M)$. Pour M assez grand, x_M est une réalisation de X .

Exemple 1.9. *Considérons une loi jointe de la forme :*

$$f(x, y) \propto C_n^x y^{x+\alpha-1} (1-y)^{n-x+\beta-1}, \quad x = 0, \dots, n, \quad 0 \leq y \leq 1.$$

On remarque que l'on peut proposer une loi binomiale de paramètres (n, y) pour $f(x|y)$ et une loi bêta de paramètres $(x + \alpha, n - x + \beta)$ pour $f(y|x)$.

Bien que dans le cas de cet exemple la loi marginale $f(x)$ est accessible,

$$f(x) \propto \int_0^1 f(x, y) dy \propto C_n^x \frac{\Gamma(x + \alpha) \Gamma(n - x + \beta)}{\Gamma(x + \beta + n)},$$

on peut toujours appliquer l'algorithme de Gibbs pour obtenir des réalisations de $f(x)$ en simulant alternativement une réalisation x^ d'une binomiale de paramètres (n, y^*) où y^* est la valeur courante de y obtenu à l'étape précédente, puis une nouvelle réalisation x d'une loi bêta de paramètre $(x^* + \alpha; n - x^*)$.*

Dans le cadre bayésien, l'algorithme de Gibbs va permettre d'obtenir une réalisation du paramètre $\theta = (\theta_1, \dots, \theta_L)$ suivant la loi a posteriori $p(\theta|x)$ dès que l'on est capable d'exprimer les lois conditionnelles $p(\theta_i|\theta_j; x)$ $j \neq i$.

L'échantillonnage de Gibbs consiste à :

Partant d'un vecteur initial $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_L^{(0)})$ où $L \geq 2$ est la dimension de $\theta^{(0)}$.

À la $(p + 1)^{\text{ème}}$ étape, disposant du vecteur $\theta^{(p)} = (\theta_1^{(p)}, \dots, \theta_L^{(p)})$,

- Simuler $\theta_1^{(p+1)} \sim p(\theta_1|\theta_2^{(p)}, \dots, \theta_L^{(p)}; x)$,

- Simuler $\theta_2^{(p+1)} \sim p(\theta_2|\theta_1^{(p+1)}, \theta_3^{(p)}, \dots, \theta_L^{(p)}; x)$,

.

TABLE 1.2 – Nombre de défaillances et temps d’observation pour dix pompes d’une centrale nucléaire

(Source : Graver and O’Muirheartaigh, 1987)

| | | | | | | | | | | |
|-------------|----|----|----|-----|---|----|-----|-----|---|----|
| Pompe | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Défaillance | 5 | 1 | 5 | 14 | 3 | 19 | 1 | 1 | 4 | 22 |
| Temps | 94 | 16 | 63 | 126 | 5 | 31 | 105 | 105 | 2 | 10 |

- Simuler $\theta_L^{(p+1)} \sim p\left(\theta_L | \theta_1^{(p+1)}, \dots, \theta_{L-1}^{(p+1)}; x\right)$.

Les itérations successives de cet algorithme génèrent successivement les états d’une chaîne de Markov $\{\theta^{(p)}, p > 0\}$ à valeurs dans $\mathbb{N}^{\otimes L}$.

Cette chaîne admet une mesure invariante qui est la loi a posteriori. Pour un nombre d’itérations suffisamment grand, le vecteur θ obtenu peut donc être considéré comme étant une réalisation de la loi a posteriori.

1.10 Exemple

L’objectif de cette section est d’appliquer les méthodes MCMC, pour obtenir les estimateurs bayésiens des paramètres inconnus et voir l’importance de choix de la loi a priori. Le problème été étudié par (Robert et Casella, 10.17). Nous allons démontrer que que choix de l’a priori joue un rôle très important dans l’estimation Bayésienne des paramètres inconnue et l’avantage de cette approche par rapport à la méthode classique.

Problématique

Le problème qui se pose décrit les défaillances répétées de dix pompes d’une centrale nucléaire; les données figurent dans le tableau(1.2).

Le modèle fait l’hypothèse que les défaillances de la pompe i suivent un processus de Poisson de paramètres λ_i ($1 \leq i \leq 10$). Pour un temps d’observation t_i , le nombre de défaillances x_i suit donc une distribution de Poisson $\mathcal{P}(\lambda_i t_i)$.

On veut effectuer une estimation bayésienne sur les paramètres λ_i en utilisant deux lois a priori différentes : une loi conjuguée et une loi non informative (Jeffreys). Par la suite

nous visons à comparer les résultats obtenus avec les estimateurs de maximum de vraisemblance.

- La vraisemblance est :

$$\begin{aligned} f(x|\lambda_i) &= \prod_{i=1}^{10} f(x_i|\lambda_i) \\ &= \prod_{i=1}^{10} \frac{e^{-\lambda_i t_i} (\lambda_i t_i)^{x_i}}{x_i!} \\ &= \left(\prod_{i=1}^{10} \frac{1}{x_i!} \right) (\exp\{-\sum_{i=1}^{10} \lambda_i t_i\}) \left(\prod_{i=1}^{10} (\lambda_i t_i)^{x_i} \right). \end{aligned}$$

1.10.1 Méthode classique

- Calculons les estimateurs du maximum de vraisemblance $\hat{\lambda}_i (1 \leq i \leq 10)$

$$\begin{aligned} \log f(x|\lambda_i) &= \log\left[\left(\prod_{i=1}^{10} \frac{1}{x_i!}\right) (\exp\{-\sum_{i=1}^{10} \lambda_i t_i\}) \left(\prod_{i=1}^{10} (\lambda_i t_i)^{x_i}\right)\right] \\ &= cte - \sum_{i=1}^{10} \lambda_i t_i + \sum_{i=1}^{10} \log(\lambda_i t_i)^{x_i} \\ &= cte - \sum_{i=1}^{10} \lambda_i t_i + \sum_{i=1}^{10} x_i \log(\lambda_i t_i). \end{aligned}$$

$$\frac{\partial \log f(X|\lambda_i)}{\partial \lambda_1} = 0 \Leftrightarrow \hat{\lambda}_1 = \frac{x_1}{t_1}.$$

•

•

•

$$\frac{\partial \log f(x|\lambda_i)}{\partial \lambda_{10}} = 0 \Leftrightarrow \hat{\lambda}_{10} = \frac{x_{10}}{t_{10}}.$$

Les estimateurs du maximum de vraisemblance sont donnés dans le tableau (1.3).

1.10.2 Méthodes bayésiennes

1. Supposons que les λ_i suivent une loi a priori conjuguée $\mathcal{G}(\alpha, \beta)$, avec, $\alpha = 1.8$ et β inconnu.

Donc, le modèle contient 11 paramètres inconnus (dix λ_i plus β).

Pour β , on propose la loi a priori $\mathcal{G}(\gamma, \delta)$, avec $\gamma = 0.01$ et $\delta = 1$.

Nous avons :

$$X \sim \mathcal{P}(\lambda_i t_i) , \quad i = \overline{1,10}.$$

$$\lambda_i \sim \mathcal{G}(\alpha, \beta) , \quad i = \overline{1,10}.$$

$$\beta \sim \mathcal{G}(\gamma, \delta).$$

• En appliquant le théorème de Bayes, on obtient la distribution a posteriori :

$$\begin{aligned} p(\lambda_1, \dots, \lambda_{10}, \beta | x_1, \dots, x_{10}, t_1, \dots, t_{10}) &\propto f(x | \lambda_i) p(\lambda_i) p(\beta) \\ &\propto \left(\prod_{i=1}^{10} \frac{e^{-\lambda_i t_i} (\lambda_i t_i)^{x_i}}{x_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_i^{\alpha-1} e^{-\lambda_i \beta} \right) \frac{\delta^\gamma}{\Gamma(\delta)} \beta^{\gamma-1} e^{-\beta \delta} \\ &\propto \left(\prod_{i=1}^{10} e^{-\lambda_i t_i} (\lambda_i t_i)^{x_i} \beta^\alpha \lambda_i^{\alpha-1} e^{-\lambda_i \beta} \right) \beta^{\gamma-1} e^{-\beta \delta} \\ &\propto \left(\prod_{i=1}^{10} \lambda_i^{x_i + \alpha - 1} e^{-(t_i + \beta) \lambda_i} \right) \beta^{10\alpha} \beta^{\gamma-1} e^{-\beta \delta} \\ &\propto \left(\prod_{i=1}^{10} \lambda_i^{x_i + \alpha - 1} e^{-(t_i + \beta) \lambda_i} \right) \beta^{10\alpha + \gamma - 1} e^{-\beta \delta}. \end{aligned}$$

Il est clair qu'il n'est pas facile de calculer cette loi a posteriori, pour cela nous avons utilisé les méthodes MCMC.

Dans ce cas, on utilise l'échantillonneur de Gibbs pour simuler.

• Déterminons les distributions conditionnelles : elles s'obtiennent en éliminant tous les termes qui ne dépendent pas du paramètre.

$$p(\lambda_i | \beta, t_i, x_i) \propto \prod_{i=1}^{10} (\lambda_i^{x_i + \alpha - 1} e^{-(t_i + \beta) \lambda_i})$$

$$\begin{aligned} p(\beta | \lambda_i, t_i, x_i) &\propto \prod_{i=1}^{10} e^{-\lambda_i \beta} \beta^{10\alpha + \gamma - 1} e^{-\beta \delta} \\ &\propto \prod_{i=1}^{10} e^{-(\lambda_i + \delta) \beta} \beta^{10\alpha + \gamma - 1} \\ &\propto \exp\left\{-\left(\sum_{i=1}^{10} \lambda_i + \delta\right)\beta\right\} \beta^{10\alpha + \gamma - 1}; \end{aligned}$$

donc

$$p(\lambda_i | \beta, t_i, x_i) \sim \mathcal{G}(x_i + \alpha, t_i + \beta), \quad i = \overline{1,10}$$

$$p(\beta | \lambda_i, t_i, x_i) \sim \mathcal{G}(10\alpha + \gamma, \delta + \sum_{i=1}^{10} \lambda_i)$$

• Les étapes de l'algorithme de Gibbs sont :

1. Donnons une valeur initiale pour β .

> $\text{beta}[1] < -1$

2. On génère $\lambda^{(1)}$ selon sa distribution conditionnelle.

> $\text{lambda}[i,] < -\text{rgamma}(\text{lenght}(x), x + \text{alpha}, t + \text{beta}[i])$

3. On génère $\beta^{(1)}$ selon sa distribution conditionnelle.

> $\text{beta}[i + 1] < -\text{rgamma}(1, 10 * \text{alpha} + \text{gamma}, \text{delta} + \text{sum}(\text{lambda}[i, 1 : \text{lenght}(x)]))$

4. On répète ces étapes en utilisant à chaque fois les valeurs de l'étapes précédente comme valeurs initiales de la nouvelle étape.

Pour un nombre de simulation $n.\text{sim} = 1000$, et pour les valeurs initiales des paramètres : $\beta = 1$, $\alpha = 1.8$, $\gamma = 0.01$ et $\delta = 1$, les résultats de la simulation de l'échantillonnage de Gibbs sont donnés dans la table (1.3).

2. Pour les λ_i suivent une loi a priori de Jeffreys

$$p(\lambda_1) \propto \frac{1}{\lambda_1}, \dots, p(\lambda_{10}) \propto \frac{1}{\lambda_{10}}.$$

On prend

$$p(\lambda_i) = \prod_{i=1}^{10} \frac{1}{\lambda_i}$$

la loi a priori de Jeffreys.

Nous avons

$$X \sim \mathcal{P}(\lambda_i t_i);$$

$$p(\lambda_i) = \prod_{i=1}^{10} \frac{1}{\lambda_i}.$$

Comme précédemment, la distribution a posteriori est

$$\begin{aligned} p(\lambda_1, \dots, \lambda_{10}, x_1, \dots, x_{10}, t_1, \dots, t_{10}) &\propto f(x|\lambda_i)p(\lambda_i) \\ &= \left(\prod_{i=1}^{10} \frac{e^{-\lambda_i t_i} (\lambda_i t_i)^{x_i}}{x_i!} \right) \prod_{i=1}^{10} \frac{1}{\lambda_i} \\ &= \prod_{i=1}^{10} e^{-\lambda_i t_i} \lambda_i^{x_i-1} \frac{t_i^{x_i}}{x_i!} \\ &\propto \exp\left\{-\sum_{i=1}^{10} \lambda_i t_i\right\} \prod_{i=1}^{10} \lambda_i^{x_i-1}. \end{aligned}$$

Dans ce cas, l'algorithme de Metropolis-Hastings sera le mieux placée pour simuler les λ_i . Comme il est déjà indiqué, l'algorithme de Metropolis-Hastings repose sur la mise en place de la distribution instrumentale. Pour cela, nous avons choisi dans ce cas la

TABLE 1.3 – Les estimateurs de maximum de vraisemblance et résultats de simulation de l'échantillonnage de Gibbs (Moyenne CG) et de Metropolis-Hastings(Moyenne NI)

| Paramètres | $\hat{\lambda}_i$ | Moyenne CG | Moyenne NI |
|----------------|-------------------|------------|------------|
| λ_1 | 0.05319149 | 0.0700731 | 0.08764225 |
| λ_2 | 0.06250000 | 0.1590098 | 0.08764225 |
| λ_3 | 0.07936508 | 0.1048920 | 0.08764225 |
| λ_4 | 0.11111111 | 0.1241629 | 0.08764225 |
| λ_5 | 2.60000000 | 2.1848445 | 0.08764225 |
| λ_6 | 0.61290323 | 0.6307686 | 0.08764225 |
| λ_7 | 1.00000000 | 1.0208779 | 0.08764225 |
| λ_8 | 1.00000000 | 1.0380995 | 0.08764225 |
| λ_9 | 2.00000000 | 1.5512630 | 0.08764225 |
| λ_{10} | 2.20000000 | 2.0278068 | 0.08764225 |

distribution $\mathcal{G}(x, \sum_{i=1}^{10} t_i)$ pour générer un candidat θ^* .

• Les étapes de l'algorithme de Metropolis-Hastings sont :

1. Déclaration de la fonction de vraisemblance et de loi instrumentale :

```
> f <- function(lambda){exp(-sum(lambda * t)) * prod(lambda)^(x - 1)};
```

```
> J <- rgamma(n.sims, x, sum(t)).
```

2. Calculons la probabilité d'acceptation :

```
> alpha(theta_i, theta_star) <- min(1, f(J[i])/f(lambda[i - 1,])).
```

3. Comparer cette probabilité avec une variable aléatoire μ suit une loi uniforme $\mathcal{U}[0; 1]$:

```
> u <- runif(1)
```

```
> if(u < alpha(theta_i, theta_star)){lambda[i, ] <- J[i]} else
```

```
{lambda[i, ] <- lambda[i - 1, ]}.
```

Pour un nombre de simulation $n.sims = 1000$, et pour les valeurs initiales suivantes des paramètres : $\lambda = 1$, $\alpha = 1.8$, $\gamma = 0.01$ et $\delta = 1$, les résultats de simulation de l'algorithme de Metropolis-Hastings sont donnés dans la table (1.3).

1.10.3 Interprétation des résultats

On remarque une différence des résultats entre les deux approches bayésiennes (informative et non informative) ; dans le cas non informative les dix estimateurs sont égaux malgré qu'ils ne doivent pas l'être. En revanche, dans le cas informative, on a gardé cette

variété. On conclut alors, que la présence de l'information a priori est très importante dans l'analyse bayésienne puisque on a d'autres informations sur le paramètre à estimer donc pourquoi ne pas les étudier ?!

D'autre part, en comparant les résultats trouvés par la méthode bayésienne et ceux de la méthode classique, on constate qu'ils sont vraiment proches. Donc, on a pas perdu d'informations.

Conclusion

Ce chapitre consiste à étudier une approche bayésienne en employant les méthodes MCMC pour réaliser l'estimation d'un modèle de poisson hiérarchique.

Comme nous l'avons vu, il est conceptuellement immédiat de passer de la vraisemblance à la distribution a posteriori. Le problème crucial reste alors le choix de la loi a priori, qui a été souvent la pierre d'achoppement de l'inférence bayésienne.

CHAPITRE 2

ANALYSE SÉQUENTIELLE

2.1 Introduction

Nous présentons les principes d'une méthode statistique qui est l'analyse séquentielle, où la taille d'échantillon soit non fixée à l'avance et qui nécessite une évaluation précise de risques. Des applications dans les essais cliniques de phase II permettent de dégager l'intérêt, les limites et le champ d'application d'une telle technique. Malgré ses imperfections, elle constitue un outil non négligeable et une façon différente de raisonner statistiquement, utile dans l'évaluation des plans d'expérience et dans le contrôle de qualité.

2.2 L'a priori séquentielle

Considérons un modèle dominé dans lequel la loi de densité sous-jacente $f(x|\theta)$ est paramétrée par θ et un échantillon x de la variable aléatoire X ayant cette loi de distribution.

Nous acceptons, par la suite, l'hypothèse que l'on connaît avec certitude la forme analytique de la densité $f(x|\theta)$, mais que nous ne connaissons pas avec précision son paramètre θ . Nous disposons d'une série de résultats d'expériences, identiquement distribués

suivant une loi paramétrée par θ .

Il existe plusieurs méthodes pour déterminer la distribution a priori du paramètre θ selon le type d'information disponible (cf. chapitre 1). Nous allons étudier et développer quelques une de ces méthode qu'on appelle séquentielle, qui peut être utilisée avec profit lorsque nous nous limitons au cas où l'information résulte d'observations passées du processus aléatoire. L'évaluation de la distribution a priori est réalisée à l'aide des idées classiques et les analyses associées sont plutôt à la frontière entre statistiques bayésienne et classique ; ainsi dans ce domaine, l'étude est peu développée.

On présente la méthode séquentielle avec une distribution, au départ, soit une mesure uniforme sur l'ensemble Θ , soit une distribution noninformative de la variable aléatoire θ (cf. paragraphe 1.2.1). Nous allons montrer qu'elle permet de retrouver la même forme de distribution et le même paramètre que celle de la conjuguée naturelle avec la fonction de vraisemblance (cf. paragraphe 1.2.2).

Malgré des restrictions sur la famille de distributions conjuguées (cf. paragraphe 1.2.4), la propriété de conjugaison permettra d'avoir une stabilité de la forme des distributions ; ainsi le passage de la distribution a priori à l'a posteriori se réduit à un changement de paramètres, afin de réaliser facilement les calculs.

Une fois la distribution a priori soit déterminée, nous pouvons examiner successivement :

- La distribution a posteriori de θ ;
- La distribution prédictive de la variable aléatoire X ;
- La distribution prédictive d'une nouvelle variable Y en donnant la valeur de la variable aléatoire X

et puis nous passerons éventuellement à l'étape de l'analyse des tests d'hypothèses.

2.2.1 Exhaustivité

Dans un problème statistique où figure un paramètre θ inconnu, un échantillon nous apporte une certaine information sur ce paramètre. Lorsque l'on résume cet échantillon x par une statistique, il ne faut pas perdre cette information sur ce paramètre θ ; une statistique qui la conserve sera qualifiée d'exhaustive.

En termes mathématiques, une statistique T est dite exhaustive par rapport à θ , si la distribution conditionnelle $\mathbb{P}_t = \mathbb{P}_\theta[\cdot | T(X) = t]$ d'un n échantillon aléatoire X , sachant

$T(X) = t$, est indépendante du paramètre θ . Cette distribution conditionnelle est portée par l'ensemble $\Omega_t = \{x \in \Omega; t = T(x)\}$, où $\Omega = \mathcal{X}^n$ l'ensemble des observations.

Si le modèle est dominé et que la statistique T est exhaustive, la fonction de vraisemblance de l'échantillon x se factorise sous certaines conditions de régularité de la manière suivante :

$$f(x|\theta) = h(x)g_\theta(T(x)) = h(x)g(T(x)|\theta) \quad (2.1)$$

où $h(x)$ est une fonction indépendante du paramètre θ . La notation $g(T(x)|\theta)$ indique que g dépend de θ , mais sa dépendance en x est seulement fonction de $t = T(x)$. Le plus souvent $h(x)$ peut être choisie comme une densité de \mathbb{P} sur Ω_t et g comme une densité de la loi de $T(x)$. Dans ce cas, nous disons que la distribution conditionnelle $f(x|\theta)$ admet une statistique exhaustive T .

Ainsi, toute l'information sur θ contenue dans $f(x|\theta)$ se retrouve dans $g(T(x)|\theta)$. Nous pouvons donc dire qu'une fois $t = T(x)$ connue, aucune valeur de l'échantillon ou d'autres statistiques ne nous apporteront des renseignements supplémentaires sur θ . Ce principe d'exhaustivité est très répandu dans l'analyse statistique.

Le principe de factorisation nous donne donc un moyen de reconnaître si une statistique est exhaustive, mais permet difficilement de la construire, ou même de savoir s'il en existe. Pour trouver des statistiques exhaustives, il existe un résultat intéressant montré par Pitman (1936) et Koopman (1936) (Zhu, 1991) : parmi les familles de distributions satisfaisant certaines conditions de régularité, une statistique exhaustive de dimension constante p ne peut exister que dans la famille exponentielle de la forme suivante :

$$f(x|\theta) = a(x)b(\theta) \exp \left[\sum_i c_i(x)d_i(\theta) \right] \quad (2.2)$$

où a et $(c_i)_{1 \leq i \leq p}$ sont des fonctions de $\Omega = \mathcal{X}^n$ dans \mathbb{R} ; b et $(d_i)_{1 \leq i \leq p}$ sont des fonctions de Θ dans \mathbb{R} ; et les fonction a , b sont positives, vérifiant : $\int_\Omega f(x|\theta)dx = 1$

Une statistique exhaustive particulière est définie par :

$$\tilde{T}(x) = \left\{ \sum_i c_1(x_i), \sum_i c_2(x_i), \dots, \sum_i c_p(x_i); n \right\} \quad (2.3)$$

ou une statistique exhaustive "normalisée" par :

$$T(x) = \left\{ \frac{1}{n} \sum_i c_1(x_i), \frac{1}{n} \sum_i c_2(x_i), \dots, \frac{1}{n} \sum_i c_p(x_i); n \right\} \quad (2.4)$$

Exemple 2.1. : Soit $x = (x_1, x_2, \dots, x_n)$ un échantillon de taille n d'une distribution gamma de paramètres α inconnu et β connu. Sa fonction de vraisemblance s'écrit :

$$f(x|\alpha) = \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right)^n \left(\prod_i x_i \right)^{\alpha-1} e^{-\beta \sum_i x_i} = \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right)^n m_G^{n\alpha} e^{-\beta n m} m_G^{-n} \quad (2.5)$$

où $m = \frac{1}{n} \sum_i x_i$ et $m_G = (\prod x_i)^{1/n}$ sont la moyenne arithmétique et la moyenne géométrique. On constate que $\{\prod x_i, n\}$ et $\{m_G, n\}$ sont des statistiques exhaustives par rapport à la variable α .

L'utilisation de la notion d'exhaustivité nous aidera à trouver une loi de distribution du paramètre θ dans l'approche bayésienne et à simplifier les procédures de recueil des données.

2.2.2 Notion de séquentialité

Soit x_1, x_2, \dots, x_n une série d'échantillons successifs, de tailles k_i , relatifs à des d'expériences aléatoires identiques. Soit $p(\theta)$ une distribution a priori. L'hypothèse séquentielle consiste à considérer la distribution a posteriori $p(\theta|x_i)$ comme une nouvelle distribution a priori pour l'échantillon suivant x_{i+1} . La distribution de θ est alors réajustée lors de la réalisation de x_{i+1} par la formule :

$$p(\theta|x_{i+1}) \propto f(x_{i+1} | \theta)p(\theta|x_i).$$

Partant du premier résultat x_1 , nous calculons les distributions de θ l'une après l'autre et nous avons finalement :

$$p(\theta|x_1, x_2, \dots, x_i) = \frac{\prod_j f(x_j | \theta)p(\theta)}{\int_{\Theta} \prod_j f(x_j | \theta)p(\theta)d\theta}.$$

Remarquons que, s'il n'existe pas de stabilité de la forme de la suite de distributions $p_i(\theta) = p(\theta|x_1, x_2, \dots, x_i), i = 1, 2, \dots, n$, la formule n'est guère exploitable de manière analytique sauf s'il existe une statistique exhaustive.

Soit y un nouvel échantillon, la distribution a posteriori $p(\theta|y)$ peut être considérée comme la $(n + 1)$ -ième distribution de la séquence :

$$p(\theta|y) = p(\theta|x_1, x_2, \dots, x_n, y)$$

Nous observons que cette méthode incorpore facilement les données consécutives et que quelle que soit la forme de la distribution a priori, si le nombre n des échantillons est grand, la modification de la distribution a posteriori est très faible. Nous sommes donc amenés, pour tenir compte des connaissances récentes, à limiter le nombre n , c'est à dire à ne pas trop cumuler d'informations disponibles (Zhu, 1991).

Exemple 2.2. *A titre d'exemple montrant que les opinions tendent à converger, considérons les données de contrôle présentées par Smith, Spiegelhalter et Parmar (1996) (Berry et al., (2011)) : sur 1934 patients en unité de soins intensifs, 566 ont développé des infections des voies respiratoires. Comme indiqué dans la table 1.1 du (cf., 1er Chapitre), pour a priori conjuguée $Beta(a; b)$, la postérieure de la probabilité p de développer une infection est $Beta(a + 566; b + 1368)$.*

Avec un échantillon aussi important, deux personnes ayant des opinions antérieures plutôt divergentes s'accorderaient plutôt bien. Par exemple, supposons que la première personne choisira $(a; b) = (1, 19)$ alors que l'autre choisit $(19, 1)$ (comme le montre la Figure 2.1). La probabilité d'infection précédente du premier patient est alors de 0.05, tandis que celle de la seconde est de 0.95. La première a postérieure est $Beta(567; 1387)$ tandis que la deuxième est $Beta(585; 1369)$, les deux étant représentées sur la Figure 2.2.

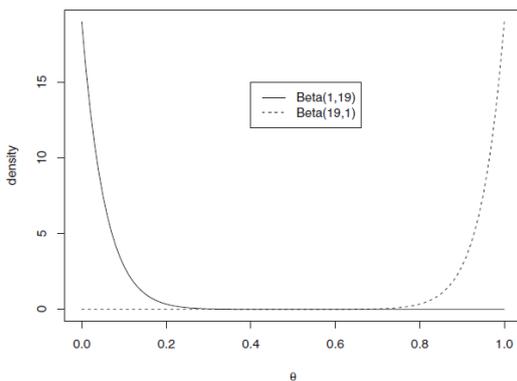


FIGURE 2.1 – Deux densités a priori de bêta assez différentes.

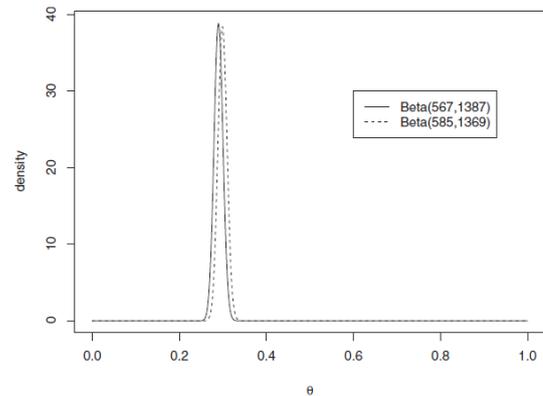


FIGURE 2.2 – Distributions a postérieure de bêta pour les a priori de la figure 2.1.

Les probabilités prédictives moyennes d'infection correspondantes sont remarquablement similaires : 0.290 et 0.299. La probabilité a priori que ces deux experts indépendants donnent une différence qui se situent à moins de 0.1 est presque égale à 0, mais la probabilité postérieure correspondante est proche de 1. Cela démontre une autre caractéristique des méthodes bayésiennes : même des chercheurs ayant des croyances antérieures extrêmement différentes peuvent finalement arriver à se mettre d'accord une fois que suffisamment de données se sont accumulées.

2.2.3 Choix de l'a priori initiale

Pour utiliser la formule séquentielle, il nous manque une distribution a priori comme point de départ.

Supposons maintenant que nous ne possédions pour tout renseignement que les échantillons antérieurs. Avant le premier échantillon, l'information a priori se traduira par le fait que nous ne voudrions privilégier aucune valeur du paramètre inconnu θ . Dans ce cas, il est naturelle que nous prenions l'a priori p comme une mesure uniforme sur l'espace mesurable (Θ, \mathcal{A}) .

Malheureusement, lorsque Θ est un ensemble de mesure non finie, la mesure uniforme est une distribution impropre, puisque la probabilité totale (ou la masse) est infinie : $\int_{\Theta} p(\theta) = +\infty$ (cf. Section 1.4, Chapitre 1).

En cas unidimensionnel, et sans perdre de généralité, nous prendrons : $\Theta =]a, b[$ avec a et $b \in \mathbb{R} \cup \{-\infty, +\infty\}$; la formule séquentielle devient :

$$\begin{aligned} p(\theta|x_1, x_2, \dots, x_i) &= \lim_{\substack{x \rightarrow a \\ y \rightarrow b}} \frac{\frac{1}{y-x} \prod_j f(x_j | \theta)}{\int_x^y \frac{1}{y-x} \prod_j f(x_j | \theta) d\theta} \\ &= \frac{\prod_j f(x_j | \theta)}{\lim_{\substack{x \rightarrow a \\ y \rightarrow b}} \int_x^y \prod_j f(x_j | \theta) d\theta} \\ &\propto \prod_j f(x_j | \theta), \end{aligned}$$

Bien entendu, nous devons supposer l'existence des limites mentionnées dans les expressions. Cette condition est vérifiée pour la famille des distributions exponentielles, ce qui est le cas en général. Cette formule peut se généraliser dans le cas multidimensionnel (Zhu, 1991).

Ces distributions $\{p_i(\theta) = p(\theta|x_1, x_2, \dots, x_i); \quad 1 \leq i \leq n\}$ appartiennent à une famille \mathfrak{S} de distributions conjuguées par rapport à la famille de fonctions de vraisemblance $\mathfrak{S} = \{f(x|\theta); x \in \mathbb{R}^k, k = 1, 2, \dots \text{ et } \theta \in \Theta\}$. En effet, un échantillon composé des n échantillons $x = x_1 \cup x_2 \cup \dots \cup x_n$. Compte tenu du fait que :

$$f(x_i|\theta) = \prod_j f(x_i^j | \theta), \quad j = 1, 2, \dots, k_i,$$

nous avons

$$p_i(\theta) \propto f(x_i|\theta) \propto \prod_j f(x_i^j | \theta)$$

$$p_n(\theta) \propto \prod_i f(x_i|\theta) = \prod_i \prod_j f(x_i^j | \theta) = f(x|\theta)$$

d'où $\mathfrak{S} = \{f(x|\theta); x \in \mathbb{R}^k, k = 1, 2, \dots \text{ et } \theta \in \Theta\}$.

Si la distribution sous-jacente $f(x|\theta)$ admet une statistique exhaustive (sous entendu dans un modèle dominé) : $f(x|\theta) = g(T(x)|\theta)$ pour un échantillon quelconque x , alors il existe une statistique exhaustive pour l'échantillon composé (ou total) des n échantillons antérieurs : $x = x_1 \cup x_2 \cup \dots \cup x_n$, qui nous permet d'écrire à la fois les deux formules suivantes :

$$p_n(\theta) \propto \prod_i f(x_i|\theta) \propto \prod_i g(T(x_i)|\theta); \quad (2.6)$$

$$p_n(\theta) \propto f(x|\theta) \propto g(T(x)|\theta) = g(T(x_1 \cup x_2 \cup \dots \cup x_n)|\theta) \quad (2.7)$$

Pour la famille exponentielle, on obtient l'égalité suivante :

$$g(\tilde{T}(x_1 \cup x_2 \cup \dots \cup x_n)|\theta) = \prod_i g(\tilde{T}(x_i)|\theta) = g(\sum_i \tilde{T}(x_i)|\theta)$$

d'où l'opération de conjugaison linéaire :

$$\tilde{T}(x_1 \cup x_2 \cup \dots \cup x_n) = \sum_i \tilde{T}(x_i)$$

Nous obtenons donc la même forme de distribution a priori que par la méthode de la conjuguée naturelle (cf. Section 2.2.1), ce qui en renforce l'intérêt.

2.3 Prédiction séquentielle

Jusqu'à présent, nous avons utilisé les informations disponibles pour procéder à une évaluation de la plausibilité du paramètre inconnu. Mais ce n'est pas notre objectif final. Dans le pratique, nous sommes intéressés par les données de l'expérience future F et on veut utiliser nos reconnaissances sur le paramètre inconnu pour prévoir les probabilités du future échantillon, et ce, est le fond de l'approche bayésienne (Aitchison and Dunsmore, 1975).

Le concept des problèmes de prédiction est de trouver la distribution de probabilité d'observation y d'un échantillon future F en donnant l'échantillon x de l'expérience informative E :

Définition 2.1 (La densité prédictive). *La fonction de densité prédictive $f(y | x)$ est définie par :*

$$\begin{aligned} f(y | x) &= \int f(y | \theta, x) p(\theta | x) d\theta \\ &= \frac{\int_{\theta} f(y | \theta) p(\theta) f(x | \theta) d\theta}{\int p(\theta) f(x | \theta) d\theta} \end{aligned}$$

Un avantage de l'approche bayésienne est la capacité de considérer les probabilités des résultats non encore observés. À titre d'illustration, nous prenons le lecteur à travers les étapes formelles de la dérivation de la fonction de densité prédictive pour le modèle bêta-binomiale :

2.3.1 Le modèle bêta-binomiale

Ici, E est décrite par une variable aléatoire $X \sim B(n, \theta)$; avec $n \in \mathbb{N}^*$ et $0 \leq \theta \leq 1$. Un résultat $x \in \{0, 1, \dots, n\}$ a pour probabilité :

$$f(x | \theta) = C_n^x \theta^x (1 - \theta)^{n-x}$$

À partir des n expériences antérieures $x_i = (x_i, n_i)$, la question qui se pose est : quelle inférence peut-on faire sur le taux θ avant d'envisager une nouvelle expérience ; et après avoir obtenu un nouveau résultat $x = (x, n)$, quelle inférence peut-on faire sur le taux ? Dans le contexte prédictif, on suppose qu'il est possible de prendre des observations supplémentaires $y = (y, K)$. Comment s'en servir ?

Il s'agit, en effet, de choisir, dans un premier temps, la distribution a priori du taux θ à partir des n expériences et de déterminer, en second lieu, la distribution du taux θ conditionnelle à $x = (x, n)$. Nous pouvons, en première approximation, accepter que la distribution a priori de θ soit une distribution uniforme sur $[0, 1]$; qui est également la version de la distribution $Beta(1, 1)$. Mais il est raisonnable d'espérer que les petites valeurs du taux θ soient plus probables que les grandes, si θ désigne le taux de défectueux en contrôle de qualité. Donc, c'est une distribution "forte" et asymétrique que nous voulons obtenir. Comme la distribution $Beta$ est la distribution a priori conjuguée pour θ , à savoir :

$$p(\theta) = \mathbf{1}_{[0,1]}(\theta)$$

Après la première expérience, nous avons l'observation x_1 de $B(n_1, \theta)$. Une simple application du Théorème de Bayes donne :

$$p(\theta | x_1) \propto \theta^{H-1} (1 - \theta)^{G-1};$$

avec $H = 1 + x_1$ et $G = 1 + n_1 - x_1$ (Table (1.1)).

L'extension, au cas où, E est constitué de k réalisations d'une $B(n, \theta)$ avec l'observation $x = ((x_1, n_1), (x_2, n_2), \dots, (x_k, n_k))$ donne une postérieure séquentielle comme suit :

$$p_k(\theta) \propto \theta^{H-1}(1-\theta)^{G-1} \text{ avec } H = 1 + \sum x_i, G = 1 + \sum (n_i - x_i).$$

Après avoir acquis une nouvelle expérience x , la distribution a posteriori de θ est la $(k+1)$ -ième distribution de la séquence $\{p_i(\theta)\}$:

$$p(\theta | X = x_{k+1}) = p_{k+1}(\theta)$$

. Soit y le résultat de l'expérience future tel que

$$Y \sim B(K, \theta);$$

K n'est pas nécessairement égale à n . Alors, la densité prédictive est donnée par :

$$f(y | x) = \frac{C_K^y \beta(y+x+1, n+K-(y+x)+1)}{\beta(x+1, n-x+1)}$$

où $y \in \{0, 1, \dots, K\}$. De manière équivalente :

$$f(y | x) \propto BeBin(K, x+1, n-x+1).$$

2.4 Tests séquentiels

Étant donnée une population P sur laquelle on mesure un caractère et qui possède, vis-à-vis, de ce caractère une fonction de répartition F , étant données des hypothèses que l'on veut vérifier sur la loi F , les tests statistiques classiques procurent un moyen de vérifier ces hypothèses, en tirant de P suivant des modalités à observer, ou plan d'échantillonnage, un échantillon E de taille donnée et en se donnant un risque de se tromper dans le refus ou l'acceptation d'une hypothèse. L'analyse séquentielle (en anglais " sequential analysis") part d'un autre point de vue :

Soit à choisir entre deux hypothèses H_0 et H_1 , peut-on construire un plan d'échantillonnage consistant à tirer, un à un, des individus de P et à s'arrêter quand on a l'information nécessaire pour choisir entre les deux hypothèses H_0 et H_1 ?

La réponse est oui. La méthode que nous allons étudier est, en général, moins coûteuse en tirages que le plan d'échantillonnage classique ; elle possède, néanmoins, des

particularités qui en restreignent sensiblement l'usage. Découverte pendant la guerre, cette technique fut, paraît-il, à cause de ce gain de temps donc d'argent (qu'on avait peut-être surestimé au départ) décrété secret militaire. Elle a fait l'objet d'une somme de travaux théoriques très importante depuis 1945 et elle est appliquée très souvent dans le contrôle des fabrications industrielles aux Etats-Unis (Millier, 1967).

Depuis quelques années, elle a fait son apparition dans les sciences biologiques, tout naturellement ; en médecine, il paraît en effet intéressant d'arrêter une expérimentation de médicament (ou autre) dès que l'on " voit " statistiquement que l'on s'est trompé.

Nous donnerons plus loin quelques exemples d'application aux essais cliniques.

Après une introduction à la notion de puissance, qui est également essentielle dans l'optimisation des plans d'expériences, nous étudierons les principales caractéristiques des tests séquentiels et les appliquer à des problèmes classiques.

2.4.1 Inférence non-décisionnelle

Il paraît nécessaire de définir la nature d'un test séquentiel par opposition aux tests classiques et pour cela, donnons les généralités sur les tests séquentiels.

Supposons que la loi de probabilité de la variable aléatoire X dépende d'un paramètre inconnu θ . On peut faire sur la valeur de ce paramètre plusieurs hypothèses.

Principe des tests statistiques séquentiels

Supposons que deux hypothèses H_0 et H_1 soient à tester.

1. Le nombre d'observations n'est pas fixe,
2. On a préalablement déterminé dans l'espace des échantillons, trois sous-ensembles mutuellement exclusifs pour chaque valeur du nombre de tirages m ,

R_m^0 : accepter H_0 ,

R_m^1 : accepter H_1 ,

R_m : continuer, c'est-à-dire procéder à un tirage supplémentaire.

On observe $E = \{x_1, x_2, \dots, x_n\}$; suivant que E appartient à R_m^0 , R_m^1 ou R_m , on applique les décisions qui en découlent.

Règle d'arrêt

D'après Robert, (2006), une règle d'arrêt τ peut être définie comme suit :
Si les expériences ξ_i produisent des observations $x_i \in \mathcal{X}_i$, avec $x_i \sim f(x_i|\theta)$, considérons la suite correspondante $\mathcal{A}_i \subset \mathcal{X}_1 \times \dots \times \mathcal{X}_i$ telle que le critère τ prend la valeur n si $(x_1, \dots, x_n) \in \mathcal{A}_n$, i.e., l'expérience s'arrête après la $n^{\text{ème}}$ observation seulement si les n premières observations sont dans \mathcal{A}_n . La vraisemblance de (x_1, \dots, x_n) est alors :

$$f(x_1, \dots, x_n|\theta) = f(x_1|\theta)f(x_2|x_1, \theta)\dots f(x_n|x_1, x_2, \dots, x_{n-1}, \theta)1_{\mathcal{A}_n}(x_1, x_2, \dots, x_n)$$

et donc dépend seulement de τ via l'échantillon x_1, \dots, x_n . Ceci implique le principe suivant.

Principe de la règle d'arrêt

Si une suite d'expériences, ξ_1, ξ_2, \dots , admet une règle d'arrêt, τ , qui implique quand doivent s'arrêter les expériences, l'inférence sur θ ne doit dépendre de τ qu'à travers l'échantillon résultant.

Caractéristiques d'un test séquentiel

Définition 2.2. La fonction OC (de l'anglais « operating characteristic »).

Soit H_0 une hypothèse prédéterminée, et θ la valeur du (ou des) paramètre (s) de la loi de probabilité; la probabilité d'accepter H_0 est évidemment une fonction de θ , l'OC est elle que :

$$Pr\{ \text{accepter } H_0 \text{ si } \theta \} = L(\theta);$$

$$Pr\{ \text{refuser } H_0 \text{ si } \theta \} = 1 - L(\theta).$$

Définition 2.3. La fonction ASN (de l'anglais « average sample number »)

Dans le test séquentiel, le nombre de tirages n'est pas fixe, c'est une variable aléatoire qui admet des moments et en particulier une moyenne.

Cette moyenne dépend de la valeur θ du (ou des) paramètre (s); c'est l'ASN = $E(N|\theta)$ (nombre moyen d'observations à prendre en compte).

Influence du test sur l'OC et E(N)

Soit à tester $\{H_0 : \theta = \theta_0\}$ contre $\{H_1 : \theta = \theta_1\}$ on a :

$$L(\theta_1) = Pr\{ \text{accepter } H_0 \text{ si } \theta = \theta_1 \} = \beta$$

de même

$$L(\theta_0) = Pr\{ \text{accepter } H_0 \text{ si } \theta = \theta_0 \} = 1 - \alpha.$$

$L(\theta)$ est donc asymptotique à $L(\theta) = 1$ et $L(\theta) = 0$ et passe par A si l'échantillonnage est classique, par A et par B si l'échantillonnage est séquentiel.

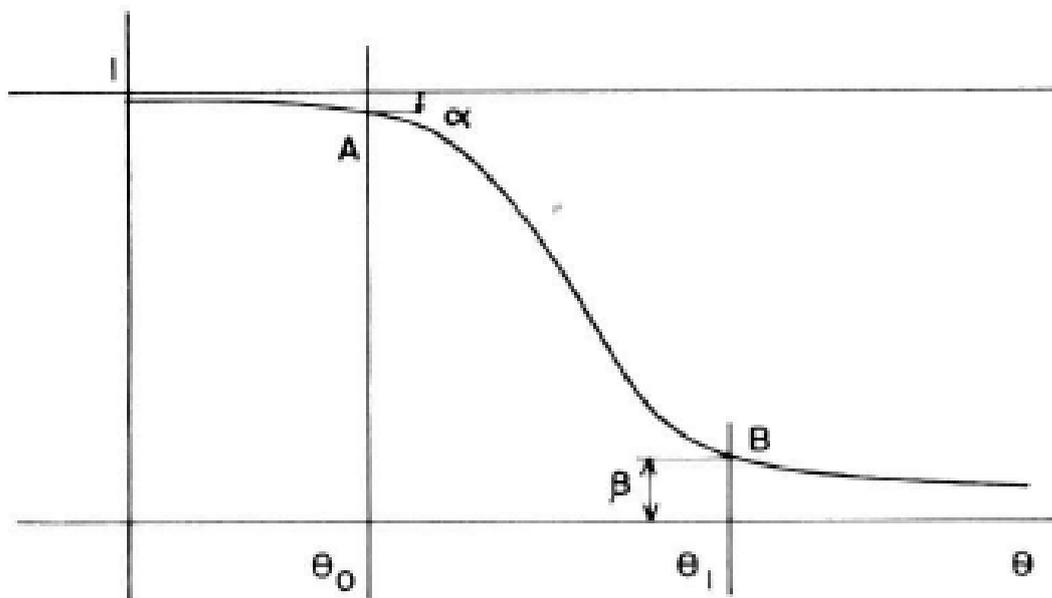


FIGURE 2.3 – Courbe OC.

— Si l'échantillonnage est classique, le choix du test et le choix du nombre de tirages conditionnent la forme de la courbe et le point B ;

— Si l'échantillonnage est séquentiel, le choix du test déterminera la forme de la courbe et le nombre moyen de tirages.

Quand nous comparions deux tests classiques, pour un nombre de tirages donné, nous préférons le test le plus puissant ($1 - \beta$ maximum), celui qui minimisait le risque de deuxième espèce.

Ici, nous choisirons le test qui demande un nombre moyen de tirages minimum (comparaison des courbes OC).

Dans la suite, nous allons exposer les propriétés des tests séquentiels dans les essais cliniques de phase II à un seul bras.

2.5 Tests séquentiels dans les essais cliniques à un seul bras

L'étude de phase II des essais cliniques joue un rôle central dans le développement des médicaments. Le but de cette phase est de détecter les traitements montrant une efficacité suffisante pour justifier la poursuite des études (Schultz et al., 1973; Setta, 2005; Goulet, 2015).

Dans cette section, notre but, est d'étudier les règles d'arrêt dans cette phase par démontrer l'efficacité et la robustesse de l'inférence bayésienne, en utilisant la probabilité prédictive pour améliorer les conceptions (designs) de cette phase.

2.5.1 Cadre général

Une grande partie de la méthodologie statistique utilisée dans les essais avec un seul bras est dérivé de l'oncologie. De tels essais sont généralement des essais non contrôlés, à savoir un seul bras; un seul traitement expérimental est évalué pour décider s'il a une efficacité suffisante pour justifier une étude plus approfondie. Ces essais utilisent, souvent, un résultat «réponse» binaire, à savoir le médicament "fonctionne" ou il ne le fait pas. Dans cette section, nous nous concentrons sur les essais à un seul bras.

Dans ce contexte, nous avons les hypothèses suivantes :

$H_0 : \theta \leq \theta_0$ où la vraie probabilité de réponse θ est inférieur à un certain niveau θ_0 inintéressant.

$H_1 : \theta > \theta_1$ où la véritable probabilité de réponse est au moins une partie souhaitable θ_1 niveau cible.

Où

θ_0 : est le taux de la réponse indésirable.

θ_1 : est le taux de la réponse désirable.

tel que $\theta_1 > \theta_0$.

2.5.2 Design à une seule étape

Dans une conception en une seule étape un nombre prédéterminé de patients sont inscrits et l'hypothèse est testée une seule fois, à savoir, à la fin de l'essai (Schultz, 1973).

Afin d'étudier l'efficacité du nouveau médicament, nous donnons une formule pour la taille de cette conception en une seule étape. Cette formule dépend du test choisi ; il existe deux méthodes différentes :

- La première est basée sur une approximation *Normale*.
- Le second est basé sur les probabilités *Binomiales* exactes.

Les deux méthodes, utilisent le nombre de succès X comme point de départ pour trouver la taille d'échantillon.

L'approximation normale

Cette procédure en une étape est basée sur la statistique $\hat{\theta} = X/N$, où N est la taille de l'échantillon et X suit une distribution *Binomiale* de paramètres (N, θ) . De plus, nous supposons que pour N suffisamment grand, $\hat{\theta}$ suit approximativement, une distribution *Normale* avec les paramètres $(\theta, \theta(1 - \theta)/N)$ et

$$Z = \frac{(\hat{\theta} - \theta)}{se(\theta)} \sim Normal(0, 1); \text{ avec } se(\theta) = \sqrt{\frac{\theta(1 - \theta)}{N}}.$$

La procédure, en une seule étape, sera, alors comme suit : rejeter H_0 lorsque $Z > Z_\alpha$; ou de façon équivalente lorsque :

$$X > Np_0 + Z_\alpha \sqrt{N\theta_0(1 - \theta_0)}.$$

Il est facile de montrer que la taille de l'échantillon de cette conception avec le niveau de signification α et la puissance $(1 - \beta)$ sur H_1 est approximativement égale à :

$$N_{\max} = \frac{(Z_\alpha + Z_\beta)^2 \bar{\theta}(1 - \bar{\theta})}{(\theta_1 - \theta_0)^2} ; \text{ avec } \bar{\theta} = \frac{(\theta_0 + \theta_1)}{2}$$

Exemple 2.3. *Supposons que nous voulons tester un traitement sur une maladie donnée et que 20% serait un taux de réponse intéressant. Alors, si la réponse est seulement de 10%, le traitement sera rejeter et on n'aura pas besoin à une étude plus approfondie.*

Pour $\alpha = 0.05$ (unilatéral), $\beta = 0.20$, $H_0 : \theta = 0.10$, et $H_1 : \theta = 0.20$, nous constatons que le nombre de patients N_{\max} est égal à 79, (Voir le programme 1 dans l'annexe).

TABLE 2.1 – Taille d'échantillon optimale dans l'exemple 2.4.

| θ_1 | θ_0 | N_{\max} | $MinN$ | $Mink$ | $Alphaa$ | $Puissance$ |
|------------|------------|------------|--------|--------|----------|-------------|
| 0.2 | 0.1 | 79 | 79 | 12 | 0.049 | 0.822 |
| 0.2 | 0.1 | 79 | 82 | 13 | 0.032 | 0.785 |
| 0.2 | 0.1 | 79 | 86 | 13 | 0.046 | 0.841 |
| 0.2 | 0.1 | 79 | 88 | 14 | 0.028 | 0.793 |
| 0.2 | 0.1 | 79 | 89 | 14 | 0.030 | 0.807 |
| 0.2 | 0.1 | 79 | 90 | 14 | 0.033 | 0.821 |

La distribution binomiale exacte

La taille de l'échantillon peut également être obtenue en utilisant la distribution *Binomiale* exacte de X .

Pour θ_0 , θ_1 , α et β spécifiés nous déterminons la taille optimale de l'échantillon par énumération, qui peuvent être résumées dans les étapes suivantes :

- Prendre la valeur obtenue dans la section précédente N_{\max} , et de varier la taille de l'échantillon N de $(3/4)N_{\max}$ à $(5/4)N_{\max}$.
- Pour chaque N , nous déterminons le minimum de k dans $(0, N)$ pour lesquels $\Pr(X \geq k) < \alpha$ sous l'hypothèse nulle (θ_0).
- Ensuite, prendre le minimum de N pour laquelle $\Pr(X \geq k) > 1 - \beta$ (sous θ_1).
- La taille de l'échantillon nécessaire sera, alors, le plus petit N et la valeur critique k correspondante qui satisfait les deux contraintes de probabilité :
 - 1- $\alpha_k = \Pr(\text{rejeter } H_0 | \theta_0) < \alpha$.
 - 2- $1 - \beta_k = \Pr(\text{rejeter } H_0 | \theta_1) > 1 - \beta$.

Exemple 2.4. Si l'on calcule, maintenant, la taille de l'échantillon en utilisant les probabilités Binomiales exactes, nous voyons dans le Tableau (2.1) ci-dessous que $N_{\max} = 79$ ($k = 12$) est le premier N dont la puissance est plus de 80%. Toutefois, pour $N = 82$ ($k = 13$) la puissance baisse et ne satisfait pas la condition requise, ce qui est un résultat de la discontinuité de la distribution Binomiale.

Tandis que $N = 89$, est le premier dont la puissance satisfait la condition et ne descend pas en dessous de 80%, (Voir le programme 2 dans l'annexe).

2.5.3 Design à plusieurs étapes

D'après Chow, (2010), il est désirable de terminer l'étude prématurément, lorsque, le traitement de test n'est pas efficace à cause des considérations éthiques. Par conséquent, le design à plusieurs étapes pour un seul bras de l'essai est employé pour déterminer si le test traité est promis comme un test final. Il existe plusieurs designs les plus connus sont celles de **Simon**.

Design à deux étapes

Comme dans la section précédente, l'hypothèse nulle à tester est que la véritable probabilité de réponse (θ) est inférieure à un certain niveau θ_0 inintéressant : $H_0 : \theta \leq \theta_0$.

L'hypothèse alternative est que la véritable probabilité de réponse est au moins une proportion souhaitable θ_1 (niveau cible) : $H_1 : \theta > \theta_1$. Lorsque H_0 ne peut être rejetée (on dit simplement «accepté»), le médicament sera rejeter et ne nécessite pas une étude plus approfondie.

Afin de réduire le nombre de patients, en cas de signes précoces de (in) efficacité, une analyse intermédiaire est effectuée. Deux modèles différents sont examinés : Un modèle fréquentiste et une conception bayésienne et les caractéristiques communes de ces designs sont donnés ci-dessous.

Tout au long, on considère que n_1, n_2 le nombre de patients dans la première et la seconde étapes, respectivement, et x_1, x_2 les nombres de réponses.

Après la première étape, lorsque les n_1 patients sont traités nous allons :

- Accepter H_0 si $x_1 \leq a_1$
- Rejeter H_0 si $x_1 \geq r_1$
- continuer si $a_1 < x_1 < r_1$

En dernière analyse, lorsque $N_{\max} = n_1 + n_2$ patients sont traités, nous allons :

- Accepter H_0 si $x_1 + x_2 \leq a_2$
- Rejeter H_0 si $x_1 + x_2 \geq r_2$ avec $r_2 = a_2 + 1$

Notons que dans l'analyse finale, une décision doit être prise, à savoir il n'y a pas de nombre de réponses qui peuvent être mis à la fin de l'essai.

Désignons par $b(x, n_1, \theta)$ et $B(a_1, n_1, \theta)$ la densité de probabilité et la fonction cumulative, respectivement, de la distribution binomiale $B(n_1, \theta)$, Les probabilités d'acceptation et de rejet après la première étape sont les suivantes :

$$\begin{aligned}\Pr(\text{accepter } H_0|\theta) &= P(x_1 \leq a_1) = B(a_1, n_1, \theta) \\ \Pr(\text{rejeter } H_0|\theta) &= P(x_1 \geq r_1) = 1 - B(r_1 - 1, n_1, \theta)\end{aligned}\quad (2.8)$$

Par conséquent, dans ces conceptions, ($PET(\theta)$) la probabilité de terminer tôt l'essai après la première étape est :

$$PET(\theta) = P(x_1 \leq a_1) + P(x_1 \geq r_1) = B(a_1, n_1, \theta) + (1 - B(r_1 - 1, n_1, \theta)) \quad (2.9)$$

En outre, ($E(N|\theta)$) la taille d'échantillon espérée ou la taille moyenne de l'échantillon est :

$$E(N|\theta) = n_1 + (1 - PET(\theta))n_2$$

Les probabilités d'acceptation et de rejet globales (à savoir la première et la deuxième étape combinées) sont :

$$\begin{aligned}\Pr(\text{accepte } H_0|\theta) &= P(x_1 \leq a_1) + P(x_1 + x_2 \leq a_2; a_1 < x_1 < r_1) \\ &= B(a_1, n_1, \theta) + \sum_{k=a_1+1}^{r_1-1} b(k, n_1, \theta)B(a_2 - k, n_2, \theta)\end{aligned}\quad (2.10)$$

$$\begin{aligned}\Pr(\text{rejeter } H_0|\theta) &= P(x_1 \geq r_1) + P(x_1 + x_2 \geq r_2; a_1 < x_1 < r_1) \\ &= 1 - B(r_1 - 1, n_1, \theta) + \sum_{k=a_1+1}^{r_1-1} b(k, n_1, \theta)(1 - B(r_2 - k - 1, n_2, \theta))\end{aligned}\quad (2.11)$$

2.6 Règles d'arrêt utilisés dans les essais cliniques

Dans l'analyse statistique des essais cliniques, on peut utiliser les deux approches : fréquentiste et bayésienne. On peut dire que les deux approche sont complémentaires et peuvent être valides. Dans cette section nous donnons les designs les plus utilisés fréquentiste et bayésienne.

2.6.1 Design fréquentiste

Simon (1989)

Dans le cadre des tests d'hypothèses de notre travail, un essai clinique de phase IIA est conçu pour tester :

$$\begin{cases} H_0 : \theta \leq \theta_0 \\ H_1 : \theta \geq \theta_1 \end{cases}$$

où θ_0 représente le taux de réponse prédéfini du traitement standard et θ_1 représente le taux de réponse cible d'un nouveau traitement. Une étude est conçue pour que

$$\begin{aligned} &P(\text{Accept New treatment} \mid H_0) \leq \alpha \\ \text{et} \quad &P(\text{Reject New treatment} \mid H_1) \leq \beta \end{aligned}$$

où α et β sont les taux d'erreur de type I et de type II respectivement.

Sachant θ_0 , θ_1 , le nombre maximal de patients, le nombre d'étapes, la taille de la cohorte à chaque étape, la région d'acceptation et celle de rejet à chaque étape, on peut calculer les taux d'erreur de type I et de type II, la probabilité d'arrêt précoce (*PET*) du traitement de l'essai et la taille d'échantillon espérée ($E(N)$) sous H_0 , en appliquant la formule récursive de Schultz et al. (1973).

Une conception souhaitable est celle dont les caractéristiques de fonctionnement satisfont les contraintes des taux d'erreur de types I et II, avec une forte probabilité d'arrêt précoce et une petite taille d'échantillon attendue sous H_0 . Notons que, nous définissons la région d'acceptation comme l'espace de résultats conduisant à l'acceptation du nouveau traitement (c'est-à-dire, rejeter l'hypothèse nulle) et la région de rejet comme l'espace de résultats conduisant au rejet de l'agent (autrement dit, ne pas rejeter l'hypothèse nulle)

L'algorithme de mise en œuvre des conceptions en deux étapes de Simon est le suivant :

Étape I : Recruter m_1 patients. Arrêter l'essai et rejeter le nouveau traitement (Arrêt pour futilité) si le nombre de réponses observés est x_1 ou moins. Sinon, continuer jusqu'au deuxième étape (Étape II).

Étape II : Recruter $(N_{\max} - m_1)$ patients pour atteindre un nombre total N_{\max} de patients. Rejeter le nouveau traitement si le nombre de réponses est x_2 ou moins. Sinon, considérer que le nouveau traitement mérite d'être développé.

Basé sur ces règles d'arrêt, PET et $E(N)$ sous H_0 peuvent être calculer par :

$$PET(\theta_0) = \Pr(\text{Early Termination} \mid H_0) = \Pr(X_1 \leq x_1 \mid H_0)$$

où, X_1 est le nombre de réponses dans m_1 ($m_1 < N_{\max}$) patients à partir du première étape, et $X_1 \sim \text{binomial}(m_1, \theta_0)$

$$E(N \mid \theta_0) = m_1 + [1 - PET(\theta_0)] \times (N_{\max} - m_1)$$

Parmi toutes les conceptions qui satisfont à la contrainte des taux d'erreur des types I et II, la conception optimale en deux étapes de Simon est obtenue lorsque $E(N \mid \theta_0)$ est la plus petite. D'autre part, la conception minimax est définie lorsque la taille maximale de l'échantillon est la plus petite.

2.6.2 Design Bayésien

Herson (1979)

Dans cette conception, la probabilité θ n'est plus fixé, mais supposée suivre une certaine distribution de probabilité $p(\theta)$. Pour les données *Binomiales*, la conjuguée *Beta* est généralement choisie qui comprend la distribution *Uniforme* comme un cas particulier (c.f Chapitre 1).

Herson a proposé un design Bayésien à deux étapes. Malheureusement, ses hypothèses sont l'inverses de **Simon** i.e. :

$H_0 : \theta = \theta_0$, θ_0 est un niveau intéressant.

$H_1 : \theta < \theta_0$, i.e., à savoir le niveau de réponse est sans intérêt.

L'idée générale d'un tel design peut être résumée par les étapes suivantes (nous allons continuer avec la notation de **Herson**) :

- Tout d'abord, pour un α donné, β et θ_0 déterminer la taille de l'échantillon N_{\max} et la région de rejet r en utilisant les probabilités *Binomiales* exactes.
- Puis, à l'analyse intérimaire, calculer la probabilité prédictive (PP) de rejeter H_0 à la fin de l'étude, compte tenu des résultats intérimaires.
- Rejeter H_0 si le PP est supérieure à un certain seuil P_0 .

1- Probabilité prédictive

Une taille d'échantillon N_{\max} appropriée peut être choisie pour un design en une seule étape et pour le test qui rejette H_0 en faveur de H_1 à chaque fois que les réponses observées sont $\leq r$. Ensuite, à n'importe quel moment au cours de l'étude, disant quand $n_1 < N_{\max}$ patients sont observés avec des réponses $x_1 \leq a_1$, il convient de calculer la probabilité que le nombre de réponses dans N_{\max} patients soit $x_1 + x_2 \leq r$. Cette probabilité est appelée la probabilité prédictive (PP). Une haute PP indiquerait qu'il est très probable que H_0 sera rejeté à la fin d'essai et, par conséquent, l'arrêt précoce pourrait être souhaitable tandis que une faible valeur de PP donnerait un peu de justification d'arrêt précoce.

Soit $p(\theta)$ la densité a priori alors :

$$\begin{aligned} P[X \text{ reponses à } N_{\max} \text{ patients} | x_1 \text{ reponses en } n_1 \text{ premier}, p(\theta)] &= P[(X, N_{\max}) | (x_1, n_1), p(\theta)] \\ &= \frac{\int_0^1 P[(x_1, n_1) | \theta] P[(X, N_{\max}) | (x_1, n_1), \theta] p(\theta) d\theta}{\int_0^1 P[(x_1, n_1) | \theta] p(\theta) d\theta} \end{aligned} \quad (2.12)$$

Pour les données binaires, prenons la distribution *Beta*, comme a priori, avec des paramètres r et s , qui suit : $p(\theta) = L\theta^{r-1}(1-\theta)^{s-1}$, L est une constante de normalisation.

Cependant, **Herson** utilisait une paramétrisation différente avec $a = r - 1$ et $b - a = s - 1$. Dans ce cas, la fonction de densité sera $p(\theta) = L\theta^a(1-\theta)^b$.

Alors, la formule (2.12) devient :

$$\begin{aligned} P[(X, N_{\max}) | (x_1, n_1), (a, b)] &= C_{N_{\max}-n_1}^{X-x_1} \frac{\int_0^1 \theta^{X+a}(1-\theta)^{(N_{\max}+b)-(X+a)} d\theta}{\int_0^1 \theta^{x_1+a}(1-\theta)^{(n_1+b)-(x_1+a)} d\theta} \\ &= C_{N_{\max}-n_1}^{X-x_1} \frac{\beta(X+a+1, N_{\max}+b-X-a+1)}{\beta(x_1+a+1, n_1+b-x_1-a+1)}. \end{aligned}$$

Où

$$C_x^y = \frac{x!}{y!(x-y)!}$$

Et

$$P[X \leq r \text{ à } N_{\max} \text{ patients} | (x_1, n_1), (a, b)] = \sum_{X=x_1}^r P[(X, N_{\max}) | (x_1, n_1), (a, b)] \quad (2.13)$$

La probabilité prédictive (PP) dans (2.13) formera la base des règles d'arrêt précoce à développer.

2- Quantification des croyances antérieures

La quantification des croyances sur θ dans l'étude est équivalente à la spécification des paramètres a, b , dans la distribution *Beta*. **Herson** a présenté l'approche suivante : L'enquêteur précise ses croyances a priori de la moyenne $\bar{\theta}$ de la distribution a priori θ et une expression du degré de confiance en $\bar{\theta}$ via le coefficient de variation $C.V(\bar{\theta})$. Cette dernière est égale à la variation divisée par la moyenne.

Une faible valeur de $C.V.$ indiquerait que l'enquêteur a un degré de confiance élevé pour que $\theta = \bar{\theta}$ alors qu'une haute $C.V.$, indique un manque de confiance.

pour la distribution *Beta* on retrouve les équations suivantes :

$$\begin{cases} \bar{\theta} = (a + 1)/(b + 2) \\ V^2 = (b - a + 1)/(a + 1)(b + 3) \end{cases}$$

D'où

$$\begin{cases} a = (b + 2)\bar{\theta} - 1 \\ b = [1 - \bar{\theta}(1 - 3V^2)]/V^2\bar{\theta} \end{cases}$$

Avec $C.V(\bar{\theta}) = V.100\%$.

3- Sélection d'un plan d'arrêt précoce

Le processus de sélection d'un *PP*-plan d'arrêt précoce peut être décrit par les étapes suivantes :

- Dériver N_{\max} et sélectionner n_1 et n_2 .
- Ensuite, pour des valeurs spécifiques de $\bar{\theta}$ et $C.V(\bar{\theta})$, dériver les paramètres a et b de la distribution *Beta*.
- Calculer le *PP*.

Pour un P_0 donné dériver a_1 en recherchant sur l'ensemble de $x_1 \in (0, r)$; pour chaque x_1 nous déterminons la valeur *PP* donnée en (2.13) et nous prenons a_1 égale à la valeur maximale de x_1 qui satisfait $PP < P_0$.

Si le *PP* est supérieure à P_0 , l'étude est terminée et que le médicament est abandonné, sinon l'essai se poursuit. Dans ce cas, après la première étape lorsque les n_1 patients sont traités, nous :

- 1- Rejetons H_0 lorsque $x_1 \leq a_1$.

Et dans l'analyse finale, nous allons :

2- Rejeter H_0 lorsque $x_1 + x_2 \leq r$.

Exemple 2.5. Les tableaux suivants donnent les mêmes exemples que dans le Tableau (??) (les mêmes probabilités de réponse, et la même α et β), la conception bayésienne de **Herson** en deux étapes basées sur les hypothèses inverses et une distribution de Beta. Avec $P_0 = 0.85$ et $C.V = 50\%$, (Voir les programmes 3, 4 et 5 dans l'annexe).

Pour la distribution Uniforme :

| | | une seule étape | | | | deux étapes | | | | | | |
|------------|------------|-----------------|-----|-----------|----------|-------------|-------|-------|-----------|----------|-----------------|-----------------|
| θ_0 | θ_1 | N_{\max} | r | $alphak1$ | $puisk1$ | r_1 | n_1 | PP | $alphak2$ | $puisk2$ | $E(N \theta_0)$ | $E(N \theta_1)$ |
| 0.3 | 0.10 | 25 | 3 | 0.033 | 0.764 | 1 | 15 | 0.858 | 0.054 | 0.790 | 24.6 | 19.5 |
| 0.4 | 0.20 | 35 | 8 | 0.026 | 0.745 | 3 | 20 | 0.913 | 0.035 | 0.760 | 34.8 | 28.8 |
| 0.2 | 0.05 | 40 | 3 | 0.028 | 0.862 | 0 | 20 | 0.952 | 0.035 | 0.868 | 39.8 | 32.8 |
| 0.4 | 0.20 | 50 | 13 | 0.028 | 0.889 | 4 | 25 | 0.952 | 0.033 | 0.894 | 49.8 | 39.5 |
| 0.5 | 0.30 | 45 | 16 | 0.036 | 0.836 | 7 | 25 | 0.903 | 0.047 | 0.847 | 44.6 | 34.8 |
| 0.5 | 0.30 | 50 | 18 | 0.032 | 0.859 | 7 | 25 | 0.898 | 0.045 | 0.870 | 49.5 | 37.2 |

Pour la distribution Beta avec $\bar{\theta} = \theta_1$:

| | | une seule étape | | | | deux étapes | | | | | | |
|------------|------------|-----------------|-----|-----------|----------|-------------|-------|-------|-----------|----------|-----------------|-----------------|
| θ_0 | θ_1 | N_{\max} | r | $alphak1$ | $puisk1$ | r_1 | n_1 | PP | $alphak2$ | $puisk2$ | $E(N \theta_0)$ | $E(N \theta_1)$ |
| 0.3 | 0.10 | 25 | 3 | 0.033 | 0.764 | 1 | 15 | 0.930 | 0.054 | 0.790 | 24.6 | 19.5 |
| 0.4 | 0.20 | 35 | 8 | 0.026 | 0.745 | 3 | 20 | 0.941 | 0.035 | 0.760 | 34.8 | 28.8 |
| 0.2 | 0.05 | 40 | 3 | 0.028 | 0.862 | 1 | 20 | 0.908 | 0.081 | 0.896 | 39.8 | 25.3 |
| 0.4 | 0.20 | 50 | 13 | 0.028 | 0.889 | 5 | 25 | 0.908 | 0.047 | 0.903 | 49.8 | 34.6 |
| 0.5 | 0.30 | 45 | 16 | 0.036 | 0.836 | 7 | 25 | 0.925 | 0.047 | 0.847 | 44.6 | 34.8 |
| 0.5 | 0.30 | 50 | 18 | 0.032 | 0.859 | 7 | 25 | 0.923 | 0.045 | 0.870 | 49.5 | 37.2 |

Pour la distribution Beta avec $\bar{\theta} = \theta_0$:

| | | une seule étape | | | | deux étapes | | | | | | |
|------------|------------|-----------------|-----|-----------|----------|-------------|-------|-------|-----------|----------|-----------------|-----------------|
| θ_0 | θ_1 | N_{\max} | r | $alphak1$ | $puisk1$ | r_1 | n_1 | PP | $alphak2$ | $puisk2$ | $E(N \theta_0)$ | $E(N \theta_1)$ |
| 0.3 | 0.10 | 25 | 3 | 0.033 | 0.764 | 0 | 15 | 0.962 | 0.035 | 0.766 | 25.0 | 22.9 |
| 0.4 | 0.20 | 35 | 8 | 0.026 | 0.745 | 3 | 20 | 0.893 | 0.035 | 0.760 | 34.8 | 28.8 |
| 0.2 | 0.05 | 40 | 3 | 0.028 | 0.862 | 0 | 20 | 0.873 | 0.035 | 0.868 | 39.8 | 32.8 |
| 0.4 | 0.20 | 50 | 13 | 0.028 | 0.889 | 4 | 25 | 0.940 | 0.033 | 0.894 | 49.8 | 39.5 |
| 0.5 | 0.30 | 45 | 16 | 0.036 | 0.836 | 7 | 25 | 0.895 | 0.047 | 0.847 | 44.6 | 34.8 |
| 0.5 | 0.30 | 50 | 18 | 0.032 | 0.859 | 7 | 25 | 0.889 | 0.045 | 0.870 | 49.5 | 37.2 |

Discussion

Remarquons que le taux d'erreur du type I dans le design à deux étapes est plus grand que celle du design à une seule étape et c'est à cause de l'analyse intérimaire qui augmente la probabilité de rejeter H_0 .

En outre, on peut observer que, sous les trois distributions a priori, la taille d'échantillon observé est toujours inférieur à la taille d'échantillon pour une seule étape surtout sous θ_1 . Il est intéressant de comparer la taille d'échantillon sous θ_1 pour les différentes distributions a priori. Prenons par exemple dans le design ($\theta_0 = 0.3$, et $\theta_1 = 0.1$), on remarque qu'on a la même taille d'échantillon espérée ($E(N|\theta_1) = 19.5$) pour les a priori *Uniforme* et *Beta* avec $\bar{\theta} = \theta_1$ par contre pour *Beta* avec $\bar{\theta} = \theta_0$ elle est plus grande ($E(N|\theta_1) = 22.9$). Ceci montre une caractéristique importante de l'approche Bayésienne c'est que lorsque l'a priori se situe proche de $\bar{\theta} = \theta_1$ (le traitement n'est pas promettant) $E(N)$ est plus petite que celle de l'a priori qui se situe proche de $\bar{\theta} = \theta_0$ (le traitement est promettant).

PP-design

Dans cette section, nous allons donner une nouvelle méthode pour calculer les probabilités prédictives, dans un contexte purement bayésien et nous décrivons leur utilisation pour améliorer le design des essais cliniques de la phase *II*. Un avantage de cette approche est qu'elle imite le processus décisionnel clinique. Basé sur les données intérimaires, la probabilité prédictive est obtenue en calculant la probabilité d'une conclusion positive (de rejeter l'hypothèse nulle) si l'essai se déroule à la taille maximale de l'échantillon prévue Lee and Liu (2008).

Donc, on peut prendre une décision de continuer ou d'arrêter l'essai en fonction de la puissance de cette probabilité prédictive.

Notre objectif est d'évaluer la proportion θ pour un nouveau médicament en testant l'hypothèse

$$H_0 : \theta \leq \theta_0 \quad \text{contre} \quad H_1 : \theta \geq \theta_1$$

dans la phase *II* d'un essai.

D'après Lee et Liu (2008), en donnant θ_0 , θ_1 , la distribution a priori du taux de réponse $p(\theta)$ et la taille de cohorte de contrôle intérimaire, nous cherchons les paramètres

de design, y compris la taille de l'échantillon maximum N_{\max} , θ_L , θ_T et θ_U pour obtenir un design satisfaisant les contraintes d'erreur du type I et du type II simultanément.

La règle de décision peut être construite comme suit :

Algorithme 2.1. (*Phase II basée sur le PP-design*)

Étape 1 : Si $PP < \theta_L$, arrêter l'essai et rejeter l'hypothèse alternative.

Étape 2 : Si $PP > \theta_U$, arrêter l'essai et rejeter l'hypothèse nulle.

Étape 3 : Sinon, passez à l'étape suivante jusqu'à ce qu'il atteigne N_{\max} patients.

Typiquement, θ_L et θ_U sont deux nombres positives choisis compris entre 0 et 1, avec θ_L est très proche de 0 et θ_U est très proche de 1.

$PP < \theta_L$ signifie qu'il est improbable que le taux de succès serait supérieur à θ_0 , d'ici à la fin de l'essai, compte tenu des données actuelles. Dans ce cas, nous pouvons également arrêter l'essai et rejeter l'hypothèse alternative à ce moment là.

$PP > \theta_U$, les données actuelles suggèrent que, si le même mouvement s'étend, nous aurons une probabilité élevée pour conclure que le traitement est efficace à la fin de l'étude.

Ceci résulte, et conduit à arrêter prématurément l'essai en raison d'efficacité.

En choisissant $\theta_L > 0$ et $\theta_U < 1$, l'essai peut se terminer prématurément en raison soit de futilité soit d'efficacité. Pour les essais de la phase II, nous préférons souvent de choisir $\theta_L > 0$ et $\theta_U = 1$, pour privilégier un arrêt prématuré en raison de futilité, mais non plus d'efficacité.

Calcul de la probabilité prédictive

Prenons toujours la distribution $Beta(a, b)$ comme a priori pour θ . La quantité $\frac{a}{(a+b)}$ est la moyenne a priori, Les quantités a et b peuvent être considérées comme le nombre de réponses et le nombre de non réponses, respectivement. Ainsi, $a + b$ peut être considéré comme une mesure de la quantité d'informations contenues dans l'a priori. Plus la valeur de $a + b$ est grande, plus l'a priori est informative et plus la croyance qu'il contient est forte.

Supposons que nous avons fixé un nombre maximum de patients N_{\max} , et suppo-

sons que X , qui est le nombre de réponses chez les patients parmi les n patients en cours ($n \leq N_{\max}$), suit une distribution $B(n, \theta)$, alors la distribution a posteriori de θ serai :

$$\theta|x \sim \text{Beta}(a + x, b + n - x).$$

L'approche de probabilité prédictive regarde vers l'avenir sur la base des données actuelles observées pour projeter si une conclusion positive à la fin de l'étude est susceptible ou non, puis prendre une décision raisonnable à l'heure actuelle par conséquent.

Soit Y le nombre de réponses dans le potentiel de futurs patients $m = N_{\max} - n$. Supposons que notre design est de déclarer l'efficacité, si la probabilité a posteriori de θ qui est supérieur à un certain niveau prédéfini θ_0 est supérieure à un seuil θ_T . Marginaliser θ de la probabilité *Binomial*, il est bien connu que Y suit une distribution de *Beta-Binomial*, $Y|X = x \sim \text{Beta} - B(m, a + x, b + n - x)$.

Quand $Y = i$, la distribution a posteriori de $\theta|(X = x, Y = i)$ est une $\text{Beta}(a + x + i, b + N_{\max} - x - i)$. La probabilité prédictive (PP) du succès de l'essai peut être alors calculée comme suit :

Soit $B_i = P(\theta > \theta_0|x, Y = i)$ et $1_i = 1_{\{B_i > \theta_T\}}$, on a :

$$\begin{aligned} PP &= E[1_{\{P(\theta > \theta_0|x, Y) > \theta_T\}}|x] \\ &= \int 1_{\{P(\theta > \theta_0|x, Y) > \theta_T\}} dP(Y|x) \\ &= \sum_{i=0}^m P(Y = i|x) \times 1_{\{P(\theta > \theta_0|x, Y=i) > \theta_T\}} \\ &= \sum_{i=0}^m P(Y = i|x) \times 1_{\{B_i > \theta_T\}} \\ &= \sum_{i=0}^m P(Y = i|x) \times 1_i. \end{aligned}$$

La quantité B_i est la probabilité que la proportion θ est plus grande que θ_0 en donnant x réponses dans n patients des données actuelles et i réponses dans les m futurs patients. La comparaison de B_i à une valeur seuil θ_T donne un indicateur 1_i pour considérer l'efficacité du traitement à la fin de l'essai compte tenu des données actuelles et des résultats potentiels des $Y = i$.

La somme pondérée des indicateurs 1_i donne la probabilité prédictive de la conclusion d'un résultat positif, d'ici à la fin de l'essai, sur la base des informations cumulées dans l'étape actuelle. La PP élevée signifie que le traitement est susceptible d'être efficace,

d'ici à la fin de l'étude, compte tenu des données actuelles. Alors qu'un faible PP suggère que le traitement ne peut avoir une activité suffisante. Par conséquent, PP peut être utilisé pour déterminer si l'essai doit être arrêté prématurément en raison d'efficacité/de futilité ou poursuivi parce que les données actuelles ne sont pas encore concluantes. C'est pour cela la règle est définie en introduisant deux seuils sur PP , $(\theta_L$ et $\theta_U)$, contrairement, à la règle de décision de **Herson**.

Exemple 2.6. *Supposons qu'un chercheur envisage de scolariser un maximum de $N_{\max} = 40$ patients dans l'étude de phase II. À un moment donné, $x = 16$ réponses sont observées dans $n = 23$ patients. Que sera la $P(\text{proportion} > 60\%)$? l'hypothèse d'une vague distribution est une $Beta(0.6, 0.4)$ pour la proportion θ et soit Y le nombre de réponse dans $m = 17$ futures observations, donc $Y|X = 16 \sim Beta - B(17, 16.6, 7.4)$. Pour chaque valeur possible de $Y = i$, l'a posteriori conditionnelle de θ suit une distribution de $Beta$, $\theta|x, Y = i \sim Beta(16.6 + i, 24.4 - i)$. Dans cet exemple, on prend $\theta_T = 0.90$. Comme on peut le voir au Tableau (??), lorsque Y est incluse dans $\{0, \dots, 11\}$, le résultat $P(\text{proportion} > 0.60)$ varie de 0.0059 à 0.8314. Par conséquent nous concluons H_0 pour $Y \leq 11$. D'autre part, lorsque Y est incluse dans $\{12, \dots, 17\}$ le résultat $P(\text{proportion} > 0.60)$ varie de 0.9022 à 0.9989. Dans ce cas, au lieu de favoriser H_1 pour $y \geq 12$, la probabilité prédictive est la moyenne pondérée de l'indicateur d'un essai qui se déroule jusqu'à la fin de l'étude. Le calcul donne $PP = 0.5656$. Si $\theta_L = 0.10$, l'essai ne serait pas arrêté à cause de la futilité car PP est supérieur à θ_L , (Voir le programme 6 dans l'annexe).*

Comme cité précédemment, nous choisissons $\theta_U = 1.0$ car si le traitement est efficace, il y a peu de raison d'arrêter tôt l'essai, inscrire plus de patients au traitement actif sera plus recommander (Spiegelhalter et al., 2004). Traiter plus de patients jusqu'à ce que la taille de l'échantillon maximum soit atteinte (généralement, moins de 100) peut aussi augmenter la précision dans l'évaluation du taux de réponse. Etant donné N_{\max} , la question est : y-a-t-il des valeurs θ_L et θ_T qui rapportent les propriétés souhaitables du design ? par la recherche sur toutes les valeurs possibles de θ_L et θ_T , l'objectif est d'identifier les combinaisons de θ_L et θ_T , pour obtenir la puissance souhaitable dans la contrainte des taux d'erreur de type α et β pour avoir des gammes de θ_L et θ_T qui satisfait les contraintes. En changeant N_{\max} de petit à grand, le design avec la plus petit N_{\max} , qui contrôle, à la fois, les erreurs du 1^{ère} et du 2^{ème} espèce au niveau nominale est celle que nous choisissons. Cette idée est similaire à trouver le design minimax, c'est-à-dire, en minimisant la taille de l'échantillon maximum.

Le cadre de cette méthode (PP -design) permet à l'enquêteur de suivre l'essai

TABLE 2.2 – Calcul de la probabilité prédictive dans l'exemple 2.4.

| $Y = i$ | $Pr(Y = i x)$ | $B_i = P(\theta > 0.6 x, Y = i)$ | $1_{(B_i > 0.90)}$ |
|---------|---------------|----------------------------------|--------------------|
| 0 | 0.0000 | 0.0059 | 0 |
| 1 | 0.0000 | 0.0120 | 0 |
| 2 | 0.0001 | 0.0263 | 0 |
| 3 | 0.0006 | 0.0526 | 0 |
| 4 | 0.0021 | 0.0963 | 0 |
| 5 | 0.0058 | 0.1624 | 0 |
| 6 | 0.0135 | 0.2529 | 0 |
| 7 | 0.0276 | 0.3651 | 0 |
| 8 | 0.0497 | 0.4910 | 0 |
| 9 | 0.0794 | 0.6186 | 0 |
| 10 | 0.1129 | 0.7354 | 0 |
| 11 | 0.1426 | 0.8314 | 0 |
| 12 | 0.1587 | 0.9022 | 1 |
| 13 | 0.1532 | 0.9488 | 1 |
| 14 | 0.1246 | 0.9760 | 1 |
| 15 | 0.0811 | 0.9901 | 1 |
| 16 | 0.0381 | 0.9964 | 1 |
| 17 | 0.0099 | 0.9989 | 1 |

sans interruption ou par n'importe quelle taille de cohorte. Lee et Liu (2008) ont choisi de commencer le calcul de PP et la prise de décisions intérimaires après les 10 premiers patients qui ont été traités et évalués leurs statut de réponse. Bien que le choix de traiter 10 patients, au minimum, est arbitraire, un nombre minimum des patients est nécessaire pour fournir suffisamment d'informations pour obtenir une bonne estimation de l'efficacité du traitement et éviter de prendre des décisions prématurées basées sur de faux résultats d'un petit nombre de patients (Saville et al., 2014; Chow, 2010). Après 10 patients, le calcul du PP se fait continûment (c'est à dire, avec la taille de la cohorte de 1) pour surveiller l'efficacité du traitement. Si PP est petite, les données indiquent que le traitement peut ne pas être efficace. PP suffisamment petite (par exemple, $PP < \theta_L$) suggère que l'essai pourrait être arrêté précocement à cause de manque d'efficacité. Noter que PP peut être calculé pour n'importe quelle taille de la cohorte et à tout moment intérimaires. Un essai peut être arrêté, à tout moment, en raison de la toxicité excessive.

Remarque 2.1. *Les résultats sont obtenus à partir d'une calculatrice "Predictive Probability Design for Phase II Studies ("PID-535)" sur le site : https://biostatistics.mdanderson.org/softwaredownload/SingleSoftware.aspx?Software_Id=10.*

CHAPITRE 3

CHOIX DE L'A PRIORI DANS LA PLANIFICATION DES ESSAIS CLINIQUES DE PHASE II

Nous développons, dans ce chapitre, une procédure hybride Bayésienne-fréquentiste pour la planification expérimentale, basé sur l'inférence prédictive. La situation expérimentale considérée est la suivante :

Après avoir recueilli un premier échantillon de données E_1 , une première conclusion ou réponse à une question d'intérêt peut être obtenue par les méthodes usuelles, tests de signification par exemple ; on cherche alors dans une deuxième étape d'obtenir cette même conclusion si on recueille de nouvelle échantillon de donnée E_2 . Nous abordons dans ce chapitre, le problème de prédiction relative à un énoncé qui est basé seulement, sur les nouvelles observations (E_2). Le problème traité est typique des problèmes abordés en planification expérimentale lors du calcul des effectifs nécessaires à l'obtention d'une conclusion souhaitée.

Nous utilisons, à cet égard, les distributions prédictives relatives à des données ultérieures et nous verrons que l'inférence prédictive met en jeu les distributions usuelles dans les tests de signification de famille exponentielle présentées au chapitre 2.

Nous proposons la procédure hybride (fréquentiste Bayésienne) basée sur le concept de l'indice de satisfaction amélioré qui est une fonction décroissante de la p-valeur du deuxième échantillon, et nous envisageons, compte tenu des données disponibles, de calculer une prédiction de satisfaction de cet indice en fonction des observations précédentes (Merabet, 2004; Labdoui, 2015; Djeridi and Merabet, 2016).

3.1 Procédure de test dans un cadre hybride

Sachant que le contexte expérimental est constitué de deux expérimentations successives, les résultats $\omega' \in \Omega'$ et $\omega'' \in \Omega''$, sont en général effectués indépendamment. Leurs distributions, construites dans le cadre d'un modèle bien établi, dépendent d'un paramètre $\theta \in \Theta$, seul ω'' , est utilisé pour trouver la conclusion officielle de l'étude et pour déterminer la satisfaction de l'utilisateur notée $\phi(\omega'')$ (et sur le choix de la fonction décroissante L sur laquelle nous reviendrons dans la Section 3.1.2).

Notons par :

P_{Θ} : La distribution a priori sur Θ .

$P_{\Theta}^{\omega'}$: La distribution a posteriori sur Θ , basée sur le résultat de la première étape.

$P_{\Omega''}^{\theta}$: La vraisemblance de la deuxième étape.

$P_{\Omega''}^{\omega'}$: Distribution de probabilité sur Ω'' , conditionnée par le résultat de la première étape.

3.1.1 Test de Neymann-Pearson

Dans ce contexte expérimental, nous allons introduire le concept de l'indice de satisfaction relative à un test d'hypothèses où l'hypothèse nulle est de type $\theta \leq \theta_0$, dans un cadre où un tel test peut être construit en utilisant une fonction de test raisonnable.

Soit un modèle $(P_{\theta})_{\theta \in \Theta}$ et testons une hypothèse nulle Θ_0 versus une alternative Θ_1 , définie pour une application $\Psi (\Theta \rightarrow \mathbb{R})$: supposons qu'il existe un point t_0 , tel que :

$$\theta \in \Theta_0 \iff \Psi(\theta) \leq t_0;$$

Sinon, nous supposons qu'on a une application réelle $\xi (\Omega'' \rightarrow \mathbb{R})$ telle que :

$$\Psi (\theta_1) < \Psi (\theta_2) \Rightarrow \forall t; \quad P_{\theta_1} [\xi \leq t] \geq P_{\theta_2} [\xi \leq t].$$

Alors un test de niveau α , de Θ_0 versus $\Theta_1 (= \{\theta; \Psi (\theta) > t_0\})$ est définie par rejeter l'hypothèse nulle si le résultat expérimentale, ω'' , vérifie que $\xi (\omega'') > g(\alpha)$, où $g(\alpha)$ est $(1 - \alpha)$ -quantile de la distribution de ξ quand $\Psi(\theta_0) = t_0$.

De plus, la région critique, R , de ce test, est l'ensemble des observations ω'' , tel que $\xi (\omega'') > g(\alpha)$ et puisque les distributions de ξ sont stochastiquement croissantes, nous avons, pour tout θ_0 tel que $\Psi(\theta_0) = t_0$,

$$\forall \theta \in \Theta_0, \quad P_{\theta} (R) \leq P_{\theta_0} (R) \leq \alpha$$

En outre, si $\theta_1 \in \Theta_1$ et $\theta_2 \in \Theta_1$ avec $\Psi (\theta_1) < \Psi (\theta_2)$, alors, $P_{\theta_1} (R) \leq P_{\theta_2} (R)$ qui signifie que la fonction de puissance augmente avec $\Psi (\theta)$.

3.1.2 Indice de satisfaction

Cette notion trouve son origine dans des situations où le statisticien, qui effectue un test, "souhaite" détecter un effet significatif, i.e., pour rejeter l'hypothèse nulle H_0 . De manière correspondante, ce statisticien est surtout plus satisfait si, en fonction des résultats expérimentaux, cet effet semble être plus significatif.

En fait, on considère la satisfaction apportée par la seconde étape de l'expérimentation et la prédire à l'aide de la première, on fait les calculs avec deux étapes indépendantes mais c'est généralisable (en particulier la deuxième étape peut intégrer la première (Chapitre 4)). On définit, tout d'abord, l'indice fruste de Grouin puis on donnera l'indice amélioré de Merabet.

Indice fruste

Étant fixé α , soit un test de niveau α défini par la région critique $\Omega_1''^{(\alpha)}$. Un premier indice de satisfaction, celui étudié par Grouin (1994), est défini par :

$$\phi (\omega'') = 1_{\Omega_1''^{(\alpha)}} (\omega''),$$

à ω' fixé, la prédiction est, alors :

$$\pi (\omega') = P_{\Omega''}^{\omega'} \left(\Omega_1''^{(\alpha)} \right) = \int_{\Omega_1''^{(\alpha)}} P_{\Omega''}^{\theta} \left(\Omega_1''^{(\alpha)} \right) P_{\Theta}^{\omega'} (d\theta),$$

où $P_{\Omega''}^{\theta} \left(\Omega_1''^{(\alpha)} \right)$ est la valeur en θ de la puissance du test.

Le défaut de l'indice fruste est qu'il exprime une satisfaction en "tout ou rien" (significatif ou non).

Indice amélioré

Il est plus intéressant de prendre en compte jusqu'à quel niveau le résultat apparaît toujours significatif. C'est ce que les utilisateurs soulignent en donnant, à la fin de la procédure de test, non seulement la conclusion en termes de "tout-rien", mais aussi la valeur de seuil inférieure pour laquelle le résultat obtenu sera considéré comme significatif, c'est-à-dire du point de vue de la théorie du test, la *p - valeur*, et c'est dans notre cas

$$p = P_{\theta_0}(\xi > \xi(\omega'')).$$

Merabet (2004) a définie un indice amélioré (*IS*), pour ce test de niveau α , comme une application de l'ensemble de résultats dans \mathbb{R}_+ ; telle que :

- Prend la valeur 0 si on ne rejette pas l'hypothèse nulle, i.e., si $\xi(\omega'') \leq g(\alpha)$,
- Et si $\xi(\omega'') > g(\alpha)$, c'est une fonction décroissante $P_{\theta_0}(\xi > \xi(\omega''))$, qu'on notera $L(P_{\theta_0}(\xi > \xi(\omega''))) = L(1 - F_{\theta_0}(\xi(\omega'')))$, où F_{θ_0} est la distribution de ξ au frontière telle que $\Psi(\theta_0) = t_0$.

3.1.3 Prédiction de satisfaction

Étant fixé α , soit un test de niveau α défini par la région critique $\Omega_1''^{(\alpha)}$. Il est plus intéressant de prendre en compte jusqu'à quel niveau le résultat apparaît toujours significatif. Nous utilisons l'indice de satisfaction, qui est défini pour les tests Bayésiens, basé sur la même a priori P_{Θ} , comme suit :

$$\phi(\omega'') = \begin{cases} 0 & \text{si } \omega'' \in \Omega_0''^{(\alpha)} \\ 1 - \inf \left\{ \beta; \omega'' \in \Omega_1''^{(\beta)} \right\} & \text{si } \omega'' \in \Omega_1''^{(\alpha)} \end{cases} \quad (3.1)$$

Une situation standard est qu'il existe une application $\Psi(\Theta \rightarrow \mathbb{R})$ telle que $\Theta_0 = \{\theta; \Psi(\theta) \leq t_0\}$ et où il existe, aussi, $\xi(\Omega'' \rightarrow \mathbb{R})$ et $g(]0, 1[\rightarrow \mathbb{R})$ telles que

$$\Omega_1''^{(\alpha)} = \{\omega''; \xi(\omega'') \leq g(\alpha)\},$$

où $g(\alpha)$ est le $(1 - \alpha)$ -quantile de la distribution de ξ quand $\Psi(\theta_0) = t_0$. De plus, on suppose que la distribution de ξ sous $P_{\Omega''}^{\theta}$ dépend seulement de $\Psi(\theta)$ (qu'on notera $Q_{\Psi(\theta)}$) et est stochastiquement croissante, de telle sorte que ξ tend, plus en plus à prendre de grandes valeurs quand $\Psi(\theta)$ devient de plus en plus grand.

Un test de niveau α de Θ_0 est, alors, défini par rejeter l'hypothèse si le résultat expérimental, qui est ω'' , vérifie $\xi(\omega'') > g(\alpha)$.

Il semble, donc, naturel d'utiliser la p -valeur pour définir les indices de satisfaction nuls si aucun effet significatif n'est détecté et, dans le cas contraire, une fonction croissante de l'indicateur classique de signification. Dans ce cas :

$$p(\omega'') = P_{\theta_0}(\xi > \xi(\omega'')).$$

Un indice de satisfaction est, alors, défini naturellement par :

$$\phi(\omega'') = \begin{cases} 0 & \text{si } \xi(\omega'') \leq g(\alpha) \\ L(p(\omega'')) & \text{si } \xi(\omega'') > g(\alpha) \end{cases} \quad (3.2)$$

Où L est une fonction décroissante. Soit F_{θ_0} une distribution de ξ sur la frontière, *i.e.*, pour tout θ_0 tel que $\Psi(\theta_0) = t_0$, l'indice de satisfaction est défini par :

$$\phi(\omega'') = \begin{cases} 0 & \text{if } p(\omega'') \geq 1 - \alpha \\ L(1 - F_{\theta_0}(\omega'')) & \text{else} \end{cases} \quad (3.3)$$

La prédiction bayésienne est, alors, donnée par :

$$\begin{aligned} \pi(\omega') &= \int_{\Omega_1''(\alpha)} \phi(\omega'') P_{\Omega''}^{\omega'}(d\omega'') \\ &= \int_{\Theta} \left(\int_{\Omega_1''(\alpha)} \phi(\omega'') P_{\Omega''}^{\theta}(\omega'') \right) P_{\Theta}^{\omega'}(d\theta) \\ &= \int_{\{\omega''; \xi(\omega'') > g(\alpha)\}} L(1 - F_{\theta_0}(\omega'')) P_{\Omega''}^{\omega'}(d\omega''). \end{aligned} \quad (3.4)$$

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

On peut généraliser cette procédure à une famille d'indices limités définis par $L(p) = (1 - p)^l$ où $l \geq 0$. Il est préférable de choisir les indices limités car ils sont faciles à interpréter.

Remarque 3.1. 1. Si $l = 1$, $1 - \phi(\omega'')$ est la p -valeur.

2. Dans le cas où $l = 0$, on trouve la fonction indicatrice de la région critique qui est l'indice fruste de Grouin (Lecoutre et al. (1995)).

3.2 Applications

Djeridi and Merabet (2016) ont proposé de calculer explicitement ou numériquement la prédiction de satisfaction pour plusieurs modèles exponentiels : Binomial, Poisson , Gamma et Gaussien pour $L(p) = (1 - p)$ et pour un test de niveau, α dans le cas où l'hypothèse nulle est de la forme $H_0 : \theta \leq \theta_0$. Nous sommes guidées par une approche Bayésienne, (mais utilisant toujours un test fréquentiste) quand la distribution a priori du paramètre inconnu θ est non-informative dans la première étape (c.f. Chapitre 1). Dans ce cas, le choix de l'a priori de Jeffreys est acceptable pour ne pas exprimer des opinions particuliers sur les mérites du traitement (Ghosh et al. 2006 ; Robert, 2007).

3.2.1 Cas de la loi Binomiale

On suppose que toutes les variables aléatoires sont indépendantes X'_i ; ($1 \leq i \leq k$) et X''_j ; ($1 \leq j \leq n$) suivant la loi de Bernoulli $B(\theta)$, où θ est inconnu.

Alors $\omega' = \sum_{i=1}^k X'_i$ est de loi binomiale $B(k, \theta)$ et $\omega'' = \sum_{j=1}^n X''_j$ est de loi binomiale $B(n, \theta)$. Soit k le nombre de patients recrutés dans la première étape et n le nombre de patients recrutés dans la deuxième.

On suppose que θ suit une a priori non-informative au sens de Jeffreys ;

$$f(\theta) = \theta^{-\frac{1}{2}} (1 - \theta)^{-\frac{1}{2}} \propto \text{Beta} \left(\frac{1}{2}, \frac{1}{2} \right).$$

Par conséquent, la densité a posteriori de θ sachant ω' est la distribution Beta $\text{Beta}(\omega' + \frac{1}{2}, k - \omega' + \frac{1}{2})$, et la distribution prédictive de ω'' sachant ω' est donnée par

$$\nu(\omega'' | \omega') = \frac{C_n^{\omega''} \beta(\omega'' + \omega' + \frac{1}{2}, n + k - (\omega'' + \omega') + \frac{1}{2})}{\beta(\omega' + \frac{1}{2}, k - \omega' + \frac{1}{2})};$$

$$\text{où } \beta(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \text{ et } C_x^y = \frac{x!}{y!(x-y)!}.$$

L'indice de satisfaction (IS) est donné par

$$\phi(\omega'') = \begin{cases} 0 & \text{si } \omega'' < q_0 \\ \sum_{t=0}^{\omega''-1} C_n^t \theta_0^t (1 - \theta_0)^{n-t} & \text{si } \omega'' \geq q_0 \end{cases}$$

où

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

et la prédiction de satisfaction (PS) est :

$$\begin{aligned}\pi(\omega') &= \sum_{\omega''=q_0}^n (\theta_0^t (1-\theta_0)^{n-t}) \nu(\omega'' | \omega') \\ &= \sum_{\omega''=q_0}^n \left(\sum_{t=0}^{\omega''-1} C_n^t \theta_0^t (1-\theta_0)^{n-t} \right) \frac{C_n^{\omega''} \beta(\omega'' + \omega' + \frac{1}{2}, n+k - (\omega'' + \omega') + \frac{1}{2})}{\beta(\omega' + \frac{1}{2}, k - \omega' + \frac{1}{2})}.\end{aligned}$$

3.2.2 Échantillonnage de Poisson

On suppose que X'_i ; ($1 \leq i \leq n$) et X''_j ($1 \leq j \leq k$) sont des variables aléatoires *i.i.d.* suivent toutes la loi de Poisson $P(\theta)$, où θ est inconnu.

Alors $\omega' = \sum_{i=1}^n X'_i$ est de loi de Poisson $P(n\theta)$ et $\omega'' = \sum_{j=1}^k X''_j$ est de loi de Poisson $P(k\theta)$.

Si θ suit une a priori non-informative :

$$f(\theta) \propto \theta^{-1};$$

alors, l'a posteriori de θ sachant ω' est donnée par :

$$f(\theta|\omega') \propto \text{Gamma}(\omega', n);$$

et la prédictive de ω'' sachant ω' est :

$$\nu(\omega'' | \omega') = \frac{\Gamma(\omega' + \omega'')}{\Gamma(\omega')\omega''!} \left(\frac{n}{n+k}\right)^{\omega'} \left(\frac{k}{n+k}\right)^{\omega''}.$$

L'indice de satisfaction est, alors, exprimé par :

$$\phi(\omega'') = \begin{cases} 0 & \text{si } \omega'' < q_0 \\ \sum_{s=0}^{\omega''-1} e^{-k\theta_0} \frac{(k\theta_0)^s}{s} & \text{si } \omega'' \geq q_0 \end{cases}$$

où

$$q_0 = \inf \left\{ s; \sum_{s=0}^{s-1} e^{-k\theta_0} \frac{(k\theta_0)^s}{s} \geq 1 - \alpha \right\};$$

et la prédiction de satisfaction est donnée par :

$$\begin{aligned}\pi(\omega') &= \sum_{\omega''=q_0}^{\infty} \left(\sum_{s=0}^{\omega''-1} e^{-k\theta_0} \frac{(k\theta_0)^s}{s} \right) \nu(\omega'' | \omega') \\ &= \sum_{\omega''=q_0}^{\infty} \left(\sum_{s=0}^{\omega''-1} e^{-k\theta_0} \frac{(k\theta_0)^s}{s} \right) \frac{\Gamma(\omega' + \omega'')}{\Gamma(\omega')\omega''!} \left(\frac{n}{n+k}\right)^{\omega'} \left(\frac{k}{n+k}\right)^{\omega''}.\end{aligned}$$

3.2.3 Modèle Gamma

On suppose que toutes les v.a. $X'_i (1 \leq i \leq k)$ et $X''_j (1 \leq j \leq n)$ sont *i.i.d.* et suivent une loi de Gamma $G(p, \theta)$ où θ est inconnu et p connu. Alors, $\omega' = \sum_{i=1}^k X'_i$ est de loi Gamma $G(kp, \theta)$ et $\omega'' = \sum_{j=1}^n X''_j$ est de loi Gamma $G(np, \theta)$. Soit $K = kp$ et $N = np$.

Si θ suit la loi a priori non informative

$$f(\theta) \propto \theta^{-1}.$$

Alors l'a posteriori de θ sachant ω' est :

$$f(\theta|\omega') \propto \text{Gamma}(K, \omega');$$

et la densité prédictive de ω'' sachant ω' est :

$$\nu(\omega'' | \omega') = \frac{1}{\beta(N, K)} \frac{(\omega'')^{N-1} (\omega')^K}{(\omega'' + \omega')^{N+K}}.$$

L'indice de satisfaction est, alors, exprimé par :

$$\phi(\omega'') = \begin{cases} 0 & \text{si } \omega'' \leq q_0 \\ F(\omega'') = \int_0^{\omega''} \frac{\theta_0^N}{\Gamma(N)} t^{N-1} \exp(-\theta_0 t) dt & \text{si } \omega'' > q_0 \end{cases}$$

où

$$F(q_0) = 1 - \alpha$$

Finalement, la prédiction de satisfaction est donnée par :

$$\pi(\omega') = \int_{q_0}^{+\infty} \left[\int_0^{\omega''} \frac{\theta_0^N}{\Gamma(N)} t^{N-1} \exp(-\theta_0 t) dt \right] \frac{1}{B(N, K)} \frac{(\omega'')^{N-1} (\omega')^K}{(\omega' + \omega'')^{N+K}} d\omega'';$$

qui peut être évaluée numériquement.

3.2.4 Modèle Gaussien

S'appuyant sur le théorème de la limite centrale, les statisticiens de la première moitié du XIXe siècle faisaient presque toujours référence à la distribution normale. Il existe évidemment de nombreux phénomènes pour lesquels un modèle normal n'est pas applicable, mais il est encore largement utilisé, notamment en économétrie et dans des

domaines où l'approximation du théorème de la limite centrale peut être justifiée (fiabilité des particules, etc.). En fait, l'approximation normale est souvent justifiée pour des raisons asymptotiques (Robert, 2007). Il est donc intéressant d'étudier en détail cette distribution particulière du point de vue bayésien. Les calculs correspondants de la prédiction étant réalisables par les méthodes de Monte-Carlo (Chapitre 1).

Nous effectuons des observations indépendantes de la même variable aléatoire normale $N(\theta, \sigma^2)$. Dans tout ce qui suit, Φ et φ (resp. $T_{(n-1)}$ et $t_{(n-1)}$) indiquent la fonction de distribution cumulative et la densité de la loi $N(0, 1)$ respectivement (resp. de la distribution de Student $\tau_1(n-1, 0, 1)$).

Le premier résultat ; $\underline{x} = (x_1, x_2, \dots, x_n)$, est une série de n observations et le second résultat est une séries ; $\underline{y} = (y_1, y_2, \dots, y_k)$.

Pour des raisons évidentes d'exhaustivité, nous fonderons tous les calculs sur $x = \frac{1}{n} \sum_{i=1}^n x_i$ et $y = \frac{1}{k} \sum_{j=1}^k y_j$, de distributions $N(\theta, \sigma_1^2)$ et $N(\theta, \sigma_2^2)$, respectivement, où $\sigma_1^2 = \frac{\sigma^2}{n}$ et $\sigma_2^2 = \frac{\sigma^2}{k}$.

On suppose, ici, que σ^2 soit inconnue (aussi σ_1^2 et σ_2^2) (La situation où σ^2 est connue a été étudiée par Merabet (2004)). On choisi comme loi a priori pour (θ, σ^2) l'a priori non-informative $\pi(\theta, \sigma^2) \propto \frac{1}{\sigma}$ (Robert, 2007). Nous souhaitons tester une hypothèse nulle de type $\theta \leq \theta_0$.

Nous utilisons un test fréquentiste basé sur \underline{y} , dont la région critique $]q_0, +\infty[$, où $q_0 = \theta_0 + S'_2 u_\alpha^+$, u_α^+ indique le α -quartile supérieur de la distribution normale standard $N(0, 1)$: $\Phi(u_\alpha^+) = 1 - \alpha$ et $S'_2 = \frac{S_2}{\sqrt{2/k}}$.

La densité a posteriori associée à l'a priori $\pi(\theta, \sigma^2) \propto \frac{1}{\sigma}$ et appliquée au deuxième étape $\underline{y} = (y_1, y_2, \dots, y_k)$ est donnée par :

$$\theta \mid \sigma, y, S_2^2 \sim N\left(y, \frac{\sigma^2}{k}\right) \text{ et } \sigma^2 \mid y, S_2^2 \sim IG\left(\frac{k-1}{2}, \frac{S_2^2}{2}\right);$$

tel que : $y = \frac{1}{k} \sum_{j=1}^k y_j$ et $S_2^2 = \frac{1}{k} \sum_{j=1}^k (y_j - y)^2$. La densité prédictive de y sachant x est, alors :

$$f_x(y) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{2\pi}\Gamma\left(\frac{n-1}{2}\right)} \frac{1}{\frac{S_1}{\frac{2}{\sqrt{nk}}}} \left(\frac{(y-x)^2}{\frac{S_1^2}{\frac{nk}{n+k}}} + 1 \right)^{-\frac{n}{2}}$$

où $S_1^2 = \frac{1}{n} \sum_{i=1}^n (x_i - x)^2$. On identifie la densité du loi de Student $\tau_1 \left(n - 1, x, \frac{S_1}{\sqrt{\frac{nk}{n+k}}} \right)$.

Finalement la prédiction de satisfaction est :

$$\pi(x) = \int_{q_0}^{+\infty} \Phi \left(\frac{y - \theta_0}{\frac{S_2}{\sqrt{k}}} \right) \frac{\Gamma \left(\frac{n}{2} \right)}{\sqrt{\pi} \Gamma \left(\frac{n-1}{2} \right)} \frac{1}{\frac{S_1}{\sqrt{\frac{nk}{n+k}}}} \left(\frac{(y-x)^2}{\frac{S_1^2}{\frac{nk}{n+k}}} + 1 \right)^{-\frac{n}{2}} dy.$$

3.3 Exemples

3.3.1 Cas des résultats binaire

Dans le cadre d'essais de phase II, un investigateur envisage d'inclure un maximum de $N_{max} = 40$ patients dans l'étude. À un moment donné, on a observé $\omega' = 16$ réponses chez $k = 23$ patients. À la lumière de ce résultat, l'investigateur devrait-il poursuivre l'essai ou l'arrêter, en utilisant l'indice de satisfaction IS et sa prédiction PS , s'il inscrit tous les patients ?

Supposons la loi a priori $Beta(1/2, 1/2)$ pour θ avec un nombre de patient dans le future échantillon $n = 17$. Alors, ω'' suit une loi $beta - binomial(17, 16.5, 7.5)$.

Pour chaque valeur possible de $\omega'' = i$, l'a posteriori de θ est une distribution $beta$ telle que ;

$$\theta | \omega', \omega'' = i \sim Beta(1/2 + \omega' + i, 1/2 + N_{max} - \omega' - i).$$

Dans cet exemple

$$\theta | \omega' = 16, \omega'' = i \sim Beta(16.5 + i, 24.5 - i).$$

Pour $\theta_0 = 60\%$ et pour un niveau de signification $\alpha = 0.05$, on a $q_0 = 13$, l'indice de satisfaction IS et sa prédiction PS sont donnés dans la table (3.1) :

| ω'' | $\phi(\omega'')$ | ω' | $\pi(\omega')$ |
|------------|------------------|-----------|----------------|
| | | 13 | 0.13 |
| | | 14 | 0.21 |
| | | 15 | 0.29 |
| < 13 | 0 | 16 | 0.40 |
| 13 | 0.9536 | 17 | 0.52 |
| 14 | 0.9877 | 18 | 0.64 |
| 15 | 0.9979 | 19 | 0.76 |
| 16 | 0.9998 | 20 | 0.86 |
| 17 | 1.0000 | 21 | 0.93 |
| | | 22 | 0.98 |
| | | 23 | 0.997 |

Table 3.1 : *IS et PS* pour différentes valeurs de ω' et ω'' .

La prédiction de satisfaction est, alors, $\pi(\omega' = 16) = 0.40$ et si nous devons rejeter H_0 comme le propose les conceptions de Simon (Simon, 1989) et la conception de probabilité prédictive (PP-design) introduite par Lee and Liu (2008), notre satisfaction sera faible. Par contre, si nous considérons $\omega' = 20$, par exemple, alors $\pi(\omega' = 20) = 0.85$ et nous devons rejeter l'hypothèse nulle avec une grande satisfaction, et nous sommes plus satisfait de l'efficacité du traitement.

3.3.2 Simulation du modèle Gaussien par Monte Carlo

Pour effectuer le calcul de $\pi(x)$ en utilisant la méthode de Monte Carlo, nous utilisons un changement de variable pour le réécrire sous la forme suivante :

$$\pi(x) = [1 - T_{n-1}(a + \gamma u_\alpha^+)] \int_{-\infty}^{+\infty} \Phi\left(\frac{z-a}{\gamma}\right) \frac{t_{n-1}(z)}{1 - T_{n-1}(a + \gamma u_\alpha^+)} \mathbf{1}_{[a + \gamma u_\alpha^+, +\infty[} dz$$

où $a = \frac{\sqrt[n-1]{n-1}}{S'_1} (\theta_0 - x)$, $\gamma = \sqrt[n-1]{n-1} \frac{S'_2}{S'_1}$ avec $S'_1 = \frac{S_1}{\sqrt{\frac{nk}{(n+k)}}}$, $S'_2 = \frac{S_2}{\sqrt{k}}$ et $\frac{t_{n-1}(z)}{1 - T_{n-1}(a + \gamma u_\alpha^+)} \mathbf{1}_{[a + \gamma u_\alpha^+, +\infty[}$ est la densité de probabilité Q à partir de la fonction de distribution cumulative de la distribution de Student conditionnée par l'événement $[a + \gamma u_\alpha^+, +\infty[$.

La méthode de Monte Carlo consiste alors à approcher $\pi(x)$ par :

$$[1 - T_{n-1}(a + \gamma u_\alpha^+)] \left[\frac{1}{N} \sum_{i=1}^N \Phi \left(\frac{Z_i - a}{\gamma} \right) \right].$$

Telle que les Z_i sont N réalisations de la probabilité Q . Le tirage de Z_i se déroule de la manière suivante :

- U_i est tiré selon la répartition uniforme sur $[0, 1]$.
- $V_i = T_{n-1}(a + \gamma u_\alpha^+) + (1 - T_{n-1}(a + \gamma u_\alpha^+)) U_i$; c'est-à-dire, les V_i suivent la loi uniforme sur $[T_{n-1}(a + \gamma u_\alpha^+), 1]$.
- $Z_i = T_{n-1}^{-1}(V_i)$, c'est-à-dire, les Z_i suivent la distribution Q .

Voir le programme 7 sous R dans l'annexe.

Représentation des résultats et discussion

Nous allons trouver ci-dessous les courbes représentatives de π en fonction des observations $x = \frac{1}{n} \sum_{i=1}^n x_i$. Nous considérons, seulement, le cas où $\theta_0 = 0$ et dans le premier comme dans le deuxième échantillon, les observations sont de la même variance unitaire σ^2 , mais où les nombres peuvent varier, étant donné qu'une modification de θ_0 et de σ^2 ne peut être traduite qu'à partir un effet de translation. Nous avons considéré, les deux cas : $\alpha = 0.05$ et $\alpha = 0.01$. D'autre part, nous avons pris $k = 10$ ou 20 ou 30 pour $n = 10$.

Les graphes 3.1 ((a)- (b)) représentent la prédiction de satisfaction quand $\alpha = 0.05$. Le premier pour $\sigma = 1$ et le second pour $\sigma = 2$. Nous voyons, clairement, que la satisfaction augmente rapidement dans le premier cas que le deuxième et il est clair que lorsque k augmente, la satisfaction augmentera rapidement.

Tandis que, les graphes 3.2 (a- b) représentant la prédiction de satisfaction lorsque $\alpha = 0.01$ et $\alpha = 0.05$ pour $n = k = 10$ et $\sigma = 1$ ou $\sigma = 2$ respectivement. Nous pouvons faire la même remarque, mais les valeurs augmentent en s'éloignant rapidement de 0 dans (a) par rapport à (b), car en changeant le degré de signification, nous avons besoin de plus d'arguments pour rejeter l'hypothèse nulle. Cela traduit bien l'intérêt de la p-valeur dans l'indice de satisfaction puisque la région de rejet soit plus informative lorsque x est plus grand, ce qui donne de l'importance à l'indice indiqué.

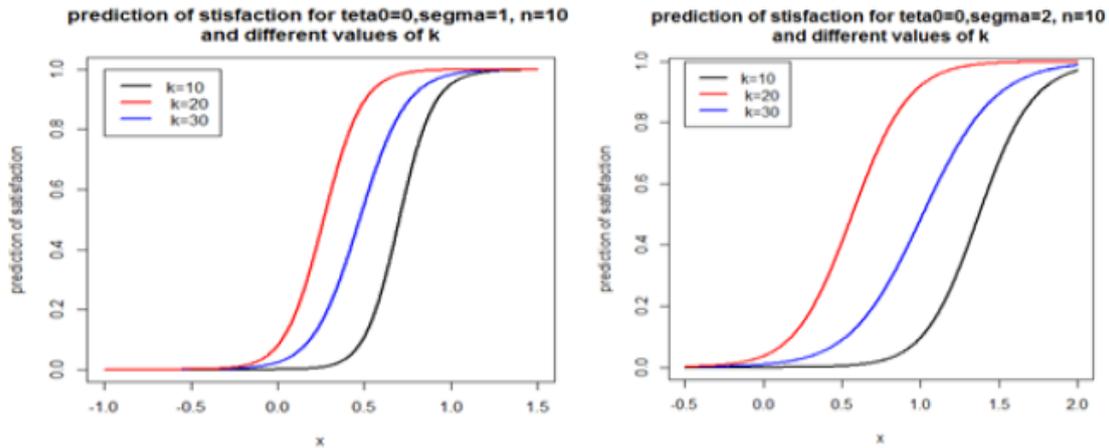


FIGURE 3.1 – Prédiction de satisfaction basé sur 5000 itérations pour $\alpha = 0.05$, $\sigma^2 = 1$ et 4 et différentes valeurs de k . Graphes avec un pas 0.001 pour x

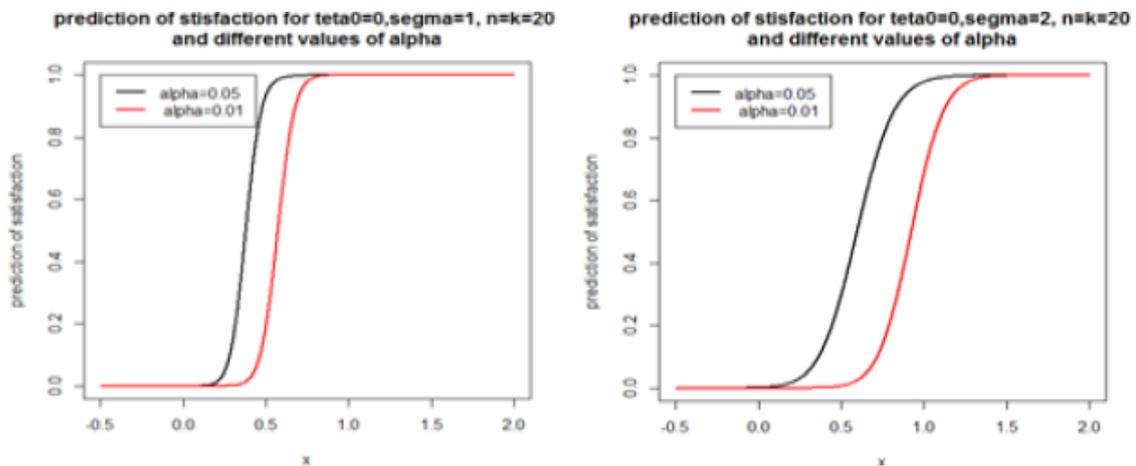


FIGURE 3.2 – Prédiction de satisfaction basé sur 5000 itérations pour $\alpha = 0.01$ et 0.05 et $\sigma^2 = 1$ et 4. Graphes avec un pas 0.001 pour x .

3.4 La règle d'arrêt PS

Cette procédure peut être utilisée, formellement, comme une règle d'arrêt pour les essais cliniques. Lors de la planification d'essai (qui peut aussi être généraliser à l'analyse intermédiaire (chapitre 4)), la résiliation a pour effet de rejeter H_0 si la prédiction de satisfaction PS au point θ_0 est élevée, formellement, si elle est supérieure à une constante spécifiée γ entre 0 et 1.

Les critères d'arrêt sont, généralement, spécifiques à chaque essai et incluent des considérations éthiques et commerciales, telles que des considérations de risque / bénéfice,

des ressources disponibles, du coût d'opportunité et du pouvoir statistique global. Dans le contexte de la surveillance intérimaire pour futilité, la prédiction de la satisfaction est naturellement attrayante, car elle répond directement à la question pertinente, à savoir si un essai est susceptible d'atteindre son objectif s'il est maintenu à la taille maximale prévue de l'échantillon.

Dans l'exemple (3.3.1), même si on prend $\gamma = 0.5$, alors que $\pi(\omega' = 17) = 0.52$ (Table (3.1)). Dans ce cas, la satisfaction sera forte et l'investigateur sera convaincu que traitement est prometteur et il peut collecter plus d'informations à ce sujet. De plus, cette a priori non-informative et cette règle d'arrêt réservent l'erreur de type I qui est $P(\omega' > 17 \mid \theta_0 = 0.60) = 0.05$. Dans ce cas, la puissance réelle de cette conception, pour une alternative de $\theta_1 = 0.80$, est de 0.84.

Les Tables (3.2) et (3.3) donnent les valeurs de la prédiction de satisfaction PS pour différentes valeurs de θ_0 , k , n , ω' pour $\alpha = 0.05$. Les valeurs plus élevées de θ_0 , par exemple, 80% ont un taux plus lent à converger vers 1 que les valeurs plus petites, par exemple 20% (voir la Figure 3.3 (a)), parce que nous avons besoin de plus d'arguments pour rejeter l'hypothèse nulle. En outre, pour l'erreur de type I, le rejet de l'hypothèse nulle sera difficile si nous augmentons le niveau de signification (Figure 3.3 (b)).

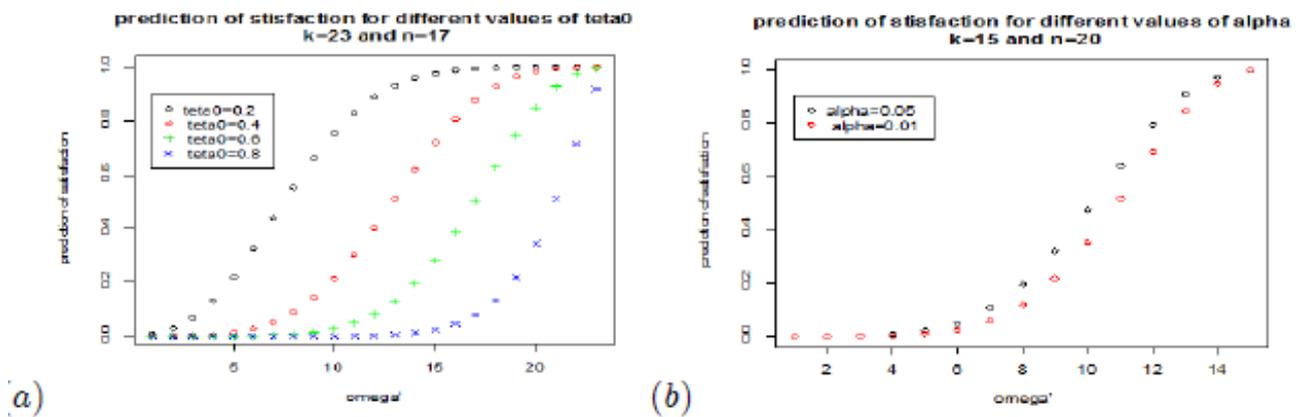


FIGURE 3.3 – La PS augmente lentement si θ_0 croit (a) et si le niveau de signification est plus grand (b)

Une critique adressée à cette procédure, qu'il ne nous donne pas d'informations bayésiennes directes sur θ , comme pourrait être fourni par un intervalle crédible. En outre, pour prouver l'efficacité du traitement, nous devrions avoir une grande probabilité de succès. Dans l'exemple (3.3.1), l'essai sera interrompu si moins de 17 succès / 23 (74%) sont observés.

| ω' | $\theta_0 = 0.2$ | $\theta_0 = 0.4$ | $\theta_0 = 0.6$ | $\theta_0 = 0.8$ |
|-----------|------------------|------------------|------------------|------------------|
| 1 | 0.0078 | 0.0001 | 0.0000 | 0.0000 |
| 2 | 0.0282 | 0.0004 | 0.0000 | 0.0000 |
| 3 | 0.0688 | 0.0017 | 0.0000 | 0.0000 |
| 4 | 0.1332 | 0.0051 | 0.0001 | 0.0000 |
| 5 | 0.2203 | 0.0127 | 0.0004 | 0.0000 |
| 6 | 0.3250 | 0.0271 | 0.0012 | 0.0000 |
| 7 | 0.4392 | 0.0517 | 0.0030 | 0.0000 |
| 8 | 0.5539 | 0.0898 | 0.0068 | 0.0001 |
| 9 | 0.6609 | 0.1439 | 0.0141 | 0.0002 |
| 10 | 0.7543 | 0.2152 | 0.0271 | 0.0006 |
| 11 | 0.8308 | 0.3025 | 0.0485 | 0.0013 |
| 12 | 0.8897 | 0.4027 | 0.0818 | 0.0029 |
| 13 | 0.9321 | 0.5101 | 0.1303 | 0.0062 |
| 14 | 0.9609 | 0.6179 | 0.1968 | 0.0125 |
| 15 | 0.9790 | 0.7191 | 0.2827 | 0.0240 |
| 16 | 0.9896 | 0.8073 | 0.3867 | 0.0444 |
| 17 | 0.9953 | 0.8783 | 0.5045 | 0.0788 |
| 18 | 0.9984 | 0.9305 | 0.6282 | 0.1343 |
| 19 | 0.9993 | 0.9651 | 0.7473 | 0.2193 |
| 20 | 0.9998 | 0.9852 | 0.8504 | 0.3428 |
| 21 | 1.0000 | 0.9951 | 0.9279 | 0.5096 |
| 22 | 1.0000 | 0.9989 | 0.9757 | 0.7132 |
| 23 | 1.0000 | 0.9999 | 0.9967 | 0.9199 |

TABLE 3.1 – PS pour différentes valeurs de θ_0 lorsque $\alpha = 0.05, k = 23$ et $n = 17$.

| ωl | $\theta_0 = 0.2$ | $\theta_0 = 0.5$ | $\theta_0 = 0.7$ |
|------------|------------------|------------------|------------------|
| 1 | 0.0274 | 0.0001 | 0.0000 |
| 2 | 0.0919 | 0.0005 | 0.0000 |
| 3 | 0.2038 | 0.0022 | 0.0001 |
| 4 | 0.3533 | 0.0076 | 0.0003 |
| 5 | 0.5178 | 0.0216 | 0.0013 |
| 6 | 0.6722 | 0.0515 | 0.0043 |
| 7 | 0.7984 | 0.1067 | 0.0122 |
| 8 | 0.8887 | 0.1951 | 0.0304 |
| 9 | 0.9456 | 0.3197 | 0.0680 |
| 10 | 0.9796 | 0.4735 | 0.1373 |
| 11 | 0.9917 | 0.6387 | 0.2508 |
| 12 | 0.9976 | 0.7901 | 0.4144 |
| 13 | 0.9995 | 0.9045 | 0.6167 |
| 14 | 0.9999 | 0.9711 | 0.8197 |
| 15 | 1.0000 | 0.9967 | 0.9649 |

TABLE 3.2 – PS pour différentes valeurs de θ_0 lorsque $\alpha = 0.05, k = 15$ et $n = 20$.

CHAPITRE 4

PS DESIGNS POUR SURVEILLER LES ESSAIS CLINIQUES EN PLUSIEURS ÉTAPES

L'utilisation d'essais adaptatifs pourrait considérablement améliorer l'efficacité du développement de médicaments ; l'intégration de l'approche bayésienne est un pas de plus dans cette direction (Chang, 2008 ; Chow et al., 2012). Les conceptions d'essais adaptatifs bayésiens constituent un outil puissant pour rationaliser le processus d'apprentissage séquentiel, prédire d'autres résultats et synthétiser les preuves sur différentes ressources. Les résultats obtenus avec les approches bayésiennes sont plus faciles à interpréter et plus utiles à la prise de décision (Jennison et Turnbull, 2000 ; *EveritetPickles*, 1999). Pour une discussion comparant les conceptions bayésienne et fréquentiste dans le contexte des essais cliniques, voir Berry (1985), Stangl et Berry (1998), Spiegelhalter et al. (1994), et Lee et Liu (2008).

Dans de nombreuses expériences ou enquêtes, au cours desquelles les données sont accumulées de manière constante sur une période donnée, il est judicieux de surveiller les résultats au fur et à mesure de leurs apparitions afin de prendre des mesures telles que l'arrêt précoce ou une modification du plan d'étude. Des méthodes statistiques séquentielles ont été développées afin d'obtenir des avantages. Pour un essai avec un

résultat positif, un arrêt précoce signifie qu'un nouveau produit peut être exploité plus tôt. Dans le cas contraire, un arrêt précoce garantit que les ressources ne sont pas gaspillées; Gould (1983) a qualifié cette situation d'"abandonner une cause perdue". Les méthodes séquentielles permettent généralement d'économiser la taille d'échantillon, le temps et le coût par rapport aux procédures standard d'échantillon fixe.

Dans ce chapitre, nous allons nous concentrer sur les essais de phase II à un seul bras avec des données binaires. Des conceptions à plusieurs étapes sont souvent mises en œuvre dans de tels contextes afin d'accroître l'efficacité de l'étude en permettant un arrêt précoce du traitement s'il est considéré inefficace. Dans ces modèles, si aucune réponse n'est observée aux étapes précédentes, l'essai est arrêté à cause de futilité.

4.1 Apeçu sur les conceptions bayésiennes de phase II

Plusieurs conceptions d'essais cliniques de phase II sont proposées dans la littérature statistique et la plupart d'entre elles se déroulent en deux étapes. Dans un cadre fréquentiste, les designs en deux étapes de Simon sont les plus populaires (Simon, 1989). Ils permettent un arrêt précoce dû au futilité. Si le traitement expérimental fonctionne bien, davantage de patients sont traités dans la deuxième étape, ce qui permet également une meilleure estimation du taux de réponse et de la toxicité du nouveau traitement avant le lancement d'un essai de phase III. Sous l'hypothèse nulle, la conception optimale en deux étapes minimise la taille attendue de l'échantillon, tandis que la conception minimax minimise la taille maximale de l'échantillon. Les deux conceptions sont soumises aux contraintes des taux d'erreur de type I et de type II (cf. Chapitre 2).

L'approche bayésienne peut être utilisée pour déterminer la meilleure stratégie disponible à l'instant. On peut l'utiliser pour surveiller les essais, prévoir les résultats, anticiper les problèmes et suggérer des solutions précoces. Zohar et al. (2008) soulignent que des approches ou des conceptions bayésiennes ont été proposées pour renforcer les essais cliniques car ils prennent en compte les informations précédentes sur la quantité d'intérêt ainsi que les données accumulées au cours d'un essai. Dans ce contexte, diverses approches ou conceptions bayésiennes de phase II ont été proposées pour des essais cliniques à un seul bras. Thall et Simon (1994) ont mis au point une conception avec un contrôle continu jusqu'à atteindre une probabilité a posteriori élevée qu'un médicament

soit prometteur ou non, ou jusqu'à atteindre la taille maximale de l'échantillon. Toujours sur la base de la probabilité a posteriori, Heitjan (1997) a préconisé l'utilisation de la «probabilité de persuasion» comme critère cohérent pour déterminer si le médicament est prometteur ou non. Tan and Machin (2002) ont construit des conceptions bayésiennes à deux étapes qui imitent les alternatives fréquentistes en calibrant les paramètres de conception sur la base de l'approche de probabilité a posteriori. Des approches similaires basées sur des a priori informatives ou des mélanges des a priori informatives ont également été rapportées (Mayo et Gajewski, 2004 ; 2006).

Les procédures prédictives bayésiennes ont largement contribué à l'inférence et l'analyse des données. Dans cette perspective, les probabilités prédictives bayésiennes sont particulièrement utiles pour communiquer avec l'enquêteur. Ils leur donnent une méthode très attrayante pour répondre à des questions essentielles telles que : «Compte tenu des données actuelles, quelle est la probabilité que le résultat final soit en quelque sorte concluant ou au contraire non concluant ? » Cette question est inconditionnelle car elle nécessite la prise en compte de toutes les valeurs possibles des paramètres. Alors que la pratique fréquentiste traditionnelle ne répond pas à ces questions, la probabilité prédictive (PP) leur donne des réponses directes et naturelles. En particulier, les procédures prédictives peuvent être utilisées pour illustrer les effets de la planification d'une expérience avec un très petit échantillon ; et d'aider à la décision d'abandonner tôt une expérience (Lecoutre, 2001). En ce qui concerne la conception des essais, Herson (1979) a été le premier à proposer l'approche de la probabilité prédictive pour la conception d'essais cliniques de phase II avec des résultats dichotomiques. La taille maximale de l'échantillon et la région critique proposées proviennent d'un plan fréquentiste à une étape. Le choix des valeurs limites pour la probabilité prédictive est quelque peu arbitraire et ne garantit pas les taux d'erreur de type I et de type II. Alors que la conception de probabilité prédictive proposée par Lee et Liu (2008) recherche les paramètres de conception à l'intérieur, les contraintes données telles que la taille et la puissance du test peuvent être garanties. (Spiegelhalter et al. (1986), Spiegelhalter et Freedman (1986), Grieve (1991), Johns et Andersen (1999), 1986, Gould (2005), Lee et Liu (2008))

L'objectif principal de ce chapitre, est de fournir une procédure hybride bayésienne-fréquentiste pour les conceptions à plusieurs étapes afin de tester l'efficacité d'un nouveau traitement. Cette procédure est basée sur le concept d'indice de satisfaction (Merabet et al., 2017) ; qui est une fonction décroissante de la p-valeur, nous envisageons, compte tenu

des données disponibles, de calculer une satisfaction prédite de cet indice en considérant les observations précédentes en utilisant l'approche bayésienne. De nombreux auteurs ont préconisé l'utilisation d'indices dans de telles situations comme Lecoutre et al. (1995), Merabet et Labdaoui (2015), Djeridi et Merabet (2016) en raison de la simplicité et de la souplesse dont ils disposent pour mesurer le degré de satisfaction dans le cas d'obtenir un résultat significatif. Cet indice est utilisé comme règle d'arrêt pour la conception d'essais cliniques de phase II.

4.2 Indice de satisfaction dans l'analyse intermédiaire

Brown et al. (1987), Grieve (1991) and Merabet et al. (2017) ont souligné que les approches ou les conceptions prédictives bayésiennes sont idéales pour les essais exploratoires dans la mesure où elles tiennent compte des informations précédentes sur la quantité d'intérêt ainsi que des données accumulées au cours de l'essai. Dans ce contexte, Merabet et al. (2017) ont mis au point une procédure hybride bayésienne fréquentiste pour les essais cliniques à un seul bras basée sur la prédiction de la satisfaction (PS) en faisant la moyenne de l'indice de satisfaction par rapport au probabilité prédictive sur les données complètes conditionnées par le résultat de la première étape. Une valeur élevée de la prévision de satisfaction indique qu'il est peu probable que le test de référence accepte H_0 , compte tenu des données accumulées. Cela nous a amenés à suggérer une règle formelle selon laquelle l'arrêt se produit à la première étape en acceptant H_0 si la prédiction de satisfaction est inférieure à un seuil prédéterminé γ dans un test unilatéral. La valeur de γ doit être comprise entre 0.5 et 1. Les critères d'arrêt spécifiques sont généralement propres à chaque essai et incluent des considérations éthiques et commerciales, telles que la considération risque / bénéfice, les ressources disponibles, le coût d'opportunité et le pouvoir statistique global déterminé par les enquêteurs de l'étude.

4.2.1 Indice de Satisfaction pour deux étapes

Cette notion trouve son origine dans des situations où le statisticien, qui effectue un test, «souhaite» détecter un effet significatif, c'est-à-dire, rejeter l'hypothèse nulle H_0 . De manière correspondante, ce statisticien est notamment plus satisfait si, en fonction des résultats expérimentaux, cet effet semble être plus significatif.

Gardant à l'esprit que le contexte expérimental consiste en deux expériences successives, avec les résultats $x \in \mathcal{X}$ et $y \in \mathcal{Y}$, qui sont généralement réalisées de manière indépendantes. On suppose que les distributions des deux résultats dépendent d'un seul paramètre $\theta \in \Theta$. Considérons $z = (x, y)$, qui est utilisé pour établir la conclusion finale de l'étude et déterminer la satisfaction de l'utilisateur, ce qui est noté $\phi(z)$. Mais, sur la base du résultat x du première étape de l'essai, il est utile de prévoir quelle sera la satisfaction après la deuxième étape. Dans notre étude, cette prévision s'effectue dans un contexte bayésien, c'est-à-dire, sur la base du choix d'une probabilité a priori sur Θ .

Étant fixé α , soit un test de niveau α défini par une région critique $R^{(\alpha)}$. Il est plus intéressant de prendre en compte à quel niveau les résultats seront toujours significatifs. Dans ce cas l'indice de satisfaction a été défini par Merabet et al. (2017), comme suit :

$$\phi(z) = \begin{cases} 0 & \text{si } z \notin R^{(\alpha)} \\ 1 - p(z) & \text{si } z \in R^{(\alpha)} \end{cases} \quad (4.1)$$

où $p(z) = \inf \{ \beta; z \in R^{(\beta)} \}$ est la p -valeur de test. Compte tenu des contraintes économiques et / ou éthiques, il n'est souvent pas intéressant de poursuivre un essai après une analyse intermédiaire si nous pouvons raisonnablement prédire que cela conduira à une conclusion significative.

De nombreux auteurs ont fortement préconisé l'utilisation de probabilités prédictives dans la prise de décisions reposant sur l'accumulation de données d'essais. L'utilisation d'une approche prédictive permet au chercheur de prendre en compte l'incertitude des résultats expérimentaux non observés, en évitant de conditionner un résultat expérimental virtuel.

Il est plus intéressant d'anticiper le degré de satisfaction par prédiction lorsqu'on doit mener une expérience en deux étapes :

- Un premier résultat x détermine si nous continuons ou non l'expérience,
- Si l'expérimentateur est très satisfait et que nous poursuivons effectivement l'expérience, alors le résultat z de la première et de la deuxième étape servira de base au test.

Soit $p(z|x)$, la probabilité prédictive de z conditionnellement à x , la prédiction de satisfaction (PS) est alors donnée par :

$$\pi(x) = E(\phi(z)|x) = \int \phi(z) p(z|x) dz. \quad (4.2)$$

4.2.2 Choix de la distribution a priori dans le cadre séquentiel

L'analyse bayésienne est guidée par la distribution a priori, de sorte que sa source et son utilisation présentent de nombreux défis (cf. Chapitre 1). Spiegelhalter et al. (2004) ont clarifié un certain nombre d'idées fausses qui peuvent apparaître. En particulier, une a priori n'est pas nécessairement spécifié avant l'expérience. Malgré le nom *a priori* suggérant une relation temporelle, il est tout à fait possible de choisir la distribution a priori après avoir pris connaissance des résultats de l'étude, dans la mesure où elle vise simplement à résumer une incertitude raisonnable compte tenu des données externes provenant de l'étude en question (Spiegelhalter et al., 2004). Les paramètres des a priori conjugués sont calculés à l'aide des données disponibles par de nombreuses procédures (Shao et al., 2007 ; Herson, 1979, ; Robert, 2007). Dans le cas des a priori non informatifs, l'a priori de Jeffreys, qui est basé sur les données disponibles, est largement utilisée.

Le design proposé est basé sur la *p-value* traditionnelle qui dépend des échantillons les «plus extrêmes» que les données observées (sous l'hypothèse nulle). Dans le cas de distributions discrètes, cela dépend de l'inclusion des données observées ou non. Par exemple, le test binomial habituel est conservateur, mais si les données observées sont exclues, le test devient libéral. Évidemment, dans ce cas, le choix d'une distribution a priori non informative ne peut pas éviter les conventions. Mais le choix particulier d'une telle distribution est la contrepartie exacte de l'arbitraire impliqué dans l'approche fréquentiste. (Lecoutre, 2001).

L'utilisation séquentielle du théorème de Bayes signifie que la distribution a posteriori après la première étape de l'étude devient simplement l'a priori de la seconde, et que la distribution a posteriori finale se présente de la même manière. Supposons que nous observons des données en deux ou plusieurs étapes, disons x suivi de y . En utilisant l'observation du première étape, la distribution a posteriori est donnée par :

$$f(\theta|x) \propto f(x|\theta) f(\theta), \quad (4.3)$$

où $f(\theta)$ est la densité a priori de θ et $f(x|\theta)$ est la vraisemblance du premier échantillon. Cette distribution a posteriori devient l'a priori de la prochaine utilisation du théorème de Bayes. Ainsi, après l'observation de l'échantillon suivant y , le conditionnement de l'a posteriori par toutes les données, c'est-à-dire, $f(\theta|x, y)$, est donné comme suit :

$$f(\theta|x, y) \propto f(y|\theta, x) f(\theta|x) \propto f(y|\theta) f(\theta|x), \quad (4.4)$$

puisque X est indépendante conditionnellement à Y sachant θ et donc $f(\theta|x)$ devient sim-

plement l'a priori pour une mise à jour bayésienne standard en utilisant la vraisemblance $f(y|\theta)$.

4.3 Utilisation séquentielle de PS pour les réponses binaires

Dans cette section, la procédure indiquée dans (4.2.1) sera généralisé pour K étapes. Dans le cadre de la prédiction de satisfaction, on peut évaluer les chances que l'essai montre un résultat concluant à la fin de l'étude, compte tenu des informations actuelles. Ensuite, la décision de poursuivre ou d'arrêter l'essai peut être prise en fonction de la force de la prédiction de satisfaction.

Dans ce cadre de travail, un nombre maximal de groupes, K , et une taille de groupe m_k ; $k = \overline{1, K}$ sont choisis. Les données accumulées sont analysées après chaque groupe de m_k réponses. Pour un test de seuil α , où l'hypothèse nulle est de type $\theta \leq \theta_0$, nous calculons l'indice de satisfaction et sa prédiction dans le cas du modèle binomial en raison d'importance de ce dernier en sciences expérimentales, en particulier pour les essais cliniques lorsque la distribution a priori du paramètre inconnu est non informative dans un cadre séquentiel (cette procédure est valable même dans le cas d'une conjuguée naturelle).

Supposons que toutes les *v.a.* Soient indépendantes $X_{l_1}^1$ ($1 \leq l_1 \leq m_1$), $X_{l_2}^2$ ($1 \leq l_2 \leq m_2$), ..., $X_{l_K}^K$ ($1 \leq l_K \leq m_K$) et qui suivent une distribution de Bernoulli $B(\theta)$, où θ est inconnu.

Alors $x_k = \sum_{l_k=1}^{m_k} X_{l_k}^k$ (pour $k = 1, \dots, K$) suit une binomiale $B(m_k, \theta)$ qui représente respectivement le nombre de résultats positifs dans les K groupes.

Au début, nous pouvons choisir une a priori uniforme *Beta* (1, 1) pour θ . Ce choix semble raisonnable dans la pratique, car souvent aucune information n'est disponible sur les paramètres; cette technique bayésienne spécifie donc un état d'ignorance. L'essai est surveillé continuellement, afin que, la probabilité a posteriori bayésienne soit actualisée après avoir observé chaque nouveau résultat. Les décisions sont prises de manière adaptative tout au long de l'essai jusqu'à ce que la taille maximale de l'échantillon soit atteinte ($N_{max} = \sum_{k=1}^K m_k$).

À tout moment au cours de l'essai, sur la base des données cumulées, il est possible d'arrêter l'essai et d'affirmer que le médicament expérimental est prometteur ou non, ou de poursuivre l'essai en raison d'un manque de preuves convaincantes permettant de prendre une décision.

L'approche adoptée, ici, a été formulée en calculant la prédiction de la satisfaction au moment d'une analyse intermédiaire ($k^{\text{ème}}$ étape). L'algorithme suivant suppose que la règle de rejet d'hypothèse nulle, lors de l'analyse intermédiaire, est que la prédiction de satisfaction est un nombre réel compris entre 0.5 et 1, et qui a été choisie par les investigateurs de l'étude.

1. Après la première étape :

L'a priori est $\theta \sim \text{Beta}(a, b)$.

L'a posteriori est $\theta | x_1 \sim \text{beta}(x_1 + a, m_1 - x_1 + b)$.

Nous voulons planifier un deuxième échantillon de taille $N_{max} - m_1$ de résultat y . La distribution prédictive de la deuxième étape compte tenu du premier résultat x_1 est

$$\nu(y | x_1) = \frac{C_{N_{max}-m_1}^y \beta(y + x_1 + a, N_{max} - (y + x_1) + b)}{\beta(x_1 + a, m_1 - x_1 + b)},$$

tel que ; $\beta(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ and $C_x^y = \frac{x!}{y!(x-y)!}$.

Après la deuxième étape : $z = (x_1, y)$ et la région de rejet est définie par $R^{(\alpha)} = \{z; z \geq q_2\}$ avec $q_2 = \inf \{u, Pr(Z \geq u | \theta_0) \leq \alpha\}$, l'indice de satisfaction sera donc :

$$\phi(z) = \Phi_{\theta_0}^B(z) I_{z > q_2};$$

où $\Phi_{\theta_0}^B(z)$ est la fonction de distribution cumulative d'une distribution binomiale (N_{max}, θ_0) .

Alors la prédiction de satisfaction est :

$$\begin{aligned} \pi(x_1) &= \sum_{y=(q_2-x_1)^+}^{N_{max}-m_1} (\phi(x_1 + y)) \nu(y | x_1) \\ &= \sum_{y=(q_2-x_1)^+}^{N_{max}-m_1} (\Phi_{\theta_0}^B(x_1 + y)) \frac{C_{N_{max}-m_1}^y \beta(y + x_1 + a, N_{max} - (y + x_1) + b)}{\beta(x_1 + a, m_1 - x_1 + b)}; \end{aligned}$$

où $u^+ = \max(0, u)$.

Si la prédiction de satisfaction $\pi(x_1)$ est inférieur à un seuil prédéfini γ , l'étude est terminée et le médicament est abandonné, sinon l'essai est poursuivi.

2. Supposons que nous décidions de poursuivre l'essai et à la fin de la deuxième étape nous avons obtenu deux résultats x_1 et x_2 . Nous devons donc plani-

fier un troisième échantillon y de taille $N_{max} - m_1 - m_2$ (en supposant toujours que la prédiction de satisfaction est un nombre réel compris entre 0.5 et 1 pour inscrire d'autres patients dans l'étude).

Ainsi, la distribution a priori sera actualisée, à la lumière des analyses intermédiaires, à savoir l'a priori sera $\theta|x_1 \sim Beta(x_1 + a, m_1 - x_1 + b)$.

L'a posteriori, qui est la distribution de $\theta|x_1, x_2$ calculée par la formule (4.4) est donc :

$$f(\theta|x_1, x_2) \propto Beta(x_1 + x_2 + a, m_1 + m_2 - (x_1 + x_2) + b).$$

La distribution prédictive de la troisième étape sachant les résultats x_1 et x_2 de la première et de la deuxième étape, respectivement, est donné par :

$$\begin{aligned} \nu(y|x_1, x_2) &= \int f(y|\theta)f(\theta|x_1, x_2)d\theta \\ &= \frac{C_{N_{max}-\sum_{i=1}^2 m_i}^y \beta(y + \sum_{i=1}^2 x_i + a, N_{max} - y - (\sum_{i=1}^2 x_i) + b)}{\beta(\sum_{i=1}^2 x_i + a, \sum_{i=1}^2 m_i - (\sum_{i=1}^2 x_i) + b)}. \end{aligned}$$

L'indice de satisfaction après la troisième étape (pour $z = (x_1, x_2, y)$) est :

$$\phi(z) = \Phi_{\theta_0}^B(z)I_{z > q_3},$$

où

$$q_3 = \inf \{u; Pr(Z \geq u|\theta_0) \leq \alpha\}.$$

Finalement, la prédiction de satisfaction pour cette troisième étape sera comme suit :

$$\begin{aligned} \pi(x_1, x_2) &= \sum_{y=(q_3-\sum_{i=1}^2 x_i)^+}^{N_{max}-\sum_{i=1}^2 m_i} (\phi(x_2 + x_1 + y)) \nu(y | x_1, x_2) \\ &= \sum_{y=(q_3-\sum_{i=1}^2 x_i)^+}^{N_{max}-\sum_{i=1}^2 m_i} \left(\Phi_{\theta_0}^B\left(\sum_{i=1}^2 x_i + y\right) \right) \\ &\times \frac{C_{N_{max}-\sum_{i=1}^2 m_i}^y \beta(y + \sum_{i=1}^2 x_i + a, N_{max} - y - (\sum_{i=1}^2 x_i) + b)}{\beta(\sum_{i=1}^2 x_i + a, \sum_{i=1}^2 m_i - (\sum_{i=1}^2 x_i) + b)}. \end{aligned}$$

3. Pour une analyse intermédiaire $k - 1$; $k = 2, \dots, K$, nous avons des données intérimaires x_1, \dots, x_{k-2} . La même procédure peut être réalisée si la prédiction de satisfaction $\pi(x_1, \dots, x_{k-2})$ est toujours dans l'intervalle $[0.5, 1]$.

L'a priori de cette étape est choisi comme étant l'a posteriori de l'étape précédente $f(\theta) = f(\theta|x_1, \dots, x_{k-2}) \propto \text{beta}(\sum_{i=1}^{k-2} x_i + a, \sum_{i=1}^{k-2} m_i - \sum_{i=1}^{k-2} x_i + b)$.

En utilisant l'indépendance des échantillons, la distribution a posteriori de $\theta|x_1, \dots, x_{k-1}$ est donnée par

$$\begin{aligned} f(\theta|x_1, \dots, x_{k-1}) &\propto f(x_{k-1}|\theta)f(x_{k-2}|\theta)\dots f(x_2|\theta)f(x_1|\theta)f(\theta) \\ &\propto \text{beta}\left(\sum_{i=1}^{k-1} x_i + a, \sum_{i=1}^{k-1} m_i - \sum_{i=1}^{k-1} x_i + b\right). \end{aligned}$$

La distribution prédictive de la $k^{\text{ème}}$ étape sachant les données x_1, \dots, x_{k-1} sera :

$$\begin{aligned} \nu(y|x_1, \dots, x_{k-1}) &= \int f(y|\theta)f(\theta|x_1, \dots, x_{k-1})d\theta \\ &= \frac{C_{N_{max}-\sum_{i=1}^{k-1} m_i}^y \beta\left(y + \sum_{i=1}^{k-1} x_i + a, N_{max} - y - \left(\sum_{i=1}^{k-1} x_i\right) + b\right)}{\beta\left(\sum_{i=1}^{k-1} x_i + a, \sum_{i=1}^{k-1} m_i - \sum_{i=1}^{k-1} x_i + b\right)}. \end{aligned}$$

pour la $k^{\text{ème}}$ étape; telle que $z = (x_1, \dots, x_{k-1}, y)$ et la région critique $R^{(\alpha)} = \{z; z \geq q_k\}$; avec $q_k = \inf \{u; Pr(Z \geq u|\theta_0) \leq \alpha\}$, l'indice de satisfaction est :

$$\phi(z) = \Phi_{\theta_0}^B(z)I_{z > q_k}.$$

Finalement, la prédiction de satisfaction, pour la $k^{\text{ème}}$ étape, est comme suit :

$$\begin{aligned} \pi(x_1, \dots, x_{k-1}) &= \sum_{y=(q_k - \sum_{i=1}^{k-1} x_i)^+}^{N_{max} - \sum_{i=1}^{k-1} m_i} \left(\phi\left(\sum_{i=1}^{k-1} x_i + y\right) \right) \nu(y | x_1, \dots, x_{k-1}) \\ &= \sum_{y=(q_k - \sum_{i=1}^{k-1} x_i)^+}^{N_{max} - \sum_{i=1}^{k-1} m_i} \left(\Phi_{\theta_0}^B\left(\sum_{i=1}^{k-1} x_i + y\right) \right) \\ &\times \frac{C_{N_{max} - \sum_{i=1}^{k-1} m_i}^y \beta\left(y + \sum_{i=1}^{k-1} x_i + a, N_{max} - y - \left(\sum_{i=1}^{k-1} x_i\right) + b\right)}{\beta\left(\sum_{i=1}^{k-1} x_i + a, \sum_{i=1}^k m_i - \sum_{i=1}^{k-1} x_i + b\right)}. \end{aligned}$$

Dans l'analyse finale, Il faut prendre une décision, i.e., à la fin de l'essai, il n'y a pas de zone grise où le nombre de réponses pourrait se situer.

4.3.1 PS-Designs pour les essais en deux étapes

Afin de clarifier la procédure PS, nous considérons la conception optimale en deux étapes comme suit :

Étant donné un taux de réponse cible θ_0 , une taille d'échantillon maximale N_{\max} , un taux d'erreur de type I (α) et un taux d'erreur de type II (β) pour un nouveau cible souhaitable θ_1 , on trouve l'ensemble de paramètres $(N_{\max}, m_1, x_1, q_2)$; satisfaisant que la prédiction de satisfaction $\pi(x_1)$ soit un nombre réel dans $[0.5, 1]$ et que les contraintes de taux d'erreurs de type I et de type II soient vérifiées en minimisant la taille de l'échantillon espérée sous θ_0 .

Dans certains cas, la conception minimax est plus attrayante que la taille minimale attendue de l'échantillon; nous pouvons donc choisir une conception en deux étapes ayant la plus petite taille maximale d'échantillon qui réponde aux contraintes de la conception.

4.3.2 Comparaison entre PS-designs et les designs en deux étapes de Simon

La comparaison entre les PS-designs et celles de Simon consiste à examiner les caractéristiques de fonctionnement du conception, telles que le taux d'erreur de type I, la puissance, $PET(\theta_0)$ et $E(N | \theta_0)$. Le concept de conception d'essais utilisant l'approche PS est illustré dans l'exemple suivant :

Exemple : Un essai sur le cancer du poumon

L'objectif principal de cette étude est d'évaluer l'efficacité d'un traitement d'association en tant que traitement de première ligne chez des patients atteints d'un cancer du poumon à non-petites cellules à un stade avancé. L'étude associe l'anticorps du facteur de croissance endothélial vasculaire à un inhibiteur du récepteur du facteur de croissance épidermique, la tyrosine kinase. Le critère principal est le taux de réponse clinique au nouveau régime thérapeutique (c'est-à-dire, le taux de réponse complète et partielle combinés).

Le traitement standard actuel donne un taux de réponse d'environ 20% (θ_0). Le taux de réponse cible du nouveau régime est 40% (θ_1).

Avec les contraintes de taux d'erreur de type I et de type II ≤ 0.1 , la conception optimal de Simon, en deux étapes, donne les résultats suivant :

$$m_1 = 17, x_1 = 3, N_{\max} = 37, x_2 = 10, PET(\theta_0) = 0.55 \text{ et } E(N | \theta_0) = 26.02;$$

avec $\alpha = 0.095$ et $\beta = 0.097$.

Le design minimax correspondant donne :

$m_1 = 19, x_1 = 3, N_{\max} = 36, x_1 + x_2 = 10, PET(\theta_0) = 0.456$ et $E(N | \theta_0) = 28.26$;
avec $\alpha = 0.086$ et $\beta = 0.098$.

Pour les PS designs, deux scénarios ont été pris en compte : Dans le premier (SC1), nous supposons que le taux de réponse θ suit l'a priori de Jeffreys $Beta(1/2, 1/2)$. L'essai est surveillé continument après l'évaluation des réponses du 10 premiers patients. Selon les valeurs spécifiées de $\theta_0, \theta_1, \alpha$ et β , les caractéristiques de la conception optimale sont les suivantes :

$m_1 = 24, x_1 = 8, N_{\max} = 37, x_1 + x_2 = 11, PET(\theta_0) = 0.9638$ et $E(N | \theta_0) = 24.47$;
avec $\alpha = 0.0892$ et $\beta = 0.0960$.

Avec une prédiction de satisfaction 0.5263. Le design minimax correspondant donne :

$m_1 = 24, x_1 = 7, N_{\max} = 35, x_1 + x_2 = 10, PET(\theta_0) = 0.9108$ et $E(N | \theta_0) = 24.98$;
avec $\alpha = 0.0892$ et $\beta = 0.0960$.

Avec une prédiction de satisfaction, dans ce cas, 0.5998.

En appliquant la même méthode, nous pouvons trouvé les caractéristiques des PS designs pour le deuxième scénario (SC2) lorsque le taux de réponse θ suit une a priori uniforme $Beta(1, 1)$. Les résultats sont présentés dans le tableau 4.1.

Par rapport à la conception minimax à deux étapes de Simon, celles du PS présentent une probabilité plus élevée d'arrêt précoce et une taille d'échantillon espérée plus petite dans le cas nulle. En effet, les résultats sont très proches dans le cas d'une a priori uniforme, car c'est une distribution plate, alors qu'il est différent dans le cas où l'a priori est de Jeffreys.

La comparaison est plus visible et les résultats prouvent que la conception est améliorée lorsqu'on choisit une a priori plus informative. Cependant, la méthodologie reste valable et nous laissons le choix de l'a priori aux utilisateurs.

| Simon's designs | | | | | | | |
|--------------------|-----------|----------------------|-----------------|-------------------|----------|---------|------|
| | x_1/m_1 | $x_1 + x_2/N_{\max}$ | $PET(\theta_0)$ | $E(N \theta_0)$ | α | β | PS |
| Optimal | 3/17 | 10/36 | 0.55 | 26.02 | 0.095 | 0.097 | 0.24 |
| Minimax | 3/19 | 10/36 | 0.456 | 28.26 | 0.086 | 0.098 | 0.17 |
| <i>PS</i> -designs | | | | | | | |
| Optimal(SC1) | 8/24 | 11/37 | 0.9638 | 24.47 | 0.0892 | 0.0960 | 0.53 |
| Minimax (SC1) | 7/24 | 10/35 | 0.9108 | 24.98 | 0.0892 | 0.0960 | 0.60 |
| Optimal(SC2) | 7/24 | 10/34 | 0.9638 | 24.47 | 0.0892 | 0.0960 | 0.53 |
| Minimax (SC2) | 7/24 | 10/34 | 0.9108 | 24.89 | 0.0892 | 0.0960 | 0.60 |

TABLE 4.1 – Caractéristiques de fonctionnement (OC) des conceptions en deux étapes de Simon et de celles de *PS* avec les taux d'erreur de type I et de type II ≤ 0.10 , $\theta_0 = 0.2$, $\theta_1 = 0.4$, a priori pour $\theta \sim Beta(0.5, 0.5)$ (SC1), a priori pour $\theta \sim Beta(1, 1)$ (SC2).

4.4 Exemple illustratif

Yanagihara et al. (2010) ont mené un essai de phase II à un seul bras de S-1 et de docétaxel chez des patients préalablement traités atteints d'un cancer du poumon non à petites cellules localement avancé ou méta-statique. Le critère d'évaluation principal était le taux de réponse globale (ORR : the overall response rate), le succès étant défini comme une réponse complète ou partielle. La taille d'échantillon prévue était de 35, en supposant l'hypothèse nulle d'un ORR de 0.09 et une hypothèse alternative d'un ORR de 0.25, avec une erreur unilatérale de type I = 0.1 et une erreur de type II = 0.1. En conséquence, les 38 patients qui étaient éligibles, ont été analysés et l'ORR final était de 18.4% (7/38). Sur les premiers patients inscrits à 10, 20, 30 et 38, des réponses partielles ont été observées chez 1, 5, 6 et 7 patients, respectivement (pas de réponse complète).

Afin d'illustrer l'approche proposée, nous spécifions les paramètres de conception tels que $\theta_0 = 0.09$, $\alpha = 0.1$, $K = 4$, $m_1 = m_2 = m_3 = 10$, $m_4 = 8$, $N_{\max} = 38$. Supposons que le seuil de probabilité d'arrêt pour inefficacité soit de 0.5.

Après la première étape et sans aucune information disponible sur le paramètre inconnu θ , il est préférable de choisir l'a priori uniforme $Beta(1, 1)$. Dans cette analyse intermédiaire, l'ORR était de 10%(1/10), de sorte que la distribution a posteriori est $Beta(2, 10)$ (Fig. 4.1 (a)). La prédiction de satisfaction est de $0.49 \simeq 0.5$. L'essai ne sera pas arrêté pour inefficacité. Après le deuxième coup d'œil intermédiaire, l'ORR a été

25%(5/20) ; cela signifie $x_1 = 1$ et $x_2 = 4$. La distribution a posteriori est une $Beta(6, 16)$ (Fig. 4.1 (b)). La prédiction de satisfaction était $\pi(x_1 + x_2 = 5) = 0.81$ (Parce que la distribution prédictive pour N_{max} est décalée vers la droite après chaque surveillance intérimaire Fig. 4.2). Bien que la prédiction de satisfaction soit excellente dans cette analyse intérimaire, il est préférable de mener l'essai jusqu'à ce que les 38 patients soient inscrits. En effet, si le traitement fonctionne, recruter davantage de patients dans le traitement actif ne serait pas contraire à l'éthique et pourrait également augmenter la précision lors de l'évaluation de l'efficacité et de la sécurité du traitement (Lee and Liu, 2008).

Après la troisième surveillance intérimaire, on avait $x_3 = 1$ dans $m_3 = 10$, la distribution a posteriori est $beta(7, 25)$ (Fig. 4.1 (c)) et la prédiction de satisfaction $\pi(x_1 + x_2 + x_3 = 6) = 0.52 > 0.5$.

En conséquence, nous pouvons conclure que le traitement expérimental est prometteur, comme le conclut également le rapport original.

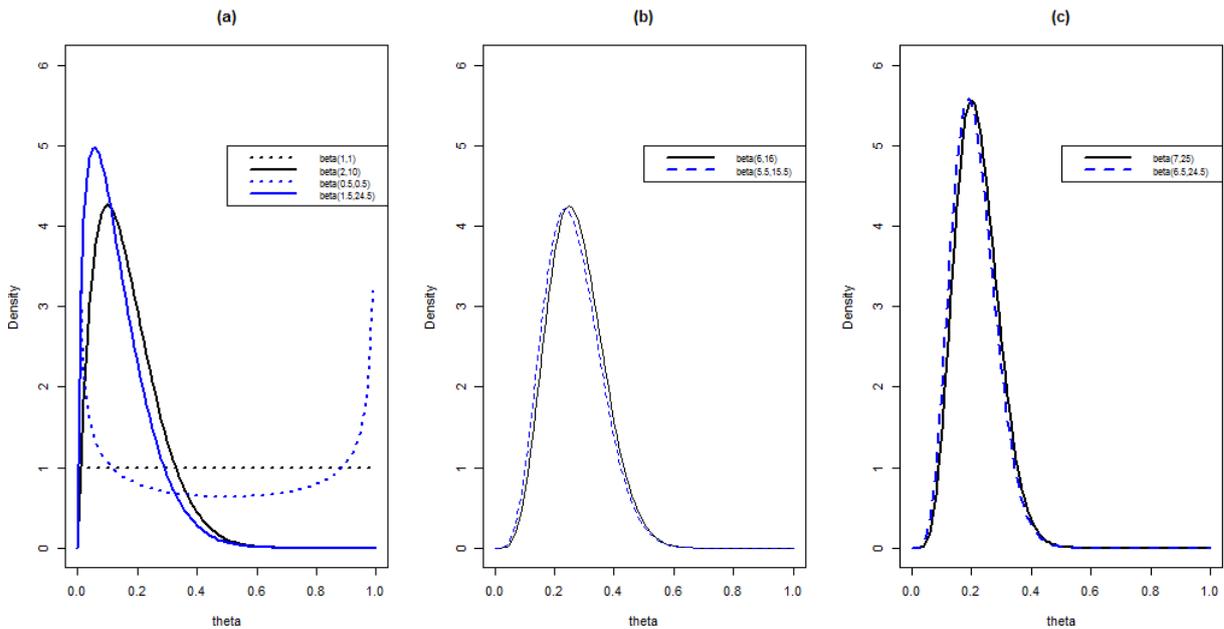


FIGURE 4.1 – L'a priori est actualisée en fonction des résultats du surveillance intermédiaire.

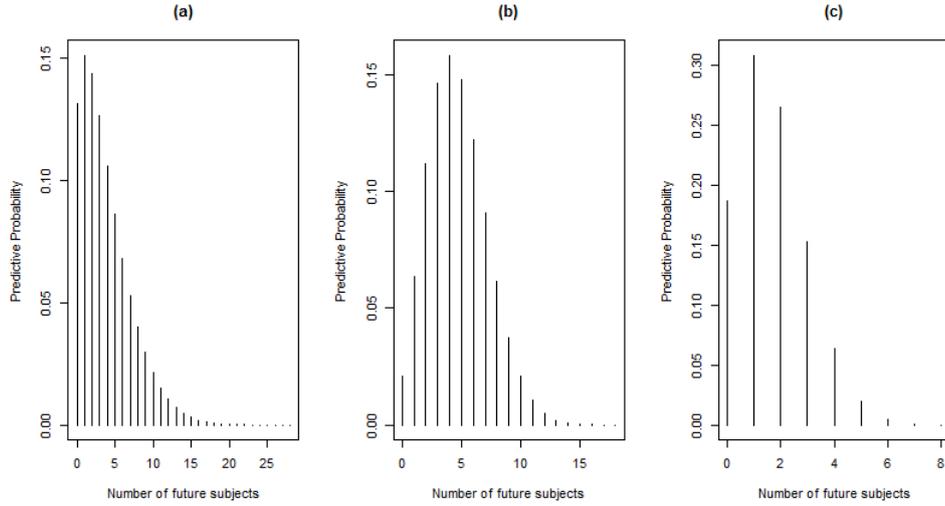


FIGURE 4.2 – La distribution prédictive pour N_{max} est décalée vers la droite après chaque surveillance intérimaire.

4.5 Utilité de PS pour futilité

La raison la plus importante de la surveillance intermédiaire des essais cliniques est l'impératif éthique d'éviter de traiter les patients avec des thérapies inefficaces ou inférieures. La prédiction de satisfaction peut être utilisée pour la surveillance intermédiaire des essais cliniques pour estimer la probabilité d'observer un effet de traitement statistiquement significatif si l'essai continue jusqu'à la taille d'échantillon maximale prédéfinie. Les méthodes bayésiennes sont utiles pour la surveillance intermédiaire, en se basant souvent la décision d'arrêt si la probabilité a posteriori d'un paramètre dépasse un seuil prédéterminé (Saville et al., 2014).

Par exemple, considérons une étude à un seul bras avec un maximum de 100 patients, mesurant une réponse à résultat binaire, dans laquelle la proportion de succès est comparée à un taux d'intérêt de 50%.

Le résultat observé est le nombre de réponses x , supposé suivre une distribution binomiale avec une probabilité θ et le nombre total de patients $N_{max} = 100$. L'essai sera considéré comme succès si la probabilité bayésienne a posteriori que la proportion dépasse le taux d'intérêt ($\theta_0 = 0.5$) est supérieure à $\eta = 0.95$ et qui est donnée par

$$P(\theta > \theta_0 | x) > \eta.$$

Supposons que $\theta \sim \text{Beta}(1, 1)$, l'essai sera considéré comme succès si 59 ou plus parmi les 100 patients montre une réponse, où $P(\theta > 0.5 | x = 59, n = 100) = 0.963$. De plus, cette a priori non informatif préserve l'erreur du type I qui est 0.044.

4.5.1 Comparaison des différentes métriques bayésienne

Dans cette section, nous comparons les différentes mesures bayésiennes utilisées pour le suivi intermédiaire, telles que : La prédiction de satisfaction, les probabilités a posteriori bayésiennes et la probabilité prédictive car le choix de la règle d'arrêt est très important dans l'analyse intermédiaire.

La probabilité a posteriori bayésienne, qui correspond à la probabilité que le paramètre soit contenu dans une région significative, est le meilleur outil pour répondre à la question traitée par la surveillance intermédiaire suivante : « Existe-t-il des preuves convaincantes en faveur de l'hypothèse nulle ou l'alternative? ». D'autre part, en utilisant des méthodes de stochastique curtalement telles que la probabilité prédictive et la prédiction de satisfaction, nous donnons des réponses à la question : « L'essai est-il susceptible de présenter des preuves convaincantes en faveur de l'hypothèse alternative si l'on recueilli des données supplémentaires? » Parce que nous traitons ici de la prévision des preuves qui seront disponibles aux étapes ultérieures de l'étude (Lecoutre, 2001).

Si la futilité est définie comme un essai ayant peu de chances d'atteindre son objectif, alors il s'agit d'un problème de prédiction ; la meilleure solution consiste à prédire la satisfaction ou la probabilité prédictive (Saville et al., 2014).

Exemple : Supposons que l'essai soit conçu avec la planification de quatre analyses intermédiaires pour futilité ; elles sont effectuées lorsque les données sont disponibles pour 20, 50, 75 et 90 patients, respectivement. Supposons que lors de la première analyse intermédiaire, 12 ont été observées sur 20 patients (60%) (p – valeur unilatérale = 0.25), tandis que 48 réponses ou plus sont nécessaires chez les 80 patients restants. afin de considérer la réussite de l'essai. Avec une a priori uniforme, la probabilité a posteriori d'intérêt est $Pr(\theta > 0.50 | x_1 = 12, m_1 = 20) = 0.82$. La distribution prédictive Bayésienne des observations futures $y_1 \sim \text{beta} - \text{binomial}(80, 13, 9)$, et la probabilité prédictive de succès de y est égale à 0.54, qui est la probabilité d'observer 48 ou plus de réponses parmi les 80 patients restants sachant les données observés. Supposons que le nombre de succès dans le deuxième, le troisième et le quatrième analyse intermédiaire sont 28

succès/50 (56%), 41/75 (55%), et 49/90 (54%), avec des probabilités a posteriori de 0.81, 0.79, et 0.80, respectivement. Compte tenu de ces résumés presque identiques des données sur le bénéfice du traitement, il n'est pas évident si l'essai doit être arrêté ou poursuivi après chacune des analyses intermédiaires. En revanche, les probabilités prédictives et la prédiction de satisfaction varient considérablement d'une évaluation à l'autre, comme le montre le tableau 4.2.

| $\sum_{i=1}^k m_i$ | $\sum_{i=1}^k x_i$ | $N_{max} - \sum_{i=1}^k m_i$ | y_i^* | $P(\theta > 0.5 x)$ | PP | $PS(\alpha = 0.05)$ | $PS(\alpha = 0.1)$ |
|--------------------|--------------------|------------------------------|---------|---------------------|------|---------------------|--------------------|
| 20 | 12 | 80 | 48 | 0.81 | 0.62 | 0.58 | 0.64 |
| 50 | 28 | 50 | 31 | 0.80 | 0.45 | 0.37 | 0.51 |
| 75 | 41 | 25 | 18 | 0.79 | 0.26 | 0.16 | 0.36 |
| 90 | 49 | 10 | 10 | 0.80 | 0.10 | 0.03 | 0.25 |

TABLE 4.2 – Comparaison entre les différentes métriques Bayésienne : $\sum_{i=1}^k m_i$ and $\sum_{i=1}^k x_i$: Le nombre de patients et succès à l'analyse intermédiaire k ; $N_{max} - \sum_{i=1}^k m_i$: nombre de patients restants à l'analyse intermédiaire k ; y_i^* : nombre minimum de réponses requiert d'atteindre le succès ; PP : probabilité prédictive Bayésienne de succès et PS : prédiction de satisfaction de succès.

Puisque la PS étant lié au taux d'erreur de type I, nous avons comparé les PP et les PS dans deux cas : $\alpha = 0.05$ et $\alpha = 0.1$.

Lorsque le nombre d'analyses intermédiaires augmente, la prédiction de satisfaction, lorsque $\alpha = 0.1$, diminue légèrement par rapport à la prédiction de satisfaction, lorsque $\alpha = 0.05$ et au PP.

Les critères d'arrêt spécifiques sont, généralement, propres à chaque essai et incluent des considérations éthiques et commerciales, telles que : risque / bénéfice des patients, les ressources disponibles, le coût d'opportunité et le pouvoir statistique global.

4.5.2 Discussion des bornes d'arrêt d'Emersson-Fleming

Considérons les limites inférieures d'Emerson et de Fleming dans l'exemple ci-dessus, dans lequel l'essai est surveillé chez 20, 50 et 75 patients, avec une analyse finale d'efficacité à 100 patients. Le test d'intérêt est un test unilatéral avec $H_0 : \theta \leq 0.5$ versus $H_1 : \theta > 0.5$, avec une erreur de Type I égale à 0.05. L'essai s'arrêtera pour futilité si

moins de 5 *successes* /20 (25%), 25/50(50%), 42/75(56%), ou 59/100(59%) sont observés. La puissance de la conception ci-dessus pour une alternative de $\theta_1 = 0.65$ est 0.93. En utilisant notre modèle Bayésien avec une a priori uniforme, les prédictions de satisfactions pour succès à 5, 25, et 42 succès pour les trois premières surveillances intérimaires (ce qui ne respecterait pas les règles d'arrêt ci-dessus) sont :

Dans le contexte du contrôle de futilité, tel est défini ci-dessus, les règles de décision

| $\sum_{i=1}^k m_i$ | $\sum_{i=1}^k x_i$ | PS |
|--------------------|--------------------|--------|
| 20 | 5 | 0.0041 |
| 50 | 25 | 0.1320 |
| 75 | 42 | 0.1874 |
| 100 | 59 | |

TABLE 4.3 – Calcul de la prédiction de satisfaction pour les limites de Emerson et Fleming.

pour futilité générées par les limites d'Emerson et de Fleming peuvent être améliorées en incluant la prédiction de satisfaction afin d'éviter la continuité des essais avec de très faibles probabilités de succès. En utilisant le modèle PS, les limites d'arrêt sont données dans le Tableau 4.4.

| $\sum_{i=1}^k m_i$ | $\sum_{i=1}^k x_i$ | PS | α -spending | puissance |
|--------------------|--------------------|--------|--------------------|-----------|
| 20 | 13 | 0.5760 | 0.057 | 0.76 |
| 50 | 30 | 0.6580 | 0.059 | 0.81 |
| 75 | 44 | 0.6025 | 0.052 | 0.89 |
| 100 | 58 | | | |

TABLE 4.4 – Bornes d'arrêt en utilisant la Prédiction de satisfaction pour $N_{max} = 100$, $\theta_0 = 0.5$, $\theta_1 = 0.65$, taux d'erreur de type I = 0.05, puissance = 0.9

CONCLUSION GÉNÉRALE

Les méthodes bayésiennes offrent de nombreuses propriétés intéressantes pour les essais cliniques ; conception et analyse. Celles-ci incluent la possibilité d'incorporer les informations obtenues avant la conception de l'étude, d'utiliser celles obtenues pendant la surveillance continue et la souplesse dans la conduite de l'essai : Tout ça est un moyen cohérent de tirer des conclusions.

Nous avons souligné la pertinence des procédures prédictives bayésiennes qui apportent une contribution importante à l'inférence et à l'analyse des données. Dans cette perspective, les probabilités prédictives bayésiennes peuvent être utilisées pour la planification et la surveillance intermédiaire d'essais expérimentaux afin d'estimer la probabilité d'observer un résultat statistiquement significatif, si l'essai devait se poursuivre jusqu'à la taille maximale prédéfinie de l'échantillon.

Comme nous l'avons vu, il est conceptuellement immédiat de passer de la vraisemblance à la distribution a posteriori. Le problème crucial reste alors le choix de la loi a priori, qui a été souvent la pierre d'achoppement de l'inférence bayésienne.

Le but de cette thèse était de proposer une conception adaptative bayésienne simple, cohérente et unifiée. Le PS-design fournit une excellente alternative pour mener des essais de phase II à plusieurs étapes. L'utilisation de probabilités prédictives bayésiennes permet de réaliser des analyses intermédiaires suffisantes et plus souples pour ces essais. La taille de l'échantillon peut être déterminée en choisissant le plus petit N_{max} parmi tous les modèles répondant aux critères de conception.

D'autre part, pour que le taux d'erreur nominal de type I soit contrôlé, nous concevons l'essai avec une p -*valeur* plus stricte à l'étape final. De plus, notre conception prend en compte l'incertitude concernant les données qui ne sont pas encore observées, puisqu'il s'agit d'un problème pré-expérimental. En même temps et afin de laisser les données guider l'analyse, nous introduisons une a priori non informative, utilisée pour obtenir des probabilités a posteriori d'efficacité du traitement.

La méthode de prédiction de satisfaction pour la surveillance intermédiaire est une méthode cohérente, efficace et flexible qui pourrait être utile dans les essais cliniques à un seul bras avec des données binaire. Elle ne nécessite pas de calcul intensif et de simples programmes de simulations peuvent être nécessaires à la conception de phase afin d'évaluer les bornes d'arrêt et les caractéristiques de fonctionnement (y compris les probabilités d'erreur de type I et de type II), du point de vue du frequentist.

Pour conclure, on dit que le choix bayésien sera, tôt ou tard, incontournable.

PERSPECTIVES

Nous envisageons la prédiction de satisfaction dans le cas séquentiel, pour un test hybride de comparaison en l'appliquant dans les essais cliniques dans deux cas :

- Le premier pour le modèle binomial en choisissant deux loi a priori différentes, une non informative pour le traitement expérimental et une autre informative pour le traitement contrôle, sans oublier de faire la randomisation pendant la sélection des patients dans les deux bras de l'essai.

- Dans le second, on considère les tests séquentiels de groupes utilisant la loi gaussienne pour trouver les frontières d'arrêt dans le cas des grands échantillons.

La deuxième perspective c'est de trouver la taille de l'échantillon optimale ainsi que les tailles des cohortes pour une prédiction de satisfaction donnée.

ANNEXE

Programme 1 :

> $\alpha = 0.05; \beta = 0.20; p_0 = 0.10; p_1 = 0.20$

> $z_\alpha = qnorm(1 - \alpha); z_\alpha; z_\beta = qnorm(1 - \beta)$

> $z_\beta; \bar{p} = (p_0 + p_1)/2; \bar{p}$

[1]1.644854

[1]0.8416212

[1]0.15

> $N_{\max} = (z_\alpha + z_\beta)^2 * (\bar{p} * (1 - \bar{p})) / (p_1 - p_0)^2$

> N_{\max}

[1]78.8276

Programme 2 : ($\beta = puissance$)

> $\alpha = 0.05; \beta = 0.80; p_0 = 0.10; p_1 = 0.20$

> $z_\alpha = qnorm(1 - \alpha); z_\alpha; z_\beta = qnorm(\beta); z_\beta; \bar{p} = (p_0 + p_1)/2; \bar{p}$

[1]1.644854

[1]0.8416212

```
[1]0.15
```

```
> Nmax = (zalpha + zbeta)^2 * (pbar * (1 - pbar)) / (p1 - p0)^2
```

```
> Nmax
```

```
[1]78.8276
```

```
> N1 = round((3/4) * Nmax); N1; N2 = round((5/4) * Nmax); N2
```

```
[1]59
```

```
[1]99
```

```
> N = 79
```

```
> for(k in 0 : N)
```

```
{alphak[k] = 1 - pbinom(k, N, p0); betak[k] = 1 - pbinom(k, N, p1)}
```

```
> for(s in 1 : N)
```

```
{
```

```
if(alphak[s] < alpha)
```

```
{alphaa[s] = alphak[s]} else {alphaa[s] = 0}
```

```
if(betak[s] > beta)
```

```
{betaa[s] = betak[s]} else {betaa[s] = 0}
```

```
}
```

Programme 3 :

```
> p0 = 0.3; p1 = 0.1; N = 25; n1 = 15; r1 = 1; r = 3; a = 0; b = 0
```

```
> #Probabilité de rejeter H0 pour la 1ère étape
```

```
> alpha1 = pbinom(r, N, p0); alpha1
```

```
[1]0.03324052
```

```
> power1 = pbinom(r, N, p1); power1
```

```
[1]0.7635914
```

```

> #Probabilité prédictive

> pp = 0

> for(X in r1 : r)

{

  s[X] = log(gamma(N - n1 + 1)) - log(gamma(X - r1 + 1)) - log(gamma(N -
n1 - X + r1 + 1)) + log(gamma(X + a + 1)) + log(gamma(N + b - X - a + 1)) -
  log(gamma(N + b + 2)) + log(gamma(n1 + b + 2)) - log(gamma(r1 + a + 1)) -
  log(gamma(n1 + b - r1 - a + 1)); s; pp = pp + exp(s[X])

}

> pp

[1]0.8575251

> #probabilité de rejeter H0 après la 2ème étape

> preject2p0 = 0

> for(k in r1 + 1 : r)

{

  px = numeric(length(r - r1 - 1))

  if(k > 0)

  px = (pbinom(k, n1, p0) - pbinom(k - 1, n1, p0)) * pbinom(r - k, N - n1, p0)

  else px = pbinom(k, n1, p0) * pbinom(r - k, N - n1, p0)

  preject2p0 = preject2p0 + px

}

> preject2p0

[1]0.0184739

> preject2p1 = 0

> for(k in r1 + 1 : r)

```

```

{
  px = numeric(length(r - r1 - 1))
  if(k > 0)
    px = (pbinom(k, n1, p1) - pbinom(k - 1, n1, p1)) * pbinom(r - k, N - n1, p1)
  else px = pbinom(k, n1, p1) * pbinom(r - k, N - n1, p1)
  preject2p1 = preject2p1 + px
}
> preject2p1
[1]0.2412689
> #alpha et la puissance après la 2ème étape
> alpha21 = pbinom(r1, n1, p0)
> alpha22 = alpha21 + preject2p0
> alpha22
[1]0.0537415
> power21 = pbinom(r1, n1, p1)
> power22 = power21 + preject2p1
> power22
[1]0.7903119
> #E(N/p)
> ASN0 = n1 + (1 - alpha21) * (N - n1)
> ASN0
[1]24.64732
> ASN1 = n1 + (1 - power21) * (N - n1)
> ASN1

```

```
[1]19.50957
```

Programme 4 :

```
> p0 = 0.3; p1 = 0.1; N = 25; n1 = 15; r1 = 1; r = 3; a = 3.1; b = 39
```

```
> #Probabilité de rejeter H0 pour la 1ère étape
```

```
> alpha1 = pbinom(r, N, p0); alpha1
```

```
[1]0.03324052
```

```
> power1 = pbinom(r, N, p1); power1
```

```
[1]0.7635914
```

```
> #Probabilité prédictive
```

```
> pp = 0
```

```
> for(X in r1 : r)
```

```
{
```

```
  s[X] = log(gamma(N - n1 + 1)) - log(gamma(X - r1 + 1)) - log(gamma(N -
n1 - X + r1 + 1)) + log(gamma(X + a + 1)) + log(gamma(N + b - X - a + 1)) -
```

```
  log(gamma(N + b + 2)) + log(gamma(n1 + b + 2)) - log(gamma(r1 + a + 1)) -
log(gamma(n1 + b - r1 - a + 1)); s; pp = pp + exp(s[X])
```

```
}
```

```
> pp
```

```
[1]0.9298985
```

```
> #probabilité de rejeter H0 après la 2ème étape
```

```
> preject2p0 = 0
```

```
> for(k in r1 + 1 : r)
```

```
{
```

```
  px = numeric(length(r - r1 - 1))
```

```
  if(k > 0)
```

```

px = (pbinom(k, n1, p0) - pbinom(k - 1, n1, p0)) * pbinom(r - k, N - n1, p0)

else px = pbinom(k, n1, p0) * pbinom(r - k, N - n1, p0)

preject2p0 = preject2p0 + px

}

> preject2p0

[1]0.0184739

> preject2p1 = 0

> for(kinr1 + 1 : r)

{

px = numeric(length(r - r1 - 1))

if(k > 0)

px = (pbinom(k, n1, p1) - pbinom(k - 1, n1, p1)) * pbinom(r - k, N - n1, p1)

else px = pbinom(k, n1, p1) * pbinom(r - k, N - n1, p1)

preject2p1 = preject2p1 + px

}

> preject2p1

[1]0.2412689

> #alpha et la puissance après la 2ème étape

> alpha21 = pbinom(r1, n1, p0)

> alpha22 = alpha21 + preject2p0

> alpha22

[1]0.0537415

> power21 = pbinom(r1, n1, p1)

> power22 = power21 + preject2p1

```

```

> power22

[1]0.7903119

> #E(N/p)

> ASN0 = n1 + (1 - alpha21) * (N - n1)

> ASN0

[1]24.64732

> ASN1 = n1 + (1 - power21) * (N - n1)

> ASN1

[1]19.50957

```

Programme 5 :

```

> p0 = 0.3; p1 = 0.1; N = 25; n1 = 15; r1 = 0; r = 3; a = 3.3; b = 12.33

> #Probabilité de rejeter H0 pour la 1ère étape

> alpha1 = pbinom(r, N, p0); alpha1

[1]0.03324052

> power1 = pbinom(r, N, p1); power1

[1]0.7635914

> #Probabilité prédictive

> pp = 0

> for(X in r1 : r)

{

  s[X] = log(gamma(N - n1 + 1)) - log(gamma(X - r1 + 1)) - log(gamma(N -
n1 - X + r1 + 1)) + log(gamma(X + a + 1)) + log(gamma(N + b - X - a + 1)) -
  log(gamma(N + b + 2)) + log(gamma(n1 + b + 2)) - log(gamma(r1 + a + 1)) -
  log(gamma(n1 + b - r1 - a + 1)); s; pp = pp + exp(s[X])

}

```

```
> pp
[1]0.9621236
> #probabilité de rejeter H0 après la 2ème étape
> preject2p0 = 0
> for(k in r1 + 1 : r)
{
px = numeric(length(r - r1 - 1))
if(k > 0)
px = (pbinom(k, n1, p0) - pbinom(k - 1, n1, p0)) * pbinom(r - k, N - n1, p0)
else px = pbinom(k, n1, p0) * pbinom(r - k, N - n1, p0)
preject2p0 = preject2p0 + px
}
> preject2p0
[1]0.03015645
> preject2p1 = 0
> for(k in r1 + 1 : r)
{
px = numeric(length(r - r1 - 1))
if(k > 0)
px = (pbinom(k, n1, p1) - pbinom(k - 1, n1, p1)) * pbinom(r - k, N - n1, p1)
else px = pbinom(k, n1, p1) * pbinom(r - k, N - n1, p1)
preject2p1 = preject2p1 + px
}
> preject2p1
```

```
[1]0.5603346
```

```
> #alpha et la puissance après la 2ème étape
```

```
> alpha21 = pbinom(r1, n1, p0)
```

```
> alpha22 = alpha21 + preject2p0
```

```
> alpha22
```

```
[1]0.03490401
```

```
> power21 = pbinom(r1, n1, p1)
```

```
> power22 = power21 + preject2p1
```

```
> power22
```

```
[1]0.7662258
```

```
> #E(N/p)
```

```
> ASN0 = n1 + (1 - alpha21) * (N - n1)
```

```
> ASN0
```

```
[1]24.95252
```

```
> ASN1 = n1 + (1 - power21) * (N - n1)
```

```
> ASN1
```

```
[1]22.94109
```

Programme 6 :

```
> #Pour  $\Pr(y = i/x)$ 
```

```
> p0 = 0.6 ; n = 23 ; N max = 40 ; m = (N max - n) ; x = 16 ; a = 0.6 ; b = 0.4
```

```
> ab = c(x + a, n - x + b)
```

```
> ys = 0 : m
```

```
> pred = pbetap(ab, m, ys)
```

```
> dist = round(cbind(ys, pred), 4)
```

```

> #Pour Bi = Pr(p > p0/x, y = 0)
> f = 1 - (pbeta(p0, a + x + 0, b + Nmax - x - 0))
> fdis = round(cbind(1, f), 4)
> #Pour Bi = Pr(p > p0/x, y = i), i = 1 : m
> p0 = 0.6; n = 23; Nmax = 40; m = (Nmax - n); x = 16; a = 0.6; b = 0.4
> g = numeric(length(m + 1))
> for (i in 1 : m) {g[i] = 1 - (pbeta(p0, a + x + i, b + Nmax - x - i)); g[i]}
> dis = round(cbind(1 : m, g), 4)
> Bi = c(fdis[, 2], dis[, 2])
> pred = c(dist[, 2])
> mat1 = cbind(ys, pred, Bi)
> #Pour la probabilite predictive
> theta.t = 0.90
> k = NA
> for(i in 1 : (m + 1)) {if(mat1[i, 3] > theta.t){k[i] = 1} else {k[i] = 0}}
> mat2 = cbind(mat1, k)
> colnames(mat2) = c("i", "Pr(Y = i/x)", "Bi", "1(Bi > 0.90)")
> mat2
> PP = 0
> for(i in 1 : (m + 1)) {if(mat2[i, 4] == 1) {PP = PP + mat2[i, 2]}}
> PP
[1]0.5655589

```

Programme 7 :

```

theta=0; sigma=1; alpha=0.01; Nsim=1000
n=40; k=30
x=0.5; s1=0.1
s1=sqrt(s1)
sprim1=(s1*sqrt(n+k))/sqrt(n*k)
yi=rt(k,n-1,(theta-x)/sprim1)
y=mean(yi)
y
s2=var(yi)
s2
s2=sqrt(s2)
sprim2=s2/sqrt(k)
a=sqrt(n-1)*(theta-x)/sprim1; t=sqrt(n-1)*sprim2/sprim1
a; t

ualpha=qnorm(1-alpha)
q0=theta+((sprim2*ualpha)/sqrt(k))
q0
u=runif(Nsim,0,1)
v=pnorm(a+t*ualpha)
for(i in 2:Nsim){
v[i]=pnorm(a+t*ualpha)+(1-pnorm(a+t*ualpha))*u[i]
}
z=qt(v,n-1)
for(i in 2:Nsim){
z[i]=qt(v[i],n-1)
}
mean(pnorm((z[i]-a)/t))
p=function(x){
  u=runif(Nsim,0,1)
  v=pt(a+t*ualpha,n-1)
  for(i in 2:Nsim){
  v[i]=pt(a+t*ualpha,n-1)+(1-pt(a+t*ualpha,n-1))*u[i]
  }
  z=qt(v,n-1)
  for(i in 2:Nsim){
  z[i]=qt(v[i],n-1)
  }
  return(cumsum(pnorm((z-a)/t))/(1:Nsim))
}
plot(p(x),xlim=c(-2,1000),type="l",ylab="prediction of satisf
title("theta=0, sigma=1, n=30, k=20
      for alpha=0.05")

```

RÉSUMÉ

L'objet de cette thèse, qui a été motivée par des problèmes méthodologiques qui se posent dans le contexte des essais expérimentaux, est de proposer une approche hybride bayésienne-fréquentiste utilisant une prédiction séquentielle bayésienne de l'indice de satisfaction, ce qui est souvent réalisé dans le cas de protocole d'essais cliniques.

Pour une analyse intermédiaire portant sur l'hypothèse de prédiction, telle que la surveillance de la futilité d'un traitement expérimental avec des résultats retardés, la prédiction de satisfaction prend en compte la quantité de données restant à observer dans un essai clinique et offre la possibilité d'intégrer des informations supplémentaires via des variables auxiliaires à l'aide de la loi a priori.

La conception de la prédiction de satisfaction garantit le taux d'erreur de type I et ne nécessite pas de calcul ni de simulation intensif.

ABSTRACT

The aim of this thesis, which was motivated by methodological problems that arise in the context of experimental trials, is to propose a hybrid Bayesian-frequentist approach using sequential Bayesian prediction of the index of satisfaction, which is often done in the case of clinical trial protocol.

For an interim analysis of the prediction hypothesis, such as fertility monitoring of the experimental treatment with delayed results, the prediction of satisfaction takes into account the amount of data remaining to be observed in a clinical trial and offers the possibility of integrating additional information via auxiliary variables by the a priori distribution.

The design of the prediction of satisfaction guarantees the type I error rate and does not require calculation or intensive simulation.

ملخص

الهدف من هذه الأطروحة و التي كانت مستوحاة من المسائل التي تظهر في سياق المحاولات التجريبية ، هو اقتراح نهج هجين يركز على التنبؤ البايزي المتسلسل لمؤشر الرضا ، والذي غالبا ما يستخدم في بروتوكول التجارب السريرية.

يأخذ التنبؤ بالرضا في الاعتبار مقدار البيانات المتبقية خلال المراقبة المؤقتة في التجارب السريرية ، كمراقبة عدم فعالية الأدوية التجريبية مع نتائج متأخرة. ويوفر أيضا إمكانية دمج معلومات إضافية عبر متغيرات مساعدة من خلال القانون المسبق.

يحد تصميم التنبؤ بالرضا معدلا الخطأ من النوع الأول ولا يتطلب حساباً أو محاكاة مكثفة.

BIBLIOGRAPHIE

- [1] Aitchison, J ; Dunsmore, I.R.(1975).*Statistical Prediction Analysis* .New York : Com-
pridge University Press. ISBN : 978-0-521-20692-1.
- [2] Bernardo, J. (1979). Reference posterior distributions for Bayesian inference (with
discussion). J. Royal Statist. Soc. Series B, 41, 113–147.
- [3] Brian S. E; Andrw P. (1999). *Statistical Aspects Of The Design And Analysis Of*
Clinical Trials. Imperial College Press London ; ISBN : 1-86094-153-2.
- [4] Berry, D. A. (1985). Interim Analyses in Clinical Trials : Classical vs. Bayesian Ap-
proaches, *Statistics in Medicine* 4 : 521–526.
- [5] Brown, B. W., Heson, J., Atkinson, N., Rozell, M. E. (1987). Projection From Previous
Studies : A Bayesian and Frequentist Compromise, *Controlled Clinical Trials* 8 : 29-
44.
- [6] Chang, M. (2008). *Adaptive Design Theory and Implementation Using SAS and R*,
Boca Raton : Chapman & Hall Crc Biostatistics.
- [7] Chow S. C. (2010). Bayesian adaptive methods for clinical trials. Chapman &
Hall/CRC Biostatistics Series, USA ; ISBN : 1-58488-776-1.
- [8] Chow, S. C., Corey, R., and Lin, M. (2012). On the Independence of Data Monitoring
Committee in Adaptive Design Clinical Trials, *Journal of biopharmaceutical statistics*
22 : 853-867.
- [9] Cox, D. R. (1999). Discussion of ‘Some statistical heresies’ (Lindsey), *Statistician* 48-
30.
- [10] DeGroot, M. (1970). Optimal Statistical Decisions. *McGraw-Hill*, New York.

- [11] Djeridi, Z. and Merabet, H. (2016). Choice of the Prior in the Bayesian Design for the Clinical Trials, *Sciences et Technologie : A* 44 : 27-31.
- [12] Djeridi, Z. and Merabet, H. (2019). A Hybrid Bayesian-Frequentist Predictive Design for Monitoring Multi-Stage Clinical Trials, *Sequential Analysis journal* , DOI :10.1080/07474946.2019.1648919, 2019.
- [13] Everitt, B. and Pickles, A. (1999). *Statistical Aspects of the Design and Analysis of Clinical Trials*, London : Imperial College Press, .
- [14] Gajewski, B. J. and Mayo, M. S. (2006). Bayesian Sample Size Calculations in Phase II Clinical Trials Using a Mixture of Informative Priors, *Statistics in Medicine* 25 : 2554–66.
- [15] German, S. et German, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6, 721–741.
- [16] Gould, A. L. (1983). Abandoning Lost Causes (Early Termination of Unproductive Clinical Trials), *Proceeding Biopharmaceutical Section, American Statistical Association* pp. 31-34.
- [17] Gould, A. L. (2005). Timing of Futility Analyses for Proof of Concept Trials, *Statistics in Medicine* 24 : 1815–1835.
- [18] Grieve, A. P. (1991). Predictive Probability in Clinical Trials, *Biometrics* 47 : 323–30.
- [19] Hamada, M.S., Wilson, A.G., Reese, C.S., Martz, H.F., (2008). Bayesian reliability, *Springer Series in statistics* ISBN 978-0-387-77948-5.
- [20] Heitjan, D. F. (1997). Bayesian Interim Analysis of Phase II Cancer Clinical Trials, *Statistics in Medicine* 16 : 1791–802.
- [21] Herson, J. (1979). Predictive Probability Early Termination Plans for Phase II Clinical Trials, *Biometrics* 35 : 775–83.
- [22] Jaynes, E. (1980). Marginalization and prior probabilities. In Zellner, A., éditeur, *Bayesian Analysis in Econometrics and Statistics*. North-Holland, Amsterdam.
- [23] Jaynes, E. (1983). *Papers on Probability, Statistics and Statistical Physics*. R.D. Rosenkrantz, Reidel, Dordrecht.
- [24] Jeffreys, H. (1961). *Theory of Probability* (3rd edition). Oxford University Press, Oxford.
- [25] Jennison, C. and Turnbull, B. W. (2000). *Group Sequential Method with Applications to Clinical Trials*, New York : Chapman & Hall/CRC Press,.

- [26] Johns, D. and Andersen, J. S. (1999). Use of Predictive Probabilities in Phase II and Phase III Clinical Trials, *Journal of Biopharmaceutical Statistics* 9 : 67–79.
- [27] Labdaoui, A. (2015). Analyse bayésienne dans les essais cliniques, *Thèse de Doctorat LMD* Université Constantine 1.
- [28] Lecoutre, B. (2001). A Bayesian Predictive Procedures for Designing and Monitoring Experiments. In *Bayesian Methods With Applications to Science, Policy and Official Statistics Luxemburg* 301-310.
- [29] Lecoutre, B., Derzko, G., and Grouin, J. M. (1995). Bayesian Predictive Approach for Inference About Proportions, *Statistics in Medecine* 14 : 1057-1063.
- [30] Lee, J. J. and Feng, L. (2005). Randomized Phase II Designs in Cancer Clinical Trials : Current Status and Future Directions, *Journal of Clinical Oncology* 23 : 4450–57.
- [31] Lee, J. J. and Liu, D. D. (2008). A Predictive Probability Design for Phase II Cancer Clinical Trials, *Clinical Trials* 5 : 93–106.
- [32] Mayo, M. S. and Gajewski, B. J. (2004). Bayesian Sample Size Calculations in Phase II Clinical Trials Using Informative Conjugate Priors, *Contemporary Clinical Trials* 25 : 157–67.
- [33] Merabet, H., (2004). Index and prevision of satisfaction in exponential models for clinical trials. *Statistica*, anno LXIV, n3, 441-453.
- [34] Merabet, H. and Labdaoui, A. (2015). Bayesian Prediction Using Two-Stages Design in Experimental Trials, *Nature & Technologie A-Sciences fondamentales et engineering* 12 : 11-23.
- [35] Merabet, H., Labdaoui, A., and Druilhet, P. (2017). Bayesian Prediction for Two-Stage Sequential Analysis in Clinical Trials, *Communication in Statistics - Theory and Methods* 46 : 9807-9816.
- [36] Millier, C., (1967). Une méthode statistique : L'analyse progressive, *Ann. Sci. forest.*24 (4), 327-343.
- [37] Raiffa, H. et Schlaifer, R. (1961). Applied Statistical decision theory. Technical report, Division of Research, Graduate School of Business Administration, Harvard Univ.
- [38] Robert, C. P. (2007). *Bayesian Choice : From Decision-Theoretic Foundations to Computational Implementation*, New York : Springer-Science+Business Media.
- [39] Robert, C. et Casella, G. (2004). Monte Carlo Statistical Methods. Springer- Verlag, New York, seconde édition.

- [40] Saville, B. R., Connor, J. T., Ayers, G. D., and Alvarez, J. (2014). The Utility of Bayesian Predictive Probabilities for Interim Monitoring of Clinical Trials, *Clinical Trials* DOI : 10.1177/1740774514531352.
- [41] Schultz, J. R., Nichol, F. R., Elfring, G. L., Weed, S. D. (1973). Multipletstage Procedures for Drug Screening, *Biometrics* 29 : 293–300.
- [42] Seidenfeld, T. (1987). Entropy and uncertainty. In MacNeill, I. et Umphrey, G., éditeurs, *Foundations of Statistical Inference*, pages 259–287. Reidel, Boston.
- [43] Shannon, C. (1948). A mathematical theory of communication. *Bell System Tech. J.*, 27, 379–423 et 623–656.
- [44] Shao, Y., Mukhi, V., and Goldberg, J. D. (2008). A hybrid Bayesian-Frequentist Approach to Evaluate Clinical Trial Designs for Tests of Superiority and Non-Inferiority, *Statistics in Medicine* 27 : 504–519.
- [45] Simon, R. (1989). Optimal Two-Stage Design for Phase II Clinical Trials, *Controlled clinical trials* 10 : 1–10.
- [46] Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Approaches to Clinical Trials and Health-Care Evaluation*. Chichester, West Sussex : Chapman John Wiley & Sons Ltd.
- [47] Spiegelhalter, D. J. and Freedman, L. S. (1986). A predictive Approach to Selecting the Size of a Clinical Trial, Based on Subjective Clinical Opinion, *Statistics in Medicine* 5 : 1–13.
- [48] Spiegelhalter, D. J., Freedman, L. S., and Blackburn, P. R. (1986). Monitoring Clinical Trials : Conditional or Predictive power? *Contemporary Clinical Trials* 7 : 8–17.
- [49] Spiegelhalter, D. J., Freedman, L. S., and Parmar, M. K. (1994). Bayesian Approaches to Randomized Trials, *Journal of the Royal Statistical Society. Series A* 157 : 357–416.
- [50] Stangl, D. K. and Berry, D. A. (1998). Bayesian Statistics in Medicine : Where are we and Where Should we be Going? *Indian Journal of Statistics* 60 : 176–95.
- [51] Tan, S. B. and Machin, D. (2002). Bayesian Two-Stage Designs for Phase II Clinical Trials, *Statistics in Medicine* 21 : 1991–2012.
- [52] Thall, P. F. and Simon, R. (1994). A Bayesian Approach to Establishing Sample Size and Monitoring Criteria for Phase II Clinical Trials, *Contemporary Clinical Trials* 15 : 463–481.
- [53] Thall, P. F. and Simon, R. (1994). Practical Bayesian Guidelines for Phase IIB Clinical Trials, *Biometrics* 50 : 337–49.

- [54] Wasserman, L. (1999). Asymptotic inference for mixture models by using data-dependent priors. *J. Royal Statist. Soc. Series B*, 61(1), 159–180.
- [55] Yanagihara, K., Yoshimura, K., Niimi, M., Yasuda, H., Sasaki, T., Nishimura, T., Ishiguro, T., Matsumoto, S., Kitano, T., Kanai, M., Misawa, A., Tada, H., Teramukai, S., Mio, T., Fukushima, M. S. (2010). Phase II Study of S-1 and Docetaxel for Previously Treated patients With Locally Advanced or Metastatic Non-Small Cell Lung Cancer, *Cancer Chemotherapy and Pharmacology* 66 :913–918.
- [56] Zhu, Q.Y., (1991). Modèles Bayésiens et application à l'estimation des caractéristiques de produits finis et aux contrôle de la qualité, *Thèse de DOCTORAT de l'École Nationale des Ponts et Chaussées*.
- [57] Zohar, S., Teramukai, S. and Zhou, Y. (2008). Bayesian Design and Conduct of Phase II Single-Arm Clinical Trials With Binary Outcomes : A Tutorial, *Contemporary Clinical Trials* 29 : 608-616.