



Faculté des Sciences Exacte et Informatique
Département de Mathématique

N° d'ordre :

N° de séries :

Mémoire de fin d'études

Présenté pour l'obtention du diplôme de

Master

Spécialité : Mathématiques.

Option : Probabilités et Statistique.

Thème

Etude des caractéristiques de fonctionnement fréquentistes dans le cadre des designs Bayésiens

Présenté par :

- BENSKAIM Yasmina.

- NOUASRA Mounia.

Devant le jury :

Président : BOUDJERDA Khawla M.A.B Université de Jijel

Encadreur : DJERIDI Zahra M.A.A Université de Jijel

Examineur : GHOUIL Djawida M.A.A Université de Jijel

✧ *Remerciements* ✧

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

*C'est avec un grand honneur que nous remercions notre enseignante et promoteur Mme **Djeridi Zahra** pour nous avoir dirigés pour la réalisation de ce travail, pour ces précieux conseils et ces encouragements..*

*Nous remercions également Mme **BOUDJERDA Khawla** de nous avoir fait l'honneur de faire partie des membres du jury et d'examiner ce travail. Nous tenons à vous remercier.*

*Nous remercions également Mme **GHOUIL Djawida** de nous avoir fait l'honneur de faire partie des membres du jury et d'examiner ce travail. Nous tenons à vous remercier.*

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

Mounia-Yasmina

Table des matières

Notations	vi
Introduction Générale	vii
1 Tests statistiques dans les essais cliniques	1
1.1 Essais cliniques	1
1.1.1 Les différentes phases d'un essai clinique	1
1.1.2 Essais de phase II	2
1.2 Tests d'hypothèses	3
1.2.1 Test randomisé	5
1.2.2 Application dans les essais clinique	7
1.3 Tests d'équivalence et de non-infériorité	9
1.3.1 Tests d'équivalence	10
1.3.2 Tests de non-infériorité	11
1.4 Analyses intermédiaires dans les essais de dispositive médicaux	14
1.4.1 Motivation	14
1.4.2 Objectifs de surveillance des essais	15
1.4.3 Situations conduisant à un arrêt prématuré	17
1.4.4 Nombre de sujets nécessaire	18

2	Conditions fréquentistes appliquées sur les tests Bayésiens	20
2.1	Modélisation bayésienne	20
2.1.1	Distribution a priori et a posteriori	22
2.1.2	Modélisation de l'information a priori	25
2.2	La densité prédictive	28
2.2.1	Puissance et nombre de sujets nécessaire	33
2.3	Règles d'arrêt bayésiennes pour les essais de phase II	36
2.3.1	Conditions de FDA sur les essais bayésiens	36
2.3.2	Méthode Bayésienne comme un outil des fréquentiste	37
2.3.3	Règle d'arrêt binaire pour futilité et efficacité	38
2.4	La probabilité prédictive dans les tests de non infériorité	44
2.4.1	Les méthodes pour calculer la probabilité prédictive	46
3	Application	49
3.1	Considérations de modélisation	50
3.1.1	Justification de l'analyse intermédiaire et tardif	50
3.1.2	Utilisation d'une a priori informative pour le contrôle	50
3.1.3	Aperçu de la procédure d'examen de non-infériorité	51
3.2	Mise en œuvre du design	52
3.2.1	Modèle statistique	52
3.2.2	Simulation	54
3.2.3	Conception à l'aide du contrôle a priori informative	55
	Conclusion générale	60
	Résumé	61

Annexe	62
Bibliographie	67

Notations

$p - valeur$	La probabilité observer.
C_n^x	Coefficient binomial.
$\pi(\theta)$	La loi a priori.
$\pi(\theta x)$	La loi a posteriori.
\propto	Proportionnelle à .
PP	Probabilité prédictive.
B	la fonction Beta.
\mathcal{B}	La loi binomiale
$\mathcal{B}e$	La loi beta
$\mathcal{B}e - \mathcal{B}$	La loi beta-binomiale
SCC	Succés clinique composite.
E	Expérimental.
C	Contrôle.
AI	Analyse intermédiaire.
FDA	Food and Drug Administration.
$I(\theta)$	L'information de fisher.
1	La fonction indicatrice

Introduction générale

Les essais cliniques sont une étape nécessaire à la mise en service d'un médicament sur le marché [22]. On va s'intéresser aux essais de phase II dont l'objectif est de déterminer la posologie optimale du produit en terme d'efficacité et de tolérance sur une population limitée et homogène de patients (quelques centaines). Nous nous concentrons particulièrement sur les essais de phase IIB qui sont des études multi-bras pour comparer l'efficacité du nouveau médicament par rapport aux médicaments standard ou d'autres médicaments expérimentaux, nous utilisons dans ce cas les tests d'équivalence et de non-infériorité [14].

Plusieurs méthodes statistiques ont été développées pour définir des règles d'arrêt en terme d'efficacité des nouveaux dispositifs médicaux dans la planification des essais cliniques de phase II mais elles reposent sur une première analyse après l'inclusion d'un nombre fixe de patients. Au cours d'un essai, l'information s'accumule progressivement au fur et à mesure des inclusions et du suivi des patients. Mais c'est seulement au terme de l'essai, après avoir recruté l'effectif prévu et suivi les patients avec le recul prévu, que la quantité d'information est suffisante et que les données peuvent être analysées. Cependant, des analyses intermédiaires à la recherche de l'effet du traitement en cours d'essai sont envisageable pour diverses raisons avant que tous les patients prévus aient été recrutés et/ou avant la fin de la période de suivi initialement prévue.

Notre objectif du mémoire c'est de donner les méthodes statistiques adaptatives pour démontrer l'efficacité ou la supériorité d'un nouveau traitement par rapport à un autre de contrôle via des tests d'hypothèse bien choisis.

On utilisera les deux approches d'un fréquentistes, et l'autre bayésiennes pour définir des règles d'arrêt dans le contrôle intermédiaire. Ces règles d'arrêt prennent en compte des considérations éthiques et économiques.

À cet égard, Food and Drug Administration (**FDA**) "l'agence américaine des produits alimentaires et médicament" recommande l'importance de choix de la méthode pour l'ana-

lyse des résultats intermédiaires et exige que les procédures réalisent les deux contraintes d'erreurs de type I et de type II des tests statistiques [25].

Ce mémoire comporte trois chapitres :

- Le premier chapitre est un aperçu sur les essais cliniques, nous avons introduit les notions des tests randomisé, et nous avons présenté les tests d'équivalence et de non-infériorité et leur utilisation dans les essais cliniques. En ce basant aussi sur l'analyse intermédiaire dans les essais des dispositifs médicaux.

- Le deuxième chapitre est consacré à l'approche bayésienne et ses méthodes adaptatives aux essais cliniques de phase II. Nous avons aussi présenté des règles d'arrêt bayésiennes qui respectent les conditions de **FDA** pour les essais bayésiens.

- Dans le troisième chapitre, nous avons illustré notre point de vue par une application inspirée de l'article de *Greg Maislin. (2011)* [11]. en choisissant de calculer la probabilité a posteriori pour démontrer la supériorité d'un nouveau dispositif médical par rapport à un autre de référence, en utilisant une seule analyse intermédiaire.

Chapitre 1

Tests statistiques dans les essais cliniques

Dans ce chapitre, nous allons introduire la notion des essais cliniques, et ses différentes phases. Ainsi l'importance d'analyses intermédiaire et l'utilisation des tests statistiques pour analyser les résultats de ces essais. Nous expliciterons aussi les tests d'équivalence ou de non-infériorité qui sont la méthode la plus répondeuse pour définir une règle d'arrêt, pour conclure la supériorité d'un traitement.

1.1 Essais cliniques

Les essais cliniques sont des études menées chez l'être humain pour démontrer la validité d'un nouveau produit thérapeutique (p.ex. un nouveau médicament, une association de médicaments ou une méthode de diagnostic) [19]. Ils ont pour but d'évaluer les produits thérapeutiques et d'estimer le bénéfice de leur utilisation favorable. Au cours du développement clinique leurs conditions optimales d'emploi sont déterminées. Le développement d'un médicament entre le brevet et la commercialisation dure en moyenne 12 ans.

1.1.1 Les différentes phases d'un essai clinique

On distingue **IV** phases dans le développement d'un produit thérapeutique :

- **Phase I** : Le nouveau produit est testé pour la première fois chez l'homme, en général

chez des volontaires sains. Dans certains cas, il n'est pas éthique de tester le produit sur les volontaires humains sains, comme en oncologie, par exemple où le produit sera administré à des malades ayant déjà reçu le traitement standard. Avant d'entamer une phase I, le produit a été évalué sur des cultures cellulaires en laboratoire et sur des animaux. Le but de cette phase est de définir la tolérance du nouveau produit et de vérifier si les résultats toxicologiques obtenus lors du développement préclinique sont comparables à ceux obtenus chez l'homme. La 1^{ère} phase permet de déterminer la dose maximale tolérée chez l'être humain.

Ces essais sont souvent menés sur quelques dizaines (20 à 100) volontaires sains [9].

• **Phase II** : Lors de la 2^{ème} phase, on s'intéresse à la détermination de la posologie optimale du produit en termes d'efficacité et de tolérance sur une population de patients limitée et homogène.

Ces essais sont souvent menés sur des petits groupes de sujets (100 à 200 sujets).

• **Phase III** : Cette étape a pour objectif de prouver l'efficacité du nouveau produit par rapport à un produit de référence déjà commercialisé ou par rapport à un placebo, c'est-à-dire un traitement sans activité pharmacologique.

Les essais concernent un plus grand nombre de patients (500 à 3000) qui sont des malades volontaires.

• **Phase IV** : Il s'agit de la seule phase qui est réalisée après la commercialisation du produit. Ces essais vont approfondir les connaissances du produit dans les conditions réelles d'utilisation et de ses effets secondaires sur une population importante.

Les phases **II**, **III** et **IV** peuvent comprendre plusieurs essais cliniques.

Définition 1.1.1 (Placebo).

C'est un produit qui ressemble en tout point à la molécule active étudiée dans l'essai clinique mais qui ne contient pas de substance active. Le placebo sert de comparateur au candidat médicament étudié.

1.1.2 Essais de phase II

Dans cet mémoire, nous abordons les considérations conceptuelles et statistiques pour le développement des essais cliniques de phase intermédiaire (phase II et phase III), où de la toxicité on passe à l'efficacité et où les essais sont menés sur des groupes beaucoup plus importants de patients (disons, 40 à 200) [4]. Nous approfondissons également le

sujet des conceptions adaptatives, sujet d'intérêt et d'utilité grandissant dans les études de phase intermédiaire. Après avoir obtenu des informations préliminaires sur le profil d'innocuité, la dose et le schéma d'administration d'un médicament au début (phase I) du développement, le prochain problème est d'examiner si un médicament a une efficacité suffisante pour justifier un développement ultérieur. Les études de phase II peuvent être divisées en deux parties :

Phase IIA : L'évaluation initiale de l'efficacité du médicament est l'objectif principal des essais de cette phase. Généralement, les essais de phase IIA sont menés en tant qu'études à un seul bras pour évaluer l'efficacité de nouveaux médicaments, dans le but d'éliminer ceux qui sont inefficaces.

Phase IIB : Les essais de phase IIB sont des études multi-bras pour comparer l'efficacité du nouveau médicament par rapport au traitement standard ou à d'autres médicaments expérimentaux, de sorte que le plus prometteur peut être sélectionné pour une évaluation à grande échelle.

Le profil de toxicité des nouveaux agents peut également être évalué dans des études de phase II.

1.2 Tests d'hypothèses

Un test d'hypothèses est une procédure basée sur l'observation d'un ou plusieurs échantillons permettant de faire un choix entre deux hypothèses formulées.

Plus précisément un test d'hypothèse consiste à décider d'accepter ou de rejeter une hypothèse spécifiant que θ appartient à un ensemble de valeurs Θ_0 , cette hypothèse de référence est appelée hypothèse nulle et est notée H_0 . Au contraire, on définit l'hypothèse alternative notée H_1 pour laquelle θ appartient à $\Theta_1 = \Theta - \Theta_0$, où $\Theta - \Theta_0$ dénote le complémentaire de Θ_0 par rapport à Θ . En bref on identifiera cette situation en écrivant que l'on teste :

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

Définition 1.2.1. *Un test d'hypothèse H_0 contre l'hypothèse H_1 de niveau α est la donnée d'un ensemble aléatoire R qui dépend uniquement de X tel que :*

$$\sup_{\theta \in \Theta_0} P_{\theta}(X \in R(X)) \leq \alpha$$

$R(X)$ s'appelle le domaine (la région) de rejet, si $X \in R(X)$ on rejette H_0 , sinon on l'accepte.

Définition 1.2.2. (Risque de première espèce)

On appelle risque de première espèce du test la probabilité de rejeter à tort l'hypothèse H_0 , c'est-à-dire, rejeter H_0 alors que H_0 est vraie,

$$\alpha(\theta) = P_{H_0}(\{\text{accepter } H_1\}) = P_{H_0}(R)$$

Où P_{H_0} est la loi de probabilité sous l'hypothèse H_0 .

Définition 1.2.3. (Risque de seconde espèce)

On appelle risque de seconde espèce du test, la probabilité de rejeter à tort l'hypothèse H_1 , c'est-à-dire, accepter H_0 alors que H_0 est non vraie,

$$\beta(\theta) = P_{H_1}(\{\text{accepter } H_0\}) = P_{H_1}(R^c)$$

Où P_{H_1} est la loi de probabilité sous l'hypothèse H_1 .

Définition 1.2.4. Le niveau du test est la probabilité maximale de rejeter à tort l'hypothèse nulle,

$$\alpha = \sup P_\theta(R), \theta \in \Theta_0.$$

On dit qu'un test est **exactement de niveau** α , $\alpha \in [0, 1]$ si et seulement si

$$\forall \theta \in \Theta_0, P_\theta(R) \leq \alpha.$$

Définition 1.2.5. La puissance du test est une fonction de Θ_1 dans $[0, 1]$ définie par

$$\text{puissance}(\theta) = P_\theta(R), \theta \in \Theta_1.$$

Il est clair que $\text{puissance}(\theta) = 1 - \beta(\theta)$.

Définition 1.2.6. Un test est dit sans biais si

$$\text{puissance}(\theta) \geq \alpha, \forall \theta \in \Theta_1,$$

α étant le niveau du test.

1.2.1 Test randomisé

Définition 1.2.7. (*Modèle statistique*)

on appelle "Modèle statistique" le triple $(\mathcal{X}; \mathcal{A}; \mathbb{P})$ où

$$\left\{ \begin{array}{l} \mathcal{X} \text{ est un ensemble appelé espace des résultats.} \\ \mathcal{A} \text{ est une tribu de parties de } \mathcal{X}. \\ \mathbb{P} \text{ est une famille de probabilité sur l'espace mesurable } (\mathcal{X}; \mathcal{A}). \end{array} \right.$$

Définition 1.2.8. On appelle test randomisé une fonction φ mesurable de $(\mathcal{X}, \mathcal{A})$ dans $[0, 1]$ telle que $\varphi(w) = \gamma$ avec

- si $\gamma = 1$: on décide H_0 ,
- si $\gamma = 0$: on décide H_1 ,
- si $0 < \gamma < 1$: on effectue un tirage au sort auxiliaire, indépendant de l'expérience, à valeurs dans $\{0, 1\}$:

$$\left\{ \begin{array}{l} P\{0\} = \gamma, \\ P\{1\} = 1 - \gamma. \end{array} \right.$$

Proposition 1.2.1. [15] Il existe toujours un test randomisé de niveau exactement α .

Remarque 1.2.1. Par opposition, il n'existe pas toujours de test non randomisé de niveau exactement α .

Définition 1.2.9. Un test randomisé est de niveau exactement α si et seulement si

$$\left\{ \begin{array}{l} \forall \theta \in \Theta_0, 1 - E_\theta(\varphi) \leq \alpha \\ \exists \theta \in \Theta_0, 1 - E_\theta(\varphi) = \alpha. \end{array} \right.$$

i) Le risque de première espèce vaut $1 - E_\theta(\varphi)$, $\theta \in \Theta_0$.

ii) Le risque de seconde espèce vaut $E_\theta(\varphi)$, $\theta \in \Theta_1$.

Définition 1.2.10. On appelle puissance du test randomisé φ la quantité $1 - E_\theta(\varphi)$, $\theta \in \Theta_1$.

Définition 1.2.11. Soient φ et φ' deux tests de niveau α , on dit que φ est plus puissant que φ' si

$$\forall \theta \in \Theta_1, \text{puissance}(\theta) = 1 - E_\theta(\varphi) \geq 1 - E_\theta(\varphi') = \text{puissance}'(\theta).$$

Définition 1.2.12. Un test φ est dit uniformément le plus puissant (UPP) de niveau α si

$$\left\{ \begin{array}{l} \varphi \text{ est de niveau } \alpha, \\ \forall \varphi' \text{ un test de niveau } \alpha, 1 - E_{\theta}(\varphi) \geq 1 - E_{\theta}(\varphi'), \theta \in \Theta_1 \end{array} \right.$$

On dit aussi qu'un tel test est optimal.

Exemple 1.2.1. Supposons que deux laboratoires A et B produisent le même type des dispositifs médicaux, mais le laboratoire A fournit un produit plus cher et d'efficacité supérieure. L'efficacité d'un dispositif médical se mesure à une entité aléatoire qui est de loi $\mathcal{N}(5; 1)$ pour le laboratoire A et $\mathcal{N}(4; 1)$ pour le laboratoire B, et ne diffère donc que par la moyenne. Un client achète le dispositif médical le plus cher par lots de 10 et désire développer un test pour contrôler qu'un lot donnée provient bien du laboratoire A. Comme accuser le producteur à tort peut avoir de graves conséquences, il doit limiter le risque correspondant et tester $H_0 : \mu = 5$ contre $H_1 : \mu = 4$, à un niveau 0.05 par exemple. Il semble naturel d'utiliser comme statistique de test la moyenne \bar{X} du lot. Sous H_0 sa loi est $\mathcal{N}(5, 1/10)$ et l'on a alors l'intervalle de probabilité 0.95 : $[5 - 1.96/\sqrt{10}, 5 + 1.96/\sqrt{10}]$, soit $[4.38, 5.62]$. D'où une règle de décision simple :

- Accepter H_0 si la réalisation \bar{x} (moyenne de lot considéré) de \bar{X} est dans $[4.38, 5.62]$.
- Rejeter sinon.

Il est possible de calculer la puissance d'un test puisque la loi de \bar{X} est connue sous H_1 : c'est la loi $\mathcal{N}(4, 1/10)$. Le risque de deuxième espèce vaut :

$$\begin{aligned} \beta &= P(4.38 < \bar{X} < 5.62 \mid H_1) = P\left(\frac{4.38 - 4}{1/\sqrt{10}} < Z < \frac{5.62 - 4}{1/\sqrt{10}}\right) \\ &= P(1.20 < Z < 5.12) \simeq 0.115, \text{ avec } Z \sim N(0,1). \end{aligned}$$

D'où une puissance d'environ 0.885. Notons que l'on peut obtenir un test plus puissant en prenant comme région d'acceptation l'intervalle de probabilité 0.95 : $[5 - 1.645/\sqrt{10}, +\infty[$ Où -1.64 est le quantile d'ordre 0.95 de la loi $\mathcal{N}(0, 1)$, soit $[4.48, +\infty[$. En effet

$$\beta = P(4.48 < \bar{X} \mid H_1) = P\left(\frac{4.48 - 4}{1/\sqrt{10}} < Z\right) = P(1.52 < Z) = P(1.52 < Z) \simeq 0.064,$$

ce qui donne une puissance de 0.936. Intuitivement, on sent bien que, dans le premier test, il est peu pertinent de borner la zone d'acceptation vers le haut car cela conduit à rejeter l'hypothèse nulle pour de très grandes valeurs de \bar{x} , au-delà de 5.62.

Définition 1.2.13. (Notion de p-valeur) La fixation du niveau α du test ne précède pas forcément la détermination de la règle de décision. En effet une démarche inverse est alors

utilisée : on calcule la probabilité, en supposant que H_0 est vraie, d'obtenir une valeur de la variable de décision au moins aussi grande que la valeur de la statistique que l'on a obtenu avec notre échantillon. Cette probabilité est appelée la *p-valeur* (*p-value*).

La connaissance de la *p-valeur* permet donc de décider si on accepte ou non H_0 .

Règle de décision : Si la *p-valeur* est supérieure au niveau α du test, on accepte H_0 , sinon on rejette H_0 .

Mais l'intérêt de la *p-valeur* est qu'elle donne plus d'information que cela : son écart avec α quantifie la marge avec laquelle on rejette H_0 dans le cas contraire.

Exemple 1.2.2. Pour $X \sim \mathcal{N}(\theta, 1)$ et $H_0 : \theta = 0$ contre $H_1 : \theta \neq 0$, la région critique (c'est-à-dire la région de rejet de H_0 du test) est de la forme $\{|x| > k_\alpha\}$ et la *p-valeur* est donnée par :

$$\begin{aligned} p(x) &= \inf_{\{|x| > k_\alpha\}} \alpha = P_X(|X| > |x|), \text{ ou } X \sim \mathcal{N}(0,1) \\ &= 1 - P_X(|X| < |x|) \\ &= 2[1 - \phi(x)]. \end{aligned}$$

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

Remarque 1.2.2. La règle d'acceptation qui précède montre que la *p-valeur* est également définie comme le niveau de test à partir duquel on rejette H_0 .

1.2.2 Application dans les essais cliniques

Mettre en place et tester des hypothèses est une partie essentielle de l'inférence statistique. Afin de formuler un tel test, des recherches antérieures ont généralement été effectuées. Cela est fait soit parce que certaines hypothèses sont vraies ou parce qu'elles doivent servir de base à un argument. La plus part des tests dans les essais cliniques sont conçus pour tester l'efficacité d'un traitement ou d'un dispositif médical et la plus part des temps, on travaille avec les modèles binomiale et leurs approximations par la loi normale [17].

La situation générale est alors la suivante. Nous avons un ensemble d'observations x_1, \dots, x_N qui sont des réalisations de variables aléatoires indépendantes identiquement distribuées (i.i.d.) X_1, \dots, X_N de la distribution paramétrique F_p avec $p \in \Theta$.

Supposons en outre que $\Theta_0 \subset \Theta$ et $\Theta_1 = \Theta/\Theta_0$. Nous voulons maintenant tester $H_0 : p \in \Theta_0$ contre l'hypothèse alternative $H_1 : p \in \Theta_1$, c'est-à-dire que nous voulons arriver à l'une des conclusions suivantes :

1. Rejeter H_0 .
2. Ne pas rejeter H_0 .

Les types d'erreurs que nous pouvons commettre, sont donnés dans le tableau 1.

Décision \ Vérité	H_0	H_1
H_0	$1 - \alpha$	β
H_1	α	$1 - \beta$

Table 1 : Résumé des résultats possibles de tout test d'hypothèse

Choisissons un niveau de signification α_0 et construisons un test tel que :

1. L'erreur de type *I* doit être plus petite ou égale à α_0 .
2. Réduisez l'erreur de type *II*.

Cela montre que nous concevons le test pour limiter l'erreur de type *I*. Pour l'erreur de type *II*, nous n'essayons pas d'atteindre une valeur prescrite. Le test peut être décrit par une région critique (ou de rejet) R avec $R \in \mathbb{R}^N$. Si $x = (x_1, \dots, x_N)$ est un vecteur de réalisations de $X = (X_1, \dots, X_N)$ alors nous avons ce qui suit :

1. Rejeter H_0 si $x \in R$.
2. Ne pas rejeter H_0 si $x \notin R$.

Par la conception du test, la probabilité de faire une erreur dans la première conclusion est limitée par α_0 . Il n'y a pas de prescription sur la probabilité de commettre une erreur dans la deuxième conclusion. D'où :

1. $F_P(x, x \in R)$ Pour chacun $p \in \Theta_0$.
2. Pour $p \notin \Theta_0$, minimiser $F_P(x, x \notin R)$, ou maximiser $F_P(x, x \in R)$.

La région de rejet R peut être dérivée en résolvant ces contraintes. Notez que la fonction $p \mapsto F_P(x, x \in R)$ s'appelle la puissance du test avec une région critique R , et $\sup_{p \in \Theta_0} F_P(x, x \in R) = \alpha$ et s'appelle la taille (le niveau de signification) du test.

Exemple 1.2.3. *Supposons que nous ayons une observation d'une distribution $\mathcal{B}(10, p)$ avec la distribution paramétrique $F_P : p \in \Theta = [0, 1]$. En outre, nous avons les hypothèses suivantes pour tester :*

$$H_0 : p \in \Theta_0 = [0, 1/4]$$

$$H_1 : p \in \Theta \setminus \Theta_0$$

Notons que les hypothèses ci-dessus peuvent également être reformulées

$$H_0 : p \leq p_0 : p \in \Theta_0$$

$$H_1 : p \geq p_1 : p_1 \in \Theta_1$$

Il est maintenant raisonnable d'utiliser une région critique de la forme $[t, 10]$, pour certains $t = 0, 1, 2, \dots, 10$. Nous choisissons le niveau de signification pour être égal à 0.05 et dériver t en résolvant les contraintes données ci-dessus. Par conséquent, nous avons $\alpha = \sup_{0 \leq p \leq 1/4} (1 - F_p(t - 1)) = 1 - F_{1/4}(t - 1)$. Puisque, X a une distribution binomiale et donc :

$$1 - F_{1/4}(t - 1) = 1 - B(t - 1, 1/4, 10) = \begin{cases} 0.0781 \dots & t=5 \\ 0.0197 \dots & t=6 \end{cases}$$

Notez que parce que la distribution binomiale est discrète, il n'est pas possible d'obtenir exactement un niveau de signification qui est égal à 0.05 et donc nous choisissons la région critique comme intervalle $[6, 10]$. Après avoir dérivé la région critique, les conclusions suivantes peuvent être faites

1. Rejeter H_0 si $x \in [6, 10]$.
2. Ne pas rejeter H_0 si $x \in [0, 5]$

Dans cet exemple nous avons vu que pour un ensemble donné N observations $\mathcal{B}(N, p)$, un niveau de signification donné et une Hypothèse donnée, nous pouvons construire un test qui rejette l'hypothèse nulle en faveur de l'hypothèse alternative. Cela a été fait en dérivant la région critique comme nous l'avons vu ci-dessus.

Inversement, comme nous le verrons dans la section 2 et plus loin dans ce mémoire, nous sommes plus intéressés par le choix de N tel que pour un donné $[p_0, p_1]$, nous avons une erreur de type I et une erreur de type II bien spécifiée.

1.3 Tests d'équivalence et de non-infériorité

Dans certains cas, un produit expérimental est comparé à un traitement de référence qui est l'objectif des essais de phase IIB et de phase III. L'objectif des tests d'équivalence et de non-infériorité dans ce cas est d'étudier l'efficacité de produit expérimental [14]. On donne les caractéristiques des deux tests statistiques [26].

1.3.1 Tests d'équivalence

Soit :

- p_e Le taux d'efficacité de traitement expérimental.
- p_s Le taux d'efficacité de traitement standard [6].

- Pour traiter une maladie, on dispose d'un traitement standard et, de plus, on étudie un nouveau traitement, expérimental.
- On juge la qualité de chacun des traitements à l'aide du pourcentage des patients guéris (ou bien soulagés, par exemple).

Le non-rejet de $H_0 : \langle p_e = p_s \rangle$ n'est pas une démonstration de H_0 .

L'échange littéral de $H_0 : \langle p_e = p_s \rangle$ et $H_1 : \langle p_e \neq p_s \rangle$ n'étant pas possible, on se contentera d'imposer un maximum d'écart entre l'efficacité du traitement standard et l'efficacité du traitement expérimental en prenant.

- $H_0 : \langle |p_e - p_s| \geq \delta \rangle$ (c'est la non-équivalence)
- $H_1 : \langle |p_e - p_s| < \delta \rangle$ (c'est l'équivalence)

Le choix de la valeur de δ relevant, certes, de l'équipe médicale et/ou du statisticien, mais essentiellement des « bonnes pratiques » préconisées par les organismes de contrôle, nationaux et internationaux.

Test d'équivalence et l'intervalles de confiance

On peut aussi écrire un test d'équivalence comme suite :

- $H_0 : \langle p_s - p_e \geq \delta \text{ ou } p_e - p_s \geq \delta \rangle$
- $H_1 : \langle p_s - p_e < \delta \text{ et } p_e - p_s < \delta \rangle$.

Donc, on obtient un test d'hypothèse de H_0 contre H_1 en rejetant H_0 lorsque l'intervalle de confiance de $p_s - p_e$ de niveau $1 - 2\alpha$ est entièrement inclus dans $] -\delta; \delta[$.

En effet, la probabilité, dans le cas où H_0 est vraie, de rejeter H_0 par erreur par cette méthode, se calcule de deux façons différentes, selon que la vraie valeur de $p_s - p_e$ est supérieure à δ ou bien inférieure à $-\delta$, ces deux cas étant les seuls possibles pour que H_0 soit vraie [23].

Dans les deux cas, le rejet ne peut avoir lieu, par définition, que si l'intervalle de confiance est inclus dans $] -\delta; \delta[$.

Rappelons aussi que c'est l'intervalle de confiance qui est aléatoire, alors que $p_s - p_e$ est une valeur de la nature, non aléatoire.

Dans le premier cas, cette probabilité est donc inférieure à la probabilité que tous les éléments de l'intervalle de confiance soient inférieurs à $p_s - p_e$, cette dernière probabilité étant par définition pour un intervalle de confiance de niveau $1 - 2\alpha$.

Dans le deuxième cas, cette probabilité est cette fois inférieure à la probabilité que tous les éléments de l'intervalle de confiance soient supérieurs à $p_s - p_e$, cette dernière probabilité étant aussi, par définition, pour un intervalle de confiance de niveau $1 - 2\alpha$.

Par disjonction des cas, a été prouvé qu'il s'agit bien d'un test de niveau inférieur ou égal à α de H_0 contre H_1 ; cependant, rien ne prouve qu'il ait une puissance optimale [6].

1.3.2 Tests de non-infériorité

Pour les études de médicaments avec contrôle actif, on utilise quasi exclusivement la version unilatérale de ce test [6]. On note p_s la probabilité de succès avec la substance active de référence, p_e celle qui correspond au produit en cours d'essai.

$$- H_0 : \langle p_e - p_s \leq -\delta \rangle \quad \text{ou encore} \quad -H_0 : \langle p_s - p_e \geq \delta \rangle$$

$$-H_1 : \langle p_e - p_s > -\delta. \quad \quad \quad -H_1 : \langle p_s - p_e < \delta \rangle$$

Rejeter H_0 signifie ici que

$$- p_e > p_s \text{ si } p_s - p_e < 0$$

$$- \text{ou, sinon, que } p_e \text{ n'est pas trop inférieur à } p_s : p_s \geq p_e > p_s - \delta.$$

Le taux d'erreur de type I pour ces tests est la probabilité que nous concluons que le groupe de traitement n'est pas inférieur au contrôle, alors qu'en fait il est inférieur. Lorsque nous contrôlons l'erreur de type I pour ces hypothèses de non-infériorité, nous protégeons contre la non-infériorité de conclusion incorrecte pour le groupe de traitement par rapport au groupe de contrôle.

Mise en oeuvre d'un test de non-infériorité

Comme on l'a vu, on définit

$$- H_0 : \langle p_s - p_e \geq \delta \rangle.$$

$$- \text{et } H_1 : \langle p_s - p_e < \delta \rangle.$$

et, donc, on peut utiliser comme valeur de la réalisation de la statistique de test asymptotiquement normale centrée réduite :

$$t = \frac{p_{s_c} - p_{e_c} - \delta}{\sqrt{\left(\frac{p_{s_c}(1-p_{s_c})}{n_s} + \frac{p_{e_c}(1-p_{e_c})}{n_e} \right)}}$$

(translation des données de δ vers la gauche).

–Ce sont les valeurs trop négatives de t qui permettent de rejeter H_0 .

–La zone de rejet est donc $] -\infty; -Z_{1-\alpha}[$, $Z_{1-\alpha}$ étant le fractile d'ordre $1 - \alpha$ de la loi $\mathcal{N}(0, 1)$, ce qui donne $] -\infty; -1,65[$ au risque $\alpha = 5\%$.

Exemple 1.3.1. Cellules souches

L'allogreffe de cellules souches hématopoïétiques est utilisée pour le traitement de certaines maladies hématologiques. Une étude de 2007 a eu pour but de comparer l'emploi de cellules sources du sang périphérique (PBSC) avec celui de la moëlle osseuse (BM).

On voit que l'avantage du sang périphérique beaucoup moins invasif pour le donneur ; de plus, des études indiquent qu'il a une meilleure efficacité que BM dans le cas des donneurs jumeaux des receveurs. Cependant, dans le cas des donneurs non apparentés aux receveurs, le risque de réaction du greffon contre l'hôte semble accru. On prend donc ici BM comme traitement expérimental, PBSC comme traitement de contrôle (de référence) et on cherche à savoir si BM n'est, au pire, que peu inférieur à PBSC pour la survie à 6 mois. Le choix a été fait de $\delta = 0.1 = 10\%$.

Mise en oeuvre du test

Données	n	mortalité au traitement
BM	583	187
PBSC	328	95

La statistique de test vaut donc

$$t = \frac{\frac{233}{328} - \frac{396}{583} - 0,1}{\sqrt{\frac{\frac{233}{328}(1-\frac{233}{328})}{328} + \frac{\frac{396}{583}(1-\frac{396}{583})}{583}}} \cong -2,18$$

- Or la zone de rejet pour un « z-test » unilatéral à gauche au risque $\alpha = 5\%$ est environ $] -\infty; -1,65[$.
- Donc, à ce risque, on rejette l'hypothèse $H_0 : \langle p_s - p_e \geq \delta \rangle$ et on déclare que BM est non-inférieur à PBSC du point de vue de la mortalité liée au traitement.

Exemple 1.3.2. « Comparaison d'anesthésiques [12]. a comparé l'efficacité et la sécurité des effets anti-émétiques de l'ondansétron par voie intraveineuse avec ceux de la metoclopramide, lors de leur administration préchirurgicale pour les patients qui a subi une cholécystectomie laparoscopique sous anesthésie générale intraveineuse (TIVA).

Dans cette étude, 80 patients ont été répartis au hasard en deux groupes recevant respectivement de l'ondansétron (4 mg) ou du dropéridol (1.25 mg), donnés en une seule dose intraveineuse immédiatement avant la mise en oeuvre d'une anesthésie générale standard.

Un (autre) indice important pour la comparaison est le score relatif au "bien-être" du patient. Celui-ci est évalué sur une échelle comportant trois valeurs : "pauvre", "modéré", et "confortable". Des études antérieures ont généralement suggéré l'absence de différence entre les scores de bien-être de l'ondansétron et du dropéridol.

Par conséquent, il est intéressant de déterminer si, pour les deux groupes, "pauvre ou modéré" et "confortable" Dropéridol et Ondansetron sont équivalents en termes de bien-être ».

Score de bien-être du dropéridol et de l'ondansetron		
	Score de bien-être	
	Pauvre ou modéré	Confortable
Dropéridol	12	28
Ondansetron	9	31

Table 2 : Résultats de la répartition des deux groupes selon le type d'anesthésie

Appliquons la méthode que nous avons vue. Borne gauche de l'intervalle de confiance de $p_e - p_s$:

$$t = \frac{28}{40} - \frac{31}{40} - z_{1-\alpha} * \sqrt{\left(\frac{\frac{28}{40} \cdot (1 - \frac{28}{40})}{40} + \frac{\frac{31}{40} \cdot (1 - \frac{31}{40})}{40}\right)}$$

$$\cong -0,24$$

Borne droite :

$$t = \frac{28}{40} - \frac{31}{40} + z_{1-\alpha} * \sqrt{\left(\frac{\frac{28}{40} \cdot (1 - \frac{28}{40})}{40} + \frac{\frac{31}{40} \cdot (1 - \frac{31}{40})}{40}\right)}$$

$$\cong 0,09$$

Ici, si l'écart maximal $\delta = 0,1$ est utilisé, on ne peut rien conclure car l'intervalle de confiance, approximativement $[-0,24; 0,09]$, n'est pas inclus dans $[-0,1; 0,1]$.

Test de non-infériorité et intervalle de confiance

En admettant que l'approximation normale soit justifiée, un intervalle de confiance de niveau $1 - \alpha$ de la différence $p_e - p_s$ est $[p_{e_c} - p_{s_c} - z_{1-\frac{\alpha}{2}}\sqrt{v}; p_{e_c} - p_{s_c} + z_{1-\frac{\alpha}{2}}\sqrt{v}]$ où $v = \frac{p_{e_c}(1-p_{e_c})}{n_{e_c}} + \frac{p_{s_c}(1-p_{s_c})}{n_{s_c}}$ et z_p est le fractile d'ordre p de $\mathcal{N}(0,1)$. On définit aussi intervalle

de confiance à gauche de niveau $1 - \alpha$:

$[p_{ec} - p_{sc} - z_{1-\alpha}\sqrt{v}; +\infty[$; de même à droite. Or, dans un test de non-infériorité, H_0 : $p_e - p_s \leq -\delta$ est rejetée si et seulement si

$$\frac{p_{ec} - p_{sc} + \delta}{\sqrt{\left(\frac{p_{ec}(1-p_{ec})}{n_e} + \frac{p_{sc}(1-p_{sc})}{n_s}\right)}} > z_{1-\alpha}$$

c'est-à-dire $p_{ec} - p_{sc} - z_{1-\alpha}\sqrt{\left(\frac{p_{sc}(1-p_{sc})}{n_s} + \frac{p_{ec}(1-p_{ec})}{n_e}\right)} > -\delta$. On voit donc que H_0 est rejetée si et seulement si l'intervalle de confiance à gauche de $p_e - p_s$ de niveau $1 - \alpha$ est inclus dans $] - \delta; +\infty[$.

Test de non-infériorité et intervalle de confiance symétrique

Si on ne veut pas utiliser d'intervalle de confiance non symétrique, on peut aussi dire que H_0 est rejetée si et seulement si l'intervalle de confiance de $p_e - p_s$ de niveau $1 - 2\alpha$ a sa borne gauche incluse dans $] - \delta; +\infty[$.

1.4 Analyses intermédiaires dans les essais de dispositifs médicaux

Les analyses intermédiaires sont des analyses des données réalisées en cours d'essai (L'information s'accumule progressivement au fur et à mesure des inclusions et du suivi des patients). Des analyses intermédiaires à la recherche de l'effet du traitement en cours d'essai sont envisageables pour diverses raisons avant que tous les patients prévus aient été recrutés et avant la fin de la période de suivi initialement prévue.

1.4.1 Motivation

D'après (ZERARI. A) [24] Le but de ces analyses intermédiaires est :

- **Arrêt pour efficacité** : Pouvoir détecter au plus tôt le bénéfice du traitement afin d'éviter de traiter des patients par un placebo alors que les données amassées sont suffisantes pour conclure à l'efficacité du traitement étudié. De plus, la confirmation au plus tôt du bénéfice apporté par un traitement permet de faire bénéficier du traitement tous les patients hors essai le plus rapidement possible.

• **Arrêt pour toxicité** : Donner les moyens de détecter au plus tôt un éventuel effet délétère du traitement afin de limiter le nombre de patients exposés au risque. Dans ces deux circonstances, le but des analyses intermédiaires est d'éviter de continuer à inclure des patients alors que l'on dispose d'une réponse suffisamment fiable à la question posée.

• **Arrêt pour futilité** : Arrêter une étude dont on peut prédire avec une certitude raisonnable qu'elle ne pourra pas aboutir. L'arrêt précoce permettra de diriger les ressources vers le test de nouvelles hypothèses. La réalisation de ces analyses pose cependant un certain nombre de problèmes méthodologiques et nécessitent une méthodologie adaptée.

1.4.2 Objectifs de surveillance des essais

Les analyses intermédiaires s'intègrent dans un processus global de surveillance des essais.

À côté de la recherche anticipée d'un effet du traitement et de la protection des personnes inclusés dans l'essai. Cette surveillance a pour objectif de vérifier le bon déroulement de l'essai.

Il s'agit d'éviter des dérives dans la réalisation de l'essai, qui, si elles n'étaient détectées qu'à la fin, rendraient l'essai inutilisable en raison de défauts de qualité rédhibitoires.

Les éléments à surveiller sont les suivants :

- Le taux d'écart au protocole : l'essai est-il de qualité ?
- Le taux d'inclusion : est-ce que l'essai pourra être réalisé dans un délai acceptable ?
- Les caractéristiques des patients inclus : le risque de base des patients effectivement inclus correspond t-il à celui initialement prévu et utilisé dans les calcul du nombre de sujets nécessaire ? Les patients recrutés correspondent-ils à la population cible de l'essai ?

Cette surveillance permet de prendre au plus tôt des mesures correctrices. Les centres investigateurs ayant des difficultés à suivre le protocole pourront rectifier le tir. En cas de taux de recrutement insuffisant, d'autres centres investigateurs pourront être recrutés afin d'éviter qu'un essai dure trop longtemps. En effet, une durée excessive limite l'intérêt d'un essai.

Cette surveillance par analyse des données amassées se superpose à la surveillance « de terrain » de l'essai (appelé parfois « monitoring ») qui est focalisée sur le contrôle de qualité des données (visite de centres, contrôles des données, audit).

Exemple 1.4.1. *Le but de cet exemple est d'étudier l'arrêt d'analyse intermédiaire [24].*

Le nombre total de comparaisons qui seront effectuées est de 4 : les 3 intermédiaires plus la comparaison finale. Le seuil à utiliser pour chacune de ces analyses est de $5\%/4 = 1.25\%$. Si une p -valeur (p) inférieure à 1.25% est obtenu à l'une des analyses intermédiaires, il est alors possible de conclure et d'arrêter l'essai sans attendre la fin du recrutement prévu.

On prend 4 cas comme suit :

1^{er} cas : *Dans la situation dépeinte par le tableau ci dessous, l'essai peut être arrêté à la 2^{ème} analyse intermédiaire. Le p obtenu lors de cette analyse est inférieur au seuil de signification corrigé et l'essai peut donc être arrêté prématurément. Cette situation met en avant tout l'intérêt des analyses intermédiaires.*

Analyses intermédiaires			Analyse finale
1	2	3	
$p = 0.10$	$p = 0.011$		

2^{ème} cas : *Dans ce deuxième exemple, $p < 5\%$ lors de la troisième analyse intermédiaire ne permet pas de conclure à une différence significative car la valeur obtenue reste supérieure au seuil corrigé pour 4 tests ($1,25\%$). L'essai va donc à son terme et lors de l'analyse finale le p devient inférieur au seuil corrigé ce qui donne donc finalement un résultat statistiquement significatif.*

Analyses intermédiaires			Analyse finale
1	2	3	
$p = 0.25$	$p = 0.08$	$p = 0.04$	$p = 0.012$

3^{ème} cas : *Le cas suivant peut paraître déroutant. Aucune analyse intermédiaire ne conduit à interrompre prématurément l'essai. Lors de l'analyse finale un p de 4% est obtenu. Cette valeur, bien qu'elle soit inférieure à 5% n'autorise pas à conclure à un résultat statistiquement significatif car elle reste supérieure au seuil corrigé. Il ne peut pas être considéré comme significatif car du risque alpha a été consommé au cours des analyses précédentes, effritant le contrôle du risque d'erreur de première espèce apporté par un $p < 5\%$ au niveau d'une comparaison donnée.*

Analyses intermédiaires			Analyse finale
1	2	3	
$p = 0.42$	$p = 0.28$	$p = 0.12$	$p = 0.04$

4^{ème} *cas* : Dans le dernier cas, aucune analyse n'atteint le seuil corrigé de signification statistique. L'essai n'obtient donc pas de résultat statistiquement significatif.

<i>Analyses intermédiaires</i>			<i>Analyse</i>
<i>1</i>	<i>2</i>	<i>3</i>	<i>finale</i>
$p = 0.89$	$p = 0.48$	$p = 0.25$	$p = 0.10$

1.4.3 Situations conduisant à un arrêt prématuré

1) L'effectif d'un essai est le nombre de sujets minimal nécessaire pour garantir une probabilité élevée de mettre en évidence l'effet du traitement.

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

2) Inflation du risque alpha : La comparaison répétée de l'efficacité de deux traitements par des tests statistiques successifs accroît le risque de conclure à tort à la supériorité de l'un par rapport à l'autre.

La réalisation de plusieurs analyses statistiques dans la même expérience, pour tester la même hypothèse, conduit à des comparaisons statistiques multiples. A chaque analyse intermédiaire un test statistique est réalisé pour rechercher un effet du traitement. La répétition à chaque test du risque d'obtenir un résultat significatif par le hasard augmente le risque global de conclure à tort à l'efficacité du traitement lors de cet essai. Enfin le risque alpha n'est plus de 5% (même si c'est le seuil retenu pour chaque test) mais il est bien supérieur.

L'utilisation de techniques statistiques adaptées est nécessaire pour empêcher cette augmentation du risque alpha, appelée en jargon statistique « inflation du risque alpha ». Le but de ces méthodes est de garantir un risque global, sur l'ensemble des comparaisons effectuées, de conclure à tort de l'efficacité du traitement de 5%. Sur l'ensemble des comparaisons effectuées le risque d'obtenir au moins un résultat significatif par le fait du hasard est contrôlé et garde sa valeur prédéfinie de 5%.

Plusieurs solutions sont possibles qui sont à la base de différentes méthodes. L'une d'entre elles consiste à diminuer le seuil de signification statistique de chacune des comparaisons intermédiaires, par exemple en divisant le risque alpha global par le nombre de comparaisons effectuées n . C'est la méthode de Bonferroni. Ainsi malgré l'inflation du risque alpha, le risque final de conclure à tort à l'efficacité restera compris dans les valeurs habituelles.

Il existe aussi d'autres méthodes bayésiennes pour résoudre ce problème qu'on verra dans le 2^{ème} chapitre.

1.4.4 Nombre de sujets nécessaire

Le nombre de sujets nécessaires n'est plus un nombre fixe puisque l'essai peut être arrêté après des tailles d'échantillon variables [24]. Deux notions sont couramment utilisées :

–**Le nombre maximal de sujets nécessaires** : est le nombre de sujets qui sera inclus si toutes les analyses intermédiaires sont négatives et si l'essai n'est pas arrêté précocement .

–**Le nombre moyen (ou attendu) de sujets** : est le nombre de sujets qui sera inclus en moyenne avant d'arrêter l'essai. Ce nombre peut se calculer sous l'hypothèse nulle d'absence d'effet du traitement, sous l'hypothèse alternative de supériorité d'un des traitements ou par une pondération de ces deux hypothèses. On pourra également rendre le nombre de sujets nécessaire sous forme de courbes en fonction de l'effet du traitement.

Comparé à un essai à analyse unique, le nombre maximal de sujets sera plus élevé, mais le nombre moyen plus faible. En d'autres termes, un centre qui n'effectuerait que des essais séquentiels conclurait en moyenne plus rapidement qu'un centre qui effectuerait les mêmes essais sans analyse intermédiaire. Mais dans certains cas, notamment lorsqu'il n'existe pas de différence entre les deux bras, l'essai ne sera pas arrêté précocement et le nombre de sujets inclus sera plus grand.

Remarque 1.4.1. *Il faut retenir ici le fait que le calcul du nombre de sujets nécessaires a été fait pour un risque d'erreur α constant et fixé à l'avance. Si on fait une analyse intermédiaire, ce sera forcément avec moins de sujets, donc avec moins de puissance que prévu initialement.*

Il semble donc logique d'utiliser une valeur de plus petite à chaque analyse intermédiaire pour conserver le niveau α final de 5%.

Remarque 1.4.2. [17] Dans le cas classique le nombre maximal de sujets donnée comme suite :

$$N_{max} = \frac{(Z_{\alpha} + Z_{\beta})^2 \bar{p}(1 - \bar{p})}{(p_1 - p_0)^2} \quad \text{Ou} \quad \bar{p} = \frac{(p_0 + p_1)}{2}$$

Ou p_0 est la proportion de réponse indésirable.

et p_1 est la proportion de réponse désirable.

Exemple 1.4.2. Supposons que nous voulons étudier un traitement pour une certaine maladie et que 20% serait un taux de réponse intéressant. D'un autre côté, si la réponse n'était que de 10%, le traitement ne mériterait pas d'être approfondi.

Pour $\alpha = 0.05$ (unilatéral), $\beta = 0.20$,

$-H_0 : p = 0.10$

et

$-H_A : p = 0.20$

On a $\bar{P} = \frac{(0.10+0.20)}{2} = 0.15$, $Z_{1-\alpha} = 1.645$, $Z_{1-\beta} = 0.845$ donc

$$N_{max} = \frac{(1.645 + 0.845)^2 0.15(0.85)}{(0.1)^2}$$

nous constatons que le nombre de patients est égal à : 79.

Chapitre 2

Conditions fréquentistes appliquées sur les tests Bayésiens

L'étude de phase **II** des essais clinique 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 étude [8] [16] [17] [24].

Dans ce chapitre, notre but, est d'étudier cette phase en démontrons l'efficacité et la robustesse de l'inférence bayésienne, en utilisant la probabilité prédictive et la distribution a posteriori, et on termine par une conception bayésienne dans les tests de non infériorité qui justifie l'utilisation de ces designs dans cette phase.

2.1 Modélisation bayésienne

L'analyse statistique bayésienne se réfère principalement à une inversion ; en effet elle vise à déterminer les causes (paramètres du mécanisme générateur) à partir des effets (observations). En d'autres mots, ayant observé un phénomène aléatoire contrôlé par le paramètre θ , une méthode statistique permet de déduire de ces observation une inférence sur θ , alors que la modélisation probabiliste caractérise le comportement des observation futures conditionnellement à θ . Une telle inversion est apparente dans la notion de fonction de vraisemblance puisque formellement, il s'agit simplement d'une densité réécrite dans le bon ordre.

$$L(\theta; x) = f(x/\theta)$$

Soit donc comme fonction de θ , qui est inconnu, dépendante de la valeur observée x .

De manière générale, l'inversion des conditionnements dans les probabilités se fait à partir du théorème de Bayes :

Théorème 2.1.1. (*Théorème de Bayes*)

Soient A et E deux évènements tels que $P(E) \neq 0$,

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

où A^c désigne le complémentaire de A . Bayes donnait aussi une version continue de ce théorème, à savoir, pour deux variable aléatoires X et Y , de densité conditionnelle $f(x/y)$ et marginale $g(y)$, la densité conditionnelle de Y sachant x est

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

Bien que ce théorème d'inversion soit naturel d'un point de vue probabiliste, Bayes et Laplace allaient plus loin et considéraient que l'incertitude sur le paramètre θ d'un modèle peut être décrite par une distribution de probabilité de densité π sur θ appelée distribution a priori. L'inférence est alors basée sur la distribution de θ conditionnelle à x , de densité $\pi(\theta/x)$, appelée distribution a posteriori et définie par :

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

Notons que $m(x) = \int_{\Theta} f(x/\theta)\pi(\theta)d\theta$ ne dépend pas de θ , c'est la densité prédictive de x . La densité de la loi de θ conditionnellement à x est proportionnelle à la densité de la loi de x conditionnellement à θ multipliée par la densité de la distribution a priori de θ .

Définition 2.1.1. *Un modèle statistique bayésien est la double donné d'un modèle paramétrique $\{f(x/\theta) : \theta \in \Theta\}$ et d'une distribution a priori pour les paramètres, de densité $\pi(\theta)$.*

Exemple 2.1.1. (*Bayes, 1763*) *Une boule de billard W roule sur ligne de longueur un, avec une probabilité uniforme de s'arrêter n'importe où.*

Supposons qu'elle s'arrête en p . Une deuxième boule O roule alors n fois dans les mêmes conditions, et on note X le nombre de fois que la boule O s'arrête à gauche de W . Connaissant X . quelle inférence pouvons-nous mener sur p ?

Dans la terminologie moderne, le problème est de déterminer la distribution a posteriori de p conditionnellement à X , quand la distribution a priori de p est uniforme sur $[0, 1]$ et $X \sim \mathcal{B}(n, p)$ une variable aléatoire binomiale. Comme

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

$$P(a < p < b \text{ et } X=x) = \int_a^b \binom{n}{x} p^x (1-p)^{n-x} dp,$$

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

nous trouvons que

$$\begin{aligned} P(a < p < b / X = x) &= \frac{\int_a^b \binom{n}{x} p^x (1-p)^{n-x} dp}{\int_0^1 \binom{n}{x} p^x (1-p)^{n-x} dp} \\ &= \frac{\int_a^b p^x (1-p)^{n-x} dp}{B(x+1, n-x+1)}, \end{aligned}$$

donc la distribution de p conditionnellement à $X = x$ est une distribution Bêta, $\mathcal{Be}(x+1, n-x+1)$.

2.1.1 Distribution a priori et a posteriori

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

Définition 2.1.2. *La loi a priori de densité $\pi(\theta)$ est la loi marginale du paramètre θ , ou en d'autres termes, la loi du paramètre avant que x ne soit observé. Elle résume, donc l'information sur θ disponible a priori (avant observation).*

Définition 2.1.3. *La loi a posteriori de densité $\pi(\theta/x)$ donne alors l'information dont on dispose sur θ , après observation. Elle représente un compromis entre l'information a priori (donné par π), et l'information tirée de l'observation de x (donnée par $f(x/\theta)$).*

Une fois que la densité de distribution a priori sur θ , $\pi(\theta)$, et celle d'échantillonnage $f(x/\theta)$, soient disponibles, nous pouvons en construire plusieurs autres densités, telles que :

1. La densité de la distribution jointe de (θ, x) ,

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

2. La densité de la distribution marginale de x ,

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

Exemple 2.1.2. (Suite de l'exemple 2.1.1) Si $X \sim \mathcal{B}(n, p)$ et $p \sim \mathcal{Be}(\alpha, \beta)$ (avec $\alpha = \beta = 1$ dans le cas particulier de Bayes)

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

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

La densité de la distribution jointe de (x, p) est alors

$$\varphi(x, p) = \frac{\binom{n}{x}}{B(\alpha, \beta)} p^{\alpha+x-1} (1-p)^{n-x+\beta-1},$$

La densité de la distribution marginale de x est

$$m(x) = \frac{\binom{n}{x}}{B(\alpha, \beta)} B(\alpha + x, n - x + \beta),$$

donc la densité de la distribution a posteriori de p est

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

qui est la densité d'une loi $\mathcal{B}\acute{e}ta$, $\mathcal{Be}(\alpha + x, \beta + n - x)$.

Exemple 2.1.3. Considérons une observation x d'une distribution $\mathcal{N}(\frac{\theta_1 + \theta_2}{2}, 1)$ avec une a priori π sur (θ_1, θ_2) telle que $\pi(\theta_1, \theta_2) = \pi_1(\theta_1 + \theta_2)\pi_2(\theta_1 - \theta_2)$. Si nous réalisons le changement de variable

$$\varsigma_1 = \frac{\theta_1 + \theta_2}{2}, \varsigma_2 = \frac{\theta_1 - \theta_2}{2}$$

La densité de la distribution a posteriori de ς_2 est alors

$$\begin{aligned} \pi(\varsigma_2/x) &\propto \int_{\mathbb{R}} \exp\{-(x - \varsigma_1)^2/2\} \pi_1(2\varsigma_1)\pi_2(2\varsigma_2) d\varsigma_1 \\ &\propto \pi_2(2\varsigma_2) \int_{\mathbb{R}} \exp\{-(x - \varsigma_1)^2/2\} \pi_1(2\varsigma_1) d\varsigma_1 \\ &\propto \pi_2(2\varsigma_2), \end{aligned}$$

pour chaque observation x . l'observation n'apporte donc pas d'information sur ς_2 .

Exemple 2.1.4. *Supposons que x suive une distribution normale avec la moyenne θ et la variance $\frac{\sigma^2}{n}$, c'est-à-dire, $X \sim \mathcal{N}(\theta, \frac{\sigma^2}{n})$. Supposons également que le paramètre d'intérêt θ suit une distribution normale avec la moyenne μ et la variance $\frac{\sigma^2}{n_0}$, c'est-à-dire, $\theta \sim \mathcal{N}(\mu, \frac{\sigma^2}{n_0})$. Ainsi, nous avons*

$$\pi(\theta/X) \propto f(X/\theta)\pi(\theta).$$

Par conséquent, la distribution postérieure de θ donnée X peut être obtenue

$$\pi(\theta/X) = C e^{-\frac{(X-\theta)^2 n}{2\sigma^2}} e^{-\frac{(\theta-\mu)^2 n_0}{2\sigma^2}} \quad (2.1)$$

où C est une constante avec θ . On peut vérifier que (2.1) est une distribution normale avec la moyenne $\theta \frac{n_0\mu + nX}{n_0+n}$ et la variance $\frac{\sigma^2}{n_0+n}$, c.-à-d.

$$\theta/X \sim \mathcal{N}\left(\theta \frac{n_0\mu + nX}{n_0+n}, \frac{\sigma^2}{n_0+n}\right).$$

Maintenant, basé sur (2.1), nous pouvons faire des prédictions concernant les valeurs futures de x en prenant en compte l'incertitude sur sa moyenne θ . Dans ce but, nous réécrivons $X = (X - \theta) + \theta$ de sorte que X est la somme de deux grandeurs indépendantes, c'est-à-dire, $(X - \theta) \sim \mathcal{N}(0, \frac{\sigma^2}{n})$ et $\theta \sim \mathcal{N}(\mu, \frac{\sigma^2}{n_0})$. En conséquence, la distribution de probabilité prédictive peut être obtenue comme :

$$X \sim \mathcal{N}\left(\mu, \sigma^2 \left(\frac{1}{n} + \frac{1}{n_0}\right)\right).$$

Notez que si nous observons les premières observations n_1 (c'est-à-dire, la moyenne des premières n_1 , x_{n_1} est connue), alors la distribution de probabilité prédictive est donnée par

$$X/x_{n_1} \sim \mathcal{N}\left(\frac{n_0\mu + n_1x_{n_1}}{n_0+n_1}, \sigma^2 \left(\frac{1}{n_0+n_1} + \frac{1}{n}\right)\right).$$

Les concepts bayésiens de base peuvent être facilement appliqués à certains modèles classiques dans des essais cliniques. Cependant, cela peut affecter la puissance et par conséquent le calcul de la taille de l'échantillon. À des fins d'illustration, nous considérons l'exemple (2.1.8).

2.1.2 Modélisation de l'information a priori

La modélisation a priori est sans doute le plus délicat de l'analyse bayésienne. deux types d'approches sont généralement considérés :

- Une approche dite subjective, ou informative, qui revient à tenir compte (lorsqu'elles existent) d'informations a priori sur le paramètre (expériences précédentes, avis d'experts, connaissances extérieures au processus d'observation, etc).
- Une approche dite objective, ou non informative, qui revient à modéliser l'absence d'information a priori.

Dans la suite, nous présenterons un modèle des lois informatives(lois a priori conjuguées) et un autre des lois non informatives(loi de Jeffreys).

A) Lois a priori conjugués

Une famille conjuguée est une famille de lois liée au processus étudié et qui est particulièrement intéressante dans des modèles paramétriques où des statistiques exhaustives qui peuvent être déniées. Il s'agit d'une famille de lois, c-à-d d'un ensemble de lois, qui soit d'une part, assez riche et flexible pour représenter l'information a priori que l'on a sur les paramètres du modèle et, d'autre part, qui se combine remarquablement avec la fonction de vraisemblance dans la formule de Bayes, pour donner a posteriori une loi sur les paramètres du modèle qui appartienne à la même famille [15].

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

Définition 2.1.4. *Une famille F de distributions de probabilité sur Θ est dite conjuguée (ou fermée par échantillonnage) par une fonction de vraisemblance $f(x/\theta)$ si pour tout $\pi \in F$, la distribution a posteriori $\pi(.|x)$ appartient également à F .*

Définition 2.1.5. Famille exponentielle

La famille exponentielle regroupe les lois de probabilités qui admettent une densité de la forme :

$$f(x/\theta) = h(x)\exp(\alpha(\theta)T(x) - \psi(\theta)), \theta \in \Theta$$

T est une statistique exhaustive. Une telle famille est dite régulière si Θ est un ouvert tel que :

$$\Theta = \left\{ \theta / \int h(x)\exp(\alpha(\theta)T(x) - \psi(\theta))d\mu(x) < \infty \right\}$$

Théorème 2.1.2. [15] Si $x \sim f(x/\theta) = h(x)\exp(\alpha(\theta)T(x) - \psi(\theta))$ alors la famille de lois a priori $\{\pi_{(\lambda,\mu)}(\theta) \propto h(x)\exp(\alpha(\theta)\mu - \lambda\psi(\theta))\}$ est conjuguée.

La loi a posteriori correspondante est :

$$\pi(\theta/\lambda + 1, \mu + T(x))$$

Le tableau 3 présente quelques lois a priori conjuguées pour quelques lois exponentielles usuelles :

$p(x/\theta)$	$\pi(\theta)$	$\pi(\theta/x)$
Normale $\mathcal{N}(\theta, \sigma^2)$	Normale $\mathcal{N}(\mu, r^2)$	$N(\rho(\sigma^2\mu + r^2x), \rho\sigma^2r^2)$ $\rho = \frac{1}{(\sigma^2+r^2)}$
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 $\mathcal{B}(n, \theta)$	Beta $\mathcal{B}e(\alpha, \beta)$	$\mathcal{B}e(\alpha + x, \beta + n - x)$
Normale $\mathcal{N}(\mu, \frac{1}{\theta})$	Gamma $\mathcal{G}(\alpha, \beta)$	$\mathcal{G}(\alpha + \frac{1}{2}, \beta + \frac{(\mu-x)^2}{2})$

Table 3 : Lois conjuguées naturelles pour quelques lois exponentielles usuelles.

Exemple 2.1.5. Loi de poisson

Pour $X \sim p(\lambda)$

$$p(x/\lambda) = \exp(-\lambda) = \frac{\lambda^x}{x!} = \frac{\exp(x \log \lambda - \lambda)}{x!}$$

Le paramètre canonique dans ce cas $\theta = \log \lambda$

B) Lois à priori non informative

La section précédente a montrée que les lois conjuguées peuvent être utilisées comme approximations des véritables lois a priori par contre lorsqu'aucune information n'est disponible sur le modèle, leur utilisation n'est justifiée que par des considérations analytiques. Dans de telles situations, il est impossible de bâtir une distribution a priori sur des considérations subjectives. On peut alors chercher à utiliser malgré tout des techniques bayésiennes qui intègrent notre ignorance sur les paramètres du modèle, de telles méthodes sont appelées de manière évidente, non informative.

Il existe plusieurs techniques basées sur les distributions invariantes [15]. On citera la plus utilisée qui est la loi a priori de Jeffreys.

Loi a priori de Jeffreys

Jeffreys (1946, 1961) propose une approche intrinsèque qui évite effectivement le besoin de prendre en compte une structure d'invariance potentielle, tout en étant souvent compatible lorsque cette structure existe [15]. Les lois a priori non informatives de Jeffreys sont fondées sur l'information de **Fisher**, donnée par :

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

dans le cas unidimensionnel. Sous certaines conditions de régularité, cette information est aussi égale à

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

La loi a priori de Jeffreys est

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

définie à un coefficient de normalisation près quand π^* est propre.

Exemple 2.1.6. Si $x \sim \mathcal{B}(n, p)$,

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

$$\frac{\partial^2 \log f(x/p)}{\partial p^2} = \frac{x}{p^2} + \frac{n-x}{(1-p)^2}$$

donc

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

Donc la loi de Jeffreys pour ce modèle est

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

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

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

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

et la loi non informative de Jeffreys est alors définie par :

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

Exemple 2.1.7. 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

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

et la densité de la loi de Jeffreys est

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

2.2 La densité prédictive

Dans le pratique, nous somme 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 [1]. 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 2.2.1. *La fonction de densité prédictive pour une expérience futur F de fonction de densité*

$$\{f(y/\theta) : \theta \in \Theta\} \text{ sur } \mathcal{Y}$$

Une expérience informative E de classe de fonctions de densité

$$\{f(x/\theta) : \theta \in \Theta\} \text{ sur } \mathcal{X}$$

Et une fonction de densité $\pi(\theta)$ sur Θ , la fonction de densité prédictive $p(y/x)$ pour F est définie par la formule (1) suivant :

$$\begin{aligned} f(y/x) &= \int f(y/\theta, x)\pi(\theta/x)d\theta \\ &= \frac{\int f(y/\theta)\pi(\theta)f(x/\theta)d\theta}{\int \pi(\theta)f(x/\theta)d\theta} \end{aligned}$$

Remarque 2.2.1. *Figure (1) Montre les étapes essentielles conduisant à la densité prédictive.*

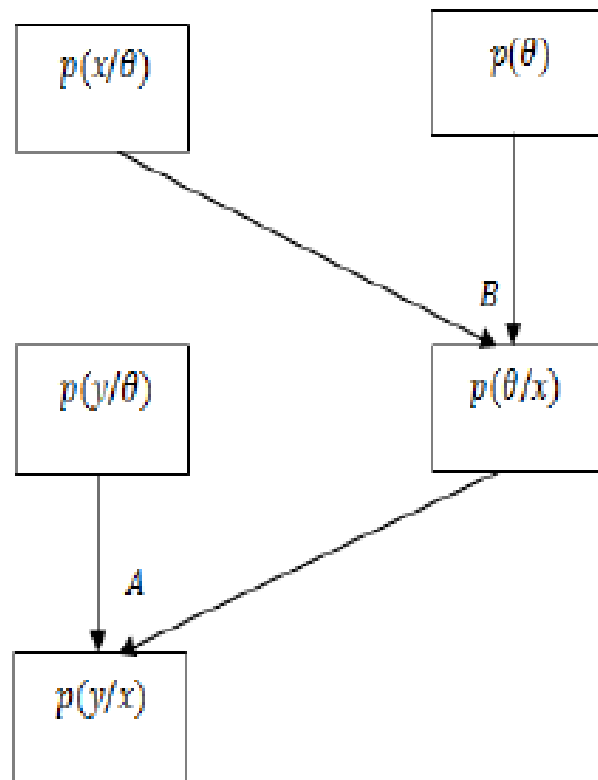


Figure (1) : Les étapes de base conduisant à la densité prédictive.

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

Exemple 2.2.1. Une étude publiée par **Freireich et Al.**, à été conçue pour évaluer les effectives de chimio thérapeutique **6-mercaptopurine** (6 - MP) pour le traitement de la leucémie aigue. Les patients ont été randomisés pour thérapie par paires. Soit θ la proportion de la population de couples dans lesquels la 6 - MP séjour des patients en rémission plus longue que le patient du groupe placebo, (Pour distinguer une probabilité θ à partir d'une distribution de probabilité sur θ , nous appellerons cela une proportion de la population ou une propension). L'hypothèse nulle est $H_0 : \theta = \frac{1}{2}$ i.e. pas d'efficacité de 6 - MP; et $H_1 : \theta \neq \frac{1}{2}$. Il y avait 21 paires de patients dans l'étude, et 18 d'entre eux favorisaient 6 - MP. Supposons que la distribution a priori est Uniforme sur $(0, 1)$: La distribution Uniforme de $(0, 1)$ est également la version de la distribution $\mathcal{Be}(1,1)$. La mise à jour de la distribution $\mathcal{Be}(a,b)$ après les y succès et $n-y$ échecs est facile, car la version Beta est une probabilité a priori conjuguée, ce qui signifie que la distribution a posteriori apparaît comme un membre de la même famille de répartition que la précédente.

Pour voir cela dans le cas des Beta - Binomiales de la règle de Bayes nous avons :

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

Nous reconnaissons cette forme comme étant proportionnelle à la densité d'une distribution $\mathcal{B}e(x+a, n-x+b)$. Ainsi, dans notre cas ($y = 18, n-y = 3$), la distribution a posteriori est $\mathcal{B}e(19, 4)$, comme c'est tracée dans la Figure (2) :

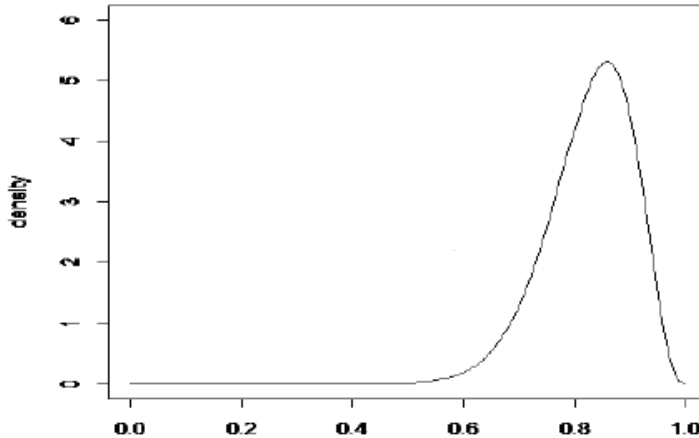


Figure (2) : La densité a posteriori de θ est $\mathcal{B}e(19, 4)$

Remarquons que cette distribution est unimodale avec une moyenne de $\frac{19}{(19+4)} = 0.8261$. Dans le contexte prédictif, on suppose qu'il est possible de prendre 5 observations supplémentaires. Comment s'en servir ? Une façon de répondre est d'évaluer la conséquence d'obtenir k succès au cours des 5 prochaines paires de patients (pour un total de 26 paires) pour $k = 1, \dots, 5$, puis de peser ces conséquences par les probabilités de prévisions des valeurs possibles de k . Pour calculer les probabilités de futures observations, on trouve premièrement ces probabilités en supposant que les paramètres sont connus, et puis on trouve la distribution a posteriori des paramètres. Nous avons ensuite la moyenne des distributions conditionnelles à l'égard de la distribution a posteriori des paramètres. Cela donne la distribution inconditionnelle prédictive d'intérêt.

Pour illustrer, premièrement, envisager une seule paire de patients, la paire de 22 patients, ayant un résultat binaire variable x_{22} . Supposons que cette paire est échangeable avec les 21 premières paires, $x = (x_1, x_2, \dots, x_{21})'$, pour laquelle $y = \sum_{i=1}^{i=21} x_i = 18$. Un membre de la nouvelle paire est affecté à la 6-MP, et l'autre est affecté au placebo. La probabilité prédictive que le patient 6-MP restera dans la remise est plus que la moyenne de la distribution a posteriori de θ . Pour la distribution $\mathcal{B}e(19, 4)$, considérée séparément, la probabilité prédictive de succès est donnée par la règle du placebo de la succession : $\frac{19}{(19+4)} = 0.8261$. D'où la distribution prédictive de x_{22} est :

$$g(x_{22}/x) = \int f(x_{22}/\theta)\pi(\theta/x)d\theta = \frac{\int \theta^{x_{22}}(1-\theta)^{1-x_{22}}}{\Gamma(19)\Gamma(4)}\theta^{18}(1-\theta)^3d\theta$$

Donc la chance que la 22^{ème} observation soit un succès ($x_{22} = 1$) est :

$$\begin{aligned} f(x_{22} = 1/x) &= \int f(x_{22}/\theta)\pi(\theta/x)d\theta \\ &= \frac{\Gamma(23)}{\Gamma(19)\Gamma(4)} \frac{\Gamma(20)\Gamma(4)}{\Gamma(20)} \int \frac{\Gamma(24)}{\Gamma(20)\Gamma(4)}\theta^{19}(1-\theta)^3d\theta \\ &= \frac{\Gamma(23)\Gamma(20)}{\Gamma(19)\Gamma(24)} \\ &= \frac{19}{23} = 0.8261 \end{aligned}$$

La 3^{ème} égalité a résulté puisque l'intégrale fixée égale à 1, et la 4^{ème} a résulté puisque $\Gamma(z) = (z-1)\Gamma(z-1)$ pour tout z .

Maintenant supposons que deux paires additionnelles de patients (les deux sont échangeables par les 21 premières paires, $x = (x_1, x_2, \dots, x_{21})'$) sont traitées, un membre de chaque paire par 6-MP et l'autre par placebo. La probabilité prédictive des deux paires soient succès pour 6-MP n'est pas le carré de 0.8261, mais plutôt la probabilité que la 1^{ère} paire soit un succès multipliée par la probabilité que la 2^{ème} paire soit un succès sachant le résultat de la 1^{ère}. En utilisant la formule (1) deux fois, nous avons :

$$\begin{aligned} f(x_{22} = 1 \wedge x_{23} = 1/x) &= f(x_{22} = 1/x)f(x_{23} = 1/x_{22} = 1, x) \\ &= \frac{19}{23} \frac{20}{24} = 0.6884 \end{aligned}$$

Pour 5 paires additionnelles, la probabilité prédictive pour 4 succès pour 6-MP est :

$$5 \left(\frac{22 \times 21 \times 20 \times 19 \times 4}{27 \times 26 \times 25 \times 24 \times 23} \right) = 0.3624$$

Le prépondérant «5» est survenu puisque il y a 5 ordres possibles auxquels les 4 succès peuvent être arrivés (SSSSE, SSSES, ..., ESSSS), dont tous les événements ont la même probabilité globale.

Remarque 2.2.2. *Le Tableau (6) dans l'annex I donne les distributions prédictives pour les modèles.*

2.2.1 Puissance et nombre de sujets nécessaire

Grâce à l'importance de nombre de sujets nécessaire (la taille de l'échantillon) et sa relation avec la puissance dans les essais cliniques.

Nous proposons d'évaluer cette relation par l'approche bayésienne, en incorporant les informations a priori.

Pour ce faire nous avons besoin de donner tout d'abord la notion de taille d'effet des échantillons normales et indépendants.

Définition 2.2.2. *Pour une population d'étude de moyenne μ et d'écart type σ , la taille effet d'écart type ajusté d'un traitement à tester est défini par :*

$$\varepsilon = \frac{\mu_e - \mu_s}{\sigma}$$

où μ_e et μ_s sont les moyennes des populations du traitement à tester et du témoin (ou de l'agent de contrôle), respectivement.

Exemple 2.2.2. *Considérons le design d'un groupe à deux bras parallèle comparant un traitement d'essai et une thérapie standard ou un agent de contrôle actif, sous l'essai de deux bras [5]. La puissance est une fonction de taille d'effet ε telle que :*

$$\text{puissance}(\varepsilon) = \Phi\left(\frac{\sqrt{n}\varepsilon}{2} - z_{1-\alpha}\right)$$

où Φ est la fonction de répartition de la distribution normale standard et $z_{1-\alpha}$ est le quantile de la loi normale centrée et réduite d'ordre $(1 - \alpha)\%$. On suppose que la distribution a priori de ε est $\pi(\varepsilon)$, donc la puissance espérée est donnée par :

$$P_{esp} = \int \Phi\left(\frac{\sqrt{n}\varepsilon}{2} - z_{1-\alpha}\right) \pi(\varepsilon) d\varepsilon \quad (2.2)$$

En pratique, l'intégration numérique est généralement utilisée pour l'évaluation de l'équation (2.2). Pour illustrer l'équation (2.2) on suppose qu'on a un test unilatéral de niveau $\alpha = 0.025$ (i.e, $z_{1-\alpha} = 1.96$), et l'a priori de ε est :

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

Par convention, on utilise la moyenne (médiane) de la taille d'effet $\bar{\varepsilon} = 0.25$ pour la taille de l'échantillon, on suppose que $\bar{\varepsilon} = 0.25$ est la vraie valeur de la taille d'effet.

Pour le design à deux bras balancé avec $\beta = 0.2$ ou la puissance=80% l'approche classique donne la taille de l'échantillon suivante :

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

D'autre part, l'approche bayésienne basée sur la puissance espérée de l'équation (2.1) est donnée par :

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

Donc la puissance espérée (66%) est inférieure à la puissance désirable de (80%). Pour obtenir la même puissance, il est nécessaire d'augmenter la taille de l'échantillon.

Si $\pi(\varepsilon)$ suit une distribution normale comme $\mathcal{N}(\mu, \frac{\sigma^2}{n_0})$, Alors la puissance attendue peut être obtenue en utilisant la distribution prédictive en évaluant la probabilité que l'événement critique se produise, c'est-à-dire, Si $\pi(\varepsilon) \sim \mathcal{N}(\mu, \frac{\sigma^2}{n_0})$. et d'après l'exemple (2.1.4).

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

Avec la taille de l'échantillon totale est :

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

Et la taille de l'échantillon espérée est :

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

Si $\pi(\varepsilon) \sim \frac{1}{b-a}$ où $a \leq \varepsilon \leq b$ On a le nombre de sujets nécessaires espéré :

$$\begin{aligned}
 n_{esp} &= \int_a^b \frac{4(z_{1-\alpha} + z_{1-\beta})^2}{\varepsilon^2} \frac{1}{b-a} d\varepsilon = \frac{1}{b-a} 4(z_{1-\alpha} + z_{1-\beta})^2 \left(\frac{-1}{b} - \left(\frac{-1}{a}\right)\right) \\
 &= \frac{1}{b-a} 4(z_{1-\alpha} + z_{1-\beta})^2 \left(\frac{b-a}{ab}\right) \\
 &= \frac{4}{ab} (z_{1-\alpha} + z_{1-\beta})^2
 \end{aligned}$$

Cela donne le rapport de taille d'échantillon :

$$R_n = \frac{n_{esp}}{n} = \frac{\frac{4}{ab}(z_{1-\alpha} + z_{1-\beta})^2}{\frac{4(z_{1-\alpha} + z_{1-\beta})^2}{\varepsilon^2}} = \frac{\varepsilon^2}{ab}$$

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

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

Cela indique que l'approche fréquentiste pourrait sous-estimer considérablement la taille de l'échantillon nécessaire pour atteindre la puissance désirée.

2.3 Règles d'arrêt bayésiennes pour les essais de phase II

Les études de phase II fournissent des étapes intermédiaires importantes pour un développement de médicament réussi. Dans l'ère post-génomique et à haut débit d'aujourd'hui, le nombre de nouveaux agents candidats augmente à pas de géant. Puisque les études en phase tardive sont longues et coûteuses, les essais en phase intermédiaire jouent un rôle essentiel dans l'élimination du «chaff» de notre collection de médicaments, de sorte que seuls les traitements les plus prometteurs sont acheminés vers le développement en phase tardive.

2.3.1 Conditions de FDA sur les essais bayésiens

Il y a un grand besoin de réflexion créative dans la conception des essais de dispositifs médicaux pour améliorer l'efficacité tout en maintenant l'intégrité scientifique des résultats des essais cliniques. [17] Les conceptions bayésiennes d'essais cliniques offrent une flexibilité supplémentaire en ce qui concerne les caractéristiques de conception, mais la validité des résultats dans le cadre réglementaire demeure préoccupante. Les indications de la **FDA**(Food and Drug Administration's) sur les essais bayésiens (février 2010) [25] indiquent que la démonstration des caractéristiques d'essai bayésiennes dans une perspective fréquentiste (c'est-à-dire les taux d'erreur attendus de type I, II et puissance) est un élément essentiel de l'évaluation bayésienne.

Par exemple dans les essais sur les dispositifs médicaux, une seule analyse intermédiaire de l'information tardive peut réduire le délai avant le dépôt réglementaire. De plus, les dispositifs de candidats contrôle sont souvent le sujet de leur propre étude réglementaire récente, sinon d'études multiples. Ignorer complètement cette information au moment de l'analyse est de plus en plus difficile à justifier. Les conceptions bayésienne a posteriori et prédictive, sont proposées pour résoudre ces problèmes. Les objectifs de conception incluent la rétention de certains randomisation pour adresser le biais caché tout en permettant une réduction substantielle de la taille de l'échantillon de contrôle grâce à l'utilisation d'une a priori bayésienne informative correctement calibré [11] [10].

2.3.2 Méthode Bayésienne comme un outil des fréquentiste

Dans le contexte de la conception et l'analyse des études enregistrées, l'approche bayésienne peut être un outil pour construire de bonnes conceptions fréquentistes. Par exemple, nous pouvons utiliser le paradigme Bayésien pour construire un essai clinique qui nécessite une plus petite taille d'échantillon espérée indépendamment des valeurs réelles des paramètres. La conception peut être compliquée, mais nous pouvons toujours trouver ses caractéristiques de fonctionnement fréquentiste en utilisant la simulation. En particulier, nous pouvons nous assurer que le taux de faux-positifs se situe dans la plage acceptable pour les organismes de réglementation [4].

Le paradigme bayésien permet d'utiliser l'information historique et les résultats des autres essais, qu'ils impliquent le même médicament, des médicaments similaires, ou peut-être le même médicament mais avec des populations de patients différentes. L'approche bayésienne est idéale pour emprunter la force à travers les groupes de patients et de maladies dans le même essai et à travers les essais. Cependant, nous mettons en garde que les informations historiques ne peuvent généralement pas être simplement considérées comme échangeables avec les informations actuelles.

Exemple 2.3.1. *Certains essais proposés par des sociétés pharmaceutiques et d'appareils sont déficients et peuvent être améliorés en adoptant une approche bayésienne. Par exemple, une entreprise peut considérer que son médicament est le plus approprié pour une maladie particulière, mais ne sait pas exactement quels sous-types de la maladie seront les plus sensibles. Ils proposent donc des essais séparés pour les différents sous-types. Pour être précis, considérons le cancer de l'ovaire avancé, une maladie particulièrement difficile pour obtenir des réponses tumorales. En explorant les effets possibles de sa drogue, supposons qu'une compagnie essaye de détecter un taux de réponse de tumeur de 10%. Il a proposé de traiter 30 patients dans un groupe et 30 patients dans le groupe complémentaire, mais de mener deux essais distincts. Tous les 60 patients seraient accumulés dans le but d'obtenir au moins une réponse tumorale. Supposons qu'il n'y ait eu aucune réponse sur les 30 patients enregistrés à l'essai 1 et aucune réponse sur 25 patients à ce jour dans l'essai 2. Leur conception, ils ajouteraient encore 5 patients de plus à l'essai 2. Mais ce serait une folie. Jusqu'à présent, nous aurions appris deux choses : d'abord, le médicament n'est pas très actif, et deuxièmement, les deux sous-groupes de patients répondent de la même façon. Une analyse de modélisation hiérarchique bayésienne permettrait de le faire,*

et une telle analyse raisonnable montrerait qu'il est fort futile (et éthiquement discutable) d'ajouter les 5 patients restants dans l'essai 2.

Remarque 2.3.1. *Les conceptions bayésiennes intègrent les méthodes séquentiel lorsque cela est possible sur le plan logistique, utilisent les probabilités prédictives des résultats futurs et la robustesse via les études des sous-groupes de patients. Ces caractéristiques bayésiennes ont des implications à la fois pour l'analyse et pour la conception. Tous les impliquent la modélisation dans la construction des fonctions de vraisemblance.*

2.3.3 Règle d'arrêt binaire pour futilité et efficacité

A) Règle d'arrêt a posteriori

D'après (Thall et Simon) [20] (Simon R. et Thall) [21], Simon et Estey (1995) présentent une classe de plans bayésien pour les essais cliniques de phase II qui incluent des règles d'arrêt basées sur des bornes de décision pour des événements cliniquement significatifs [4].

Pour illustrer, soit $x_i \in \{0, 1\}$ un indicateur de réponse pour le i^{me} patient. Soit p_e et p_s la probabilité de réponse sous la thérapie expérimentale (e) et la norme de soins (s), respectivement. Les études de phase IIA ne comprennent pas la randomisation pour contrôler.

Dans de tels cas, nous supposons que p_s est connu, suit une distribution a priori informatif $\pi(p_s)$.

Définition 2.3.1. *Soit $x = (x_1, \dots, x_n)$ toutes les données jusqu'au patient n . Nous pouvons évaluer de manière significative les probabilités a posteriori qui sont de la forme*

$$\pi_n = P(p_e > p_s + \delta/x).$$

La probabilité π_n est la probabilité a posteriori que la probabilité de réponse sous la thérapie expérimentale domine celle sous la norme de soins d'au moins δ . Le résultat est déterminé par l'investigateur et devrait refléter l'amélioration minimale cliniquement significative. Cela dépend aussi de la nature de la réponse, de la maladie et support de p_s . La probabilité π_n est actualisée après chaque patient (ou cohorte de patients), et est ensuite utilisée pour définir des règles d'arrêt séquentielles comme suit :

Algorithme

$$\text{décision} = \begin{cases} \text{Arrêter et déclarer } E \text{ prometteur} & \text{si } \pi_n > U_n \\ \text{Continuer à inscrire des patients} & \text{si } L_n < \pi_n < U_n \\ \text{Arrêter et déclarer } E \text{ non prometteur} & \text{si } \pi_n < L_n \end{cases}$$

Les limites de décision $\{(L_n, U_n), n = 1, 2, \dots\}$ sont les paramètres de la conception. Par exemple, on pourrait utiliser $L_n = 0.05$ et $U_n = 0.95$ pour tout n .

Les considérations qui entrent dans le choix de ces limites sont similaires à celles du choix des limites d'arrêt pour les conceptions séquentielles de groupes fréquentistes.

Remarque 2.3.2. *Dans la pratique, on commence par un premier choix raisonnable, on évalue les caractéristiques de fonctionnement fréquentistes et on ajuste itérativement les bornes de décision jusqu'à ce que les caractéristiques de fonctionnement désirées soient atteintes.*

Par exemple, nous pourrions commencer avec $L_n = 1\%$ et $U_n = 80\%$. Ensuite, nous calculons les caractéristiques de fonctionnement. Nous pourrions considérer deux scénarios : un scénario nul S_0 avec $p_e = p_s$, et un scénario alternatif S_1 avec $p_e > p_s + \delta$ comme vérité de simulation.

L'erreur de type I est alors la probabilité de terminer l'essai avec la conclusion (fausse) que E est prometteur, alors que la puissance est la probabilité, par rapport au simulation répétée de l'essai possible sous S_1 , que l'essai se termine avec une conclusion correcte que E est promettant.

Supposons que nous trouvons que l'erreur de type I est de 8%, un peu plus grande que souhaité. Nous essayons ensuite d'augmenter la borne inférieure L_n pour réduire l'erreur de type I. Maintenant, nous pourrions avoir une erreur de type I acceptable sous S_0 , mais une puissance de seulement 70% sous S_1 . Pour augmenter la puissance, nous pourrions maintenant essayer de réduire la borne supérieure, disons à $U_n = 75\%$. Une séquence de telles corrections itératives conduira éventuellement à un ensemble de limites qui atteignent des caractéristiques de fonctionnement souhaitables.

B) Règle d'arrêt utilisant la probabilité prédictive

La probabilité prédictive joue un rôle essentiel dans la conception et la surveillance des essais. Par exemple, le conditionnement de ce qui est connu à propos des covariables et les résultats des patients à tout moment pendant un essai, permettent de déterminer

la probabilité d'obtenir une signification statistique à la fin d'essai [4].

Dans cette partie, nous allons donner une nouvelle méthode pour calculer les probabilités prédictives, 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 rejeté d'hypothèse nulle) si l'essai se déroule à la taille maximale de l'échantillon prévue [13].

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

Notre objectif est d'évaluer la proportion p pour un nouveau médicament en testant les deux l'hypothèses :

$$H_0 : p \leq p_0$$

$$H_1 : p \geq p_1$$

Où p_0 représente un taux de réponse pré spécifié du traitement standard et p_1 représente un taux de réponse cible d'un nouveau traitement. dans la phase II d'un essai.

En donnant p_0 , p_1 , la distribution a priori du taux de réponse $\pi(p)$ 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 de l'erreur du 1ère et du 2ème espèce simultanément.

La règle de décision peut être construite comme suit :

Algorithme(*phase II basée sur la Conception de PP*).

$$\text{décision} = \begin{cases} \text{Arrêter et déclarer } E \text{ prometteur} & \text{si } PP > \theta_U \\ \text{Continuer à inscrire des patients} & \text{si } \theta_L < PP < \theta_U \\ \text{Arrêter et déclarer } E \text{ non prometteur} & \text{si } PP < \theta_L \end{cases}$$

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 p_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.

Remarque 2.3.3. *En choisissant $\theta_L > 0$ et $\theta_U < 1$, l'essai peut se terminer prématurément en raison soit de futilité soit d'efficacité. Pour l'essai 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 pour l'efficacité [4].*

Calcul de la probabilité prédictive

Supposons que la distribution a priori de p est, $\pi(p)$, la distribution *Beta* (a, b). La quantité $\frac{a}{a+b}$ indique la moyenne a priori, les quantités a et b peuvent être considérées comme le nombre de réponses a priori favori ou pas, respectivement, $a + b$ peut être considérée comme une mesure de la précision a priori.

Supposons que nous avons fixé un nombre maximum de patients N_{max} , et supposons que X qui est le nombre de réponses chez les patients parmi les n patients en cours ($n \leq N_{max}$) suite une distribution $B(n, p)$, alors la distribution a posteriori de p serai :

$$p/x \sim \mathcal{B}e(a + x, b + n - x).$$

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 conséquence.

Soit Y le nombre de réponses dans le potentiel de futur patients $m = N_{max} - n$. Supposons que notre design est de déclarer l'efficacité, si la probabilité a posteriori de p qui est supérieure à un certain niveau prédéfini p_0 est supérieure à un seuil θ_T . Marginaliser p de la probabilité *binomiale*, il est bien connu que Y suit une distribution de *Beta-Binomiale* [1],

$$Y/X = x \sim \mathcal{B}e - \mathcal{B}(m, a + x, b + n - x).$$

Quand $Y = i$, la distribution a postérieure de

$$p/(X = x, Y = i) \sim \mathcal{B}e(a + x + i, b + N_{max} - x - i).$$

La probabilité prédictive (PP) du succès de l'essai peut être alors calculé comme suit :

Soit : $B_i = P(p > p_0/x, Y = i)$ et $1_i = 1_{\{B_i > \theta_T\}}$, on a :

$$\begin{aligned}
 PP &= E [1_{\{P(p > p_0/x, Y) > \theta_T\}}/x] \\
 &= \int 1_{\{P(p > p_0/x, Y) > \theta_T\}} dP(Y/x) \\
 &= \sum_{i=0}^m P(Y = i/x) \times 1_{\{P(p > p_0/x, Y) > \theta_T\}} \\
 &= \sum_{i=0}^m P(Y = i/x) \times 1_{\{B_i > \theta_T\}} \\
 &= \sum_{i=0}^m P(Y = i/x) \times 1_i.
 \end{aligned}$$

La quantité B_i est la probabilité que le proportion p est plus grand que p_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 conclusion d'un résultat positif, d'ici à la fin de l'essai, sur la base du informations cumulées dans la phase actuelle.

- La PP élevé signifie que le traitement susceptible 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 de l'efficacité /de futilité ou poursuivi parce que les données actuelles ne sont pas encore concluantes. C'est pour cela est définit en introduisant deux seuils sur PP , (θ_L et θ_U).

Exemple 2.3.2. *Supposons qu'un chercheur envisage de recruter 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 $\mathcal{Be}(0.6, 0.4)$ pour la proportion p et soit

Y le nombre de réponse dans $m = 17$ futures observations, donc $Y/X = 16 \sim \mathcal{B}e - \mathcal{B}(17, 16.6, 7.4)$.

Pour chaque valeur possible de $Y = i$, l'a posteriori conditionnelle de p suit une distribution de Beta, $p/x, Y = i \sim \mathcal{B}e(16.6 + i, 24.4 - i)$.

Dans cet exemple, on prend $\theta_T = 0.90$. Comme on peut le voir au tableau (4), lorsque $0 \leq Y \leq 11$, $0.0059 \leq P(p > 0.60) \leq 0.8314$. Par conséquent nous concluons H_0 pour $Y \leq 11$. D'autre part, lorsque $12 \leq Y \leq 17$; $0.9022 \leq P(p > 0.60) \leq 0.9989$. Dans ce cas, au lieu de favoriser H_1 pour $y \geq 12$ la probabilité prédictive est la moyenne pondérée à 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 1 dans l'annexe).

$Y=i$	$Pr(Y=i/x)$	$B_i = P(p > 0.60/x, Y = i)$	$I_{(B_i > 0.90)}$
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

Table 4 : Calcul de la probabilité prédictive bayésienne pour $p_0 = 0.60, \theta_T = 0.90, N_{max} = 40, x = 16, n = 23$, et une distribution a priori Beta (0.6, 0.4) sur p .

2.4 La probabilité prédictive dans les tests de non infériorité

Guosheng Yin, Nan Chen, et J. Jack Lee (2012) proposent une conception d'essai clinique de phase II randomisée basée sur la randomisation adaptative bayésienne et la surveillance prédictive. La randomisation adaptative assigne plus de patients à un bras de traitement plus efficace, en comparant les probabilités a posteriori d'efficacité entre différents bras. On surveille continuellement l'essai en utilisant la probabilité prédictive. L'essai est terminé prématurément lorsqu'il démontre qu'un traitement est extrêmement supérieur aux autres, ou que tous les traitements sont équivalents [10].

Deux méthodes utilisées ont été proposées pour calculer la probabilité prédictive en considérant l'incertitude de la taille de l'échantillon des données futures. Cette conception contrôle également les erreurs de type I et de type II et offre une approche bayésienne alternative à la conception séquentielle de groupe fréquentiste [2]. La conception proposée est comme suit :

Supposons que nous comparions les K traitements dans un essai de phase II randomisé sur K -bras. Soit p_k le taux de réponse du k^{eme} traitement, et assignons une distribution a priori $p_k \sim \mathcal{Be}(\alpha_k, \beta_k)$, pour $k = 1, \dots, K$.

Si, parmi les n_k sujets traités dans le k^{eme} bras, on observe des réponses x_k , alors

$$X_k \sim \mathcal{B}(n_k, p_k)$$

et la distribution a posteriori de p_k est

$$p_k / (X_k = x_k) \sim \mathcal{Be}(\alpha_k + x_k, \beta_k + n_k - x_k).$$

Si la taille maximal de l'échantillon du k^{eme} bras est N_k , alors le nombre de réponses pour les futurs patients $N_k - n_k$, Y_k , suit une distribution beta-binomiale [1] :

$$Y_k / (X_k = x_k) \sim \mathcal{Be} - \mathcal{B}(N_k - n_k, \alpha_k + x_k, \beta_k + n_k - x_k).$$

Lorsque $Y_k = y_k$, la distribution a posteriori du taux de réponse compte tenu des données actuelles et futures est

$$p_k / (X_k = x_k, Y_k = y_k) \sim \mathcal{Be}(\alpha_k + x_k + y_k, \beta_k + N_k - x_k - y_k).$$

Pour faciliter l'exposition, nous considérons deux traitements pour illustrer la conception, c'est-à-dire $K = 2$. Nous spécifions une différence de traitement cliniquement significative δ et une probabilité de seuil θ_T . Si

$$P(|p_2 - p_1| > \delta / X_1 = x_1, X_2 = x_2, Y_1 = y_1, Y_2 = y_2) \geq \theta_T.$$

nous revendiquons la non-équivalence des deux traitements, c'est-à-dire qu'un traitement est supérieur à l'autre. Cependant, Y_1 et Y_2 sont les données futures, qui n'ont toujours pas été observées au stade actuel de la prise de décision. Nous calculons la moyenne sur le caractère aléatoire du Y_1 et Y_2 en calculant la PP comme suit :

$$\begin{aligned} PP &= E_{Y_1 Y_2} [1_{\{P(|p_2 - p_1| > \delta / X_1 = x_1, X_2 = x_2, Y_1, Y_2) \geq \theta_T\}}] \\ &= \sum_{y_1=0}^{N_1 - n_1} \sum_{y_2=0}^{N_2 - n_2} P(Y_1 = y_1 / X_1 = x_1) P(Y_2 = y_2 / X_2 = x_2) \\ &\quad \times 1_{\{P(|p_2 - p_1| > \delta / X_1 = x_1, X_2 = x_2, Y_1, Y_2) \geq \theta_T\}} \end{aligned} \tag{2.3}$$

où 1 est la fonction indicatrice et PP désigne la probabilité prédictive pour revendiquer la supériorité à la fin d'essai. Nous devons spécifier les probabilités de seuil inférieur (θ_L) et supérieur (θ_U) pour la prise de décision adaptative dans la conduite de l'essai. La règle de décision basée sur la PP est la suivante :

- a) Arrêt pour équivalence : Si $PP < \theta_L$, nous arrêtons l'essai et acceptons l'hypothèse nulle pour revendiquer l'équivalence du traitement.
- b) Arrêt pour supériorité : Si $PP > \theta_U$, on arrête l'essai et rejette l'hypothèse nulle pour réclamer un bras de traitement supérieur.

Nous pouvons maintenir les taux d'erreur fréquentistes ; de type I et de type II en calibrant les paramètres de conception ($N, \delta, \theta_T, \theta_L, \theta_U$), où N est la taille maximale de l'échantillon, $N = N_1 + N_2$.

Une conception d'essai basée sur la PP permet une surveillance continue ; si les deux traitements ont des effets similaires sur l'efficacité, ou si un traitement est largement supérieur à l'autre, l'essai peut être arrêté rapidement lorsque des preuves suffisantes se sont accumulées. Cela se traduirait par une plus petite taille d'échantillon prévue, et donc un essai plus efficace. À la fin de l'essai, nous déclarons qu'un traitement est meilleur que l'autre ou l'équivalence de deux traitements.

2.4.1 Les méthodes pour calculer la probabilité prédictive

On a deux façons différentes de calculer la PP . La première méthode est plus rigoureuse mais plus intensive en calcul, et la seconde applique une approximation, mais elle est relativement rapide.

Méthode 1

Une fois la randomisation adaptative est en vigueur, le nombre total de sujets dans le 1^{er} bras et 2^{eme} bras, N_1 et N_2 , deviennent aléatoires, alors que le nombre de sujets restants dans l'essai, m , est fixe, si l'essai n'est pas autorisé pour l'arrêt précoce. Soit Z le nombre de sujets qui seraient assignés au 2^{eme} bras, alors $Z \sim \mathcal{B}(m, \pi)$ avec

$$\pi = \frac{P(p_2 > p_1 / X_1 = x_1, X_2 = x_2)^\tau}{P(p_2 > p_1 / X_1 = x_1, X_2 = x_2)^\tau + \{1 - P(p_2 > p_1 / X_1 = x_1, X_2 = x_2)\}^\tau}$$

Nous assignons la prochaine cohorte de patients au 2^{eme} bras avec la probabilité π , et au 1^{er} bras avec la probabilité $1 - \pi$. Nous utilisons le paramètre d'accord τ pour contrôler le taux de la randomisation adaptative : si $\tau = 0$, alors $\pi = 0.5$, conduisant à une randomisation égale. Une plus grande valeur de τ conduirait à un déséquilibre plus élevé dans l'allocation des patients entre les deux bras et vice versa.

Et donc :

$$P_Z(z / X_1 = x_1, X_2 = x_2) = \binom{m}{z} \pi^z (1 - \pi)^{m-z} \quad (2.4)$$

Pour obtenir la PP , nous calculons d'abord la moyenne sur Y_1 et Y_2 sur $Z = z$, puis sur Z moyenne selon la distribution binomiale dans l'équation (2.4). D'où

$$\begin{aligned} PP &= \sum_{z=0}^m \sum_{y_1=0}^{m-z} \sum_{y_2=0}^z P_Z(z / X_1 = x_1, X_2 = x_2) \\ &\quad \times P(Y_1 = y_1 / X_1 = x_1, Z = z) P(Y_2 = y_2 / X_2 = x_2, Z = z) \\ &\quad \times 1_{\{P(|p_2 - p_1| > \delta / X_1 = x_1, Y_1 = y_1, X_2 = x_2, Y_2 = y_2) \geq \theta_T\}} \end{aligned}$$

Ce qui peut être très intensif en calcul en raison de la somme additionnelle marginalisée sur Z . Cette méthode énumère toutes les possibilités des tailles d'échantillons futures.

Méthode 2

La première méthode implique trois sommations intégrées et est coûteuse en termes de calcul. La seconde approche consiste à approximer $N_k - n_k$ par le nombre attendu de sujets assignés au k^{eme} bras pour $k = 1, 2$, c'est-à-dire $N_1 - n_1 = m(1 - \pi)$ et $N_2 - n_2 = m\pi$. Ceci est une approximation directe basée sur les données actuellement observées qui n'imposent plus de difficultés de calcul.

Bien que la taille totale de l'échantillon soit fixe, la taille de l'échantillon restant n'est pas fixe si l'essai est autorisé pour une résiliation anticipée. L'arrêt d'un essai est une caractéristique supplémentaire d'un plan d'étude. Les paramètres de conception sont étalonnés selon une procédure séquentielle en deux étapes : Nous choisissons d'abord δ et θ_T sans arrêt anticipé, puis sélectionnons les paramètres d'arrêt précoce θ_U et θ_L . Les paramètres de conception sont étalonnés dans un ordre séquentiel afin d'éviter les effets entrelacés d'un arrêt précoce.

Exemple 2.4.1. (*Comparaison entre les deux méthodes*)

Les cinq paramètres de conception ($N, \delta, \theta_T, \theta_L, \theta_U$) doivent être calibrés sur la base du taux d'erreur de type I souhaité et de la puissance de l'essai.

On effectuée une série d'études de simulation avec différentes valeurs de (θ_L, θ_U) , et on comparé les taux d'erreur de type I et les puissances correspondants. Choisisant $N = 160$, bien que la taille réelle de l'échantillon puisse être beaucoup moindre en raison de l'arrêt prématurée de l'essai. Les 40 premiers patients ($n_1 = n_2 = 20$) ont été randomisés de manière égale dans les deux bras et les patients ont ensuite été randomisés de manière adaptative sur la base des probabilités a posteriori de comparaison entre les taux de réponse des deux traitements après chaque observation, pour contrôler à la fois le taux d'erreur de type I (10% ou moins) et la puissance (au moins 80%). Pour

Un taux d'efficacité pour le traitement standard $p_1 = 0.2$.

Un taux de réponse pour le nouveau traitement $p_2 = 0.4$.

On a simulé 10000 essais cliniques.

En fixant $\delta = 0.05$ et $\theta_T = 0.85$, (θ_L, θ_U) sont calibrées, pour déterminer l'arrêt précoce d'un essai en raison d'équivalence ou de supériorité, respectivement. Bien que la conception permette une surveillance après que chaque résultat soit disponible, du point de vue informatique et pratique, la taille de cohorte est choisit de 10.

La table 5 donne toutes les valeurs de l'erreur du type I et la puissance selon les deux méthodes de calcul.

θ_U	Results for the following null cases θ_L :					Results for the following alternative cases θ_L :				
	0.00	0.05	0.10	0.15	0.20	0.00	0.05	0.10	0.15	0.20
<i>Method 1</i>										
0.95	0.126	0.116	0.112	0.106	0.095	0.839	0.823	0.792	0.783	0.747
0.96	0.120	0.120	0.105	0.107	0.096	0.828	0.815	0.799	0.772	0.739
0.97	0.115	0.106	0.097	0.092	0.092	0.827	0.813	0.791	0.765	0.734
0.98	0.107	0.105	0.092	0.091	0.082	0.827	0.814	0.791	0.764	0.734
0.99	0.105	0.099	0.095	0.080	0.079	0.820	0.803	0.796	0.766	0.723
1.00	0.101	0.097	0.088	0.083	0.074	0.822	0.806	0.789	0.755	0.733
<i>Method 2</i>										
0.95	0.125	0.116	0.113	0.104	0.099	0.840	0.822	0.793	0.782	0.747
0.96	0.121	0.115	0.106	0.102	0.088	0.829	0.816	0.799	0.772	0.734
0.97	0.112	0.110	0.099	0.096	0.082	0.829	0.813	0.789	0.765	0.735
0.98	0.112	0.101	0.096	0.092	0.085	0.826	0.815	0.796	0.765	0.731
0.99	0.102	0.096	0.087	0.076	0.071	0.819	0.802	0.796	0.767	0.722
1.00	0.101	0.093	0.088	0.080	0.070	0.820	0.806	0.790	0.760	0.731

Tableau 5 : Taux d'erreur de type I et valeurs de puissance en faisant varier les paramètres de conception θ_L et θ_U en utilisant la méthode 1 et la méthode 2 (fixation $\delta = 0,05$ et $\theta_T = 0,85$)

Dans le tableau 5, nous pouvons voir que les taux d'erreur de type I et les puissances obtenues à partir des méthodes 1 et 2 sont très proches, ce qui implique que l'utilisation du nombre attendu de futurs sujets dans la méthode 2 donne une très bonne approximation des résultats futurs tailles. Notre objectif est toujours de maintenir un taux d'erreur de type I au plus 10% et d'atteindre une puissance au moins 80% lorsque l'essai est autorisé à se terminer tôt. Il y a plusieurs paires de (θ_L, θ_U) qui satisfont nos exigences de conception, comme indiqué par les valeurs dans les zones teintées du tableau 5, à partir des quelles nous pouvons sélectionner $\theta_L = 0.05$ et $\theta_U = 0.99$.

Dans la conception d'essai clinique de phase II basée sur la randomisation adaptative bayésienne et la surveillance prédictive, et d'après cette section on peut voir que la probabilité prédictive joue un rôle très important pour l'arrêt précoce d'essai, donc nous utilisons PP pour contrôler les erreurs de type I et de type II, ce qui nécessite un calcul approfondi au stade de la planification, et aussi pour surveiller le test et prendre la décision d'arrêter l'essai comme suit :

- Si un traitement est supérieur à l'autre, arrêtez l'essai et déclarez la supériorité.
- Si les deux traitements ont une efficacité similaire, arrêtez l'essai et déclarer l'équivalence.
- Sinon, poursuivez l'essai jusqu'à ce que la taille maximale de l'échantillon ait été atteinte.

Chapitre 3

Application

Dans notre travail, on discute une classe de modèles d'essais cliniques randomisés pour la démonstration de la non-infériorité clinique entre un dispositif expérimental (E) et un dispositif de contrôle (C) basé sur un critère de succès clinique composite (SCC) évalué post opératoire. Pour plus de simplicité, les exemples dans ce chapitre supposent que l'état clinique sera évalué au 24e mois après l'intervention. La conception de l'essai comprend une seule analyse intermédiaire (AI) de l'information tardive (section 2.2.1); l'objectif est de réduire considérablement le délai de dépôt réglementaire. De plus, les dispositifs de candidats contrôle font souvent l'objet de leur propre étude réglementaire récente, voire d'études multiples. Ignorer complètement cette information au moment de l'analyse, est difficile à justifier lorsque les ressources ne sont pas illimitées. Pour accommoder l'utilisation de cette information, la conception de probabilité a posteriori bayésienne est étendue pour incorporer une a priori informatif sur le dispositif de contrôle seulement. Les objectifs de conception comprennent la rétention d'une certaine randomisation pour adresser le biais caché tout en permettant une réduction substantielle de la taille de l'échantillon de contrôle grâce à l'utilisation d'une a priori informatif bayésienne correctement calibrée. Les caractéristiques de fonctionnement fréquentistes des conceptions avec et sans utilisation d'une information informative de contrôle (c-informative) pour plusieurs exemples de conceptions sont explorées et comparées en fonction de la taille de l'échantillon et de la manière dont le contrôle est supposé être informatif.

3.1 Considérations de modélisation

3.1.1 Justification de l'analyse intermédiaire et tardif

La raison d'être d'une analyse intermédiaire (*AI*) est de réduire le fardeau des répondants associé au délai de soumission qui résulte du fait d'attendre que tous les patients évaluable sur le plan de l'efficacité deviennent théoriquement responsables du suivi du 24ème mois. Les paramètres de conception peuvent être ajustés de sorte que le plan expérimental ait une probabilité relativement élevée de permettre une soumission précoce (c'est-à-dire aboutir à une conclusion de non-infériorité clinique à l'*AI*), mais contrôle l'erreur de type I (α). Il existe un autre problème important qui motive l'utilisation d'une *AI* à temps d'information tardif : Maintenir en double aveugle dans les essais cliniques randomisés des dispositifs médicaux expérimentaux n'est souvent pas faisable et, si possible, n'est pas pratique pour des raisons techniques. Idéalement, une *AI* est réalisée et les résultats sont interprétés par un « tiers indépendant » afin de ne pas influencer le comportement ultérieur des patients ou des investigateurs. Afin d'éliminer la possibilité que la dissémination des résultats de l'*AI* influe sur les tendances d'inscription futures, l'*AI* ne devrait pas être effectuée avant que les tailles d'échantillon maximales ciblées dans chaque groupe d'appareils aient été randomisées. De cette manière, il n'est pas possible que la publication des résultats de l'analyse d'impact influence la future inscription à l'étude, évitant ainsi le biais de sélection. Par exemple, si le dernier patient est randomisé 6 mois avant l'*AI* planifiée, alors plus de 18 mois devront s'écouler avant que l'opportunité ne se présente pour l'analyse finale.

3.1.2 Utilisation d'une a priori informative pour le contrôle

Il est naturel d'utiliser une distribution a priori non informative sur le *CCS* du mois 24 pour *E*. En revanche, les dispositifs candidats *C* font souvent l'objet de leur propre étude réglementaire récente, si ce n'est de multiples études. De plus en plus, il peut y avoir des candidats *C* pour lesquels des données récapitulatives suffisantes sont disponibles pour guider la construction d'une a priori raisonnable. Ignorer complètement cette information au moment de l'analyse est de plus en plus difficile à justifier lorsque les ressources ne sont pas illimitées. La conception bayésienne des probabilités a posteriori avec un *AI*-information-temps unique est proposée pour résoudre ce problème. Les objectifs de conception incluent la rétention de certains randomisation pour adresser le biais caché tout

en permettant une réduction substantielle de la taille de l'échantillon de contrôle grâce à l'utilisation d'une a priori bayésienne informative correctement calibrée. De cette manière, l'approche bayésienne améliore l'efficacité en n'ignorant pas les preuves scientifiques valables concernant la performance du dispositif de contrôle lors des comparaisons avec un nouveau dispositif expérimental. L'approche bayésienne comprend une façon formelle d'incorporer ces informations dans les comparaisons de groupes de traitement et d'influencer les résultats statistiques concernant la non-infériorité clinique. Ceci est en opposition avec l'approche fréquentiste qui permet seulement l'utilisation de telles informations à la phase de conception de l'étude mais pas à la phase d'analyse. Par conséquent, il est essentiel que le commanditaire justifie suffisamment la distribution a priori particulière à utiliser et qu'il parvienne à un consensus avec les conditions de **FDA**.

3.1.3 Aperçu de la procédure d'examen de non-infériorité

Les méthodes utilisées reposent sur le modèle statistique relativement familier et direct basé sur les distributions bêta-binomiales.

$$\text{Soit } p_e \sim \mathcal{Be}(\frac{1}{2}, \frac{1}{2}), \text{ et } p_c \sim \mathcal{Be}(a, b) \text{ avec } X \sim \mathcal{B}(N, p)$$

Les distribution a posteriori

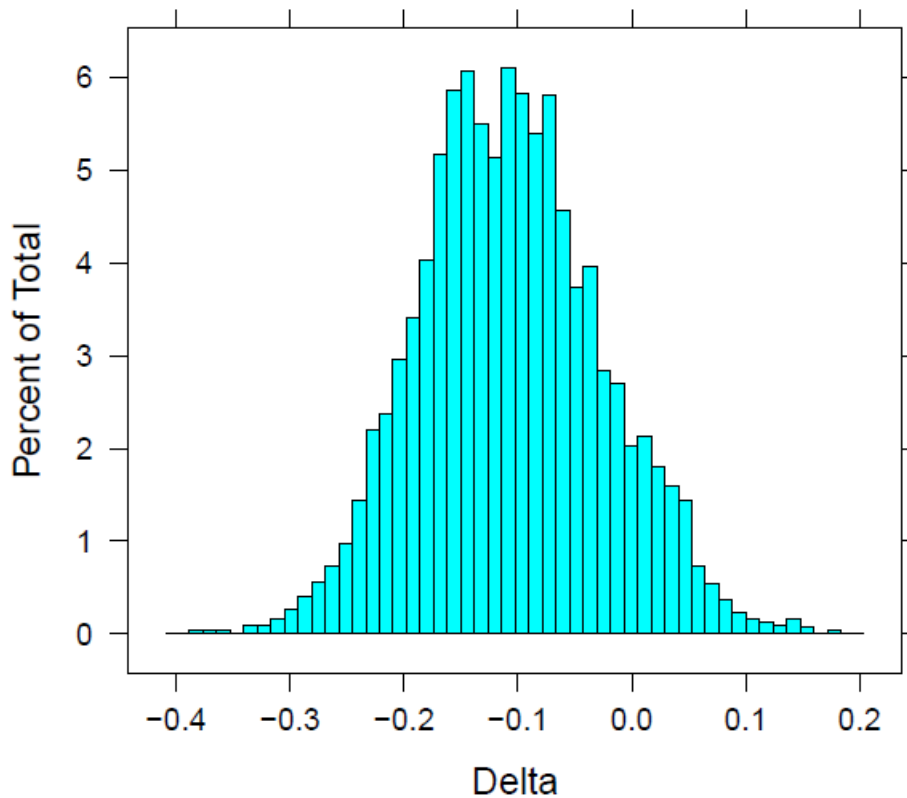
$$p_e/(X_e, N_e) \sim \mathcal{Be}(\frac{1}{2} + X_e, \frac{1}{2} + (N_e - X_e))$$

et

$$p_c/(X_c, N_c) \sim \mathcal{Be}(a + X_c, b + (N_c - X_c))$$

telle que (X_e, N_e) et (X_c, N_c) sont les données observées.

Un programme dans R pour calculer cette probabilité a posteriori est donné dans l'annexe (programme 2 et 3).



Figure(1.3) : La distribution a posteriori pour la différence d'appareil dans. 24 SCC

La distribution a posteriori de la différence de groupe peut être déterminée à partir de la différence de ces distributions bêta mises à jour de manière appropriée. Le support de cette distribution est $[-1, +1]$, la figure (1.3) est simulée par le programme(2 et 3) de l'annexe qui est une distribution gaussienne. L'intégrale de -0.10 à 1 est la probabilité a posteriori bayésienne que le taux de SCC au 24 mois pour les patients implantés avec E ne soit pas inférieur à -0.10 à celui de C .

3.2 Mise en œuvre du design

3.2.1 Modèle statistique

L'approche adoptée ici a été formulée en termes de détermination de la probabilité a posteriori de non-infériorité au moment de l'AI et de nouveau à l'analyse finale s'il n'y avait pas suffisamment de preuves favorisant la non-infériorité au moment de l'AI. Les étapes suivantes supposent des a priori non informatifs pour les deux périphériques.

Une section suivante examine l'impact d'une a priori informative uniquement pour le dispositif de contrôle. La description de l'algorithme suivant suppose que le critère de rejet de l'hypothèse d'infériorité à l'*AI* est une probabilité a posteriori de non-infériorité au moins égale à 0.95. Si la probabilité postérieure est inférieure à 0.95, elle sera recalculée après l'ajout des données des patients évaluables d'efficacité restants après qu'ils soient tous théoriquement dus pour le suivi du 24^e mois. Les seuils a posteriori bayésiens réels utilisés à l'analyse *AI* et finale doivent être considérés comme des paramètres de réglage qui permettent aux conceptions de varier en termes d'erreur de type I et de type II.

1. Supposons des distributions a priori non informatives (Jeffreys) pour les probabilités de succès des groupes expérimentaux (p_e) et des dispositifs de contrôle (p_c).
2. En utilisant les résultats d'une analyse intermédiaire, déterminer la distribution bivariée a posteriori bayésienne pour la différence ($p_e - p_c$). Individuellement, p_e et p_c ont des distributions bêta. La distribution bivariée nécessaire peut être obtenue de manière empirique par des simulations dans lesquelles une valeur de p_e est obtenue à partir de sa distribution (spécifique au groupe de dispositifs) et une valeur de p_c est obtenue à partir de sa distribution. La différence ($p_e - p_c$) peut alors être calculée. Si ce processus est répété un grand nombre de fois, la distribution de fréquence résultante de l'observation ($p_e - p_c$) est une estimation empirique de la distribution de probabilité nécessaire.
3. A partir de la distribution a posteriori bivariée obtenue à l'étape 2, déterminer la probabilité que $(p_e - p_c) > -\delta$, (soit $\delta = 0.10$). Cette probabilité peut être déterminée en comptant le nombre de simulations dans lesquelles cette condition est remplie.
4. Si le 5^{ème} centile de la distribution bivariée de $(p_e - p_c)$ dépasse -0.10 , alors la probabilité que le taux de réussite pour E ne soit pas plus de 0.10 inférieur à C est d'au moins 0.95. Cette constatation appuie la décision d'autoriser l'enregistrement et l'approbation précoces des appareils puisque des renseignements suffisants sont disponibles au moment de l'*AI* pour appuyer la demande de non-infériorité.
5. Si le 5^{ème} centile de $(p_e - p_c)$ ne dépasse pas -0.10 , alors la probabilité que le taux de réussite de E ne soit pas plus de 0.10 inférieur à celui de C est inférieure à 0.95. Par conséquent, le statut du point final pour les patients évaluables restants doit être obtenu et un calcul de probabilité postérieur final effectué après que tous ces patients sont théoriquement dus pour le suivi du 24^{ème} mois en répétant les étapes décrites ci-dessus.

3.2.2 Simulation

Cet exemple est une étude visant à démontrer la non-infériorité clinique d'un nouveau dispositif médical expérimental par rapport au contrôle.

Par conséquent, trois scénarios ont été initialement évalués. Celles-ci étaient $(p_e0, p_c0) = (0.85, 0.85)$, $(0.74, 0.85)$ et $(0.87, 0.85)$. Ceux-ci reflètent la véritable équivalence, la vraie infériorité et la légère supériorité. Les *second et troisième* scénarios ont été utilisés pour obtenir des estimations de puissance statistique et le *premier* a été utilisé pour obtenir des estimations d'erreur de type I.

Dans les trois scénarios le nombre d'échantillons simulés $numsim = 3000, N_e1 = 140, N_c1 = 70, N_e2 = 60, N_c2 = 30$ et $delta = -0.10$.

Les résultats ont été pris de l'article de **Greg M (2011)**

scénarios 1 : pour $(p_e0, p_c0) = (0.74, 0.85)$

Type de rejet	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulée
Rejeter à titre provisoire	102	3.40	102	3.40
Rejeter en finale	52	1.73	154	5.13
Ne pas rejeter	2846	94.87	3000	100.00

scénarios 2 : pour $(p_e0, p_c0) = (0.85, 0.85)$

Type de rejet	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulée
Rejeter à titre provisoire	1991	66.37	1991	66.37
Rejeter en finale	468	15.60	2459	81.97
Ne pas rejeter	541	18.03	3000	100.00

scénarios 3 : pour $(p_e0, p_c0) = (0.87, 0.85)$

Type de rejet	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulée
Rejeter à titre provisoire	2278	75.93	2278	75.93
Rejeter en finale	438	14.60	2716	90.53
Ne pas rejeter	284	9.47	3000	100.00

Il y a aussi un autre scénarios lorsque Ce cas a été comparé à un cas où le critère de probabilité postérieure AI a été augmenté de 0.95 à 0.975 à titre d'exemple de paramètres de conception «d'ajustement» pour améliorer les caractéristiques de fonctionnement fréquentistes

Les résultats pour "l'infériorité limite" définie comme $(p_e0, p_c0) = (0.75, 0.85)$ étaient :

Type de rejet	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulée
Rejeter à titre provisoire	154	5.13	154	5.13
Rejeter en finale	65	2.17	219	7.30
Ne pas rejeter	2781	92.70	3000	100.00

Discussion :

Dans le premier scénarios le taux d'erreur de type I proche de 0.05(5.13%) lorsque. D'après le second scénarios si les vraies probabilités de succès sont toutes les deux égales à 0.85, la probabilité estimée d'atteindre le critère d'enregistrement précoce est de 0.664. Il'ya une probabilité de 0.156 que l'étude échouera à rejeter la non-infériorité pendant l' AI , mais sera alors capable de rejeter la non-infériorité durant l'analyse finale. Par conséquent, la puissance statistique globale est estimée à 82.0%.

Cependant, si $(p_e0, p_c0) = (0.87, 0.85)$ (*scénarios 3*), la probabilité d'un arrêt précoce augmente de 0.663 (*dans scénarios 2*) à 0.759 (*dans scénarios 3*). De plus, il ya une probabilité de 0.146 que l'étude échoue à rejeter la non-infériorité à l' AI , mais soit ensuite capable de rejeter la non-infériorité au cours de l'analyse finale. Par conséquent, la puissance statistique globale de ce scénario est de 90.5%.

Pour le dernier scénario, l'erreur de type I est un peu plus grande que 5%(7,3%). L'erreur de type I peut être contrôlée en ajustant les seuils de probabilité a posteriori. Par exemple, le seuil AI pourrait être augmenté de 0.95 à 0.975 comme illustré ci-dessous.

3.2.3 Conception à l'aide du contrôle a priori informative

Les Caractéristiques de fonctionnement sous une a priori informative pour C

L'interprétation de l'erreur de type I n'est pas aussi simple dans le cas des a priori

informatives que dans celle des a priori non informatifs. Pour cette raison, dans les essais bayésiens, voici quelques points à considérer concernant les erreurs de type I : La **FDA** considère les erreurs de type I, ainsi que d'autres caractéristiques opérationnelles pour la conception de l'essai, dans l'évaluation de la soumission. Nous nous efforçons de contrôler raisonnablement le taux d'erreur de type I. Une caractérisation adéquate des caractéristiques de fonctionnement de toute conception particulière peut nécessiter des simulations approfondies.

Lors de l'utilisation d'informations a priori, il peut être approprié de contrôler les erreurs de type I à un niveau moins strict que lorsqu'aucune information a priori n'est utilisée. Par exemple, si l'information a priori est favorable, l'essai en cours peut ne pas avoir besoin de fournir autant d'informations concernant la sécurité et l'efficacité. Le degré auquel nous pourrions relâcher le contrôle des erreurs de type I est une décision au cas par cas qui dépend de nombreux facteurs, principalement la confiance que nous avons dans le précédent. Nous pouvons recommander d'actualiser l'information historique / a priori si la distribution antérieure est trop informative par rapport à l'étude actuelle. Ce qui constitue "trop informatif" est aussi une décision au cas par cas.

Les choix de la priori bayésiens

Toutes les distributions bêta considérées, à l'exception de l'a priori non informatif de Jeffreys, ont des valeurs attendues égales à $\frac{a}{(a+b)} \cong 0.72$. Le «+0.5» supplémentaire dans le numérateur et le dénominateur de ces ratios sont des facteurs de «rétrécissement» qui réduisent légèrement le taux. Les variables a priori de contrôle informatif varient en fonction de la quantité d'information qu'elles contiennent ($a + b$). Pour les primitives de contrôle informatives, ces valeurs sont 1, 10, 20, 30, 40, 50, 60 et 70, respectivement. En utilisant la stratégie de simulation.

Les valeurs suivantes ont été déterminées en utilisant le programme *R*. Pour ces simulations, des tirages 3000/3000 beta - binomial ont été utilisés pour chaque simulation.

Figure 1
Interim Analysis Rejection Probabilities
for Different Amounts of Prior Control Information
in Terms of Beta Parameters $a+b$

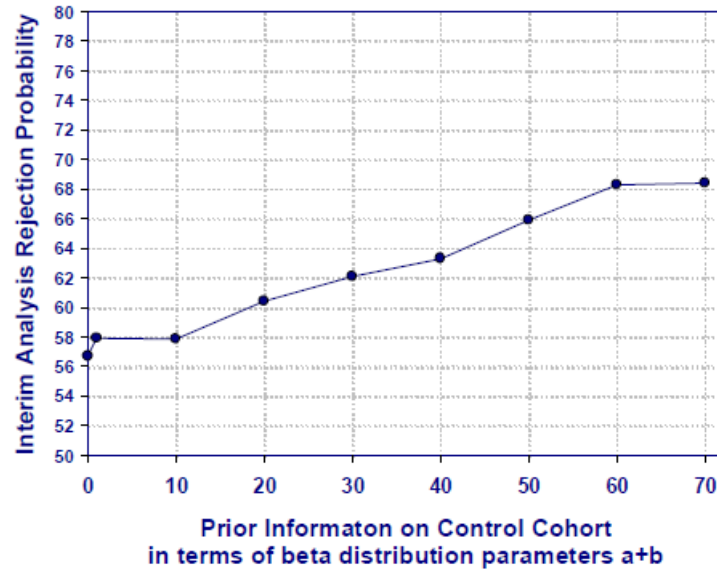
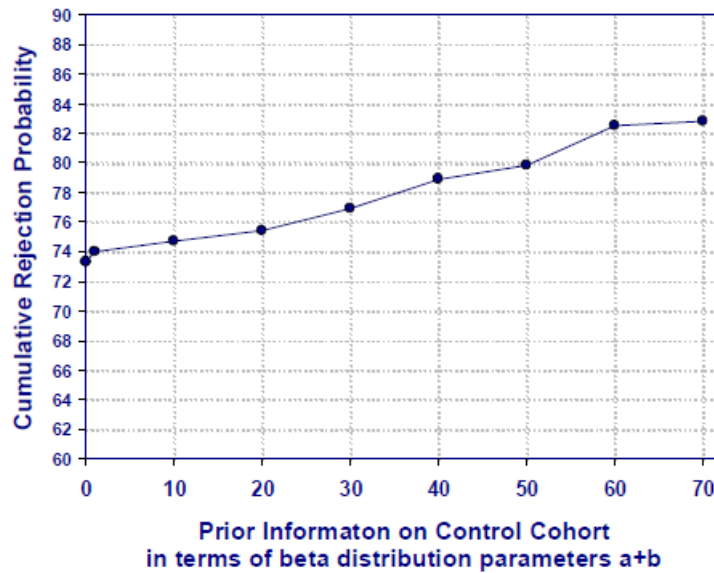


Figure 2
Cumulative Rejection Probabilities
for Different Amounts of Prior Control Information
in Terms of Beta Parameters $a+b$



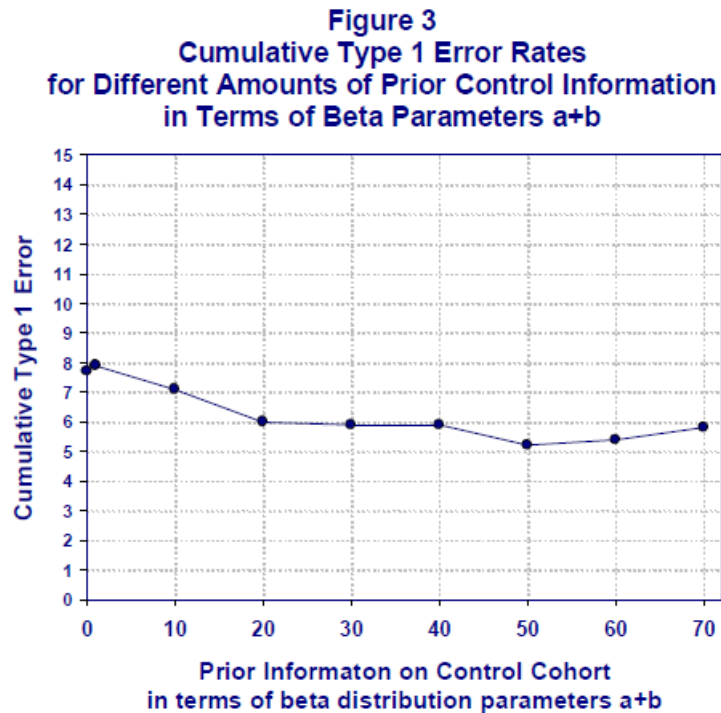


Figure 1 : Résume l'analyse intermédiaire de la puissance fréquentiste en utilisant une distribution a priori non informative pour E et une a priori informative pour C, avec des sommes de $a + b$ égales à $1 + 1, 10 + 1, \dots, 70 + 1$. Les termes "+1" représentent les termes 0.5 qui forment le précédent de Jeffreys. La puissance AI est augmentée de 56.6% à 68.3% grâce à l'utilisation d'une a priori sur les commandes pour le choix final de la conception "60 + 1".

Figure 2 : Résume le pouvoir fréquentiste cumulatif en utilisant une distribution non informative et pour le contrôle informatif avec des sommes de $a + b$ égales à $1 + 1, 10 + 1, 20 + 1, 30 + 1, 40 + 1, 50 + 1, 60 + 1$ et $70 + 1$. La puissance cumulée est augmentée de 73.3% à 82.7% grâce à l'utilisation d'une a priori informative sur les commandes pour le choix final du design "60 + 1". Au moins 80% de la puissance est conservée pour une conception qui compare 182E vs 91C à l'AI sur la base d'une probabilité a posteriori calculée avec un contrôle informatif a priori défini par bêta ($a = 43.7, b = 17.3$) et comprend alors 78 dispositifs supplémentaires E et 39C à ajouter à l'échantillon d'analyse après que tous les patients évaluable sont théoriquement dus pour le suivi du 24 mois.

Figure 3 : Résume l'erreur de type I pour chaque conception calculée en supposant que les taux réels de SCC au 24 mois étaient de 62% et de 72%, respectivement, pour les dispositifs E et C. Pour cette conception particulière, il n'a pas été observé qu'il existe une forte association positive entre la quantité d'informations dans l'a priori et le taux

d'erreur de type I. Pour l'a priori définie par bêta ($a = 43.7, b = 17.3$), le taux d'erreur de type I estimé est de 5.4%. Étant donné que ce taux estimé est si proche de 5%, on a estimé qu'aucun ajustement de la probabilité a posteriori bayésienne au moment de l'AI ne devait être effectué afin de mieux contrôler l'erreur de type I.

Ce chapitre montre qu'une conception probabiliste a posteriori bayésienne avec une seule analyse intermédiaire en temps réel sans ou avec un dispositif de contrôle informatif antérieur a un potentiel substantiel pour raccourcir le délai d'approbation tout en contrôlant les erreurs de type I et de type II.

Conclusion générale

Chaque phase joue un rôle très important dans le cycle de développement des médicaments. Ce travail est basé sur l'étude de phase II qui consiste à tester l'efficacité d'un nouveau traitement par rapport à un autre contrôle.

Dans notre travail nous avons étudié l'importance de tests d'équivalence et de non-infériorité dans le concept des essais cliniques, soit en utilisant l'approche fréquentiste ou bayésienne pour étudier les plans adaptatifs prenant en compte les réponses disponibles.

Nous avons compris que la probabilité a posteriori et la probabilité prédictive fournissent un résumé utile pour étudier la supériorité d'un nouveau traitement par rapport à un traitement de référence, dans un concept séquentiel.

Bien que les approches fréquentistes ont historiquement dominé le domaine des designs et l'analyse des essais cliniques [7], les méthodes bayésiennes fournissent plusieurs propriétés attirantes [3]. Il s'agit notamment de la capacité d'incorporer l'information obtenue avant et pendant l'essai, aussi les tests d'équivalence et de non-infériorité devenus l'un des meilleurs tests statistiques pour étudier l'efficacité ou bien la supériorité de nouveaux traitements.

Résumé

L'objectif de ce mémoire, motivé par des problèmes méthodologiques qui se posent dans le contexte des essais thérapeutiques, est de donner quelques méthodes adaptées à l'analyse et au contrôle de ces essais clinique.

Nous avons essayé dans ce travail, d'exposer les méthodes statistiques fréquentistes et bayésiennes utilisés dans l'évaluations des résultats cliniques.

Nous avons parlé des tests d'équivalence et de non-infériorité qui sont utilisées pour comparer l'efficacité d'un nouveaux traitement par rapport à un autre de référence en utilisant des analyse intermédiaires pour arrêter le plus tôt possible l'essai pour des considérations éthiques et commerciales.

Annexe

Programme :1

```
>#Pour Pr(y = i/x)
>p0 = 0.6; n = 23; Nmax = 40; m = (Nmax-n); x = 16; a = 0.6; b = 0.4
>ab = c(x + a, n - x + b)
>ys = 0 : m
>pred = pbeta(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)
>fdis
>#Pour Bi = Pr(p > p0/x, y = i), i = 1 : m
>f = 1 - (pbeta(p0, a + x + 0, b + Nmax-x - 0))
>fdis = round(cbind(1, f), 4)
>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

```

Programme :2

```

>library(lattice)
>posterior <-function(ai0,bi0,ac0,bc0,xi1,ni1,xc1,nc1,delta,numsim) {
ai1 <- ai0 + xi1
bi1 <- bi0 + (ni1 - xi1)
ac1 <- ac0 + xc1
bc1 <- bc0 + (nc1 - xc1)
pi1 <- rbeta(numsim,ai1,bi1)
pc1 <- rbeta(numsim,ac1,bc1)
pstdif <- pi1 - pc1;
pstdif.ge.delta <- pstdif > delta
sum.pstdif.ge.delta <- sum(pstdif.ge.delta)
posterior.prob <- sum.pstdif.ge.delta / length(pstdif.ge.delta)
rate i <- xi1/ni1
rate c <- xc1/nc1
rate s <- c(rate i, rate c)
require(stats)
out=list(Rates=rate,
Posterior prob=posterior.prob,
histogram(pstdif, nint=50,xlab='Delta',main="Posterior Distribution for
Device Difference in Mo. 24 CCS")
)
out
}
>posterior(0.5,0.5,0.5,0.5,xi1,ni1,xc1,nc1,delta,numsim)
pi0=0.74; pc0=0.85; ni1=140; nc1=70; ni2=60; nc2=30; numsim=3000;delta=-0.10

```

Programme 3 :

```

>library(lattice)
>posterior <-function(ai0,bi0,ac0,bc0,xi1,ni1,xc1,nc1,delta,numsim) {
ai1 <- ai0 + xi1
bi1 <- bi0 + (ni1 - xi1)

```

```

ac1 <- ac0 + xc1
bc1 <- bc0 + (nc1 - xc1)
pi1 <- rbeta(numsim,ai1,bi1)
pc1 <- rbeta(numsim,ac1,bc1)
pstdif <- pi1 - pc1;
pstdif.ge.delta <- pstdif > delta
sum.pstdif.ge.delta <- sum(pstdif.ge.delta)
posterior.prob <- sum.pstdif.ge.delta / length(pstdif.ge.delta)
rate i <- xi1/ni1
rate c <- xc1/nc1
rates <- c(rate i, rate c)
require(stats)
out=list(Rates=rates,
Posterior prob=posterior.prob,
histogram(pstdif, nint=50,xlab= 'Delta',main="Posterior Distribution for
Device Difference in Mo. 24 CCS")
)
out
}
>posterior(0.5,0.5,0.5,0.5,xi1,ni1,xc1,nc1,delta,numsim)
pi0=0.85; pc0=0.85; ni1=140; nc1=70; ni2=60; nc2=30; numsim=3000;delta=-0.10

```

Programme 4 :

```

>library(lattice)
>posterior <-function(ai0,bi0,ac0,bc0,xi1,ni1,xc1,nc1,delta,numsim) {
ai1 <- ai0 + xi1
bi1 <- bi0 + (ni1 - xi1)
ac1 <- ac0 + xc1
bc1 <- bc0 + (nc1 - xc1)
pi1 <- rbeta(numsim,ai1,bi1)
pc1 <- rbeta(numsim,ac1,bc1)
pstdif <- pi1 - pc1;
pstdif.ge.delta <- pstdif > delta
sum.pstdif.ge.delta <- sum(pstdif.ge.delta)
posterior.prob <- sum.pstdif.ge.delta / length(pstdif.ge.delta)
rate i <- xi1/ni1

```

```

rate c <- xc1/nc1
rates <- c(rate i, rate c)
require(stats)
out=list(Rates=rates,
Posterior prob=posterior.prob,
histogram(pstdif, nint=50,xlab= 'Delta',main="Posterior Distribution for
Device Difference in Mo. 24 CCS")
)
out
}
>posterior(0.5,0.5,0.5,0.5,xi1,ni1,xc1,nc1,delta,numsim)
pi0=0.87; pc0=0.85; ni1=140; nc1=70; ni2=60; nc2=30; numsim=3000;delta=-0.10

```

Évaluation d'erreur de type I :

```

gm_adapt2(hypPs=0.85,hypPc=0.85,simPs=0.85,simPc=0.85,ns1=140,nc1=70,ns2=60,nc2=30
,as0=0.5,bs0=0.5,ac0=0.5,bc0=0.5,delta<-0.10,postc1<-.95,postc2<-.95)

Binomial Draws  Beta Draw  IA Post.>  Final Post.>  Set Ps
"5000.000"      5000.000"  "0.950"   "0.950"       "0.850"

Set Pc  IA Stopping Prob Cond.  Final Reject. Prob  Overall Reject. Prob  Max N
"0.850"  "0.6566"  "0.148"  "0.8046"  "300.000"

Expected N
"240.906"

```

Évaluation de puissance :

```

gm_adapt2(hypPs=0.85,hypPc=0.85,simPs=0.75,simPc=0.85,ns1=140,nc1=70,ns2=60,nc2=30
,as0=0.5,bs0=0.5,ac0=0.5,bc0=0.5,delta<-0.10,postc1<-.95,postc2<-.95)

Binomial Draws  Beta Draw  IA Post.>  Final Post.>  Set Ps
"5000.000"      5000.000"  "0.950"   "0.950"       "0.750"

Set Pc  IA Stopping Prob Cond.  Final Reject. Prob  Overall Reject. Prob  Max N
"0.850"  "0.049"  "0.019"  "0.068"  "300.000"

Expected N
"295.590"

```

Évaluation de puissance :

```

gm_adapt2(hypPs=0.85,hypPc=0.85,simPs=0.75,simPc=0.85,ns1=140,nc1=70,ns2=60,nc2=30
,as0=0.5,bs0=0.5,ac0=0.5,bc0=0.5,delta<-0.10,postc1<-.975,postc2<-.95)

Binomial Draws  Beta Draw  IA Post.>  Final Post.>  Set Ps
"5000.000"      5000.000"  "0.975"   "0.950"       "0.750"

Set Pc  IA Stopping Prob Cond.  Final Reject. Prob  Overall Reject. Prob  Max N
"0.850"  "0.0274"  "0.0352"  "0.0623"  "300.000"

Expected N
"298.390"

```

Annexe I

Notation and name	Random variable and domain	Probability density function	Parameter restrictions
1. BeBi(n, g, h) Beta-binomial	$y = 0, 1, \dots, n$	$\binom{n}{y} \frac{B(g+y, h+n-y)}{B(g, h)}$	n positive integer, $g > 0, h > 0$.
2. NeBi(n, l) Negative-binomial	$y = 0, 1, 2, \dots$	$\binom{y+n-1}{n-1} l^n (1-l)^n$	n positive integer, $0 < l < 1$.
3. InBe(k, g, h) Inverse-beta	$y > 0$	$\frac{h^g y^{g-1}}{B(k, g)(h+y)^{g+k}}$	$k > 0, g > 0, h > 0$.
4. DiMu(n, g, h) Dirichlet-multinomial	$y = (y_1, y_2, \dots, y_d)$ $y_i = 0, 1, \dots, n$ $\sum y_i \leq n$	$\binom{n}{y} \frac{D(g+y, h+n-\sum y_i)}{D(g, h)}$	$g = (g_1, \dots, g_d)$, $g_i \geq 0$, $h > 0$.
5. St(k, b, c) Student	$m \in \mathbb{R}^1$	$\frac{1}{B(\frac{1}{2}, \frac{1}{2}k)(kc)^{1/2} \{1 + (kc)^{-1}(m-b)^2\}^{(k+1)/2}}$	$b \in \mathbb{R}^1$, $c > 0, k > 0$.
Si(k, g, h) Siegel	$v > 0$	$\frac{v^{(g/2)-1}}{B(\frac{1}{2}k, \frac{1}{2}g) h^{g/2} (1+h^{-1}v)^{(k+g)/2}}$	$k > 0$, $g > 0, h > 0$.
StSi($k; b, c; g, h$) Student-Siegel	(m, v) $m \in \mathbb{R}^1, v > 0$	$\frac{v^{(g/2)-1}}{D(\frac{1}{2}, \frac{1}{2}k, \frac{1}{2}g)(kc)^{1/2} h^{g/2} \{1 + (kc)^{-1}(m-b)^2 + h^{-1}v\}^{(k+g+1)/2}}$	$b \in \mathbb{R}^1, c > 0$, $k > 0, g > 0, h > 0$.
6. St _d (k, b, c) Student	$m \in \mathbb{R}^d$	$\frac{1}{D(\frac{1}{2}, \frac{1}{2}k, \frac{1}{2}(k-d+1)) kc ^{1/2} \{1 + (m-b)'(kc)^{-1}(m-b)\}^{(k+1)/2}}$	$b \in \mathbb{R}^d, c \in \mathbb{S}^d$, $k > d - 1$.
Si _d (k, g, h) Siegel	$v \in \mathbb{S}^d$	$\frac{ v ^{(g-d-1)/2}}{B_d(\frac{1}{2}k, \frac{1}{2}g) h ^{g/2} 1 + h^{-1}v ^{(k+g)/2}}$	$k > d - 1, g > d - 1$, $h \in \mathbb{S}^d$.
StSi _d ($k; b, c; g, h$) Student-Siegel	(m, v) $m \in \mathbb{R}^d, v \in \mathbb{S}^d$	$\frac{\Gamma_d(\frac{1}{2}(k+g+1)) v ^{(g-d-1)/2}}{\Gamma_d(\frac{1}{2}k) \Gamma_d(\frac{1}{2}g) h ^{d/2} h ^{g/2} 1 + (kc)^{-1}(m-b)'(m-b) + h^{-1}v ^{(k+g+1)/2}}$	$k > d - 1, b \in \mathbb{R}^d$, $c \in \mathbb{S}^d, g > d - 1$, $h \in \mathbb{S}^d$.

Table 6 : Les distributions prédictives pour quelque modèles

Bibliographie

- [1] Aitchison J and Dunsmore I.R. (2008). *Statistical Prediction Analysis*, British ; 74-25649.
- [2] C. Jennison and Bruce W. Turnbull. (2000). *Group Sequential Methods with Applications to Clinical Trials*. Chapman Hall/CRC USA ; ISBN 0-8493-0316-8.
- [3] Carlin P. B, and Louis T. A. (2000). *Bayes And Empirical Bayes Methods For Data +Analysis*. Second edition. Chapman Hall/CRC Statistical Science Series, USA.
- [4] Chow S. C. (2010). *adaptative methods for clinical trials*. Champman Hall/CRC Biostatistics Series, USA ; ISBN : 1-58488-776-1.
- [5] Chow S. C, et Chang M. (2007). *Adaptative design methods in clinical trials*. Champman Hall/CRC Biostatistics Series, USA.
- [6] Faisant J. (2014). *Les tests d'équivalence et de non-infériorité*, IREM de Caen - BN, groupe « Histoire des probabilités et de la statistique ».
- [7] Friedman L. M, Furberg C. D, et Demets D. L. (2010). *Fundamentals of clinical trials*. 4^{eme} édition. Springer Science+Business Media, New York.
- [8] Goulet V. (2015). *Méthodes Numériques en Actuariat Partie I Programmation en R*, Université Laval.
- [9] Guillaume M.(2013). *Modélisation du processus d'inclusion de patients dans un essai clinique multicentrique*, Thèse de Doctorat, Université Toulouse III Paul Sabatier (UT3 Paul Sabatier).
- [10] Guosheng Yin, Nan Chen, et J. Jack Lee (2012). *Phase II trial design with Bayesian adaptive randomization and predictive probability* ;J R Stat Soc Ser C Appl Stat. 61(2) 10.1111/j.1467-9876.2011.01006.x.
- [11] Greg M. (2011). *Frequentist Operating Characteristics of Bayesian Posterior Probability Designs for Medical Device Trials that Include a Single, Late-information-time, Interim Analysis*, 1 Biomedical Statistical Consulting, 1357 Garden Road, Wynnewood, PA 19096.

-
- [12] Helmy. (1999). *Statistical inference for equivalence trials with ordinal responses*. Computational Statistics Data Analysis; 51 (2007) 5918-5926.
- [13] Lee J. J and Liu D. D. (2008). *A predictive probability design for phase II cancer clinical trials*. Clinical Trials, 5(2) : 93 106.
- [14] Piaggio. G, Elbourne.D.R, Pocock. S. J, Evans. S. J. W, Altman. D. J. (2012). *Reporting of Noninferiority and Equivalence Randomized Trials* ; Journal of American Medical Association, 308(24) :2594-2604.
- [15] Robert C. P. (2006). *Le choix bayésien : principe et pratique*. Springer-Verlag France, Paris ; ISBN-10 :2-287-25173-1.
- [16] Schultz J. R, Nichol F. R, Elfring G. L and Weed S. D. (1973). *Multiple-stage procedures for drug screening*. Biometrics ; 29 : 293-300.
- [17] Setta M. (2005). *Sequential Designs In Phase II Clinical Trials*, Thèse de doctorat. Doctorat en informatique. Université de Amsterdam.[[https : //esc.fnwi.uva.nl/thesis/centraal/les/f930801131.doc](https://esc.fnwi.uva.nl/thesis/centraal/les/f930801131.doc)].
- [18] Simon R. (1989). *Optimal two-stage designs for phase II Clinical Trials*. New York. Cont Clin Trials; 10 : 1-10.
- [19] Tessy G. (2011). *Essai clinique comparant la combinaison Nifurtimox-Eflornithine au traitement de référence Eflornithine pour le traitement de la trypanosomiase humaine africaine à Trypanosoma Brucei Gambiense en phase meningo-encéphalique*, Université de Strasbourg Id : dumas-00623076.
- [20] Thall, P.F. and Simon, R.M. (1994). Practical Bayesian guidelines for phase IIB clinical trials. Biometrics, 50, 337349.
- [21] Thall, P.F. Simon, R.M. and Estey, E.H. (1995). Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes. Statistics in Medicine, 14, 357379.
- [22] Thomas R. Fleming. (1982). *One-Sample Multiple Testing Procedure for Phase II Clinical Trials*. Biometrics, Vol. 38, 1, 143-151. International Biometric Society.
- [23] Wilfred J. Westlake. (1976). *Symmetrical Confidence Intervals for Bioequivalence Trials*. Biometrics ; 32 : 741-744.
- [24] Zerari A. (2013). *Approche Bayésienne aux essais cliniques séquentiels*, Thèse de Magister. Département De Mathématiques, Faculté Des Sciences Exactes, Universitaire Mentouri Constantine.

- [25] *Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials*, Document issued on : February 5, 2010,
<http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071072.htm>
- [26] Ich Harmonised Tripartite Guideline. (5 February 1998). *Statistical Principles For Clinical Trials* E9 : Current Step 4 version. (<http://www.ich.org>).