

PATRICK GAGNON

Intervalles de confiance pour une différence de deux proportions

Mémoire présenté
à la Faculté des études supérieures de l'Université Laval
dans le cadre du programme de maîtrise en statistique
pour l'obtention du grade de maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES ET DE GÉNIE
UNIVERSITÉ LAVAL
QUÉBEC

Décembre 2006

©Patrick Gagnon, 2006

Résumé

L'intervalle de confiance le plus connu pour une différence de deux proportions est l'intervalle de Wald. Cet intervalle a l'avantage d'être simple à construire, mais il est anti-conservateur. Il existe plusieurs intervalles alternatifs à l'intervalle de Wald qui performant beaucoup mieux. Dans ce mémoire, on s'intéressera particulièrement à l'intervalle d'Agresti-Coull et à l'intervalle bayésien approximatif. Ces intervalles performant très bien tout en étant simples à construire. On regardera d'abord la performance de ces intervalles lorsqu'on a deux échantillons indépendants de tailles fixées au départ. On regardera aussi leur performance lorsque le nombre d'observations dépend des vraies proportions, soit dans une expérience à étapes multiples, soit dans une expérience à allocations séquentielles où un plan adaptatif est utilisé.

Avant-Propos

J'aimerais d'abord remercier mon directeur, monsieur Claude Bélisle, qui m'a donné de judicieux conseils lors de la rédaction de ce mémoire et sans qui je n'aurais jamais eu l'idée d'aller passer cinq mois au pays des Kiwis. Je garde d'excellents souvenirs de ce voyage grâce à Jean-Hubert Smith-Lacroix et Nathalie Savard, avec qui j'ai séjourné dans une petite maison rose à l'ouest d'Auckland.

Mon passage de cinq ans au Département de mathématiques et de statistique a été très enrichissant tant sur le plan professionnel que sur le plan personnel. Je me considère chanceux d'avoir travaillé au Service de consultation statistique, au Centre de dépannage et d'apprentissage en mathématique et statistique (CDA) et comme auxiliaire d'enseignement pour deux cours de statistique. Ces expériences ont été très formatrices. Sur le plan personnel, il y a eu plusieurs moments inoubliables au cours de ces cinq années. Je garderai d'excellents souvenirs de tous les camps, des Vachon Partys (à l'époque où il y avait plus de 10 personnes, il y a très longtemps), des 5 à ...

Et bien sûr, je remercie mes parents de qui j'ai toujours reçu l'appui dont j'ai eu besoin.

Table des matières

Résumé	ii
Avant-Propos	iii
Table des matières	v
Liste des tableaux	vii
Table des figures	viii
1 Introduction	1
2 Intervalle de confiance pour une proportion	5
2.1 L'intervalle de Wald	5
2.1.1 Raisons des faibles probabilités de recouvrement de l'intervalle de Wald	9
2.2 Intervalles alternatifs	11
2.2.1 L'intervalle de Wald avec correction pour la continuité	11
2.2.2 L'intervalle de Wilson	12
2.2.3 L'intervalle d'Agresti-Coull	13
2.2.4 L'intervalle bayésien approximatif	14
2.2.5 L'intervalle du rapport des vraisemblances	15
2.3 Comparaison des intervalles	16
2.3.1 Probabilités de recouvrement des intervalles	16
2.3.2 Longueurs des intervalles	18
2.4 Conclusion	20
3 Intervalle de confiance pour une différence de deux proportions	21
3.1 L'intervalle de Wald	21
3.1.1 Raisons des faibles probabilités de recouvrement de l'intervalle de Wald	22
3.2 Intervalles alternatifs	23
3.2.1 L'intervalle de Wald avec correction pour la continuité	24

3.2.2	L'intervalle d'Agresti-Coull	24
3.2.3	L'intervalle bayésien approximatif	25
3.3	Comparaison des intervalles	25
3.3.1	Probabilités de recouvrement des intervalles	26
3.3.2	Longueurs des intervalles	28
3.4	Conclusion	29
4	Règle d'arrêt d'une expérience	31
4.1	Expérience à taille fixe	31
4.2	Expérience à étapes multiples	37
4.2.1	Tests répétés	38
4.2.2	Intervalles de confiance répétés	45
4.3	Conclusion	53
5	Plans adaptatifs dans une expérience à allocations séquentielles	54
5.1	Hypothèses et notation	55
5.2	Critères d'évaluation des plans d'allocation	57
5.2.1	Inférence statistique	57
5.2.2	Biais de sélection	58
5.2.3	Taux de succès dans l'expérience	59
5.3	Plans d'allocation adaptatifs	60
5.3.1	Allocations dépendant des assignations antérieures	60
5.3.2	Allocations dépendant des assignations et des réponses antérieures	62
5.4	Comparaison des plans	66
5.4.1	Biais de \hat{p}_1 et \hat{p}_2	68
5.4.2	Biais de sélection	68
5.4.3	Taux de succès	71
5.4.4	Estimation par intervalle de confiance de $p_1 - p_2$	71
5.4.5	Puissance	74
5.4.6	Puissance et probabilité de recouvrement des intervalles de confiance répétés	76
5.5	Conclusion	80
6	Conclusion	82
	Bibliographie	84

Liste des tableaux

2.1	Statistiques descriptives des probabilités de recouvrement des six intervalles étudiés de niveau 95 % pour 90 couples (n, p)	18
4.1	Probabilités de recouvrement et puissances lorsque n provient de l'équation (4.10) et la puissance désirée est 0,8.	35
4.2	Probabilités de recouvrement et puissances lorsque n est obtenu par l'équation (4.10) en prenant V_0 de l'équation (4.12).	36
4.3	Probabilités de recouvrement et puissances exactes lorsque n est obtenu par l'équation (4.10) en prenant V_0 de l'équation (4.12) et V_1 de l'équation (4.14).	37
4.4	Puissance pour différents K des tests de Pocock.	41
4.5	Puissance pour différents K des tests de O'Brien-Fleming.	44
4.6	Puissance pour différents K des tests indépendants.	46
4.7	Probabilités de recouvrement des intervalles de Wald, AC et BA de niveau 95 % construits après une expérience où on a utilisé les tests de Pocock de niveau 95 %.	48
4.8	Probabilités de recouvrement des intervalles de Wald, AC et BA de niveau 95 % construits après une expérience où on a utilisé les tests de O'Brien-Fleming de niveau 95 %.	49
4.9	Probabilités de recouvrement et puissances des intervalles de Wald, AC et BA où les α_k sont ceux des tests de Pocock de niveau 95 %.	50
4.10	Probabilités de recouvrement et puissances des intervalles de Wald, AC et BA où les α_k sont ceux des tests de O'Brien-Fleming de niveau 95 %.	51
5.1	Moyenne des 10 000 biais de \hat{p}_1 et \hat{p}_2 pour les plans étudiés pour différentes valeurs de (p_1, p_2) lorsque $n = 30$	69
5.2	Valeurs moyennes de $\frac{\lambda_A}{n_1} + \frac{\lambda_B}{n_2}$ pour les plans étudiés pour différentes valeurs de (p_1, p_2) lorsque $n = 30$	70
5.3	Valeurs moyennes des taux de succès pour les plans étudiés pour différentes valeurs de (p_1, p_2) lorsque $n = 30$	71
5.4	Probabilités de recouvrement de l'intervalle de Wald de niveau 95 % pour les plans étudiés pour différentes valeurs de (p_1, p_2) lorsque $n = 30$	72

5.5	Probabilités de recouvrement de l'intervalle AC de niveau 95 % pour les plans étudiés pour différentes valeurs de (p_1, p_2) lorsque $n = 30$	73
5.6	Probabilités de recouvrement de l'intervalle BA de niveau 95 % pour les plans étudiés pour différentes valeurs de (p_1, p_2) lorsque $n = 30$	73
5.7	Puissances de l'intervalle AC de niveau 95 % pour les plans étudiés pour différentes valeurs de (p_1, p_2) lorsque la puissance désirée est 0,75.	75
5.8	Puissances de l'intervalle BA de niveau 95 % pour les plans étudiés pour différentes valeurs de (p_1, p_2) lorsque la puissance désirée est 0,75.	75
5.9	Probabilités de recouvrement de l'intervalle AC pour les plans étudiés lorsque $p_1 = p_2$ où les α_k sont ceux des tests de Pocock de niveau 95 % avec $K = 2$	77
5.10	Probabilités de recouvrement de l'intervalle AC pour les plans étudiés lorsque $p_1 = p_2$ où les α_k sont ceux des tests de Pocock de niveau 95 % avec $K = 5$	77
5.11	Probabilités de recouvrement de l'intervalle AC pour les plans étudiés lorsque $p_1 \neq p_2$ où les α_k sont ceux des tests de Pocock de niveau 95 % avec $K = 2$	78
5.12	Probabilités de recouvrement de l'intervalle AC pour les plans étudiés lorsque $p_1 \neq p_2$ où les α_k sont ceux des tests de Pocock de niveau 95 % avec $K = 5$	78
5.13	Puissances de l'intervalle AC pour les plans étudiés où les α_k sont ceux des tests de Pocock de niveau 95 % avec $K = 2$	79
5.14	Puissances de l'intervalle AC pour les plans étudiés où les α_k sont ceux des tests de Pocock de niveau 95 % avec $K = 5$	79

Table des figures

2.1	Probabilité de recouvrement de l'intervalle de Wald de niveau 95 % lorsque $n = 20$	8
2.2	Probabilité de recouvrement de l'intervalle de Wald de niveau 95 % lorsque $n = 30$	9
2.3	Probabilité de recouvrement de l'intervalle de Wald de niveau 95 % lorsque $p = 0, 200$	9
2.4	Probabilités de recouvrement des six intervalles étudiés de niveau 95 % lorsque $n = 20$	17
2.5	Diagrammes en boîte des probabilités de recouvrement des six intervalles étudiés de niveau 95 % pour 90 couples (n, p)	18
2.6	Longueurs espérées de cinq intervalles de niveau 95 % lorsque $n = 20$. .	19
2.7	Longueurs espérées de cinq intervalles de niveau 95 % lorsque $n = 100$.	19
3.1	Probabilités de recouvrement des quatre intervalles étudiés de niveau 95 % en fonction de p_1 lorsque $n_1 = n_2 = 20$ et $p_2 = 0, 5$	26
3.2	Probabilités de recouvrement des quatre intervalles étudiés de niveau 95 % en fonction de p_1 lorsque $n_1 = n_2 = 20$ et $p_2 = 0, 2$	27
3.3	Probabilités de recouvrement des quatre intervalles étudiés de niveau 95 % en fonction de p_1 lorsque $n_1 = n_2 = 20$ et $p_1 - p_2 = 0, 25$	27
3.4	Probabilités de recouvrement des quatre intervalles étudiés de niveau 95 % en fonction de n_1 lorsque $n_2 = 30$, $p_1 = 0, 5$ et $p_2 = 0, 2$	27
3.5	Diagrammes en boîte des probabilités de recouvrement des quatre intervalles étudiés de niveau 95 % pour 330 valeurs de (n_1, p_1, n_2, p_2)	28
3.6	Longueurs espérées des quatre intervalles étudiés de niveau 95 % en fonction de p_1 lorsque $n_1 = n_2 = 20$ et $p_2 = 0, 5$	29
3.7	Longueurs espérées des quatre intervalles étudiés de niveau 95 % en fonction de p_1 lorsque $n_1 = n_2 = 20$ et $p_2 = 0, 2$	29
3.8	Longueurs espérées des quatre intervalles étudiés de niveau 95 % en fonction de p_1 lorsque $n_1 = n_2 = 20$ et $p_1 - p_2 = 0, 25$	29
3.9	Longueurs espérées des quatre intervalles étudiés de niveau 95 % en fonction de n_1 lorsque $n_2 = 30$, $p_1 = 0, 5$ et $p_2 = 0, 2$	29

Chapitre 1

Introduction

Dans bien des applications de la statistique, on est intéressé à estimer la valeur d'un certain paramètre, disons θ , à partir de données recueillies. À défaut d'être capable de l'estimer avec exactitude, on l'estime généralement par intervalle. Dans ce qui suit, $\tilde{X} = (X_1, \dots, X_n)$ sera un vecteur de variables aléatoires prenant ses valeurs dans \mathcal{X} . Les définitions présentées ci-dessous sont celles de Casella et Berger (2002).

Définition 1. *Supposons que $\tilde{X} = \tilde{x}$ soit observé. Une estimation par intervalle d'un paramètre θ est n'importe quelle paire de fonctions $L(\tilde{x})$ et $U(\tilde{x})$ satisfaisant $L(\tilde{x}) \leq U(\tilde{x})$ pour tout $\tilde{x} \in \mathcal{X}$. On appellera estimateur par intervalle l'intervalle aléatoire $(L(\tilde{X}), U(\tilde{X}))$.*

On a supposé dans la définition précédente que l'intervalle était ouvert, mais rien n'empêche qu'il soit fermé ou demi-ouvert. Rien n'empêche aussi que l'intervalle soit *unilatéral*, c'est-à-dire que $L(\tilde{X}) = -\infty$ ou $U(\tilde{X}) = +\infty$. Autrement, l'intervalle est *bilatéral*. Dans ce mémoire, on s'intéressera uniquement aux intervalles bilatéraux.

Passons maintenant à la définition de probabilité de recouvrement.

Définition 2. *Pour un estimateur par intervalle $(L(\tilde{X}), U(\tilde{X}))$ d'un paramètre θ , la probabilité de recouvrement de $(L(\tilde{X}), U(\tilde{X}))$ est donnée par l'équation suivante :*

$$\mathbb{P}_\theta \left[\theta \in (L(\tilde{X}), U(\tilde{X})) \right]. \quad (1.1)$$

La probabilité de recouvrement est donc la probabilité que le paramètre θ soit dans l'estimateur par intervalle. Dans certains cas, cette probabilité sera la même peu importe

la valeur de θ . Par exemple, supposons que \tilde{X} soit un vecteur de variables aléatoires indépendantes et identiquement distribuées de la loi Normale(θ, σ^2) avec σ^2 connu. Soit z_γ le quantile d'ordre $(1 - \gamma)$ de la loi Normale(0,1) et $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$. C'est bien connu que si $(L(\tilde{X}), U(\tilde{X})) = (\bar{X} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{X} + z_{\alpha/2}\sigma/\sqrt{n})$, alors la probabilité de recouvrement sera $(1 - \alpha)$ peu importe la valeur de θ .

Pour les cas où la probabilité de recouvrement dépend de θ , il est particulièrement important de définir le coefficient de confiance.

Définition 3. Pour un estimateur par intervalle $(L(\tilde{X}), U(\tilde{X}))$ d'un paramètre θ , le coefficient de confiance de $(L(\tilde{X}), U(\tilde{X}))$ est donné par l'équation suivante :

$$\inf_{\theta} \mathbb{P}_{\theta} \left[\theta \in (L(\tilde{X}), U(\tilde{X})) \right]. \quad (1.2)$$

Dans le cas où la probabilité de recouvrement ne dépend pas de θ , il n'y a pas de différence entre probabilité de recouvrement et coefficient de confiance. Passons maintenant à la définition théorique d'un intervalle de confiance.

Définition 4. On appelle intervalle de confiance de niveau $(1 - \alpha)$ un estimateur par intervalle ayant un coefficient de confiance égal à $(1 - \alpha)$.

En pratique, il arrivera qu'un intervalle de confiance ait un coefficient de confiance approximativement égal à $(1 - \alpha)$. C'est le cas lorsque l'intervalle est construit à partir d'un résultat asymptotique. Par exemple, lorsque \tilde{X} contient des variables aléatoires indépendantes et identiquement distribuées d'une loi quelconque avec variance σ^2 connue et $\theta \equiv \mathbb{E}[X_1]$ inconnu, on pourra dire, grâce au théorème limite central, que le coefficient de confiance de l'intervalle $(\bar{X} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{X} + z_{\alpha/2}\sigma/\sqrt{n})$ sera approximativement égal à $(1 - \alpha)$ si n est assez grand. On appellera *niveau de confiance nominal* la valeur désirée du coefficient de confiance. En pratique, un intervalle de confiance de niveau $(1 - \alpha)$ est un estimateur par intervalle pour lequel le niveau de confiance nominal est $(1 - \alpha)$. On utilisera cette dernière définition dans ce mémoire.

Dans ce mémoire, le paramètre qui nous intéressera particulièrement est $\Delta \equiv p_1 - p_2$ où p_1 et p_2 sont des probabilités de succès. L'information sur p_1 et p_2 sera recueillie à partir d'observations X_1, \dots, X_{n_1} indépendantes et identiquement distribuées de la loi Bernoulli(p_1) et Y_1, \dots, Y_{n_2} indépendantes et identiquement distribuées de la loi Bernoulli(p_2). On a aussi que les variables X_1, \dots, X_{n_1} sont indépendantes des variables Y_1, \dots, Y_{n_2} . Un intervalle de confiance grandement utilisé pour Δ est l'intervalle de

Wald. Cet intervalle est l'extension pour une différence de deux proportions de l'intervalle de Wald pour une seule proportion qui est lui aussi très utilisé et souvent présenté dans les cours de base en statistique. Bien que ces intervalles soient très utilisés, ils ne performant pas très bien. Leurs probabilités de recouvrement sont généralement trop petites surtout lorsque n n'est pas très grand.

Ce problème est bien connu et plusieurs autres intervalles corrigent les défauts de l'intervalle de Wald pour une proportion et pour la différence de deux proportions. Par exemple, Newcombe (1998a) présente six autres intervalles pour une proportion et Newcombe (1998b) présente dix autres intervalles pour une différence de deux proportions. Certains de ces intervalles ont une forme très compliquée, d'autres n'ont même pas de forme explicite, mais certains ont une forme aussi simple que les intervalles de Wald. On s'intéressera particulièrement à ces derniers car un intervalle a plus de chances d'être utilisé à grande échelle s'il n'a pas une forme trop compliquée. C'est probablement pourquoi les intervalles de Wald sont encore beaucoup utilisés. On verra qu'il existe d'autres intervalles pratiquement aussi simples à construire que les intervalles de Wald, mais qui performant beaucoup mieux.

Les intervalles étudiés pour une différence de deux proportions seront des extensions d'intervalles pour une proportion. C'est pourquoi le chapitre 2 sera consacré aux intervalles de confiance pour une proportion et le chapitre 3 aux intervalles de confiance pour le paramètre qui nous intéresse vraiment, la différence de deux proportions. Dans les deux chapitres, n_1 et n_2 seront des constantes.

Par la suite, on supposera que p_1 et p_2 sont des taux de succès de deux traitements médicaux et que les observations sont recueillies pour trouver quel est le meilleur traitement. Ici, une observation est la réponse d'un patient à un des deux traitements. Un nouvel élément à considérer est le choix de n_1 et n_2 . On les voudra assez grands pour avoir assez d'information, mais pas trop pour ne pas faire durer l'expérience trop longtemps inutilement. Le chapitre 4 sera consacré aux règles d'arrêt d'une telle expérience. On regardera entre autre comment performant les intervalles vus au chapitre 3 dans une expérience à étapes multiples. Dans une expérience à étapes multiples, le moment où l'on arrête l'expérience dépend de la différence entre les performances observées des deux traitements. Cela veut dire que n_1 et n_2 sont des variables aléatoires dépendant de p_1 et p_2 et cela implique que les estimations de p_1 et p_2 peuvent être biaisées. Cela rend l'estimation par intervalle de confiance difficile.

Supposons que les patients arrivent un à la fois dans l'expérience et que toutes les observations antérieures sont connues au moment de décider lequel des deux traitements sera administré au présent patient. L'administration d'un traitement peut donc

dépendre de la performance observée jusqu'à maintenant des deux traitements. On pourra ainsi assigner plus d'individus au traitement qui performe le mieux. Cela veut dire que n_1 et n_2 sont des variables aléatoires dépendant de p_1 et p_2 et cela implique encore que les estimations de p_1 et p_2 peuvent être biaisées. Au chapitre 5, on comparera, par rapport à divers critères, différents plans d'allocations adaptatifs, c'est-à-dire des plans où la probabilité d'administrer un traitement à un patient dépend des assignations antérieures, et on examinera la performance des intervalles vus au chapitre 3 lors de l'utilisation de tels plans.

Chapitre 2

Intervalle de confiance pour une proportion

Le présent chapitre concerne l'estimation par intervalle de confiance d'une proportion p à partir d'observations X_1, \dots, X_n indépendantes et identiquement distribuées Bernoulli(p), ce qui implique que $\sum_{i=1}^n X_i \sim \text{Binomiale}(n, p)$. Tout au long du mémoire, l'estimateur du maximum de vraisemblance de p , $\frac{1}{n} \sum_{i=1}^n X_i$, sera noté \hat{p} , à moins d'avis contraire.

2.1 L'intervalle de Wald

L'intervalle de confiance approximatif le plus connu pour p est l'intervalle de Wald. Cet intervalle découle du théorème limite central (TLC). Ce théorème dit qu'avec des observations Y_1, Y_2, Y_3, \dots indépendantes et identiquement distribuées avec $\mathbb{E}[Y_1] = \mu < \infty$ et $0 < \text{Var}[Y_1] = \sigma^2 < \infty$, on a

$$\mathcal{L} \left(\frac{\frac{1}{n} \sum_{i=1}^n Y_i - \mu}{\sigma / \sqrt{n}} \right) \rightarrow N(0, 1) \quad (2.1)$$

lorsque $n \rightarrow \infty$. Soit z_α le quantile d'ordre $1 - \alpha$ de la loi Normale(0,1). L'équation

précédente implique que

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[-z_{\alpha/2} < \frac{\frac{1}{n} \sum_{i=1}^n Y_i - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2} \right] = 1 - \alpha. \quad (2.2)$$

On peut donc obtenir un intervalle de confiance approximatif de niveau $(1 - \alpha)100\%$ pour μ en gardant tous les μ tels que

$$\left| \frac{\frac{1}{n} \sum_{i=1}^n Y_i - \mu}{\sigma/\sqrt{n}} \right| < z_{\alpha/2}. \quad (2.3)$$

Si σ est connu, il est facile de voir qu'on gardera les μ inclus dans l'intervalle

$$\left(\frac{1}{n} \sum_{i=1}^n Y_i \pm z_{\alpha/2} \sigma / \sqrt{n} \right), \quad (2.4)$$

ce qui revient à la forme d'un intervalle de confiance pour μ à partir d'observations indépendantes de la loi Normale(μ, σ^2) avec σ^2 connu.

Dans le cas où Y_1, \dots, Y_n sont indépendantes et identiquement distribuées de la loi Bernoulli(p), $\mu = p$ et $\sigma^2 = p(1 - p)$. L'équation (2.3) devient alors

$$\left| \frac{\frac{1}{n} \sum_{i=1}^n Y_i - p}{\sqrt{p(1-p)/n}} \right| < z_{\alpha/2}. \quad (2.5)$$

Le problème ici est que $\sigma = \sqrt{p(1-p)}$ est inconnu. L'intervalle de confiance de Wald s'obtient en remplaçant σ par $\hat{\sigma} = \sqrt{\hat{p}(1-\hat{p})}$. On garde alors les p tels que

$$\left| \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \right| < z_{\alpha/2}. \quad (2.6)$$

On peut facilement voir que l'intervalle de Wald inclut tous les p dans l'intervalle

$$\left(\hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n} \right). \quad (2.7)$$

On présente ici l'intervalle de Wald comme un intervalle ouvert, mais on aurait aussi bien pu le présenter comme un intervalle fermé.

Il ne faut pas oublier que l'équation (2.2) implique que l'intervalle de Wald a du sens lorsque n est assez grand. On peut donc se demander à partir de quelle valeur de n il est correct d'utiliser cet intervalle. Brown et al. (2001) ont dressé une liste de règles que l'on retrouve dans des livres de base en statistique. Parmi celles-ci, il y a $n \geq 30$, $np(1-p) \geq 5$ et $(n \times \min(p, 1-p)) \geq 10$. En plus de n , on tiendra souvent compte de p puisque la vitesse de convergence dans l'équation (2.1) dépend de p . Cependant, les règles contenant p ne sont pas applicables puisque p est inconnu. Dans la pratique, on remplace p par \hat{p} dans ces règles.

On pourrait penser que pour $np(1-p) \geq 5$, la probabilité de recouvrement de l'intervalle de Wald devrait être près du niveau de confiance nominal et que pour un p fixé, plus n augmente, plus la probabilité de recouvrement s'approche du niveau de confiance nominal, mais ce n'est pas nécessairement le cas. On verra avec les figures qui suivent qu'il existe des couples (n, p) chanceux, c'est-à-dire qu'il y a des couples où n n'est pas grand pour lesquels la probabilité de recouvrement de l'intervalle de Wald est très près du niveau de confiance nominal, et d'autres qui sont malchanceux, c'est-à-dire des couples où n est grand et p n'est pas trop près de 0 ou 1 pour lesquels la probabilité de recouvrement est nettement inférieure au niveau de confiance nominal.

Les prochaines figures montrent des probabilités de recouvrement exactes de l'intervalle de Wald avec niveau de confiance 95 % pour différents n et p . Pour un couple (n, p) donné, la probabilité de recouvrement de l'intervalle de Wald, que l'on notera $W(n, p)$, s'obtient de la façon suivante :

$$\begin{aligned}
 W(n, p) &= \mathbb{P} \left[p \in \left(\hat{p} \pm z_{0,025} \sqrt{\hat{p}(1-\hat{p})/n} \right) \right] \\
 &= \mathbb{P} \left[p \in \left(\frac{Y}{n} \pm z_{0,025} \sqrt{\frac{Y}{n} \left(1 - \frac{Y}{n} \right) / n} \right) \right] \\
 &\quad (\text{où } Y \sim \text{Binomiale}(n, p)) \\
 &= \mathbb{E} [I_W(n, p, Y)] \\
 &= \sum_{j=0}^n I_W(n, p, j) \binom{n}{j} p^j (1-p)^{n-j}, \tag{2.8}
 \end{aligned}$$

où

FIG. 2.1 – Probabilité de recouvrement de l'intervalle de Wald de niveau 95 % lorsque $n = 20$. La ligne hachurée est le niveau de confiance nominal 95 %.

$$I_W(n, p, j) = \begin{cases} 1 & \text{si } p \in \left(\frac{j}{n} \pm z_{0,025} \sqrt{\frac{j}{n} \left(1 - \frac{j}{n}\right) / n} \right) \\ 0 & \text{sinon} \end{cases}$$

pour $0 \leq p \leq 1$ et $n \in \mathbb{N}^*$.

Dans ce chapitre, toutes les probabilités de recouvrement présentées seront des probabilités de recouvrement exactes, calculées selon l'équation (2.8).

La figure 2.1 montre la probabilité de recouvrement de l'intervalle de Wald de niveau 95 % en fonction de p lorsque $n = 20$. On a calculé la probabilité de recouvrement pour $p = 0,001, 0,002, \dots, 0,999$. On voit que la probabilité de recouvrement est nettement trop basse lorsque p est près de 0 ou 1 et qu'elle a tendance à croître vers le niveau de confiance nominal lorsque p s'approche de $1/2$. Cependant, cette montée vers le niveau de confiance nominal ne se fait pas de façon monotone. Cela est dû au fait que l'intervalle pour p est obtenu à partir d'une loi discrète, la loi binomiale. Un exemple de couple chanceux est le couple $n = 20$ et $p = 0,145$, où la probabilité de recouvrement est 0,952. Cela est très bien d'autant plus que $np(1 - p) \cong 2,48 < 5$. Cependant, en prenant $p = 0,146$, la probabilité de recouvrement descend à 0,807, ce qui est nettement sous le niveau de confiance nominal. Pour l'ensemble des p , on voit que l'intervalle de Wald n'est pas assez conservateur, c'est-à-dire que les probabilités de recouvrement sont généralement trop petites. Sur les 999 valeurs de p utilisées, 740 ont une probabilité de recouvrement plus petite que 0,930, plus de la moitié, soit 534, en ont une plus petite que 0,920 et 410 en ont une plus petite que 0,900. Il faut dire que lorsque $n = 20$, $np(1 - p)$ est toujours inférieur ou égal à 5, avec égalité seulement lorsque $p = 1/2$. Regardons maintenant ce qui se passe pour un n un peu plus grand.

La figure 2.2 montre la probabilité de recouvrement de l'intervalle de Wald de niveau 95 % lorsque $n = 30$. Lorsque p est entre les deux lignes verticales, toutes les règles mentionnées précédemment, soit $n \geq 30$, $np(1 - p) \geq 5$ et $(n \times \min(p, 1 - p)) \geq 10$, sont respectées. Malgré cela, pour certaines de ces valeurs, la probabilité de recouvrement est trop faible. Pour les 577 valeurs de p utilisées entre les lignes verticales du graphiques, 196 (34 %) ont une probabilité de recouvrement plus petite que 0,930, 122 (21 %) en ont une plus petite que 0,920 et la probabilité de recouvrement peut être plus petite que 0,890. Même dans les cas où les règles usuelles sont respectées, l'intervalle de Wald peut être anti-conservateur.

FIG. 2.2 – Probabilité de recouvrement de l’intervalle de Wald de niveau 95 % lorsque $n = 30$. La ligne hachurée est le niveau de confiance nominal 95 %. Les valeurs de p contenues entre les deux lignes verticales vérifient les règles $n \times \min(p, 1 - p) \geq 10$ et $np(1 - p) \geq 5$.

Il faut aussi considérer que dans bien des cas pratiques, on ne peut se permettre d’avoir beaucoup d’observations ; il arrive fréquemment dans certains domaines que l’on doive se contenter d’une vingtaine d’observations au maximum. Dans ces cas, les règles vu précédemment ne sont pas toutes respectées. Il serait bien d’avoir une ou des alternatives à l’intervalle de Wald dans ces cas là.

La figure 2.3 illustre la probabilité de recouvrement de l’intervalle de Wald de niveau 95 % pour n allant de 20 à 100 lorsque $p = 0,200$. On voit que la probabilité de recouvrement a tendance à croître vers le niveau de confiance nominal, mais qu’il y a beaucoup d’oscillations. On peut voir que la probabilité de recouvrement est plus loin du niveau de confiance nominal pour $n = 92$ (0,922) que pour $n = 23$ (0,951). On voit là un autre défaut de l’intervalle de Wald. En fait, même avec $p = 0,500$, c’est seulement à partir de $n \geq 194$ que l’on est sûr d’avoir une probabilité de recouvrement supérieure à 0,940 lorsque le niveau de confiance nominal est 95 % (Brown et al., 2001).

2.1.1 Raisons des faibles probabilités de recouvrement de l’intervalle de Wald

Comme l’a remarqué Xia (2002), la probabilité que l’intervalle de Wald soit un ensemble vide est non-négligeable, surtout lorsque p est près de 0 ou 1. Cela arrive lorsque toutes les observations ont la même valeur. En fait, dans ces cas, l’intervalle de Wald est un ensemble vide s’il est ouvert et il contient une seule valeur, 0 ou 1, s’il est fermé. Qu’il soit ouvert ou fermé, pour tout $0 < p < 1$, la vraie valeur de p n’est pas contenue dans l’intervalle si toutes les observations sont égales. Cela implique que la probabilité de recouvrement est toujours inférieure ou égale à $1 - (p^n + (1 - p)^n)$ et ce peu importe le niveau de confiance nominal. Par exemple, si $n = 20$ et $p = 0,9$ la

FIG. 2.3 – Probabilité de recouvrement de l’intervalle de Wald de niveau 95 % lorsque $p = 0,200$. La ligne hachurée est le niveau de confiance nominal 95 %.

probabilité de recouvrement de l'intervalle de Wald ne pourra être supérieure à 0,878, peu importe le niveau de confiance nominal.

Ce n'est pas la seule raison pour laquelle l'intervalle de Wald donne des probabilités de recouvrement trop petites. Rappelons qu'avec l'intervalle de Wald, on suppose que

$$W_{n,p} \equiv \frac{\hat{p} - p}{\sqrt{\hat{p}(1 - \hat{p})/n}} \sim N(0, 1). \quad (2.9)$$

Dans Brown et al. (2002), on montre qu'en moyenne, $W_{n,p}$ n'est pas centrée à 0. Ce biais est causé par l'approximation de la variance.

Le biais de $W_{n,p}$ sera donné par $\mathbb{E}[W_{n,p}]$. Avec un peu de calcul, on peut voir que

$$W_{n,p} \equiv g(Z_{n,p}) = \frac{Z_{n,p}}{\sqrt{1 + (1 - 2p)Z_{n,p}/\sqrt{np(1-p)} - Z_{n,p}^2/n}}, \quad (2.10)$$

où $Z_{n,p} = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$. La variable $Z_{n,p}$ n'étant pas trop loin de 0 et la fonction $g(u)$ étant suffisamment lisse autour de 0, on peut écrire

$$g(Z_{n,p}) \cong g(0) + g^{(1)}(0)Z_{n,p} + \frac{g^{(2)}(0)}{2!}Z_{n,p}^2 + \frac{g^{(3)}(0)}{3!}Z_{n,p}^3 + \frac{g^{(4)}(0)}{4!}Z_{n,p}^4. \quad (2.11)$$

On obtient donc

$$\begin{aligned} \mathbb{E}[W_{n,p}] = \mathbb{E}[g(Z_{n,p})] &\cong g(0) + g^{(1)}(0) \mathbb{E}[Z_{n,p}] + \frac{g^{(2)}(0)}{2} \mathbb{E}[Z_{n,p}^2] \\ &\quad + \frac{g^{(3)}(0)}{6} \mathbb{E}[Z_{n,p}^3] + \frac{g^{(4)}(0)}{24} \mathbb{E}[Z_{n,p}^4]. \end{aligned}$$

Comme $\mathbb{E}[Z_{n,p}] = 0$, $\mathbb{E}[Z_{n,p}^2] = 1$, $\mathbb{E}[Z_{n,p}^3] = \frac{1-2p}{\sqrt{np(1-p)}}$ et $\mathbb{E}[Z_{n,p}^4] = 3 + \frac{1-6p(1-p)}{np(1-p)}$ et comme $g(0) = 0$, $g^{(1)}(0) = 1$, $g^{(2)}(0) = -\frac{1-2p}{\sqrt{np(1-p)}}$, $g^{(3)}(0) = \frac{3}{n} + \frac{9(1-2p)^2}{4np(1-p)}$ et $g^{(4)}(0) = -\frac{15}{2} \left(\frac{1-2p}{\sqrt{np(1-p)}}\right)^3 - \frac{18}{n} \frac{1-2p}{\sqrt{np(1-p)}}$, on obtient

$$\mathbb{E}[W_{n,p}] \cong \frac{p - 1/2}{\sqrt{np(1-p)}} \left\{ 1 + \frac{7}{2n} + \frac{9(p - 1/2)^2}{2npq} \right\}. \quad (2.12)$$

Le biais est donc négatif pour $p < 1/2$ et positif pour $p > 1/2$. Il faudrait donc que le centre de l'intervalle soit ramené vers $1/2$.

Il y a aussi l'approximation de la loi Binomiale par la loi Normale qui peut causer problème. On sait que la loi Binomiale est asymétrique lorsque $p \neq 1/2$.

Soit $Y \sim \text{Binomiale}(n, p)$. Posons pour $0 < \alpha < 1$

$$L_\alpha = \sup \{y \in \mathbb{N}; \mathbb{P}[Y < y] \leq \alpha/2\} \quad (2.13)$$

et

$$S_\alpha = \inf \{y \in \mathbb{N}; \mathbb{P}[Y > y] \leq \alpha/2\}. \quad (2.14)$$

Si la variable Y était normalement distribuée, on aurait $\mathbb{E}[Y] - L_\alpha = S_\alpha - \mathbb{E}[Y]$. Dans le cas où $Y \sim \text{Binomiale}(n, p)$, on peut montrer que $\mathbb{E}[Y] - L_\alpha < S_\alpha - \mathbb{E}[Y]$ si $p < 1/2$ et $\mathbb{E}[Y] - L_\alpha > S_\alpha - \mathbb{E}[Y]$ si $p > 1/2$. Donc, même si le biais causé par l'estimation de la variance était nul, on aurait quand même intérêt à rapprocher le centre de l'intervalle de $1/2$.

2.2 Intervalles alternatifs

Dans la présente section, on présentera cinq alternatives à l'intervalle de Wald. Pour les quatre premières, on utilise des approches qui ressemblent à celle utilisée pour construire l'intervalle de Wald. La dernière vient du test du rapport des vraisemblances.

2.2.1 L'intervalle de Wald avec correction pour la continuité

On peut réécrire l'équation (2.2) de la manière suivante :

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[n\mu - z_{\alpha/2} \sqrt{n}\sigma \leq \sum_{i=1}^n Y_i \leq n\mu + z_{\alpha/2} \sqrt{n}\sigma \right] = 1 - \alpha. \quad (2.15)$$

Cette équation est identique à l'équation (2.2), à l'exception que les bornes sont maintenant incluses dans l'intervalle. Supposons que les bornes soient des entiers et que

les Y_i soient indépendantes et identiquement distribuées Bernoulli(p). Dans ce cas, on peut écrire

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[n\mu - z_{\alpha/2}\sqrt{n}\sigma - 1/2 \leq \sum_{i=1}^n Y_i \leq n\mu + z_{\alpha/2}\sqrt{n}\sigma + 1/2 \right] = 1 - \alpha, \quad (2.16)$$

où $\mu = p$ et $\sigma = \sqrt{p(1-p)}$. On peut alors montrer que l'on a un intervalle de confiance approximatif de niveau $(1 - \alpha)100\%$ pour p en gardant toutes les valeurs incluses dans l'intervalle suivant :

$$\hat{p} \pm (z_{\alpha/2}\sigma/\sqrt{n} + 1/(2n)). \quad (2.17)$$

Comme pour l'intervalle de Wald, on remplace σ par $\sqrt{\hat{p}(1-\hat{p})}$ pour obtenir

$$\hat{p} \pm \left(z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} + 1/(2n) \right). \quad (2.18)$$

Cet intervalle a la même forme que l'intervalle de Wald, à l'exception que l'on ajoute $1/(2n)$ de chaque côté de l'intervalle pour corriger son caractère anti-conservateur.

2.2.2 L'intervalle de Wilson

Comme pour l'intervalle de Wald, l'intervalle de Wilson, aussi appelé l'intervalle de *score*, utilise la convergence de la loi Binomiale vers la loi Normale. L'équation (2.2) suggère que l'on peut obtenir un intervalle de niveau $(1 - \alpha)100\%$ pour p en incluant toutes les valeurs de p satisfaisant

$$\left| \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \right| < z_{\alpha/2}. \quad (2.19)$$

En remplaçant le dénominateur $\sqrt{p(1-p)/n}$ par $\sqrt{\hat{p}(1-\hat{p})/n}$ et en isolant p , il est facile de voir que l'on obtient l'intervalle de Wald. L'intervalle de Wilson s'obtient quant à lui en isolant p directement dans l'équation (2.19). On obtient alors

$$\left| \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \right| < z_{\alpha/2} \quad (2.20)$$

$$\Leftrightarrow \frac{(\hat{p}-p)^2}{p(1-p)/n} < z_{\alpha/2}^2 \quad (2.21)$$

$$\Leftrightarrow p^2 - 2p\hat{p} + \hat{p}^2 < \frac{z_{\alpha/2}^2}{n}(p - p^2) \quad (2.22)$$

$$\Leftrightarrow \left(1 + \frac{z_{\alpha/2}^2}{n}\right)p^2 - \left(2\hat{p} + \frac{z_{\alpha/2}^2}{n}\right)p + \hat{p}^2 < 0 \quad (2.23)$$

$$\Leftrightarrow p \in \frac{1}{2\left(1 + \frac{z_{\alpha/2}^2}{n}\right)} \left\{ \left(2\hat{p} + \frac{z_{\alpha/2}^2}{n}\right) \pm \sqrt{\left(2\hat{p} + \frac{z_{\alpha/2}^2}{n}\right)^2 - 4\left(1 + \frac{z_{\alpha/2}^2}{n}\right)\hat{p}^2} \right\} \quad (2.24)$$

$$\Leftrightarrow p \in \left(\frac{n\hat{p} + z_{\alpha/2}^2(\frac{1}{2})}{n + z_{\alpha/2}^2} \pm z_{\alpha/2} \sqrt{\left(\frac{n}{n + z_{\alpha/2}^2}\right) \frac{\hat{p}(1-\hat{p})}{n + z_{\alpha/2}^2} + \left(\frac{z_{\alpha/2}^2}{n + z_{\alpha/2}^2}\right) \frac{(\frac{1}{2})(\frac{1}{2})}{n + z_{\alpha/2}^2}} \right) \quad (2.25)$$

On remarque que cet intervalle n'est pas centré à \hat{p} , mais plutôt au point $\frac{n\hat{p} + z_{\alpha/2}^2(\frac{1}{2})}{n + z_{\alpha/2}^2}$, que l'on notera \tilde{p} . On peut réécrire \tilde{p} autrement pour obtenir

$$\tilde{p} = \left(\frac{n}{n + z_{\alpha/2}^2}\right)\hat{p} + \left(\frac{z_{\alpha/2}^2}{n + z_{\alpha/2}^2}\right)\left(\frac{1}{2}\right). \quad (2.26)$$

On voit clairement que \tilde{p} est une moyenne pondérée entre \hat{p} et $1/2$. On voit aussi que ce qui est dans la racine carrée est une moyenne pondérée de la variance de l'estimateur du maximum de vraisemblance de p obtenu à partir d'un échantillon de taille $(n + z_{\alpha/2}^2)$ lorsque $p = \hat{p}$ et lorsque $p = 1/2$.

2.2.3 L'intervalle d'Agresti-Coull

Cet intervalle est celui que Agresti et Coull (1998) appellent *add 2 successes and 2 failures adjusted Wald interval*. Il découle de l'intervalle de Wilson, avec l'avantage d'être beaucoup plus simple d'utilisation. On utilise le fait que dans plusieurs cas, on veut des intervalles de niveau 95 % (donc avec $\alpha = 0,05$), ce qui donne $z_{\alpha/2} = z_{0,025} = 1,96 \approx 2$. En remplaçant $z_{\alpha/2}$ par 2 dans l'équation de \tilde{p} , on obtient $\tilde{p} = \frac{\sum_{i=1}^n X_i + 2}{n + 4}$. On peut voir \tilde{p} comme l'estimateur du maximum de vraisemblance de p d'un échantillon contenant $(n + 4)$ observations : X_1, \dots, X_n en plus de 4 nouvelles données dont 2 sont égales à 1 (2 succès) et 2 sont égales à 0 (2 échecs). L'intervalle d'Agresti-Coull a la même forme

que celui de Wald où l'on remplace \hat{p} par \tilde{p} . On obtient alors l'intervalle suivant :

$$\left(\tilde{p} \pm z_{\alpha/2} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{(n+4)}} \right). \quad (2.27)$$

Pour un niveau de confiance 95 %, l'intervalle d'Agresti-Coull est centré pratiquement au même endroit que celui de Wilson.

2.2.4 L'intervalle bayésien approximatif

Cet intervalle utilise l'estimation bayésienne. L'estimation bayésienne d'un certain paramètre θ se fait de la façon suivante : on choisit d'abord une distribution *a priori* (avant de prendre connaissance des données) pour θ que l'on notera $\xi(\theta)$. Aussi, notons $\xi(\theta|X_1, \dots, X_n)$ la distribution *a posteriori* de θ , $f(X_1, \dots, X_n|\theta)$ la distribution de X_1, \dots, X_n sachant θ et $f(X_1, \dots, X_n)$ la distribution de X_1, \dots, X_n sans connaissance de θ . En utilisant la formule de Bayes, on obtient

$$\xi(\theta|X_1, \dots, X_n) = \frac{\xi(\theta)f(X_1, \dots, X_n|\theta)}{f(X_1, \dots, X_n)}. \quad (2.28)$$

En pratique $f(X_1, \dots, X_n)$ peut être difficile à évaluer, mais on n'a pas besoin de le faire puisque ce terme est constant par rapport à θ . Pour trouver la distribution *a posteriori* de θ , il suffit donc de voir que l'expression $\xi(\theta)f(X_1, \dots, X_n|\theta)$ est proportionnelle à une certaine distribution.

Pour l'estimation d'une proportion p , il est naturel de choisir comme distribution *a priori* de p la loi Uniforme(0,1). Comme $f(X_1, \dots, X_n|p) = p^{\sum_{i=1}^n X_i} (1-p)^{n-\sum_{i=1}^n X_i}$, on a pour $0 \leq p \leq 1$

$$\begin{aligned} \xi(p)f(X_1, \dots, X_n|p) &= p^{\sum_{i=1}^n X_i} (1-p)^{n-\sum_{i=1}^n X_i} \\ &\propto \frac{\Gamma(n+2)}{\Gamma(\sum_{i=1}^n X_i + 1) \Gamma(n - \sum_{i=1}^n X_i + 1)} p^{\sum_{i=1}^n X_i} (1-p)^{n-\sum_{i=1}^n X_i}, \end{aligned}$$

ce qui veut dire que $(p|X_1, \dots, X_n) \sim \text{Beta}(\sum_{i=1}^n X_i + 1, n - \sum_{i=1}^n X_i + 1)$. On sait que si $Y \sim \text{Beta}(\alpha, \beta)$, alors $\mathbb{E}[Y] = \frac{\alpha}{\alpha+\beta}$ et $\text{Var}[Y] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$. On a donc

$p^* \equiv \mathbb{E}[p|X_1, \dots, X_n] = \frac{\sum_{i=1}^n X_i + 1}{n+2}$ et $Var[p|X_1, \dots, X_n] = \frac{p^*(1-p^*)}{n+3}$. Notons que p^* est l'estimateur du maximum de vraisemblance de p d'un échantillon contenant les X_i en plus de 2 nouvelles données dont une est égale à 1 et l'autre est égale à 0.

Pour obtenir l'intervalle bayésien approximatif, on utilise le résultat suivant.

Proposition 1. *Si $Y_{a,b} \sim Beta(a, b)$ et si $\mu_{a,b}$ et $\sigma_{a,b}$ dénotent la moyenne et l'écart-type de la loi $Beta(a, b)$, alors quand $a \rightarrow \infty$ et $b \rightarrow \infty$*

$$\frac{Y_{a,b} - \mu_{a,b}}{\sigma_{a,b}} \rightarrow N(0, 1) \quad (2.29)$$

en loi.

Donc si a et b sont suffisamment grands, on a

$$Y_{a,b} \approx \text{Normale}(\mu_{a,b}, \sigma_{a,b}^2). \quad (2.30)$$

Dans notre cas, a et b suffisamment grands veut dire qu'on a suffisamment de succès et d'échecs, ce qui arrive lorsque n est suffisamment grand et p pas trop près de 0 ou 1. Sous ces conditions, on a

$$(p|X_1, \dots, X_n) \approx \text{Normale}(p^*, p^*(1-p^*)/(n+3)). \quad (2.31)$$

L'intervalle bayésien approximatif s'écrit donc de la manière suivante :

$$\left(p^* \pm z_{\alpha/2} \sqrt{\frac{p^*(1-p^*)}{n+3}} \right). \quad (2.32)$$

2.2.5 L'intervalle du rapport des vraisemblances

Cet intervalle utilise le fait que asymptotiquement, $2 \{l(\hat{p}) - l(p)\} \sim \chi_1^2$ où $l(\cdot)$ est le logarithme de la fonction de vraisemblance étant donné un échantillon X_1, \dots, X_n . L'intervalle du rapport des vraisemblances contient tous les $p \in [0, 1]$ tels que

$$2 \{l(\hat{p}) - l(p)\} < \chi_{1,\alpha}^2, \quad (2.33)$$

où $\chi_{1,\alpha}^2$ est le quantile d'ordre $(1 - \alpha)$ de la loi χ_1^2 . Comme

$$l(\hat{p}) = \log \left\{ \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^{\sum_{i=1}^n X_i} \left(1 - \frac{1}{n} \sum_{i=1}^n X_i \right)^{n - \sum_{i=1}^n X_i} \right\} \quad (2.34)$$

$$= \left(\sum_{i=1}^n X_i \right) \log \left(\frac{1}{n} \sum_{i=1}^n X_i \right) + \left(n - \sum_{i=1}^n X_i \right) \log \left(1 - \frac{1}{n} \sum_{i=1}^n X_i \right) \quad (2.35)$$

et de même

$$l(p) = \left(\sum_{i=1}^n X_i \right) \log(p) + \left(n - \sum_{i=1}^n X_i \right) \log(1 - p), \quad (2.36)$$

l'équation (2.33) devient

$$2 \left[\left(\sum_{i=1}^n X_i \right) \left\{ \log \left(\frac{1}{n} \sum_{i=1}^n X_i \right) - \log p \right\} + \left(n - \sum_{i=1}^n X_i \right) \left\{ \log \left(1 - \frac{1}{n} \sum_{i=1}^n X_i \right) - \log(1 - p) \right\} \right] < \chi_{1,\alpha}^2. \quad (2.37)$$

Cet intervalle a le désavantage de ne pas avoir de forme explicite. En pratique, on obtient les bornes de cet intervalle en trouvant les deux valeurs de p qui font en sorte que la partie à gauche de l'équation (2.37) soit égale à $\chi_{1,\alpha}^2$.

2.3 Comparaison des intervalles

On conserve les six intervalles présentés pour les comparaisons : celui de Wald, de Wald avec correction pour la continuité (WCC), de Wilson, d'Agresti-Coull (AC), bayésien approximatif (BA) et du rapport des vraisemblances (RV). On comparera les intervalles sur deux fronts : la probabilité de recouvrement et la longueur des intervalles.

2.3.1 Probabilités de recouvrement des intervalles

La figure 2.4 présente les probabilités de recouvrement exactes avec niveau de confiance nominal 95 % des six intervalles étudiés pour $n = 20$. Elles sont calculées

FIG. 2.4 – Probabilités de recouvrement des six intervalles étudiés de niveau 95 % lorsque $n = 20$. La ligne hachurée est le niveau de confiance nominal 95 %.

de la même façon que pour l'intervalle de Wald, avec l'équation (2.8) adaptée selon l'intervalle.

On remarque d'abord que les cinq alternatives donnent de meilleures probabilités de recouvrement que l'intervalle de Wald. L'intervalle WCC donne des probabilités de recouvrement plus élevées que l'intervalle de Wald pour les valeurs de p près de 0 ou 1, mais elles sont encore trop faibles. Pour les p entre 0,2 et 0,8, les probabilités de recouvrement sont toujours plus élevées que 0,9 et elles ne sont pas toujours plus petites que le niveau de confiance nominal 0,95 comme c'était le cas avec l'intervalle de Wald. Les intervalles de Wilson, AC et BA montrent de bons résultats. Pour des p pas trop près de 0 ou de 1, les probabilités de recouvrement sont très près du niveau de confiance nominal et pour les p près de 0 ou 1, les intervalles de Wilson, AC et BA sont un peu trop conservateurs, mais quand même meilleurs que l'intervalle de Wald avec ou sans correction de continuité. Dans l'ensemble, l'intervalle RV donne des probabilités de recouvrement comparables à celles des intervalles AC et BA, sauf aux alentours de 0,1 et de 0,9, où la probabilité de recouvrement tombe sous 0,85. Pour les cas $n = 20$ et niveau de confiance nominal 95 %, les meilleurs intervalles par rapport à la probabilité de recouvrement sont les intervalles de Wilson, AC et BA, suivis de près par l'intervalle RV, qui est quand même meilleur que l'intervalle WCC et finalement l'intervalle de Wald ferme la marche.

Outre son caractère anti-conservateur, l'intervalle de Wald a le défaut de produire des probabilités de recouvrement qui ne s'approchent pas nécessairement du niveau de confiance nominal si on fait croître n ou si on approche p de 0,5. Cela a été illustré par les oscillations dans les figures de la section 1.1. On voit dans la figure 2.4 qu'aucune des alternatives étudiées ne corrige ce défaut entièrement. Cependant, pour les intervalles AC et BA, les oscillations sont moins prononcées qu'avec l'intervalle de Wald.

Comparer les intervalles pour un seul n n'est pas suffisant. On a calculé les probabilités de recouvrement pour 90 couples (n, p) : il y a neuf n différents (20, 30, ..., 100) et pour chaque n , il y a dix p différents (0,05, 0,10, ..., 0,50). On n'a pas inclus de $p > 0,5$ parce que pour tous les intervalles étudiés, pour un n et un p fixés, la probabilité de

FIG. 2.5 – Diagrammes en boîte des probabilités de recouvrement des six intervalles étudiés de niveau 95 % pour 90 couples (n, p) . Le niveau de confiance nominal pour tous les intervalles est 95 %

Intervalle	Médiane	Moyenne	Écart-type	Minimum	Maximum
Wald	0,940	0,925	0,045	0,639	0,959
WCC	0,955	0,945	0,041	0,641	0,967
Wilson	0,954	0,953	0,012	0,925	0,978
AC	0,956	0,957	0,010	0,935	0,984
BA	0,948	0,949	0,014	0,922	0,989
RV	0,949	0,946	0,018	0,858	0,984

TAB. 2.1 – Statistiques descriptives des probabilités de recouvrement des six intervalles étudiés de niveau 95 % pour 90 couples (n, p) .

recouvrement pour le couple (n, p) est la même que pour le couple $(n, 1 - p)$.

Les diagrammes en boîte de ces 90 valeurs sont présentés dans la figure 2.5 et le tableau 2.1 présente quelques statistiques descriptives de ces valeurs.

On voit que les mesures de tendance centrale, soit la médiane et la moyenne, sont près du niveau de confiance nominal pour tous les intervalles à l'exception de l'intervalle de Wald. Les intervalles de Wilson, AC et BA se démarquent du fait qu'on ne retrouve aucune des 90 probabilités de recouvrement trop loin du niveau de confiance nominal.

2.3.2 Longueurs des intervalles

La longueur d'un intervalle de confiance est un facteur très important puisqu'elle représente la précision de la connaissance que l'on a du paramètre p . Les longueurs des intervalles de Wald, WCC, de Wilson, AC et BA sont comparées. Pour l'intervalle RV, il est souvent difficile d'évaluer les bornes des intervalles. Comme cette méthode est la plus difficile d'application et qu'elle n'est pas celle offrant les meilleures probabilités de recouvrement, on l'abandonne pour cette section.

Dans cette section, on calcule les longueurs espérées des intervalles. Pour un intervalle et un couple (n, p) donnés, on calcule la longueur espérée, notée $L(n, p)$, de la

FIG. 2.6 – Longueurs espérées de cinq intervalles de niveau 95 % lorsque $n = 20$. La ligne continue la plus épaisse représente l'intervalle de Wald, celle plus mince l'intervalle WCC, celle hachurée l'intervalle de Wilson, celle pointillée l'intervalle AC et celle mi-hachurée mi-pointillée l'intervalle BA.

FIG. 2.7 – Longueurs espérées de cinq intervalles de niveau 95 % lorsque $n = 100$. La ligne continue la plus épaisse représente l'intervalle de Wald, celle plus mince l'intervalle WCC, celle hachurée l'intervalle de Wilson, celle pointillée l'intervalle AC et celle mi-hachurée mi-pointillée l'intervalle BA.

façon suivante :

$$L(n, p) = \sum_{j=0}^n L(n, p, j) \binom{n}{j} p^j (1-p)^{n-j}, \quad (2.38)$$

où

$L(n, p, j)$ = la longueur de l'intervalle lorsque $\sum_{i=1}^n X_i = j$ pour $0 \leq p \leq 1$ et $n \in \mathbb{N}^*$.

Reprenons le cas où $n = 20$. La figure 2.6 montre les longueurs espérées des intervalles de niveau 95 %.

On voit que l'intervalle WCC est toujours plus long de 0,05 ($1/n$) par rapport à l'intervalle de Wald et que les longueurs des intervalles de Wilson, AC et BA sont comparables, quoi que l'intervalle de Wilson est toujours légèrement plus court que l'intervalle BA. Pour p entre 0,25 et 0,75, les intervalles de Wilson, AC et BA sont plus courts que les deux autres, tout en ayant des meilleures probabilités de recouvrement. Les intervalles de Wald et WCC sont plus courts que les deux autres près de 0 ou 1, mais on a vu que ces intervalles sont très anti-conservateurs dans ces cas-là.

La figure 2.7 montre les longueurs espérées des intervalles de niveau 95 % lorsque $n=100$.

Pour p pas trop près de 0 ou 1, les intervalles de Wald, de Wilson, AC et BA ont des longueurs comparables et l'intervalle WCC est le plus long. Les intervalles de Wilson et BA ont pratiquement la même longueur pour toutes les valeurs de p . Pour p près de 0 ou 1, les intervalles de Wald et WCC sont encore les plus courts, mais comme pour

le cas $n = 20$, leurs probabilités de recouvrement sont nettement trop faibles dans ces cas-là.

2.4 Conclusion

On a vu que l'intervalle de Wald pour une proportion a le principal défaut d'être très anti-conservateur, surtout lorsque p est près de 0 ou 1 ou lorsque n est petit. Il possède néanmoins l'avantage d'être très simple d'application. Cependant, on a vu qu'il existe d'autres intervalles tout aussi simples qui performant beaucoup mieux. Un de ceux-là, l'intervalle WCC, ne fait qu'allonger l'intervalle de Wald en gardant le même centre. Il donne des probabilités de recouvrement plus près du niveau de confiance nominal que l'intervalle de Wald, mais est encore trop anti-conservateur. Les meilleurs intervalles considérés sont les intervalles de Wilson, AC et BA qui ramènent le centre un peu plus près de $1/2$. Lorsque p n'est pas trop près de 0 ou 1, leurs probabilités de recouvrement sont très près du niveau de confiance nominal si on les compare à celles des intervalles de Wald et WCC, tout en ayant des longueurs espérées pas plus élevées que ces deux derniers intervalles. Les intervalles AC et BA sont des alternatives particulièrement intéressantes puisqu'ils sont simples d'applications. On verra dans le prochain chapitre si l'extension de ces intervalles pour le cas d'une différence de proportions donne des résultats semblables.

Chapitre 3

Intervalle de confiance pour une différence de deux proportions

Le présent chapitre concerne l'estimation par intervalle de confiance d'une différence de deux proportions $p_1 - p_2$. Les intervalles qu'on étudiera ici seront des extensions des intervalles du chapitre 2. Tout au long du chapitre, à moins d'avis contraire, on supposera que X_1, \dots, X_{n_1} sont des observations indépendantes et identiquement distribuées Bernoulli(p_1) et que Y_1, \dots, Y_{n_2} sont indépendantes et identiquement distribuées Bernoulli(p_2) avec X_1, \dots, X_{n_1} indépendantes de Y_1, \dots, Y_{n_2} . Les estimateurs du maximum de vraisemblance de p_1 et p_2 , soit $\frac{1}{n_1} \sum_{i=1}^{n_1} X_i$ et $\frac{1}{n_2} \sum_{j=1}^{n_2} Y_j$, seront notés \hat{p}_1 et \hat{p}_2 respectivement.

3.1 L'intervalle de Wald

Par le théorème limite central, on sait que

$$\mathcal{L} \left(\frac{\hat{p}_1 - p_1}{\sqrt{p_1(1-p_1)/n_1}} \right) \rightarrow \text{Normale}(0, 1) \quad (3.1)$$

$$\mathcal{L} \left(\frac{\hat{p}_2 - p_2}{\sqrt{p_2(1-p_2)/n_2}} \right) \rightarrow \text{Normale}(0, 1) \quad (3.2)$$

lorsque $n_1 \rightarrow \infty$ et $n_2 \rightarrow \infty$. Comme les X_1, \dots, X_{n_1} sont indépendantes de

Y_1, \dots, Y_{n_2} , on a

$$\mathcal{L} \left(\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}} \right) \rightarrow \text{Normale}(0, 1) \quad (3.3)$$

lorsque $n_1 \rightarrow \infty$ et $n_2 \rightarrow \infty$. On a donc

$$\lim_{\substack{n_1 \rightarrow \infty \\ n_2 \rightarrow \infty}} \mathbb{P} \left[\left| \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}} \right| < z_{\alpha/2} \right] = 1 - \alpha, \quad (3.4)$$

ce qui veut dire qu'on peut obtenir un intervalle de confiance approximatif de niveau $(1 - \alpha)100\%$ pour $(p_1 - p_2)$ en gardant tous les $(p_1 - p_2)$ tels que

$$\left| \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}} \right| < z_{\alpha/2}. \quad (3.5)$$

L'intervalle de confiance de Wald s'obtient en remplaçant les p_1 et p_2 contenus au dénominateur par \hat{p}_1 et \hat{p}_2 respectivement. L'intervalle de Wald s'écrit donc de la façon suivante :

$$\left(\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2} \right). \quad (3.6)$$

3.1.1 Raisons des faibles probabilités de recouvrement de l'intervalle de Wald

L'intervalle de Wald pour une différence de deux proportions conserve certaines propriétés de celui pour une proportion. Par exemple, il est encore possible que l'intervalle soit un ensemble vide. Pour cela il faut que

$$\left[\left\{ \left(\sum_{i=1}^{n_1} X_i = 0 \right) \cup \left(\sum_{i=1}^{n_1} X_i = n_1 \right) \right\} \cap \left\{ \left(\sum_{i=1}^{n_2} Y_i = 0 \right) \cup \left(\sum_{i=1}^{n_2} Y_i = n_2 \right) \right\} \right]. \quad (3.7)$$

Cela implique que pour tout $0 < p < 1$, la probabilité de recouvrement est inférieure ou égale à $1 - \{p_1^{n_1} + (1-p_1)^{n_1}\} \{p_2^{n_2} + (1-p_2)^{n_2}\}$ (Xia, 2002) peu importe le niveau de confiance nominal.

L'intervalle de Wald est construit en supposant que

$$W_{n_1, p_1, n_2, p_2} \equiv \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}_1(1 - \hat{p}_1)/n_1 + \hat{p}_2(1 - \hat{p}_2)/n_2}} \sim N(0, 1). \quad (3.8)$$

En procédant comme à la section 2.1.1, on obtient l'approximation suivante pour l'espérance de W_{n_1, p_1, n_2, p_2} :

$$\begin{aligned} \mathbb{E}[W_{n_1, p_1, n_2, p_2}] &\cong \frac{p_1 - 1/2}{n_1 u_1 u^{1/2}} \left\{ 1 + \frac{1}{n_1} \left(\frac{9}{2u_1} - 1 \right) \right\} \\ &\quad - \frac{p_2 - 1/2}{n_2 u_2 u^{1/2}} \left\{ 1 + \frac{1}{n_2} \left(\frac{9}{2u_2} - 1 \right) \right\} \\ &\quad + \frac{3(p_1 - p_2)}{2n_1 n_2 u_1 u_2 u^{1/2}} \\ &\quad - \frac{15(p_1 - 1/2)(p_2 - 1/2)}{2n_1 n_2 u_1 u_2 u^{3/2}} \left(\frac{p_2 - 1/2}{n_2} - \frac{p_1 - 1/2}{n_1} \right), \end{aligned}$$

où $u = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$, $u_1 = 1 + \frac{p_2(1-p_2)n_1}{p_1(1-p_1)n_2}$ et $u_2 = 1 + \frac{p_1(1-p_1)n_2}{p_2(1-p_2)n_1}$. Voir Xia (2002) pour les détails. On voit que si p_1 et p_2 sont tous les deux près de $1/2$, le biais sera faible. Comme pour l'intervalle de Wald pour une proportion, on a intérêt à rapprocher les estimateurs de p_1 et p_2 de $1/2$. On voit aussi que le biais est faible si p_1 et p_2 sont près l'un de l'autre, à condition que n_1 et n_2 soient comparables.

3.2 Intervalles alternatifs

Dans la présente section, on présentera quelques alternatives à l'intervalle de Wald. Ces alternatives sont les extensions de certaines alternatives du chapitre 2. Puisque le côté gauche de l'inégalité (3.5) ne peut pas s'écrire comme une fonction de $p_1 - p_2$ uniquement, l'intervalle de Wilson pour une proportion n'a pas d'analogue dans le cas d'une différence de deux proportions. L'intervalle du rapport des vraisemblances a aussi été abandonné ; on a vu au chapitre 2 qu'il était le plus compliqué des intervalles étudiés sans performer mieux que les intervalles d'Agresti-Coull et bayésien approximatif. Il existe plusieurs autres intervalles pour la différence de deux proportions, voir Newcombe (1998b).

3.2.1 L'intervalle de Wald avec correction pour la continuité

Cet intervalle a la même forme que l'intervalle de Wald, à l'exception que l'on ajoute $1/(2n_1) + 1/(2n_2)$ de chaque côté de l'intervalle pour la même raison que pour un intervalle de confiance pour une proportion. La forme de l'intervalle est alors

$$\hat{p}_1 - \hat{p}_2 \pm \left(z_{\alpha/2} \sqrt{\hat{p}_1(1 - \hat{p}_1)/n_1 + \hat{p}_2(1 - \hat{p}_2)/n_2 + 1/(2n_1) + 1/(2n_2)} \right). \quad (3.9)$$

3.2.2 L'intervalle d'Agresti-Coull

On a vu au chapitre 2 pourquoi l'intervalle d'Agresti-Coull pour une proportion est centré à $\frac{\sum_{i=1}^n X_i + 2}{n+4}$. Dans le cas d'une différence de deux proportions, on pourrait penser que l'intervalle d'Agresti-Coull s'écrit comme l'intervalle de Wald en remplaçant \hat{p}_1 par $\frac{\sum_{i=1}^{n_1} X_i + 2}{n_1 + 4}$ et \hat{p}_2 par $\frac{\sum_{i=1}^{n_2} Y_i + 2}{n_2 + 4}$, autrement dit, en ajoutant huit observations au total : 2 succès et 2 échecs à chaque échantillon.

Agresti et Caffo (2000) ont testé quatre possibilités ; ils ont construit des intervalles ayant la forme de l'intervalle de Wald, mais en ajoutant $t/2$ observations à chaque échantillon pour $t = \{2, 4, 6, 8\}$ dont la moitié sont des succès et l'autre moitié sont des échecs. Si $t/2$ est impair, on suppose alors que la dernière observation ajoutée dans un échantillon prend la valeur $1/2$ et qu'il y a autant de succès que d'échecs parmi les autres observations ajoutées dans l'échantillon. Après avoir calculé des probabilités de recouvrement sur plusieurs (n_1, p_1, n_2, p_2) , ils arrivent à la conclusion que le meilleur choix est $t = 4$ parce que les probabilités de recouvrement obtenues sont généralement près du niveau de confiance nominal et qu'elles ne s'en éloignent jamais trop.

Suite à cela, on définit dans ce mémoire l'intervalle d'Agresti-Coull pour une différence de deux proportions comme un intervalle ayant la même forme que l'intervalle de Wald, mais en remplaçant \hat{p}_1 et \hat{p}_2 par $\tilde{p}_1 = \frac{\sum_{i=1}^{n_1} X_i + 1}{n_1 + 2}$ et $\tilde{p}_2 = \frac{\sum_{i=1}^{n_2} Y_i + 1}{n_2 + 2}$ respectivement. On obtient donc

$$\left(\tilde{p}_1 - \tilde{p}_2 \pm z_{\alpha/2} \sqrt{\tilde{p}_1(1 - \tilde{p}_1)/(n_1 + 2) + \tilde{p}_2(1 - \tilde{p}_2)/(n_2 + 2)} \right). \quad (3.10)$$

3.2.3 L'intervalle bayésien approximatif

Comme pour l'intervalle bayésien approximatif pour une proportion, on obtient, après avoir choisi comme loi *a priori* de (p_1, p_2) la loi uniforme sur le carré $(0, 1)^2$, les lois *a posteriori* de p_1 et p_2 :

$$(p_1 | X_1, \dots, X_{n_1}) \sim \text{Beta} \left(\sum_{i=1}^{n_1} X_i + 1, n_1 - \sum_{i=1}^{n_1} X_i + 1 \right) \quad (3.11)$$

$$(p_2 | Y_1, \dots, Y_{n_2}) \sim \text{Beta} \left(\sum_{i=1}^{n_2} Y_i + 1, n_2 - \sum_{i=1}^{n_2} Y_i + 1 \right) \quad (3.12)$$

avec $(p_1 | X_1, \dots, X_{n_1})$ indépendant de $(p_2 | Y_1, \dots, Y_{n_2})$. Ensuite, comme au chapitre précédant, on utilise l'approximation normale, c'est-à-dire que l'on suppose que

$$(p_1 | X_1, \dots, X_{n_1}) \sim \text{Normale}(\tilde{p}_1, \tilde{p}_1(1 - \tilde{p}_1)/(n_1 + 3)) \quad (3.13)$$

$$(p_2 | Y_1, \dots, Y_{n_2}) \sim \text{Normale}(\tilde{p}_2, \tilde{p}_2(1 - \tilde{p}_2)/(n_2 + 3)) \quad (3.14)$$

où \tilde{p}_1 et \tilde{p}_2 ont les mêmes valeurs que pour l'intervalle d'Agresti-Coull. L'intervalle est alors donné par l'équation suivante :

$$\left(\tilde{p}_1 - \tilde{p}_2 \pm z_{\alpha/2} \sqrt{\tilde{p}_1(1 - \tilde{p}_1)/(n_1 + 3) + \tilde{p}_2(1 - \tilde{p}_2)/(n_2 + 3)} \right). \quad (3.15)$$

On voit facilement que pour une différence de deux proportions, l'intervalle bayésien approximatif est inclus dans l'intervalle d'Agresti-Coull.

3.3 Comparaison des intervalles

On compare quatre intervalles : celui de Wald, de Wald avec correction pour la continuité (WCC), d'Agresti-Coull (AC) et bayésien approximatif (BA). Encore une fois, on comparera les intervalles par rapport à la probabilité de recouvrement et par rapport à la longueur des intervalles.

3.3.1 Probabilités de recouvrement des intervalles

Les probabilités de recouvrement présentées dans les prochaines figures sont des probabilités de recouvrement exactes des différents intervalles de niveau 95 % pour différents n_1 , n_2 , p_1 et p_2 . Soit $P_A(n_1, p_1, n_2, p_2)$ la probabilité de recouvrement exacte de l'intervalle A lorsque $X \sim \text{Binomiale}(n_1, p_1)$ et $Y \sim \text{Binomiale}(n_2, p_2)$. Cette probabilité de recouvrement est calculée de la façon suivante :

$$\begin{aligned} P_A(n_1, p_1, n_2, p_2) &= \mathbb{P}[p_1 - p_2 \in A] \\ &= \mathbb{E}[I_A(n_1, n_2, p_1, p_2, X, Y)] \\ &= \sum_{j=0}^{n_1} \sum_{k=0}^{n_2} I_A(n_1, n_2, p_1, p_2, j, k) \\ &\quad \times \binom{n_1}{j} p_1^j (1-p_1)^{n_1-j} \binom{n_2}{k} p_2^k (1-p_2)^{n_2-k}, \end{aligned}$$

où

$$I_A(n_1, n_2, p_1, p_2, j, k) = \begin{cases} 1 & \text{si } p_1 - p_2 \text{ est inclus dans l'intervalle } A \\ & \text{lorsque } X = j \text{ et } Y = k; \\ 0 & \text{sinon.} \end{cases}$$

La figure 3.1 montre les probabilités de recouvrement pour les quatre intervalles de niveau 95 % en fonction de p_1 lorsque $n_1 = n_2 = 20$ et $p_2 = 0,5$. On voit que l'intervalle de Wald est anti-conservateur comme il l'était au chapitre 2. Sa probabilité de recouvrement n'atteint jamais le niveau de confiance nominal. À l'inverse, l'intervalle WCC est trop conservateur. Les intervalles AC et BA sont ceux ayant des probabilités de recouvrement le plus près du niveau de confiance nominal pour la plupart des valeurs de p_1 . Il arrive cependant que la probabilité de recouvrement de l'intervalle BA soit inférieure à celle de l'intervalle de Wald.

FIG. 3.1 – Probabilités de recouvrement des quatre intervalles étudiés de niveau 95 % en fonction de p_1 lorsque $n_1 = n_2 = 20$ et $p_2 = 0,5$. La ligne continue la plus épaisse représente l'intervalle de Wald, celle plus mince l'intervalle WCC, celle hachurée l'intervalle AC et celle pointillée l'intervalle BA.

La figure 3.2 montre les probabilités de recouvrement pour les quatre intervalles encore une fois en fonction de p_1 et avec $n_1 = n_2 = 20$, mais avec $p_2 = 0,2$. Les

commentaires donnés sur la figure 3.1 s'appliquent encore. On remarque cependant que les oscillations des probabilités de recouvrement sont cette fois-ci plus prononcées lorsque p_1 est près de 1, spécialement pour les intervalles de Wald et WCC.

La figure 3.3 montre les probabilités de recouvrement pour les quatre intervalles de niveau 95 % en fonction de p_1 lorsque $n_1 = n_2 = 20$ et $p_1 - p_2 = 0,25$. On voit que l'intervalle BA performe moins bien dans ce cas là. Il présente des probabilités de recouvrement inférieures à celles de l'intervalle de Wald pour p_1 entre 0,492 et 0,758. Pour p_2 près de 0 et p_1 près de 1, ce sont les intervalles AC et BA qui performent le mieux; autrement, l'intervalle AC donne des probabilités de recouvrement toujours meilleures que les autres intervalles.

FIG. 3.2 – Probabilités de recouvrement des quatre intervalles étudiés de niveau 95 % en fonction de p_1 lorsque $n_1 = n_2 = 20$ et $p_2 = 0,2$. La ligne continue la plus épaisse représente l'intervalle de Wald, celle plus mince l'intervalle WCC, celle hachurée l'intervalle AC et celle pointillée l'intervalle BA.

FIG. 3.3 – Probabilités de recouvrement des quatre intervalles étudiés de niveau 95 % en fonction de p_1 lorsque $n_1 = n_2 = 20$ et $p_1 - p_2 = 0,25$. La ligne continue la plus épaisse représente l'intervalle de Wald, celle plus mince l'intervalle WCC, celle hachurée l'intervalle AC et celle pointillée l'intervalle BA.

La figure 3.4 montre les probabilités de recouvrement pour les quatre intervalles de niveau 95 % en fonction de n_1 lorsque $p_1 = 0,5$, $p_2 = 0,2$ et $n_2 = 30$. Comme pour la figure 3.1, l'intervalle de Wald est anti-conservateur, l'intervalle WCC est trop conservateur et les intervalles AC et BA sont ceux qui performent le mieux. Il est intéressant de remarquer que pour l'intervalle de Wald, la probabilité de recouvrement n'a pas tendance à se rapprocher du niveau de confiance nominal lorsque n_1 se rapproche de 100. Cela montre une certaine faiblesse de l'intervalle de Wald lorsque un échantillon est beaucoup plus petit que l'autre.

FIG. 3.4 – Probabilités de recouvrement des quatre intervalles étudiés de niveau 95 % en fonction de n_1 lorsque $n_2 = 30$, $p_1 = 0,5$ et $p_2 = 0,2$. La ligne continue la plus épaisse représente l'intervalle de Wald, celle plus mince l'intervalle WCC, celle hachurée l'intervalle AC et celle pointillée l'intervalle BA.

On a calculé les probabilités de recouvrement des quatre intervalles avec niveau de confiance 95 % pour 330 valeurs de (n_1, p_1, n_2, p_2) . Les cinq valeurs choisies pour n_1

FIG. 3.5 – Diagrammes en boîte des probabilités de recouvrement des quatre intervalles étudiés de niveau 95 % pour 330 valeurs de (n_1, p_1, n_2, p_2) .

et n_2 sont 20, 40, 60, 80 et 100. Les sept valeurs choisies pour p_1 et p_2 sont 0,05, 0,1, 0,25, 0,5, 0,75, 0,9 et 0,95. On obtient 330 combinaisons de (n_1, p_1, n_2, p_2) au lieu des $(5 \times 7)^2 = 1225$ combinaisons possibles de (n_1, p_1, n_2, p_2) puisque certaines combinaisons donnent les mêmes probabilités de recouvrement. En effet, pour tous les intervalles considérés, la probabilité de recouvrement d'un (n_1, p_1, n_2, p_2) donné est la même que celle de (n_2, p_2, n_1, p_1) , de $(n_1, 1 - p_1, n_2, 1 - p_2)$ et de $(n_2, 1 - p_2, n_1, 1 - p_1)$.

La figure 3.5 montre les diagrammes en boîte pour les 330 valeurs des quatre intervalles. Comme au chapitre 2, on voit que ce sont les intervalles AC et BA qui performant le mieux. La plus petite valeur de probabilité de recouvrement observée est 0,943 pour l'intervalle AC et 0,937 pour l'intervalle BA alors qu'elle est de 0,871 pour l'intervalle WCC et 0,729 pour l'intervalle de Wald. Il est intéressant de remarquer que ces plus petites valeurs de probabilités de recouvrement observées ne proviennent pas des mêmes valeurs de (n_1, p_1, n_2, p_2) . La probabilité de recouvrement minimale provient du point $(n_1, p_1, n_2, p_2) = (20, 0, 5, 100, 0, 9)$ pour l'intervalle AC, du point $(20, 0, 5, 80, 0, 05)$ pour l'intervalle BA, du point $(20, 0, 05, 20, 0, 95)$ pour l'intervalle WCC et du point $(20, 0, 05, 100, 0, 05)$ pour l'intervalle de Wald.

3.3.2 Longueurs des intervalles

Les figures 3.6 à 3.9 montrent les longueurs espérées des quatre intervalles de niveau 95 % pour les quatre cas vus précédemment, soit le cas où p_1 varie avec $n_1 = n_2 = 20$ et $p_2 = 0, 5$ (figure 3.6), le cas où p_1 varie avec $n_1 = n_2 = 20$ et $p_2 = 0, 2$ (figure 3.7), le cas où p_1 varie avec $n_1 = n_2 = 20$ et $p_1 - p_2 = 0, 25$ (figure 3.8) et le cas où n_1 varie avec $n_2 = 30$, $p_1 = 0, 5$ et $p_2 = 0, 2$ (figure 3.9). Soient les variables $X \sim \text{Binomiale}(n_1, p_1)$ et $Y \sim \text{Binomiale}(n_2, p_2)$. Pour un intervalle et un (n_1, p_1, n_2, p_2) donnés, on calcule la longueur espérée, notée $L(n_1, p_1, n_2, p_2)$, de la façon suivante :

$$L(n_1, p_1, n_2, p_2) = \sum_{j=0}^{n_1} \sum_{k=0}^{n_2} L(n_1, p_1, n_2, p_2, j, k) \times \binom{n_1}{j} p_1^j (1 - p_1)^{n_1 - j} \binom{n_2}{k} p_2^k (1 - p_2)^{n_2 - k},$$

où

$L(n_1, p_1, n_2, p_2, j, k)$ = la longueur de l'intervalle lorsque $X = j$ et $Y = k$.

Les trois figures où p_1 varie se ressemblent. Les longueurs des intervalles AC et BA sont pratiquement les mêmes, avec BA toujours légèrement plus court que AC. L'intervalle de Wald est un peu plus long que les intervalles AC et BA pour p_1 pas trop près de 0 ou 1, autrement, il est le plus court. L'intervalle WCC est toujours le plus long des quatre.

Dans le cas où n_1 varie (figure 3.9), on voit que les intervalles de Wald, AC et BA sont pratiquement de la même longueur et l'intervalle WCC est toujours le plus long.

FIG. 3.6 – Longueurs espérées des quatre intervalles étudiés de niveau 95 % en fonction de p_1 lorsque $n_1 = n_2 = 20$ et $p_2 = 0,5$. La ligne continue la plus épaisse représente l'intervalle de Wald, celle plus mince l'intervalle WCC, celle hachurée l'intervalle AC et celle pointillée l'intervalle BA.

FIG. 3.7 – Longueurs espérées des quatre intervalles étudiés de niveau 95 % en fonction de p_1 lorsque $n_1 = n_2 = 20$ et $p_2 = 0,2$. La ligne continue la plus épaisse représente l'intervalle de Wald, celle plus mince l'intervalle WCC, celle hachurée l'intervalle AC et celle pointillée l'intervalle BA.

FIG. 3.8 – Longueurs espérées des quatre intervalles étudiés de niveau 95 % en fonction de p_1 lorsque $n_1 = n_2 = 20$ et $p_1 - p_2 = 0,25$. La ligne continue la plus épaisse représente l'intervalle de Wald, celle plus mince l'intervalle WCC, celle hachurée l'intervalle AC et celle pointillée l'intervalle BA.

FIG. 3.9 – Longueurs espérées des quatre intervalles étudiés de niveau 95 % en fonction de n_1 lorsque $n_2 = 30$, $p_1 = 0,5$ et $p_2 = 0,2$. La ligne continue la plus épaisse représente l'intervalle de Wald, celle plus mince l'intervalle WCC, celle hachurée l'intervalle AC et celle pointillée l'intervalle BA.

3.4 Conclusion

Les propriétés des intervalles observées au chapitre 2 s'appliquent encore pour une différence de deux proportions. L'intervalle de Wald est encore anti-conservateur,

surtout lorsqu'une des deux probabilités s'approche de 0 ou de 1. L'intervalle WCC améliore un peu la probabilité de recouvrement en étant plus long que l'intervalle de Wald, mais il donne encore trop souvent des probabilités de recouvrement trop faibles par rapport au niveau de confiance nominal lorsque les deux probabilités sont près de 0 ou 1. Les intervalles AC et BA, qui modifient plus le centre que la longueur de l'intervalle de Wald, sont encore une fois les plus performants. Ils ne donnent pas des probabilités de recouvrement trop loins du niveau de confiance nominal tout en n'étant pas trop longs.

Il est intéressant de voir que les probabilités de recouvrement les plus faibles sont généralement obtenues lorsque n_1 et n_2 sont éloignés. Supposons que l'on fait une expérience où l'on doit choisir n_1 et n_2 observations indépendantes issues des lois Bernoulli de paramètre p_1 et p_2 respectivement avec $n_1 + n_2$ fixé au départ. Cela veut dire que si l'unique critère lors du choix de n_1 et n_2 est de ne pas avoir des probabilités de recouvrement trop faibles, on a intérêt à choisir des valeurs de n_1 et n_2 les plus proches possibles.

Dans le prochain chapitre, on étudiera des règles qui permettront de choisir le nombre d'observations nécessaires dans une expérience. On regardera aussi la performance des intervalles AC et BA selon la règle utilisée.

Chapitre 4

Règle d'arrêt d'une expérience

Dans ce chapitre, on verra diverses règles qui permettront de choisir le nombre total d'observations nécessaires dans une expérience où l'on veut comparer deux traitements, disons A et B . Pour chaque observation, la réponse au traitement assigné sera dichotomique (succès ou échec) avec probabilité de succès p_1 pour le traitement A et p_2 pour le traitement B . Le nombre d'observations nécessaires sera toujours déterminé en fonction de la probabilité d'erreur de première espèce, soit la probabilité de détecter une différence entre les traitements lorsque $p_1 = p_2$, et la probabilité d'erreur de deuxième espèce, soit la probabilité de ne pas détecter une différence entre les traitements lorsque $\Delta \equiv p_1 - p_2 \neq 0$. On notera la probabilité d'erreur de première espèce désirée α et la probabilité d'erreur de deuxième espèce désirée $\beta(\Delta)$. Les valeurs de α et $\beta(\Delta)$ seront toujours fixées au début de l'expérience.

On notera aussi la proportion d'individus assignés au traitement A par π . On s'intéressera particulièrement au cas où $\pi = 1/2$. On verra deux types d'expérience : celles à taille fixe, c'est-à-dire celles où le nombre total d'observations est fixe en fonction de π , α et $\beta(\Delta)$, et celles à étapes multiples, c'est-à-dire celles où le nombre total d'observations dépend de la performance des traitements dans l'expérience.

4.1 Expérience à taille fixe

On veut ici trouver le nombre total d'observations qui fait en sorte que l'on atteigne la puissance désirée tout en ayant une probabilité d'erreur de première espèce inférieure ou égale à α . On notera n le nombre total d'observations. L'inférence sur Δ se fait une

fois que l'expérience est terminée.

Supposons un instant que les résultats soient issus de la loi Normale(θ_1, σ_1^2) pour le traitement A et de la loi Normale(θ_2, σ_2^2) pour le traitement B avec σ_1^2 et σ_2^2 connus. Supposons qu'il y ait $n\pi$ observations pour le traitement A et $n(1-\pi)$ pour le traitement B . On notera \bar{X} et \bar{Y} les estimateurs du maximum de vraisemblance de θ_1 et θ_2 . Les lois de ces estimateurs sont la loi Normale($\theta_1, \frac{\sigma_1^2}{n\pi}$) et la loi Normale($\theta_2, \frac{\sigma_2^2}{n(1-\pi)}$) respectivement. Sous l'hypothèse $H_0 : \theta_1 = \theta_2$, on a

$$\mathbb{P} \left[|\bar{X} - \bar{Y}| < \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{\frac{\sigma_1^2}{\pi} + \frac{\sigma_2^2}{1-\pi}} \right] = 1 - \alpha, \quad (4.1)$$

où z_γ est le quantile d'ordre $1-\gamma$ de la loi Normale(0,1). Cela veut dire que si $|\bar{X} - \bar{Y}| < \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{\frac{\sigma_1^2}{\pi} + \frac{\sigma_2^2}{1-\pi}}$, on ne rejettera pas l'hypothèse H_0 disant que $\theta_1 = \theta_2$; si $\bar{X} - \bar{Y} < -\frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{\frac{\sigma_1^2}{\pi} + \frac{\sigma_2^2}{1-\pi}}$, on conclura que $\theta_1 < \theta_2$ et si $\bar{X} - \bar{Y} > \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{\frac{\sigma_1^2}{\pi} + \frac{\sigma_2^2}{1-\pi}}$, on conclura que $\theta_1 > \theta_2$.

Supposons maintenant que H_0 soit fausse et que $\theta_1 = \theta_2 + \Delta$ avec $\Delta \neq 0$. Dans tous les calculs qui suivront, on supposera sans perte de généralité que $\Delta > 0$. Pour avoir une erreur de deuxième espèce d'au plus $\beta(\Delta)$ pour un Δ assez grand, il faut trouver une valeur de n telle que

$$\mathbb{P} \left[|\bar{X} - \bar{Y}| < \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{\frac{\sigma_1^2}{\pi} + \frac{\sigma_2^2}{1-\pi}} \right] \cong \mathbb{P} \left[\bar{X} - \bar{Y} < \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{\frac{\sigma_1^2}{\pi} + \frac{\sigma_2^2}{1-\pi}} \right] \leq \beta(\Delta). \quad (4.2)$$

La dernière équation implique que

$$\mathbb{P} \left[\frac{\sqrt{n}(\bar{X} - \bar{Y} - \Delta)}{\sqrt{\frac{\sigma_1^2}{\pi} + \frac{\sigma_2^2}{1-\pi}}} < z_{\alpha/2} - \frac{\sqrt{n}\Delta}{\sqrt{\frac{\sigma_1^2}{\pi} + \frac{\sigma_2^2}{1-\pi}}} \right] \leq \beta(\Delta). \quad (4.3)$$

Sous l'hypothèse $\theta_1 = \theta_2 + \Delta$, $\frac{\sqrt{n}(\bar{X} - \bar{Y} - \Delta)}{\sqrt{\frac{\sigma_1^2}{\pi} + \frac{\sigma_2^2}{1-\pi}}}$ suit la loi Normale(0,1) et cela implique

que l'équation précédente est satisfaite si et seulement si

$$z_{\alpha/2} - \frac{\sqrt{n}\Delta}{\sqrt{\frac{\sigma_1^2}{\pi} + \frac{\sigma_2^2}{1-\pi}}} \leq -z_{\beta(\Delta)}. \quad (4.4)$$

Avec un peu de calcul, on peut voir que

$$n \geq \frac{(z_{\alpha/2} + z_{\beta(\Delta)})^2 \left(\frac{\sigma_1^2}{\pi} + \frac{\sigma_2^2}{1-\pi} \right)}{\Delta^2}. \quad (4.5)$$

En connaissant σ_1 , σ_2 et π au départ, on peut trouver la valeur de n au début de l'expérience.

Revenons au cas où les réponses aux traitements sont dichotomiques avec probabilités de succès p_1 pour le traitement A et p_2 pour le traitement B . Lorsque le nombre d'observations est suffisamment grand, on sait que \hat{p}_1 et \hat{p}_2 , les estimateurs du maximum de vraisemblance de p_1 et p_2 , sont approximativement normalement distribués. On peut donc approcher la valeur de n nécessaire à partir des équations précédentes. Sous l'hypothèse $H_0 : p = p_1 = p_2$, l'équation (4.1) devient

$$\mathbb{P} \left[|\hat{p}_1 - \hat{p}_2| < \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{\frac{p(1-p)}{\pi} + \frac{p(1-p)}{1-\pi}} \right] \cong 1 - \alpha \quad (4.6)$$

pour n suffisamment grand. Le problème ici c'est que non seulement on ne connaît pas la variance de $(\hat{p}_1 - \hat{p}_2)$, mais elle dépend de p . On s'occupera de ce problème un peu plus loin. Pour l'instant, on posera $V_0 = \frac{p(1-p)}{\pi} + \frac{p(1-p)}{1-\pi}$ et on le supposera connu. En supposant que \hat{p}_1 et \hat{p}_2 soient normalement distribués avec V_0 connu, on détectera que p_1 est plus grand que p_2 si et seulement si $\hat{p}_1 - \hat{p}_2 > \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{V_0}$ et que p_2 est plus grand que p_1 si et seulement si $\hat{p}_1 - \hat{p}_2 < -\frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{V_0}$.

Supposons maintenant que H_0 soit fautive et que $p_1 = p_2 + \Delta$ avec $0 < \Delta < 1 - p_2$. Pour avoir une erreur de deuxième espèce d'au plus $\beta(\Delta)$ pour Δ assez grand, il faut trouver une valeur de n telle que

$$\mathbb{P} \left[|\hat{p}_1 - \hat{p}_2| < \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{V_0} \right] \cong \mathbb{P} \left[\hat{p}_1 - \hat{p}_2 < \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{V_0} \right] \leq \beta(\Delta). \quad (4.7)$$

Comme $\mathbb{E}[\hat{p}_1 - \hat{p}_2] = \Delta$ et $V_1 \equiv \text{Var}[\hat{p}_1 - \hat{p}_2] = \frac{(p_2 + \Delta)(1 - p_2 - \Delta)}{\pi} + \frac{p_2(1 - p_2)}{1 - \pi}$, $\frac{\sqrt{n}(\hat{p}_1 - \hat{p}_2 - \Delta)}{\sqrt{V_1}}$ est approximativement distribué selon la loi Normale(0,1) lorsque n est suffisamment grand. À partir de l'équation (4.7) et du dernier résultat, on obtient

$$\mathbb{P} \left[\frac{\sqrt{n}(\hat{p}_1 - \hat{p}_2 - \Delta)}{\sqrt{V_1}} < \frac{z_{\alpha/2}\sqrt{V_0} - \sqrt{n}\Delta}{\sqrt{V_1}} \right] \leq \beta(\Delta), \quad (4.8)$$

ce qui mène à la condition suivante :

$$\frac{z_{\alpha/2}\sqrt{V_0} - \sqrt{n}\Delta}{\sqrt{V_1}} \leq -z_{\beta(\Delta)}. \quad (4.9)$$

Avec un peu de calcul, on peut voir que

$$n \geq \frac{(z_{\alpha/2}\sqrt{V_0} + z_{\beta(\Delta)}\sqrt{V_1})^2}{\Delta^2}. \quad (4.10)$$

Il y a deux problèmes liés à l'application de cette formule. Premièrement, cette formule provient de l'approximation de la loi Binomiale par la loi Normale. On a vu dans les chapitres 2 et 3 avec l'intervalle de confiance de Wald que l'approximation normale peut donner des résultats loin de la réalité. On peut se demander si l'approximation normale donne les résultats escomptés, autrement dit, si la valeur de n obtenue à partir de l'équation (4.10) donne des probabilités d'erreurs de première et deuxième espèces inférieures ou égales à α et $\beta(\Delta)$. Il se pourrait que l'une ou l'autre des probabilités erreurs soit trop élevée ou que les deux probabilités d'erreurs soient trop petites par rapport aux valeurs désirées, ce qui veut dire que le nombre d'observations est trop grand.

Le tableau 4.1 présente le nombre nécessaire d'observations obtenu à partir de l'équation (4.10), les probabilités de recouvrement et puissances exactes pour différentes valeurs de p_1 , p_2 lorsque $\pi = 1/2$, $\alpha = 0,05$ et la puissance désirée est 0,8. On met les vraies valeurs de p_1 , p_2 et de puissance désirée dans l'équation (4.10) pour obtenir le nombre d'observations désiré. On a posé $p = p_2$ dans l'expression de V_0 .

Par exemple, dans le cas où le traitement A donne un taux de succès de 0,8 et le

p_1	p_2	Nombre d'obs.	Probabilité de recouvrement			Puissance exacte		
			Wald	AC	BA	Wald	AC	BA
0,4	0,1	40	0,932	0,955	0,948	0,673	0,585	0,619
0,65	0,35	80	0,939	0,954	0,954	0,795	0,795	0,795
0,8	0,5	84	0,944	0,951	0,947	0,850	0,841	0,844
0,6	0,1	14	0,883	0,964	0,950	0,609	0,516	0,516
0,75	0,25	24	0,922	0,963	0,936	0,769	0,766	0,766
0,9	0,4	28	0,930	0,951	0,948	0,894	0,845	0,848

TAB. 4.1 – Probabilités de recouvrement et puissances lorsque n provient de l'équation (4.10) et la puissance désirée est 0,8. Dans tous les cas, $\pi = 1/2$ et $\alpha = 0,05$.

traitement B un taux de succès de 0,5, le nombre d'observations requis est donné par

$$n \geq \frac{\left(1,96 \times \sqrt{\frac{0,5 \times 0,5}{0,5} + \frac{0,5 \times 0,5}{0,5}} + 0,84 \times \sqrt{\frac{0,5 \times 0,5}{0,5} + \frac{0,8 \times 0,2}{0,5}}\right)^2}{0,3^2}. \approx 82,24. \quad (4.11)$$

Il faudrait donc 83 observations pour avoir la puissance désirée. On en prend 84 pour pouvoir avoir autant d'observations pour les deux traitements (42 de chaque côté).

Pour les probabilités de recouvrement et puissances exactes, on utilise trois intervalles vu au chapitre 3, soient les intervalles de Wald, d'Agresti-Coull (AC) et bayésien approximatif (BA). Pour un intervalle donné, la puissance est la probabilité que la borne inférieure soit plus grande que 0 puisque dans tous les cas, $p_1 > p_2$.

On voit que la puissance désirée est atteinte (ou presque atteinte) seulement lorsque p_2 n'est pas trop loin de 0,5. Dans les deux cas où $p_2 = 0,1$, on est très loin de la puissance désirée. Il faut dire que supposer que la variance dépend seulement de p_2 sous H_0 est une décision arbitraire puisque dans les cas présentés, H_0 n'est pas vraie. Pour corriger le manque de puissance observé, on peut poser $p = 1/2$ au lieu de p_2 dans l'expression de V_0 pour obtenir

$$V_0 = \frac{1}{4\pi(1-\pi)} \quad (4.12)$$

et ainsi maximiser la valeur de V_0 .

p_1	p_2	Nombre d'obs.	Probabilité de recouvrement			Puissance exacte		
			Wald	AC	BA	Wald	AC	BA
0,4	0,1	78	0,945	0,954	0,954	0,907	0,883	0,894
0,65	0,35	86	0,940	0,946	0,946	0,840	0,788	0,811
0,8	0,5	84	0,944	0,951	0,947	0,850	0,841	0,844
0,6	0,1	28	0,930	0,951	0,948	0,894	0,845	0,848
0,75	0,25	30	0,935	0,960	0,939	0,826	0,804	0,808
0,9	0,4	28	0,930	0,951	0,948	0,894	0,845	0,848

TAB. 4.2 – Probabilités de recouvrement et puissances lorsque n est obtenu par l'équation (4.10) en prenant V_0 de l'équation (4.12). Dans tous les cas, $\pi = 1/2$, $\alpha = 0,05$ et la puissance désirée est 0,8.

Le tableau 4.2 montre le nombre observations nécessaires obtenu par l'équation (4.10) en prenant V_0 de l'équation (4.12), les probabilités de recouvrement et puissances exactes lorsque $\pi = 1/2$, $\alpha = 0,05$ et la puissance désirée est 0,8.

Dans les cas où p_2 est loin de $1/2$, le nombre d'observations et la puissance se voient nettement augmentés. Le défaut de poser $p = 1/2$ dans V_0 est que dans certains cas cela peut demander un peu plus d'observations que ce qui serait vraiment requis pour atteindre la puissance désirée, mais au moins la puissance désirée est presque toujours atteinte.

Dans les tableaux 4.1 et 4.2, le nombre d'observations est déterminé à partir des vraies valeurs de p_1 et p_2 , ce qui est impossible en pratique. C'est le deuxième problème lié à l'application de l'équation (4.10). Cependant, il y a moyen de contourner ce problème. Par exemple, il peut arriver que l'on ait une connaissance de p_2 avant le début de l'expérience. Supposons par exemple que le traitement A est un nouveau traitement que l'on veut comparer avec le traitement B qui est un traitement utilisé depuis un certain temps. On pourrait utiliser les connaissances que l'on a sur l'efficacité du traitement B pour estimer V_0 et V_1 . On pourrait aussi estimer p_1 et p_2 à partir des observations obtenues tôt dans l'expérience.

Sinon, une façon très simple de procéder est de poser $p = 1/2$ dans l'expression de V_0 pour ainsi obtenir (4.12) et poser $p_2 = 1/2 - \Delta(1 - \pi)$ dans l'expression de V_1 pour obtenir

$$V_1 = \frac{(1/2 + \Delta\pi)(1/2 - \Delta\pi)}{\pi} + \frac{\{1/2 - \Delta(1 - \pi)\} \{1/2 + \Delta(1 - \pi)\}}{(1 - \pi)}. \quad (4.13)$$

p_1	p_2	Nombre d'obs.	Probabilité de recouvrement			Puissance exacte		
			Wald	AC	BA	Wald	AC	BA
0,4	0,1	86	0,944	0,952	0,950	0,928	0,912	0,919
0,65	0,35	86	0,940	0,946	0,946	0,840	0,788	0,811
0,8	0,5	86	0,942	0,953	0,952	0,861	0,851	0,856
0,6	0,1	30	0,919	0,954	0,952	0,899	0,844	0,875
0,75	0,25	30	0,935	0,960	0,939	0,826	0,804	0,808
0,9	0,4	30	0,919	0,954	0,952	0,899	0,844	0,875

TAB. 4.3 – Probabilités de recouvrement et puissances exactes lorsque n est obtenu par l'équation (4.10) en prenant V_0 de l'équation (4.12) et V_1 de l'équation (4.14). Dans tous les cas, $\pi = 1/2$, $\alpha = 0,05$ et la puissance désirée est 0,8.

Cela maximisera les valeurs de V_0 et V_1 et du même coup, cela donnera peut-être un peu plus d'observations que ce qui est réellement nécessaire, mais on s'assurera que les probabilités d'erreurs de première et deuxième espèces ne soient pas trop grandes. Dans le cas $\pi = 1/2$, l'équation (4.13) peut s'écrire

$$V_1 = (1 + \Delta)(1 - \Delta). \quad (4.14)$$

On voit dans le tableau 4.3 les probabilités de recouvrement et les puissances exactes lorsque n est obtenu par l'équation (4.10) en prenant V_0 de l'équation (4.12) et V_1 de l'équation (4.14). Dans tous les cas, $\pi = 1/2$, $\alpha = 0,05$ et la puissance désirée est 0,8.

Comme attendu, les puissances sont souvent plus élevées que la puissance désirée, surtout lorsque p_1 ou p_2 est loin de 1/2, mais au moins, les puissances seront généralement assez grandes et les nombres d'observations nécessaires du tableau 4.3 ne sont pas beaucoup plus grand que ceux du tableau 4.2.

4.2 Expérience à étapes multiples

Dans plusieurs situations, on sera intéressé à ce que l'expérience ne dure pas trop longtemps. Par exemple, si les observations coûtent cher à récolter ou si un traitement est nettement meilleur que l'autre. Dans ce dernier cas, une expérience qui dure trop longtemps allouerait inutilement des patients au moins bon traitement, ce qui d'un point de vue éthique n'est pas acceptable. Si un des deux traitements performe nette-

ment mieux que l'autre tôt dans l'expérience, on pourrait déclarer que ce traitement est le meilleur et arrêter l'expérience. Cela éviterait que d'autres patients reçoivent inutilement le moins bon traitement. Dans une expérience à étapes multiples, on regarde à plusieurs reprises dans l'expérience, au moyen de tests ou d'intervalles de confiance, si la différence entre les deux traitements est significative. Si c'est le cas, on arrête l'expérience. On supposera que l'on inclut un maximum de n observations au total avec possibilité de K tests ou intervalles durant l'expérience pour détecter s'il y a une différence significative entre les deux traitements.

4.2.1 Tests répétés

Une façon de découvrir rapidement si un traitement est meilleur que l'autre est de faire des tests statistiques à plusieurs moments dans l'expérience. Notons les statistiques utilisées pour les tests par $(Z_k, k \in \{1, \dots, K\})$. On rejettera l'hypothèse d'égalité des proportions p_1 et p_2 si pour au moins un $k \in \{1, \dots, K\}$, $|Z_k| > c_k$. Pour simplifier la notation, notons pour tout événement Ω et pour tout $-p_2 \leq \Delta \leq 1 - p_2$

$$\mathbb{P}_\Delta[\Omega] = \mathbb{P}[\Omega | p_1 = p_2 + \Delta]. \quad (4.15)$$

On voudra que

$$\mathbb{P}_0 \left[\bigcup_{k=1}^K \{|Z_k| > c_k\} \right] \leq \alpha \quad (4.16)$$

et

$$\mathbb{P}_\Delta \left[\bigcap_{k=1}^K \{|Z_k| \leq c_k\} \right] \leq \beta(\Delta) \quad (4.17)$$

pour un certain $\Delta \neq 0$. Pour $k \in \{1, \dots, K\}$, posera

$$\alpha_k = \mathbb{P}_0[|Z_1| \leq c_1, \dots, |Z_{k-1}| \leq c_{k-1}, |Z_k| > c_k]. \quad (4.18)$$

Le paramètre α_k est la probabilité d'erreur de première espèce dépensée au k^e test, c'est-à-dire la probabilité que l'expérience s'arrête après avoir fait le k^e test si $p_1 = p_2$. Les α_k pourront être choisis arbitrairement, à condition que $\sum_{k=1}^K \alpha_k \leq \alpha$.

On verra ici deux tests répétés qui donnent des α_k différents : les tests de Pocock et de O'Brien-Fleming.

Tests répétés de Pocock

Ces tests ont été proposés par Pocock (1977) et il les a utilisés lorsque le paramètre d'intérêt est la différence de deux espérances provenant de populations normales avec variance connue et identique pour les deux traitements. Il a aussi abordé brièvement le cas de populations dichotomiques.

Pour ces tests, on choisit $c \equiv c_1 = \dots = c_K$ de sorte que l'équation (4.16) soit satisfaite. Dans Pocock (1977), on suppose qu'il y a m observations de chaque traitement entre chaque test. Supposons que l'on ait des observations X_1, X_2, \dots indépendantes et identiquement distribuées de la loi Normale(μ_1, σ^2) et des observations Y_1, Y_2, \dots indépendantes et identiquement distribuées de la loi Normale(μ_2, σ^2). Les statistiques ($Z_k, k \in \{1, \dots, K\}$) sont définies par

$$Z_k = \frac{\sqrt{km}(\bar{X}_{km} - \bar{Y}_{km})}{\sqrt{2}\sigma}, \quad (4.19)$$

où $\bar{X}_{km} = \frac{1}{km} \sum_{i=1}^{km} X_i$ et $\bar{Y}_{km} = \frac{1}{km} \sum_{i=1}^{km} Y_i$. Il est facile de voir que la loi de Z_k est la loi Normale(0, 1), peu importe les valeurs de m et de k . Cependant, cela ne veut pas dire que la valeur c soit facile à trouver. La principale difficulté est qu'il faut tenir compte de la dépendance entre les ($Z_k, k \in \{1, \dots, K\}$). Pocock (1977) a calculé dans son tableau 1 les valeurs de c et de α_1 pour $\alpha = 0,05$ et $0,01$ et pour K allant de 2 à 20.

Posons $n = 2mK$, le nombre maximum d'individus dans l'expérience. La puissance des tests dépend de m et de K . Les résultats présentés dans le tableau 2 de Pocock (1977) montrent que pour atteindre une certaine puissance et avec une erreur de première espèce fixe, la valeur de n augmente lorsque K augmente. Il est donc possible que l'expérience dure plus longtemps lorsque l'on fait des tests répétés ($K \geq 2$) que lorsque l'on fait un seul test à la fin de l'expérience ($K = 1$). Par contre, les tableaux 2 et 3 de Pocock (1977) montrent que pour atteindre une certaine puissance et avec une erreur de première espèce fixe, plus la valeur de K est grande, plus le nombre moyen d'observations nécessaires est petit.

Donc, lorsqu'on a des données normalement distribuées avec variance connue et où

les tests sont toujours espacés par le même nombre d'observations (qui est égal pour chaque traitement), on sait que les expériences à tests répétés de Pocock performant mieux que les expériences à test unique, dans le sens où ils nécessitent moins d'observations en moyenne. On peut se demander comment performant ces tests si l'on déroge du cadre proposé. Par exemple, on peut se demander ce qui se passe lorsque les observations sont issues de la loi Bernoulli.

On utilisera les tests répétés de Pocock dans le cas où X_1, X_2, \dots sont indépendantes et identiquement distribuées de la loi Bernoulli(p_1) et Y_1, Y_2, \dots sont indépendantes et identiquement distribuées de la loi Bernoulli(p_2). On aura encore m observations pour chaque traitement entre chaque test. On comparera les statistiques suivantes :

$$Z_k^{(u)} = \frac{\sqrt{km}(\hat{p}_{1,km} - \hat{p}_{2,km})}{\sqrt{\hat{p}_{1,km}(1 - \hat{p}_{1,km}) + \hat{p}_{2,km}(1 - \hat{p}_{2,km})}} \quad (4.20)$$

où $\hat{p}_{1,km} = \frac{1}{km} \sum_{i=1}^{km} X_i$ et $\hat{p}_{2,km} = \frac{1}{km} \sum_{i=1}^{km} Y_i$ et

$$Z_k^{(p)} = \frac{\sqrt{km}(\hat{p}_{1,km} - \hat{p}_{2,km})}{\sqrt{2\hat{p}_{2km}(1 - \hat{p}_{2km})}} \quad (4.21)$$

où \hat{p}_{2km} est la moyenne des $2km$ premières observations dans l'expérience (peu importe le traitement). Dans notre cas, pour $k \in \{1, \dots, K\}$, $\hat{p}_{2km} = \frac{\hat{p}_{1,km} + \hat{p}_{2,km}}{2}$ puisqu'il y a autant d'observations pour chaque traitement après $2km$ assignations. Le u dans $Z_k^{(u)}$ est pour *unpooled variance estimator* et le p dans $Z_k^{(p)}$ est pour *pooled variance estimator*. Dans $Z_k^{(u)}$, la variance de $\hat{p}_1 - \hat{p}_2$ est estimée à partir de \hat{p}_1 et \hat{p}_2 alors que dans $Z_k^{(p)}$, on utilise le fait que sous H_0 , toutes les observations suivent la même loi et qu'il vaut mieux toutes les utiliser pour estimer $p = p_1 = p_2$ dans la variance de $\hat{p}_1 - \hat{p}_2$. Dans Pocock (1977), on propose d'utiliser $Z_k^{(p)}$ et Lin et al. (1991) ont comparé les erreurs de première espèce obtenues à partir des statistiques $Z_k^{(u)}$ et $Z_k^{(p)}$ pour $m = 5, 10$ et 15 et $p_1 = p_2 = 0, 1, 0,2 \dots, 0,5$ lorsque $\alpha = 0,05$ et $K = 3$. On voit que les erreurs de premières espèces sont trop élevées lorsqu'on utilise $Z_k^{(u)}$. Sur les 15 erreurs de première espèce obtenues à partir de $Z_k^{(u)}$, seulement deux sont inférieures à $0,05$ ($m = 5$ et 10 , $p_1 = p_2 = 0, 1$) et la plus élevée est de $0,146$ ($m = 5$, $p_1 = p_2 = 0, 5$). Pour celles obtenues à partir de $Z_k^{(p)}$, la plupart sont inférieures à $0,05$ et la plus élevée est de $0,054$ ($m = 15$, $p_1 = p_2 = 0, 5$). Il n'est pas surprenant de voir que les tests avec $Z_k^{(u)}$ rejettent l'hypothèse $H_0 : p_1 = p_2$ trop souvent puisque les valeurs de $\hat{p}_{1,km}$ et $\hat{p}_{2,km}$ pour lesquelles $\left| Z_k^{(u)} \right| > z_{\gamma/2}$ sont celles pour lesquelles la valeur 0 est exclue de l'intervalle de Wald pour une différence de deux proportions de niveau $(1 - \gamma)100$ %. Or, on a vu au

Puissance désirée = 0,75						Puissance désirée = 0,90					
K	m	p_1	p_2	Total moyen	Puissance exacte	K	m	p_1	p_2	Total moyen	Puissance exacte
1	39	0,4	0,1	78	0,895	1	59	0,4	0,1	118	0,978
2	22	0,4	0,1	63,5	0,902	2	33	0,4	0,1	81,2	0,980
5	10	0,4	0,1	58,4	0,910	5	15	0,4	0,1	65,6	0,984
1	39	0,65	0,35	78	0,778	1	59	0,65	0,35	118	0,917
2	22	0,65	0,35	70,4	0,753	2	33	0,65	0,35	89,4	0,919
5	10	0,65	0,35	67,0	0,777	5	15	0,65	0,35	82,0	0,929
1	39	0,8	0,5	78	0,813	1	59	0,8	0,5	118	0,938
2	22	0,8	0,5	67,4	0,822	2	33	0,8	0,5	88,0	0,943
5	10	0,8	0,5	64,5	0,829	5	15	0,8	0,5	76,3	0,951
1	14	0,6	0,1	28	0,848	1	22	0,6	0,1	44	0,962
2	8	0,6	0,1	24,1	0,864	2	12	0,6	0,1	30,8	0,967
5	4	0,6	0,1	24,3	0,896	5	6	0,6	0,1	26,7	0,983
1	14	0,75	0,25	28	0,762	1	22	0,75	0,25	44	0,938
2	8	0,75	0,25	25,2	0,784	2	12	0,75	0,25	33,2	0,936
5	4	0,75	0,25	24,5	0,860	5	6	0,75	0,25	29,0	0,959

TAB. 4.4 – Puissance pour différents K des tests de Pocock. Dans tous les cas, $\pi = 1/2$ et $\alpha = 0,05$. Le nombre maximal d'observations au total est $2mK$. “Total moyen” est le nombre d'observations nécessaires en moyenne pour les deux échantillons.

chapitre 3 que l'intervalle de Wald était anti-conservateur, y compris lorsque $p_1 = p_2$.

Pour ce qui est de la puissance, il est intéressant de comparer le cas $K = 1$ avec $K \geq 2$ dans le cas de populations dichotomiques. Pocock (1977) l'a fait, mais pour des populations normales. Le tableau 4.4 présente le nombre moyen d'observations nécessaire, les probabilités de recouvrement et puissances exactes pour différentes valeurs de p_1 , p_2 et de puissance désirée lorsque $\pi = 1/2$ et $\alpha = 0,05$. Les statistiques $(Z_k^{(p)}, k \in \{1, \dots, K\})$ sont utilisées pour les tests et le nombre d'observations séparant chaque test est obtenu à partir du tableau 2 de Pocock (1977) en posant $\sigma = \sqrt{1/2(1 - 1/2)} = 1/2$ pour maximiser la variance.

On voit que la puissance a tendance à augmenter avec le nombre de tests effectués, même si en général le nombre moyen d'observations diminue plus on fait de tests. La probabilité de conclure à tort que p_2 est plus grand que p_1 est très légèrement plus élevée en général si on fait cinq tests au lieu d'un seul, mais c'est négligeable. Pour les valeurs de K , p_1 et p_2 étudiées, c'est lorsque $K = 5$, $p_1 = 0,65$ et $p_2 = 0,35$ que la probabilité de conclure à tort que p_2 est plus grand que p_1 est la plus élevée et elle est

environ égale à 6×10^{-5} .

Il est intéressant de voir ce qui se passe si le nombre d'observations est $m_{1,k}$ pour le traitement A et $m_{2,k}$ pour le traitement B après k tests où $m_{1,k} \neq m_{2,k}$. Si on revient au cas où les observations sont normalement distribuées avec variance connue, on peut trouver la valeur de c qui donne exactement le seuil α désiré puisque les paramètres inconnus μ_1 et μ_2 n'interviennent pas dans les calculs. La valeur de c dépend entre autre de la covariance entre les différences des moyennes de deux étapes successives. Dans le cas où les observations sont issues de la loi Bernoulli, la covariance de la différence de proportions entre deux étapes peut dépendre de p_1 et p_2 . Posons $\hat{\Delta}_k$, la différence de proportions observée après k tests ($k \in \{1, \dots, K\}$). Le coefficient de corrélation entre $\hat{\Delta}_k$ et $\hat{\Delta}_l$ ($1 \leq k < l \leq K$) est donnée par

$$\sqrt{\frac{p_1(1-p_1)/m_{1,k} + p_2(1-p_2)/m_{2,k}}{p_1(1-p_1)/m_{1,l} + p_2(1-p_2)/m_{2,l}}}, \quad (4.22)$$

(voir Lin et al., 1991). On peut voir que cette expression n'est pas fonction de p_1 et p_2 si et seulement si $m_{1,k}/m_{2,k} = m_{1,l}/m_{2,l}$. Donc, si la proportion d'individus assignés au traitement A est toujours la même d'un test à l'autre, la covariance ne dépendra pas des paramètres inconnus p_1 et p_2 . Si on tient compte seulement de l'inférence statistique, il est souhaitable de toujours avoir les mêmes proportions allant aux traitements A et B entre chaque test.

Lin et al. (1991) ont observé les probabilités d'erreur de première espèce pour des expériences avec 100 observations au total lorsque $p_1 = p_2$, $K = 3$, $\alpha = 0,05$ et pour différentes valeurs de $m_{1,k}$ et $m_{2,k}$ ($k \in \{1, 2, 3\}$) en utilisant les même valeurs critiques pour les tests que si on avait $m = m_{1,k} = m_{2,k}$. Dans tous les cas, les probabilités de recouvrement sont quand même assez près du niveau de confiance nominal lorsque la statistique $Z_k^{(p)}$ est utilisée. On regardera plus en détail la performance des tests répétés de Pocock lorsque $m_{1,k}/m_{2,k} \neq m_{1,l}/m_{2,l}$ au prochain chapitre.

Les tests répétés de Pocock donnent toujours $\alpha_1 > \dots > \alpha_K$. Si les tests sont toujours séparés par le même nombre d'observations, cela implique que ces tests donnent plus d'importance aux premières observations qu'aux dernières. Dans le cas où p_1 et p_2 ne sont pas égales, cela implique aussi que la probabilité que l'expérience se termine très rapidement est relativement élevée. Par contre, si on ne détecte pas l'inégalité très tôt dans l'expérience, alors il deviendra difficile de la détecter plus loin dans l'expérience.

Tests répétés de O'Brien-Fleming

Ces tests ont été proposés par O'Brien et Fleming (1979) et ressemblent aux tests de Pocock. La différence est que maintenant on a $c_k = \sqrt{K/k} c_K$ pour $k \in \{1, \dots, K\}$. En fait, si la statistique utilisée était la différence des sommes des observations des deux traitements et qu'il y avait autant d'observations de chaque traitement entre chaque étape, la valeur critique serait la même à toutes les étapes. Ces tests donnent moins d'importance aux premières observations que les tests de Pocock.

Supposons encore une fois qu'il y ait m observations de chaque traitement entre chaque test. Comme pour les tests de Pocock, les valeurs de c_k sont obtenues de sorte que l'équation (4.16) soit satisfaite en supposant que les observations soient normalement distribuées et qu'on utilise la statistique (4.19). On utilisera ces tests dans le cas où les observations sont dichotomiques.

Lin et al. (1991) ont aussi obtenu les probabilités d'erreur de première espèce à partir des statistiques $Z_k^{(u)}$ et $Z_k^{(p)}$ pour $m = 5, 10$ et 15 et $p_1 = p_2 = 0, 1, 0,2 \dots, 0,5$ lorsque $\alpha = 0,05$ et $K = 3$ pour les tests de O'Brien-Fleming. Les statistiques $Z_k^{(u)}$ donnent encore des probabilités d'erreur de première espèce trop élevées, quoi qu'un peu moins élevées que pour les tests de Pocock. Les statistiques $Z_k^{(p)}$ donnent encore des probabilités d'erreur de première espèce près de $0,05$.

Pour ce qui est de la puissance, le tableau 4.5, comme le tableau 4.4, montre le nombre d'observations moyen, les probabilités de recouvrement et puissances pour différentes valeurs de p_1, p_2 et de puissance désirée lorsque $\pi = 1/2$ et $\alpha = 0,05$, mais cette fois-ci pour les tests de O'Brien-Fleming. Encore une fois, les statistiques ($Z_k^{(p)}, k \in \{1, \dots, K\}$) sont utilisées pour les tests. Comme Pocock (1977) l'a fait pour ses tests, on a calculé la valeur de $\Delta\sqrt{n}/\sqrt{2}\sigma$ nécessaire pour atteindre une certaine puissance étant donné une probabilité d'erreur de première espèce α et le nombre de tests prévu K lorsqu'on fait les tests de O'Brien-Fleming en supposant que les observations soient Normales avec variance connue. On a posé $\sigma = 1/2$.

On remarque d'abord que, comme pour les tests de Pocock, le fait de faire plusieurs tests diminue le nombre d'observations moyen sans diminuer la puissance. Aussi, la probabilité de détecter à tort que p_2 est plus grand que p_1 est encore négligeable. Si on compare les résultats obtenus à partir des tests de O'Brien-Fleming (tableau 4.5) avec ceux obtenus à partir des tests de Pocock (tableau 4.4), on constate que les tests de O'Brien-Fleming ont souvent besoin de plus d'observations en moyenne que les tests de Pocock sans donner de meilleures puissances. Cela est surtout vrai lorsque la puissance désirée est $0,90$.

Puissance désirée = 0,75						Puissance désirée = 0,90					
K	m	p_1	p_2	Total moyen	Puissance exacte	K	m	p_1	p_2	Total moyen	Puissance exacte
1	39	0,4	0,1	78	0,895	1	59	0,4	0,1	118	0,978
2	20	0,4	0,1	70,0	0,902	2	30	0,4	0,1	91,1	0,978
5	8	0,4	0,1	60,2	0,893	5	13	0,4	0,1	82,1	0,984
1	39	0,65	0,35	78	0,778	1	59	0,65	0,35	118	0,917
2	20	0,65	0,35	71,8	0,797	2	30	0,65	0,35	99,2	0,924
5	8	0,65	0,35	66,2	0,748	5	13	0,65	0,35	91,6	0,933
1	39	0,8	0,5	78	0,813	1	59	0,8	0,5	118	0,938
2	20	0,8	0,5	71,9	0,818	2	30	0,8	0,5	97,8	0,940
5	8	0,8	0,5	63,4	0,818	5	13	0,8	0,5	89,0	0,952
1	14	0,6	0,1	28	0,848	1	22	0,6	0,1	44	0,962
2	7	0,6	0,1	26,8	0,848	2	11	0,6	0,1	35,7	0,963
5	3	0,6	0,1	24,3	0,852	5	5	0,6	0,1	32,7	0,978
1	14	0,75	0,25	28	0,762	1	22	0,75	0,25	44	0,938
2	7	0,75	0,25	26,6	0,763	2	11	0,75	0,25	36,7	0,938
5	3	0,75	0,25	24,3	0,813	5	5	0,75	0,25	34,9	0,950

TAB. 4.5 – Puissance pour différents K des tests de O'Brien-Fleming. Dans tous les cas, $\pi = 1/2$ et $\alpha = 0,05$.

Pour les mêmes raisons que pour les tests de Pocock, on peut douter de la bonne performance des tests de O'Brien-Fleming lorsque la proportion d'individus assignés au traitement A n'est pas toujours la même d'un test à l'autre. Comme ils l'ont fait pour les tests de Pocock, Lin et al. (1991) ont montré pour quelques valeurs de $m_{1,k}$ et $m_{2,k}$ ($k \in \{1, 2, 3\}$) que les erreurs de premières espèces étaient près du niveau de confiance nominal (0,05) lorsque les statistiques $Z_k^{(p)}$ étaient utilisées.

REMARQUE : On peut se demander si on obtiendrait la même puissance sans avoir besoin de plus d'observations si on faisait des tests indépendants, c'est-à-dire que pour le k^e test, on utilise seulement les observations obtenues entre le $(k-1)^e$ et le k^e test. Il serait alors beaucoup plus facile de trouver la valeur critique c utilisée pour chaque test et la valeur de $D \equiv \Delta\sqrt{m}/\sqrt{2}\sigma$ permettant d'obtenir la puissance désirée. Toujours en supposant que les observations soient normalement distribuées avec variance σ^2 connue, on peut montrer facilement que

$$c = \Phi^{-1} \left(\frac{\alpha^{1/K} + 1}{2} \right) \quad (4.23)$$

et

$$1 - \beta(\Delta) = 1 - \{\Phi(c - D) - \Phi(-c - D)\}^K. \quad (4.24)$$

Le tableau 4.6 montre la performance des tests indépendants dans le cas où les observations sont dichotomiques. La valeur de m est encore une fois déterminée en posant $\sigma = 1/2$.

On voit que les tests indépendants nécessitent souvent beaucoup plus d'observations en moyenne sans donner de meilleures puissances. En plus, le nombre d'observations moyen est souvent plus grand lorsqu'on fait des tests multiples ($K > 1$) que lorsqu'on fait un seul test ($K = 1$). Il n'est donc pas souhaitable d'utiliser cette approche.

4.2.2 Intervalles de confiance répétés

Les tests répétés montrent directement si l'hypothèse H_0 doit être rejetée ou non. Dans notre cas, cette hypothèse est l'égalité entre p_1 et p_2 . En plus de s'intéresser à savoir si $p_1 = p_2$, on peut vouloir obtenir un intervalle de confiance sur la différence entre p_1 et p_2 . Soit I_k le k^e intervalle de confiance pour Δ ($k \in \{1, \dots, K\}$). On peut décider d'arrêter l'expérience aussitôt que la valeur 0 soit à l'extérieur d'un des I_k . Le

Puissance désirée = 0,75						Puissance désirée = 0,90					
K	m	p_1	p_2	Total moyen	Puissance exacte	K	m	p_1	p_2	Total moyen	Puissance exacte
1	39	0,4	0,1	78	0,895	1	59	0,4	0,1	118	0,978
2	28	0,4	0,1	74,8	0,887	2	41	0,4	0,1	94,6	0,976
5	20	0,4	0,1	98,6	0,896	5	28	0,4	0,1	101,9	0,979
1	39	0,65	0,35	78	0,778	1	59	0,65	0,35	118	0,917
2	28	0,65	0,35	84,2	0,746	2	41	0,65	0,35	108,4	0,897
5	20	0,65	0,35	124,9	0,743	5	28	0,65	0,35	132,8	0,912
1	39	0,8	0,5	78	0,813	1	59	0,8	0,5	118	0,938
2	28	0,8	0,5	81,3	0,796	2	41	0,8	0,5	102,1	0,940
5	20	0,8	0,5	112,2	0,825	5	28	0,8	0,5	125,3	0,932
1	14	0,6	0,1	28	0,848	1	22	0,6	0,1	44	0,962
2	11	0,6	0,1	29,9	0,873	2	15	0,6	0,1	36,4	0,955
5	7	0,6	0,1	41,6	0,783	5	11	0,6	0,1	44,9	0,958
1	14	0,75	0,25	28	0,762	1	22	0,75	0,25	44	0,938
2	11	0,75	0,25	32,0	0,792	2	15	0,75	0,25	39,4	0,903
5	7	0,75	0,25	40,2	0,808	5	11	0,75	0,25	52,3	0,911

TAB. 4.6 – Puissance pour différents K des tests indépendants. Dans tous les cas, $\pi = 1/2$ et $\alpha = 0,05$.

dernier intervalle de confiance produit sera celui que l'on conservera pour l'inférence. On voudra alors que

$$\mathbb{P}_0 [\{0 \in I_1\}, \dots, \{0 \in I_{k-1}\}, \{0 \notin I_k\}] = \alpha_k \quad (4.25)$$

avec $\sum_{k=1}^K \alpha_k \leq \alpha$ et

$$\mathbb{P}_\Delta [\cap_{k=1}^K \{0 \in I_k\}] \leq \beta(\Delta) \quad (4.26)$$

pour un certain $\Delta \neq 0$.

On peut aussi décider de faire des tests répétés et une fois l'expérience terminée, construire l'intervalle de confiance de niveau $(1 - \alpha)$. On voudra alors que

$$\mathbb{P}_0 [0 \notin I_T] \leq \alpha \quad (4.27)$$

et

$$\mathbb{P}_\Delta [0 \in I_T] \leq \beta(\Delta) \quad (4.28)$$

où T est le nombre de tests effectués dans l'expérience. Dans un cas comme dans l'autre, le problème c'est que les intervalles standards, c'est-à-dire ceux utilisés lorsque $K = 1$, ne performant pas nécessairement bien. Par exemple, dans le cas où X_1, X_2, \dots sont indépendantes et identiquement distribuées de la loi Normale(μ_1, σ^2) et Y_1, Y_2, \dots sont indépendantes et identiquement distribuées de la loi Normale(μ_2, σ^2) avec σ^2 connu, Tsiatis et al. (1984) ont présenté les probabilités de recouvrement de l'intervalle de confiance standard pour $\mu_1 - \mu_2$, soit

$$\frac{1}{km} \sum_{i=1}^{km} X_i - \frac{1}{km} \sum_{i=1}^{km} Y_i \pm z_{\alpha/2} \sqrt{2\sigma / \sqrt{km}}, \quad (4.29)$$

lorsqu'on fait les tests de Pocock et les tests de O'Brien-Fleming avec $K = 5$ pour différentes valeurs de $\mu_1 - \mu_2$. Dans tous les cas présentés, les probabilités de recouvrement sont près du niveau de confiance nominal, mais pas égales à ce niveau de confiance,

K	m	p_1	p_2	Probabilités de recouvrement		
				Wald	AC	BA
1	39	0,1	0,1	0,941	0,970	0,970
2	22	0,1	0,1	0,933	0,958	0,958
5	10	0,1	0,1	0,935	0,956	0,956
1	39	0,25	0,25	0,940	0,955	0,948
2	22	0,25	0,25	0,932	0,938	0,938
5	10	0,25	0,25	0,923	0,931	0,929
1	39	0,5	0,5	0,946	0,946	0,946
2	22	0,5	0,5	0,919	0,943	0,941
5	10	0,5	0,5	0,921	0,921	0,921

TAB. 4.7 – Probabilités de recouvrement des intervalles de Wald, AC et BA de niveau 95 % construits après une expérience où on a utilisé les tests de Pocock de niveau 95 %. Dans tous les cas, $\pi = 1/2$ et la statistique $Z_k^{(p)}$ a été utilisée.

comme cela aurait été le cas si on avait eu $K = 1$. La raison pour laquelle les probabilités de recouvrement ne sont pas identiques au niveau de confiance nominal est que la probabilité de recouvrement de I_k ($k = 1, \dots, n$) n'est pas indépendante de T et T n'est pas indépendante de $\mu_1 - \mu_2$. Revenons au cas où les observations sont issues des lois Bernoulli(p_1) et Bernoulli(p_2). On a que

$$\mathbb{P}_\Delta[p_1 - p_2 \in I_T] = \sum_{k=1}^K \mathbb{P}_\Delta[p_1 - p_2 \in I_k | T = k] \mathbb{P}_\Delta[T = k]. \quad (4.30)$$

Si on avait $\{p_1 - p_2 \in I_k\}$ indépendant de $\{T = k\}$, on pourrait prendre des I_k ayant tous un niveau de confiance de $(1 - \alpha)$, mais ce n'est pas le cas. Si on procède de sorte que l'équation (4.27) soit satisfaite, la probabilité de recouvrement sera trop petite si la valeur de $|p_1 - p_2|$ est petite, en particulier si $p_1 = p_2$, puisqu'il est fort probable que $p_1 - p_2$ soit dans l'intervalle I_T seulement si $T = K$ et on a $\mathbb{P}_0[T = K] < 1$. Les tableaux 4.7 et 4.8 montrent les probabilités de recouvrement des intervalles de Wald, AC et BA de niveau 95 % construits après avoir fait des tests de Pocock et de O'Brien-Fleming de niveau 95 %. Le nombre d'observations entre les tests a été choisi en supposant que l'on veuille détecter une différence de 0,3 entre p_1 et p_2 avec probabilité 0,75.

On voit que les probabilités de recouvrement ont tendance à diminuer lorsqu'on augmente K . On observe une probabilité de recouvrement de 0,921 pour les intervalles AC et BA construits après avoir fait des tests de Pocock pour $p_1 = p_2 = 0,5$ avec $K = 5$. On n'a observé des probabilités de recouvrement aussi faibles pour aucune des

K	m	p_1	p_2	Probabilités de recouvrement		
				Wald	AC	BA
1	39	0,1	0,1	0,941	0,970	0,970
2	20	0,1	0,1	0,941	0,969	0,969
5	8	0,1	0,1	0,937	0,962	0,962
1	39	0,25	0,25	0,940	0,955	0,948
2	20	0,25	0,25	0,945	0,952	0,949
5	8	0,25	0,25	0,939	0,944	0,942
1	39	0,5	0,5	0,946	0,946	0,946
2	20	0,5	0,5	0,941	0,941	0,941
5	8	0,5	0,5	0,938	0,938	0,938

TAB. 4.8 – Probabilités de recouvrement des intervalles de Wald, AC et BA de niveau 95 % construits après une expérience où on a utilisé les tests de O'Brien-Fleming de niveau 95 %. Dans tous les cas, $\pi = 1/2$ et la statistique $Z_k^{(p)}$ a été utilisée.

330 valeurs de (n_1, p_1, n_2, p_2) étudiées au chapitre 3. On voit que les probabilités de recouvrement sont plus près du niveau de confiance nominal avec les tests de O'Brien-Fleming puisque lorsque $p_1 = p_2$, $\mathbb{P}[T = K]$ est plus près de 1.

Si on procède de sorte que l'équation (4.25) soit satisfaite, lorsque la valeur de $|p_1 - p_2|$ est grande, la probabilité que l'expérience se termine rapidement est élevée et la probabilité que $p_1 - p_2$ soit dans l'intervalle sachant que l'expérience se termine rapidement sera plus élevée que $(1 - \alpha)$. La probabilité de recouvrement sera donc trop élevée, ce qui peut avoir comme effet de donner une puissance trop petite. On voit dans les tableaux 4.9 et 4.10 les probabilités de recouvrement et les puissances des intervalles de Wald, AC et BA lorsque les α_k sont les mêmes que ceux des tests de Pocock (tableau 4.9) et que ceux des tests de O'Brien-Fleming (tableau 4.10) de niveau 95 %.

Cette fois-ci, on voit que les probabilités de recouvrement augmentent lorsque K augmente pour les intervalles AC et BA. Pour l'intervalle de Wald, cela n'est pas nécessairement le cas puisque lorsque K est grand, on a besoin de moins d'observations en moyenne et cet intervalle est anti-conservateur dans ces cas là. Dans les deux tableaux, on voit que les probabilités de recouvrement et les puissances sont plus élevées lorsque les α_k sont ceux des tests de Pocock. Il est intéressant de remarquer que pour les intervalles AC et BA, les puissances n'ont pas tendance à diminuer lorsque K augmente même si les probabilités de recouvrement augmentent.

Une autre approche est de construire un intervalle après avoir fait des tests répétés qui ne tient pas seulement compte de la statistique observée lors de la fin de l'expérience,

K	m	p_1	p_2	Probabilité de recouvrement			Puissance		
				Wald	AC	BA	Wald	AC	BA
1	39	0,4	0,1	0,945	0,954	0,954	0,907	0,883	0,894
2	22	0,4	0,1	0,965	0,974	0,973	0,909	0,902	0,890
5	10	0,4	0,1	0,967	0,985	0,984	0,922	0,912	0,903
1	39	0,65	0,35	0,943	0,953	0,943	0,778	0,778	0,778
2	22	0,65	0,35	0,956	0,972	0,972	0,820	0,768	0,753
5	10	0,65	0,35	0,948	0,983	0,978	0,820	0,809	0,805
1	39	0,8	0,5	0,942	0,949	0,946	0,826	0,802	0,813
2	22	0,8	0,5	0,964	0,974	0,974	0,839	0,830	0,819
5	10	0,8	0,5	0,955	0,984	0,981	0,850	0,837	0,829
1	14	0,6	0,1	0,930	0,951	0,948	0,894	0,845	0,848
2	8	0,6	0,1	0,933	0,971	0,971	0,901	0,875	0,858
5	4	0,6	0,1	0,908	0,988	0,988	0,934	0,910	0,891
1	14	0,75	0,25	0,929	0,953	0,953	0,862	0,762	0,762
2	8	0,75	0,25	0,928	0,974	0,974	0,872	0,855	0,782
5	4	0,75	0,25	0,895	0,989	0,989	0,881	0,862	0,860

TAB. 4.9 – Probabilités de recouvrement et puissances des intervalles de Wald, AC et BA où les α_k sont ceux des tests de Pocock de niveau 95 %. Dans tous les cas, $\pi = 1/2$ et la puissance désirée est 0,75.

K	m	p_1	p_2	Probabilité de recouvrement			Puissance		
				Wald	AC	BA	Wald	AC	BA
1	39	0,4	0,1	0,945	0,954	0,954	0,907	0,883	0,894
2	20	0,4	0,1	0,968	0,968	0,966	0,906	0,894	0,887
5	8	0,4	0,1	0,973	0,974	0,974	0,908	0,892	0,883
1	39	0,65	0,35	0,943	0,953	0,943	0,778	0,778	0,778
2	20	0,65	0,35	0,958	0,969	0,969	0,798	0,797	0,797
5	8	0,65	0,35	0,972	0,974	0,975	0,809	0,773	0,753
1	39	0,8	0,5	0,942	0,949	0,946	0,826	0,802	0,813
2	20	0,8	0,5	0,967	0,969	0,966	0,829	0,819	0,813
5	8	0,8	0,5	0,971	0,975	0,976	0,830	0,821	0,816
1	14	0,6	0,1	0,930	0,951	0,948	0,894	0,845	0,848
2	7	0,6	0,1	0,960	0,963	0,963	0,885	0,847	0,809
5	3	0,6	0,1	0,825	0,967	0,967	0,899	0,853	0,851
1	14	0,75	0,25	0,929	0,953	0,953	0,862	0,762	0,762
2	7	0,75	0,25	0,963	0,970	0,970	0,865	0,769	0,759
5	3	0,75	0,25	0,808	0,977	0,977	0,842	0,815	0,815

TAB. 4.10 – Probabilités de recouvrement et puissances des intervalles de Wald, AC et BA où les α_k sont ceux des tests de O'Brien-Fleming de niveau 95 %. Dans tous les cas, $\pi = 1/2$ et la puissance désirée est 0,75.

mais aussi du nombre de tests effectués.

Considérons (W_T, T) où W_T est la valeur de la statistique du T^e test. Soient (w_{t_1}, t_1) et (w_{t_2}, t_2) deux valeurs de (W_T, T) . On dira que (w_{t_1}, t_1) est plus grand que (w_{t_2}, t_2) si et seulement si une des trois conditions suivantes est satisfaite :

1. $t_1 = t_2$ et $w_{t_1} > w_{t_2}$;
2. $t_1 < t_2$ et $w_{t_1} > c_{t_1}$;
3. $t_1 > t_2$ et $w_{t_2} < -c_{t_2}$.

Revenons au cas où les observations sont Normales avec espérances μ_1 et μ_2 inconnues et variance σ^2 connue. Pour une valeur $(W_T, T) = (w_t, t)$ observée, Tsiatis et al. (1984) obtiennent un intervalle de confiance pour $\mu_1 - \mu_2$ de niveau $(1 - \alpha)100$ %. La borne inférieure de leur intervalle est le Δ_I pour lequel

$$\mathbb{P}[(W_T, T) > (w_t, t) | \Delta_I = \mu_1 - \mu_2] = \alpha/2 \quad (4.31)$$

et la borne supérieure est le Δ_S pour lequel

$$\mathbb{P}[(W_T, T) < (w_t, t) | \Delta_S = \mu_1 - \mu_2] = \alpha/2. \quad (4.32)$$

Dans le cas où les observations sont dichotomiques, on peut obtenir l'intervalle de confiance de la forme (Δ_I, Δ_S) de niveau $(1 - \alpha)100$ % pour $p_1 - p_2$ en trouvant les valeurs Δ_I et Δ_S de la façon suivante (Lin et al., 1991) :

$$\Delta_I = \inf_{-1 \leq \Delta \leq 1} \left\{ \Delta : \sup_{p_2} \mathbb{P}_\Delta[(W_T, T) \geq (w_t, t)] \geq \alpha/2 \right\} \quad (4.33)$$

et

$$\Delta_S = \sup_{-1 \leq \Delta \leq 1} \left\{ \Delta : \sup_{p_2} \mathbb{P}_\Delta[(W_T, T) \leq (w_t, t)] \geq \alpha/2 \right\}, \quad (4.34)$$

où p_2 doit être dans l'intervalle $[\max(0, -\Delta), \min(1, 1 - \Delta)]$. On peut trouver ces probabilités en calculant la loi exacte de (W_T, T) sachant Δ ou en l'approximant. Dans un cas comme dans l'autre, cela ne mène pas à des calculs très simples. De plus, on voit dans le tableau 4 de Lin et al. (1991) que cet intervalle est lui aussi conservateur.

4.3 Conclusion

On a vu dans ce chapitre que les expériences à étapes multiples peuvent demander moins d'observations que celles à taille fixe, sans pour autant augmenter les erreurs de première et de deuxième espèces. On s'est intéressé particulièrement à deux tests répétés, soient les tests de Pocock et de O'Brien-Fleming. Ces deux tests répartissent la probabilité d'erreur de première espèce différemment sur les K tests. Les tests de Pocock la dépensent très tôt dans l'expérience alors que les tests de O'Brien-Fleming la dépensent surtout vers la fin de l'expérience. Il n'y a pas de règle absolue dictant quel test est le meilleur. Rien n'empêche de choisir arbitrairement les valeurs de α_k , tant que l'équation (4.16) soit respectée.

On a aussi vu que les intervalles AC et BA, qui performaient bien au chapitre 3, performant un peu moins bien dans les expériences à étapes multiples. Cela est causé par le fait que le moment où l'on arrête l'expérience dépend du paramètre Δ qui est inconnu. Tous les intervalles qui tiennent compte seulement des estimations de p_1 et p_2 ont ce problème. Utiliser un intervalle standard de niveau $(1 - \alpha)$ après avoir fait des tests répétés n'est pas une bonne idée puisque les probabilités d'erreur de première espèce seront trop élevées.

Par contre, utiliser les intervalles standards de niveau $(1 - \alpha_k)$ et arrêter lorsque 0 est exclus de l'intervalle donne des résultats respectables. Cela a pour effet de donner des probabilités de recouvrement élevées surtout lorsque $|p_1 - p_2|$ est grand, ce qui veut dire que l'intervalle sera long, mais on est quand même en mesure de détecter qu'il y a une différence entre p_1 et p_2 . On peut aussi construire un intervalle qui utilise la statistique (W_T, T) , mais il n'est pas clair que le gain en performance qu'on pourrait aller chercher pour construire cet intervalle est assez important pour compenser sa complexité.

On a surtout étudié les cas où $\pi = 1/2$, mais il y a des cas où l'on souhaitera avoir $\pi \neq 1/2$. Il y a aussi des cas où il sera impossible de savoir à l'avance combien d'observations il y aura à chaque traitement. Dans le prochain chapitre, on verra des façons d'assigner les traitements aux individus qui tiennent compte de certains critères comme le biais de sélection et le taux de succès dans l'expérience. On regardera entre autre si les tests et intervalles de confiance pour les expériences à étapes multiples vus dans ce chapitre performant bien même si on ne peut prédire π .

Chapitre 5

Plans adaptatifs dans une expérience à allocations séquentielles

Dans ce chapitre, on s'intéresse à trouver un bon plan d'allocation à utiliser lors d'une expérience où l'on veut comparer deux traitements A et B . On supposera que les candidats pour l'expérience arrivent un à la fois. Pour chaque candidat, il y a trois options : lui donner le traitement A , lui donner le traitement B ou ne pas l'inclure dans l'expérience. Une telle expérience est dite à *allocations séquentielles*.

On supposera qu'il y a une personne, que l'on appellera *expérimentateur*, qui décide si le candidat peut être inclus ou non dans l'expérience et une autre, que l'on appellera *statisticien*, qui détermine, pour un candidat inclus, lequel des deux traitements lui sera assigné. On suppose que le statisticien n'a aucune information sur le candidat lors de l'assignation du traitement et que l'expérimentateur connaît le plan d'allocation utilisé par le statisticien.

Une façon très simple d'assigner les individus aux traitements est d'y aller par alternance : par exemple, le premier individu inclus dans l'expérience reçoit le traitement A , le second inclus reçoit le traitement B , le troisième inclus le traitement A et ainsi de suite. Ce plan d'allocation à l'avantage de donner une différence entre les nombres d'observations des traitements A et B d'au maximum 1. Cela veut dire que si le nombre d'individus inclus est assez grand, on est sûr d'avoir assez d'information sur chacun des traitements.

Le plan précédent a cependant deux défauts importants : il peut mener à un biais de

sélection important, puisque l'expérimentateur saura toujours après le premier individu inclus quel sera le prochain traitement assigné, et il n'assigne pas plus d'individus au meilleur traitement.

Il y a biais de sélection lorsque l'expérimentateur agit de manière à favoriser un des traitements. Supposons que l'expérimentateur veuille favoriser le traitement A . S'il sait que le prochain individu inclus sera assigné par le statisticien au traitement A , il inclura seulement un candidat qui, selon ses connaissances, a de bonnes chances de bien répondre au traitement. S'il sait que le prochain individu inclus sera assigné au traitement B , il inclura seulement un candidat qui a de faibles chances de bien répondre au traitement. Les résultats seront alors biaisés en faveur du traitement A .

Pour l'autre défaut, supposons que les réponses possibles aux deux traitements soient survie et décès. Si, par exemple, les taux de survie (inconnus) des traitements A et B sont de 95 % et de 10 % respectivement, alors une expérience allouant le même nombre d'individus aux deux traitements devrait donner un taux de décès plus élevé qu'une expérience allouant plus d'individus au traitement A . D'un point de vue éthique, on ne peut négliger l'intérêt des individus prenant part à l'expérience. Bien sûr, on ne peut pas décider à l'avance d'assigner plus d'individus au traitement A puisque au départ, on ne connaît pas les taux de survie des deux traitements, mais on peut se servir de l'information obtenue lors des assignations précédentes.

On verra dans ce chapitre des plans d'allocation qui font un compromis entre avoir une quantité suffisante d'information pour chaque traitement (dans le but de tirer les bonnes conclusions sur l'efficacité des traitements), minimiser le biais de sélection et assigner plus d'individus au meilleur traitement. Avant cela, voici les hypothèses et la notation utilisée dans ce chapitre.

5.1 Hypothèses et notation

Cette section présente les hypothèses et la notation qui seront utilisées tout au long du chapitre.

On supposera que les réponses aux deux traitements sont dichotomiques (succès ou échecs) et qu'elles sont connues immédiatement après l'assignation du traitement, ce qui veut dire que la réponse d'un individu est connue avant que le prochain individu se présente dans l'expérience.

Notons $(A_1, B_1), (A_2, B_2), \dots$ les réponses potentielles des candidats de l'expérience où A_i est la réponse au traitement A et B_i est celle au traitement B pour le i^e candidat de l'expérience. Les deux valeurs possibles pour A_i et B_i sont 1 pour succès et 0 pour échec. On ne pourra connaître qu'au plus une valeur entre A_i et B_i puisqu'un individu ne peut pas recevoir les deux traitements. On notera D_i l'assignation du i^e candidat. Il y a trois valeurs possibles pour D_i : a si le i^e candidat est assigné au traitement A , b s'il est assigné au traitement B et 0 s'il n'est pas inclus dans l'expérience. Notons aussi $\mathcal{F}_n = \sigma(A_1, \dots, A_{n-1}, B_1, \dots, B_{n-1}, D_1, \dots, D_n)$ pour $n \geq 2$ et $\mathcal{F}_1 = \sigma(D_1)$ où $\sigma(\mathcal{C})$ est la σ -algèbre engendrée par la famille de variables aléatoires \mathcal{C} . Pour tout $n \geq 1$, on supposera que (A_n, B_n) est indépendant de \mathcal{F}_n . Cela implique que les réponses aux traitements sont indépendantes entre les individus, c'est-à-dire que pour tous $i, j \geq 1$, avec $i \neq j$, on a (A_i, B_i) indépendant de (A_j, B_j) . On n'imposera pas l'indépendance entre A_i et B_i pour tout i . On posera

$$p_1 = \mathbb{P}[A_i = 1] \quad \forall i;$$

$$p_2 = \mathbb{P}[B_i = 1] \quad \forall i.$$

Ici, p_1 et p_2 sont les taux de succès des traitements A et B respectivement. On les suppose identiques pour tous les candidats. En fait, ils ne sont pas vraiment identiques ; ils le sont du point de vue du statisticien qui n'a aucune connaissance sur les candidats. Pour l'expérimentateur, il peut, avec ses connaissances sur les candidats, être capable d'approximer les vraies probabilités de succès pour chaque candidat. Cependant, l'inférence statistique faite sur p_1 et p_2 ne tiendra pas compte de ses connaissances, ce qui veut dire que l'on suppose qu'il n'y a pas de biais de sélection causé par l'expérimentateur.

Bien que l'on suppose la bonne foi de l'expérimentateur lorsque l'on fera l'inférence statistique sur p_1 et p_2 , on supposera que l'expérimentateur cherche à deviner le plus d'assignations possibles. Cet élément sera considéré lors de la comparaison des plans d'allocation.

Posons maintenant

$$X_k = A_{m_k}, \tag{5.1}$$

où

$$m_k = \min \left\{ n \geq 1; \sum_{i=1}^n 1_a(D_i) = k \right\}. \quad (5.2)$$

avec $1_a(D_i) = 1$ si $D_i = a$, 0 sinon. Ici, X_k est la k^e valeur observée du traitement A . Posons aussi

$$Y_k = B_{n_k}, \quad (5.3)$$

où

$$n_k = \min \left\{ n \geq 1; \sum_{i=1}^n 1_b(D_i) = k \right\}. \quad (5.4)$$

Ici, Y_k est la k^e valeur observée du traitement B .

5.2 Critères d'évaluation des plans d'allocation

Dans cette section, on verra les critères d'évaluation d'un plan d'allocation séquentielle et les variables utilisées pour tenir compte de chacun d'eux.

5.2.1 Inférence statistique

Lorsqu'on fait une expérience pour comparer deux traitements, on veut recueillir suffisamment d'information sur les deux traitements. Il va de soi qu'on préférera un plan d'allocation qui a tendance à donner l'information la plus juste possible sur p_1 et p_2 .

On peut se demander si l'analyse des résultats obtenus à partir d'un plan adaptatif peut être fait de la même façon que lorsqu'on a deux échantillons indépendants issus d'un plan non-adaptatif. Par exemple, est-ce que les intervalles de confiance pour une différence de deux proportions vus au chapitre 3 performant de la même façon pour un plan adaptatif? Certains auteurs, dont Wei et al. (1990), suggèrent d'utiliser des intervalles de confiance qui tiennent compte du plan adaptatif utilisé lorsque n n'est pas très grand.

Bélisle et Melfi (2007) montrent que, sous certaines conditions qui sont respectées ici, étant donné un échantillon contenant n_1 observations du traitement A et n_2 du traitement B , alors les observations sont indépendantes. Cependant, cela ne veut pas dire que le plan utilisé n'a pas d'impact sur l'inférence que l'on fait de p_1 et p_2 . Par exemple, d'après ce qu'on a vu au chapitre 3, un intervalle de confiance produit à partir d'un plan qui a tendance à donner deux échantillons de tailles différentes pourrait contenir moins souvent la vraie valeur de $p_1 - p_2$ qu'un autre intervalle produit à partir d'un plan qui a tendance à donner des tailles d'échantillons égales. De plus, il est possible que les estimateurs du maximum de vraisemblance de p_1 et p_2 obtenus à partir d'un plan adaptatif soient biaisés.

Dans ce chapitre, on observera les biais approximatifs des estimateurs du maximum de vraisemblance de p_1 et p_2 et les probabilités de recouvrement approximatives des intervalles de Wald, Agresti-Coull (AC) et bayésien approximatif (BA). On observera aussi la puissance approximative de chaque intervalle, c'est-à-dire la probabilité approximative que 0 soit en dehors de l'intervalle lorsque $p_1 - p_2 \neq 0$. On regardera les probabilités de recouvrement et les puissances dans une expérience à taille fixe et dans une expérience à étapes multiples. Cela nous donnera une idée sur l'efficacité des plans d'allocation à bien faire ressortir l'information que l'on a sur p_1 et p_2 . Cela nous donnera aussi l'occasion de voir si les intervalles qui performaient bien au chapitre 3, soit les intervalles AC et BA, performent encore bien lorsqu'on utilise un plan d'allocation adaptatif.

5.2.2 Biais de sélection

L'analyse des résultats est faite en supposant que les données sont indépendantes et identiquement distribuées à l'intérieur d'un traitement. Cela implique qu'on suppose que l'expérimentateur choisit les candidats qui seront inclus dans l'expérience sans vouloir favoriser un traitement.

Si l'expérimentateur désire favoriser un traitement, il pourra le faire plus facilement s'il est capable de deviner les prochaines assignations. On cherchera donc un plan d'allocation qui donne un petit nombre espéré d'assignations correctement devinées.

On pourrait naïvement compter le nombre d'assignations correctement devinées, mais on préférera tenir compte de $\frac{\lambda_A}{n_1} + \frac{\lambda_B}{n_2}$ où λ_A et λ_B sont le nombre de fois que l'expérimentateur a deviné correctement que la prochaine assignation soit aux traitements A et B respectivement et n_1 et n_2 sont le nombre d'assignations aux traitements A et B . On peut justifier cela de la façon suivante :

Supposons qu'il n'y a pas de différence entre les traitements, donc que le taux de succès des deux traitements soit $p = p_1 = p_2$. Supposons que l'expérimentateur souhaite favoriser le traitement A et qu'il sait qu'il y a certains patients qui, à cause de facteurs indépendants des traitements, ont une probabilité de succès de $(p + \delta)$ et d'autres qui ont une probabilité de succès de $(p - \delta)$ avec $0 < \delta < \min(p, 1 - p)$. Dans bien des cas, la façon optimale pour l'expérimentateur d'arriver à ses fins est d'inclure un individu dans l'expérience ayant une probabilité de succès de $(p + \delta)$ si celui-ci a plus de chance d'être assigné au traitement A . Si la prochaine assignation a plus de chance d'aller au traitement B , il devrait inclure un individu ayant une probabilité de succès de $(p - \delta)$. On aura donc

$$\mathbb{E}[\hat{p}_1] = \mathbb{E} \left[\frac{\lambda_A(p + \delta) + (n_1 - \lambda_A)(p - \delta)}{n_1} \right] \quad (5.5)$$

et

$$\mathbb{E}[\hat{p}_2] = \mathbb{E} \left[\frac{\lambda_B(p - \delta) + (n_2 - \lambda_B)(p + \delta)}{n_2} \right], \quad (5.6)$$

ce qui mène à

$$\mathbb{E}[\hat{p}_1 - \hat{p}_2] = 2\delta \left(\mathbb{E} \left[\frac{\lambda_A}{n_1} + \frac{\lambda_B}{n_2} \right] - 1 \right). \quad (5.7)$$

Dans ce cas là, on souhaiterait avoir $\mathbb{E}[\hat{p}_1 - \hat{p}_2] = 0$ puisqu'il n'y a pas de différence entre les traitements. Cela arrive lorsque $\mathbb{E} \left[\frac{\lambda_A}{n_1} + \frac{\lambda_B}{n_2} \right] = 1$. Pour plusieurs plans d'allocation, le minimum pour $\mathbb{E} \left[\frac{\lambda_A}{n_1} + \frac{\lambda_B}{n_2} \right]$ est 1 si l'expérimentateur utilise la méthode optimale décrite plus haut.

5.2.3 Taux de succès dans l'expérience

Dans la recherche du meilleur traitement, on n'a pas le choix d'assigner des individus aux deux traitements. On peut voir les individus assignés au moins bon traitement comme des individus sacrifiés au profit de la recherche. Cependant, un bon plan d'allocation peut assigner plus d'individus au meilleur traitement tout en recueillant suffisamment d'information sur p_1 et p_2 . Assigner plus d'individus au meilleur traitement devrait se traduire pour un taux de succès plus élevé dans l'expérience. On tiendra compte du nombre espéré de succès pour évaluer la performance d'un plan d'allocation par rapport à ce critère.

5.3 Plans d'allocation adaptatifs

On verra ici différents plans d'allocation adaptatifs, c'est-à-dire des plans où les assignations sont faites en tenant compte de l'information recueillie depuis le début de l'expérience. On séparera ces plans en deux catégories : ceux où les allocations dépendent seulement des assignations antérieures et ceux où les allocations dépendent des assignations et des réponses antérieures. Dans cette section, on notera n le nombre d'individus inclus dans l'expérience.

5.3.1 Allocations dépendant des assignations antérieures

Ces plans tentent de trouver un compromis entre avoir un nombre suffisant d'observations pour chaque traitement et minimiser le biais de sélection. Voici quelques uns de ces plans.

Le plan binomial tronqué généralisé

Ce plan découle du plan binomial tronqué lui-même découlant du plan binomial. Le plan binomial s'explique très simplement : pour chacune des n assignations, on choisit un ou l'autre des deux traitements avec probabilité $1/2$. Ce plan est optimal par rapport au biais de sélection, mais peut mener à une expérience trop débalancée, c'est-à-dire qu'il se peut qu'un traitement compte beaucoup moins de données que l'autre.

Le plan binomial tronqué, introduit par Blackwell et Hodges (1957), force les deux traitements à avoir le même nombre d'assignations. Il faut bien sûr que le nombre total d'assignations soit pair. On procède comme pour le plan binomial jusqu'à ce qu'un des deux traitements ait $n/2$ assignations. Toutes les assignations restantes vont à l'autre traitement.

Le plan binomial tronqué généralisé contrôle le nombre d'observations pour les deux traitements, sans toutefois forcer les deux traitements à avoir le même nombre d'observations comme le fait le plan binomial tronqué. On tolère maintenant une différence d'au maximum d observations entre les traitements à la fin de l'expérience. Si n est pair, d sera pair ; si n est impair, d sera impair. La procédure est comme pour le plan binomial jusqu'à ce qu'un des deux traitements ait $n/2 + d/2$ observations. Les assignations restantes vont à l'autre traitement. On notera ce plan PBTG(n, d).

Le plan binomial et le plan binomial tronqué sont des cas particuliers du PBTG(n, d) avec $d = n$ et $d = 0$ respectivement.

Le plan aléatoire biaisé

Ce plan a été proposé par Efron (1971) sous le nom de *Biased Coin Design*. On procède comme suit :

Soient $n_1(k-1)$ et $n_2(k-1)$ le nombre d'assignations aux traitements A et B respectivement après $k-1$ assignations ($k \in \{1, \dots, n\}$). Si $n_1(k-1) = n_2(k-1)$, alors la k^e assignation sera à l'un ou l'autre des traitements avec probabilité $1/2$. Si $n_1(k-1) < n_2(k-1)$, alors la k^e assignation sera au traitement A avec probabilité p_0 ($p_0 > 1/2$) et au traitement B avec probabilité $1 - p_0$. Si $n_1(k-1) > n_2(k-1)$, alors la k^e assignation sera au traitement A avec probabilité $1 - p_0$ et au traitement B avec probabilité p_0 . On notera ce plan PAB(p_0).

Si p_0 est près de $1/2$, le biais de sélection sera minime, mais l'expérience risque d'être débalancée. En fait, si $p_0 = 1/2$, on revient au plan binomial. Si p_0 est près de 1, l'inverse se produira; on aura une expérience balancée, mais les assignations seront facilement devinées par l'expérimentateur.

Le plan aléatoire adaptatif biaisé

Ce plan est le *Adaptive Biased Coin Design* proposé par Wei (1978). Il ressemble au plan précédent, sauf qu'on ne se soucie plus seulement de savoir quel traitement a le plus d'observations jusqu'à maintenant, mais aussi de savoir quelle est la différence du nombre d'observations entre les deux traitements jusqu'à maintenant.

Posons $C_0 = 0$ et $C_k = n_1(k)/k - n_2(k)/k$ pour $k \in \{1, \dots, n-1\}$ où $n_1(k)$ et $n_2(k)$ ont la même définition qu'à la sous-section précédente. Prenons $h : [-1, 1] \rightarrow [0, 1]$ une fonction décroissante telle que pour tout $c \in [-1, 1]$, $h(c) = 1 - h(-c)$. La probabilité que la k^e assignation soit au traitement A est $h(C_{k-1})$. On notera ce plan PAAB(h).

Le PAB(p_0) est un cas particulier du PAAB(h) avec

$$h(x) = \begin{cases} p_0 & \text{si } -1 \leq x < 0; \\ 1/2 & \text{si } x = 0; \\ 1 - p_0 & \text{si } 0 < x \leq 1. \end{cases}$$

5.3.2 Allocations dépendant des assignations et des réponses antérieures

Ces plans peuvent faire un compromis entre les trois facteurs mentionnés au début du chapitre, soit avoir une quantité suffisante d'information pour chaque traitement, minimiser le biais de sélection et assigner plus d'individus au meilleur traitement. Cependant, ces plans ont le défaut de biaiser les estimateurs du maximum de vraisemblance de p_1 et p_2 . Voici quatre de ces plans.

Le plan adaptatif randomisé

Ce plan est le *Randomized Adaptive Design* de Xia (2002). Ici, on utilise les réponses obtenues antérieurement pour obtenir un échantillon qui minimisera la variance, sans se soucier du taux de succès dans l'expérience. Posons \hat{p}_{1k} et \hat{p}_{2k} comme les estimateurs du maximum de vraisemblance de p_1 et p_2 respectivement après k assignations. On sait que, conditionnellement à $n_1(k)$, on a $n_1(k)\hat{p}_{1k} \sim \text{Binomiale}(n_1(k), p_1)$, ce qui veut dire que $\text{Var}[n_1(k)\hat{p}_{1k}] = n_1(k)p_1(1 - p_1)$ et donc que $\text{Var}[\hat{p}_{1k}] = p_1(1 - p_1)/n_1(k)$. De même, conditionnellement à $n_2(k)$, $\text{Var}[\hat{p}_{2k}] = p_2(1 - p_2)/n_2(k)$ et sous l'hypothèse d'indépendance des observations, on a

$$\text{Var}[\hat{p}_{1k} - \hat{p}_{2k}] = \frac{p_1(1 - p_1)}{n_1(k)} + \frac{p_2(1 - p_2)}{n_2(k)}. \quad (5.8)$$

Soit π_k la proportion des k premières observations qui sont assignées au traitement A. L'équation précédente devient

$$\text{Var}[\hat{p}_{1k} - \hat{p}_{2k}] = \frac{p_1(1 - p_1)}{k\pi_k} + \frac{p_2(1 - p_2)}{k(1 - \pi_k)}. \quad (5.9)$$

Chercher un échantillon de n observations avec $\text{Var}[\hat{p}_1 - \hat{p}_2]$ minimale revient à chercher le $\pi \equiv \pi_n$ minimisant $\text{Var}[\hat{p}_1 - \hat{p}_2]$. Pour le trouver, on dérive $\text{Var}[\hat{p}_1 - \hat{p}_2]$ par rapport à π :

$$\begin{aligned}
\frac{dVar[\hat{p}_1 - \hat{p}_2]}{d\pi} &= \frac{-p_1(1-p_1)}{n\pi^2} + \frac{p_2(1-p_2)}{n(1-\pi)^2} = 0 \\
&\leftrightarrow p_1(1-p_1)(1-\pi)^2 = p_2(1-p_2)\pi^2 \\
&\leftrightarrow \sqrt{p_1(1-p_1)}(1-\pi) = \sqrt{p_2(1-p_2)}\pi \\
&\leftrightarrow \pi = \frac{\sqrt{p_1(1-p_1)}}{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}}.
\end{aligned}$$

Il est facile de voir que pour tout $\pi \in [0, 1]$, $\frac{d^2Var[\hat{p}_1 - \hat{p}_2]}{d\pi^2} > 0$ ce qui veut dire que $\frac{\sqrt{p_1(1-p_1)}}{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}}$ est un minimum.

Comme on ne connaît pas p_1 et p_2 , on peut utiliser ce résultat seulement à partir des estimations de p_1 et p_2 obtenues à partir des assignations précédentes. On pourrait par exemple assigner le $(k+1)^e$ candidat retenu au traitement A avec probabilité

$$\frac{\sqrt{\hat{p}_{1,k}(1-\hat{p}_{1,k})}}{\sqrt{\hat{p}_{1,k}(1-\hat{p}_{1,k})} + \sqrt{\hat{p}_{2,k}(1-\hat{p}_{2,k})}}. \quad (5.10)$$

Cependant, cette méthode comporte un défaut important : le premier traitement pour lequel on observera au moins un succès et un échec se verra assigner tous les autres candidats retenus. Pour contrer ce problème, Melfi et al. (2001) proposent d'ajouter une pseudo-observation à chaque traitement, 1/2 succès et 1/2 échec, pour estimer p_1 et p_2 . On assignera alors le $(k+1)^e$ candidat retenu au traitement A avec probabilité

$$\frac{\sqrt{\tilde{p}_{1,k}(1-\tilde{p}_{1,k})}}{\sqrt{\tilde{p}_{1,k}(1-\tilde{p}_{1,k})} + \sqrt{\tilde{p}_{2,k}(1-\tilde{p}_{2,k})}}, \quad (5.11)$$

où $\tilde{p}_{1,k} = \frac{n_1(k)\hat{p}_{1,k}+1/2}{n_1(k)+1}$ et $\tilde{p}_{2,k} = \frac{n_2(k)\hat{p}_{2,k}+1/2}{n_2(k)+1}$. On voit que la première assignation a la même probabilité d'aller au traitement A qu'au traitement B et que si il y a un traitement qui n'a pas encore reçu d'assignation, la prochaine assignation sera faite à ce traitement avec probabilité plus grande ou égale à 1/2. Ce plan sera noté PAR.

Le plan aléatoire biaisé doublement adaptatif

Ce plan est le *Doubly Adaptive Biased Coin Design* proposé par Eisele (1994). On dit *doublement* adaptatif parce qu'en plus de tenir compte du nombre d'observations d'écart entre les deux traitements pour la prochaine assignation, on tient aussi compte de la performance observée jusqu'à maintenant.

Prenons $\rho \equiv \rho(p_1, p_2) : [0, 1]^2 \rightarrow [0, 1]$ une fonction de p_1 et p_2 retournant la proportion désirée d'observations assignées au traitement A . On pose $\hat{\rho}_k \equiv \rho(\hat{p}_{1k}, \hat{p}_{2k})$. Prenons une fonction $q : [0, 1]^2 \rightarrow [0, 1]$ ayant les quatre propriétés suivantes :

1. q est continue ;
2. $q(a, a) = a \quad \forall a \in [0, 1]$;
3. $q(a, b)$ est strictement décroissante en a et strictement croissante en b sur $[0, 1]^2$;
4. Les dérivées de q sont bornées pour les deux arguments.

Prenons un $n_0 < n/2$. Eisele (1994) propose d'assigner les n_0 premiers candidats inclus dans l'expérience au traitement A et les n_0 suivants au traitement B . Ensuite, pour $k \in \{2n_0 + 1, \dots, n\}$, la probabilité que la k^e assignation sera au traitement A est $q(n_1(k-1)/(k-1), \hat{\rho}_{k-1})$. Dans ce mémoire, les conditions initiales d'utilisation de ce plan dépendront de (q, ρ) . On notera ce plan PABDA(q, ρ).

Le PAAB(h) est un cas particulier du PABDA(h, ρ) avec $\rho = 1/2$.

Le *play-the-winner* randomisé

Ce plan introduit par Wei et Durham (1978) provient du *play-the-winner rule* (PW) proposé par Zelen (1969). Le plan PW est simple : le premier individu inclus dans l'expérience reçoit l'un ou l'autre des deux traitements avec probabilité $1/2$; si le résultat est un succès, le second inclus recevra le même traitement que le premier, sinon (si le résultat est un échec) le second recevra l'autre traitement et ainsi de suite. Ce plan assigne plus d'individus au traitement qui performe le mieux. Cependant, il peut mener à un important biais de sélection, l'expérimentateur connaissant toujours, à partir de la deuxième assignation, le prochain traitement à être assigné.

Le *play-the-winner* randomisé s'attaque à ce problème. Voici comment on procède :

On place μ boules marquées d'un A et μ autres boules marquées d'un B dans une urne. On tire une boule et le premier individu inclus dans l'expérience reçoit le traitement inscrit sur la boule. On remet la boule dans l'urne. Si le résultat est un succès, on rajoute dans l'urne β boules marquées du traitement reçu par le premier individu et α boules (où $\alpha < \beta$) marquées de l'autre traitement. Si le résultat est un échec, on rajoute α boules marquées du traitement reçu par le premier individu et β boules marquées de l'autre traitement. On continue ainsi de suite. On notera ce plan $\text{RPW}(\mu, \alpha, \beta)$.

Pour α et β fixes et $\mu \rightarrow \infty$, le plan $\text{RPW}(\mu, \alpha, \beta)$ revient au plan binomial. On notera $\text{RPW}(\mu, \beta)$ pour désigner le plan $\text{RPW}(\mu, 0, \beta)$.

On s'est beaucoup intéressé à ce plan dans la littérature et plusieurs résultats asymptotiques ont été démontrés pour ce plan. Wei et Durham (1978) ont montré que la probabilité qu'un individu soit assigné au traitement A tend vers

$$\frac{\alpha p_2 + \beta q_2}{\alpha(p_1 + p_2) + \beta(q_1 + q_2)}, \quad (5.12)$$

où $q_i = 1 - p_i$ ($i = 1, 2$). Si $\alpha = 0$, la cela revient à

$$\frac{q_2}{q_1 + q_2}. \quad (5.13)$$

De plus, toujours lorsque $\alpha = 0$, on peut montrer que la proportion d'échecs dans l'expérience tend presque sûrement vers $\frac{2q_1q_2}{q_1+q_2}$ (Coad et Rosenberger, 1999). Pour un plan ayant pour but d'assigner autant d'observations aux deux traitements, la proportion d'échecs tend vers $\frac{q_1+q_2}{2}$. Il est facile de voir que

$$\frac{2q_1q_2}{q_1 + q_2} \leq \frac{q_1 + q_2}{2} \Leftrightarrow 0 \leq (q_1 - q_2)^2. \quad (5.14)$$

L'expression à droite du *si et seulement si* étant toujours vraie, cela implique que pour $n \rightarrow \infty$, où n est le nombre d'observations au total, la proportion d'échecs pour le plan $\text{RPW}(\mu, \beta)$ est presque sûrement inférieure ou égale à la proportion d'échecs d'un plan ayant pour but d'assigner autant d'observations aux deux traitements, avec égalité seulement si $p_1 = p_2$ ($q_1 = q_2$). Cela implique aussi que plus la différence entre p_1 et p_2 est importante, plus le plan $\text{RPW}(\mu, \beta)$ diminue la proportion d'échecs par rapport à un plan balancé.

Un problème avec le plan $\text{RPW}(\mu, \beta)$ c'est qu'il peut mener à des expériences très déséquilibrées, surtout lorsque p_1 ou p_2 est grand. En effet, si on commence l'expérience avec un traitement ayant une forte probabilité de succès, alors il y a de fortes chances d'ajouter une boule de ce traitement dans l'urne, donc plus de chances d'être choisi à nouveau et ainsi de suite. C'est ce qui s'est passé lors d'une expérience réalisée à l'Université du Michigan comparant deux traitements contre les difficultés respiratoires de nouveaux-nés (voir Wei et al., 1990). On a utilisé le plan $\text{RPW}(1,1)$ pour les assignations et sur les dix nouveaux-nés inclus dans l'expérience, neuf ont reçu le nouveau traitement contre seulement un pour l'ancien. Il est difficile d'estimer $p_1 - p_2$ dans un cas comme celui-là.

Le *drop-the-loser*

Ce plan ressemble au plan $\text{RPW}(\mu, \beta)$, mais au lieu de récompenser la bonne performance des traitements, on pénalise la mauvaise performance des traitements. Voici comment on procède (Ivanova, 2003) :

Dans une urne, on commence avec a boules marquées d'un 0, C boules marquées d'un A et C boules marquées d'un B . On tire une boule. Si c'est une boule marquée d'un 0, on la remet dans l'urne et on ajoute M boules marquées d'un A et M marquées d'un B . Si c'est une boule marquée d'un A (respectivement B), l'individu recevra le traitement A (respectivement B) et si le résultat est un succès, la boule sera remise dans l'urne, ce qui veut dire que la composition de l'urne demeure inchangée ; si le résultat est un échec, la boule ne sera pas remise dans l'urne. On continue ainsi de suite.

On notera ce plan $\text{DL}(a, C, M)$. Pour ce plan, le traitement ayant obtenu moins d'échecs a le plus de chances d'être choisi à la prochaine assignation et ce peu importe le nombre de succès obtenu par les deux traitements. Cela a pour effet de favoriser le meilleur traitement sans toutefois donner une expérience trop déséquilibrée comme cela peut arriver pour le plan RPW .

5.4 Comparaison des plans

Dans cette section, on comparera les plans d'allocation dépendant des assignations et des réponses antérieures. Ces plans sont utiles surtout lorsque la différence entre p_1 et p_2 est grande. Si on pense que la différence entre p_1 et p_2 est petite, les plans où

les allocations dépendent seulement des assignations antérieures font l'affaire. Pour une analyse plus approfondie sur ces plans, voir Auclair (1999).

On comparera les plans étudiés par rapport au biais de \hat{p}_1 et \hat{p}_2 , au biais de sélection et au taux de succès dans l'expérience. On regardera aussi la facilité de détecter la (vraie) différence des deux proportions au moyen des intervalles de Wald, AC et BA dans une expérience à taille fixe et à étapes multiples pour les plans étudiés.

Les plans comparés seront le RPW(1,1), le RPW(5,1), le DL(1,1,1), le DL(1,5,1), le PAR, le PABDA($h, \frac{1}{2}$), qui est en fait le PAAB(h), et le PABDA($h, \frac{q_2}{q_1+q_2}$) ($q_i = 1 - p_i$) avec pour tout $0 < a < 1$

$$h(x, a) = \begin{cases} 1 - \left(\frac{1-a}{a}\right) x & \text{si } 0 \leq x \leq a \\ \left(\frac{a}{1-a}\right) - \left(\frac{a}{1-a}\right) x & \text{si } a \leq x \leq 1. \end{cases}$$

Pour un $0 < a < 1$, la fonction h est une droite passant du point (0,1) au point (a, a) et du point (a, a) au point (1,1).

Pour les plans PABDA($h, \frac{1}{2}$) et PABDA($h, \frac{q_2}{q_1+q_2}$), la première assignation sera au traitement A avec probabilité $1/2$ et la k^e ($k \geq 2$) sera au traitement A avec probabilité $q(n_1(k-1)/(k-1), \hat{p}_{k-1})$. Pour le plan PABDA($h, \frac{q_1}{q_1+q_2}$), au lieu de poser $\hat{p}_k = \frac{1-\hat{p}_{2,k}}{(1-\hat{p}_{1,k})+(1-\hat{p}_{2,k})}$, on posera $\hat{p}_k = \frac{(1-\tilde{p}_{2,k})}{(1-\tilde{p}_{1,k})+(1-\tilde{p}_{2,k})}$ où $\tilde{p}_{1,k} = \frac{n_1(k)\hat{p}_{1,k}+1/2}{n_1(k)+1}$ et $\tilde{p}_{2,k} = \frac{n_2(k)\hat{p}_{2,k}+1/2}{n_2(k)+1}$, comme proposé dans Melfi et al. (2001).

Il arrivera qu'on comparera les résultats obtenus pour les plans utilisés avec ceux qu'on aurait obtenus avec le plan binomial. Ce plan est plus simple que les plans adaptatifs étudiés. Il est intéressant de voir sa performance pour pouvoir juger si cela vaut la peine de chercher des plans d'allocation plus compliqués. Les plans PAR et PABDA($h, \frac{1}{2}$) ne s'attaquent pas au taux de succès, mais cherchent plutôt à favoriser l'inférence statistique. Le plan PAR le fait en tentant de minimiser la variance de $\hat{p}_1 - \hat{p}_2$ et le plan PABDA($h, \frac{1}{2}$) le fait en essayant de donner une expérience pas trop déséquilibrée. Les plans RPW(1,1), DL(1,1,1) et PABDA($h, \frac{q_1}{q_1+q_2}$) devraient donner de bons taux de succès, peut-être au détriment de la quantité d'information et du biais de sélection. Les plans RPW(5,1) et DL(1,5,1) font un compromis entre taux de succès, biais de sélection et quantité d'information.

Pour chaque couple (p_1, p_2) et pour chacun des plans étudiés, on fait 10 000 simulations pour obtenir les statistiques présentées dans les tableaux qui suivront.

5.4.1 Biais de \hat{p}_1 et \hat{p}_2

Le tableau 5.1 présente les biais de \hat{p}_1 et \hat{p}_2 pour les plans étudiés pour différentes valeurs de (p_1, p_2) lorsque $n = 30$. Pour un couple (p_1, p_2) et un plan donnés, le biais présenté de \hat{p}_i ($i = 1, 2$) est en fait la moyenne des 10 000 valeurs de $(\hat{p}_i - p_i)$ obtenues à partir des simulations.

En théorie, les biais sont nuls pour les plans binomial (pas présenté dans le tableau) et PABDA($h, 1/2$) puisqu'ils ne tiennent pas compte de la performance des traitements dans l'expérience et ils sont très près de 0 pour le plan PAR. Pour les autres plans, les biais sont tous négatifs. Les plans RPW(5,1) et DL(1,5,1) ont des biais comparables qui varient de -0,017 à -0,001. Le plan DL(1,1,1) donne des biais qui sont en général un peu plus élevés que ceux des plans RPW(5,1) et DL(1,5,1). Ce sont les plans PABDA($h, \frac{q_2}{q_1+q_2}$) et RPW(1,1) qui donnent les biais les plus importants. Dans le pire cas, soit l'estimation de p_2 lorsque $(p_1, p_2) = (0,9, 0,4)$, le biais est de -0,038 pour le plan PABDA($h, \frac{q_2}{q_1+q_2}$) et de -0,046 pour le plan RPW(1,1).

5.4.2 Biais de sélection

Le tableau 5.2 montre la moyenne des 10 000 valeurs de $\frac{\lambda_A}{n_1} + \frac{\lambda_B}{n_2}$ obtenues à partir des simulations pour chaque paire (p_1, p_2) lorsque $n = 30$. On suppose ici que l'expérimentateur prédira que la prochaine assignation sera au traitement A (respectivement B) si la probabilité que la prochaine assignation soit au traitement A (respectivement B) est plus grande que $1/2$. Si cette probabilité est $1/2$, l'expérimentateur fera son choix au hasard.

Le plan binomial donne en théorie $\mathbb{E}[\frac{\lambda_A}{n_1} + \frac{\lambda_B}{n_2}] = 1$ et n'est pas présenté dans ce tableau. Le plan PAR donne des $\frac{\lambda_A}{n_1} + \frac{\lambda_B}{n_2}$ moyens peu élevés pour l'ensemble des valeurs de (p_1, p_2) utilisées. C'est beaucoup mieux que l'autre plan désirant favoriser l'inférence statistique, le plan PABDA($h, 1/2$), qui donne des $\frac{\lambda_A}{n_1} + \frac{\lambda_B}{n_2}$ moyens de 1,15 ou 1,16.

(p_1, p_2)	PAR	PABDA $(h, 1/2)$	PABDA $\left(h, \frac{q_2}{q_1+q_2}\right)$	RPW $(1,1)$	RPW $(5,1)$	DL $(1,1,1)$	DL $(1,5,1)$
(0,1 , 0,1)	-0,004	0,001	-0,003	-0,004	-0,003	-0,004	-0,003
	-0,007	0,000	-0,001	-0,004	-0,002	-0,004	-0,002
(0,25 , 0,25)	-0,009	0,002	-0,007	-0,009	-0,005	-0,007	-0,008
	-0,008	-0,003	-0,006	-0,008	-0,005	-0,008	-0,006
(0,5 , 0,5)	0,000	0,001	-0,011	-0,013	-0,010	-0,013	-0,012
	0,002	0,000	-0,014	-0,014	-0,009	-0,013	-0,012
(0,75 , 0,75)	0,008	0,000	-0,021	-0,017	-0,009	-0,012	-0,010
	0,009	0,001	-0,021	-0,019	-0,007	-0,011	-0,010
(0,9 , 0,9)	0,006	-0,001	-0,025	-0,015	-0,006	-0,008	-0,003
	0,006	0,000	-0,026	-0,016	-0,005	-0,007	-0,005
(0,4 , 0,1)	-0,002	0,000	-0,008	-0,008	-0,003	-0,006	-0,008
	-0,008	0,000	-0,005	-0,005	-0,003	-0,007	-0,004
(0,65 , 0,35)	0,005	0,000	-0,011	-0,012	-0,006	-0,008	-0,004
	-0,005	-0,001	-0,017	-0,020	-0,008	-0,013	-0,012
(0,8 , 0,5)	0,007	-0,001	-0,010	-0,008	-0,007	-0,006	-0,006
	-0,002	0,002	-0,026	-0,028	-0,015	-0,020	-0,017
(0,6 , 0,1)	0,001	0,000	-0,008	-0,008	-0,007	-0,004	-0,006
	-0,007	0,000	-0,003	-0,008	-0,004	-0,006	-0,006
(0,75 , 0,25)	0,006	0,000	-0,008	-0,006	-0,004	-0,005	-0,004
	-0,007	-0,001	-0,019	-0,022	-0,009	-0,016	-0,014
(0,9 , 0,4)	0,009	-0,001	-0,005	-0,003	-0,004	-0,003	-0,002
	-0,002	0,001	-0,038	-0,046	-0,015	-0,025	-0,017
(0,9 , 0,1)	0,006	0,000	-0,004	-0,002	-0,001	-0,002	-0,001
	-0,004	0,001	-0,011	-0,015	-0,007	-0,011	-0,008

TAB. 5.1 – Moyenne des 10 000 biais de \hat{p}_1 et \hat{p}_2 pour les plans étudiés pour différentes valeurs de (p_1, p_2) lorsque $n = 30$. Pour chaque couple (p_1, p_2) , le biais de \hat{p}_1 est écrit en haut et le biais de \hat{p}_2 est écrit en bas.

(p_1, p_2)	PAR	PABDA $(h, 1/2)$	PABDA $(h, \frac{q_2}{q_1+q_2})$	RPW $(1,1)$	RPW $(5,1)$	DL $(1,1,1)$	DL $(1,5,1)$
(0,1 , 0,1)	1,07	1,15	1,19	1,15	1,11	1,31	1,27
(0,25 , 0,25)	1,04	1,15	1,20	1,12	1,09	1,29	1,24
(0,5 , 0,5)	1,02	1,15	1,22	1,07	1,04	1,22	1,17
(0,75 , 0,75)	1,05	1,15	1,24	1,02	0,99	1,13	1,10
(0,9 , 0,9)	1,08	1,15	1,21	1,00	0,96	1,07	1,05
(0,4 , 0,1)	1,05	1,16	1,18	1,08	1,06	1,27	1,21
(0,65 , 0,35)	1,03	1,15	1,20	1,06	1,04	1,19	1,14
(0,8 , 0,5)	1,04	1,15	1,21	1,05	1,02	1,13	1,09
(0,6 , 0,1)	1,05	1,16	1,17	1,06	1,03	1,20	1,15
(0,75 , 0,25)	1,04	1,15	1,18	1,06	1,03	1,14	1,10
(0,9 , 0,4)	1,05	1,15	1,22	1,07	1,03	1,09	1,06
(0,9 , 0,1)	1,07	1,15	1,22	1,07	1,03	1,09	1,06

TAB. 5.2 – Valeurs moyennes de $\frac{\lambda_A}{n_1} + \frac{\lambda_B}{n_2}$ pour les plans étudiés pour différentes valeurs de (p_1, p_2) lorsque $n = 30$.

Les plans RPW(1,1) et RPW(5,1) donnent aussi des $\frac{\lambda_A}{n_1} + \frac{\lambda_B}{n_2}$ moyens pas trop élevés. Ceux du plan RPW(1,1) sont toujours un peu plus élevés que ceux du plan RPW(5,1) et ils sont plus élevés lorsque p_1 et p_2 sont petits. On remarque que pour le plan RPW(5,1), la valeur moyenne de $\frac{\lambda_A}{n_1} + \frac{\lambda_B}{n_2}$ est inférieure à 1 lorsque (p_1, p_2) est égale à (0,75 , 0,75) et (0,9 , 0,9). Cela est dû au fait que lorsque p_1 et p_2 sont grands, il y a de fortes chances que le premier traitement assigné soit toujours celui qui aura le plus de chances d'être assigné à toutes les assignations suivantes. Donc, même si l'expérimentateur peut deviner la majorité des assignations, il ne pourra avantager un des deux traitements puisque toutes ses prédictions iront au même traitement. On aura donc λ_A ou λ_B égale à 0 et cela veut dire que $\frac{\lambda_A}{n_1} + \frac{\lambda_B}{n_2}$ est plus petit ou égal à 1.

Les plans DL(1,1,1) et DL(1,5,1) donnent des $\frac{\lambda_A}{n_1} + \frac{\lambda_B}{n_2}$ moyens plus élevés que les plans RPW(1,1) et RPW(5,1). Ceux du plan DL(1,1,1) sont toujours un peu plus élevés que ceux du plan DL(1,5,1) et ils sont plus élevés lorsque p_1 et p_2 sont petits. Le PABDA $(h, \frac{q_2}{q_1+q_2})$ donne aussi des $\frac{\lambda_A}{n_1} + \frac{\lambda_B}{n_2}$ moyens élevés, surtout lorsque p_1 et p_2 sont grands.

(p_1, p_2)	PAR	PABDA $(h, 1/2)$	PABDA $\left(h, \frac{q_2}{q_1+q_2}\right)$	RPW $(1,1)$	RPW $(5,1)$	DL $(1,1,1)$	DL $(1,5,1)$
$(0,4, 0,1)$	0,272	0,250	0,273	0,276	0,272	0,275	0,274
$(0,65, 0,35)$	0,501	0,500	0,534	0,534	0,525	0,531	0,531
$(0,8, 0,5)$	0,637	0,651	0,698	0,695	0,675	0,684	0,678
$(0,6, 0,1)$	0,386	0,350	0,427	0,433	0,408	0,427	0,421
$(0,75, 0,25)$	0,500	0,500	0,597	0,600	0,567	0,584	0,576
$(0,9, 0,4)$	0,615	0,650	0,787	0,773	0,723	0,741	0,726
$(0,9, 0,1)$	0,499	0,502	0,757	0,757	0,668	0,701	0,677

TAB. 5.3 – Valeurs moyennes des taux de succès pour les plans étudiés pour différentes valeurs de (p_1, p_2) lorsque $n = 30$.

5.4.3 Taux de succès

Le tableau 5.3 montre la moyenne des 10 000 taux de succès obtenus à partir des simulations pour chaque paire (p_1, p_2) lorsque $n = 30$.

Le plan binomial (pas inclus dans le tableau) et le plan PABDA($h, 1/2$) donnent tous les deux des taux de succès de $\frac{p_1+p_2}{2}$. Pour $(p_1, p_2) = (0,4, 0,1)$, tous les plans donnent des taux de succès presque identiques sauf le plan PABDA($h, 1/2$). Cela devrait être toujours vrai lorsque la différence entre les deux traitements n'est pas trop grande et lorsque $p_1 + p_2 < 1$. Lorsque $p_1 + p_2 > 1$, le plan PAR donne moins de succès que tous les autres plans, même le plan binomial. Cela n'est pas acceptable d'un point de vue éthique. Ce sont les plans PABDA $\left(h, \frac{q_2}{q_1+q_2}\right)$ et RPW(1,1) qui présentent les meilleurs taux de succès. Pour un Δ donné, ces deux plans performant mieux si $p_1 + p_2$ est grand. Dans ces cas là, ces plans ont tendance à assigner beaucoup plus d'observations au meilleur traitement. Les plans RPW(5,1) et DL(1,5,1) ont des taux de succès comparables quoique ceux du plan DL(1,5,1) sont légèrement supérieurs. Le plan DL(1,1,1) donne des taux de succès entre ceux des plans RPW(1,1) et DL(1,5,1).

5.4.4 Estimation par intervalle de confiance de $p_1 - p_2$

Il peut être difficile d'estimer $p_1 - p_2$ par intervalle de confiance lorsqu'un plan adaptatif est utilisé. Premièrement, parce que certains plans peuvent mener à des expériences

(p_1, p_2)	PAR	PABDA $(h, 1/2)$	PABDA $\left(h, \frac{q_2}{q_1+q_2}\right)$	RPW $(1,1)$	RPW $(5,1)$	DL $(1,1,1)$	DL $(1,5,1)$
(0,1 , 0,1)	0,900	0,923	0,917	0,915	0,921	0,913	0,917
(0,25 , 0,25)	0,887	0,935	0,915	0,913	0,918	0,910	0,917
(0,5 , 0,5)	0,917	0,932	0,929	0,917	0,928	0,921	0,924
(0,75 , 0,75)	0,887	0,930	0,952	0,916	0,928	0,940	0,933
(0,9 , 0,9)	0,893	0,923	0,934	0,876	0,918	0,920	0,923
(0,4 , 0,1)	0,908	0,922	0,918	0,912	0,923	0,911	0,907
(0,65 , 0,35)	0,896	0,915	0,913	0,896	0,915	0,916	0,908
(0,8 , 0,5)	0,895	0,924	0,913	0,898	0,921	0,922	0,921
(0,6 , 0,1)	0,910	0,922	0,895	0,894	0,907	0,905	0,912
(0,75 , 0,25)	0,889	0,920	0,868	0,839	0,895	0,885	0,891
(0,9 , 0,4)	0,908	0,921	0,846	0,816	0,898	0,903	0,899
(0,9 , 0,1)	0,833	0,858	0,671	0,674	0,800	0,775	0,819

TAB. 5.4 – Probabilités de recouvrement l’intervalle de Wald de niveau 95 % pour les plans étudiés pour différentes valeurs de (p_1, p_2) lorsque $n = 30$.

très débalancées et alors l’information recueillie sur le moins bon traitement est restreinte. Deuxièmement, on a vu que les plans adaptatifs peuvent biaiser les estimateurs du maximum de vraisemblance de p_1 et p_2 . Ces estimateurs sont utilisés dans plusieurs intervalles de confiance pour une différence de deux proportions, dont ceux vus au chapitre 3.

Wei et al. (1990) ont étudié divers intervalles tenant compte du plan utilisé. Avant d’envisager l’utilisation de ces intervalles, qui n’ont pas de formes explicites, on regardera comment performant les intervalles vus au chapitre 3. Les tableaux 5.4 à 5.6 présentent les probabilités de recouvrement approximatives des intervalles de Wald, AC et BA de niveau 95 % pour les plans étudiés pour diverses valeurs de (p_1, p_2) lorsque $n = 30$. Ces probabilités de recouvrement sont en fait la proportion des 10 000 simulations où la vraie valeur de $p_1 - p_2$ est incluse dans l’intervalle.

Pour l’intervalle de Wald, les probabilités de recouvrement sont presque toujours inférieures au niveau de confiance nominal. Cela n’est pas nécessairement dû au fait qu’on utilise des plans adaptatifs puisque le caractère anti-conservateur de l’intervalle de Wald a déjà été montré au chapitre 3. Pour les intervalles AC et BA, les probabilités de recouvrement sont près du niveau de confiance nominal, peu importe le plan utilisé. Les

(p_1, p_2)	PAR	PABDA $(h, 1/2)$	PABDA $\left(h, \frac{q_2}{q_1+q_2}\right)$	RPW $(1,1)$	RPW $(5,1)$	DL $(1,1,1)$	DL $(1,5,1)$
(0,1 , 0,1)	0,994	0,989	0,991	0,993	0,993	0,991	0,992
(0,25 , 0,25)	0,958	0,962	0,962	0,964	0,962	0,958	0,960
(0,5 , 0,5)	0,944	0,955	0,952	0,946	0,952	0,947	0,950
(0,75 , 0,75)	0,958	0,957	0,959	0,957	0,956	0,955	0,949
(0,9 , 0,9)	0,994	0,988	0,973	0,972	0,984	0,982	0,986
(0,4 , 0,1)	0,958	0,956	0,966	0,963	0,965	0,966	0,961
(0,65 , 0,35)	0,940	0,950	0,950	0,952	0,950	0,954	0,948
(0,8 , 0,5)	0,947	0,953	0,944	0,950	0,948	0,948	0,947
(0,6 , 0,1)	0,955	0,951	0,959	0,965	0,959	0,963	0,962
(0,75 , 0,25)	0,951	0,955	0,968	0,967	0,958	0,960	0,964
(0,9 , 0,4)	0,955	0,951	0,960	0,962	0,949	0,952	0,947
(0,9 , 0,1)	0,971	0,970	0,969	0,971	0,972	0,972	0,967

TAB. 5.5 – Probabilités de recouvrement l’intervalle AC de niveau 95 % pour les plans étudiés pour différentes valeurs de (p_1, p_2) lorsque $n = 30$.

(p_1, p_2)	PAR	PABDA $(h, 1/2)$	PABDA $\left(h, \frac{q_2}{q_1+q_2}\right)$	RPW $(1,1)$	RPW $(5,1)$	DL $(1,1,1)$	DL $(1,5,1)$
(0,1 , 0,1)	0,993	0,989	0,991	0,993	0,993	0,991	0,992
(0,25 , 0,25)	0,953	0,958	0,958	0,961	0,958	0,957	0,960
(0,5 , 0,5)	0,936	0,946	0,944	0,938	0,943	0,940	0,943
(0,75 , 0,75)	0,951	0,954	0,945	0,943	0,944	0,943	0,941
(0,9 , 0,9)	0,994	0,988	0,960	0,961	0,980	0,975	0,982
(0,4 , 0,1)	0,953	0,947	0,958	0,957	0,959	0,958	0,953
(0,65 , 0,35)	0,937	0,948	0,945	0,945	0,945	0,949	0,944
(0,8 , 0,5)	0,941	0,946	0,937	0,939	0,939	0,939	0,937
(0,6 , 0,1)	0,949	0,945	0,953	0,959	0,953	0,960	0,959
(0,75 , 0,25)	0,942	0,944	0,959	0,959	0,946	0,947	0,951
(0,9 , 0,4)	0,950	0,943	0,952	0,953	0,937	0,937	0,932
(0,9 , 0,1)	0,964	0,961	0,956	0,961	0,963	0,963	0,960

TAB. 5.6 – Probabilités de recouvrement l’intervalle BA de niveau 95 % pour les plans étudiés pour différentes valeurs de (p_1, p_2) lorsque $n = 30$.

plus petites probabilités de recouvrement observées sont de 0,937 pour l'intervalle BA et de 0,940 pour l'intervalle AC. On remarque aussi que pour les intervalles AC et BA, les différences des probabilités de recouvrement sont plus prononcées entre les valeurs de (p_1, p_2) que entre les plans. En fait, les moyennes des probabilités de recouvrement de chaque plan sont presque identiques, ce qui n'est pas le cas pour l'intervalle de Wald.

Les intervalles AC et BA ne sont pas dérangés ici par les deux problèmes potentiels des plans adaptatifs mentionnés plus haut. On a vu au chapitre 3 que les intervalles AC et BA performaient bien pour les plans débalancés. Pour le biais des estimateurs de p_1 et p_2 , il faut dire que dans la plupart des cas, les biais de \hat{p}_1 et \hat{p}_2 sont très faibles. Dans les cas où le biais est un peu plus élevé, le biais n'est pas assez important pour affecter significativement les probabilités de recouvrement.

5.4.5 Puissance

Au chapitre 4, l'équation (4.10) nous donne le nombre d'observations nécessaires pour atteindre une certaine puissance donnée lorsqu'on connaît l'écart-type sous H_0 , l'écart-type sous H_1 et π , la proportion d'observations allant au traitement A . Toujours au chapitre 4, on a vu que la puissance désirée était généralement atteinte lorsqu'on fixe $\pi = 1/2$ et qu'on remplace l'écart-type sous H_0 par $1/2$ et celle sous H_1 par $(1 + \Delta)(1 - \Delta)$. On n'a pas regardé ce qui se passe lorsque π n'est pas connu à l'avance, comme c'est le cas lorsqu'on utilise un plan adaptatif.

Les tableaux 5.7 et 5.8 montrent les puissances approximatives des intervalles AC et BA pour différentes valeurs de (p_1, p_2) lorsque la puissance désirée est de 0,75. Ces puissances sont en fait la proportion des 10 000 simulations où la borne inférieure de l'intervalle est supérieure à 0. Comme au chapitre 4, on suppose que $\pi = 1/2$ pour calculer le nombre d'observations nécessaires. Les puissances de l'intervalle de Wald ne sont pas présentées puisque ses probabilités de recouvrement n'atteignent que très rarement le niveau de confiance désiré.

On remarque d'abord que la puissance désirée est presque toujours atteinte. Il est intéressant de comparer les puissances obtenues avec celles qu'on aurait obtenues si on avait utilisé un plan avec $\pi = 1/2$ fixé au début de l'expérience. Ces puissances sont dans la partie gauche du tableau 4.7 ou 4.8 du chapitre 4 lorsque $K = 1$. Pour tous

(p_1, p_2)	Nb d'obs.	PAR	PABDA $(h, 1/2)$	PABDA $\left(h, \frac{q_2}{q_1+q_2}\right)$	RPW (1,1)	RPW (5,1)	DL (1,1,1)	DL (1,5,1)
(0,4 , 0,1)	78	0,879	0,885	0,880	0,875	0,879	0,876	0,876
(0,65 , 0,35)	78	0,770	0,773	0,752	0,737	0,756	0,759	0,742
(0,8 , 0,5)	78	0,813	0,809	0,772	0,753	0,788	0,788	0,788
(0,6 , 0,1)	28	0,828	0,823	0,801	0,790	0,808	0,822	0,819
(0,75 , 0,25)	28	0,774	0,769	0,731	0,702	0,750	0,759	0,760
(0,9 , 0,4)	28	0,830	0,829	0,730	0,730	0,799	0,814	0,820

TAB. 5.7 – Puissances de l'intervalle AC de niveau 95 % pour les plans étudiés pour différentes valeurs de (p_1, p_2) lorsque la puissance désirée est 0,75. Pour calculer le nombre d'observations nécessaires, on suppose $\pi = 1/2$.

(p_1, p_2)	Nb d'obs.	PAR	PABDA $(h, 1/2)$	PABDA $\left(h, \frac{q_2}{q_1+q_2}\right)$	RPW (1,1)	RPW (5,1)	DL (1,1,1)	DL (1,5,1)
(0,4 , 0,1)	78	0,883	0,888	0,884	0,879	0,883	0,880	0,880
(0,65 , 0,35)	78	0,775	0,776	0,760	0,744	0,763	0,766	0,750
(0,8 , 0,5)	78	0,817	0,812	0,778	0,763	0,792	0,793	0,793
(0,6 , 0,1)	28	0,835	0,831	0,812	0,803	0,818	0,830	0,827
(0,75 , 0,25)	28	0,798	0,787	0,755	0,734	0,772	0,784	0,786
(0,9 , 0,4)	28	0,838	0,837	0,743	0,742	0,806	0,819	0,825

TAB. 5.8 – Puissances de l'intervalle BA de niveau 95 % pour les plans étudiés pour différentes valeurs de (p_1, p_2) lorsque la puissance désirée est 0,75. Pour calculer le nombre d'observations nécessaires, on suppose $\pi = 1/2$.

les (p_1, p_2) étudiés, les puissances obtenues lorsque $\pi = 1/2$ sont comparables à celles obtenues pour les plans PAR et PABDA($h, 1/2$), mais elles sont toujours plus grandes que celles des autres plans. Au moins, la puissance désirée est atteinte dans la plupart des cas. C'est le plan RPW(1,1) qui donne les moins bonnes puissances, la puissance désirée n'étant pas atteinte pour trois des six valeurs de (p_1, p_2) utilisées.

5.4.6 Puissance et probabilité de recouvrement des intervalles de confiance répétés

Au chapitre précédent, on a examiné la performance des intervalles répétés de Wald, AC et BA dans le cas où $\pi = 1/2$. On examine maintenant la performance de l'intervalle AC lorsque des plans adaptatifs sont utilisés. On laissera tomber les intervalles de Wald et BA pour les comparaisons. Le nombre d'observations entre chaque étape sera déterminé en supposant que $\pi = 1/2$ comme au chapitre 4. Les α_k des intervalles AC seront toujours ceux des tests de Pocock de niveau 95 %.

REMARQUE : On a aussi calculé les biais de \hat{p}_1 et \hat{p}_2 , les $\frac{\lambda_A}{n_1} + \frac{\lambda_B}{n_2}$ et les taux de succès moyens obtenus à partir d'expérience à étapes multiples. Dans certains cas, les biais de \hat{p}_1 et \hat{p}_2 sont plus élevés que ceux observés à partir d'une expérience à taille fixe (tableau 5.1). Cela est probablement causé par le fait que les expériences à étapes multiples peuvent se terminer avec relativement peu d'observations. Autrement, les résultats sont comparables à ceux obtenus à partir d'une expérience à taille fixe.

Les tableaux 5.9 et 5.10 présentent les probabilités de recouvrement approximatives des intervalles répétés AC lorsque $K = 2$ et $K = 5$ pour $p_1 = p_2$. Elles sont calculées comme à la section 5.4.4. Le nombre d'observations entre chaque étape est déterminé de sorte que l'on puisse détecter une différence entre p_1 et p_2 de 0,3 avec probabilité 0,75.

On voit que les probabilités de recouvrement sont généralement au moins aussi grandes que le niveau de confiance nominal. Elles sont très grandes surtout lorsque p_1 et p_2 sont loin de 1/2. Pour toutes les valeurs de (p_1, p_2) utilisées, la différence entre les probabilités de recouvrement des plans utilisés n'est pas élevée. Voyons ce qui se passe lorsque $p_1 \neq p_2$. Les tableaux 5.11 et 5.12 présentent les probabilités de recouvrement des intervalles répétés AC lorsque $K = 2$ et $K = 5$ pour $p_1 \neq p_2$. Le

(p_1, p_2)	PAR	PABDA $(h, 1/2)$	PABDA $\left(h, \frac{q_2}{q_1+q_2}\right)$	RPW $(1,1)$	RPW $(5,1)$	DL $(1,1,1)$	DL $(1,5,1)$
$(0,1, 0,1)$	0,984	0,980	0,979	0,978	0,980	0,978	0,978
$(0,25, 0,25)$	0,944	0,954	0,955	0,955	0,952	0,958	0,954
$(0,5, 0,5)$	0,946	0,951	0,945	0,944	0,946	0,946	0,948
$(0,75, 0,75)$	0,947	0,957	0,954	0,952	0,959	0,950	0,955
$(0,9, 0,9)$	0,981	0,980	0,977	0,974	0,978	0,978	0,977

TAB. 5.9 – Probabilités de recouvrement de l'intervalle AC pour les plans étudiés lorsque $p_1 = p_2$ où les α_k sont ceux des tests de Pocock de niveau 95 % avec $K = 2$. Le nombre d'observations nécessaires est calculé en supposant que $\pi = 1/2$.

(p_1, p_2)	PAR	PABDA $(h, 1/2)$	PABDA $\left(h, \frac{q_2}{q_1+q_2}\right)$	RPW $(1,1)$	RPW $(5,1)$	DL $(1,1,1)$	DL $(1,5,1)$
$(0,1, 0,1)$	0,985	0,984	0,984	0,987	0,987	0,985	0,987
$(0,25, 0,25)$	0,946	0,957	0,956	0,955	0,955	0,952	0,957
$(0,5, 0,5)$	0,931	0,943	0,945	0,942	0,942	0,940	0,940
$(0,75, 0,75)$	0,947	0,956	0,952	0,952	0,952	0,952	0,953
$(0,9, 0,9)$	0,987	0,984	0,979	0,981	0,983	0,982	0,984

TAB. 5.10 – Probabilités de recouvrement de l'intervalle AC pour les plans étudiés lorsque $p_1 = p_2$ où les α_k sont ceux des tests de Pocock de niveau 95 % avec $K = 5$. Le nombre d'observations nécessaires est calculé en supposant que $\pi = 1/2$.

(p_1, p_2)	PAR	PABDA $(h, 1/2)$	PABDA $\left(h, \frac{q_2}{q_1+q_2}\right)$	RPW $(1,1)$	RPW $(5,1)$	DL $(1,1,1)$	DL $(1,5,1)$
(0,4 , 0,1)	0,976	0,975	0,976	0,978	0,976	0,977	0,977
(0,65 , 0,35)	0,965	0,971	0,971	0,967	0,970	0,970	0,969
(0,8 , 0,5)	0,970	0,967	0,969	0,967	0,969	0,965	0,966
(0,6 , 0,1)	0,978	0,975	0,981	0,978	0,977	0,976	0,977
(0,75 , 0,25)	0,975	0,973	0,981	0,982	0,982	0,985	0,982
(0,9 , 0,4)	0,977	0,973	0,981	0,984	0,977	0,983	0,980

TAB. 5.11 – Probabilités de recouvrement de l’intervalle AC pour les plans étudiés lorsque $p_1 \neq p_2$ où les α_k sont ceux des tests de Pocock de niveau 95 % avec $K = 2$. La puissance désirée est 0,75 et le nombre d’observations nécessaires est calculé en supposant que $\pi = 1/2$.

(p_1, p_2)	PAR	PABDA $(h, 1/2)$	PABDA $\left(h, \frac{q_2}{q_1+q_2}\right)$	RPW $(1,1)$	RPW $(5,1)$	DL $(1,1,1)$	DL $(1,5,1)$
(0,4 , 0,1)	0,988	0,987	0,990	0,991	0,990	0,991	0,989
(0,65 , 0,35)	0,981	0,987	0,989	0,989	0,987	0,986	0,984
(0,8 , 0,5)	0,987	0,987	0,986	0,988	0,984	0,982	0,982
(0,6 , 0,1)	0,990	0,986	0,988	0,988	0,990	0,990	0,990
(0,75 , 0,25)	0,990	0,988	0,991	0,990	0,992	0,991	0,990
(0,9 , 0,4)	0,991	0,990	0,988	0,988	0,990	0,988	0,987

TAB. 5.12 – Probabilités de recouvrement de l’intervalle AC pour les plans étudiés lorsque $p_1 \neq p_2$ où les α_k sont ceux des tests de Pocock de niveau 95 % avec $K = 5$. La puissance désirée est 0,75 et le nombre d’observations nécessaires est calculé en supposant que $\pi = 1/2$.

nombre d’observations entre chaque étape est déterminé de sorte à avoir une puissance de 0,75.

Comme au chapitre 4 où $\pi = 1/2$, les probabilités de recouvrement sont souvent toutes plus grandes que le niveau de confiance nominal et elles sont plus grandes lorsque $K = 5$. Voyons si les puissances sont affectées par cela. Les tableaux 5.13 et 5.14 présentent les puissances des intervalles AC lorsque $K = 2$ et $K = 5$. Elles sont calculées comme à la section 5.4.5. Le nombre d’observations entre chaque étape est déterminé de sorte à avoir une puissance de 0,75.

(p_1, p_2)	PABDA		RPW		DL		
	PAR	$(h, 1/2)$	$\left(h, \frac{q_2}{q_1+q_2}\right)$	(1,1)	(5,1)	(1,1,1)	(1,5,1)
(0,4 , 0,1)	0,895	0,887	0,887	0,885	0,884	0,884	0,894
(0,65 , 0,35)	0,768	0,762	0,748	0,746	0,755	0,758	0,762
(0,8 , 0,5)	0,822	0,816	0,772	0,765	0,790	0,794	0,789
(0,6 , 0,1)	0,841	0,841	0,813	0,794	0,808	0,814	0,813
(0,75 , 0,25)	0,795	0,778	0,737	0,723	0,758	0,768	0,772
(0,9 , 0,4)	0,848	0,840	0,708	0,724	0,803	0,808	0,818

TAB. 5.13 – Puissances de l'intervalle AC pour les plans étudiés où les α_k sont ceux des tests de Pocock de niveau 95 % avec $K = 2$. La puissance désirée est 0,75 et le nombre d'observations nécessaires est calculé en supposant que $\pi = 1/2$.

(p_1, p_2)	PABDA		RPW		DL		
	PAR	$(h, 1/2)$	$\left(h, \frac{q_2}{q_1+q_2}\right)$	(1,1)	(5,1)	(1,1,1)	(1,5,1)
(0,4 , 0,1)	0,908	0,900	0,901	0,904	0,904	0,909	0,900
(0,65 , 0,35)	0,792	0,790	0,770	0,767	0,779	0,773	0,774
(0,8 , 0,5)	0,834	0,831	0,787	0,775	0,797	0,802	0,810
(0,6 , 0,1)	0,891	0,895	0,876	0,871	0,888	0,883	0,884
(0,75 , 0,25)	0,860	0,852	0,810	0,798	0,835	0,828	0,829
(0,9 , 0,4)	0,896	0,897	0,741	0,756	0,852	0,864	0,872

TAB. 5.14 – Puissances de l'intervalle AC pour les plans étudiés où les α_k sont ceux des tests de Pocock de niveau 95 % avec $K = 5$. La puissance désirée est 0,75 et le nombre d'observations nécessaires est calculé en supposant que $\pi = 1/2$.

Comme dans le cas où le nombre total d'observations est fixe, ce sont les plans PAR et PABDA($h, 1/2$) qui donnent les meilleures puissances. Les plans PABDA($h, \frac{q_2}{q_1+q_2}$) et RPW(1,1) donnent les moins bonnes puissances. On voit, comme au chapitre 4, que même si les probabilités de recouvrement ont tendance à augmenter si K augmente, les puissances augmentent aussi.

5.5 Conclusion

On ne peut pas dire qu'il existe un plan d'allocation qui performe mieux que les autres dans toutes les situations. Chacun comporte ses forces et faiblesses. Le plan PAR a l'avantage d'être difficile à prédire pour l'expérimentateur, de donner des estimateurs du maximum de vraisemblance de p_1 et p_2 pas trop biaisés et de donner des bonnes puissances. Son défaut principal est que lorsque $p_1 + p_2 > 1$, il assigne en moyenne plus d'observations au moins bon traitement. On ne devrait pas utiliser ce plan dans ces cas là. Lorsque $p_1 + p_2 < 1$, il conserve les qualités mentionnées plus haut et il assignera en moyenne plus d'observations au meilleur traitement.

Le plan PABDA($h, 1/2$) donne des estimateurs de p_1 et p_2 non biaisés, mais c'est un plan facile à prédire. Le plan PABDA($h, \frac{q_2}{q_1+q_2}$) donne des bons taux de succès, mais lui aussi est facile à prédire. Il donne aussi des estimateurs de p_1 et p_2 plus biaisés que la plupart des autres plans étudiés.

Les plans RPW(μ, α, β) et DL(a, C, M) sont des plans intéressants puisqu'ils sont relativement simples et qu'ils peuvent être adaptés si on a plus de deux traitements. Le plan RPW(1,1) donne de très bons taux de succès. Cependant, il donne des estimateurs de p_1 et p_2 biaisés et si p_1 ou p_2 est grand, il y a une probabilité non-négligeable qu'un des deux traitements reçoive très peu d'observations. C'est pourquoi les puissances observées pour ce plan sont plus petites que pour les autres plans, spécialement lorsque p_1 est grand. L'importance des défauts du plan RPW(1,1) peut être diminuée si on augmente la valeur de μ ou si on pose $\alpha > 0$. Cependant, cela aura aussi pour effet de diminuer le taux de succès.

Le plan DL(1,1,1) donne des biais pour p_1 et p_2 moins élevés que ceux du plan RPW(1,1), mais donne des $\frac{\lambda_A}{n_1} + \frac{\lambda_B}{n_2}$ moyens plus élevés, surtout lorsque p_1 et p_2 sont petits. Ce défaut peut être corrigé en augmentant a ou M , mais le taux de succès sera diminué.

Il est intéressant de voir que les intervalles AC et BA performant aussi bien dans toutes les méthodes adaptatives vues dans ce chapitre qu'au chapitre 3 où n_1 et n_2 sont fixés au départ. Dans Wei et al. (1990), on recommande, pour $n < 50$, d'utiliser des intervalles de confiance qui tiennent compte du plan adaptatif utilisé. Ces recommandations sont faites après avoir regardé les probabilités de recouvrement approximatives pour différentes valeurs de (n, p_1, p_2) des intervalles de Wald et du rapport des vraisemblances lorsque les résultats sont obtenus à partir du plan RPW(1,1). On y voit que pour des petites valeurs de n , l'intervalle du rapport des vraisemblances performe moyennement bien et que l'intervalle de Wald est nettement anti-conservateur. Or, on a déjà signalé la piètre qualité de l'intervalle de Wald au chapitre 3 pour un plan d'allocation non-adaptatif. Cela amène à penser que l'important n'est pas nécessairement de tenir compte du plan adaptatif utilisé, mais de choisir un intervalle de confiance qui performe déjà bien dans un plan non-adaptatif.

Pour tous les plans étudiés, les intervalles de confiance répétés AC donnent des probabilités de recouvrement et des puissances plus élevées lorsque K est grand comme c'était le cas au chapitre 4 où $\pi = 1/2$.

Chapitre 6

Conclusion

On a vu que l'intervalle de Wald pour une proportion ne performe pas très bien. Il donne des probabilités de recouvrement très souvent inférieures au niveau de confiance nominal surtout lorsque le nombre d'observations est petit, lorsque la vraie proportion est près de 0 ou 1 et même lorsque les règles d'utilisation proposées dans les manuels de base en statistique sont respectées. Les défauts de l'intervalle de Wald ont été signalés dans plusieurs articles et on y déplore le fait que cet intervalle est souvent le seul présenté dans les cours de base en statistique.

Il existe plusieurs alternatives à l'intervalle de Wald pour une proportion et on en a présenté cinq. Parmi celles-ci, l'intervalle d'Agresti-Coull et l'intervalle bayésien approximatif sont particulièrement intéressants puisqu'ils donnent des probabilités de recouvrement près du niveau de confiance nominal et ils ne sont pas trop longs, tout en étant pratiquement aussi simples à construire que l'intervalle de Wald.

Pour une différence de deux proportions, on a vu à nouveau que l'intervalle de Wald était anti-conservateur. Il donne des probabilités de recouvrement trop petites, surtout lorsque les tailles d'échantillons sont petites, lorsque les taux de succès sont près de 0 ou 1 et aussi lorsque l'expérience est débalancée. L'intervalle d'Agresti-Coull et l'intervalle bayésien approximatif performant bien aussi dans le cas d'une différence de deux proportions.

Dans les expériences à intervalles répétés, les probabilités de recouvrement dépendent de la différence des deux proportions inconnues si l'intervalle utilisé tient seulement compte des estimations de p_1 et p_2 . Dans ce mémoire, on ne s'est pas attardé à trouver un intervalle de confiance qui est conçu spécialement pour les expériences à étapes multiples. On s'est plutôt intéressé aux intervalles standards et on a vu que l'intervalle

d'Agresti-Coull et l'intervalle bayésien approximatif étaient un peu plus conservateurs pour ce genre d'expérience, mais que la puissance n'en était pas diminuée.

Dans une expérience à allocations séquentielles, on a vu que les probabilités de recouvrement de l'intervalle d'Agresti-Coull et de l'intervalle bayésien approximatif étaient près du niveau de confiance, peu importe le plan adaptatif choisi. Les puissances de ces intervalles étaient plus influencées par le plan utilisé. Les plans qui mettent trop l'emphase sur le taux de succès ont le défaut de donner des puissances plus petites que les autres. Il vaut mieux choisir un plan qui fait un compromis entre les deux. Il faut comprendre qu'il ne sert à rien d'avoir une expérience avec un bon taux de succès si on n'est pas capable de dire quel est le meilleur traitement à la fin de l'expérience.

Le but de ce mémoire était de montrer que ce n'est pas parce l'intervalle de Wald ne performe pas bien que tous les autres intervalles simples à construire ne performent pas bien. Dans certains articles, on a jugé nécessaire de construire des intervalles compliqués parce que l'intervalle de Wald ne performait pas bien. On est conscient que l'intervalle d'Agresti-Coull et l'intervalle bayésien approximatif ne sont pas conçus pour tous les types d'expériences et qu'il peut exister de meilleurs intervalles dans certains cas. On voulait simplement montrer qu'ils se débrouillaient bien dans plusieurs situations.

Bibliographie

Bibliographie

- [1] Agresti, A. et Caffo, B. (2000). Simple and effective confidence intervals for proportions and difference of proportions result from adding two successes and two failures. *The American Statistician*, vol. 54, No. 4, pages 280-288.
- [2] Agresti, A. et Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, vol. 52, No. 2, pages 119-126.
- [3] Auclair, J. (1999). *Méthodes randomisées adaptatives pour l'allocation séquentielle des patients à différents traitements*, Essai de maîtrise. Département de mathématiques et de statistique, Faculté des Sciences et de Génie, Université Laval.
- [4] Bélisle, C. et Melfi, V.F. (2007). Independence after adaptive allocation. À paraître dans *Statistics and Probability Letters*.
- [5] Brown, L.D., Cai, T.T. et DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, vol. 16, No. 2, pages 101-133.
- [6] Brown, L.D., Cai, T.T. et DasGupta, A. (2002). Confidence intervals for a binomial proportion and asymptotics expansions. *The Annals of Statistics*, vol. 30, No. 1, pages 160-201.
- [7] Casella, G. et Berger, R.L. (2002). *Statistical Inference*, Second Edition. DUXBURY, Pacific Grove.
- [8] Coad, D.S. et Rosenberger, W.F. (1999). A comparison of the randomized play-the-winner rule and the triangular test for clinical trials with binary responses. *Statistics in Medicine*, vol. 18, pages 761-769.
- [9] Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika*, vol. 58, No. 3, pages 403-417.
- [10] Eisele, J.R. (1994). The doubly adaptive biased coin design for sequential clinical trials. *Journal of Statistical Planning and Inference*, vol. 38, pages 249-262.

- [11] Ivanova, A. (2003). A play-the-winner rule urn design with reduced variability. *Metrika*, vol. 58, pages 1-13.
- [12] Lin, D.Y., Wei, L.J. et DeMets, D.L. (1991). Exact statistical inference for group sequential trials. *Biometrics*, vol. 47, No. 4, pages 1399-1408.
- [13] Melfi, V.F., Page, C. et Geraldès, M. (2001). An adaptive randomized design with application to estimation. *La Revue Canadienne de Statistique*, vol. 29, No. 1, pages 107-116.
- [14] Newcombe, R.G. (1998a). Two-sided confidence intervals for the single proportion : comparison of seven methods. *Statistics in Medicine*, vol. 17, pages 857-872.
- [15] Newcombe, R.G. (1998b). Interval estimation for the difference between independent proportions : comparison of eleven methods. *Statistics in Medicine*, vol. 17, pages 873-890.
- [16] O'Brien, P.C. et Fleming, T.R. (1979). A multiple testing procedure for clinical trials. *Biometrics*, vol. 35, No. 3, pages 549-556.
- [17] Pocock, S.J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, vol. 64, No.2, pages 191-199.
- [18] Tsiatis, A.A., Rosner, G.L. et Mehta, C.R. (1984). Exact confidence intervals following a group sequential test. *Biometrics*, vol. 40, No.3, pages 797-803.
- [19] Wei, L.J. et Durham, S. (1978). The randomized play-the-winner rule in medical trials. *Journal of the American Statistical Association*, vol. 73, No. 364, pages 840-843.
- [20] Wei, L.J., Smythe, R.T., Lin, D.Y. et Park, T.S. (1990). Statistical inference in data-dependent treatment allocation rules. *Journal of the American Statistical Association*, vol. 85, No. 409, pages 156-162.
- [21] Xia, Y. (2002). *Interval estimation for the difference of two binomial proportions in non-adaptive and adaptive designs*, Thèse de doctorat. *Department of Statistics and Probability, Michigan State University*.
- [22] Zelen, M. (1969). Play-the-winner rule and the controlled clinical trial. *Journal of the American Statistical Association*, vol. 64, No. 325, pages 131-146.