

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE Mohamed Seddik Ben Yahia – Jijel

Faculté des Sciences Exactes et Informatique

Département de Mathématiques



Mémoire

Pour l'obtention du diplôme de : Master

Spécialité : Mathématiques Appliquées

Option : Probabilités et statistique

Thème

Les designs Bayésiens séquentiels de groupe

Présenté par :

Benhamouda Fatiha

Boukhellout Nadjat

Devant le jury :

Présidente : D. Azzam-Laouir Pro Univ. Jijel

Encadreur : Z. Djeridi M.A.A Univ. Jijel

Examinatrice : N. Sellami M.A.A Univ. Jijel

Examinatrice : Z. Abdi M.A.A Univ. Jijel

Promotion 2016/2017

Remerciement

Tout la gratitude et le merci au dieu le tout puissant qui a fait que ce travail soit mené et terminé.

Nous adressons nos vifs remerciements à notre promotrice consultante Madame DJERIDI ZAHRA pour nous avoir diligenté tout au long de ce travail, pour sa compréhension, sa patience, sa compétence, et ces remarques qui nous ont été précieuses.

Nous tenons à remercier également les membres du jury pour l'honneur qu'ils nous font en acceptant de juger et d'évaluer notre travail: Professeur AZZAM-LAOUIR DALILA, Madame ABDI ZINEB et Madame SELLAMI NAWEL.

Nous présentons nos chaleureux remerciements aux enseignants du département de mathématiques pour leurs aides et orientations durant notre formation.

Nos derniers remerciements, vont à tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail.

Fatiha-Nadjet

Table des matières

Introduction Générale	3
Notations	5
1 Préliminaires	6
1.1 Aperçu sur les essais cliniques	6
1.1.1 Ethiques et Randomisation	7
1.1.2 Puissance	8
1.1.3 Puissance et nombre de sujets nécessaire	10
1.1.4 Règle d'arrêt	14
1.2 Analyses intermédiaires	15
1.2.1 Objectifs des analyses intermédiaires	15
1.3 Test séquentiel de groupe	16
1.3.1 Concepts de base	17
1.3.2 Une formulation unifiée	18
1.4 Application dans les essais cliniques	19
1.4.1 Test d'une moyenne pour une population simple	19
1.4.2 Comparaison de deux traitements parallèles	20
1.4.3 Comparaison de deux traitements appaillés	21
1.4.4 Test d'O'Brien et Fleming	22
2 Tests séquentiels Bayésiens de groupe	27
2.1 Le prespective Bayésien	27
2.2 Paradigme Bayésien	29
2.2.1 Utilisation séquentielles du théorème de Bayses	32
2.2.2 Prédiction	33
2.3 Test séquentiel Bayésien de groupe	37
2.3.1 Règles d'arrêt Bayésiennes	37
2.3.2 Choix de distribution a priori	41

3	Designs séquentiels de groupe	43
3.1	Puissance conditionnelle	43
3.2	Puissance Prédictive	49
3.3	Application	53
	Conclusion générale	59
	Résumé	60
	Annexe	61
	Bibliographie	70

Introduction générale

L'analyse séquentielle ou le test d'hypothèse séquentiel est une analyse statistique où la taille de l'échantillon n'est pas fixée à l'avance. Plutôt, les données sont évaluées au fur et à mesure qu'elles sont recueillies et l'échantillonnage s'arrête selon une règle prédéfinie, dès que des résultats significatifs sont observés. Ainsi une conclusion peut parfois être atteinte à un stade beaucoup plus précoce que ce qui serait possible avec des tests d'hypothèses ou des estimations plus classiques, à un coût financier ou humain par conséquent inférieure.

Cette méthode a été développée initialement pour des installations industrielles (**Shewhart(1931)**), ensuite elle a été modifiée pour les essais cliniques (**C. Jennison(2000)**).

Dans ce travail nous avons choisit la méthode d'OBrien et Fleming comme exemple des méthodes séquentiels pour les tests d'hypothèses utilisant la statistique normalisée afin de trouver des frontières d'arrêts plus adéquates.

L'analyse statistique séquentielle standard n'est pas bien équipée pour gérer une règle d'arrêt non rigide et certaines recherches ont impliqué des approches plus souples, telle que les procédure de la stochastique curtailements qui surveillent la probabilité qu'un test d'échantillon fixe spécifique atteigne un résultat significatif((**Lan et al.(1982)**, **Choi et al(1985)**, **Spiegelhalter et al.(1986)**). L'approche Bayésienne offre, également, la possibilité d'un suivi souple des résultats des essais, et nous devons contraster et comparer les approches bayésienne et fréquentiste. Pour se faire, nous avons choisit la notion de puissance prédictive et puissance conditionnelle pour analyser les données des survie.

Ce mémoire comporte trois chapitres :

- Le premier chapitre est un aperçu sur les essais cliniques, et les analyse séquentielles quant à leur motivation, au principe de leurs méthodes notamment l'étude des tests séquentiels de groupe. Des applications numériques ont été étudiées pour illustrer les méthodes utilisées.

- Dans le deuxième chapitre, nous développons les méthodes Bayésiennes qui donnent beaucoup de souplesse au monitoring des essais cliniques.

•Finalement, dans le troisième chapitre, nous avons choisit la procédure stochastique curtalement (**PSC**) qui utilise la notion de puissance afin de trouver des designs plus souples pour bien surveiller les essais cliniques en utilisant les méthodes séquentielles de groupes fréquentistes comme bayésiennes en terminant par un exemple d'application dans le domaine des données de survie.

Notations

C_n^x	Coefficient binomial.
$v.a$	Variable aléatoire.
$p - valeur$	La probabilité observer.
\propto	Proportionnelle à .
$\pi(\theta)$	La loi a priori.
$\pi(\theta x)$	La loi a posteriori.
$f(x y)$	La loi prédictive.
$DSMB$	Données et sûreté monitoring board.
EMV	L'estimateur du maximum de vraisemblance.
RBO	Le rapport des odds a posteriori.
PSC	Procédure de stochastique curtalement.
I	Le niveau d'information.
K	Nombre de groupes (analyses, regards, étapes) d'une procédure séquentielle de g
ϕ	La fonction de distribution cumulative de la loi normale standard.

Chapitre 1

Préliminaires

L'évaluation de l'efficacité d'un nouveau traitement doit reposer sur des études cliniques.

À côté de la recherche d'un effet du traitement, il faut prendre, en compte, la protection des patients depuis la période de **Hippocrate**.

Dans ce chapitre, nous allons introduire la notion des essais cliniques, avec ses différentes phases, la puissance d'un test et sa relation avec le nombre du sujet nécessaire. Ainsi l'importance des analyses intermédiaire. Nous expliciterons aussi les tests séquentiels qui sont la méthode la plus répandue pour définir la règle d'arrêt, pour conclure l'effet du traitement.

1.1 Aperçu sur les essais cliniques

Les essais cliniques sont implicites dans le développement de médicaments et sont menés pour recueillir des données sur la sécurité et l'efficacité des interventions sanitaire (**Brian S. Everitt et Andrew Pikles (1999)**). Le développement clinique des médicaments est structuré en quatre phases :

- Les essais de phase I : incluent des études introductives pour étudier le métabolisme et les interventions pharmacologiques chez les humains et les effets secondaires associés aux doses croissantes. Ils sont également tenus de fournir des preuves précoces d'efficacité.

- Les essais de phase II : sont des études contrôlées menées pour évaluer l'efficacité du médicament pour une indication particulière chez les patients de la maladie spécifique à l'étude et pour déterminer les effets secondaires et les risques à court terme courants.

- Les essais de phase III : sont des études contrôlées élargies qui sont réalisées une fois que des preuves préliminaires, suggérant l'efficacité du médicament, ont été obtenues. Ils visent également à recueillir des informations supplémentaires sur la relation bénéfico-risque globale du médicament.

- Les essais de phase VI : sont des études post-commercialisation qui sont menées pour obtenir des informations supplémentaires, y compris les risques ; les avantages et l'utilisation optimale du médicament.

Selon **De Martini (2013)**, au cours des dernières années, en moyenne (environ) 2600 phase I, 3700 phase II, 2300 phase III et 1800 essais de phase VI ont été présentés annuellement pour approbation dans le cadre de la Food and Drug Administration (**FDA**) des États-Unis (source : clinicaltrial.gov).

Ces essais s'élèvent à environ 60% de ceux qui se déroulent à l'échelle mondiale chaque année. En règle générale, la quantité totale d'essais dans le monde entier se divise comme suit : 60% dans le cadre de la Food and Drug Administration (**FDA**) des États-Unis ; 30% dans le cadre de l'agence européenne des médicaments (**EMA**) ; Et les 10% restants sous d'autres agences, principalement l'Agence japonaise des produits pharmaceutiques et des dispositifs médicaux (**PMDA**). Il s'ensuit qu'un nombre impressionnant d'épreuves sont simultanément opérationnelles dans le monde entier chaque année.

1.1.1 Ethiques et Randomisation

Depuis la période de **Hippocrate**, les médecins occidentaux ont pris un serment, dans du lequel, ils jurent pour protéger leur " patients de celui qui soit délétère et malfaisant ". Malheureusement un tel serment n'est pas parvenu à arrêter beaucoup de thérapies dommageables ou à diminuer la persistance des pratiques barbares telles que l'abondance du sang (**Brain S. Everitt et Andrew Pikles (1999)**).

Les issues morales, dans la médecine en général, et dans les essais cliniques en particulier, sont clairement de grande importance et présentent un champ de mines potentiel particulièrement pour les statisticiens qui sont les plus impliqués et peut-être d'avantage intéressés aux problèmes pragmatiques de l'analyse des données produites dans les essais. Néanmoins, avec tout le personnel impliqué dans les essais, le statisticien doit partager la responsabilité générale de conduite morale d'un essai.

Et il y a, en outre, quelques secteurs de conduite d'essai où le statisticien doit prendre la responsabilité particulière de s'assurer que la conduite proposée et la conduite réelle de l'essai soient appropriées.

Une question morale souvent identifiée avec des essais cliniques est celle

de randomisation. Les essais contrôlés randomisés sont maintenant largement répandues dans la recherche médicale. Deux exemples récents de nombreuses essais prise tous les ans incluent :

- Une étude multicentre d'un régime pauvre en protéine sur la progression échec rénal chronique chez les enfants (**Wingen et al. (1997)**).
- Une étude de l'immunothérapie pour l'asthme allergique chez les enfants (**Ad-kinson et al. (1997)**).

Si la randomisation est la première priorité dans un essai clinique acceptable, l'aveuglement vient une deuxième étroite. L'idée fondamentale de l'aveuglement est que les patients d'essai, les personnes impliquées de leur gestion et ceux rassemblant des données cliniques des études, ne devraient pas être influencés par la connaissance du traitement assigné. L'aveuglement est nécessaire pour la possibilité de biais du patient, du médecin et de l'évaluation. Il y a un certain nombre de niveaux de l'aveuglement dont les plus importants sont :

- **Simple-aveugle** : Habituellement utilisé pour la situation dans la quelle le patient est ignorant de quel traitement il reçoit.
- **Double-aveugle** : Ici le patient et l'investigateur sont gardé aveugles au traitement du patient. Pour beaucoup d'essais, c'est l'arrangement du choix. Dans les essais de médicaments, l'aveuglement est généralement relativement facile à organiser, mais l'aveuglement des traitements physiques, par exemple, les intervention chirurgicales, est souvent plus difficile.

L'essai contrôlé randomisé en anonymat (double-aveugle) est "l'étalon-or" pour juger la qualité des essais cliniques en général. Mais de telles essais sont encore mal compris par les cliniciens et il y a beaucoup de questions sur la persistance ou non de leur éthique.

Il existe d'autres problèmes éthiques dans les essais cliniques qui concernent directement l'un ou l'autre des aspects statistiques de la conception et de l'analyse. Un exemple consiste à déterminer la taille de l'échantillon appropriée au moyen d'une analyse de puissance, en utilisant un échantillon trop petit ou trop grand serait contraire à l'éthique, un point qui sera abordé dans les prochaines sections.

1.1.2 Puissance

Un essai clinique de type explicatif permet de comparer toutes choses égales par ailleurs, les effets d'un traitement(A) à ceux d'un comparateur pertinent (B), placebo ou traitement de référence. La valeur de la différence observée entre les effets de A et de B fait l'objet d'un test statistique. Il s'agit de confronter cette valeur à celles peuvent résulter de simples d'échantillonnage, au cas où il n'existerait pas de réelle différence entre les effets des

deux traitements. Le test statistique consiste donc, en fait, à tester l'in vraisemblance d'une hypothèse nulle H_0 selon laquelle il n'existe pas de différence réelle entre les effets de A et de B , dans le but de rejeter cette hypothèse H_0 . Cette hypothèse sera rejetée si la probabilité d'observer, une différence au moins aussi grande, de suffit du hasard que celle observée entre les effets de A et de B est inférieure ou égale à 5%, (5% est le seuil de signification que l'on s'impose).

On sait qu'un tel test est assorti de deux risques :

- i) **Le risque**(α) : appelé risque d'erreur du type I, qui est celui de rejeter l'hypothèse nulle au vu des résultats, alors que les effets des traitements comparés ne diffèrent pas dans la réalité. Si l'autorisation de mise sur le marché d'un nouveau produit ne dépendait que de son efficacité (sans tenir compte de sa tolérance) et ne s'appréciait que sur un seul essai, ce serait le risque de mettre sur le marché un traitement inefficace. C'est le risque de rejeter l'hypothèse H_0 alors qu'elle est vraie et qui favorise donc l'hypothèse H_1 , s'appelle seuil de signification ou "risque de première espèce"

$$\alpha = P\{\text{rejeter } H_0/H_0 \text{ vraie}\}$$

- ii) **Le risque** (β) : appelé défaut de puissance risque d'erreur du type II, ou manque de puissance, est le risque de ne pas rejeter H_0 au vu des résultats et donc de ne pas conclure. C'est le risque de "passer à côté" de la démonstration de l'efficacité d'un produit réellement efficace, "risque de deuxième espèce".

$$\beta = P\{\text{ne pas rejeter } H_0/H_1 \text{ vraie}\}$$

- La puissance d'un test $1 - \beta$ est la probabilité de conclure à l'existence d'une différence entre les effets de deux traitements si cette différence existe réellement. Elle est l'aptitude d'un essai à démontrer l'efficacité ou la supériorité d'un produit donné.

Les différentes situations sont résumées dans la figure (1.1) :

La figure (1.1) visualise ces risques si on suppose que les fluctuations d'échantillonnage des valeurs de la différence entre les effets de A et de B ont une distribution normale. La puissance d'un test (qui vaut $1 - \beta$) est donc la probabilité de conclure à l'existence d'une différence entre les effets de deux traitements si cette différence existe réellement. Elle est l'aptitude d'un test (d'un essai) à démontrer l'efficacité ou la supériorité d'un produit donné. On conçoit donc, bien qu'en matière de comparaison de traitements, le

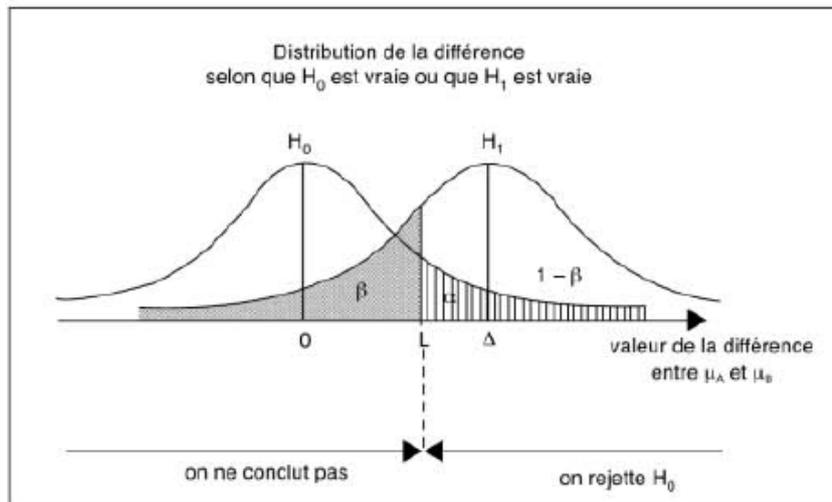


FIGURE 1.1 – Distribution de la différence observée $d = (m_A - m_B)$ selon que H_0 est vraie ou que H_1 est vraie

manque de puissance d'un test puisse être cause du non-rejet de l'hypothèse nulle au même titre que l'équivalence vraie des traitements comparés. C'est pourquoi ne pas mettre en évidence une différence significative entre les effets de A et de B n'autorise pas à affirmer leur équivalence. L'équivalence d'efficacité de deux traitements ne peut être envisagée, en toute rigueur, qu'à la suite de la réalisation d'un essai dit d'équivalence ou de non-infériorité (c'est alors l'hypothèse H_1 qui est testée, pour être rejetée) ou, dans certains cas, à la suite d'un essai classique dont l'effectif des patients inclus garantit une «puissance suffisante». Cette notion de puissance suffisante est, en fait, une exigence plus générale. À quoi servirait, en effet, de réaliser l'essai clinique d'un produit efficace si le test statistique de cet essai manquait à ce point de puissance, que la probabilité de démontrer l'efficacité du produit n'était que de l'ordre de 10 à 20% ? C'est pourquoi la puissance minimale que tout essai clinique devrait pouvoir garantir est consensuellement fixée à au moins 80% C'est-à-dire que le risque bêta ne doit pas dépasser 20%.

1.1.3 Puissance et nombre de sujets nécessaire

La puissance d'essai dépend étroitement du nombre de sujets inclus, plus l'effectif des patients recrutés est important, plus grande est la puissance. Le calcul du nombre de sujets nécessaire pour garantir une puissance suffisante et donc préalable à toute prise de décision de réalisation. Il est importante en effet

de s'avoir, dès les premiers stades de la conception du protocole de l'essai, si les exigences de puissance imposent un effectif compatibles avec les possibilités réalistes de recrutement. Dans le cas contraire, la réalisation de l'essai est illusoire, c'est ainsi qu'un certain nombre de projets théoriquement pertinents doivent être abandonnés. Pour des raisons pratiques et même économiques, ils ne pourraient aboutir (**Bouvenot. G (2006)**).

Le calcul du nombre de sujets nécessaires, fondé sur le choix d'une puissance suffisante, répond aussi à des considérations éthiques (voir la section 1.1.1) : il ne serait pas raisonnables d'enrôler (et dans certain cas "d'exposer") plus de patients qu'il n'est besoin ou, à l'inverse, de conduire des études inutiles d'effectifs insuffisants pour permette la démonstration de l'activité d'un produit efficace. Une autre démarche consiste, dans des cas pathologies rares et orphelines, non plus à calculer un nombre de sujets nécessaire pour garantir une certaine puissance mais, étant donné des possibilités limitées de recrutement, à déterminer la puissance que garantirait ce recrutement et à décider ensuite de l'intérêt et de l'opportunité de montrer de telles études.

Deux remarques méritent encord d'être faites :

- La première concerne la répartition de l'effectif étotale des patients entre les deux groupes à comparer. Pour un effectif total donné, la puissance de la comparaison est maximale lorsque les groupes à comparer ont un effectif identique ou très voisin. Si l'on veut garantir une telle puissance pour la comparaison de deux groupes de taille différente, il faut un nombre globale de patients plus élevé ;
- La seconde est que le nombre de sujets nécessaire variant en sens inverse du carré de la taille de la différence escomptée entre les effets des deux traitement, il faut, pour une puissance donnée, davantage de patients pour démontrer l'efficacité (si elle existe) d'un produit peu différent de son comparateur que pour démontrer l'efficacité d'un produit très supérieur a son comparateur. C'est ainsi qu'on a besoin de moins de sujets pour démontrer l'efficacité versus placebo d'une substance active que pour mettre en évidence sa supériorité éventuelle sur un produit de référence validé : la taille de la différence attendue entre les effets des traitements comparés n'est pas la même dans les deux cas.

Comparaison de deux traitements normales avec une variance connue

Le test bilatéral est un test d'une hypothèse contre une alternative bilatéral. Nous limitons l'attention à déterminer une différence dans la réponse moyenne de deux traitement quand les observations sont gaussiennes avec une

variance commune connue.

$$\begin{cases} H_0 : \theta = 0 \\ H_a : \theta \neq 0 \end{cases}$$

H_a correspondent à un traitement supérieur ou inférieur à l'autre.

Dans cette comparaison, la statistique du test, Z est distribué symétriquement autour de 0 sous H_0 , et la règle de décision sera rejeter H_0 si $|Z| > c$ pour certaines constantes c . Le signe de Z détermine quel traitement doit être préféré quand H_0 est rejeté.

L'erreur du type I est défini pour être la probabilité de rejeter incorrectement l'hypothèse nulle,

$$P_{\theta=0}\{|Z| > c\} = \alpha.$$

La puissance d'un essai est la probabilité de rejeter l'hypothèse nulle quand elle ne se tient pas,

$$P_{\theta}\{|Z| > c\}$$

pour des valeurs de $\theta \neq 0$. La puissance dépend de θ . Elle augmente si θ éloigne de 0, mais il est comode d'énoncer une puissance conditionnelle à une valeur spécifique de θ , par exemple,

$$P_{\theta=\delta}\{|Z| > c\} = P_{\theta=-\delta}\{|Z| > c\} = 1 - \beta,$$

où δ représente une différence de traitement que les investigateurs espéraient détecter avec la probabilité élevée.

En pratique, il n'est pas souhaitable de rejeter H_0 en faveur de $\theta < 0$ quand réellement $\theta > 0$ car ceci impliquerait une recommandation du subordonné des deux traitements.

À cet effet, la puissance conditionnelle ci-dessus peut être remplacé par :

$$P_{\theta=\delta}\{Z > c\} = P_{\theta=-\delta}\{Z < -c\} = 1 - \beta,$$

Sois X_{Ai} et X_{Bi} , $i = 1, 2, \dots$ dénotent les réponses des sujets assignés à deux traitement A et B . Supposons que

$X_{Ai} \sim \mathcal{N}(\mu_A, \sigma^2)$ et $X_{Bi} \sim \mathcal{N}(\mu_B, \sigma^2)$, $i = 1, 2, \dots$, et toutes les informations sont indépendantes. Considérez le problème d'évaluer l'hypothèse nulle sans différence de traitement

$$H_0 : \mu_A = \mu_B \text{ contre } H_a : \mu_A \neq \mu_B$$

avec l'erreur du type I, α , et la puissance $1 - \beta$, $\mu_A - \mu_B = \pm\delta$

Si n sujets sont assignés à chaque traitement, la statistique du test sera

$$Z = \frac{1}{\sqrt{2n\sigma^2}} \left(\sum_{i=1}^n X_{Ai} - \sum_{i=1}^n X_{Bi} \right) \\ \sim \mathcal{N} \left((\mu_A - \mu_B) \sqrt{n/2\sigma^2}, 1 \right)$$

Ainsi sous H_0 , $Z \sim \mathcal{N}(0, 1)$, et rejetez H_0 si :

$$|Z| > \phi^{-1}(1 - \alpha/2)$$

où ϕ dénote la fonction de distribution cumulative normal.

Pour satisfaire la puissance conditionnelle, nous avons également

$$P\{|Z| > \phi^{-1}(1 - \alpha/2)\} = 1 - \beta$$

Quand $Z \sim \mathcal{N}(\pm\delta\sqrt{n/2\sigma^2}, 1)$.

Nous pouvons ignorer les probabilités très petites telles que

$$Z < -\phi^{-1}(1 - \alpha/2) \text{ quand } \mu_A = \mu_B = \delta$$

ou

$$Z > \phi^{-1}(1 - \alpha/2) \text{ quand } \mu_A = \mu_B = -\delta,$$

ainsi nous avons besoin de

$$E(Z) = \phi^{-1}(1 - \alpha/2) + \phi^{-1}(1 - \beta)$$

à l'alternative positive, $\mu_A - \mu_B = \delta$, et égalons cette valeur prévue à $\delta\sqrt{n/2\sigma^2}$, nous trouvons la taille de l'échantillon nécessaire

$$n_f(\alpha, \beta, \delta, \sigma^2) = \{\phi^{-1}(1 - \alpha/2) + \phi^{-1}(1 - \beta)\}^2 2\sigma^2 / \delta^2$$

sujet par bras de traitement. Dans la pratique n_f doit être arrondi pour avoir une taille d'échantillon de nombre entier.

Exemple 1.1.1. Supposons que $\sigma^2 = 4$, ainsi $X_{Ai} \sim \mathcal{N}(\mu_A, 4)$ et $X_{Bi} \sim \mathcal{N}(\mu_B, 4)$, $i = 1, 2, \dots$, et on exige pour examiner $H_0 : \mu_A = \mu_B$, avec $\alpha = 0.05$ et $1 - \beta = 0.9$ quand $\mu_A - \mu_B = \pm 1$.

La taille de l'échantillon nécessaire pour ce test est :

$$\{\phi^{-1}(0.975) + \phi^{-1}(0.9)\}^2 \times 2 \times 4 = 84.1$$

sujets pour chaque traitement, qu'on arrondi à 85. L'hypothèse nulle sera rejetée si

$$\left| \sum_{i=1}^{85} X_{Ai} - \sum_{i=1}^{85} X_{Bi} \right| \geq 1.96\sqrt{85 \times 2 \times 4} = 51.1.$$

calculer α et $1 - \beta$ de ce test :

$$\begin{aligned}
 1 - \beta &= P(\text{accepter } H_1 \mid H_1 \text{ est vraie}) \\
 &= P(|Z| > \phi^{-1}(1 - \alpha/2) \mid \mu_A - \mu_B = 1) \\
 &= P\left(Z' > \phi^{-1}(1 - \alpha/2) - \sqrt{n/2\sigma^2}\right) + P\left(Z' < -\phi^{-1}(1 - \alpha/2) - \sqrt{n/2\sigma^2}\right) \\
 &= P(Z' > -1.29) + P(Z' < -5.22) \\
 &= \phi(1.29) \\
 &= 0.9
 \end{aligned}$$

$$\begin{aligned}
 \alpha &= P(\text{rejetter } H_0 \mid H_0 \text{ est vraie}) \\
 &= P(|Z| > \phi^{-1}(1 - \alpha/2) \mid \mu_A - \mu_B = 0) \\
 &= P(Z > 1.96) + P(Z < -1.96) \\
 &= 2 - 2\phi(1.96) \\
 &= 0.05
 \end{aligned}$$

Nous reviendrons à cet exemple pour illustrer et comparer les différents types de tests séquentielles de groupes.

1.1.4 Règle d'arrêt

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

Principe de la règle d'arrêt

Si une suite d'expérience ζ_1, ζ_2, \dots admet une règle d'arrêt γ , qui indique quand les expériences doivent s'arrêter. L'inférence sur θ ne doit dépendre de γ qu'à travers l'échantillon résultant.

Règle d'arrêt précoce

D'après **Chow S. C et Chang M. (2007)**, il est désirable d'arrêter l'étude lorsque l'efficacité du médicament devient remarquable durant l'étude. Pour arrêter l'essai prématurément, nous limitons le nombre de sujet aléatoires. On utilise au moins une de ces règles :

1) **Règle pour utilité :**

L'écart du taux de succès entre le groupe d'étude et le groupe témoin dépasse un seuil, et l'extrémité inférieure correspondante à l'intervalle

de confiance de probabilité à 95% et inférieure à un certain seuil.

2) **Règle pour futilité :**

L'écart du taux de succès entre le groupe du traitement et le groupe témoin est inférieur à un seuil, et l'extrémité supérieure correspondante à l'intervalle de confiance de probabilité à 95% est inférieure à un seuil précis .

1.2 Analyses intermédiaires

Les analyses intermédiaires sont des analyses de données réalisées au 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 prévus (**Société canadienne du cancer (2007)**)[www.cancer.ca] .

1.2.1 Objectifs des analyses intermédiaires

Le but de ces analyses intermédiaires selon Société canadienne du cancer (2007) est :

- De 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 (arrêt pour efficacité). 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.

- Donner les moyens de détecter au plus tôt un éventuel effet délétère du traitement afin de limiter le nombre de patients exposés au risque (arrêt pour toxicité). 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 faible à la question posée.

- Arrêter une étude dont on peut prédire avec une certitude raisonnable qu'elle ne pourra pas aboutir (arrêt pour futilité). L'arrêt précoce permettra de diriger les ressources vers le test de nouvelles hypothèses.

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 1.2.1. *Dans un essai devant recruter 300 patients avec un suivi d'un mois, deux analyses intermédiaires sont réalisées : la première après l'inclusion des 100 premiers patients, la seconde avec 200 patients. L'analyse finale porte, comme prévu, sur 300 patients. Dans un essai de 1000 patients avec un suivi de 5 ans, une analyse intermédiaire est réalisée après recrutement de l'ensemble des patients, mais à mi-parcours, c'est-à-dire avec un recul de 2.5 ans.*

1.3 Test séquentiel de groupe

Les analyses séquentielles de groupe, permettent de répéter plusieurs fois, en cours des essais, et d'arrêter l'essai dès qu'il est possible de rejeter l'hypothèse nulle. L'analyse séquentielle de groupe est équivalente à la réalisation de nombreuses analyses intermédiaires. (C. Jennison et Bruce W. Turnbull (2000)).

À chaque nouvelle paire de patients (un patient traité avec le traitement étudié et un avec le traitement contrôle), on calcule une nouvelle valeur de

la statistique du test que l'on compare à une valeur maximale et une valeur minimale, si la valeur maximale est dépassée, on rejette l'hypothèse nulle. Si la valeur est inférieure à la valeur minimale, on rejette l'hypothèse alternative ce qui permet de conclure à l'équivalence à un delta près. Dans ces deux cas, l'essai s'arrête. Tant que la valeur de la statistique du test reste entre ces deux limites, on continue à recueillir des observations.

En statistiques, l'analyse séquentielle de groupe ou test d'hypothèse séquentiel est une analyse statistique où la taille de l'échantillon n'est pas fixée à l'avance. Au lieu de cela des données sont évaluées pendant qu'elles sont rassemblées, et davantage prélèvement est arrêté selon une règle d'arrêt pré-définie dès qu'on observera des résultats significatifs.

1.3.1 Concepts de base

Un test séquentiel est un test réalisé sur la base des données cumulées après l'obtention de chaque nouvelle observation (**Chow S. C, et Chang M.(2007)**). Les concepts de base sont :

L'inflation de l'erreur : Pour un essai, en une seule étape, conventionnelle unilatéral avec $\alpha = 0.025$, l'hypothèse nulle H_0 est rejetée si $Z \geq \phi^{-1}(1 - \alpha) = 1.96$. Pour un essai séquentiel avec K analyses, si à la k^{eme} analyse ($k = 1, 2, \dots, K$), $|Z_k|$ est suffisamment grande, nous rejeterons H_0 et arrêterons l'essai. Il n'est pas approprié d'appliquer simplement un essai unilatéral de niveau α à chaque analyse puisque les essais multiples mène- raient à une inflation de taux d'erreur du type I. En fait, le niveau réel α est indiqué par $1 - (1 - \alpha^k)$. Ainsi, pour $K = 5$, le niveau réel α est 0.071, presque trois fois plus grand que celle du 0.025 le niveau de signification appliqué à chaque analyse individuelle.

Frontière d'arrêt : Les frontières d'arrêt se composent d'un ensemble de valeurs critiques avec les quelles les statistiques de test calculées à partir des données réelles seront comparées pour déterminer si le test devrait être terminée ou continuée. Par exemple, la figure (1.2) fournit un ensemble de valeurs critiques comme frontière d'arrêt. En d'autres termes, si la moyenne d'échantillon observé à une étape donnée tombe en dehors des frontières, nous terminerons l'essai, autrement, l'essai continue.

L'échelles des frontières : Beaucoup de différentes échelles peut être employées pour construire les frontières d'arrêt. Les quatre échelles utili- sées généralement sont la Z-statistique normalisée, l'échelle de la moyenne d'échantillon, l'échelle d'erreur-dépense, et l'échelle somme- moyenne. En principe, les échelles sont équivalentes après transformation appropriée (**Chow S.C (2007)**). Comme exemple, considérons l'hypothèse pour examiner la dif- férence entre les moyennes de deux échantillons indépendants. Ces échelles

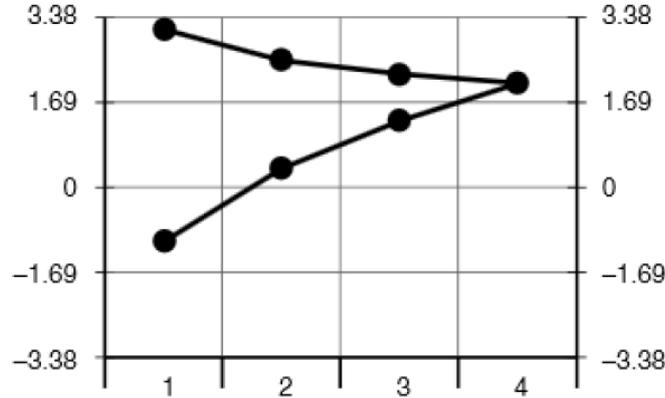


FIGURE 1.2 – Frontière d'arrêt.

sont définies comme suit :

- L'échelle de la moyenne : $\theta_k = \bar{x}_{A_k} - \bar{x}_{B_k}$
- Z-statistique normalisée : $Z_k = \theta_k \sqrt{I_k}$, où $I_k = n_k / \sigma^2$ est le niveau de l'information.
- Échelle de la somme -moyenne : $D_k = \sum_{i=1}^k x_{A_i} - \sum_{i=1}^k x_{B_i}$
- Échelle d'erreur-dépense : $\alpha(s_k)$, qui est également connu comme probabilité de balance.

Remarque 1.3.1. *Quand le nombre d'analyses intermédiaires augmente, il y a trop de conceptions possibles avec α et la puissance désirée. Par conséquent, dans la pratique, il est difficile de choisir une conception plus appropriée qui peut satisfaire notre besoin. Ainsi, on suggère qu'une fonction simple préférée pour définir les frontières d'arrêt soit employée. Une telle fonction avec quelques paramètres tels que les fonctions de frontière d'**O'Brien-Fleming** (qu'on verra dans la section 1.4.4), et autre que voir (Cleophas T.J, Zwinderman A. H, and Cleophas T. F. (2006)).*

1.3.2 Une formulation unifiée

C. Jennison (2000) a présenté une approche unifiée pour le design de test séquentielle de groupe. Cette approche unifiée est brièvement décrite ci-dessous :

On Considère une étude séquentielle de groupe jusqu'au K analyses donnant la suite des statistiques de test $\{Z_1, \dots, Z_K\}$.

Définition 1.3.1. *Nous disons que les statistiques $\{Z_1, \dots, Z_K\}$ ont la distribution commune canonique avec les niveaux d'informations I_1, \dots, I_K pour le paramètre θ si :*

$$\begin{cases} (Z_1, \dots, Z_K) \sim \mathcal{N}_K \\ E(Z_k) = \theta\sqrt{I_k}; \quad k = 1, \dots, K \\ \text{cov}(Z_{k_1}, Z_{k_2}) = \sqrt{I_{k_1}/I_{k_2}}; \quad 1 < k_1 \leq k_2 \leq K \end{cases} \quad (1.1)$$

Remarque 1.3.2. *La suite $\{Z_1, \dots, Z_K\}$ est une chaîne de Markov (**Brian S.E et Andrew P (1999)**). Cette propriété importante simplifie les calculs pour les tests séquentiels de groupe.*

1.4 Application dans les essais cliniques

Dans cette section, nous allons donner les tests séquentiels les plus utilisés dans les essais cliniques.

1.4.1 Test d'une moyenne pour une population simple

Soit les observations $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $i = 1, 2, \dots$, d'une population commun, tel que σ^2 est connu. Nous souhaitons évaluer :

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_a : \mu \neq \mu_0 \end{cases}$$

avec l'erreur du type I, α , et la puissance $1 - \beta$. Si les n_k observations sont disponibles à l'analyse k , nous estimons μ par :

$$\bar{X}^{(k)} = \frac{1}{n_k} \sum_{i=1}^{n_k} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n_k}\right)$$

et définissons par :

$$I_k = (\text{var}(\bar{X}^{(k)}))^{-1} = n_k/\sigma^2.$$

Les statistiques du test sont alors :

$$Z_k = (\bar{X}^{(k)} - \mu_0)\sqrt{I_k}, \quad k = 1, \dots, K$$

L'hypothèse nulle est rejetée si :

$$|Z_k| \geq z_{\alpha/2}$$

Puisque chaque Z_k est une combinaison linéaire des variables aléatoires normales indépendantes X_i , le vecteur (Z_1, \dots, Z_K) est normal multivarié.

Marginalement,

$$Z_k \sim \mathcal{N}(\theta\sqrt{I_k}, 1), k = 1, \dots, K$$

où $\theta = \mu - \mu_0$. Finalement, pour $k_1 \leq k_2$,

$$\begin{aligned} \text{cov}(Z_{k_1}, Z_{k_2}) &= \text{cov}\left(\{\bar{X}^{(k_1)} - \mu_0\}\sqrt{I_{k_1}}, \{\bar{X}^{(k_2)} - \mu_0\}\sqrt{I_{k_2}}\right) \\ &= \frac{1}{n_{k_1}} \frac{1}{n_{k_2}} n_{k_1} \sigma^2 \sqrt{I_{k_1}} \sqrt{I_{k_2}} \\ &= \sqrt{I_{k_1}/I_{k_2}} \end{aligned}$$

et nous voyons que $\{Z_1, \dots, Z_K\}$ suit la distribution commune canonique (1.1) avec les niveaux d'informations $\{I_1, \dots, I_K\}$, pour $\theta = \mu - \mu_0$.

1.4.2 Comparaison de deux traitements parallèles

Dans cette section, nous permettons une différence de variances et de nombres des sujets sur les deux bras de traitement. Ainsi nous avons des réponses $X_{A_i} \sim \mathcal{N}(\mu_A, \sigma_A^2)$, $i = 1, 2, \dots$ au traitement A et $X_{B_i} \sim \mathcal{N}(\mu_B, \sigma_B^2)$, $i = 1, 2, \dots$ au traitement B . Nous souhaitons évaluer :

$$\begin{cases} H_0 : \mu_A = \mu_B \\ H_a : \mu_A \neq \mu_B \end{cases}$$

avec l'erreur du type I, α , et la puissance $1 - \beta$.

Pour $k = 1, \dots, K$, soient n_{A_k} et n_{B_k} le nombre cumulatif d'observation sur les traitements A et B , respectivement, à l'heure de la k^{eme} analyse l'estimateur naturel de $\mu_A - \mu_B$ est :

$$\begin{aligned} \bar{X}_A^{(k)} - \bar{X}_B^{(k)} &= \frac{1}{n_{A_k}} \sum_{i=1}^{n_{A_k}} X_{A_i} - \frac{1}{n_{B_k}} \sum_{i=1}^{n_{B_k}} X_{B_i} \\ &\sim \mathcal{N}\left(\mu_A - \mu_B, \frac{\sigma_A^2}{n_{A_k}} + \frac{\sigma_B^2}{n_{B_k}}\right) \end{aligned}$$

Notons que la distribution de $\overline{X}_A^{(k)} - \overline{X}_B^{(k)}$ est une distribution marginale sachant n_{A_k} et n_{B_k} données.

Nous définissons l'information de $\mu_A - \mu_B$ par :

$$I_k = \left(\frac{\sigma_A^2}{n_{A_k}} + \frac{\sigma_B^2}{n_{B_k}} \right)^{-1}$$

employant ceci pour créer la statistique normalisée a l'analyse k pour tester $H_0 : \mu_A = \mu_B$, comme suit :

$$Z_k = (\overline{X}_A^{(k)} - \overline{X}_B^{(k)})\sqrt{I_k}, \quad k = 1, \dots, K$$

L'hypothèse nulle est rejetée si :

$$|Z_k| \geq z_{\alpha/2}$$

le vecteur (Z_1, \dots, Z_K) est normal multivarié puisque chaque Z_k est une combinaison linéaire de variables aléatoires normales indépendantes, X_{A_i} et X_{B_i} , $i = 1, 2, \dots$, et marginalement

$$Z_k \sim \mathcal{N}(\theta\sqrt{I_k}, 1), \quad k = 1, \dots, K$$

où $\theta = \mu_A - \mu_B$. La covariance de Z_{k_1} pour $k_1 \leq k_2$, est donnée par :

$$\begin{aligned} \text{cov}(Z_{k_1}, Z_{k_2}) &= \text{cov} \left(\{\overline{X}_A^{(k_1)} - \overline{X}_B^{(k_1)}\}\sqrt{I_{k_1}}, \{\overline{X}_A^{(k_2)} - \overline{X}_B^{(k_2)}\}\sqrt{I_{k_2}} \right) \\ &= \left(\frac{1}{n_{A_{k_1}}} \frac{1}{n_{A_{k_2}}} n_{A_{k_1}} \sigma_A^2 + \frac{1}{n_{B_{k_1}}} \frac{1}{n_{B_{k_2}}} n_{B_{k_1}} \sigma_B^2 \right) \sqrt{I_{k_1}} \sqrt{I_{k_2}} \\ &= (I_{k_2})^{-1} \sqrt{I_{k_1}} \sqrt{I_{k_2}} \\ &= \sqrt{I_{k_1}/I_{k_2}} \end{aligned}$$

Ainsi $\{Z_1, \dots, Z_K\}$ ayant la distribution commune canonique avec les niveaux d'infarmations $\{I_1, \dots, I_K\}$, pour $\theta = \mu_A - \mu_B$.

1.4.3 Comparaison de deux traitements appaillés

Dans une comparaison de deux traitements, il peut être avantageux de comander la variance dans la réponse attribuable aux facteurs pronostiques, connus en employant un design de paires assorties. Des sujets sont appareillés de sort que les deux sujet dans la même paire aient les valeurs semblables des facteurs pronostiques, un sujet dans chaque paire qui est aléatoirement choisi recoit le traitement A et l'autre recoit le traitement B . Soit X_{A_i} et

X_{B_i} les réponses des sujets pour les traitements de réception A et B dans la paire i , respectivement.

Nous supposons que les différences dans les paires sont normalement distribuées,

$$X_{A_i} - X_{B_i} \sim \mathcal{N}(\mu_A - \mu_B, \tilde{\sigma}^2), i = 1, 2, \dots$$

où la variance $\tilde{\sigma}^2$ est connu. Si la variance d'une observation individuelle est σ^2 et la corrélation entre les réponses des sujets dans des paires est ρ , alors $\tilde{\sigma}^2 = 2(1 - \rho)\sigma^2$. Ainsi, si l'assortiment des sujets dans les paires réalise une corrélation positive modérément grande, $\tilde{\sigma}^2$ peut être de manière significative moins que $2\sigma^2$, de la différence dans la réponse entre deux sujets aléatoirement choisis assignés aux traitement A et B .

Nous considérons le problème d'évaluer l'hypothèse nulle $H_0 : \mu_A = \mu_B$ dans un essai séquentiel de groupe où des observations sont prises en K groupes. Si les paires de n_k d'observations sont disponibles dans la k^{eme} analyse, nous estimons $\theta = \mu_A - \mu_B$ par :

$$\frac{1}{n_k} \sum_{i=1}^{n_k} (X_{A_i} - X_{B_i}) \sim \mathcal{N}(\mu_A - \mu_B, \frac{\tilde{\sigma}^2}{n_k})$$

l'information pour θ est $I_k = \frac{n_k}{\tilde{\sigma}^2}$, la réciproque de la variance de l'évaluation, et la statistique du test sera :

$$Z_k = \frac{1}{\sqrt{n_k \tilde{\sigma}^2}} \sum_{i=1}^{n_k} (X_{A_i} - X_{B_i}), \quad k = 1, \dots, K \quad (1.2)$$

La probabilité de rejeter l'hypothèse nulle est :

$$P_{\mu_A - \mu_B = 0} \{|Z_k| > c\} = \alpha_k$$

et la puissance est :

$$P_{\mu_A - \mu_B = \theta} \{|Z_k| > c\} = 1 - \beta_k$$

Comme un exemple d'application des tests séquentiels de groupes, on choisit le test d'**O'Brien et Fleming**, pour comparer deux traitements appariés (**C. Jennison(2000)**).

1.4.4 Test d'O'Brien et Fleming

O'Brien et Fleming(1979) ont proposé un test dans lequel les niveaux de signification nominale nécessaires pour rejeter H_0 à chaque analyse augmentent à mesure que l'étude progresse, et c'est une alternative au test

de signification répétée de **Pocook(1977)**. L'entrée patiente est divisée en groupes également classés de K contenant de m sujets sur chaque traitement, et les données sont analysées après qu'on ait observé chaque nouveau groupe d'observations.

Ainsi, dans leur test il est plus difficile de rejeter $H_0 : \mu_A - \mu_B = 0$ aux premières analyses mais plus facile plus tard. Dans ce test les sujets sont attribués, aux traitements en groupes et les données accumulées sont analysées après la disponibilité de groupes de données complets.

Le test d'**O'Brien et Fleming** est également défini en termes de statistique normale Z_k .

$$Z_k = \frac{1}{\sqrt{2mk\sigma^2}} \left(\sum_{i=1}^{mk} X_{Ai} - \sum_{i=1}^{mk} X_{Bi} \right) \quad k = 1, \dots, K$$

et H_0 est rejeté après le groupe k si $|Z_k| \geq c_k$ pour un ordre des valeurs critiques c_1, \dots, c_K Formalement, le test est réalisé comme suit :

- Après le k^{eme} groupe ($k = 1, \dots, K - 1$) :
si $|Z_k| \geq C_B(K, \alpha)\sqrt{K/k}$ arrêter et rejeter H_0 ,
autrement continue au groupe ($k + 1$).
- Après le K^{eme} groupe :
si $|Z_K| \geq C_B(K, \alpha)$ arrêter et rejeter H_0 ,
autrement arrêter et accepter H_0 .

Ainsi, $c_k = C_B(K, \alpha)\sqrt{K/k}$, $k = 1, \dots, K$.

Les valeurs $C_B(K, \alpha)$ sont fournies dans le tableau(1.1). En termes de niveaux de signification nominale, H_0 est rejeté à l'analyse k , $k = 1, \dots, K$, si le niveau de signification bilatéral de H_0 est :

$$\alpha'_k = 2[1 - \phi\{C_B(K, \alpha)\sqrt{K/k}\}]$$

Remarque 1.4.1.

1) Puisque $C_B(K, \alpha)\sqrt{K/k}$ déminue avec l'augmentation de k , alors α'_k croît. Pour $K = 5$ et $\alpha = 0.05$, les niveaux de signification nominale appliqués aux analyses de 1 à 5 sont :

$$\begin{aligned} \alpha'_1 &= 2(1 - \phi(C_B(5, 0.05)\sqrt{5/1})) \\ &= 2(1 - \phi(2.04\sqrt{5})) \\ &= 2(1 - \phi(4.56)) \\ &= 2(1 - 0.9999975) \\ &= 0.000005 \end{aligned}$$

K	$C_B(K, \alpha)$		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
1	2.576	1.960	1.645
2	2.580	1.977	1.678
3	2.595	2.004	1.710
4	2.609	2.024	1.733
5	2.621	2.040	1.751
6	2.631	2.053	1.765
7	2.640	2.063	1.776
8	2.648	2.072	1.786
9	2.654	2.080	1.794
10	2.660	2.087	1.801
11	2.665	2.092	1.807
12	2.670	2.098	1.813
15	2.681	2.110	1.826
20	2.695	2.126	1.842

TABLE 1.1 – Test d’O’Brien et de Fleming : constantes $C_B(K, \alpha)$ de test bilatéral avec des groupes de K d’observations, l’erreur du type I, (α).

$$\begin{aligned}
\alpha'_2 &= 2(1 - \phi(2.040\sqrt{5/2})) \\
&= 2(1 - \phi(3.22)) \\
&= 2(1 - 0.99931) \\
&= 0.0013
\end{aligned}$$

$$\alpha'_3 = 0.0084, \quad \alpha'_4 = 0.0225, \quad \alpha'_5 = 0.0413.$$

2) Puisque les $c_K = C_B(K, \alpha)$, comparer les $C_B(K, \alpha)$ aux $C_B(1, \alpha)$, montre combien la valeur de $|Z_k|$ soit grande par rapport à valeur critique qui serait appliquée dans l’analyse des données final. Il est évident que cette augmentation, soit tout à fait petite pour les tests d’**O’Brien et Fleming** (voir le tableau 1.1).

Le tableau (1.2) contient $R_B(K, \alpha, \beta)$, pour l’usage dans le calculs des tailles de l’échantillons. La taille de l’échantillon maximum sur chaque bras de traitement est $R_B(K, \alpha, \beta)n_f(\alpha, \beta, \delta, \sigma^2)$ où

$$n_f(\alpha, \beta, \delta, \sigma^2) = \{\phi^{-1}(1 - \alpha/2) + \phi^{-1}(1 - \beta)\}^2 2\sigma^2 / \delta^2$$

est le nombre d’observations par traitement requis pour un test de taille d’échantillon fixé.

Ainsi, la taille nécessaire de groupe est :

$$m = R_B(K, \alpha, \beta)n_f(\alpha, \beta, \delta, \sigma^2) / K$$

K	$R_B(K, \alpha, \beta)$					
	$1 - \beta = 0.8$			$1 - \beta = 0.9$		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
1	1.000	1.000	1.000	1.000	1.000	1.000
2	1.001	1.008	1.016	1.001	1.007	1.014
3	1.007	1.017	1.027	1.006	1.016	1.025
4	1.011	1.024	1.035	1.010	1.022	1.032
5	1.015	1.028	1.040	1.014	1.026	1.037
6	1.017	1.032	1.044	1.016	1.030	1.041
7	1.019	1.035	1.047	1.018	1.032	1.044
8	1.021	1.037	1.049	1.020	1.034	1.046
9	1.022	1.038	1.051	1.021	1.036	1.048
10	1.024	1.040	1.053	1.022	1.037	1.049
11	1.025	1.041	1.054	1.023	1.039	1.051
12	1.026	1.042	1.055	1.024	1.040	1.052
15	1.028	1.045	1.058	1.026	1.042	1.054
20	1.030	1.047	1.061	1.029	1.045	1.057

TABLE 1.2 – Test d’O’Brien et de Fleming : constantes $R_B(K, \alpha, \beta)$ pour déterminer des tailles de groupe de test bilatéral avec des groupes de K d’observations, l’erreur du type I, α , et de puissance $1 - \beta$.

par bras de traitement, convonablement arrondi.

Exemple 1.4.1. Nous examinons $H_0 : \mu_A = \mu_B$, avec $\alpha = 0.05$, $1 - \beta = 0.9$, $\mu_A - \mu_B = \pm 1$, $\sigma^2 = 4$, et la taille de l’échantillon fixe exigée est $n_f(0.05, 0.1, 1, 4) = 84.1$, $R_B(5, 0.05, 0.1) = 1.026$, par conséquent, un essai d’**O’Brien et Fleming** avec 5 groupes d’observations a nécessite une taille d’échantillon maximal par bras de traitement qui vaut :

$$R_B(5, 0.05, 0.1)n_f(0.05, 0.1, 1, 4) = 1.026 \times 84.1 = 86.3$$

et la taille nécessaire par groupe est $m = 86.3/5 = 17.3$ par traitement, qui sera arrondi jusqu’à 18. L’essai cesse de rejeter H_0 à l’analyse k , $k = 1, \dots, 5$ si :

$$\left| \sum_{i=1}^{18k} X_{Ai} - \sum_{i=1}^{18k} X_{Bi} \right| \geq 2.040 \sqrt{5/k} \sqrt{18k \times 2 \times 4} = 54.74$$

et accepter H_0 s’il n’est pas rejeté par l’analyse 5.

Tout le nombre maximum des sujets pour les deux bras de traitement qui peuvent être exigés par **O’Brien et Fleming** est 180, une petite augmentation au-dessous des 170 requise pour un essai témoin fixe. Le nombre prévu de sujets est réduit sensiblement quand $|\mu_A - \mu_B|$ est suffisamment grand.

La frontière de l’essai d’**O’Brien et Fleming** pour cet exemple est illustrée sur la figure suivante :

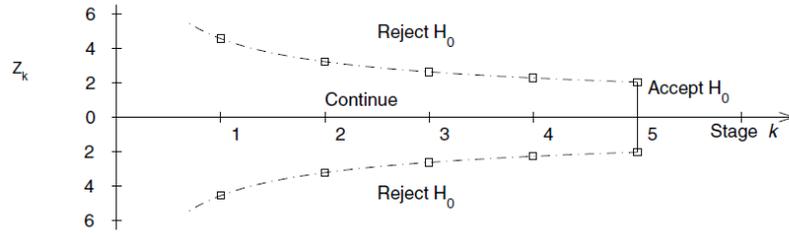


FIGURE 1.3 – Test d’O’Brien et de Fleming pour 5 groupes d’observations.

Dans la figure(1.3), où l’essai est décrit en terme de valeurs critiques pour Z_k , les valeurs diminuent nettement aux analyses successives.

Exemple 1.4.2. (Comparaison appariés) Considérons la comparaison entre deux traitements appariés. Les paires d’observations X_{A_i} et X_{B_i} rapportent des différences de moyenne $\mu_A - \mu_B$. Nous définissons $|\mu_A - \mu_B| = 1$ représentant une différence médicalement importante de traitement et nous décidons d’examiner $H_0 : \theta = 0$ avec $\alpha = 0.05$ et puissance $1 - \beta = 0.9$ pour $\theta = \pm 1$

En utilisant **O’Brien et Fleming** avec 5 groupes d’observations.

Les statistiques normalisées du test sont données par(1.2) et les niveaux d’information sont $I_k = n_k/\tilde{\sigma}^2, k = 1, \dots, K$ où n_k est le nombre de paires d’observations dans les premiers k groupes et $\tilde{\sigma}^2 = \text{var}(X_{A_i} - X_{B_i})$. Supposons qu’on sait que $\tilde{\sigma}^2 = 6$. L’équation :

$$I_{f,2} = \{\phi^{-1}(1 - \alpha/2) + \phi^{-1}(1 - \beta)\}^2/\delta^2$$

donne $I_{f,2} = (1.960 + 1.282)^2/1^2 = 10.51$ et prenant $R = R_B(5, 0.05, 0.1) = 1.026$ du tableau (1.2) nous trouvons les niveaux désirés de l’information qui sont $I_k = n_k/6 = (k/5)1.026 \times 10.51, k = 1, \dots, 5$. Ceci exige $n_k = 12.94k$, c.à.d, 12.9 paires d’observation par groupe, qui se rond vers 13. Les valeurs cliniques pour un essai d’OBrien et Fleming sont $C_B(5, 0.05) = 2.040$ du tableau (1.1).

Ainsi, aux analyses $k = 1, \dots, 5$, l’essai cesse de rejeter $H_0 : \theta = 0$ si :

$$\left| \frac{1}{\sqrt{13k \times 6}} \sum (X_{A_i} - X_{B_i}) \right| \geq 2.040 \sqrt{\frac{5}{k}},$$

c.à.d, si $|\sum (X_{A_i} - X_{B_i})| \geq 40.29$, et H_0 est accepté s’il n’a pas été rejeté par la 5^{eme} analyse.

Chapitre 2

Tests séquentiels Bayésiens de groupe

L'approche Bayésienne aux essais cliniques, et de grand intérêt pour la communauté médicale de développement d'un traitement, parcequ'elle offre une manière simple et élégante d'obtenir l'information valide en quelques sorte, qui est potentiellement de diminuer le temps.

En paradigme Bayésienne, la croyance initiale au sujet d'un paramètre d'intérêt (discret ou continue) est exprimée par une distribution a priori. L'évidence des données est alors modélisée par une fonction de probabilité pour le paramètre.

En particulier, les conceptions séquentielles Bayésiennes de groupes peut être considérées comme une combinaison naturelle de la prédiction et la prise de décision.

Dans ce chapitre nous discutons le perspective Bayésien en indiquant le contre exemple de **Freedman (1994)**. Nous expliciterons aussi le raisonnement des règles d'arrêt Bayésiennes dans les essais cliniques avec le choix de la distribution a priori.

2.1 Le perspective Bayésien

N'importe quelle discussion des procédures séquentielles de groupes serait inachevée sans reconnaître que quelques statisticiens considèrent l'approche séquentielle est presque fatalement déféctueuse.

Freedmain et al. (1994), trouvaient des raisons de s'inquiéter dans plusieurs domaines :

- Maladresse philosophique .
- Comment dessiner des inférences si la règle d'arrêt n'est pas suivie.

- Comment estimer des effets de traitement à la fin d'un essai séquentielle de groupe.

Freedmain et al. (1994) illustrent leurs difficultés philosophiques avec les approches habituelles frequentistes aux analyses intermédiaires par la situation (hypothétique) suivante :

Supposons qu 'un clinicien Dr.C vient au statisticien, Dr S.

C a conduit un essai clinique d'un nouveau traitement pour le **SIDA**. Il a traité et évalué 200 patients, et l'analyse des résultats suggère un avantage du nouveau traitement qui est statistiquement significatif (p -valeur = 0.02).

Dr . S, qui est également un frequentist, demande combien de fois Dr.C projette analyser les résultats pendant que l'essai progresse. Nous considérons maintenant deux réponses possibles .

- 1) Dr. C peut répondre que c'est la première et seulement l'analyse qui a fait et elle est suffisante. Il a attendu jusqu' à ce que les données soient complètes avant de les analyser, Dr .S approuve l'analyse et la p -valeur.
- 2) Dr. C peut au lieu de celà, répondre qu 'il a l'intention d'inclure 1000 patients dans cette essai , et que c'est la première de 5 analyses projetées .

D'ailleurs le statisticien, qui a aidé à concevoir l'essai, avait préconisé une frontière d'**O'Brien et de Fleming**. Dr. S conseille alors Dr . C que les résultats ne sont pas encore statistiquement significatifs et pour celà la p -valeur nominale (p -valeur=0.02) ne peut pas être prise puisque ce n'est qu'une seule valeur parmi une série d'essais qui sont planifiés.

Freedmain et al. (1994) se plaignent qu'il semble peu raisonnable que différentes inférences devraient être faites par Dr .S dépendant du plan pour d'avantage d'analyse .

Ils suggèrent alors une approche Bayésienne alternative pour améliorer les résultats de l'approche séquentielle impliquait les étapes suivantes :

- Une distribution a priori, représentant sa croyance au sujet de la différence de traitement est indiquée. En fait deux a priori sont recommandées :

La première pour représenter un sceptique raisonnable, la seconde pour représenter un fervent raisonnable .

- Des données sont alors recueillies pendant l'essai menant à une évaluation de la différence de traitement avec un intervalle de confiance .

- Le théorème de Bayes est alors appliqué pour calculer une distribution a posteriori qui représente sa croyance courante au sujet de la différence des traitements .

• Des recommandations concernant la suite de l'essai sont basées sur la distribution a posteriori . La balance de différence de traitement est divisée en trois gammes :

- 1) Différences qui mèneraient à un choix du traitement standard .
- 2) Différences qui mèneraient à un choix du nouveau traitement.
- 3) Une gamme intermédiaire dans laquelle les avantages du nouveau traitement sont compensés par la toxicité accrue, des inconvénients ou des coûts .

Les probabilités a posterioris de l'effet de traitement de chacune de ces trois régions peuvent être employées pour prendre des décisions au sujet du futur de l'essai .

Freedman et al. (1994) donnaient un exemple d'une telle approche dans un essai clinique étudiant l'effet de la combinaison de traitement de S-flurouracil et de le vamisole sur la longueur de la survie des patients avec le cancer colorectal .

Bien que les propriétés de l'approche Bayésienne soient attrayantes en fournissant une vue intégrée de tous les aspects de l'arrêt , jusqu'ici elle n'a pas fait un impact important sur la surveillance des essais cliniques . **Machin (1994)** croyait que ceci puisse être partiellement dû au scepticisme de cliniciens concernant la manière dont les différentes a priori prétendant représenter la croyance, peuvent influencer l'interprétation des résultats .

2.2 Paradigme Bayésien

Le paradigme Bayésien fournit un cadre très différent d'inférence, pour la conception et l'analyse des essais séquentiels de groupes. En ce paradigme l'incertitude environ θ (le paramètre ou les paramètres d'intérêt) est exprimée par une distribution de probabilité, appelée la distribution a priori de densité $\pi(\theta)$ en considérant θ comme une v. a .

Définition 2.2.1 (la loi a priori).

On entend par information a priori sur le paramètre θ toute information disponible sur θ , en dehors de celle apportée par les observations. L'information a priori est entachée d'incertitude (si ce n'était pas le cas, le paramètre serait connu avec certitude et on n'aurait pas à l'estimer). Il est naturel de modéliser cette information a priori à travers une loi de probabilité appelée loi a priori, notée $\pi(\theta)$.

Pendant que les données s'accumulent, cette a priori est mis à jour pour former la distribution a postérieure qui résume l'incertitude courante au sujet

de la valeur θ .

Définition 2.2.2 (la loi a posteriori).

C'est la loi conditionnelle de θ sachant x sa densité notée $\pi(\theta | x)$, appelée distribution a posteriori et est définie par :

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

Et on a aussi :

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

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

b) *La loi marginale de x : sa densité est notée $m(x)$,*

$$m(x) = \int f(\theta | x)\pi(\theta)d\theta$$

Exemple 2.2.1. *Une boule de billard W roule sur une ligne de longueur 1, avec une probabilité uniforme de s'arrêter n'importe où.*

Supposons qu'elle s'arrête en θ . 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 θ ?

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

$$\begin{aligned} P(X = x | \theta) &= C_n^x \theta^x (1 - \theta)^{n-x} \\ P(a < \theta < b, X = x) &= \int_a^b C_n^x \theta^x (1 - \theta)^{n-x} d\theta \\ \text{et } P(X = x) &= \int_0^1 C_n^x \theta^x (1 - \theta)^{n-x} d\theta \end{aligned}$$

Nous trouvons que

$$\begin{aligned} P(a < \theta < b \mid X = x) &= \frac{\int_a^b C_n^x \theta^x (1 - \theta)^{n-x} d\theta}{\int_0^1 C_n^x \theta^x (1 - \theta)^{n-x} d\theta} \\ &= \frac{\int_a^b \theta^x (1 - \theta)^{n-x} d\theta}{\mathcal{B}(x + 1, n - x + 1)} \end{aligned}$$

Donc la distribution de θ conditionnellement à $X = x$ est une distribution Bêta, $\mathcal{B}(x + 1, n - x + 1)$. Si $x \sim \mathcal{B}(n, \theta)$ et $\theta \sim \mathcal{Beta}(\alpha, \beta)$ (avec $\alpha = \beta = 1$ dans le cas particulier de Bayes)

$$\begin{aligned} f(x \mid \theta) &= C_n^x \theta^x (1 - \theta)^{n-x}, x = 0, 1, \dots, n \\ \pi(\theta) &= \frac{1}{\mathcal{B}(\alpha, \beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}, 0 \leq \theta \leq 1 \end{aligned}$$

la densité de la distribution jointe de (θ, x) est alors

$$\varphi(\theta, x) = \frac{C_n^x}{\mathcal{B}(\alpha, \beta)} \theta^{\alpha+x-1} (1 - \theta)^{n-x+\beta-1}$$

et densité de la distribution marginale de x est

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

donc la densité de la distribution a posteriori de θ est

$$\pi(\theta \mid x) = \frac{\theta^{\alpha+x-1} (1 - \theta)^{\beta+n-x-1}}{\mathcal{B}(\alpha + x, \beta + n - x)},$$

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

Exemple 2.2.2. Soit $X = (x_1, x_2, \dots, x_n)$ un échantillon aléatoire de taille n , de la loi $\mathcal{N}(\theta, \sigma^2)$ où σ^2 est connu, et la densité de X est donnée par :

$$\begin{aligned} f(X \mid \theta) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left[- \sum_{i=1}^n (x_i - \theta)^2 / 2\sigma^2 \right] \\ &\propto \exp \left[\frac{-1}{2} \theta^2 \left(\frac{n}{\sigma^2} \right) + \theta \bar{x} \left(\frac{n}{\sigma^2} \right) \right] \\ &\propto \exp \left[\frac{-1}{2} (\bar{x} - \theta)^2 \left(\frac{n}{\sigma^2} \right) \right] \end{aligned}$$

Donc $X | \theta \sim \mathcal{N}(\theta, \sigma^2/n)$. On choisit l'a priori pour θ qui est la loi $\mathcal{N}(\mu_0, \sigma_0^2)$.

Puisque on a une a priori conjuguée (voir Jennison et Turnbull. (2000)), la distribution a posteriori de θ est donnée par :

$$\theta | X \sim \mathcal{N}(\mu_n, \sigma_n^2)$$

Où

$$\begin{aligned} \mu_n &= \rho\left(\frac{\sigma^2}{n}\mu_0 + \sigma_0^2\bar{x}\right), \quad \rho = \frac{1}{\frac{\sigma^2}{n} + \sigma_0^2} \\ &= \frac{\frac{\sigma^2}{n}\mu_0 + \sigma_0^2\bar{x}}{\frac{\sigma^2}{n} + \sigma_0^2} \\ &= \frac{\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \end{aligned}$$

et

$$\begin{aligned} \sigma_n^2 &= \rho \frac{\sigma^2}{n} \sigma_0^2 \\ &= \frac{\frac{\sigma^2}{n} \sigma_0^2}{\frac{\sigma^2}{n} + \sigma_0^2} \\ \text{Alors } \sigma_n^{-2} &= n\sigma^{-2} + \sigma_0^{-2} \end{aligned}$$

$$\text{et } \bar{X} = n^{-1} \sum_{i=1}^n x_i .$$

Remarque 2.2.1. À n'importe quelle étape, cette distribution a posteriori peut être employée pour dessiner des inférences au sujet de θ . Par exemple nous pouvons construire des intervalles crédibles pour θ avec une probabilité a posteriori égale à un certain niveau présécifié $1 - 2\epsilon$ (**C. Jennison et Bruce W. Turnbull (2000)**). Si θ est unidimensionnel et $\epsilon = 0.025$ l'intervalle est :

$$P(\theta_L < \theta < \theta_u | \text{donnes}) = 0.95 \quad .$$

2.2.1 Utilisation séquentielles du théorème de Bayes

Supposons que nous observons des données dans deux étapes ou plus, disons y_m suivi du y_n . Alors après qu'on observe le premier échantillon la loi

a posteriori sera :

$$\pi(\theta | y_m) \propto f(y_m | \theta)\pi(\theta) \quad (2.1)$$

Cette a posteriori devient la distribution a priori pour la prochaine utilisation du théorème de Bayes, ainsi après qu'on observe la prochaine y_n du 2^{ème} échantillon, la distribution a posteriori conditionnée par toutes les données, c.à.d. $\pi(\theta | y_n, y_m)$ est donnée par :

$$\pi(\theta | y_n, y_m) \propto f(y_n | \theta, y_m)\pi(\theta | y_m) \quad (2.2)$$

Combinons les deux expressions (2.1) et (2.2) nous aurons :

$$\pi(\theta | y_n, y_m) \propto f(y_n | \theta, y_m)f(y_m | \theta)\pi(\theta)$$

Ceci peut également être dérivé en considérant une utilisation unique du théorème de Bayes de données y_n, y_m en mettant en facteur la vraisemblance comme suit :

$$f(y_n, y_m | \theta) = f(y_n | \theta, y_m)\pi(\theta | y_m)$$

Dans la plupart des situations le premier terme dans (2.2) ne dépendra pas de y_m (c-à-d y_n est indépendante de y_m sachant θ), et ainsi $\pi(\theta | y_m)$ devient simplement une a priori pour une mise à jour bayésienne standard en utilisant la vraisemblance $f(y_n | \theta)$.

2.2.2 Prédiction

La réalisation de prédictions est l'un des objectifs fondamentaux de la modélisation statistique, et une approche Bayésienne peut rendre cette tâche raisonnablement directe. Supposons que nous souhaitons prédire certaines observations futures x sur la base des données actuellement observées y .

Définition 2.2.3 (La densité prédictive). *La fonction de densité prédictive $f(x | y)$ est définie par :*

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

Prédiction des données normales :

Les prédictions sont particulièrement facile pour les distributions normales. Par exemple supposons que $Y_n \sim \mathcal{N}(\theta, \sigma^2/n)$ pour quelques données futures Y_n , et la loi a priori $\theta \sim \mathcal{N}(\mu, \sigma^2/n_0)$.

Nous souhaitons faire des prédictions au sujet des valeurs futures Y_n , tenir compte de notre incertitude sur la moyenne θ . Nous pouvons écrire $Y_n = (Y_n - \theta) + \theta$, Y_n est la somme de deux quantités indépendantes normales : $Y_n - \theta \sim \mathcal{N}(0, \sigma^2/n)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2/n_0)$.

On sait que la somme de deux quantités indépendantes normales est normale avec la somme de moyennes et de variances, et par conséquent Y_n aurons une distribution prédictive comme suit :

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

En utilisant les expression pour la mayenne et la variance (**David J. (2004)**(section(2.2.2)))

$$E(Y_n) = E_\theta[E(Y_n|\theta)] = E_\theta[\theta] = \mu,$$

$$V(Y_n) = V_\theta[E(Y_n|\theta)] + E_\theta[V(Y_n|\theta)] = V_\theta(\theta) + E_\theta(\sigma^2/n) = \sigma^2\left(\frac{1}{n} + \frac{1}{n_0}\right).$$

Supposons maintenant que nous avons déjà observé des données Y_m et par conséquent notre distribution est :

$$\theta|y_m \sim \mathcal{N}\left(\frac{n_0\mu + my_m}{n_0 + m}, \frac{\sigma^2}{n_0 + m}\right)$$

$$Y_n|y_m \sim \mathcal{N}\left(\frac{n_0\mu + my_m}{n_0 + m}, \sigma^2\left(\frac{1}{n_0 + m} + \frac{1}{n}\right)\right)$$

L'utilisation de cette expression est illustrée dans l'exemple ci-dessous.

Exemple 2.2.3. *L'essai **GREAT** (du traitement précoce pour l'infarctus du myocarde, fournit les données dans le tableau (2.1)).*

Event	traitement Nouveau	traitement contrôle	le nombre d'evenements
mort	13	23	36
pas de mort	150	125	275
	163	148	311

TABLE 2.1 – GREAT : normal de probabilité

*Supposons que nous voulons prolonger l'essai **GREAT**(**David J. (2004)**(section (2.4.1))), pour inclure 100 patients plus encore sur chaque bras. Quelle*

sera la valeur prédite pour odds-ratio (OR) observé pour les patients de future échantillon, avec et sans utiliser les information a priori de l'essai ?

Il est important de rappeler que la prédiction de l'estimateur de OR dépend de nombre des événements (des morts) observés seulement.

L'évaluation du maximum de vraisemblance de la chance de la mort sous la nouvelle interposition est $13/150$ (Le nombre de mort par le nombre de survivants), Sous le contrôle est $23/125$. Et la rapport de odds ratio $OR = \frac{13/150}{23/125}$. En prenant $\theta = \log(OR)$, on peut l'estimée par :

$$\hat{\theta} = \log \left[\frac{(13 + \frac{1}{2})(125 + \frac{1}{2})}{(23 + \frac{1}{2})(150 + \frac{1}{2})} \right]$$

donne un (OR) estimé $y_m = -0.736$, avec,

$$\begin{aligned} V(\hat{\theta}) &= \frac{1}{13 + 1/2} + \frac{1}{23 + 1/2} + \frac{1}{150 + 1/2} + \frac{1}{125 + 1/2} \\ &= 0.131 \\ &= 0.362^2. \end{aligned}$$

(**David (2004)**(section(2.4.1))).

Supposons que $\theta = \log(OR)$ observé dans ces futurs patients est $Y_n \sim \mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right)$, et $\theta \sim \mathcal{N}(-0.31, 4/236.7)$ où le futur nombre d'événement est n et $\sigma = 2$: avec 100 patients dans chaque bras nous pouvons nous attendre a $n \simeq 20$ événements, sachant que le taux courante de mortalité est environ de 10%. La distribution a postérieure courante est $\theta|Y_m \sim \mathcal{N}\left(-0.31, \frac{\sigma^2}{n_0+m}\right)$ où $n_0 + m = 267.2$. La distribution prédictive du $\log(OR)$ et de moyenne

$$E(\log(OR)_F | \log(OR)) = (-0.31)$$

et la variance

$$\begin{aligned} \text{Var}(\log(OR)_F | \log(OR)) &= \sigma^2(1/267.2 + 1/20) \\ &= \sigma^2/18.6 \\ &= 0.46^2. \end{aligned}$$

Ceci est montré sur la figure (2.1). La grande incertitude dans les futures observations est évidente.

En considérons les donnée de l'essai seulement c'est équivalent à dire que $n_0 = 0$, et en utilisant une a priori "flate", et par conséquent la distribution a postérieure courante sera, $\theta | Y_m \sim \mathcal{N}(-0.74, \sigma^2/m)$, où $m = 30.5$ (le nombre

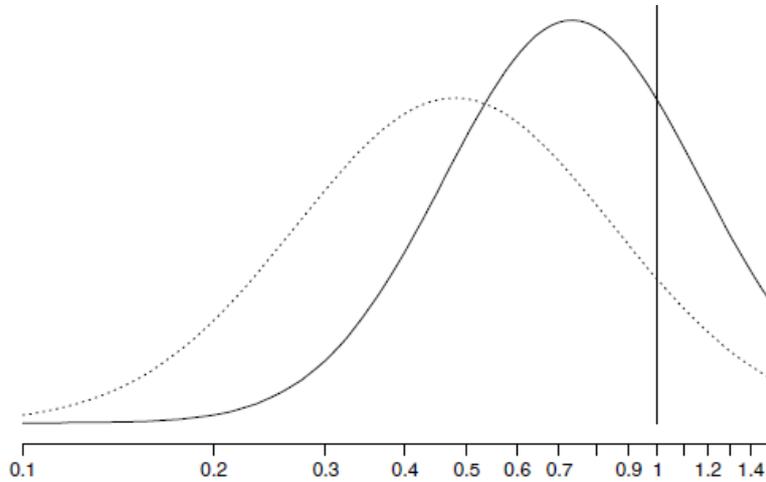


FIGURE 2.1 – On observera les distributions prédictives pour OR observe dans de les 100 patients futures randomisés à chaque bras dans l'essai **GREAT**, assumant autour 20 événements : avec l'information a priori (la ligne en lisse), et sans l'information a priori (la ligne en tiré)

d'événements) par conséquent, ignorant l'opinion de l'expert, la distribution prédictive de $\log(OR)$ ayant la moyenne

$$E(\log(OR)_F | \log(OR)) = (-0.74),$$

et la variance

$$\begin{aligned} \text{Var}(\log(OR)_F | \log(OR)) &= \sigma^2(1/30.5 + 1/20) \\ &= \sigma^2/12.1 \\ &= 0.58^2. \end{aligned}$$

La figure (2.1) prouve que cette distribution prédictive est considérablement plus plate que dans le cas où on prend en considération l'avis de l'expert.

Nous pouvons employer les distributions prédictives pour calculer la chance de n'importe quels résultat d'intérêt. Disons que $\log(OR)$ inférieure à 0.50 des composantes futures de l'essai.

En utilisant l'information a priori $\theta \sim \mathcal{N}(-0.31, \sigma^2/n_0)$ assez sceptique, cette probabilité sera :

$$\begin{aligned} P(Y_n < \log(0.50) | Y_m) &= \phi\left(\frac{-0.69 + 0.31}{0.4}\right) \\ &= \phi(-0.83) \\ &= 0.21, \end{aligned}$$

et si la distribution a priori est ignorée on aura :

$$\phi((-0.69 + 0.74)/0.58) = \phi(0.08) = 0.53.$$

Donc l'avis a priori nous pousse à douter que l'avantage courante sera observé dans les futures patients si l'essai est prolongé.

2.3 Test séquentiel Bayésien de groupe

Pour illustrer les calculs, considérons le cas des données normales provoquant les statistiques normalisées Z_1, Z_2, \dots avec la distribution commune canonique (1.1). À toute étape $k = 1, 2, \dots$ la vraisemblance pour le paramètre θ est la densité de la loi $\mathcal{N}(\theta, I_k^{-1})$, évalué à

$$\hat{\theta}^{(k)} = Z_k / \sqrt{I_k}$$

Nous discuterons le choix a priori en détail, dans la section(2.2.2). Considérons le choix commode d'une distribution a priori normale conjuguée $\mathcal{N}(\mu_0, \sigma_0)$.

En utilisant (2.1) la distribution a postérieure pour θ à l'étape k sera :

$$\mathcal{N}\left(\frac{\hat{\theta}^{(k)} I_k + \mu_0 \sigma_0^{-2}}{I_k + \sigma_0^{-2}}, \frac{1}{I_k + \sigma_0^{-2}}\right) \quad (2.3)$$

et l'intervalle crédible pour θ au niveau 95% est donné comme suit par

$$\left(\frac{\hat{\theta}^{(k)} I_k + \mu_0 \sigma_0^{-2}}{I_k + \sigma_0^{-2}} \pm 1.96 \frac{1}{\sqrt{I_k + \sigma_0^{-2}}}\right) \quad (2.4)$$

Notons que cet interval " est rétréci " vers μ_0 tandis que l'intervale de confiance habituel de l'évaluation est centré autour de l'estimateur $\hat{\theta}^{(k)}$. Souvent la moyenne a priori μ_0 prend la valeur 0, la valeur à l'hypothèse nulle.

Dans ce cas l'intervalle crédible pour θ est rétréci vers 0, une propriété considérée comme souhaitable par quelques auteurs (**Hughes et Pocock (1988)**).

Si $\sigma_0 \rightarrow \infty$ alors l'a priori est de référence ou non-informative uniform et la distribution a postérieure pour θ se réduit au $\mathcal{N}(\hat{\theta}^{(k)}, I_k^{-1})$.

2.3.1 Règles d'arrêt Bayésiennes

Bien que des inférences Bayésiennes sur l'arrêt soient facilement dérivées, les problèmes de la conception, et en particulier de la décision du moment où on doit s'arrêter ne sont pas francs.

Nous considérerons l'approche théorique de la décision Bayésienne dans ce qui suit. Cependant, cette méthodologie exige des coûts et des utilités, qui sont difficiles à évaluer, et une approche est souvent adaptée basée seulement sur la distribution a posteriori (définition (2.2.2)). Par exemple, on peut arrêter un essai tôt si à une certaine étape intermédiaire k

$$(i)P\{\theta < 0|donnes\} < \epsilon \text{ ou } (ii)P\{\theta > 0|donnes\} < \epsilon, \quad (2.5)$$

telle que ϵ pourrait être 0.025 c'est équivalent à l'arrêt quand à $1 - \epsilon$ la région crédible exclut zero. **Metha et Cain(1984)** proposaient des règles semblables.

La conception séquentielle n'affecte pas l'inférence Bayésienne, ainsi l'intervalle crédible(2.4) est toujours un sommaire correct dans le paradigme Bayésien, à l'heure de la k^{me} analyse, soit en s'arrêtant ou pas. Par exemple, si la condition (i) dans (2.5) est satisfaite alors la probabilité a posteriori que $\theta < 0$ est moins que ϵ , indépendamment de n'importe quelle règle d'arrêt formel cette propriété mène au conclusion de (**Berry(1987)**) qu'il n'y a aucun prix, à payer de regarder des données en utilisant une approche Bayésienne.

Néanmoins, les propriétés fréquentistes des arrangements de surveillance Bayésiens, peuvent tout à fait être étonnantes, en particulier, le taux d'erreur du type I n'est pas commandée, et peut être excessivement gonflées.

Exemple 2.3.1. *Reprenons l'exemple (2.2.2) sur la loi normal avec :*

1) **a priori non-informative :**

En faisant $\sigma_0 \rightarrow \infty$ dans (2.3) nous voyons que la procédure (2.5) avec $\epsilon = 0.025$ est équivalente à la règle répétée du test de signification qui s'arrête la première fois que $|Z_k| > 1.96$.

La table (2.2) donne le taux d'erreur du type I de cette procédure en fonction de le nombre maximum des analyses. C'est également la probabilité, sous $\theta = 0$, que 95% intervalle crédible sur l'arrêt $(\hat{\theta}^{(k)} \pm 1.96I_k^{-1})$ ce qui coïncide avec l'intervalle non ajusté de confiance fréquentist de 95% de ne pas inclure la valeur vraie $\theta = 0$.

Le comportement semblable des taux d'erreur du type I, pour cette forme de procédure se produit également pour d'autres choix d'a priori.

2) **a priori conjuguée normal de moyenne null et une variance $\sigma_0^{-2} = I_K/4$:**

Ici on assume que les analyses sont équidistantes avec $I_k = (k/K)I_K$, $k = 1, \dots, K$.

Number of analyses		Type I error probability
K	Non-informative prior	
1	0.05	
2	0.08	
3	0.11	
4	0.13	
5	0.14	
10	0.19	
20	0.25	
50	0.32	
∞	1.00	

TABLE 2.2 – Taux des faux positifs pour les règles d'arrêt (2.5) et (2.6) basé sur des probabilités a postériori Bayésiennes. Les analyses sont équidistantes.

cette a priori est appelée "handicap" prior (**C. Jennison (2000)**).

Pour $k \rightarrow \infty$, **Cornfield (1966)** a proposé une procédure en utilisant des a priori mixed en donnant une masse de probabilité pour $\theta = 0$ et la probabilité restante $(1-p) \sim \mathcal{N}(0, \sigma_0^2)$. Donc l'a postériori des odds (chances) en faveur de l'hypothèse $H_0 : \theta = 0$ est défini par :

$$\lambda = \frac{P\{\theta = 0 | \text{donne}\}}{P\{\theta \neq 0 | \text{donne}\}} = \frac{p}{1-p} RBO$$

où

$$RBO = \sqrt{(1 + I_k \sigma_0^2)} \exp\left(\frac{-Z_k^2}{2\{1 + (I_k \sigma_0^2)^{-1}\}}\right)$$

qui est le rapport des odds postériori par rapport aux antérieurs pour H_0 . Ce rapport est appelée "le facteur de Bayes" par (**Dickey (1973)**).

Supposons qu'on utilise la règle d'arrêt précédente, si $|Z_k| > 1.96$. Alors, en s'arrêtant avec I_k large, les odds a postériori

$$\lambda \simeq \{p/(1-p)\} \sigma_0 \sqrt{I_k} \exp(-1.96^2/2)$$

qui favorisent H_0 et pas H_1 si I_k est suffisamment grand.

Remarque 2.3.1. Cette procédure n'est pas subjective car, le rejet de H_0 est presque certain.

Mais, (**Cornfield(1966b)**) a continué pour proposer une règle d'arrêt basé sur les frontières d'arrêt pour RBO ou d'une manière équivalent, les odds postérieurs, λ .

Pour une telle procédure les frontières d'arrêts basées sur Z_k , divergent avec la taille de l'échantillon à la différence des frontières d'arrêt d'**O'Brien et Fleming** (section 1.4.4) qui deviennent plus étroites avec la taille de l'échantillon (**C. Jennison (2000)**).

Lachin (1981) a prolongé le modèle de **Cornfield** à une hypothèse null composée en remplaçant l'a priori discrète par autre continue soutenu sur un petit intervalle autour de zéro.

Remarque 2.3.2. *La plupart des Bayésiens ne sont pas satisfaits des a priori mélangées. Il faut mentionner que les conceptions basées sur RBO n'ont pas été adaptées dans la pratique récentes et les recommandations Bayésiennes ont été basées sur la règle(2.5).*

Il a été noté que les inférences Bayésiennes, basées sur des probabilités a posteriori, ne dépendent pas des règles de surveillance ou d'arrêt.

Par conséquent, dans le paradigme Bayésien, il n'est pas nécessaire de définir une taille d'échantillon maximale. Cependant, il est utile d'avoir une taille d'échantillon cible et cela peut être basé sur des analyses "préposées" utilisant des distributions prédictives.

Règle d'arrêt basée sur l'intervalle crédible :

Pour l'arrêt, la décision est basée sur l'intervalle crédible (voir la remarque 2.2.1), (θ_L, θ_U) . $\theta > 0$ correspond à une situation où nous souhaitons décider qu'un nouveau traitement est préférable, notre règle de décision doit recommander le nouveau traitement si $\theta_L > 0$, recommander contre le nouveau traitement si $\theta_U < 0$, et soient "neutre" si $\theta_L \leq 0 \leq \theta_U$. Pour une taille d'échantillon donnée, la probabilité conditionnelle de la conclusion avec une recommandation pour le nouveau traitement, c.à.d, $\theta > 0$, peut être calculé par :

$$\frac{\int_0^{+\infty} P\{\theta_L > 0 | \theta\} \pi(\theta) d\theta}{\int_0^{+\infty} \pi(\theta) d\theta}$$

Une expression semblable peut être calculée pour la probabilité conditionnelle de trouver $\theta_U < 0$, et le rejet du nouveau traitement sachant qu'il est inférieur, c.à.d $\theta < 0$. Une taille de l'échantillon cible peut être alors déterminée en s'assurant que tous les deux probabilités excèdent un certain seuil prescrit, (par exemple 90%). Alternativement, la taille de l'échantillon

cible peut être basée sur la probabilité prédictive d'une épreuve concluante c.à.d :

$$1 - P\{\theta_L \leq 0 \leq \theta_U\} = 1 - \int_0^{+\infty} P\{\theta_L \leq 0 \leq \theta_U\} \pi(\theta) d\theta$$

2.3.2 Choix de distribution a priori

La clef de l'approche Bayésienne est que les croyances antérieures, et les preuves externes peuvent être récapitulées mathématiquement et s'expriment, sous la forme d'une distribution a priori pour les paramètres inconnus. Ce choix est particulièrement important dans le contexte de la surveillance séquentielle de groupes parce que l'arrêt, dépendant des données, peut augmenter considérablement la sensibilité des intervalles crédibles bayésiens à la spécification erronée de l'a priori (**Rosenbaum et Rubin (1984)**).

Le choix de l'a priori est subjectif, et il n'y aura aucun choix objectif sur lequel tout peut convenir. Une méthode simple est de rapporter la fonction de vraisemblance que chaque consommateur combine, avec son choix a priori personnel afin de faire des inférences et des décisions .

Cependant, c'est habituellement impraticable. Au lieu de cela, une stratégie recommandée est de baser les décisions sur une collection d'analyses qui résultent d'une collection ou d'une "communauté" des a prioris. (**Kass et Greenhouse (1980)**) souligne que le but d'un essai clinique est "d'apporter des avis différents au consensus" et considération simultanée d'une variété de priori peuvent aider à ce but. **Fayers, Ashby et Parmar (1997)** déclarent que les distributions a priori "devraient refléter le niveau du scepticisme, qui est exprimé par des cliniciens au quels, on cherche à influencer". Nous énumérons maintenant les types des a priori qui peuvent être considérés :

- 1) **A prioris cliniques** : sont censés de refléter l'opinion des experts, qui sont obtenus par l'intermédiaire d'un questionnaire (par exemple **Freedman et Spiegelhalter(1983)**) visé pour des cliniciens compétents sur le terrain.

Alternativement, des a prioris cliniques peuvent être basés sur des méta-analyses ou des aperçus d'informations historiques mais, en raison du biais de publication, celles-ci peuvent être trop optimistes.

- 2) **Les a prioris sceptiques** : devraient représenter une extrémité dans la gamme de l'opinion. Comme méthode de travail de construction. **Spiegelhalter et al. (1994)** suggèrent une a priori symétrique cen-

trée à zero avec seulement une petite probabilité, par exemple 5%, que le bénéfice θ dépasse une valeur alternative indiquée dans l'hypothèse alternative $\theta_1 > 0$.

- 3) **Les a priori enthousiastes** : agissent comme un contre poids aux a priori sceptiques. Sur le plan opérationnel une a priori enthousiaste pourrait être construite en centrant la distribution à l'alternative θ_1 , avec seulement une petite probabilité que $\theta < 0$. Le but de la présence des a priori sceptiques et enthousiastes dans une collection est de diminuer la pression d'arrêter prématurément face aux résultats positifs ou négatifs résultats, respectivement.
- 4) **A priori non informative ou de référence** : essayent d'être objectifs du fait qu'elles essayent de représenter un manque d'opinion a priori. Dans l'exemple des données normale nous avons utilisé une a priori uniforme (la limite lorsque $\sigma_0 \rightarrow \infty$). Le choix de l'a priori non informative dépendent de la paramétrisation, ainsi ce qui est censée d'être non informative pour θ peut être tout à fait pour une fonction $g(\theta)$.

Chapitre 3

Designs séquentiels de groupe

L'approche du stochastique curtalement a pour principe d'extrapoler, à partir des résultats observés lors d'une analyse intermédiaire, ce que pourrait être le résultat final de l'essai et de calculer les probabilités d'obtenir une différence significative sous différentes hypothèses d'effet du traitement. Parmi ces méthodes, on peut trouver l'approche de puissance conditionnelle et puissance prédictive.

Cette approche de surveillance séquentielle a été évaluée à partir de l'idée de la restriction simple par laquelle une expérience peut être terminée dès que le résultat final devient inévitable.

3.1 Puissance conditionnelle

Dans ce cas, nous souhaitons évaluer une hypothèse nulle :

$$H_0 : \theta = \theta_0$$

concernant le paramètre θ .

D'abord un "test de référence", est indiqué. C'est typiquement un test simple avec l'erreur du type I, α sous $\theta = \theta_0$ et de puissance $1 - \beta$ à une alternative $\theta = \theta_1$ pour un test unilatéral, ou aux valeurs θ_1 et θ_2 pour un test bilatéral. Cependant le test de référence peut être un test séquentiel ou séquentiel de groupe.

À une étape intermédiaire, k , pendant l'étude, soit $D(k)$ les données accumulées. La puissance conditionnelle à l'étape k est définie par :

$$p_k(\theta) = P_\theta\{\text{rejette } H_0 | D(k)\}$$

Remarque 3.1.1.

- 1) pour $k = 0$ on trouve la fonction habituelle de puissance.
- 2) À l'arrêt prévu de l'étude, c.à.d. à l'étape K , cette probabilité est 0 ou 1.

À une étape k , intermédiaire, la puissance conditionnelle peut être tracée en fonction de θ et des valeurs θ_0, θ_1 et θ_2 (pour un essai bilatéral), et $\hat{\theta}(k)$, l'évaluation courante de maximum de vraisemblance de θ , sont d'intérêt particulier.

La fonction conditionnelle de puissance est un dispositif utile de communication avec les investigateurs cliniques. Par exemple, elle peut être employée pour illustrer les effets du compte de régularisation ou pour faciliter la prise de décision pour abandonner une étude si la puissance conditionnelle semble faible.

Une valeur élevée de $p_k(\theta_0)$ indique que l'essai de référence est peut susceptible d'accepter H_0 sachant les données accumulées, même si H_0 est vrai. **Lan et al. (1982)** ont proposé une règle formelle par laquelle l'arrêt se produit à l'étape k , pour rejeter H_0 , si $p_k(\theta_0) \geq \gamma$, pour une constante prédéfinie γ telle que $0.5 < \gamma < 1$ (les valeurs 0.8 ou 0.9 sont recommandés).

De même, on peut permettre l'arrêt tôt d'accepter H_0 si :

$$1 - p_k(\theta_1) \geq \gamma'$$

dans un test unilatéral et pour un test bilatéral si :

$$1 - p_k(\theta_1) \geq \gamma'$$

et

$$1 - p_k(\theta_2) \geq \gamma',$$

où $0.5 < \gamma' < 1$. La quantité $1 - p_k(\theta_1)$ est dite l'indice de futilité (**Ware et al. (1985)**).

En employant les règles ci-dessus, les probabilités d'erreurs du test peuvent être prévus pour montrer que la décision finale de test de référence parfois sera incorrectement anticiper.

Nous nous référons à un test formellement défini, basé sur les règles ci-dessus comme procédure de stochastique curtalement (**PSC**).

On peut calculer les probabilités d'erreur d'un **PSC** avec un programme donné d'analyse intermédiaire.

Propriétés 3.1.1 (Propriétés stochastique de la puissance conditionnelle).

1) **Lan, Simon et Halperin (1982)** ont montré avec un argument simple que le type d'erreur I ne sera pas plus que α/γ . En effet,

$$\{p_k(\theta), k = 1, 2, \dots\}$$

est une martingale par rapport à la filtration définie par $\{D(1), D(2), \dots\}$. Soit ν le temps d'arrêt du test de référence (avec la possibilité qu'il soit un test séquentiel). D'abord, supposons que $\gamma' = 1$, alors il n'y a aucun arrêt pour acceptation de H_0 . Dans ce cas, l'étape auquel le **PSC** s'arrêtera sera prématuré

$$\tau = \min(\nu, \min\{k : p_k(\theta) \geq \gamma\}),$$

$$E_{\theta_0}\{p_\tau(\theta_0)\} = E_{\theta_0}\{p_0(\theta_0)\} = p_0(\theta_0) = \alpha.$$

Or, de la définition de τ

$$\begin{aligned} E_{\theta_0}\{p_\tau(\theta_0)\} &\geq \gamma P_{\theta_0}\{p_k(\theta_0) \geq \gamma, \text{ pour } k \leq \nu\} \\ &= \gamma P_{\theta_0}\{\text{Rejet } H_0\} \end{aligned}$$

et par conséquent

$$P_{\theta_0}\{\text{Rejet } H_0\} \leq \alpha/\gamma$$

2) Si $\gamma' < 1$, alors l'erreur du type I, peut seulement être réduite plus loin et ainsi il est encore plus proche de α/γ . De même l'erreur du type II du **PSC** n'est pas plus que β/γ' , où β est l'erreur du type II du test de référence.

De cette manière, on peut s'assurer que les deux types d'erreur I et II soient au plus α et β pour le test de **PSC**.

Calcul de la fonction conditionnelle de puissance :

Supposons que des analyses successives donnent une suite de statistiques de test Z_1, \dots, Z_K avec la distribution commune canonique (1.1) donnant les informations I_1, \dots, I_K pour θ .

Puisque Z_k est une statistique suffisante pour θ à l'étape k qui peut remplacer $D(k)$. Sans perte de généralité, soit $\theta_0 = 0$ et $\theta_1 = \delta$. Supposons que le test de référence est un test simple unilatéral, $H_0 : \theta_0 = 0$ avec α , l'erreur du type I et la puissance $1 - \beta$ à $\theta = \delta$, le test continue à l'étape finale K où on rejette H_0 si

$$Z_K > z_\alpha,$$

où $z_\alpha = \phi^{-1}(1 - \alpha)$.

La puissance $1 - \beta$ à $\theta = \delta$ est réalisée en choisissant la taille de l'échantillon qui assure un niveau d'information :

$$I_K = I_{f,1} = \{\phi^{-1}(1 - \alpha) + \phi^{-1}(1 - \beta)\}^2 / \delta^2$$

pour chaque $k = 1, \dots, K - 1$.

Dans ce cas la distribution conditionnelle de Z_K sachant Z_k est :

$$Z_K | Z_k \sim \mathcal{N}(Z_k \sqrt{I_k / I_K} + \{\theta(I_K - I_k) / \sqrt{I_K}\}, (1 - I_k) / I_K).$$

par conséquent, la puissance conditionnelle à l'analyse k sera

$$p_k(\theta) = \phi \left\{ \frac{Z_k \sqrt{I_k} - z_\alpha \sqrt{I_K} + (I_K - I_k)\theta}{\sqrt{I_K - I_k}} \right\} \quad k = 1, \dots, K - 1 \quad (3.1)$$

Exemple 3.1.1. La figure (3.1) donne les courbes de puissance d'un test de référence (la ligne lisse) c.à.d, $p_k(\theta) : k = 0$, et la puissance conditionnelle (la ligne en tiré) calculer dans une étape intermédiaire.

Dans cet exemple $\alpha = 0.05$ et $I_K = 214.1$, soit la courbe de puissance

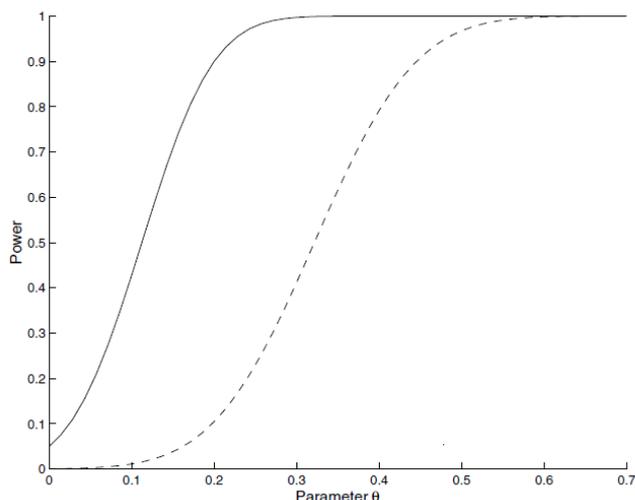


FIGURE 3.1 – Courbe de la puissance (la ligne lisse) et puissance conditionnelle (la ligne en tiré) pour un test unilatéral

conditionnelle pour l'étape k pour la quelle la moitié des observations ont été prises, ainsi $I_k = I_K / 2 = 107.1$, et l'EMV courant de θ est

$$\hat{\theta}^{(k)} = \delta / 2 = 0.1$$

Ceci donne une valeur $Z_k = 0.1 \sqrt{107.1} = 1.035$ (voir la section 2.3).

Nous voyons que sous l'alternative $\theta = 0.2$, la probabilité de rejeter H_0 si l'expérience continue à la fin, comme été prévu, a été réduit de 0.9 à 0.1. Si on définit une règle d'arrêt comme ci-dessus alors :

Arrêter et rejeter H_0 si $p_k(\theta_0) \geq \gamma$, pour $\theta_0 = 0$ est équivalente à :

$$Z_k \geq z_\alpha \sqrt{I_K/I_k} + z_{1-\gamma} \sqrt{(I_K - I_k)/I_k} \quad (3.2)$$

Arrêter et accepter H_0 si

$p_k(\theta_1) \leq 1 - \gamma'$, pour $\theta_1 = \delta$ est équivalente à :

$$Z_k \leq z_\alpha \sqrt{I_K/I_k} - z_{1-\gamma'} \sqrt{(I_K - I_k)/I_k} - \delta(I_K - I_k)/\sqrt{I_k} \quad (3.3)$$

On peut voir ces frontières d'arrêt sur la figure (3.2) pour le test unilatéral défini par la figure(3.1) pour $\alpha = 0.05, 1 - \beta = 0.9, \delta = 0.2, I_k = 214.1$, et $\gamma = \gamma' = 0.8$. On remarque que quel que soit le plan d'analyse intermédiaire,

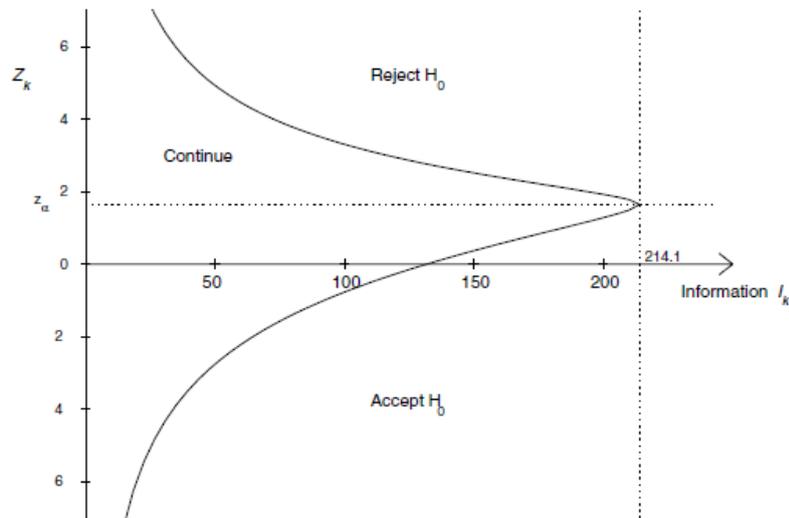


FIGURE 3.2 – Frontière d'arrêt pour un test unilatéral stochastiquement raccourci en utilisant l'approche de puissance conditionnelle, avec $\alpha = 0.05, 1 - \beta = 0.9$ à $\theta = 0.2$, le niveau d'information $I_{f,1} = 214.1$ et $\gamma = \gamma' = 0.8$.

la probabilité d'erreur du type I, α , de cette procédure n'exédèra pas $\alpha/\gamma = 0.05/0.8 = 0.0625$ et la puissance lorsque $\theta = 0.2$ est garantie pour être au moins $1 - (0.1/0.8) = 0.875$.

Supposons maintenant que le test de référence est bilatéral tel que $H_0 : \theta = 0$, la probabilité d'erreur de type I, α et la puissance $1 - \beta$ pour $\theta = \pm\delta$, le test prend une seule décision à l'étape final K :

rejette H_0 si

$$|Z_K| \geq z_{\alpha/2}$$

et accepter H_0 autrement.

La taille de l'échantillon devrait être choisie pour rapporter un niveau $I_K = I_{f,2}$.

La puissance conditionnelle à l'analyse k est :

$$p_k(\theta) = \phi \left\{ \frac{Z_k \sqrt{I_k} - z_{\alpha/2} \sqrt{I_K} + (I_K - I_k)\theta}{\sqrt{I_K - I_k}} \right\} + \phi \left\{ \frac{-Z_k \sqrt{I_k} - z_{\alpha/2} \sqrt{I_K} - (I_K - I_k)\theta}{\sqrt{I_K - I_k}} \right\} \quad (*)$$

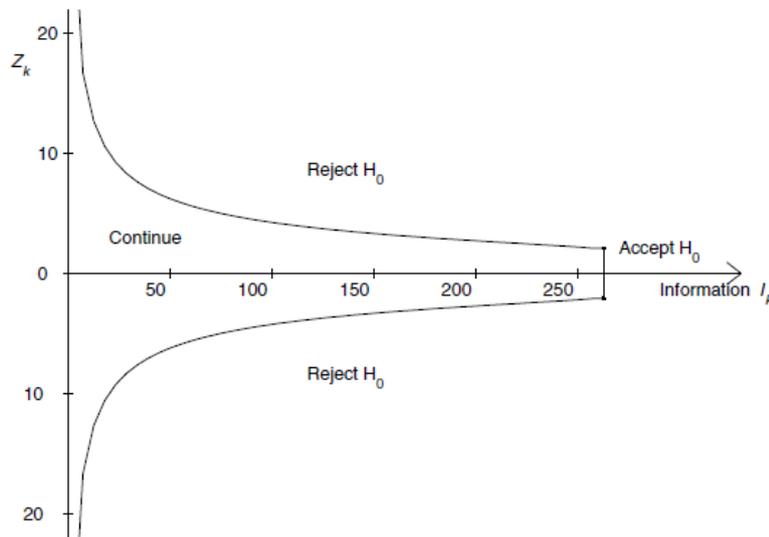


FIGURE 3.3 – Frontière d'arrêt pour un test bilatérale stochastique en utilisant l'approche conditionnelle de puissance. le test de référence est un test simple bilatéral avec $\alpha = 0.05$, et le niveau d'information $I_{f,2} = 262.7$, $1 - \beta = 0.90$ à $\theta = 0.2$ et $\gamma = 0.8$

Supposons que nous employons la **PSC** pour définir une règle d'arrêt formelle qui se termine tôt en rejetant H_0 si $p_k(\theta) \geq \gamma$, d'après les propriétés stochastique de cette procédure, la probabilité d'erreur du type I de cette **PSC** est au plus α/γ . La frontière d'arrêt pour $\alpha = 0.05$, $\gamma = 0.8$ et $I_k = 262.7$ est montrée dans la figure (3.3).

La puissance du test est 0.9 quand $\theta = \delta = \pm 0.2$, est ainsi directement comparable aux frontières des essais de **d'O'Brien et de Fleming** avec $K = 4$ groupes d'observations avec les mêmes probabilités d'erreurs.

Remarque 3.1.2. Si $\gamma = 0.5$, la **PSC** est la version continue du monitoring d'un essai **d'O'Brien et de Fleming**, cette **PSC** également coïncide avec d'autres tests (**C. Jennison (2000)**).

Conclusion 3.1.1. *L'importance de cette méthode est la capacité de faire des analyses intermédiaires arbitraires non planifiées, même si les données dépendants du temps. Si la synchronisation des analyses est fixée à l'avance, des méthodes numériques peuvent être employées pour obtenir une valeur exacte de l'erreur du type I et des frontières moins conservatrices pourraient être dérivées.*

3.2 Puissance Prédicative

Une critique de la méthode, précédente est quelle est basée sur la puissance conditionnelle calculée sous des valeurs de θ qui ne peuvent être pas soutenues par les données courantes. Par exemple, le critère arrêt pour rejeter H_0 implique $p_k(\theta)$ évalué à $\theta = \theta_0$, une valeur de θ de peu vraisemblable si le rejet est considéré.

Nous présentons maintenant une procédure qui évite ce problème, en faisant la moyenne de la puissance conditionnelle par rapport aux valeurs de θ dans un calcul Bayésien, mais toujours avec un test fréquentiste à l'esprit.

En utilisant comme précédemment, nous supposons un test de référence de $H_0 : \theta = 0$ avec l'erreur du type I, α , et la puissance $1 - \beta$ à $\theta = \theta_1$ dans le cas d'un test unilatéral ou $\theta = \theta_1$ et θ_2 pour un test bilatéral.

Au lieu d'employer la puissance conditionnelle à des valeurs θ , nous définissons " " la puissance prédictive " " à l'étape k d'être la moyenne de la fonction conditionnelle de puissance

$$p_k = \int p_k(\theta)\pi(\theta | D(k))d\theta \quad (3.4)$$

où $p_k(\theta)$ est définie dans la section précédente.

La fonction de masse π est la densité a posteriori de θ pour une distribution à priori de θ et les données $D(k)$ accumulées par l'analyse k , donc π reflète la croyance courante au sujet de la valeur de θ .

Cette approche a été préconisée par **Herson (1979)**, **Lhoi et al. (1985)**, **Spiegelhalter et al. (1986)**. La puissance prédictive est encore une quantité utile à communiquer aux chercheurs appliquée à une étape intermédiaire d'un essai .

Comme la puissance conditionnelle, elle peut être employée pour définir une règle d'arrêt formelle par l'arrêt à l'étape k pour :

rejeter H_0 si $p_k > \gamma$

ou pour accepter H_0 si $p_k \leq 1 - \hat{\gamma}$

Exemple 3.2.1. *Considérons encore l'exemple de la section précédente, on assume que Z_1, Z_2, \dots, Z_K suivent la distribution (1.1) avec I_1, \dots, I_K , et soit un test unilatéral avec la probabilité d'erreur du type I, α , et la puissance $1 - \beta$ à $\theta = \delta$ conduit avec le niveau d'information $I_K = I_{f,1}$. Ce test rejette $H_0 : \theta = \theta_0$ si et seulement si :*

$$Z_K \geq z_\alpha$$

Supposons que nous employons l'a priori impropre $\pi(\theta) = 1$, pour tous θ , pour représenter un premier manque de la connaissance sur θ , puis à l'étape k , quand, l'EMV de θ est $\hat{\theta} = Z_k/\sqrt{I_k}$, la distribution a posteriori pour θ est donnée par

$$\theta \mid Z_K \sim \mathcal{N}(Z_k/\sqrt{I_k}, 1/I_k)$$

Substituant cette densité dans (3.4) et effectuant l'intégration, nous obtenons

$$p_k = \phi \left\{ \frac{Z_k \sqrt{I_K} - Z_\alpha \sqrt{I_k}}{\sqrt{(I_K - I_k)}} \right\} \quad (3.5)$$

où ϕ est la distribution cumulative de la loi normale standard, après θ à été remplacé par son EMV mais, en fait, il diffère par le facteur $\sqrt{I_k/I_K}$ une certaine différence doit être prévu car θ est fixé dans le calcul de la puissance conditionnelle tandis que l'EMV est le centre de la distribution a posteriori dans le cas de la puissance prédictive.

Les critères d'arrêts $p_k \geq \gamma$ et $p_k \leq 1 - \gamma$ sont équivalents à : rejeter H_0 si :

$$Z_k \geq Z_\alpha \sqrt{I_k/I_K} + Z_{1-\gamma} \sqrt{(I_K - I_k)/I_K} \quad (3.6)$$

et acceptter H_0 si :

$$Z_k \leq Z_\alpha \sqrt{I_k/I_K} - Z_{1-\gamma} \sqrt{(I_K - I_k)/I_K} \quad (3.7)$$

La figure (3.4) montre ces frontière quand $\gamma = \hat{\gamma} = 0.8$ et le test de référence est le même utilisé dans la figure (3.2) bien que les critères (3.6) et (3.7) soient semblables aux règles d'arrêt de la puissance conditionnelle données dans des figure (3.2) et (3.3), une comparaison des figure (3.4) et (3.2) montre que l'approche de puissance prédictive donne des frontières beaucoup plus étroites pour les même valeurs de γ et γ' .

L'arrêt prématuré est autorisé plus aisément puisque les probabilités conditionnelles sont basées sur une estimation de θ obtenu à partir des données plutôt que les valeurs présumées $\theta = \theta_0$ et $\theta = \delta$. Bienque les resultat analytiques simples ne soient pas disponibles, on peut utiliser les calculs numériques pour trouver l'effet de cet arrêt prématuré sur l'erreurs du type I et II.

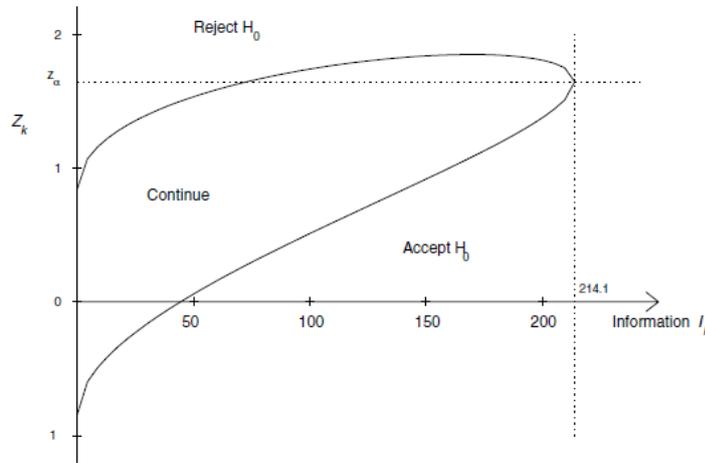


FIGURE 3.4 – Frontière d’arrêt pour un test unilatéral stochastiquement, raccourci en utilisant l’approche de puissance prédictive, avec une a priori uniforme. Le test de référence est un test unilatéral simple, avec $\alpha = 0.5$, $I_{f,1} = 214.1$, $1 - \beta = 0.9$ à $\theta = 0.2$, $\gamma = \hat{\gamma} = 0.8$.

Dans le problème d’un test bilatéral de la section (3.1). Le test de référence au niveau $I_K = I_{f,2}$ rejette $H_0 : \theta = 0$ si $|Z_K| \geq z_{\alpha/2}$. Le critère $P_k \geq \gamma$ réclame maintenant une décision pour arrêter et rejeter H_0 à l’étape k si

$$\phi \left\{ \frac{|Z_k| \sqrt{I_K} - z_{\alpha/2} \sqrt{I_k}}{\sqrt{(I_K - I_k)}} \right\} + \phi \left\{ \frac{-|Z_k| \sqrt{I_K} - z_{\alpha/2} \sqrt{I_k}}{\sqrt{(I_K - I_k)}} \right\} \geq \gamma$$

La frontière pour $\gamma = 0.8$ est montrée sur la **figure (3.5)**. Ici, le test de référence est un test bilatéral, la probabilité d’erreur du type I, $\alpha = 0.05$, et $I_K = 262.7$, pour réaliser la puissance $1 - \beta = 0.9$ à $\theta = 0.2$ comme dans l’exemple employé pour illustrer l’approche conditionnelle de puissance de la figure(3.3).

Une fois de plus, la région de continuation est plus étroite dans les premières de analyse que de l’approche de la puissance conditionnelle, qui mène à un arrêt prématuré plus facile. Si un plan particulier des analyses est projeté, les taux d’erreurs atteints pourraient être calculés numériquement et γ ou I_K croit si on désire de réduire ces taux d’erreur. Nous avons seulement employé le critère $P_k \geq \gamma$ un arrêt précoce. Une première décision pour accepter H_0 si $P_k \leq 1 - \hat{\gamma}$ pourrait également être autorisé.

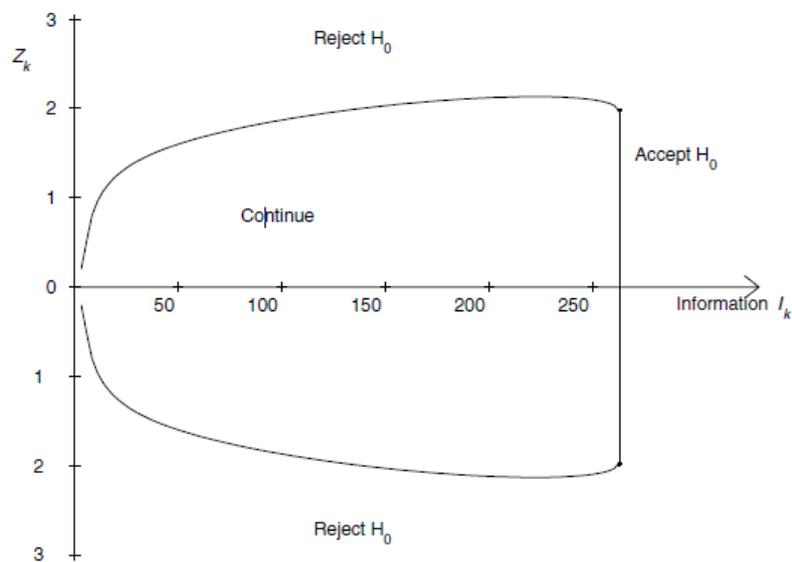


FIGURE 3.5 – Frontière d'arrêt pour un test bilatéral stochastiquement raccourci en utilisant l'approche de puissance prédictive avec une a priori uniforme. Le test de référence est un test bilatéral simple, $\alpha = 0.05$, et niveau de l'information $I_{f,2} = 262.7$, pour réaliser la puissance 0.9 à $\theta = 0.2$. Le paramètre stochastique de terminaison $\gamma = 0.8$.

3.3 Application

Dans cette application, nous illustrerons le calcul de la puissance conditionnelle en utilisant des données intermédiaires dans un essai clinique randomisée conçue pour comparer les temps de survie des sujets sur deux bras de traitements. Ces calculs sont adaptés à ceux demandés par le conseil de surveillance des données et de la sécurité (**DSMB**) lors d'une analyse intermédiaire de l'essai de prévention nutritionnelle du cancer et ils suivent le développement d'**Andersen (1987)**.

Supposons des temps de survie d'être exponentiellement distribués avec une moyenne μ_A pour les patients recevant le traitement A et μ_B pour ceux recevant le traitement B . Soient $\lambda_A = 1/\mu_A$ et $\lambda_B = 1/\mu_B$ les taux de risque sur les traitements A et B respectivement, et définissons $\psi = \lambda_B/\lambda_A$ le rapport de risque ou "le risque relatif".

L'hypothèse nulle à examiner sera : $H_0 : \psi = 1$.

Nous décrivons d'abord un test d'échantillon fixe de H_0 . Soit X_A le nombre total de patient-années observées sur le traitement A , c'est à dire la somme des temps de survie de chaque patient jusqu'à l'échec ou la censure, et que N_A le nombre d'échec pour patients observés sur le traitement A . De même, on définit X_B et N_B pour le traitement B . Nous estimons les taux de risque par

$$\hat{\lambda}_A = N_A/X_A \text{ et } \hat{\lambda}_B = N_B/X_B$$

et le rapport de risque par

$$\hat{\psi} = \hat{\lambda}_B/\hat{\lambda}_A$$

un test de H_0 peut être basé sur la statistique

$$W = \log(\hat{\psi}) \sim \mathcal{N}(\log(\psi), \frac{1}{N_A} + \frac{1}{N_B}).$$

Ainsi un test bilatéral, avec l'erreur du type I est approximativement égal à α , rejette H_0 si

$$|W| \geq z_{\alpha/2} \sqrt{N_A^{-1} + N_B^{-1}}.$$

En utilisant la notation des sections précédentes, nous avons : l'information $I = (N_A^{-1} + N_B^{-1})^{-1}$, la statistique normalisée $Z = W\sqrt{I}$, et le paramètre canonique $\theta = \log(\psi)$.

Il est possible de déterminer le nombre de personnes-années qui devraient être observées dans l'étude (X_A et X_B). En donnant une estimation initiale à λ_A , l'erreur de type I, α , et la puissance $1 - \beta$ d'être réalisé à une alternative indiquée $\psi = \psi^*$.

La puissance du test ci-dessus, à une valeur donnée ψ (d'après la formule (*)), est approximativement

$$\phi \left\{ -z_{\alpha/2} + \frac{\log(\psi)}{\sqrt{N_A^{-1} + N_B^{-1}}} \right\} + \phi \left\{ -z_{\alpha/2} - \frac{\log(\psi)}{\sqrt{N_A^{-1} + N_B^{-1}}} \right\}. \quad (3.8)$$

Si l'étude accumule un total de T -années-personnes, réparties également entre les deux traitements, les valeurs prévues du N_A et N_B sont $\lambda_A T/2$ et $\psi^* \lambda_A T/2$, respectivement.

La substitution de ces valeurs (N_A et N_B) dans (3.8) donne une expression qui peut être évaluée pour trouver la valeur de T en donnant la puissance approximative $1 - \beta$ quand $\psi = \psi^*$ (George et Desu (1974)). La connaissance des taux d'accumulation et de renvoi peut être alors employée pour estimer les tailles d'échantillons exigées et pour étudier la durée de l'étude pour atteindre cette valeur de T années-personnes. Divers diagrammes, et programmes sont disponibles pour faciliter ce calcul; par exemple, **George et Desu(1974)**, **schoenfeld et Richter(1982)** et **Makuch et Simon(1982)**.

Dans notre exemple, le traitement A était un groupe de contrôle et un taux d'échec 3.14% par an a été prévu d'une expérience précédente. Le temps étant mesuré en années, cela donne $\lambda_A = 0.0314$. Une réduction du taux de risque de 25% a été considérée comme amélioration significative et par conséquent nous avons établi $\psi^* = 0.75$. Pour $\alpha = 0.05$, $1 - \beta = 0.8$, nous donne un total de $T = 14,116$ des années-personnes nécessaires. L'équation (3.8) en fonction de $\psi > 0$ avec $N_A = \lambda_A T/2$ et $N_B = \psi^* \lambda_A T/2$, où $\lambda_A = 0.0314$, $\psi^* = 0.75$ et $T = 14.116$, donne la fonction de puissance pour la conception original qui est montrée comme ligne solide dans la figure (3.6)

À une étape intermédiaire de l'essai, nous pouvons considérer la distribution à l'arrêt de la statistique $W = \log(\hat{\psi})$, compt tenu des données courantes.

Soit n_A le nombre de patients qui ont échoué jusqu'à l'instant, x_A toutes les années-personnes observées jusque là et s_A les années-personnes additionnelles projetées pour des sujets sur le traitement A . Définissons de même pour le traitement B , n_B , x_B et s_B et soit $s = s_A + s_B$.

Notons D_A et D_B les variables aléatoires présentant le nombres d'autres décès qui seront observées sur les traitements A et B , respectivement, dans les temps additionnels de suivi de s_A et s_B années personnes.

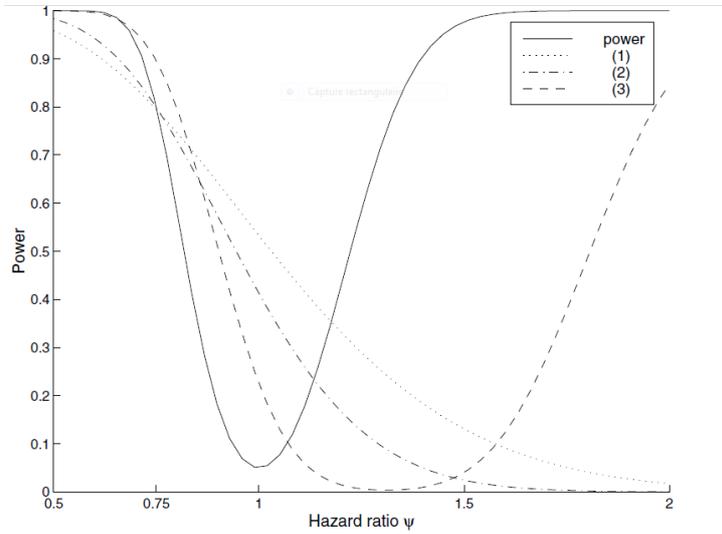


FIGURE 3.6 – Courbe de puissance et de puissance conditionnelle. La fonction de puissance pour la conception originale est montrée par la ligne solide. Les autres courbes représentent des fonctions de puissance conditionnelle de trois scénarios différents, (1), (2) et (3), continué après l'analyse intermédiaire.

À l'arrêt, nous aurons

$$\begin{aligned}
 W &= \log(\hat{\psi}) \\
 &= \log(\hat{\lambda}_B) - \log(\hat{\lambda}_A) \\
 &= \log\left(\frac{n_B + D_B}{x_B + s_B}\right) - \log\left(\frac{n_A + D_A}{x_A + s_A}\right)
 \end{aligned}$$

La distribution conditionnelle de cette statistique en donnant les valeurs actuellement observées n_A , x_A , n_B et x_B est approximativement normale de moyenne

$$\log\left(\frac{n_B + \psi\lambda_A s_B}{x_B + s_B}\right) - \log\left(\frac{n_A + \lambda_A s_A}{x_A + s_A}\right)$$

et de variance

$$\frac{\psi\lambda_A s_B}{(\psi\lambda_A s_B + n_B)^2} + \frac{\lambda_A s_A}{(\lambda_A s_A + n_A)^2}$$

(Andersen (1987)). Au temps d'arrêt on rejettera H_0 si

$$|W| \geq Z_{\alpha/2} \sqrt{(n_A + D_A)^{-1} + (n_B + D_B)^{-1}} \quad (3.9)$$

Par conséquent, nous pouvons estimer, la puissance conditionnelle de la distribution conditionnelle W indiquée ci-dessus, en substituant les estimateurs $s_A \hat{\lambda}$ pour D_A et $s_B \hat{\lambda}$ pour D_B dans le membre droit de (3.9), où $\hat{\lambda} = \frac{(n_A+n_B)}{(x_A+x_B)}$ est l'estimateur courant du taux commun de risque sous H_0 .

L'ors d'une réunion intermédiaire, le **DSMB** a été confronté aux données de mortalité dans le tableau (3.1).

Bras.de.traitement	échec	Années.personnes
A, Control	nA = 118	xA = 3896
B, traitement Active	nB = 89	xB = 3943
total	n = 207	x = 7839

TABLE 3.1 – Données de mortalité intermédiaire pour l'étude.

Une analyse conduite à ce point intermédiaire, (comme si c'étaient l'analyse finale) rapporte une estimation du rapport de risque

$$\hat{\psi} = \frac{\left(\frac{89}{3943}\right)}{\left(\frac{118}{3896}\right)} = 0.7452 \quad ,$$

Indiquant une meilleure survie observée sur le traitement actif. Donc :

$$\begin{aligned} W &= \log(0.7452) \\ &= -0.2940, \end{aligned}$$

et la statistique normalisée du test sera

$$\begin{aligned} Z &= \frac{W}{\sqrt{\frac{1}{118} + \frac{1}{89}}} \\ &= -2.094, \end{aligned}$$

pour lequel la p-valeur pour le test bilatéral est $P(|Z| > 2.094) = 0.036$.

Naturellement, cette p-valeur n'est pas ajusté à l'effet du multiples analyses (**C. Jennison (2000)** chapitre 9) pour plus d'informations). Cependant, le protocole ait indiqué que l'essai devrait continuer, le **DSMB** a en le sujet d'inquiétude et demande que des courbes conditionnelles de puissance, Soient calculées pour un test de référence à l'arrêt sous les trois scénarios suivants :

- 1) une année de suivi.
- 2) deux années supplémentaires de suivi.
- 3) continuation à l'arrêt comme il a été prévu.

La table(3.2) indique les années-personnes additionnelles et totales du suivi estimées sous chaque scénario, en utilisant des taux observés accumulés et les nombres de sujets actuellement actifs. Des courbes conditionnelles de puissance ont été alors calculées en utilisant les formules (3.8) et (3.9). Celles-ci sont tracées sur la **figure (3.6)** pour chacun des trois scénarios.

Scenario	Années- personnes additionnelles	Années-personnes total
	sA sB s	xA+SA xB+sB x+s
(1) Une d'année de plus	375 375 750	4271 4318 8589
(2) Deux d'années	815 815 1630	4711 4758 9469
(3) Plan original	3138 3139 6276	7034 7082 14116

TABLE 3.2 – Années-personnes additionnelles sous trois scénarios possibles pour continuer l'étude.

Discussion des résultats

Avec la fonction de puissance pour la conception originale, chacune des trois courbes conditionnelles de puissance est de forme U , mais ceux pour les scénarios (1) et (2) ne commencent pas à se lever du membre droit jusqu'à ce que θ commence à avoir des valeurs plus élevé autour $\psi = 3$ de (2) et $\psi = 5$ de (1). D'après ces courbes nous pouvons remarque que pour $\hat{\psi} = 0.75$ où l'estimateur courante $\hat{\psi} = 0.7452$, il y avait une probabilité conditionnelle élevée d'un résultat significatif qui seront 0.79, 0.80 et 0.90 pour des scénarios (1).(2). (3), respectivement.

Les calculs exacts sont fournis par les programmes (1,2 et 3) de R dans l'annexe. Même si l'hypothèse nulle étaient vraie et $\psi = 1$, la probabilité de rejeté H_0 était aussi grande que 53%, 41% et 23% sous les trois scénarios. Les calculs exacts sont fournis par les programmes (4,5 et 6) de R dans l'annexe.

Donc il est préférable de continuer l'essai.

En fait, l'essai a été terminée tôt à la prochaine analyse intermédiaire (**Anderson (1987)**).

Maintenant considérons le calcul de la puissance prédictive dans le cadre Bayésien. En choisissant une a priori impropre $\pi(\theta) = 1$, pour tout θ .

Alors la distribution a posteriori sera :

$$W = \log(\hat{\psi}) \mid Z_K \sim \mathcal{N} \left(Z_k \sqrt{((n_A + D_A)^{-1} + (n_B + D_B)^{-1})}, ((n_A + D_A)^{-1} + (n_B + D_B)^{-1}) \right)$$

et d'après la formule (3.5), la puissance prédictive à l'étape K sera :

$$\phi \left\{ \frac{|Z_k| \sqrt{I_K} - z_{\alpha/2} \sqrt{I_k}}{\sqrt{(I_K - I_k)}} \right\} + \phi \left\{ \frac{-|Z_k| \sqrt{I_K} - z_{\alpha/2} \sqrt{I_k}}{\sqrt{(I_K - I_k)}} \right\}$$

avec $I_K = I_{f,2} = \{\phi^{-1}(1-\alpha/2) + \phi^{-1}(1-\beta)\}^2 / \delta^2$ et $I_k = ((n_A + D_A)^{-1} + (n_B + D_B)^{-1})$
 Les calculs exacts sont fournis par les programmes (7,8 et 9) de R dans l'annexe.

Alors les résultats sont données dans le tableau (3.3).

Scénario	Puissance conditionnelle	Puissance prédictive
(1) Une d'année de plus	0,7797311	0,9799765
(2) Deux d'années	0,7906049	0,9827406
(3) Plan original	0,894965	0,9905724

TABLE 3.3 – La comparaison entre la puissance conditionnelle et la puissance prédictive

Il est claire, que la puissance prédictive est plus grande que la puissance conditionnelle et çadue au faite qu'on a pris la moyenne de la puissance conditionnelle sur toutes les valeurs du θ et non seulement pour une seule valeur qui etait son **EMV**.

De ce fait, nous aurons la puissance des règles Bayésiennes, qui nous donne plus de chance pour continuer l'étude et ne pas s'arrêter a vaut la fin prévu et décrite dans le protocole, en s'assurant, qu'on a pas gaspillé les ressources humaines et monitaires de l'étude.

Conclusion générale

Dans notre travail nous avons étudié l'importance des tests séquentiels de groupe qui joue un rôle très important dans le développement d'un médicament.

Ces méthodes séquentielles permettent de limiter la durée d'un essai randomisé et donc de recommander plus rapidement la meilleure attitude thérapeutique.

L'avantage de l'approche Bayésienne dans le concept des essais cliniques est la souplesse de l'utilisation de la prédiction pour prendre une meilleure décision d'efficacité/futilité d'un traitement.

Ainsi que la procédure de stochastique curtalement permet d'obtenir une différence significative sous différentes hypothèses d'effet du traitement, à partir des résultats observés lors d'une analyse intermédiaire. Parmi ces méthodes on peut trouver la notion de puissance conditionnelle et puissance prédictive, qui sont adaptées à la décision d'arrêt d'un essai clinique, notamment avec les procédures séquentielles : testes séquentiels de groupes, en générale, et bayésiennes en particulier.

Résumé

Dans ce travail, nous avons essayé d'exposer les tests séquentiels de groupe et en particulier les designs classiques et Bayésiennes de groupe et l'avantage de la prédiction pour prévoir le meilleur traitement dans les essais cliniques.

Notre point de vu a été illustré par des applications numérique aux données de survie, en comparant la puissance conditionnelle avec la puissance prédictive utilisé dans le contrôle intermédiaire des essais cliniques.

Annexe

```
> #Programme 1:
> # Pour scénario (1)
> alpha=0.05;lambdaa=0.0314;psi=0.75;xA=3896;xB=3943;NB=89;
> nA=118;sA=375;sB=375
> elambda=(nA+NB)/(xA+xB)
> elambda
[1] 0.02640643
> DA= sA*elambda
> DA
[1] 9.902411
> DB=sB*elambda
> DB
[1] 9.902411
> elambdaz=(NB+DB)/(xB+sB)
> elambdaz
[1] 0.02290468
> elambdan=(nA+DA)/(xA+sA)
> elambdan
[1] 0.02994671
> epsi=elambdaz/elambdan
> epsi
[1] 0.7648479
> za=qnorm(1-alpha/2)
> za
[1] 1.959964
> racine=sqrt((nA+DA)^(-1)+(NB+DB)^(-1))
> W=log((psi*lambdaa*sB+NB)/(xB+sB))-log((lambdaa*sA+nA)/(xA+sA))
> sigma=sqrt(psi*lambdaa*sB/(psi*lambdaa*sB+NB)^2+(lambdaa*sA)/(lambdaa*sA+nA))
> puissance=1-pnorm(za*racine,W,sigma)+pnorm(-za*racine,W,sigma)
> puissance
[1] 0.7797311
~ |
```

```
> #Programme 2:
> # Pour scénario (2)
> alpha=0.05;lambdaa=0.0314;psi=0.75;xA=3896;xB=3943;NB=89;
> nA=118;sA=815;sB=815
> elambda=(nA+NB)/(xA+xB)
> elambda
[1] 0.02640643
> DA= sA*elambda
> DA
[1] 21.52124
> DB=sB*elambda
> DB
[1] 21.52124
> elambdaz=(NB+DB)/(xB+sB)
> elambdaz
[1] 0.02322851
> elambdan=(nA+DA)/(xA+sA)
> elambdan
[1] 0.02961606
> epsi=elambdaz/elambdan
> epsi
[1] 0.7843214
> za=qnorm(1-alpha/2)
> za
[1] 1.959964
> racine=sqrt((nA+DA)^(-1)+(NB+DB)^(-1))
> W=log((psi*lambdaa*sB+NB)/(xB+sB))-log((lambdaa*sA+nA)/(xA+sA))
> sigma=sqrt(psi*lambdaa*sB/(psi*lambdaa*sB+NB)^2+(lambdaa*sA)/(lambdaa*sA+nA))
> puissance=1-pnorm(za*racine,W,sigma)+pnorm(-za*racine,W,sigma)
> puissance
[1] 0.7906049
>
< |
```

```
> #Programme 3:
> # Pour scénario (3)
> alpha=0.05;lambdaa=0.0314;psi=0.75;xA=3896;xB=3943;NB=89;
> nA=118;sA=3138;sB=3139
> elambda=(nA+NB)/(xA+xB)
> elambda
[1] 0.02640643
> DA= sA*elambda
> DA
[1] 82.86338
> DB=sB*elambda
> DB
[1] 82.88978
> elambdaz=(NB+DB)/(xB+sB)
> elambdaz
[1] 0.02427136
> elambdan=(nA+DA)/(xA+sA)
> elambdan
[1] 0.02855607
> epsi=elambdaz/elambdan
> epsi
[1] 0.8499546
> za=qnorm(1-alpha/2)
> za
[1] 1.959964
> racine=sqrt((nA+DA)^(-1)+(NB+DB)^(-1))
> W=log((psi*lambdaa*sB+NB)/(xB+sB))-log((lambdaa*sA+nA)/(xA+sA))
> sigma=sqrt(psi*lambdaa*sB/(psi*lambdaa*sB+NB)^2+(lambdaa*sA)/(lambdaa*sA+nA)^2)
> puissance=1-pnorm(za*racine,W,sigma)+pnorm(-za*racine,W,sigma)
> puissance
[1] 0.894965
>
\ |
```

```
> #Programme 4:
> # Pour scénario (1)
> alpha=0.05;lambdaa=0.0314;psi=1;xA=3896;xB=3943;NB=89;
> nA=118;sA=375;sB=375
> elambda=(nA+NB)/(xA+xB)
> elambda
[1] 0.02640643
> DA= sA*elambda
> DA
[1] 9.902411
> DB=sB*elambda
> DB
[1] 9.902411
> elambdaz=(NB+DB)/(xB+sB)
> elambdaz
[1] 0.02290468
> elambdan=(nA+DA)/(xA+sA)
> elambdan
[1] 0.02994671
> epsi=elambdaz/elambdan
> epsi
[1] 0.7648479
> za=qnorm(1-alpha/2)
> za
[1] 1.959964
> racine=sqrt((nA+DA)^(-1)+(NB+DB)^(-1))
> W=log((psi*lambdaa*sB+NB)/(xB+sB))-log((lambdaa*sA+nA)/(xA+sA))
> sigma=sqrt(psi*lambdaa*sB/(psi*lambdaa*sB+NB)^2+(lambdaa*sA)/(lambdaa*sA+nA)^2)
> risq=1-pnorm(za*racine,W,sigma)+pnorm(-za*racine,W,sigma)
> risq
[1] 0.513095
>
. |
```

```
> #Programme 5:
> # Pour scénario (2)
> alpha=0.05;lambdaa=0.0314;psi=1;xA=3896;xB=3943;NB=89;
> nA=118;sA=815;sB=815
> elambda=(nA+NB)/(xA+xB)
> elambda
[1] 0.02640643
> DA= sA*elambda
> DA
[1] 21.52124
> DB=sB*elambda
> DB
[1] 21.52124
> elabdaz=(NB+DB)/(xB+sB)
> elabdaz
[1] 0.02322851
> elambdan=(nA+DA)/(xA+sA)
> elambdan
[1] 0.02961606
> epsi=elabdaz/elambdan
> epsi
[1] 0.7843214
> za=qnorm(1-alpha/2)
> za
[1] 1.959964
> racine=sqrt((nA+DA)^(-1)+(NB+DB)^(-1))
> W=log((psi*lambdaa*sB+NB)/(xB+sB))-log((lambdaa*sA+nA)/(xA+sA))
> sigma=sqrt(psi*lambdaa*sB/(psi*lambdaa*sB+NB)^2+(lambdaa*sA)/(lambdaa*sA+nA)^2)
> risq=1-pnorm(za*racine,W,sigma)+pnorm(-za*racine,W,sigma)
> risq
[1] 0.4017444
> |
```

```
> #Programme 6:
> # Pour scénario (3)
> alpha=0.05;lambdAA=0.0314;psi=1;xA=3896;xB=3943;NB=89;
> nA=118;sA=3138;sB=3139
> elambda=(nA+NB)/(xA+xB)
> elambda
[1] 0.02640643
> DA= sA*elambda
> DA
[1] 82.86338
> DB=sB*elambda
> DB
[1] 82.88978
> elambdaZ=(NB+DB)/(xB+sB)
> elambdaZ
[1] 0.02427136
> elambdaN=(nA+DA)/(xA+sA)
> elambdaN
[1] 0.02855607
> epsi=elambdaZ/elambdaN
> epsi
[1] 0.8499546
> za=qnorm(1-alpha/2)
> za
[1] 1.959964
> racine=sqrt((nA+DA)^(-1)+(NB+DB)^(-1))
> W=log((psi*lambdAA*sB+NB)/(xB+sB))-log((lambdAA*sA+nA)/(xA+sA))
> sigma=sqrt(psi*lambdAA*sB/(psi*lambdAA*sB+NB)^2+(lambdAA*sA)/(lambdAA*sA+nA)^2)
> risq=1-pnorm(za*racine,W,sigma)+pnorm(-za*racine,W,sigma)
> risq
[1] 0.2235817
> ,
```

```
> #Programme 7:
> #Pour scénario 1
> alpha=0.05;lambdAA=0.0314;psi=0.75;xA=3896;xB=3943;NB=89;
> nA=118;sA=375;sB=375;beta=0.2
> elambda=(nA+NB)/(xA+xB)
> elambda
[1] 0.02640643
> DA= sA*elambda
> DA
[1] 9.902411
> DB=sB*elambda
> DB
[1] 9.902411
> elambdaZ=(NB+DB)/(xB+sB)
> elambdaZ
[1] 0.02290468
> elambdaN=(nA+DA)/(xA+sA)
> elambdaN
[1] 0.02994671
> epsi=elambdaZ/elambdaN
> epsi
[1] 0.7648479
> za=qnorm(1-alpha/2)
> za
[1] 1.959964
> IK=(qnorm(1-alpha/2)+qnorm(1-beta))^2/log(epsi)^2
> Ik=(nA+DA)^(-1)+(NB+DB)^(-1)
> puispre=pnorm((-log(epsi)*sqrt(Ik*IK)-za*sqrt(Ik))/sqrt(IK-Ik))+
+ pnorm((log(epsi)*sqrt(Ik*IK)-za*sqrt(Ik))/sqrt(IK-Ik))
> puispre
[1] 0.9799765
> |
```

```
> #Programme 8:
> #Pour scénario 2
> alpha=0.05;lambdAA=0.0314;psi=0.75;xA=3896;xB=3943;NB=89;
> nA=118;sA=815;sB=815;beta=0.2
> elambda=(nA+NB)/(xA+xB)
> elambda
[1] 0.02640643
> DA= sA*elambda
> DA
[1] 21.52124
> DB=sB*elambda
> DB
[1] 21.52124
> elambdaZ=(NB+DB)/(xB+sB)
> elambdaZ
[1] 0.02322851
> elambdaN=(nA+DA)/(xA+sA)
> elambdaN
[1] 0.02961606
> epsi=elambdaZ/elambdaN
> epsi
[1] 0.7843214
> za=qnorm(1-alpha/2)
> za
[1] 1.959964
> IK=(qnorm(1-alpha/2)+qnorm(1-beta))^2/log(epsi)^2
> Ik=(nA+DA)^(-1)+(NB+DB)^(-1)
> puispre=pnorm((-log(epsi)*sqrt(Ik*Ik)-za*sqrt(Ik))/sqrt(IK-Ik))+
+ pnorm((log(epsi)*sqrt(Ik*Ik)-za*sqrt(Ik))/sqrt(IK-Ik))
> puispre
[1] 0.9827406
> |
```

```
>
> #Programme 9:
> #Pour scénario 3
> alpha=0.05;lambdaa=0.0314;psi=0.75;xA=3896;xB=3943;NB=89;
> nA=118;sA=3138;sB=3139;beta=0.2
> elambda=(nA+NB)/(xA+xB)
> elambda
[1] 0.02640643
> DA= sA*elambda
> DA
[1] 82.86338
> DB=sB*elambda
> DB
[1] 82.88978
> elambdaz=(NB+DB)/(xB+sB)
> elambdaz
[1] 0.02427136
> elambdan=(nA+DA)/(xA+sA)
> elambdan
[1] 0.02855607
> epsi=elambdaz/elambdan
> epsi
[1] 0.8499546
> za=qnorm(1-alpha/2)
> za
[1] 1.959964
> IK=(qnorm(1-alpha/2)+qnorm(1-beta))^2/log(epsi)^2
> Ik=(nA+DA)^(-1)+(NB+DB)^(-1)
> puispre=pnorm((-log(epsi)*sqrt(Ik*IK)-za*sqrt(Ik))/sqrt(IK-Ik))+
+ pnorm((log(epsi)*sqrt(Ik*IK)-za*sqrt(Ik))/sqrt(IK-Ik))
> puispre
[1] 0.9905724
> |
```

Bibliographie

- [1] Adkinson, N.F., Eggleston, P.A., Eney, D., Goldstein, E.O., Schubert, K.C., Bacon, J.R., Hamilton, R.G., Weiss, M.E., Arshand, H., Meinert, C.L., Tonascra, J. and Wheeler, B. (1997). *A controlled trial of immunotherapy for asthma in allergic children*, *New Engl. J. Med.* 336 : 324-331.
- [2] Andersen, P.K. (1987). *Conditional power calculations as an aid in the decision whether to continue a clinical trial*. *Contr. Clin. Trials*, 8, 67-74.
- [3] Berger, J.O. and Berry, D.A. (1988). *Statistical analysis and the illusion of objectivity*. *Amer. Scientist*, 76, 159-165.
- [4] Berry, D.A. and Fristedt, B. (1985). *Bandit Problems : Sequential Allocation of Experiments*. London : Chapman and Hall.
- [5] Berry, D.A. (1987). *Interim analyses in clinical trials : The role of the likelihood principle*. *The Amer. Statistician*, 41, 117-122.
- [6] Bouvenot G. and vray M. (2006). *Essais cliniques (théorie, pratique et critique) Médecine Sciences Publications*. Paris, Lavoisier SAS ; ISBN : 978-2-257-14224-5.
- [7] Brian S. E. and Andrw P. (1999). *Statistical Aspects Of The Design And Analysis Of Clinical Trials*. Imperial College Press London ; ISBN : 1-86094-153-2.
- [8] C. Jennison and Bruce W. Turnbull. (2000). *Group Sequential Methods with Applications to Clinical Trials*. Chapman Hall/CRC USA ; ISBN 0-8493-0316-8.
- [9] Chow S. C, et Chang M. (2007). *Adaptative design methods in clinical trials*. Chapman et Hall/CRC Biostatistics Series, USA.
- [10] Cleophas T.J, Zwinderman A. H, and Cleophas T. F. (2006). *Statistics Applied to Clinical Trials, Third Edition*. Springer Pharmaceutical Medicine Lyon, France.
- [11] Choi, S.C., Smith, P.J. and Becker, D.P. (1985). *Early decision in clinical trials when treatment differences are small*. *Contr. Clin. Trials*, 6, 280-288.

-
- [12] Cornfield, J. (1966a). *Sequential trials, sequential analysis and the likelihood principle*. The Amer. Statistician, 20, 18-23.
- [13] Cornfield, J. (1966b). *A Bayesian test of some classical hypotheses - with applications to sequential clinical trials*. J. Amer. Statist. Assoc., 61, 577-594.
- [14] David J. S, Keith R., Jonathan P. Myles. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. England; ISBN 0-471-49975-7.
- [15] De Martin, D. (2013). *Success Probability Estimation With Applications to Clinical Trials*. USA; ISBN : 978-1-118-33578-9.
- [16] Dickey, J. (1973). *Scientific reporting and personal probabilities : Student's hypothesis*. J. Roy. Statist. Soc. B., 35, 285-305.
- [17] Freedmain, L.S., Spiegelhalter, D.J. and Parmar, M.K.B. (1994). *The what, why and how of Bayesian Clinical trials monitoring*. Statist. Med. 13 : 1371-1384.
- [18] Freedmain, L.S. and Spiegelhalter, D.J. (1989). *Comparison of Bayesian with group sequential for monitoring clinical trials*. Contr. Clin. Trials, 10, 357-367.
- [19] Fayers, P.M., Ashby, D. and Parmar, M.B. (1997). *Bayesian data monitoring in clinical trials*. Statist. Med., 16, 1413-1430.
- [20] Freedmain, L.S. and Spiegelhalter, D.J. (1983). *The assessment of subjective opinion and its use in relation to stopping rules for clinical trials*. The Statistician, 32, 153- 160.
- [21] George, S.L. and Desu, M.M. (1974). *Planning the size and duration of a clinical trial studying the time to some critical event*. J. Chron. Diseases, 27, 15-24.
- [22] Herson, J. (1979). *Predictive probability early termination for phase II clinical trials*. Biometrics, 35, 775-783.
- [23] Hughes M.D. and Pocock, S.J. (1988). *Stopping rules and estimation problems in clinical trials*. Statist. Med., 7, 1231-1242.
- [24] Kass, R.E. and Greenhouse, J.B. (1989). *Comments on Investigating therapies of potentially great benefit : ECMO (by J.H. Ware)*. Statistical Science 4, 310-317.
- [25] Lachin, J.M. (1981). *Sequential clinical trials for normal variates using interval composite hypotheses*. Biometrics, 37, 87-101.
- [26] Lan, K.K.G., Simon, R. and Halperin, M. (1982). *Stochastically curtailed tests in long-term clinical trials*. Commun. Statist. C, 1, 207-219.

- [27] Lindley D.V. (1965). *Introduction to Probability and Statistics. from a Bayesian Viewpoint*; Part 2, Inference., Cambridge : Cambridge Univ. Press.
- [28] Machin, D. (1994). *Discussions of The What, Why and How of Bayesian Clinical Trials Monitoring*. Statist. Med. 13 : 1385-1390.
- [29] Mehta, C.R. and Cain, K.C. (1984). *Charts for early stopping of pilot studies*. J. Clinical Oncology, 2, 676-682.
- [30] Makuch, R. and Simon, R. (1982). *Sample size requirements for comparing time-tofailure among k treatment groups*. J. Chronic Diseases, 35, 861-867.
- [31] O'Brien, P.C. and Fleming, T.R. (1979). *A multiple testing procedure for clinical trials*. Biometrics, 35, 549-556.
- [32] Robert C. P. (2006). *Le choix bayésien : principes et pratique*. Springer-Verlag France, Paris.
- [33] Rosenbaum, P.R. and Rubin, D.B. (1984). *Sensitivity of Bayes inference with data dependent sampling rules*. The Amer. Statistician, 38, 106-109.
- [34] Société canadienne du cancer. (2007). *Les essais cliniques : Guide à l'intention des personnes atteintes du cancer*. Le cancer : une lutte à finir. 1888 939-3333 [www.cancer.ca].
- [35] Spiegelhalter, D.J. and Freedmain, L.S. (1986). *A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion*. Statist. Med., 5, 1-13.
- [36] Spiegelhalter, D.J., Freedmain, L.S. and Parmar, M.K.B. (1994). *Bayesian approaches to clinical trials (with discussion)*. J. Roy. Statist. Soc. A., 157, 357-416.
- [37] Schoenfeld, D.A. and Richter, J.R. (1982). *Nomograms for calculating the number of patients needed for a clinical trial with survival as an endpoint*. Biometrics, 38, 163-170.
- [38] Shewhart, W.A. (1931). *Economic Control of Manufactured Product*. Van Nostrand, New York.
- [39] Ware, J.H., Muller, J.E. and Braunwald, E. (1985). *The futility index : An approach to the cost-effective termination of randomized clinical trials*. Amer. J. Med., 78, 635-643.
- [40] Wingen, A.M., FabianBach, C., Schaefer, F. and Mehls, O. (1997). *Randomized multicentre study of low-protein diet on the progression of chronic renal failure in children*. Lancet 349 : 1117-1123.
- [41] Zerari A. (2013). *Approche Bayésienne aux essais cliniques séquentiels*. Magister. Constantine.