HAL open science

Lois bayésiennes a priori dans un plan binomial séquentiel

Pierre Bunouf

▶ To cite this version:

Pierre Bunouf. Lois bayésiennes a priori dans un plan binomial séquentiel. Mathématiques [math]. Université de Rouen, 2006. Français. NNT : . tel-00539868

HAL Id: tel-00539868

https://theses.hal.science/tel-00539868v1

Submitted on 25 Nov 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de doctorat de l'université de Rouen

Spécialité : Mathématique

présentée par

Pierre Bunouf

pour obtenir le titre de Docteur de l'université de Rouen

Lois bayésiennes a priori dans un plan binomial séquentiel

Soutenue le 1er mars 2006 devant le jury composé de

Dominique Fourdrinier, Professeur univ. Rouen :	Président
Bruno Lecoutre, Directeur de Recherche CNRS :	Directeur
Rodolfo de Cristofaro, Professeur univ. Florence :	Rapporteur
Christian Robert, Professeur univ. Dauphine :	Rapporteur
Jean-Marie Grouin, Maître conf. univ. Rouen :	Examinateur

Remerciements

Ce chapitre est un rituel auquel je me plie volontiers tant l'expérience de cette thèse m'aura enthousiasmé et apporté personnellement et professionnellement.

Je remercie donc chaleureusement Bruno Lecoutre, qui a dirigé ma thèse, pour sa gentillesse et son ouverture d'esprit. La justesse de son pilotage exprimé à travers des suggestions pertinentes (mais aussi parfois des silences évocateurs) a largement contribué au résultat.

Je remercie également Dominique Fourdrinier et l'équipe du LMRS pour la confiance qu'ils m'ont accordée. J'ai pu notamment bénéficié d'une organisation et d'un environnement optimal conjuguant efficacité et convivialité. Je souhaite sincèrement un avenir à la mesure du plaisir que m'apportait chacune de mes visites.

Je remercie naturellement Jean-Marie Grouin, générateur puissant d'idées et contradicteur de haute volée, pour son support technique, parfois moral, et puis le reste.

J'exprime toute ma gratitude à Rodolfo de Cristofaro pour avoir accepté d'être rapporteur. Dès le début j'ai eu la chance de bénéficier d'informations clés de sa part.

J'adresse des remerciements chaleureux à Christian Robert qui a également accepté le rôle de rapporteur.

Je n'oublie pas non plus mes supporters locaux, Gérard Tap et Jean-René Mathieu, du Laboratoire Statistique-Probabilité à l'UPS Toulouse.

Enfin une mention particulière aux proches ou autres collègues, qui m'ont supportés (dans les deux sens du termes) et qui se reconnaîtront. J'espère qu'ils jugeront que parfois mes absences et mon silence durant ces trois années étaient partiellement justifiés.

Mot-clés

Principe de vraisemblance, règle d'arrêt, plan Binomial séquentiel, information de Fisher, inférence bayésienne, propriétés fréquentistes, principe de translation de la vraisemblance, principe d'impartialité du plan expérimental, règle de Jeffreys, estimation ponctuelle, intervalle de confiance, test d'hypothèse, facteur de Bayes, principe d'égalité aux bornes.

Key-words

Likelihood principle, stopping rule, sequential binomial design, Fisher information, Bayesian inference, frequentist properties, likelihood translation principle, impartiality of the design principle, Jeffreys rule, point estimation, confidence interval, hypothesis test, Bayes factor, equality at bounds principle.

Résumé

La reformulation du théorème de Bayes par R. de Cristofaro permet d'intégrer l'information sur le plan expérimental dans la loi a priori. En acceptant de transgresser les principes de vraisemblance et de la règle d'arrêt, un nouveau cadre théorique permet d'aborder le problème de la séquentialité dans l'inférence bayésienne. En considérant que l'information sur le plan expérimental est contenue dans l'information de Fisher, on dérive une famille de lois a priori à partir d'une vraisemblance directement associée à l'échantillonnage. Le cas de l'évaluation d'une proportion dans le contexte d'échantillonnages Binomiaux successifs conduit à considérer la loi Bêta-J. L'étude sur plusieurs plans séquentiels permet d'établir que l'"a priori de Jeffreys corrigé" compense le biais induit sur la proportion observée. Une application dans l'estimation ponctuelle montre le lien entre le paramétrage des lois Bêta-J et Bêta dans l'échantillonnage fixe. La moyenne et le mode des lois a posteriori obtenues présentent des propriétés fréquentistes remarquables. De même, l'intervalle de Jeffreys corrigé montre un taux de recouvrement optimal car la correction vient compenser l'effet de la règle d'arrêt sur les bornes. Enfin, une procédure de test, dont les erreurs s'interprètent à la fois en terme de probabilité bayésienne de l'hypothèse et de risques fréquentistes, est construite avec une règle d'arrêt et de rejet de H_0 fondée sur une valeur limite du facteur de Bayes. On montre comment l'a priori de Jeffreys corrigé compense le rapport des évidences et garantit l'unicité des solutions, y compris lorsque l'hypothèse nulle est composite.

Abstract

Bayesian priors in sequential binomial design. The reformulation by R. de Cristofaro of the Bayes theorem allows to integrate the information on the experimental design in the prior distribution. In accepting to transgress the Likelihood and Stopping Rules principles, a new framework allows to move on the issue of sequenciality in the Bayesian inference. Considering that the information on the design is contained in the Fisher information, a new family of priors is derived from a likelihood directly related to the sampling rule. The case of the study of a proportion in the sequential context of successive Binomial samplings leads to consider the Beta-J distribution. The study on several sequential designs allows to state that the "corrected Jeffreys prior" compensates the bias induced on the observed proportion. An application in the estimation shows the relationship between the parameters of the Beta-J and Beta distributions in the fixed sampling. The mean and mode of the posterior distributions show remarkable frequentist properties. As well, the corrected Jeffreys interval has an optimal covering rates as the correction compensates the effect of stopping rule on the limits. Last, a test procedure, whose the errors are interpreted in terms of both Bayesian probabilities of hypotheses and frequentist risks, is designed with a rule for stopping and rejecting H_0 based on a limit value of the Bayes factor. It is shown how the corrected Jeffreys prior compensates the ratio of evidences and guarantees the unicity of solutions, even when the null hypothesis is composite.

Table des matières

Introduction **26**

1 Principe de Vraisemblance **29**

1.1 Évidence statistique 29

1.2 Principe de Vraisemblance 33

1.3 Principe de la règle d'arrêt 36

1.4 Echantillonnage séquentiel dans l'approche Bayésienne classique . . . 38

2 Loi a priori **40**

2.1 Loi a priori et principe de vraisemblance 40

2.1.1 A priori informatif 40

2.1.2 A priori non-informatif 41

2.2 Prise en compte du plan d'expérimental 48

2.3 A priori de Jeffreys corrigé pour un plan expérimental comportant des analyses intermédiaires 51

2.4 Influence de la correction dans un plan expérimental comportant une analyse intermédiaire 53

2.4.1 Propriété de la loi a priori de Jeffreys corrigée 54

2.4.2 Etude du biais induit sur l'EMV lorsque $n_1 = 1$ 57

2.5 Propriété de la loi Bêta-J d'ordre 1 60

3 Application à l'estimation ponctuelle 66

3.1 Procédure de Whitehead et Estimateur VUMSB 66

3.2 Equation du nouvel estimateur 69

3.3 Cas d'un arrêt conditionné à l'observation d'au moins un événement . 73

3.3.1 Principe de résolution graphique 73

3.3.2 Dérivation d'une forme analytique non-linéaire 77

3.3.3 Dérivation d'une forme analytique linéaire 81

3.3.4 Approximation du biais de l'estimateur du Mode 83

3.3.5 Etude comparative du biais et de l'efficacité relative lorsque $n_1 = n_2$ 92

3.3.6 Etude comparative du biais lorsque $n_1 \neq n_2$ 94

3.3.7 Dérivation de l'estimateur du Mode Corrigé 98

3.4 Généralisation à d'autres plans séquentiels 100

3.4.1 Dérivation de l'estimateur du Mode Corrigé 100

3.4.2 Etude du biais et de l'efficacité relative dans différents plans bornés unilatéralement 106

3.4.3 Propriété de l'estimateur du Mode Corrigé dans un plan borné bilatéralement 110

3.4.4 Comparaison du biais et de l'efficacité relative avec d'autres estimateurs 113

3.5 Conclusion 117

4 Application à l'estimation par intervalle 118

4.1 Différentes méthodes dans un échantillonnage fixe 118

4.2 Méthode fréquentistes non-conditionnelles dans un plan séquentiel . . 122

4.3 Taux de recouvrement et longueur moyenne 124

4.4 Construction de l'intervalle de Jeffreys corrigé 136

4.5 Performance des méthodes usuelles dans le cas d'un arrêt conditionné à l'observation d'au moins un événement 145

4.6 Généralisation à d'autres plans expérimentaux 153

4.7 Conclusion 164

5 Application au test d'hypothèses 166

5.1 Tests fréquentistes séquentiels fondés sur des fonctions de dépense des risques 167

5.2 Test fréquentiste conditionnel séquentiel ($TFCS^*$) 169

5.2.1 Construction du test 169

5.2.2 Application à un test de supériorité 172

5.3 Test fréquentiste non-conditionnel séquentiel ($TFNCS$) 178

5.3.1 Construction du test 178

5.3.2 Exemple 183

5.4 Application à une problématique d'équivalence 188

5.4.1 Détermination des tests $TFCS^*$ et $TFNCS$ 188

5.4.2 Test symétrique aux bornes ($TFNCS^{SB}$) 192

5.4.3 Etude de l'erreur de type 1 du test $TFNCS^{SB}$ en fonction du décentrage de Θ_0 197

5.5 Conclusion 200

5.6 Discussion 202

Conclusion **204**

Discussion et perspectives **208**

A Notions de base **215**

A.1 Éléments de probabilité 215

A.2 Échantillonnage 217

A.3 Réduction des données 220

A.4 Information de Fisher 223

B Inférence statistique **228**

B.1 Approches bayésienne et fréquentiste 228

B.2 Estimation ponctuelle 230

B.3 Estimation par intervalle 232

B.4 Test d'hypothèses 236

B.4.1 Tests dans l'inférence fréquentiste 236

B.4.2 Tests dans l'inférence bayésienne 238

B.5 Inférence fréquentiste conditionnelle 241

B.5.1 Conditionnement et partitionnement 241

B.5.2 Test Conditionnel BBW 242

B.5.3 Test Conditionnel d'une proportion 247

B.5.4 Calibration de la fonction 'p' non-conditionnelle 250

C Figures et Tableaux **253**

D Bayesian priors in sequential binomial design **283**

E Bayesian estimators in sequential binomial design **285**

F Bayesian confidence intervals in sequential binomial design **287**

G Bayesian testing procedures in sequential binomial design **289**

Liste des tableaux

5.1 *Test de Dass avec l'a priori de Jeffreys classique pour* $n=10$ 173

5.2 *Densités prédictives et facteur de Bayes de* H_0 *sur* H_1 *avec les a priori de Jeffreys classique et corrigé pour* $\alpha_r^* = 0.2$ *et* $n_1=n_2=10$ 174

5.3 *Décisions et risques fréquentistes conditionnels du test* $TFCS^*$ *avec les a priori de Jeffreys classique et corrigé pour* $\alpha_r^* = 0.2$ *et* $n_1 = n_2 = 10$ 177

5.4 *Bornes* $(S_{1,L}, S_{2,L})$ *avec les a priori de Jeffreys classique et corrigé pour* $\alpha_r^* = 0.20$ *(i.e.* $B_r = 1/4$*) et* $n_1=n_2=10$, 20, 30*, 40*, 50*, 60*, 80*, 100* 179

5.5 *Bornes* $(S_{1,L}, S_{2,L})$ *avec les a priori de Jeffreys classique et corrigé pour* $\alpha_r^* = 0.10$ *(i.e.* $B_r = 1/9$*) et* $n_1=n_2=10$, 20*, 30, 40*, 50*,60*, 80*, 100* 179

5.6 *Bornes* $(S_{1,L}, S_{2,L})$ *avec les a priori de Jeffreys classique et corrigé pour* $\alpha_r^* = 0.05$ *(i.e.* $B_r = 1/19$*) et* $n_1=n_2=10$, 20, 30*, 40*, 50*, 60*, 80*, 100* 183

5.7 *Décisions et risques associés au test* $TFCS^*$ *et au test* $TFNCS$ *avec l'a priori de Jeffreys corrigé pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2 = 30$ 186

5.8 *Bornes* $(S_{1,L}, S_{1,U}, S_{2,L}, S_{2,U})$ *avec les a priori de Jeffreys classique et corrigé pour* $\alpha_r^* = 0.10$ *(i.e.* $B_r = 1/9$*) et* $n_1 = n_2 = 30$, 40, 50*, 60, 80, 100*, 120, 150 189

5.9 *Bornes* $(S_{1,L}, S_{1,U}, S_{2,L}, S_{2,U})$ *avec les a priori de Jeffreys classique et corrigé pour* $\alpha_r^* = 0.05$ *(i.e.* $B_r = 1/19$*) et* $n_1 = n_2 = 60$, 80, 100, 120, 150 189

5.10 *Bornes* $(S_{1,L}, S_{1,U}, S_{2,L}, S_{2,U})$ *pour* $\alpha_r^* = 0.05$ *(i.e.* $B_r = 1/19$*) et* $n_1 = n_2 = 100$ 193

5.11 *Probabilités de rejeter* H_0 *avec l'a priori de Jeffreys corrigé pour* $\alpha_r^* = 0.05$ *(i.e.* $B_r = 1/19$*) et* $n_1 = n_2 = 100$ 193

5.12 *Facteurs de Bayes et risques associés au test* $TFNCS^{SB}$ *et à la variante correspondante du test* $TFCS^*$ *pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2 = 100$ 195

C.1 *Décisions et risques associés au test* $TFCS^*$ *et au test* $TFNCS$ *avec l'a priori de Jeffreys classique pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2 = 30$ 279

Table des figures

2.1 *Noyau de l'a priori de Jeffeys pour un échantillonnage Binomial (– –) et de l'a priori de Jeffeys (impropre) pour un échantillonnage Binomial Négatif (—)* 47

2.2 *Fonctions de répartition de la loi Binomiale* $CBin(n_1, \theta; J = [0, S_{1,U} - 1])$ *avec* $S_{1,U} = 1, 2, .., 5$ 55

2.3 *Noyaux de la loi a priori de Jeffreys classique dans l'échantillonnage Binomial (- - -) et de la loi a priori de Jeffreys corrigée (—) dans les échantillonnages séquentiels* $d_{Seq_{(1,1)}}$*=(*n_1*=5,*$S_{1,U} = s_{1,U}$*,*n_2*=5) avec* $s_{1,U}$*=1,2,3,4,5* 55

2.4 *Noyau de la loi a priori de Jeffreys classique dans les échantillonnages Binomial (- - -) et Binomial Négatif (·−·), et corrigé dans les échantillonnages séquentiels* $d_{Seq_{(1,1)}}$*=(*n_1*=1,*$S_{1,U}$*=1,*n_2*=1) (—) et* $d_{Seq_{(1,+\infty)}}$*=(*n_1*=1,*$S_{1,U}$*=1,*$n_2 \to +\infty$*) (– –)* 57

2.5 *Espérance de* $\hat{\theta}^{MV} = Y/N$ *dans l'échantillonnage Binomial (- - -), Binomial Négatif (·−·), séquentiel avec* $d_{Seq_{(1,1)}}$*=(*n_1*=1,*$S_{1,U}$*=1,*n_2*=1) (—) et* $d_{Seq_{(1,+\infty)}}$*=(*n_1*=1,*$S_{1,U}$*=1,*$n_2 \to +\infty$*) (– –)* 59

3.1 *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) pour* $\mathcal{Y}_{(1,y)} = \{y = y_1; y=1, 2, .., 5\}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$ 74

3.2 *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) pour* $\mathcal{Y}_{(2,y)} = \{y = y_1+y_2; y=0, 1, .., 5\}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$ . . 74

3.3 *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) pour* $\mathcal{Y}_{(1,y)} = \{y = y_1; y=1, 2, .., 5\}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$ 75

3.4 *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) pour* $\mathcal{Y}_{(2,y)} = \{y = y_1+y_2; y=0, 1, .., 5\}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$ . . 75

3.5 *Estimateur du Mode en fonction de* $(m, y/n)$ *pour* $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 1, 2, .., 5\}$ *et* $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 5\}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 77

3.6 *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) pour* $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 1, 2, .., 5\}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 78

3.7 *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) pour* $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 5\}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ . . 78

3.8 *Fonctions holomorphes* $g(\theta)$, $g_{0.5}(\theta)$, $g_1(\theta)$ *(- - -) et* $h(\theta)$ *(—) pour* $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 1, 2, .., 5\}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 80

3.9 *Estimations* $\hat{\theta}^{Mod}_{(m,y)}$ *(—),* $\hat{\theta}^{ModA}_{(m,y)}$ *(- - -),* $\hat{\theta}^{ModA_{0.5}}_{(m,y)}$ *(· − ·), et* $\hat{\theta}^{ModA_1}_{(m,y)}$ *(· · ·) soustraites de* y/n *pour* $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 1, 2, .., 5\}$ *et* $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 5\}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ . . 81

3.10 *Estimations* $\hat{\theta}^{Mod}_{(m,y)}$ *(—),* $\hat{\theta}^{ModL}_{(m,y)}$ *(- - -),* $\hat{\theta}^{ModL_{0.5}}_{(m,y)}$ *(· − ·) et* $\hat{\theta}^{ModL_1}_{(m,y)}$ *(· · ·) soustraites de* y/n *pour* $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 1, 2, .., 5\}$ *et* $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 5\}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ . . 83

3.11 *Biais de* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1, S_{1,U} = 1, n_2)$ *avec* $n_1 = n_2 = 2, 5, 10$ 86

3.12 *Biais de* $\hat{\theta}^{Mod}$ *dans le plan séquentiel* $(n_1, S_{1,U} = 1, n_2)$ *avec* $n_1 = n_2 = 2, 5, 10$ 86

3.13 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{Mod}$ *(—),* $\hat{\theta}^{ModA}$ *(- - -),* $\hat{\theta}^{ModA_{0.5}}$ *(· − ·) et* $\hat{\theta}^{ModA_1}$ *(· · ·) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 88

3.14 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{Mod}$ *(—),* $\hat{\theta}^{ModL}$ *(- - -),* $\hat{\theta}^{ModL_{0.5}}$ *(· − ·) et* $\hat{\theta}^{ModL_1}$ *(· · ·) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 89

3.15 *Approximation du biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{Mod}$ *(—),* $\hat{\theta}^{ModL}$ *(- - -),* $\hat{\theta}^{ModL_{0.5}}$ *(· − ·) et* $\hat{\theta}^{ModL_1}$ *(· · ·) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$. 91

3.16 *EQM de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{Mod}$ *(—),* $\hat{\theta}^{ModA}$ *(- - -) et* $\hat{\theta}^{VUMSB}$ *(· − ·) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 93

3.17 *Efficacité Relative de* $\hat{\theta}^{Mod}$ *(—),* $\hat{\theta}^{ModA}$ *(- - -) et* $\hat{\theta}^{VUMSB}$ *(· − ·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 94

3.18 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{Mod}$ *(—) et* $\hat{\theta}^{ModA}$ *(- - -) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=1,2,..,10)$ 95

3.19 *Efficacité Relative de* $\hat{\theta}^{Mod}$ *(—),* $\hat{\theta}^{VUMSB}$ $(\cdot - \cdot)$ *versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=1,2,..,10)$ 96

3.20 *Efficacité Relative de* $\hat{\theta}^{Mod}$ *(—) et* $\hat{\theta}^{ModA}$ *(- - -) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=1,2,..,10)$ 96

3.21 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{Mod}$ *(—) et* $\hat{\theta}^{ModA}$ *(- - -) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5,10,20,40,..,120,150)$ 97

3.22 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{Mod_{R=1}}$ *(—) et* $\hat{\theta}^{ModA_{R=1}}$ *(- - -) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5,10,20,40,..,120,150)$ 98

3.23 *Efficacité Relative de* $\hat{\theta}^{Mod_{R=1}}$ *(—),* $\hat{\theta}^{ModA_{R=1}}$ *(- - -) et de* $\hat{\theta}^{VUMSB}$ $(\cdot - \cdot)$ *versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5,10,20,40,..,120,150)$ 99

3.24 *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) dans* $\mathcal{Y}_{(1,y)} = \{y = y_1; y=2,3,4,5\}$ *et* $h(\theta)$ $(\cdots)$ *dans* $\mathcal{Y}_{(2,y)} = \{y=y_1+y_2; y=0,1,..,6\}$ *pour le plan séquentiel* $(n_1=5, S_{1,U}=2, n_2=5)$ 102

3.25 *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) dans* $\mathcal{Y}_{(1,y)} = \{y = y_1; y=3,4,5\}$ *et* $h(\theta)$ $(\cdots)$ *dans* $\mathcal{Y}_{(2,y)} = \{y=y_1+y_2; y=0,1,..,7\}$ *pour le plan séquentiel* $(n_1=5, S_{1,U}=3, n_2=5)$ 102

3.26 *Estimation de* $\hat{\theta}^{ModC}_{(m,y)}$ *(—),* $\hat{\theta}^{ModCA}_{(m,y)}$ *(- - -) et* $\hat{\theta}^{VUMSB}_{(m,y)}$ $(\cdot - \cdot)$*, soustraite de* $\hat{\theta}^{MV}_{(m,y)}$*, dans* $\mathcal{Y}_{(1,y)}=\{y=y_1; y=1,2,..,5\}$ *et* $\mathcal{Y}_{(2,y)}=\{y=y_1+y_2; y=0,1,..,5\}$ *pour le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$ 104

3.27 *Estimation de* $\hat{\theta}^{ModC}_{(m,y)}$ *(—),* $\hat{\theta}^{ModCA}_{(m,y)}$ *(- - -) et* $\hat{\theta}^{VUMSB}_{(m,y)}$ $(\cdot - \cdot)$*, soustraite de* $\hat{\theta}^{MV}_{(m,y)}$*, dans* $\mathcal{Y}_{(1,y)}=\{y=y_1; y=2,3,4,5\}$ *et* $\mathcal{Y}_{(2,y)}=\{y=y_1+y_2; y=0,1,..,6\}$ *pour le plan séquentiel* $(n_1=5, S_{1,U}=2, n_2=5)$ 105

3.28 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -) dans le plan séquentiel* $(n_1=5, S_{1,U}=s_{1,U}, n_2=5)$ *avec* $s_{1,U}=1,2,3,4,5$ 107

3.29 *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{ModCA}$ *(- - -) et* $\hat{\theta}^{VUMSB}$ $(\cdot - \cdot)$ *versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=2, n_2=5)$ 108

3.30 *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{ModCA}$ *(- - -) et* $\hat{\theta}^{VUMSB}$ $(\cdot - \cdot)$ *versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=3, n_2=5)$ 108

3.31 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -) dans le plan séquentiel* $(n_1=5, S_{1,U}=s_{1,U}, n_2=2)$ *avec* $s_{1,U}=1,2,3,4,5$ 109

3.32 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -) dans le plan séquentiel* $(n_1=5, S_{1,U}=s_{1,U}, n_2=10)$ *avec* $s_{1,U}=1,2,3,4,5$ 109

3.33 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -) dans le plan séquentiel* $(n_1=5, S_{1,U}=s_{1,U}, n_2=50)$ *avec* $s_{1,U}=1,2,3,4,5$ 110

3.34 *Fonctions holomorphes* $g(\theta)$ *(- - -),* $h(\theta)$ *(—) dans* $\mathcal{Y}_{(1,y)}=\{y=y_1; y=1,2,..,5\}$ *et* $h(\theta)$ $(\cdots)$ *dans* $\mathcal{Y}_{(2,y)}=\{y=y_1+y_2; y=0,1,..,10\}$ *pour les plans séquentiels* $(n_1=5, S_{1,L}=s_{1,L}, S_{1,U}=s_{1,L}+3, n_2=5)$ *avec* $s_{1,L}=-1,0,1,2,3$ 111

3.35 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -) dans le plan séquentiel* $(n_1=5, S_{1,L}=0, S_{1,U}=3, n_2=5)$ 112

3.36 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -) dans le plan séquentiel* $(n_1=5, S_{1,L}=1, S_{1,U}=4, n_2=5)$ 112

3.37 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et de* $\hat{\theta}^{Whi}$ $(\cdot - \cdot)$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$ 115

3.38 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et de* $\hat{\theta}^{Whi}$ $(\cdot - \cdot)$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=2, n_2=5)$ 115

3.39 *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et* $\hat{\theta}^{Whi}$ $(\cdot - \cdot)$ *versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$ 116

3.40 *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et* $\hat{\theta}^{Whi}$ $(\cdot - \cdot)$ *versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=2, n_2=5)$ 116

4.1 *Fonctions* $P_\theta(Y \geq i)$ $i=1,..,5$ *où* $Y \sim Bin(\theta, n=5)$ *avec un découpage par les bornes* $\theta_J^{inf}(y)$ *et* $\theta_J^{sup}(y)$ *de l'intervalle de Jeffreys 90%* 126

4.2 *Taux de recouvrement de l'intervalle de Jeffreys 90% (n=5)* 126

4.3 *Taux de recouvrement de l'intervalle de Jeffreys modifié 90%* $(n=5)$. 128

4.4 *Taux de recouvrement de l'intervalle de Clopper-Pearson 90%* $(n=5)$ 128

4.5 *Taux de recouvrement de l'intervalle de Wilson 90%* $(n=5)$ 129

4.6 *Taux de recouvrement de l'intervalle d'Agresti-Coull 90%* (n=5) . . . 129

4.7 *Taux de recouvrement de l'intervalle corrigé au second-ordre 90%* (n=5) 130

4.8 *Longueur moyenne des intervalles de confiance à 90%* (n=5). *Les intervalles de* CI_J, CI_{JM} *et* CI_{CP} *(—) sont respectivement représentés de bas en haut avec également les intervalles* CI_W *(· · ·),* CI_{AC} *(· − ·) et* CI_2 *(- - -).* 131

4.9 *Ensemble des fonctions* $P_\theta(Y_1 \geq i)$ $i = 1, .., 5$ *(—) et* $P_\theta(Y_1 \geq i) + P_\theta(Y_1 = 0)P_\theta(Y_2 \geq j)$ $i = 1, .., 5$ *et* $j = 0, .., 5$ *(− −) avec* n_1=n_2=5 . . 134

4.10 *Fonctions supports des taux de recouvrement de l'intervalle de Jeffreys 90% dans le plan* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ *avec le découpage par les bornes* $\hat{\theta}_J^{inf}(1, y)$ *et* $\hat{\theta}_J^{sup}(1, y)$ *avec* $y \in \mathcal{Y}_{(1)}$ *(—) et par* $\hat{\theta}_J^{inf}(2, y)$ *et* $\hat{\theta}_J^{sup}(2, y)$ *avec* $y \in \mathcal{Y}_{(2)}$ *(- - -)* 135

4.11 *Taux de recouvrement des intervalles 90% de Jeffreys (—) dans le plan séquentiel* (n_1=5, $S_{1,U}$=1, n_2=5) 136

4.12 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,1) dans le plan séquentiel* (n_1= 5, $S_{1,U}$=1, n_2=5) 137

4.13 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,2) dans le plan séquentiel* (n_1= 5, $S_{1,U}$=1, n_2=5) 138

4.14 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,3) dans le plan séquentiel* (n_1= 5, $S_{1,U}$=1, n_2=5) 138

4.15 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,4) dans le plan séquentiel* (n_1= 5, $S_{1,U}$=1, n_2=5) 139

4.16 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,5) dans le plan séquentiel* (n_1= 5, $S_{1,U}$=1, n_2=5) 139

4.17 *Bornes des intervalles de Jeffreys 90% classique (- - -) et corrigé (—) dans le plan séquentiel* (n_1=5, $S_{1,U}$=1, n_2=5) 141

4.18 *Bornes des intervalles de Jeffreys 90% modifié classique (- - -) et Jeffeys modifié corrigé (—) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 142

4.19 *Taux de recouvrement des intervalles de Jeffreys 90% classique (- - -) et corrigé (—) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 142

4.20 *Taux de recouvrement des intervalles de Jeffreys modifiés 90% classique (- - -) et corrigé (—) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 143

4.21 *Longueur moyenne des intervalles de Jeffreys et Jeffreys modifiés 90% classique (- - -) et corrigé (—) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 144

4.22 *Bornes des intervalles 90% de Wilson (- - -), d'Agresti-Coull (—) et de l'intervalle corrigé au second-ordre* $(\cdot - \cdot)$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 146

4.23 *Bornes des intervalles 90% de Clopper-Pearson (—) et Jennison-Turnbull* $(\cdot - \cdot)$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 147

4.24 *Taux de recouvrement de l'intervalle de Wilson 90% dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 148

4.25 *Taux de recouvrement de l'intervalle d'Agresti 90% dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 149

4.26 *Taux de recouvrement de l'intervalle du second-ordre 90% dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 149

4.27 *Taux de recouvrement de l'intervalle de Clopper-Pearson 90% dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 150

4.28 *Taux de recouvrement de la borne inférieure des intervalles de Clopper-Pearson 90% classique (- - -) et corrigé (—) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 151

4.29 *Taux de recouvrement de l'intervalle de Jennison-Turnbull 90% dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ 152

4.30 *Longueur moyenne des intervalles de confiance à 90% dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 1, n_2 = 5$). *Les intervalles de* CI_J, CI_{JM} *et* CI_{CP} *(—) sont respectivement représentés de bas en haut, avec également les intervalles* CI_{JT} *(– –),* CI_W *(· · ·),* CI_{AC} *(· – ·) et* CI_2 *(- - -).* 152

4.31 *Taux de recouvrement des intervalles de Jeffreys 90% classique (- - -) et corrigé (—) dans le plan séquentiel* ($n_1 = 10, S_{1,U} = 2, n_2 = 10$) . . . 154

4.32 *Taux de recouvrement des intervalles de Jeffreys 90% classique (- - -) et corrigé (—) dans le plan séquentiel* ($n_1 = 20, S_{1,U} = 4, n_2 = 20$) . . . 154

4.33 *Bornes des intervalles de Jeffreys 90% classique (- - -) et corrigé (—), et Jennison-Turnbull (· – ·) dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 2, n_2 = 5$) 155

4.34 *Taux de recouvrement des intervalles de Jeffreys 90% classique (- - -) et corrigé (—) dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 2, n_2 = 5$) 156

4.35 *Taux de recouvrement de la borne inférieure de l'intervalle de Clopper-Pearson 90% dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 2, n_2 = 5$) 157

4.36 *Taux de recouvrement de la borne inférieure de l'intervalle de Jennison-Turnbull 90% dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 2, n_2 = 5$) 157

4.37 *Taux de recouvrement des intervalles de Jeffreys 90% classique (- - -) et corrigé (—) dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5$) 159

4.38 *Taux de recouvrement des intervalles 90% de Jeffreys classique (- - -) et corrigé (—) dans le plan séquentiel* ($n_1 = 10, S_{1,U} = 2, S_{1,L} = 8, n_2 = 10$) 160

4.39 *Taux de recouvrement des intervalles 90% de Jeffreys classique (- - -) et corrigé (—) dans le plan séquentiel* ($n_1 = 20, S_{1,U} = 4, S_{1,L} = 16, n_2 = 20$) 160

4.40 *Taux de recouvrement de l'intervalle de Wilson 90% dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5$) 161

4.41 *Taux de recouvrement de l'intervalle d'Agresti-Coull 90% dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5$) 161

4.42 *Taux de recouvrement de l'intervalle du second-ordre 90% dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5$) 162

4.43 *Taux de recouvrement de l'intervalle de Clopper-Pearson 90% dans le plan séquentiel* ($n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5$) 162

4.44 *Taux de recouvrement de l'intervalle de Jennison-Turnbull 90% dans le plan séquentiel* ($n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5$) 163

4.45 *Longueur moyenne des intervalles de Jeffreys et Jeffreys modifié 90% classique (- - -) et corrigé (—) dans le plan séquentiel* ($n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5$) 163

5.1 *Risques fréquentistes conditionnels du test $TFCS^*$ avec les a priori de Jeffreys classique (- - -) et corrigé (—) pour $\alpha_r^* = 0.2$ et $n_1=n_2=10$* 176

5.2 *Courbes de puissance du test $TFNCS$ avec les a priori de Jeffreys classique ($\cdot - \cdot$) et corrigé (—) pour $\alpha_r^* = 0.20$ et $n_1=n_2=10$, 20, 30*, 40*, 50*, 60*, 80*, 100** 180

5.3 *Courbes de puissance du test $TFNCS$ avec les a priori de Jeffreys classique ($\cdot - \cdot$) et corrigé (—) pour $\alpha_r^* = 0.10$ et $n_1=n_2=$ 10, 20*, 30, 40*, 50*, 60*, 80*, 100** 181

5.4 *Courbes des effectifs moyens du test $TFNCS$ avec les a priori de Jeffreys classique ($\cdot - \cdot$) et corrigé (—) pour $\alpha_r^* = 0.10$ et $n_1=n_2=10$, 20*, 30, 40*, 50*,60*, 80*, 100** 182

5.5 *Courbes de puissance du test $TFNCS$ avec les a priori de Jeffreys classique ($\cdot - \cdot$) et corrigé (—) pour $\alpha_r^* = 0.05$ et $n_1=n_2=10$, 20, 30*, 40*, 50*, 60*, 80*, 100** 184

5.6 *Courbes des effectifs moyens du test $TFNCS$ avec les a priori de Jeffreys classique ($\cdot - \cdot$) et corrigé (—) pour $\alpha_r^* = 0.05$ et $n_1=n_2=10$, 20, 30, 40, 50*, 60*, 80*, 100** 184

5.7 *Risques fréquentistes conditionnels associés au test $TFCS^*$ avec les a priori de Jeffreys classique (- - -) et corrigé (—) pour $\alpha_r^* = 0.05$ et $n_1=n_2=30$* 187

5.8 *Courbes de puissance du test $TFNCS$ avec les a priori de Jeffreys classique ($\cdot - \cdot$) et corrigé (—) pour $\alpha_r^* = 0.10$ et $n_1 = n_2=30$, 40, 50*, 60, 80, 100*, 120, 150* 190

5.9 *Courbes de puissance du test $TFNCS$ avec les a priori de Jeffreys classique et corrigé (—) pour $\alpha_r^* = 0.05$ et $n_1 = n_2$=60, 80, 100, 120, 150* 191

5.10 *Courbes des effectifs moyens du test $TFNCS$ avec les a priori de Jeffreys classique et corrigé (—) pour $\alpha_r^* = 0.05$ et $n_1 = n_2$=60, 80, 100, 120, 150* 191

5.11 *Courbes de puissance du test $TFNCS^{SB}$ (—) et du test $TFNCS$ avec la pondération $(w_a, w_b) = (0.5, 0.5)$ de l'a priori de Jeffreys (- - -) pour $\alpha_r^* = 0.05$ et $n_1 = n_2 = 100$* 194

5.12 *Courbes des effectifs moyens du test $TFNCS^{SB}$ (—) et du test $TFNCS$ avec la pondération $(w_a, w_b) = (0.5, 0.5)$ de l'a priori de Jeffreys (- - -) pour $\alpha_r^* = 0.05$ et $n_1 = n_2 = 100$* 194

5.13 *Risques fréquentistes conditionnels du test $TFCS^*$ associé au test $TFNCS^{SB}$ obtenus avec les a priori de Jeffreys classique (- - -) et corrigé (—) pour $\alpha_r^* = 0.05$ et $n_1 = n_2 = 100$* 196

5.14 *Pondérations $(\tilde{w}_a, \tilde{w}_b)$ du test $TFNC^{SB}$ en fonction du paramètre de décentrage dec dans l'hypothèse $H_0 : \{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ pour $\alpha_r^* = 0.05$ et $n_1 = n_2 = 100$* 198

5.15 *Erreurs de type1 du test $TFNC^{SB}$ en fonction du paramètre de décentrage dec dans l'hypothèse $H_0 : \{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ pour $\alpha_r^* = 0.05$ et $n_1 = n_2 = 100$* 198

5.16 *Bornes $(S_{1,L}, S_{1,U})$ de la région de rejet de H_0 du test $TFNC^{SB}$ en fonction du paramètre de décentrage dec dans l'hypothèse $H_0 : \{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ pour $\alpha_r^* = 0.05$ et $n_1 = n_2 = 100$* 199

5.17 *Bornes $(S_{2,L}, S_{2,U})$ de la région de rejet de H_0 du test $TFNC^{SB}$ en fonction du paramètre de décentrage dec dans l'hypothèse $H_0 : \{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ pour $\alpha_r^* = 0.05$ et $n_1 = n_2 = 100$* 199

B.1 *Valeur de r sur la fonction $F_0(B)$ (—) si H_0 et H_1 sont symétriques (i.e. $c = 1$) et $(\cdot - \cdot)$ sinon (ex. $c > 1$), et, valeur de a sur la fonction $1 - F_1(B)$* 244

C.1 *Fonctions holomorphes* $g(\theta)$ *(- - -),* $h(\theta)$ *(—) dans* $\mathcal{Y}_{(1,y)}=\{y=y_1; y=1,2,..,5\}$ *et* $h(\theta)$ *(· · ·) dans* $\mathcal{Y}_{(2,y)}=\{y=y_1+y_2; y=0,1,..,5\}$ *pour le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$ 254

C.2 *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) dans* $\mathcal{Y}_{(1,y)}=\{y=y_1; y=4,5\}$ *et* $h(\theta)$ *(· · ·) dans* $\mathcal{Y}_{(2,y)}=\{y=y_1+y_2; y=0,1,..,8\}$ *pour le plan séquentiel* $(n_1=5, S_{1,U}=4, n_2=5)$ 254

C.3 *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) dans* $\mathcal{Y}_{(1,y)}=\{y=y_1; y=5\}$ *et* $h(\theta)$ *(· · ·) dans* $\mathcal{Y}_{(2,y)}=\{y=y_1+y_2; y=0,1,..,9\}$ *pour le plan séquentiel* $(n_1=5, S_{1,U}=5, n_2=5)$ 255

C.4 *Estimation de* $\hat{\theta}^{ModC}_{(m,y)}$ *(—),* $\hat{\theta}^{ModCA}_{(m,y)}$ *(- - -) et* $\hat{\theta}^{VUMSB}_{(m,y)}$ *(· − ·), soustraite de* $\hat{\theta}^{MV}_{(m,y)}$*, dans* $\mathcal{Y}_{(1,y)}=\{y=y_1; y=3,4,5\}$ *et* $\mathcal{Y}_{(2,y)}=\{y=y_1+y_2; y=0,1,..,7\}$ *pour le plan séquentiel* $(n_1=5, S_{1,U}=3, n_2=5)$ 256

C.5 *Estimation de* $\hat{\theta}^{ModC}_{(m,y)}$ *(—),* $\hat{\theta}^{ModCA}_{(m,y)}$ *(- - -) et* $\hat{\theta}^{VUMSB}_{(m,y)}$ *(· − ·), soustraite de* $\hat{\theta}^{MV}_{(m,y)}$*, dans* $\mathcal{Y}_{(1,y)}=\{y=y_1; y=4,5\}$ *et* $\mathcal{Y}_{(2,y)}=\{y=y_1+y_2; y=0,1,..,8\}$ *pour le plan séquentiel* $(n_1=5, S_{1,U}=4, n_2=5)$ 257

C.6 *Estimation de* $\hat{\theta}^{ModC}_{(m,y)}$ *(—),* $\hat{\theta}^{ModCA}_{(m,y)}$ *(- - -) et* $\hat{\theta}^{VUMSB}_{(m,y)}$ *(· − ·), soustraite de* $\hat{\theta}^{MV}_{(m,y)}$*, dans* $\mathcal{Y}_{(1,y)}=\{y=y_1; y=5\}$ *et* $\mathcal{Y}_{(2,y)}=\{y=y_1+y_2; y=0,1,..,9\}$ *pour le plan séquentiel* $(n_1=5, S_{1,U}=5, n_2=5)$ 258

C.7 *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{ModCA}$ *(- - -) et* $\hat{\theta}^{VUMSB}$ *(· − ·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=4, n_2=5)$ 259

C.8 *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{ModCA}$ *(- - -) et* $\hat{\theta}^{VUMSB}$ *(· − ·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=5, n_2=5)$ 259

C.9 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -)dans le plan séquentiel* $(n_1=5, S_{1,L}=-1, S_{1,U}=2, n_2=5)$ 260

C.10 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -) dans le plan séquentiel* $(n_1=5, S_{1,L}=2, S_{1,U}=5, n_2=5)$ 260

C.11 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -) dans le plan séquentiel* $(n_1=5, S_{1,L}=3, S_{1,U}=6, n_2=5)$ 261

C.12 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et de* $\hat{\theta}^{Whi}$ *(· − ·) dans le plan séquentiel* $(n_1=5, S_{1,U}=3, n_2=5)$ 261

C.13 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et de* $\hat{\theta}^{Whi}$ *(· – ·) dans le plan séquentiel* $(n_1=5, S_{1,U}=4, n_2=5)$ 262

C.14 *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et de* $\hat{\theta}^{Whi}$ *(· – ·) dans le plan séquentiel* $(n_1=5, S_{1,U}=5, n_2=5)$ 262

C.15 *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et* $\hat{\theta}^{Whi}$ *(· – ·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=3, n_2=5)$ 263

C.16 *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et* $\hat{\theta}^{Whi}$ *(· – ·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=4, n_2=5)$ 263

C.17 *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et* $\hat{\theta}^{Whi}$ *(· – ·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=5, n_2=5)$ 264

C.18 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,0) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$ 264

C.19 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,1) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$ 265

C.20 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,2) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$ 265

C.21 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,3) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$ 266

C.22 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,4) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$ 266

C.23 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,5) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$ 267

C.24 *Taux de recouvrement de l'intervalle de Wilson 90% dans le plan séquentiel* $(n_1=5, S_{1,U}=2, n_2=5)$ 268

C.25 *Taux de recouvrement de l'intervalle d'Agresti 90% dans le plan séquentiel* $(n_1=5, S_{1,U}=2, n_2=5)$ 268

C.26 *Taux de recouvrement de l'intervalle du second-ordre 90% dans le plan séquentiel* ($n_1=5, S_{1,U}=2, n_2=5$) 269

C.27 *Taux de recouvrement de l'intervalle de Clopper-Pearson 90% dans le plan séquentiel* ($n_1=5, S_{1,U}=2, n_2=5$) 269

C.28 *Longueur moyenne des intervalles de confiance 90% dans le plan séquentiel* ($n_1=5, S_{1,U}=2, n_2=5$). *Les intervalles* CI_J, CI_{JM} *et* CI_{CP} *(—) sont respectivement représentés de bas en haut, avec également les intervalles* CI_{JT} *(− − −),* CI_W *(· · ·),* CI_{AC} *(· − ·) et* CI_2 *(- - -).* 270

C.29 *Longueur moyenne des intervalles de Jeffreys modifiés 90% classique (- - -) et corrigé (—) dans le plan séquentiel* ($n_1=5, S_{1,U}=2, n_2=5$) . 270

C.30 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,1) dans le plan séquentiel* ($n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5$) 271

C.31 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,2) dans le plan séquentiel* ($n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5$) 271

C.32 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,3) dans le plan séquentiel* ($n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5$) 272

C.33 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,4) dans le plan séquentiel* ($n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5$) 272

C.34 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,0) dans le plan séquentiel* ($n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5$) 273

C.35 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,1) dans le plan séquentiel* ($n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5$) 273

C.36 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,2) dans le plan séquentiel* ($n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5$) 274

C.37 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,3) dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5$) 274

C.38 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,4) dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5$) 275

C.39 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,5) dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5$) 275

C.40 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,6) dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5$) 276

C.41 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,7) dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5$) 276

C.42 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,8) dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5$) 277

C.43 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,9) dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5$) 277

C.44 *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,10) dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5$) 278

C.45 *Longueur moyenne des intervalles de confiance 90% dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5$). *Les intervalles* CI_J, CI_{JM} *et* CI_{CP} *(—) sont respectivement représentés de bas en haut avec également les intervalles* CI_{JT} *(— — —),* CI_W *(· · ·),* CI_{AC} *(· — ·) et* CI_2 *(- - -).* 278

C.46 *Valeurs limites de* $Pr(RejetH_0|\theta_0 = dec)$ *(—) et* $Pr(RejetH_0|\theta_0 = dec{+}0.2)$ *(- - -) telles que* $Pr(RejetH_0|\theta_0 = dec) < Pr(RejetH_0|\theta_0 = dec{+}0.2)$ *en fonction du paramètre de décentrage dec dans l'hypothèse* $H_0 : \{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ *pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2 = 100$ . . . 280

C.47 *Valeurs limites de* $Pr(RejetH_0|\theta_0 = dec)$ *(—) et* $Pr(RejetH_0|\theta_0 = dec{+}0.2)$ *(- - -) telles que* $Pr(RejetH_0|\theta_0 = dec) > Pr(RejetH_0|\theta_0 = dec{+}0.2)$ *en fonction du paramètre de décentrage dec dans l'hypothèse* $H_0 : \{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ *pour* $\alpha_r^* = 0.05$ *et* $n_1{=}n_2{=}100$ . . . 280

C.48 *Pondérations limites* (w_a, w_b) *en fonction du paramètre de décentrage dec dans l'hypothèse* $H_0 : \{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ *avec* $Pr(RejetH_0|\theta_0 = dec) > Pr(RejetH_0|\theta_0 = dec{+}0.2)$ *(—) et* $Pr(RejetH_0|\theta_0 = dec) < Pr(RejetH_0|\theta_0 = dec{+}0.2)$ *(- - -) pour* $\alpha_r^* = 0.05$ *et* $n_1{=}n_2{=}100$ 281

C.49 *Bornes* $(S_{1,L}, S_{1,U})$ *de la région de rejet de* H_0 *en fonction du paramètre de décentrage dec dans l'hypothèse* $H_0 : \{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ *avec* $Pr(RejetH_0|\theta_0 = dec) > Pr(RejetH_0|\theta_0 = dec + 0.2)$ *(—) et* $Pr(RejetH_0|\theta_0 = dec) < Pr(RejetH_0|\theta_0 = dec + 0.2)$ *(- - -) pour* $\alpha_r^* = 0.05$ *et* $n_1{=}n_2{=}100$ 282

C.50 *Bornes* $(S_{2,L}, S_{2,U})$ *de la région de rejet de* H_0 *en fonction du paramètre de décentrage dec dans l'hypothèse* $H_0 : \{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ *avec* $Pr(RejetH_0|\theta_0 = dec) > Pr(RejetH_0|\theta_0 = dec + 0.2)$ *(—) et* $Pr(RejetH_0|\theta_0 = dec) < Pr(RejetH_0|\theta_0 = dec + 0.2)$ *(- - -) pour* $\alpha_r^* = 0.05$ *et* $n_1{=}n_2{=}100$ 282

Introduction

La détermination et la prise en compte dans l'analyse des facteurs ayant une influence sur l'observation constituent une préoccupation majeure dans les sciences expérimentales. Il est par exemple naturel de considérer que l'arrêt prématuré d'une expérience motivé par la nature des résultats affecte l'information sur le paramètre d'intérêt.

Lorsque le plan expérimental est prédéterminé, la performance des procédures fréquentistes se mesure en moyennant par les probabilités d'échantillonnage sur l'espace des observations. Dans un plan séquentiel, la notion de biais dans l'estimation ponctuelle permet d'envisager des mesures correctrices pour l'estimateur classique du maximum de vraisemblance. De même, les tests séquentiels fondés sur des fonctions de dépense des risques garantissent les risques nominaux α et β prédéfinis.

Dans l'approche bayésienne, la notion de biais est plus diffuse car l'inférence sur le paramètre est portée par la loi a posteriori que l'on obtient par une simple application du théorème de Bayes. Lorsque la loi a priori n'intègre aucune information sur les observations, la formulation classique de ce théorème est en accord avec le principe de vraisemblance (voir chapitre (1)). La loi a posteriori ne tient pas compte de la nature du plan expérimental car cette information n'est pas non plus contenue dans la vraisemblance.

En notant que le plan expérimental constitue une partie de l'évidence, R. de Cristofaro propose une reformulation du théorème de Bayes qui permet d'intégrer cette

information dans la loi a priori. En introduisant le principe d'impartialité du plan expérimental, il offre en plus une nouvelle légitimité à la règle de Jeffreys.

Le travail mené dans le cadre de cette thèse est d'abord une tentative pour apporter une réponse bayésienne cohérente au problème de la règle d'arrêt dans un plan expérimental. L'idée de base développée dans le chapitre (2) est de considérer que l'information sur le plan expérimental est contenue dans l'information de Fisher. A partir de la vraisemblance associée au plan d'échantillonnage, on dérive une famille de lois a priori dont fait partie l'"a priori de Jeffreys corrigé". Le cas de l'évaluation d'une proportion conduit à considérer la loi Bêta-J.

Dans le chapitre (3), une première application dans l'estimation ponctuelle permet d'établir un lien direct entre le paramétrage de la loi Bêta-J et celui de la loi Bêta dans l'échantillonnage fixe. La forme analytique d'un nouvel estimateur bayésien aux propriétés remarquables en terme de biais est obtenue dans un plan expérimental comportant un arrêt optionnel.

Dans le chapitre (4), on étudie l'effet de la règle d'arrêt sur le taux de recouvrement et la longueur moyenne des intervalles de confiance habituellement utilisés. La comparaison des performances avec l'intervalle de crédibilité basé sur l'a priori de Jeffreys corrigé met en évidence le mécanisme de compensation du biais induit par la correction.

Enfin, un test séquentiel dont la règle d'arrêt et de rejet de l'hypothèse nulle est fondée sur une valeur limite du facteur de Bayes est développé dans le chapitre (5). Ce nouveau test permet des interprétations de l'erreur en terme de probabilités bayésiennes et de risques fréquentistes conditionnels et non-conditionnels. La correction de l'a priori de Jeffreys permet en plus de dériver le facteur de Bayes corrigé qui constitue une mesure quantitative pertinente du rapport des évidences de deux hypothèses. Le cas d'une hypothèse nulle disjointe (i.e. $H_0 = H_0^a \cup H_0^b$) permet enfin d'utiliser toute la potentialité d'une procédure bayésienne pour optimiser les

propriétés fréquentistes du test. On définit un principe général pour minimiser les risques de type 1 pour chaque composante H_0^i.

Le lecteur trouvera en annexe trois propositions d'articles dérivés de cette thèse. Le premier présente les caractéristiques générales de l'a priori de Jeffreys corrigé. On y trouve une première démonstration de la correction du biais induit sur la loi a posteriori lorsque le plan séquentiel comprend un nombre maximum d'analyses $K > 2$. Le deuxième article concerne l'estimation ponctuelle et le troisième l'estimation par intervalle. On y retrouve certains éléments mentionnés respectivement dans les chapitres (3) et (4) de ce document, mais développés dans un cadre expérimental plus général avec des applications lorsque $K > 2$.

Chapitre 1

Principe de Vraisemblance

Le principe de vraisemblance est défini dans Birnbaum (1962) [11] puis reformalisé dans Berger et Wolpert (1988) [9]. Ce principe est d'abord décrit dans la section (1.2) pour ensuite aborder celui de la règle d'arrêt dans la section (1.3) qui en est une conséquence majeure. On montre dans la section (1.4) que l'approche bayésienne classique ne transgresse pas ces principes.

1.1 Évidence statistique

L'*évidence* permet de caractériser l'information qu'il est possible d'obtenir sur le paramètre étudié dans le cadre d'une *expérience*. Dans le modèle statistique général $(\mathcal{X}, P_\theta; \theta \in \Theta)$ où $\mathcal{X}$ est l'espace des observations, Θ l'espace des paramètres et P_θ une famille de loi de probabilité sur $\mathcal{X}$ indicée par $\theta \in \Theta$, la formalisation de l'évidence (Birnbaum (1962) [11]) met en jeu deux composantes : *l'expérience* caractérisée par le triplet $E = (\mathcal{X}, \Theta, p_\theta(x))$ et l'observation x qui prend ses valeurs dans $\mathcal{X}$.

L'évidence $Ev(E, x)$ se base exclusivement sur l'information en provenance de l'observation. Dans le contexte d'une expérience $E = \{\mathcal{X}, \Theta, p(x|\theta)\}$, l'inférence est la conclusion sur θ de l'*issue* (E, x). L'évidence définie donc l'ensemble des inférences sur θ.

Suivant le type d'inférence, l'évidence $Ev(E, x)$ se traduit par un ensembles d'estimations de θ, un ensemble de fonctions de vraisemblance en θ, ou encore des décisions de rejeter ou non une hypothèse suivant les valeurs de $x \in \mathcal{X}$: on parle alors d'une expérience binaire. Dans le contexte bayésien où des assertions probabilistes sont portées sur θ, $Ev(E, x)$ s'exprime par un ensemble de lois a posteriori sur ϑ. Dans la suite, on supposera que la fonction de probabilité $p_\theta(x)$ utilisée dans l'expérience est la loi d'échantillonnage $p(x|\theta)$.

Dans le modèle Binomial où la *va* Y correspond au nombre de succès de n observations binaires (succès, échec), l'évidence statistique associée aux réalisations de $Y = y$ est décrite par le doublet $Ev_{Bin} = (E_{Bin}, y)$ où $E_{Bin} = (\{0, 1, .., n\}, [0, 1], Bin(\theta))$ caractérise l'expérience. Par exemple, si la conclusion de l'expérience sur θ repose sur la fonction de vraisemblance, Ev_{Bin} est définie par les $n + 1$ fonctions en θ

$$L(\theta; y) = \binom{n}{y} \theta^y (1 - \theta)^{(n-y)}.$$

Trois conditions permettant de conclure à l'*équivalence mathématique* de deux évidences (i.e. $Ev(E, x) = Ev(E', y)$) sont présentées dans cette section. Chacune conduit à des interprétations qui disqualifient les procédures statistiques les plus couramment utilisées.

La première condition est reliée à la notion de statistique exhaustive.

Principe d'exhaustivité (PE). *Soit $E = (\mathcal{X}, \Theta, p(x|\theta))$ une expérience associée à l'observation x, si E' est une autre expérience dérivée de E pour laquelle chaque valeur de x dans E correspond à une valeur de la statistique exhaustive $T(x)$ pour θ dans E' alors $Ev(E, x) = Ev(E', T(x))$ $\forall x \in \mathcal{X}$.*

Le PE sous-entend que l'évidence associée à l'observation x est équivalente à celle associée à une statistique exhaustive T fonction de x. Le PE écarte de fait les procédures dépendantes des valeurs d'une statistique ancillaire (voir section (A.3)). Bien qu'un conditionnement sur une statistique ancillaire soit courant dans les procédures

fréquentistes, les fondements de l'inférence fréquentiste ne contredisent pas le PE. D'autre part, Howson et Urbach (1993) [34] montrent l'égalité des lois bayésiennes a posteriori $p(\theta|x) = p(\theta|T(x))$, concluant à l'équivalence des deux conditionnements dans l'inférence bayésienne.

Démonstration. D'après le théorème de Bayes,

$$p(x|\theta, t) = p(x|t)\frac{p(\theta|x,t)}{p(\theta|t)}. \tag{1.1.1}$$

Comme T est une statistique exhaustive pour θ, on a $p(x|\theta, t) = p(x|t)$; d'autre part comme t est déterminé par x, on a $p(\theta|x,t) = p(\theta|x)$. Ceci nous permet de réécrire l'expression (B.1.1) avec

$$p(x|t) = p(x|t)\frac{p(\theta|x)}{p(\theta|t)},$$

de laquelle on déduit $p(\theta|x) = p(\theta|t)$. □

La version faible du PE conclut à l'équivalence des évidences de deux expériences produisant des réalisations identiques de la statistique exhaustive.

Principe faible d'exhaustivité (PEF). *Soient $E = (\mathcal{X}, \Theta, p(x|\theta))$ une expérience et $T(x)$ une statistique exhaustive pour θ, si $T(x_1) = T(x_2)$ alors $Ev(E, x_1) = Ev(E, x_2)$.*

La deuxième condition est reliée au concept de conditionnement sur une expérience de référence. On suppose que l'expérience E^M est une expérience mixte, mélange de plusieurs expériences E_k. La probabilité associée à chaque observation dans E^M est égale à la probabilité d'observer d'abord l'indice k d'une expérience à partir d'une *va* K dont la loi de probabilité est connue et indépendante de θ, et, d'observer ensuite la valeur x_k dans l'expérience E_k. Dans ce cas, $(E^M, (k, x_k))$ est une écriture alternative de l'issue (E_k, x_k).

Principe de conditionnement (PC). *Si E^M est une expérience mixte, mélange de plusieurs expériences E_k sur le même espace de paramètre Θ, alors pour chaque*

observation (k, x_k) *de* E^M *on a*

$$Ev(E^M, (k, x_k)) = Ev(E_k, x_k).$$

En d'autres termes, l'évidence résultante d'une observation (k, x_k) de E^M est équivalente à celle résultante d'une observation x_k dans l'expérience E_k effectivement réalisée, ignorant de fait la structure de mélange des expériences. Le choix de l'expérience E_k parmi l'ensemble des expériences obéit à un processus ancillaire indépendant de θ, soulignant le rôle majeur de l'ancillarité dans le PC.

Par opposition à une statistique exhaustive T où la distribution d'une *va* X conditionnellement à T ne dépend pas de θ, le fait de baser l'inférence sur une distribution paramétrée par θ conditionnellement à la réalisation d'une statistique ancillaire permet de conserver l'intégrité de l'information sur θ.

L'inférence bayésienne sur θ s'effectue par la loi a posteriori $p(\theta|x)$. C'est donc une approche conditionnée à l'observation x. L'inférence fréquentiste est également une approche conditionnelle mais Fisher (1956) [31] suggère un conditionnement sur une statistique ancillaire pour ne pas dénaturer l'information sur θ qui est contenue dans la vraisemblance. Dans le modèle Binomial, la statistique ancillaire par rapport à θ est l'effectif n observé.

L'exemple ci-dessous, qui introduit le principe faible de conditionnement, souligne le manque de pertinence de l'expérience non réalisée. Considérons deux expériences, $E_1 = (X_1, \Theta, p_1(x_1|\theta))$ et $E_2 = (X_2, \Theta, p_2(x_2|\theta))$, et une *va* K telle que $Pr(K=1) = Pr(K=2) = \frac{1}{2}$ quelle que soient les valeurs de θ, x_1 et x_2. On peut imaginer le lancé {pile,face} d'une pièce de monnaie. L'expérience mixte consiste à procéder à E_1 si $k = 1$ ou à E_2 si $k = 2$. Elle s'écrit comme le triplet $E^M = (\mathcal{X}^{\mathcal{M}}, \Theta, p^M((k, x_k)|\theta))$ où $\mathcal{X}^{\mathcal{M}} = (1, 2) \times (\mathcal{X} \cup \mathcal{X})$, et

$$p^M((k, x_k)|\theta) = \begin{cases} \frac{1}{2}\, p_1(x_1|\theta) & \text{si } k = 1 \\ \frac{1}{2}\, p_2(x_2|\theta) & \text{si } k = 2 \end{cases}$$

On note une nouvelle fois l'indépendance par rapport à θ du mécanisme ancillaire déterminant le choix de l'expérience E_1 ou E_2. La version faible du principe de conditionnement suffit pour admettre que l'évidence de E^M sur θ correspond à l'évidence de l'expérience, E_1 ou E_2, effectivement réalisée.

Principe faible de conditionnement. *Si E^M est une expérience mixte, mélange de deux expériences $E_1 = (X_1, \theta, p_1(x_1|\theta))$ et $E_2 = (X_2, \theta, p_2(x_2|\theta))$ sur le même espace de paramètre Θ avec chacune une même probabilité de survenue (indépendante de θ, x_1 et x_2), alors*

$$Ev(E, (k, x_k)) = Ev(E_k, x_k).$$

1.2 Principe de Vraisemblance

La fonction de vraisemblance $L(\theta; x)$ relie l'information préexpérimentale exprimée dans le choix du modèle statistique, à celle fournie par l'observation x. Certains statisticiens considèrent que $L(\theta; x)$ contient l'ensemble de l'information nécessaire pour l'inférence sur θ. Il acceptent en cela une conséquence majeur du principe de vraisemblance qui constitue une troisième condition pour conclure à l'équivalence de deux évidences.

Principe de vraisemblance (PV). *Soient $E_1 = (\mathcal{X}, \Theta, p_1(x_1|\theta))$ et $E_2 = (\mathcal{X}, \Theta, p_2(x_2|\theta))$ deux expériences sur le même espace de paramètre Θ, si il existe deux observations $X_1 = x_1^*$ et $X_2 = x_2^*$ ($x_1^* \in E_1$ et $x_2^* \in E_2$) satisfaisant l'égalité*

$$p_1(x_1^*|\theta)) = cp_2(x_2^*|\theta)$$

$$(i.e.\ L_1(\theta; x_1^*) = cL_2(\theta; x_2^*))$$

avec $c = c(x_1, x_2)$ pour tout θ, alors

$$Ev(E_1, x_1^*) = Ev(E_2, x_2^*).$$

Le PC et le PE sont généralement bien acceptés. Il y a en revanche plus de réticence à accepter le PV. Pourtant (Birnbaum (1962) [11]) démontre que le PV est une

conséquence directe du PFC et du PFE (preuve reformalisée dans Berger (1988) [9]).

Démonstration. Considérons l'expérience mixte $E^M = (\mathcal{X}^{\mathcal{M}}, \Theta, p^M(x|\theta))$, on définit une statistique $T : \mathcal{X}^{\mathcal{M}} \mapsto \mathcal{X}^{\mathcal{M}}$ telle que

$$T((k, x_k)) = \begin{cases} (1, x_1^*) & \text{si } j = 2 \text{ et } x_j = x_2^* \\ (i, x_i) & \text{sinon.} \end{cases}$$

La statistique T est donc à valeur dans E_1 et E_2. Pour les deux observations $x_1^* \in E_1$ et $x_2^* \in E_2$, on a la relation $T(x_1^*) = T(x_2^*)$ en plus de la supposition initiale $L_1(\theta; x_1^*) = cL_2(\theta; x_2^*)$; sinon T laisse les autres $x_j \in \mathcal{X}_j$ inchangés. Pour montrer que T est une statistique exhaustive, on doit établir que la loi de probabilité de x^* conditionnellement à une valeur $T(x^*) = t$ ne dépend pas de θ.

Pour $t \neq (1, x_1^*)$, on a

$$Pr(X^* = (i, x_i) \,|\, T(x^*) \neq (1, x_1^*)) = \begin{cases} 1 & \text{si } t = (j, x_j) \\ 0 & \text{si } t \neq (j, x_j). \end{cases}$$

En utilisant les égalités $Pr(E_k = E_1) = Pr(E_k = E_2) = \frac{1}{2}$ et $p_1(x_1^*|\theta)) = cp_2(x_2^*|\theta)$, lorsque $t = (1, x_1^*)$ on obtient

$$\begin{aligned} Pr(X^* = (i, x_i) \,|\, T(x^*) = (1, x_1^*)) &= 1 - Pr(x^* = (2, x_2) \,|\, T(x^*) = (1, x_1^*)) \\ &= \frac{\frac{1}{2} p_1(x_1^*|\theta)}{\frac{1}{2} p_1(x_1^*|\theta) + \frac{1}{2} p_2(x_2^*|\theta)} \\ &= \frac{c}{c+1} \end{aligned}$$

En récapitulant dans le système

$$Pr(X^* = (i, x_i) \,|\, T(x^*) = t) = \begin{cases} \frac{c}{c+1} & \text{si } t = (1, x_1^*) \text{ et } j = 1 \text{ et } x_j = x_1^* \\ \frac{1}{c+1} & \text{si } t = (1, x_1^*) \text{ et } j = 2 \text{ et } x_j = x_2^* \\ 1 & \text{si } t \neq (1, x_1^*) \text{ et } t = (j, x_j) \\ 0 & \text{si } t \neq (1, x_1^*) \text{ et } t \neq (j, x_j), \end{cases}$$

il vient que T est exhaustive car $Pr(X^* = (i, x_i) \,|\, T(x^*) = t)$ ne dépend pas de θ.

On démontre que $PFE + PFC \longrightarrow PV$ en justifiant les égalités suivantes :

$$\begin{aligned}
Ev(E_1, x_1^*) &= Ev(E^*, (1, x_1^*)) && \text{par le } PFC \\
&= Ev(E^*, T(1, x_1^*)) \\
&= Ev(E^*, T(2, x_2^*)) && \text{par le } PFE \\
&= Ev(E^*, (2, x_2^*)) \\
&= Ev(E_2, x_2^*) && \text{par le } PFC
\end{aligned}$$

□

On déduit du PV une autre assertion que l'on qualifie de version faible du PV ; deux ensembles d'observations obtenues dans une expérience ou deux expériences différentes et qui produisent des vraisemblances proportionnellement identiques sur le même espace de paramètre Θ conduisent à la même inférence sur θ. Une conséquence importante est que l'information provenant de l'observation x est entièrement contenue dans la vraisemblance $L(\theta; x)$ qui reste la seule source d'information utilisée pour l'inférence sur θ.

Cette interprétation est par définition compatible avec l'inférence fiduciaire (voir section (B.1)). Elle l'est également avec l'inférence bayésienne puisque la densité a posteriori $p(\theta|x)$ se dérive à partir de la densité a priori $p(\theta)$, indépendante de x, et de la fonction de vraisemblance $L(\theta; x)$ qui reste donc l'unique source d'information provenant de l'observation.

Certains fondements de l'inférence fréquentiste ne sont pas compatibles avec le PV car les procédures utilisées nécessitent une intégration sur l'espace des observations $\mathcal{X}$, avec une pondération $p(x|\theta)$ donnée par des valeurs x autres que x^* observée. Les propriétés de ces procédures, comme la mesure du biais dans l'estimation de θ, impliquent également une telle intégration.

En revanche, la prédétermination du plan d'échantillonnage sur laquelle repose l'inférence fréquentiste n'est pas en contradiction avec le PV mais cette informa-

tion ne doit pas être utilisée dans l'inférence sur θ. Dans le modèle Binomial avec $Y \sim Bin(n, \theta)$, cette information est contenue dans le facteur combinatoire $\binom{n}{y}$ qui n'apparaît pas dans la vraisemblance $L(\theta; y)$.

1.3 Principe de la règle d'arrêt

Le principe de la règle d'arrêt est une conséquence directe du PV. Il met en avant la notion d'impartialité de la règle d'arrêt de l'expérience dans l'inférence sur θ.

Si l'on note $X_j \sim p(x|\theta)$ une *va* associée à l'expérience $E = (\mathcal{X}, \Theta, p(x|\theta))$, on peut définir $\mathcal{X}^j = \prod_{k=1}^{j} \mathcal{X}^k$ l'espace des observations associé au j-uplet $X^j = (X_1, \ldots, X_j)$. Avec la condition d'indépendance, la probabilité jointe d'échantillonnage de X^j s'écrit

$$p(x^j|\theta) = \prod_{k=1}^{j} p(x_k|\theta).$$

Pour tout $m \geq 1$, on peut également définir une expérience

$$E^m = (\mathcal{X}^m, \Theta, p(x^m|\theta)),$$

elle même composée de m expériences successives avec un effectif prédéfini pour chacune.

Définition. *Une règle d'arrêt aléatoire $\tau = \{\tau^1, \ldots, \tau^m\}$ définie sur une expérience E^m est une séquence de fonctions $\tau^k : \mathcal{X}^k \to [0, 1]$ qui caractérise la séquentialité du plan expérimental. A chaque période m, l'arrêt de l'expérience est associée à une probabilité τ^m, sinon une nouvelle observation $X_{m+1} \sim p(x|\theta)$ est effectuée en période $m + 1$.*

Une règle d'arrêt est dite propre si l'expérience s'arrête presque sûrement. Elle est dite non-aléatoire si $\tau^k \in \{0, 1\}$ pour chaque k.

Sur la base d'une règle d'arrêt propre, on peut définir une expérience séquentielle

$$E^\tau = (\mathcal{X}^\tau, \Theta, p(x^\tau|\theta))$$

avec

$$\begin{aligned}\mathcal{X}^\tau &= N \times \cup_{k=0}^{\infty} \mathcal{X}^k \\ p^\tau((m, x^m|\theta)) &= \prod_{k=0}^{m-1} (1 - \tau^k(x^k))\tau^m(x^m) \prod_{k=1}^{m} p(x^k|\theta).\end{aligned}$$

La règle d'arrêt

$$\tau^k(x^k) = \begin{cases} 0 & \text{si } k < m \\ 1 & \text{si } k \geq m \end{cases}$$

conduit à l'égalité des expériences $E^\tau = E^m$.

Le principe de la règle d'arrêt stipule que l'observation recueillie lors d'une expérience avec une règle d'arrêt fournit la même évidence que la même observation recueillie dans le contexte d'une expérience sans règle d'arrêt.

Principe de la règle d'arrêt (PRA). *Pour toute règle d'arrêt τ définie sur une expérience $E^m = (\mathcal{X}^m, \Theta, p(x^m|\theta))$*

$$Ev(E^\tau, (m, x^m)) = Ev(E^m, x^m).$$

On démontre qu'une procédure en accord avec le PV l'est également avec le PRA car il y a proportionnalité des vraisemblances dans les deux cas :

$$\begin{aligned}p^\tau((m, x^m|\theta)) &= \prod_{k=0}^{m-1} (1 - \tau^k(x^k))\tau^m(x^m) \prod_{k=1}^{m} p(x^k|\theta) \\ &\propto \prod_{k=1}^{m} p(x^k|\theta) \\ &= p^m((m, x^m|\theta)).\end{aligned}$$

L'existence d'une règle d'arrêt lorsque le plan expérimental prévoit des analyses successives a des conséquences différentes dans les approches fréquentistes et bayésiennes. Sur la base d'une règle d'arrêt prédéterminée, l'inférence fréquentiste impose

des ajustements ou des pénalités pour palier les biais d'estimation ou l'évaluation des risques d'une procédure de test. C'est une transgression du PRA car l'inférence sur θ dépend de cette règle d'arrêt.

1.4 Echantillonnage séquentiel dans l'approche Bayésienne classique

Dans l'approche bayésienne classique (i.e. résultante de l'application du théorème de Bayes formulé dans (B.1.1), l'information apportée par les données est exclusivement contenue dans la vraisemblance. Le PV n'est pas transgressé tant que la détermination de la loi a priori s'effectue indépendamment de la nature des données et de la manière dont elles sont recueillies. Si plusieurs expériences indépendantes sont menées, l'information globale sur θ se retrouve dans la vraisemblance jointe. La nature de l'inférence est ainsi redéfinie en fonction des observations et indépendamment d'une règle d'arrêt de l'expérience, prédéfinie ou non. Les problèmes de biais d'estimation ne sont pas considérés comme rédhibitoires car l'inférence sur θ est fondée sur les caractéristiques de la loi a posteriori toute entière.

Dans le modèle Binomial bi-dimensionnel $(\{0,1,\ldots,n_1\}\times\{0,1,\ldots,n_2\}, \prod_{i=1}^{2} Bin(n_i,\theta); \theta \in [0,1])$ où les résultats sont associés aux *va* $Y_1 \sim Bin(n_1,\theta)$ et $Y_2 \sim Bin(n_2,\theta)$, la vraisemblance jointe s'écrit

$$L(\theta; y_1, y_2) = L(\theta; y_1)L(\theta; y_2) = \binom{n_1}{y_1}\theta^{y_1}(1-\theta)^{n_1-y_1}\binom{n_2}{y_2}\theta^{y_2}(1-\theta)^{n_2-y_2}$$

qui est proportionnelle à $\theta^y(1-\theta)^{n-y}$ où $n = n_1+n_2$ et $y = y_1+y_2$. La vraisemblance $L(\theta; y_1, y_2)$ s'incrémente séquentiellement et indépendamment de l'ordre avec lequel les expériences sont réalisées. La nature de l'inférence est ainsi redéfinie en fonction de l'accumulation de l'évidence.

La multiplication des vraisemblances permet d'écrire

$$r(\theta) \propto r_1(\theta)r_2(\theta)$$

où la vraisemblance relative $r(\theta)$ (voir (B.3)) résultante des deux expériences est proportionnelle au produit des vraisemblances relative pour chacune.

L'approche bayésienne est compatible avec ce principe de réajustement de l'information sur θ en fonction de l'accumulation de l'évidence. Si l'on considère une observation $x = \{x_1, x_2\}$ sur deux échantillons indépendants, on a

$$\begin{aligned} p(\theta|x) &\propto p(\theta)p(x|\theta) \\ &= p(\theta)p(x_1, x_2|\theta) \\ &= p(\theta)p(x_1|\theta)p(x_2|\theta) \end{aligned}$$

Comme $p(\theta|x_1) \propto p(\theta)p(x_1|\theta)$, on trouve finalement

$$p(\theta|x) \propto p(\theta|x_1)p(x_2|\theta).$$

La loi a posteriori résultante de la première expérience prend le statut de loi a priori dans le modèle bayésien pour la deuxième expérience.

Dans le modèle Binomial bi-dimensionnel, la loi a posteriori de θ s'écrit

$$\vartheta/y = \vartheta/y_1, y_2 \sim Be(y_1 + y_2 + 1, n_1 + n_2 - y_1 - y_2 + 1) = Be(y + 1, n - y + 1).$$

Plus généralement, la conjugaison de la loi a priori $Be(a, b)$ avec une observation Binomiale (n,y) se traduit par un a posteriori du type $\vartheta/y \sim Be(a + y, b + n - y)$, offrant une interprétation implicite des coefficients a, b. La loi a priori $Be(a, b)$ correspond à la loi a posteriori résultant d'une observation Binomiale antérieure, imaginaire ou non, qui aurait générée $a-1$ succès sur un effectif de $a+b-2$ données dans l'échantillon. Dans cette perspective, l'a priori Uniforme $\vartheta \sim Be(1, 1)$ traduit une absence complète de telles données imaginaires.

Chapitre 2

Loi a priori

2.1 Loi a priori et principe de vraisemblance

2.1.1 A priori informatif

L'inférence bayésienne pose en premier lieu le problème du choix de la loi a priori qui caractérise l'incertitude sur les valeurs du paramètre avant l'expérience. Un a priori informatif peut se construire le plus objectivement possible sur la base de données historiques ou bien en introduisant une part de subjectivité. Dans ce dernier cas, l'évidence vient corriger un excès d'optimisme ou de pessimisme de l'a priori, rejoignant ainsi la démarche du test d'hypothèses (voir Freedman et Spiegelhalter (1983) [33] et Berry (1985) [3]).

Ce positionnement de l'évidence est critiquée dans Whitehead (1993) [55] car on attribut à l'a priori un rôle habituellement imparti à une fonction de risque. De plus, a priori subjectif et évidence sont étroitement associés à travers le degré de confirmation ou d'infirmation que l'on veut attribuer à l'observation. Par exemple dans le modèle Binomial avec $Y \sim Bin(n, \theta)$, les paramètres de la loi a priori $\vartheta \sim Be(a, b)$ s'interprètent directement en terme d'effectif d'une expérience précédente fictive (voir section (1.4)). Suivant les pondérations données à l'a priori et à l'évidence dans la construction de la loi a posteriori, on ajuste l'effectif "préexpérimental" des pseudo-

observations en fonction de l'effectif à observer. Ce type de conditionnement sur l'effectif se retrouve dans les procédures fréquentistes avec le calcul préexpérimental de la puissance d'un test. L'utilisation d'une loi bayésienne a priori informative ou bien la justification de la taille de l'échantillon à observer dans une procédure fréquentiste sont une transgression du PV car des éléments de l'information future sont utilisés dans une perspective préexpérimentale.

Enfin, une loi a priori est souvent choisie telle que sa *conjugaison* avec la vraisemblance produise une loi a posteriori de la même famille. Ce choix de l'a priori en fonction du type d'échantillonnage enfreint le PV une nouvelle fois.

2.1.2 A priori non-informatif

L'inexistence d'un a priori traduisant un état d'ignorance absolu sur les valeurs du paramètre fait l'objet d'un consensus (Box et Tiao (1992)[12]). D'autre part, l'idée d'une connaissance à minima n'a de sens que par rapport à l'information qui sera apportée par l'expérience, ce qui conduit à rejeter le PV (cette assertion est rediscutée section 2.2).

Dans le modèle Binomial avec $Y \sim Bin(n, \theta)$, l'a priori Uniforme $\vartheta \sim Be(1, 1)$ fournit une information semblable à celle d'une expérience fictive d'effectif nul. L'uniformité de la densité sur l'intervalle $\Theta = [0, 1]$ satisfait la propriété de domination de la vraisemblance et - par définition - la propriété d'uniformité locale. Cette dernière propriété sous-entend que la densité de la loi a priori ne change pas brutalement pour les valeurs de θ correspondant à une vraisemblance $L(\theta)$ élevée et se maintient à un niveau faible en dehors de ces valeurs.

Cependant, si une paramétrisation plus pertinente du modèle nécessite une transformation $h(\theta)$ du type $1/\theta$, $log(\theta)$, $logit(\theta)$,..., l'a priori $\vartheta \sim Be(1, 1)$ ne conserve

plus la propriété d'uniformité locale pour la *va* transformée $h(\vartheta)$, car sa densité est

$$p(h(\theta)) = p(\theta) \left| \frac{\partial \theta}{\partial\ h(\theta)} \right|.$$

En d'autre terme, l'inférence sur $h(\vartheta)$ n'est plus de la même nature que celle sur ϑ, ce qui traduit un problème d'invariance.

Se pose alors le problème du choix de la transformation (échelle ou métrique) dans laquelle la propriété d'uniformité locale peut être satisfaite.

Le principe suivant constitue un critère majeur dans ce choix. Il démontre en plus toute l'ambiguïté qu'il y a à considérer un a priori Uniforme ou localement uniforme comme non-informatif lorsque la forme de la vraisemblance $L(\theta; x)$ dépend de l'observation x.

Principe de translation de la vraisemblance (PTV). *Sous la supposition que la vraisemblance ne dépend de l'observation x que par l'intermédiaire d'une statistique exhaustive T (i.e. $L(\theta; x) = L(\theta; T(x))$; si l'impact de l'observation se traduit uniquement par une translation en θ de la vraisemblance, alors toute l'information provenant de l'observation est contenue dans cette translation.*

Lorsque l'apport de l'observation se traduit uniquement par une translation de la vraisemblance sur l'espace Θ, le pouvoir discriminant de $L(\theta; x)$ entre θ et $\theta + dt$ reste constant à la translation prés. Une loi a priori Uniforme exprime <u>dans ce cas</u> un état d'ignorance préalable.

Une vraisemblance en θ translatée par $T(x)$ exhaustive peut s'écrire

$$L(\theta; x) = L(\theta; T(x)) = g(\theta - T(x)).$$

Certaines vraisemblances sont naturellement translatées car leurs formes ne dépend pas de l'observation (c'est le cas d'un échantillonnage Normal avec une variance constante). En revanche la vraisemblance $L(\theta; y) = \theta^y (1-\theta)^{n-y}$ associée à $Y \sim Bin(n, \theta)$ ne peut pas prendre la forme $g(\theta - y/n)$.

Il convient alors de déterminer une métrique Ψ dans laquelle la vraisemblance en $\Psi(\theta)$ est translatée. On obtient la forme

$$L(\theta; x) = L(\theta; T(x)) = g(\Psi(\theta) - T(x)).$$

Par référence au PTV, c'est dans cette métrique que l'uniformité d'un a priori prend tout son sens.

Un développement limité d'ordre 2 en $\hat{\theta}^{MV}_{(x)}$ de la vraisemblance permet d'obtenir la métrique par approximation. En se rappelant que

$$\left(\frac{\partial}{\partial\theta} LogL(\theta; x)\right)_{\hat{\theta}^{MV}_{(x)}} = 0,$$

la vraisemblance est approchée par

$$\begin{aligned} Log(L(\theta; x)) &= Log(\prod_{i=1}^{n} L(\theta; x_i)) \\ &\simeq Log(L(\hat{\theta}^{MV}_{(x)}; x)) - \frac{n}{2}(\theta - \hat{\theta}^{MV}_{(x)})^2 \left(-\frac{1}{n}\frac{\partial^2}{\partial\theta^2} LogL\right)_{\hat{\theta}^{MV}_{(x)}} \end{aligned}$$

Si $\left(-\frac{1}{n}\frac{\partial^2}{\partial\theta^2} LogL\right)_{\hat{\theta}^{MV}_{(x)}}$ est une constante, $Log(L(\theta; x))$ se comporte comme une combinaison linéaire de $(\theta - \hat{\theta}^{MV}_{(x)})^2$ avec des coefficients dépendants de l'observation x. La vraisemblance est donc approximativement translatée localement, pour les valeurs de θ aux alentours de $\hat{\theta}^{MV}_{(x)}$.

Dans le cas contraire, on détermine la métrique Ψ pour que

$$\left(-\frac{1}{n}\frac{\partial^2}{\partial\Psi^2} LogL\right)_{\Psi(\hat{\theta}^{MV}_{(x)})} = \text{constante}. \tag{2.1.1}$$

Si cette condition est satisfaite, la vraisemblance est translatée en $\Psi(\theta)$.

En posant

$$\mathfrak{J}(\hat{\theta}^{MV}_{(x)}) = \left(-\frac{1}{n}\frac{\partial^2}{\partial\theta^2} LogL\right)_{\hat{\theta}^{MV}_{(x)}},$$

on explicite la relation entre $\mathfrak{J}(\hat{\theta}^{MV}_{(x)})$ et $\mathfrak{J}(\Psi(\hat{\theta}^{MV}_{(x)}))$:

$$\begin{aligned}
\mathfrak{J}(\Psi(\hat{\theta}^{MV}_{(x)})) &= \left(-\frac{1}{n}\frac{\partial^2}{\partial\Psi^2}LogL\right)_{\Psi(\hat{\theta}^{MV}_{(x)})} \\
&= \left(-\frac{1}{n}\frac{\partial^2}{\partial\theta^2}LogL\right)_{\hat{\theta}^{MV}_{(x)}} \left(\frac{\partial\theta}{\partial\Psi}\right)^2_{\hat{\theta}^{MV}_{(x)}} \\
&= \mathfrak{J}(\hat{\theta}^{MV}_{(x)}) \left(\frac{\partial\theta}{\partial\Psi}\right)^2_{\hat{\theta}^{MV}_{(x)}}
\end{aligned}$$

Pour maintenir $\mathfrak{J}(\Psi(\hat{\theta}^{MV}_{(x)}))$ constante, on choisit donc Ψ telle que

$$\left|\frac{\partial\theta}{\partial\Psi}\right|_{\hat{\theta}^{MV}_{(x)}} \propto \mathfrak{J}^{-\frac{1}{2}}(\hat{\theta}^{MV}_{(x)}).$$

Cette condition en $\hat{\theta}^{MV}_{(x)}$ est forcément satisfaite si $\forall\,\theta \in \Theta$

$$\left|\frac{\partial\theta}{\partial\Psi}\right| \propto \mathfrak{J}^{-\frac{1}{2}}(\theta).$$

Ensuite par intégration on obtient la métrique

$$\Psi \propto \int^{\theta} \mathfrak{J}^{\frac{1}{2}}(t)dt$$

dans laquelle la vraisemblance est approximativement translatée en $\Psi(\theta)$ pour des valeurs autour de $\Psi(\hat{\theta}^{MV}_{(x)})$.

On déduit la densité a priori de θ dans son échelle d'origine par

$$p(\theta) \propto \left|\frac{d\Psi}{d\theta}\right| \propto \mathfrak{J}^{\frac{1}{2}}(\theta).$$

Dans le modèle Binomial avec $Y \sim Bin(n,\theta)$, on trouve successivement

$$\begin{aligned}
\mathfrak{J}(\hat{\theta}^{MV}_{(y)}) &= \left(-\frac{1}{n}\frac{\partial^2}{\partial\Psi^2}LogL\right)_{\hat{\theta}^{MV}_{(y)}} = \frac{1}{\hat{\theta}^{MV}_{(y)}(1-\hat{\theta}^{MV}_{(y)})} = \frac{1}{\frac{y}{n}(1-\frac{y}{n})} \\
p(\theta) &\propto \mathfrak{J}^{\frac{1}{2}}(\theta) \propto \theta^{\frac{1}{2}}(1-\theta)^{\frac{1}{2}} \propto Be(\frac{1}{2},\frac{1}{2}) \\
\Psi &= \int^{\theta} t^{-\frac{1}{2}}(1-t)^{-\frac{1}{2}}dt \propto sin^{-1}\sqrt{\theta}.
\end{aligned}$$

La loi a priori Uniforme satisfait les conditions du PTV si elle est associée à la métrique de référence $sin^{-1}\sqrt{\theta}$ dans laquelle la vraisemblance est approximativement translatée.

D'une manière générale, $\mathfrak{J}(\hat{\theta}^{MV}_{(x)})$ est uniquement fonction de $\hat{\theta}^{MV}_{(x)}$ si le modèle statistique est un modèle exponentiel où la probabilité d'échantillonnage est du type

$$p(x|\theta) = h(x)w(\theta)exp[c(\theta)u(x)].$$

Sinon, si $\mathfrak{J}(\hat{\theta}^{MV}_{(x)})$ est fonction de l'observation tout entière et non plus seulement de $\hat{\theta}^{MV}_{(x)}$, la métrique Ψ n'est plus fondée sur $\mathfrak{J}(\hat{\theta}^{MV}_{(x)})$ mais sur l'information de Fisher $I(\theta|x)$ qui l'approxime localement (voir démonstration section A.4). Cette substitution conduit à la relation

$$\left|\frac{\partial\Psi}{\partial\theta}\right| \propto I^{\frac{1}{2}}(\theta|x)$$

qui permet la détermination de Ψ par

$$\Psi \propto \int^{\theta} I^{\frac{1}{2}}(t)dt.$$

On déduit la densité a priori de θ dans son échelle d'origine

$$p(\theta) \propto \left|\frac{d\Psi}{d\theta}\right| \propto I^{\frac{1}{2}}(\theta|x),$$

avec la propriété que pour toute transformation h sur θ la densité a priori est

$$p(h(\theta)) \propto I^{\frac{1}{2}}(h(\theta)|x).$$

Cet argument d'invariance est le fondement de la règle de Jeffreys [36].

Règle de Jeffreys. *Une loi a priori sur θ est approximativement non-informative si*

$$p(\theta) \propto I^{\frac{1}{2}}(\theta|x).$$

Si $\mathbf{x} = \{x_1, ..., x_n\}$ est un échantillon constitué de n observations indépendantes l'a priori de Jeffreys est proportionnelle à l'information de Fisher pour une observation x parmi les x_i

$$p(\theta) \propto I^{\frac{1}{2}}(\theta|\mathbf{x}) \propto I^{\frac{1}{2}}(\theta|x).$$

Les propriétés de l'information de Fisher (voir Chap A.3) conduisent à l'interprétation que la règle de Jeffreys favorise a priori les valeurs de θ où l'information en provenance du modèle est la plus forte. Dans le modèle Binomial, la densité en U de la loi a priori $\theta \sim Be(\frac{1}{2}, \frac{1}{2})$ est une mesure du degré de certitude d'une prévision basée sur les informations disponibles pour θ.

On introduit maintenant une propriété remarquable de l'a priori de Jeffreys en terme d'interprétation. La loi Binomiale Négative se construit à partir de la probabilité $\theta^y(1-\theta)^z$ associée à y succès et z échecs. Si l'on décide d'observer le nombre z d'échecs intervenus avant le y^{ieme} succès, le nombre de combinaisons possibles est $\binom{y+z-1}{z}$ car les z échecs associés aux $y-1$ succès précédents peuvent intervenir dans n'importe quel ordre.

La probabilité associée à une observation de la va $Z \sim NBin(y, \theta)$ est

$$p(z|\theta) = \binom{y+z-1}{z} \theta^y (1-\theta)^z.$$

A partir de la dérivée seconde de la log-vraisemblance et de l'espérance $E(Z) = y(1-\theta)/\theta$, on obtient successivement l'information de Fisher

$$I(\theta|z) = \frac{y}{\theta^2(1-\theta)},$$

l'a priori de Jeffreys $\theta \sim Be(0, \frac{1}{2})$ avec

$$p(\theta) \propto I^{\frac{1}{2}}(\theta|z) \propto \theta^{-1}(1-\theta)^{-\frac{1}{2}},$$

et la métrique

$$\Psi(\theta) = log\left(\frac{1-\sqrt{1-\theta}}{1+\sqrt{1-\theta}}\right)$$

dans laquelle la vraisemblance est approximativement et localement translatée.

L'a priori de Jeffreys $Be(\frac{1}{2}, \frac{1}{2})$ obtenu dans le modèle Binomial est caractérisé par la courbe symétrique en U de sa densité. La figure (2.1) montre que la loi a priori $Be(0, 0.5)$ dans le modèle Binomial Négatif vient charger les valeurs de θ proches de

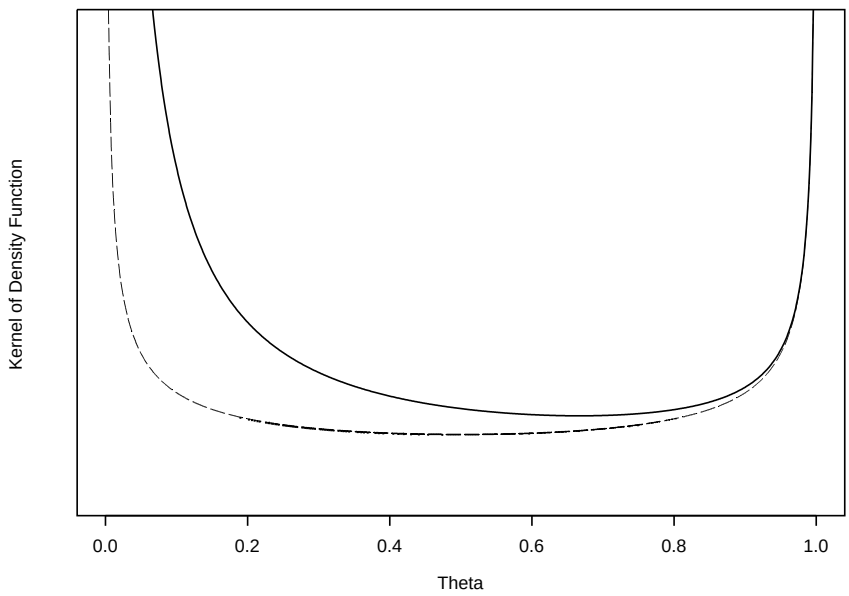

FIG. 2.1 – *Noyau de l'a priori de Jeffeys pour un échantillonnage Binomial (– –) et de l'a priori de Jeffeys (impropre) pour un échantillonnage Binomial Négatif (—)*

0, offrant ainsi une correction du biais causé par la surestimation du taux de succès après arrêt de l'expérience sur un cas favorable (voir de Cristofaro (2005) [27]).

L'EMV $\hat{\theta}^{MV} = Y/n$ est un estimateur sans biais dans l'échantillonnage Binomial (i.e. $E_\theta(\frac{Y}{n}) = \theta$). Le biais induit par l'MV dans l'échantillonnage Binomial Négatif (également appelé échantillonnage de Pascal) se dérive analytiquement lorsque l'on fixe la valeur $y = 1$ dans $Z \sim NBin(y, \theta)$. La fonction de vraisemblance se réduit à $\theta(1-\theta)^{n-1}$. La *va* N associée aux nombre d'observations nécessaires pour obtenir un succès est distribuée suivant une distribution géométrique. L'espérance de $\hat{\theta}^{MV} = y/N = 1/N$ devient

$$\begin{aligned} E_\theta(\frac{1}{N}) &= \sum_{n=1}^{\infty} \frac{1}{n}\theta(1-\theta)^{n-1} \\ &= \frac{\theta}{1-\theta} Log\frac{1}{\theta} > \theta \end{aligned} \tag{2.1.2}$$

La critique principale à l'encontre de la règle de Jeffreys tient dans sa transgression

du PV car la loi est déterminée sur la base des caractéristiques du plan d'échantillonnage. L'information sur θ des observations futures n'est plus exclusivement contenue dans la vraisemblance. On voit dans la section (2.2) que cette critique est liée à une formulation incomplète du théorème de Bayes lorsque les caractéristiques du plan expérimental sont connues avant le recueil des données.

2.2 Prise en compte du plan d'expérimental

Les méthodes statistiques compatibles avec le PV n'intègrent pas l'information sur le plan expérimental dans l'inférence sur θ. Lorsqu'une la loi Bayésienne a priori contient ce type d'information, la lecture habituelle du théorème de Bayes conduit à une transgression du PV.

A l'inverse, on peut souligner l'incohérence à ne considérer que la vraisemblance dans l'inférence sur θ alors que l'information sur un plan expérimental défini avant le recueil des données constitue une partie de l'évidence statistique. Cette attitude n'est pas non plus conforme avec le PV.

Si l'on conçoit l'évidence $e = Ev(E, x)$ d'une expérience $E == \{\mathcal{X}, \Theta, p(x|\theta)\}$ comme l'ensemble de l'information sur θ pouvant être obtenue de E (voir section 1.1), de Cristofaro (1996) [24] suggère que l'inférence sur θ soit établie conditionnellement à e. En conditionnant sur l'espace des observations $\mathcal{X}$, on ne porte de conclusion sur θ qu'en référence à l'information disponible dans une expérience donnée. En cela, la démarche de de Cristofaro rejoint celle de Carnap (1962) [18] selon lequel un degré de certitude pour les valeurs de θ sur la base de e doit être déterminé analytiquement par l'intermédiaire d'une fonction du type $c(\theta, e)$.

Ce conditionnement est en accord avec l'approche analytique de l'induction statistique où l'inférence est une procédure logique fondée sur une axiomatique, en l'occurrence celle de la théorie des probabilités (voir discussion dans de Cristofaro (2004)

[26]). Lindley (2000) [43] va plus loin en affirmant qu'une procédure en désaccord avec la théorie des probabilités n'est pas explicitement une inférence.

Si maintenant on considère que les observations sont obtenues par un processus de génération de données dans le cadre d'une expérience E, la probabilité associée à $i = (E, x)$ une observation (instance) prédite dans l'expérience E est influencée par la valeur de θ et par l'évidence e. La règle de Bayes qui permet de dériver la probabilité a posteriori de θ conditionnellement à e et à l'instance i s'écrit

$$p(\theta|e, i) = p(\theta|e)p(i|\theta, e). \tag{2.2.1}$$

En poussant un peu plus loin l'interprétation, l'instance i est prédite suivant un processus nécessitant seulement la connaissance de la valeur de θ et du plan expérimental "design" d. L'évidence e est donc constituée de l'information sur d et des connaissances sur θ contenue dans l'évidence préexpérimentale e_0. En substituant e par e_0, d dans (2.2.1), on obtient une nouvelle formulation de la règle de Bayes

$$p(\theta|i, e_0, d) = p(\theta|e_0, d)p(i|\theta, e_0, d). \tag{2.2.2}$$

Cette écriture souligne le rôle du plan expérimental dans la construction de la loi a priori. Si d est déterminé avant le recueil des données, il n'y a conceptuellement aucune raison de ne pas tenir compte de cette information pour construire la loi a posteriori. La formulation (2.2.2) du théorème de Bayes permet donc de faire figurer dans la loi a priori l'information sur d qui est perdue dans la vraisemblance.

Il est d'ailleurs incohérent d'ignorer la spécificité d'un plan expérimental qui avantage un ensemble de valeurs dans Θ. Dans le cas informatif, cette démarche est faîte intuitivement. Dans le cas non-informatif, c'est un préalable à la dérivation de l'a priori de Jeffreys.

La probabilité a priori $p(\theta|e_0, d)$ dans (2.2.2) confirme l'idée qu'un état d'ignorance a priori pour les valeurs de θ ne peut être exprimé que relativement au plan expérimental. De Cristofaro (1996) [24] suggère l'utilisation d'un a priori équiprobable si

le plan expérimental reste neutre dans l'inférence sur θ. Pour cela, d doit satisfaire le principe d'impartialité qu'il définit comme suit.

Principe d'impartialité du plan expérimental (PIPE). *Un plan expérimental est impartial pour θ si la vraisemblance est translatée.*

Le critère de vraisemblance translatée semble être celui qui exprime le mieux l'idée d'impartialité d'un plan expérimental. Il qualifie les vraisemblances symétriques autour de $\theta_{(x)}^{MV}$. La vraisemblance associée à une valeur de θ présentant un écart dt par rapport à $\theta_{(x)}^{MV}$ n'est pas affectée par la valeur de $\theta_{(x)}^{MV}$, donc par l'échantillonnage. Ce dernier ne favorise aucune valeur particulière de θ.

L'a priori de Jeffreys pour θ est associé à un a priori localement uniforme dans une métrique $\Psi(\theta)$, dans laquelle la vraisemblance est localement et approximativement translatée (2.1.2). Le $PIPE$ offre une nouvelle interprétation à l'a priori de Jeffreys car la métrique $\Psi(\theta)$ garantit de fait l'impartialité du plan expérimental. Cette assertion réhabilite également l'a priori Uniforme $\vartheta \sim Be(1,1)$ dans le modèle Binomial, lorsque l'effectif est suffisamment grand pour justifier une approximation Normale (qui est naturellement translatée).

Le $PIPE$ fait que l'a priori de Jeffreys est une réponse au problème de la détermination de l'évidence implicitement contenue dans le plan expérimental. Par la propriété induite de vraisemblance translatée, il modifie également la notion d'exhaustivité : seul un EMV associé à un plan expérimental impartial contient l'ensemble de l'information pour l'inférence sur θ.

Une inférence basée sur la seule vraisemblance ne peut pas prendre en compte l'information sur le plan expérimental car celle-ci est contenue dans la constante de proportionnalité. L'utilisation d'un a priori du type $p(\theta|e_0, d)$ transgresse le PV dans sa version formelle. Une version réactualisée du PV pourrait mentionner qu'une information sur le plan expérimental est également contenue dans la vraisemblance : l'impact du plan expérimental dans l'inférence sur θ se mesure par le non-respect du

PTV.

2.3 A priori de Jeffreys corrigé pour un plan expérimental comportant des analyses intermédiaires

Les arguments donnés précédemment supportent l'utilisation de la règle de Jeffreys pour définir une loi a priori $p(\theta|d)$ car elle garantit l'impartialité du plan expérimental. Dans cette section, on étend l'utilisation de cette règle à un plan expérimental comportant des analyses intermédiaires pouvant conduire à un arrêt de l'expérience.

L'analyse séquentielle groupée est associée à un plan expérimental où sont menées des analyses intermédiaires sur des effectifs prédéfinis pouvant conduire à l'arrêt de l'expérience. Dans la suite, on traite le cas de l'évaluation d'une proportion dans le contexte d'échantillonnages Binomiaux successifs. K désigne le nombre maximum d'analyses intermédiaires initialement prévu, n_k et Y_k respectivement le nombre fixé d'observations et le nombre de succès observés pendant la période k ($1 \leq k \leq K$). Enfin, M est la *va* associée au nombre total d'analyses effectuées (i.e. l'expérience s'arrête en période M) et on note respectivement $N = \sum_{k=1}^{M} n_k$ et $Y = \sum_{k=1}^{M} Y_k$ le nombre d'observations et de succès enregistrés après l'arrêt de l'expérience.

Le plan expérimental est paramétré par des bornes inférieures et supérieures predéfinies, respectivement $S_{k,L}$ et $S_{k,U}$, pour chaque réalisation de $Y^k = Y_1 + Y_2 + \ldots + Y_k$. L'expérience s'arrête à la k^{ieme} analyse ou en k^{ieme} période (i.e. $m = k$) si $Y^k \leq S_{k,L}$ ou si $Y^k \geq S_{k,U}$ avec la condition $S_{K,L} = S_{K,U} - 1$. On considère alors un échantillonnage Binomial $Y^k \sim Bin(n_1 + n_2 + \ldots + n_k, \theta)$.

La règle de Jeffreys préconise l'utilisation d'un a priori proportionnel à $I^{\frac{1}{2}}(\theta|m, y)$. La vraisemblance associée à (M, Y) s'écrit

$$L(\theta; m, y_1, .., y_K) = [Bin(n_1, \theta; y_1)]^{1_{m=1}} [Bin(n_1 + n_2, \theta; y_1 + y_2)]^{1_{m=2}} \times \ldots$$
$$\ldots \times \left[Bin(n_1 + n_2 + \ldots + n_K, \theta; y^K)\right]^{1_{m=K}}$$

où 1 désigne la fonction indicatrice.

Par une Log-transformation, on obtient

$$\begin{aligned}
Log(L(\theta; m, y_1, y_2, .., y_K)) &= 1_{m=1} Log\left(\binom{n_1}{y_1}\theta^{y_1}(1-\theta)^{n_1-y_1}\right) \\
&\quad + 1_{m=2} Log\left(\binom{n_1+n_2}{y_1+y_2}\theta^{y_1+y_2}(1-\theta)^{n_1+n_2-y_1-y_2}\right) \\
&+ \ldots + 1_{m=K} Log\left(\binom{n_1+..+n_K}{y^K}\theta^{y^K}(1-\theta)^{n_1+..+n_k-y^K}\right) \\
&= h(y_1, y_2, .., y_K) + 1_{m=1}[y_1 Log(\theta) + (n_1-y_1)Log(1-\theta)] \\
&\quad + 1_{m=2}[(y_1+y_2)Log(\theta) + (n_1+n_2-y_1-y_2)Log(1-\theta)] \\
&+ \ldots + 1_{m=K}[\,y^K Log(\theta) + (n_1+..+n_K-y^K)Log(1-\theta)] \\
&= h(y_1, y_2, .., y_K) + [y_1 Log(\theta) + (n_1-y_1)Log(1-\theta)] \\
&\quad + 1_{m\geq 2}[y_2 Log(\theta) + (n_2-y_2)Log(1-\theta)] \\
&+ \ldots + 1_{m=K}[y_K Log(\theta) + (n_K-y_K)Log(1-\theta)].
\end{aligned}$$

Comme

$$\frac{\partial}{\partial\theta}1_{m=i} = 0 \qquad \text{et} \qquad E_\theta[1_{M\geq i}] = Pr(M\geq i),$$

on déduit l'information de Fisher

$$\begin{aligned}
I(\theta|m, y) &= -E_\theta\left[\frac{\partial^2}{\theta^2}LogL(\theta; m, y_1, .., y_K)\right] \\
&= \frac{n_1}{\theta(1-\theta)} + Pr(M\geq 2)\frac{n_2}{\theta(1-\theta)} + \ldots + Pr(M=K)\frac{n_K}{\theta(1-\theta)} \\
&= \frac{1}{\theta(1-\theta)}\,[n_1 + n_2 Pr(M\geq 2) + \ldots + n_K Pr(M=K)].
\end{aligned}$$

La forme générale de l'a priori de Jeffreys corrigé dans le plan expérimental séquentiel d_{seq} est proportionnelle à $I^{\frac{1}{2}}(\theta|m, y)$. Si on note $r_k = n_k/n_1$, elle s'écrit

$$\vartheta|d_{Seq} \propto I^{\frac{1}{2}}(\theta|m, y) \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}[1+r_2 Pr(M\geq 2)+\ldots+r_K Pr(M=K)]^{\frac{1}{2}} \quad (2.3.1)$$

La densité de l'a priori de Jeffreys corrigé $p(\theta|d_{Seq})$ dépend de n_1 et des rapports des effectifs $r_k = n_k/n_1$. Le conditionnement sur la composante d_{Seq} garantit l'impartialité du plan expérimental dans l'inférence sur θ. L'information caractérisant

la règle d'arrêt du plan expérimental est contenue dans le terme $[1 + r_2 Pr(M \geq 2) + \ldots + r_K Pr(M = K)]$.

Chaque probabilité $Pr(M \geq k)$ est la somme des probabilités associées aux séquences $(y_1, y_2, \ldots, y_{k-1})$ à savoir

$$\begin{aligned} p(y_1, .., y_{k-1}|\theta) &= Pr(Y_1 = y_1, .., Y_{k-1} = y_{k-1}) \\ &= \binom{n_1}{y_1} .. \binom{n_{k-1}}{y_{k-1}} \theta^{y^{k-1}} (1-\theta)^{n_1 + .. + n_{k-1} - y^{k-1}} \end{aligned}$$

sur la restriction de dimension $k-1$

$$\mathcal{R}_{(k,y)} = \{(y_1, .., y_{k-1}) : S_{i,L} + 1 \geq y^i \geq S_{i,U} - 1 \quad \forall i \in [1, k-1]\}.$$

On obtient

$$Pr(M = K) = \sum_{\mathcal{R}_{(K,y)}} \binom{n_1}{y_1} .. \binom{n_{K-1}}{y_{K-1}} \theta^{y^{K-1}} (1-\theta)^{n_1 + .. + n_{K-1} - y^{K-1}}$$

$$Pr(M \geq k) = \sum_{\mathcal{R}_{(k,y)}} \binom{n_1}{y_1} .. \binom{n_{k-1}}{y_{k-1}} \theta^{y^{k-1}} (1-\theta)^{n_1 + .. + n_{k-1} - y^{k-1}}.$$

Par ailleurs, en notant la fonction Binomiale cumulative

$$CBin(n, \theta; x) = \sum_{j=0}^{x} \binom{n}{x} \theta^j (1-\theta)^{n-j},$$

on obtient la probabilité

$$Pr(M \geq 2) = Pr(S_{1,L} < Y_1 < S_{1,U}) = CBin(n_1, \theta; S_{1,U} - 1) - CBin(n_1, \theta; S_{1,L}).$$

2.4 Influence de la correction dans un plan expérimental comportant une analyse intermédiaire

L'influence de la correction de l'a priori de Jeffreys est étudiée dans un plan expérimental comportant un arrêt optionnel ($K = 2$).

2.4.1 Propriété de la loi a priori de Jeffreys corrigée

En notant J la partie de $[0, n_1]$ qui désigne la zone de continuation de l'expérience de la période 1 vers la période 2 (i.e. l'expérience est prolongée en période 2 si $Y_1 \in J$), on écrit

$$Pr(M=2) = Pr(Y_1 \in J) = CBin(n_1, \theta; J) = \sum_{j \in J} \binom{n_1}{j} \theta^j (1-\theta)^{n_1 - j}.$$

L'impact de la composante $[1 + r_2 Pr(M=2)]^{\frac{1}{2}}$ sur la densité a priori

$$p(\theta | d_{Seq}) = \theta^{-\frac{1}{2}} (1-\theta)^{-\frac{1}{2}} [1 + r_2 Pr(M=2)]^{\frac{1}{2}} \tag{2.4.1}$$

est examiné dans le cas où $n_1 = n_2 = 5$ (i.e. $r = r_2 = 1$).

En remplaçant $Pr(M = 2)$ dans (2.4.1), on obtient l'expression de l'a priori de Jeffreys corrigé

$$p(\theta | d_{Seq}) = C\theta^{-\frac{1}{2}} (1-\theta)^{-\frac{1}{2}} [1 + CBin(n_1, \theta; J)]^{\frac{1}{2}} \tag{2.4.2}$$

La composante $[1 + CBin(n_1, \theta; J)]^{\frac{1}{2}}$ est une fonction croissante de $CBin(n_1, \theta; J)$. La figure (2.2) décrit les courbes de $CBin(n_1, \theta; J)$ avec $J = [0, S_{1,U} - 1]$ en fonction de θ pour les cinq valeurs possibles de $S_{1,U}$=1,2,3,4,5. La valeur $S_{1,U} = 1$ produit la courbe qui décroît le plus tôt. La fonction $CBin(n_1, \theta; [0, S_{1,U} - 1])$ est décroissante en θ ($\forall$ $S_{1,U}$=1,2,3,4,5) et croissante en $S_{1,U}$ ($\forall$ $\theta \in [0, 1]$).

Le noyau de la densité a priori (2.4.2) est obtenu en multipliant $\theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}$ le noyau de la densité de la loi $Be(\frac{1}{2}, \frac{1}{2})$ par $(1 + rCBin(n_1, \theta; [0, S_{1,U} - 1]))^{\frac{1}{2}}$. Les courbes associées aux cinq valeurs $S_{1,U}$=1,2,3,4,5 sont décrites dans la figure (2.3).

La courbe correspondant à la valeur $S_{1,U} = 1$ est la plus proche de l'axe horizontal. Pour chaque valeur $S_{1,U} = 1, 2, 3, 4$, la courbe est monotone décroissante en θ. Lorsque $S_{1,U}$=5, la charge se stabilise à partir de $\theta \simeq 0.4$ et réaugmente légèrement

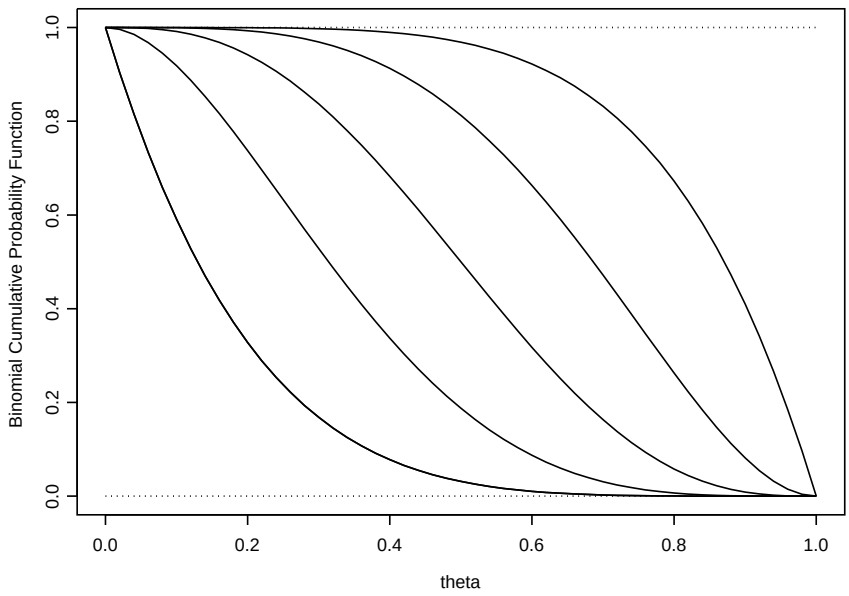

FIG. 2.2 – *Fonctions de répartition de la loi Binomiale* $CBin(n_1, \theta; J = [0, S_{1,U} - 1])$ *avec* $S_{1,U} = 1, 2, .., 5$

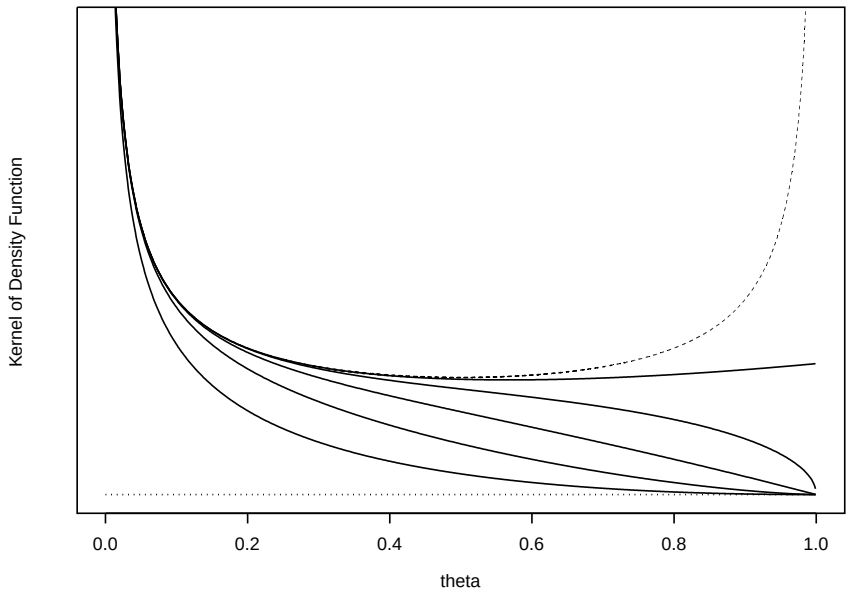

FIG. 2.3 – *Noyaux de la loi a priori de Jeffreys classique dans l'échantillonnage Binomial (- - -) et de la loi a priori de Jeffreys corrigée (—) dans les échantillonnages séquentiels* $d_{Seq_{(1,1)}}$*=(*n_1*=5,*$S_{1,U} = s_{1,U}$*,*n_2*=5) avec* $s_{1,U}$*=1,2,3,4,5*

pour les valeurs de θ proches de 1. Si on lève toute condition à la poursuite de l'expérience en période 2 (i.e. $S_{1,U}$=6 et N=n_1+n_2=10), on retrouve l'a priori de Jeffreys $Be(\frac{1}{2},\frac{1}{2})$ pour un échantillonnage Binomial fixé. Cet a priori correspond également au cas où aucun prolongement en période 2 n'est possible (i.e. $S_{1,U}$=0 et N=n_1=5).

Dans le cas particulier où $n_1 = S_{1,U} = 1$, on a la condition

$$\begin{cases} m = 1 \text{ si } y_1 = 1 \\ m = 2 \text{ si } y_1 = 0. \end{cases}$$

La densité de l'a priori de Jeffreys corrigé s'écrit

$$\begin{aligned} p(\theta|d_{Seq}) &\propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}(1+n_2(1-\theta))^{\frac{1}{2}} \\ &= C\,\theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}(1-\frac{n_2}{n_2+1}\theta)^{\frac{1}{2}}. \end{aligned} \tag{2.4.3}$$

Les noyaux des lois a priori de Jeffreys corrigée qui garantissent l'impartialité des plans séquentiels $d_{Seq_{(1,1)}}$=(n_1=1,$S_{1,U}$=1,n_2=1) et $d_{Seq_{(1,+\infty)}}$=(n_1=1,$S_{1,U}$=1,$n_2 \to +\infty$) figurent dans les expressions (2.4.4). On trouve respectivement

$$\begin{aligned} \vartheta|d_{Seq_{(1,1)}} &\propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}(1-\frac{\theta}{2})^{\frac{1}{2}} \\ \vartheta|d_{Seq_{(1,+\infty)}} &\sim Be(\frac{1}{2},1). \end{aligned} \tag{2.4.4}$$

Les courbes de ces fonctions ainsi que celle du noyau de la loi de Jeffreys classique dans l'échantillonnage Binomial ($\vartheta \sim Be(\frac{1}{2},\frac{1}{2})$) sont représentées dans la figure (2.4). La courbe du noyau de $\vartheta|d_{Seq_{(1,1)}}$ possède une forme en U légèrement asymétrique, avec une pondération moins forte pour les valeurs de θ proches de 1, alors que la courbe du noyau de $\vartheta|d_{Seq_{(1,+\infty)}}$ est strictement décroissante.

La figure (2.4) décrit également le noyaux de la loi de Jeffreys classique dans l'échantillonnage Binomial Négatif, qui montre une pondération plus forte pour les valeurs de θ proches de 1 que l'a priori $d_{Seq_{(1,+\infty)}}$. Ce cas est justifié dans la sous-section suivante.

D'une manière générale, les densités des a priori de Jeffreys classiques et corrigés

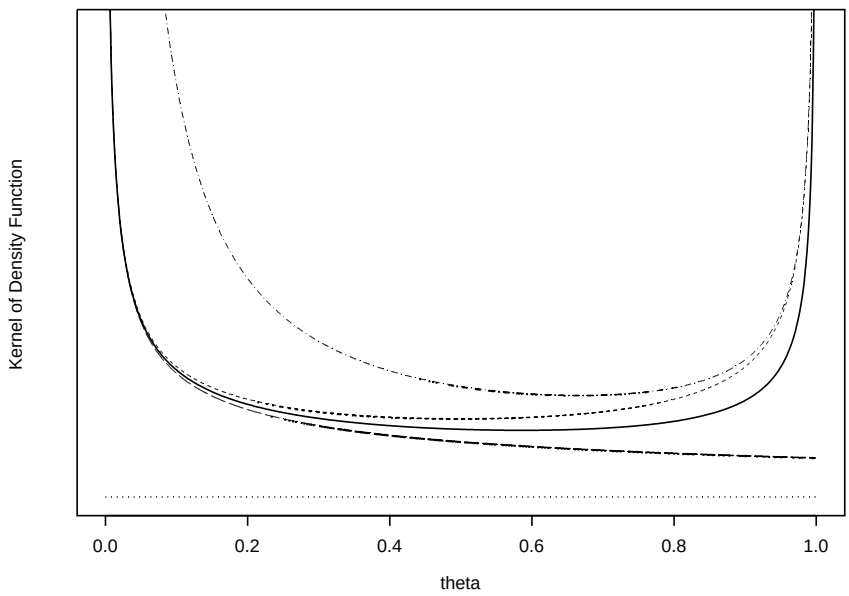

FIG. 2.4 – *Noyau de la loi a priori de Jeffreys classique dans les échantillonnages Binomial (- - -) et Binomial Négatif (· – ·), et corrigé dans les échantillonnages séquentiels* $d_{Seq_{(1,1)}}$*=(*n_1*=1,*$S_{1,U}$*=1,*n_2*=1) (—) et* $d_{Seq_{(1,+\infty)}}$*=(*n_1*=1,*$S_{1,U}$*=1,*$n_2 \to +\infty$*) (– –)*

laissent apparaître une charge plus importante sur les valeurs de θ qui sont désavantagées par l'échantillonnage.

2.4.2 Etude du biais induit sur l'EMV lorsque $n_1 = 1$

Les a priori de Jeffreys classiques et corrigés offrent une compensation du biais induit par la nature séquentielle des plans d'expérimentaux sur l'EMV. On vérifie cette assertion dans cette sous-section en déterminant les formes analytiques des biais lorsque $n_1 = 1$.

On estime le biais $B_{Seq,\hat{\theta}^{MV}}(\theta)$ de l'EMV dans le plan séquentiel $d_{Seq_{(1,1)}}$=(n_1=1,

$S_{1,U}$=1,n_2=1) en calculant l'espérance de $\hat{\theta}^{MV} = Y/N$.

$$\begin{aligned} E_{Seq_{(1,1)},\theta}\left(\frac{Y}{N}\right) &= Pr(Y_1 = 1) + Pr(Y_1 = 0)\sum_{i=0}^{n_2}\frac{i}{n_2+1}\binom{n_2}{i}\theta^i(1-\theta)^{n_2-i} \\ &= \theta + (1-\theta)\frac{n_2}{n_2+1}\theta \end{aligned}$$

Il vient ensuite

$$B_{Seq_{(1,1)},\hat{\theta}^{MV}}(\theta) = E_{Seq,\theta}\left(\frac{Y}{N}\right) - \theta = \frac{n_2}{n_2+1}\,\theta(1-\theta).$$

La fonction $B_{Seq_{(1,1)},\hat{\theta}^{MV}}(\theta)$ est décroissante en n_2. Elle trouve son minimum si $n_2 = 1$ avec

$$B_{Seq_{(1,1)},\hat{\theta}^{MV}}(\theta) = \frac{1}{2}\,\theta(1-\theta), \tag{2.4.5}$$

et son maximum si $n_2 \to +\infty$ avec

$$B_{Seq_{(1,+\infty)},\hat{\theta}^{MV}}(\theta) = \theta(1-\theta).$$

La figure (2.5) montre les courbes des espérances de $\hat{\theta}^{MV} = Y/N$ en fonction de θ dans les échantillonnages Binomial, Binomial Négatif, séquentiel avec $d_{Seq_{(1,1)}}$=(n_1=1, $S_{1,U}$=1,n_2=1) et $d_{Seq_{(1,+\infty)}}$=(n_1=1, $S_{1,U}$=1,$n_2 \to +\infty$).

L'estimateur $\hat{\theta}^{MV} = Y/N$ est sans-biais pour θ dans l'échantillonnage Binomial (i.e. $B_{Bin,\hat{\theta}^{MV}}(\theta)$=0) ; la courbe de $E_{Bin,\theta}(\hat{\theta}^{MV})$ est confondue avec la première bissectrice. Les biais de $\hat{\theta}^{MV}$ induits dans les autres échantillonnages se mesurent par l'écart des courbes des espérances par rapport à cette bissectrice. La courbe de $E_{NBin,\theta}(\hat{\theta}^{MV})$ croise celle de $E_{Seq_{(1,+\infty)},\theta}(\hat{\theta}^{MV})$, caractérisant un biais induit inférieur dans l'échantillonnage Binomial Négatif par rapport au plan $d_{Seq_{(1,+\infty)}}$=(n_1=1, $S_{1,U}$=1,$n_2 \to +\infty$) pour les grandes valeurs de θ. Cette observation justifie la pondération plus forte offerte pour ces valeurs par l'a priori de Jeffreys classique dans l'échantillonnage Binomial Négatif par rapport l'a priori de Jeffreys corrigé dans $d_{Seq_{(1,+\infty)}}$.

D'autre part, le biais de $\hat{\theta}^{MV} = Y/N$ dans l'échantillonnage Binomial Négatif est uniformément supérieur au biais induit dans l'échantillonnage séquentiel avec

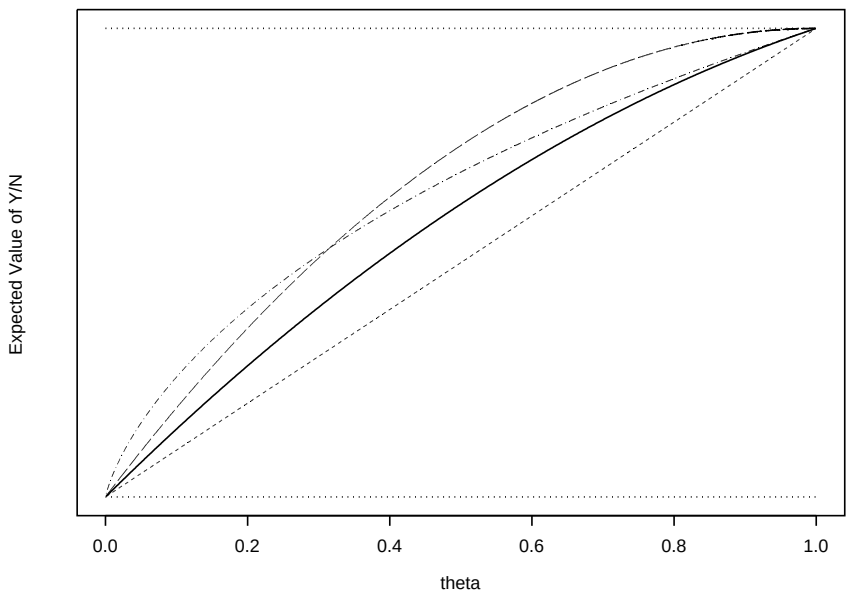

FIG. 2.5 – *Espérance de* $\hat{\theta}^{MV} = Y/N$ *dans l'échantillonnage Binomial (- - -), Binomial Négatif (· – ·), séquentiel avec* $d_{Seq_{(1,1)}}$*=(*n_1*=1,*$S_{1,U}$*=1,*n_2*=1) (—) et* $d_{Seq_{(1,+\infty)}}$ *=(*n_1*=1,*$S_{1,U}$*=1,*$n_2 \to +\infty$*) (– –)*

$d_{Seq_{(1,1)}}$=(n_1=1,$S_{1,U}$=1,n_2=1) (i.e. $B_{NBin,\hat{\theta}^{MV}}(\theta) > B_{Seq_{(1,1)},\hat{\theta}^{MV}}(\theta)$). En utilisant (2.1.2) et (2.4.5), on le vérifie analytiquement avec

$$\begin{aligned}
B_{NBin,\hat{\theta}^{MV}}(\theta) &\geq B_{Seq_{(1,1)},\hat{\theta}^{MV}}(\theta) \\
\frac{\theta}{1-\theta} Log\frac{1}{\theta} &\geq \frac{1}{2}\theta(1-\theta) \\
Log\frac{1}{\theta} &\geq \frac{(1-\theta)^2}{2} \\
\theta &\leq exp(-\frac{1}{2}(1-\theta)^2) \\
&\leq 1 - \frac{1}{2}(1-\theta)^2 + o((1-\theta)^4) \quad \text{avec} \quad o((1-\theta)^4) \geq 0 \\
\theta &\leq \frac{1}{2} + \theta - \frac{\theta^2}{2} + o((1-\theta)^4) \\
\theta^2 &\leq 1 + o((1-\theta)^4) \qquad \forall \theta \in [0,1]
\end{aligned}$$

2.5 Propriété de la loi Bêta-J d'ordre 1

La loi a priori de Jeffreys pour le plan séquentiel d_{Seq} appartient à une classe plus large de distributions notée Bêta-J, paramétrée par trois constantes a, b, c, les effectifs n_k $(k = 1, .., K)$ et les rapports $r_k = n_k/n_1$ $(k = 2, .., K)$. Les zones de continuation de l'expérience de la période $k-1$ à la période k sont les intervalles J_k $(k = 1, .., K-1)$. L'ordre $K-1$ de la loi Bêta-J est déterminé par le nombre maximum d'analyses intermédiaires prévues avant l'analyse finale.

La *va* $\varphi \sim Be^J(a, b, c; n_1, r_2, r_3, .., r_K; J_1, J_2, .., J_{K-1})$ a pour densité

$$p(\phi) = C\ \phi^{a-1}(1-\phi)^{b-1}[1 + r_2 Pr(M \geq 2) + \ldots + r_K Pr(M = K)]^c,$$

où

$$Pr(M \geq k) = \sum_{\mathcal{R}_{(k,y)}} \binom{n_1}{y_1} .. \binom{n_{k-1}}{y_{k-1}} \theta^{y^{k-1}} (1-\theta)^{n_1 + .. + n_{k-1} - y^{k-1}}$$

sur la restriction de dimension $k-1$

$$\mathcal{R}_{(k,y)} = \{(y_1, .., y_{k-1}) : y^i \in J_i \quad \forall i \in [1, k-1]\}.$$

Dans cette section on examine les propriétés de la loi Bêta-J d'ordre 1 (i.e. $K = 2$). En simplifiant l'écriture avec $r = r_2$ et $J = J_1$, on note que la *va*

$$\varphi \sim Be^J(a, b, c; n_1, r; J)$$

a pour densité

$$p(\phi) = C\ \phi^{a-1}(1-\phi)^{b-1}[1 + rCBin(n_1, \phi; J)]^c.$$

Dans le cas particulier $c = 0$, φ est distribuée suivant une loi $Be(a, b)$ de densité

$$p(\phi) = \frac{1}{B(a, b)}\ \phi^{a-1}(1-\phi)^{b-1}.$$

Les moments de φ sont

$$E(\varphi^r) = \frac{B(a, b)}{B(a + r, b)}.$$

Pour $c = 1$, la constante C est donnée par

$$C = \frac{1}{B(a,b) + \sum_{j \in J} \binom{n_1}{j} B(a+j, b+n_1-j)}.$$

Les moments de φ sont

$$E(\varphi^r) = \frac{B(a+r,b) + \sum_{j \in J} \binom{n_1}{j} B(a+j+k, b+n_1-j)}{B(a,b) + \sum_{j \in J} \binom{n_1}{j} B(a+j, b+n_1-j)}.$$

Dans le cas où $c = \frac{1}{2}$, Lecoutre (2004) [42] montre empiriquement qu'on obtient une très bonne approximation de la densité en prenant la moyenne des densités pour $c = 0$ et $c = 1$. On obtient une approximation des moments de la même façon.

Comme pour la loi $Be(a,b)$, la classe des lois a priori du type

$$\vartheta | d_{Seq} \sim Be^J(a, b, c; n_1, r; J)$$

forme une famille conjuguée avec la loi d'échantillonnage Binomiale $Y \sim Bin(n, \theta)$. La loi a posteriori ainsi formée est

$$\vartheta | d_{Seq}, y \sim Be^J(a+y, b+n-y, c; n_1, r; J),$$

de densité

$$p(\theta | d_{Seq}, y) = C\ \phi^{a+y-1}(1-\phi)^{b+n-1}(1 + rCBin(n_1, \phi; J))^c. \tag{2.5.1}$$

L'intégration en θ de la loi $Be^J(a+y, b+n-y, c; n_1, r; J)$ nécessite de déterminer la constante de proportionnalité C. Dans le cas particulier où on a $n_1 = 1$ et la condition

$$\begin{cases} m = 1 \text{ si } y_1 = 1 \\ m = 2 \text{ si } y_1 = 0, \end{cases}$$

la densité de la loi a priori

$$p(\theta | d_{Seq}) \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}(1 - \frac{n_2}{n_2+1}\theta)^{\frac{1}{2}}$$

est associée à la densité a posteriori

$$p(\theta|d_{Seq}, y) = C\ \theta^{a+y-1}(1-\theta)^{b+n-y-1}(1-\frac{n_2}{n_2+1}\theta)^c.$$

La détermination de la constante C peut être faite par identification à la fonction

$${}_2F_1(A,B;C;x) = \frac{\Gamma(C)}{\Gamma(A)\Gamma(C-A)}\int_0^1 v^{A-1}(1-v)^{C-A-1}(1-xv)^{-B}dv.$$

Les 2 systèmes d'équations suivants sont équivalents :

$$\begin{cases} v = \theta \\ A = a+y \\ C-A = b+n-y \\ -B = c \\ x = \frac{n_2}{n_2+1} \end{cases} \iff \begin{cases} \theta = v \\ A = a+y \\ C = a+b+n \\ B = -c \\ x = \frac{n_2}{n_2+1} \end{cases}$$

Il vient

$$\begin{aligned} C &=_2 F_1(a+y,-c;a+b+n;\frac{n_2}{n_2+1})\frac{\Gamma(a+y)\Gamma(b+n-y)}{\Gamma(a+b+n)} \\ &=_2 F_1(a+y,-c;a+b+n;\frac{n_2}{n_2+1})\frac{a+y}{b+n-y}. \end{aligned}$$

L'algorithme d'acceptation-rejet AR peut être utilisé pour l'intégration en θ sur la loi a posteriori $Be^J(a+y, b+n-y, c; n_1, r; J)$. On cherche d'abord une expression qui domine une fonction du noyau. Si on note $f(\theta)$ cette fonction, on a

$$\begin{aligned} f(\theta) &= \frac{1}{B(a+y,b+n-y)}\theta^{a+y-1}(1-\theta)^{b+n-y-1}(1+rCBin(n_1,\theta;J))^c \\ &= g(\theta)(1+rCBin(n_1,\theta;J))^c \end{aligned}$$

où $g(\theta)$ est la densité d'une loi $Be(a+y, b+n-y)$.

La fonction $f(\theta)$ est dominée par $g(\theta)(1+r)^c$. En notant $M = (1+r)^c$, on applique

l'algorithme AR

Repeat {

$\varphi \sim Be(a+y, b+n-y)$

$U \sim Unif[0,1]$

Si $U < \dfrac{1}{M}(1 + rCBin(n_1, \theta; J))^c \longrightarrow$ accepter$(\theta = \varphi)$

compt=compt+1

Si compt=$nbsim \longrightarrow$ STOP }

On obtient alors un ensemble de $nbsim$ valeurs de θ associées à une va

$$\vartheta \sim Be^J(a+y, b+n-y, c; n_1, r; J) \propto f(\theta).$$

Démonstration. On suppose $\vartheta \sim Be^J(a+y, b+n-y, c; n_1, r; J)$ et les fonctions $f(\theta)$, $g(\theta)$ comme définies plus haut,

$$\begin{aligned}
Pr(\vartheta \leq \theta_0 | U \leq \frac{f(\vartheta)}{M\, g(\vartheta)}) &= \frac{Pr(\vartheta \leq \theta_0, U \leq \frac{f(\vartheta)}{M\, g(\vartheta)}))}{Pr(U \leq \frac{f(\vartheta)}{M\, g(\vartheta)})} \\
&= \frac{\int_0^{\theta_0} \int_0^{\frac{f(\theta)}{Mg(\theta)}} dU\, g(\theta) d\theta}{\int_0^1 \int_0^{\frac{f(\theta)}{Mg(\theta)}} dU\, g(\theta) d\theta} \\
&= \frac{\int_0^{\theta_0} \frac{f(\theta)}{Mg(\theta)} g(\theta) d\theta}{\int_0^1 \frac{f(\theta)}{Mg(\theta)} dU\, g(\theta) d\theta} \\
&= \frac{\int_0^{\theta_0} f(\theta) d\theta}{\int_0^1 f(\theta) d\theta} \\
&= \int_0^{\theta_0} f(\theta) d\theta \\
&= Pr(\vartheta \leq \theta_0)
\end{aligned}$$

□

Par moyennage, on obtient

$$Pr(\vartheta \in \Theta) \simeq \frac{1}{nbsim} \sum_{i=1}^{nbsim} 1_{\theta \in \Theta}$$

$$E(\vartheta^r) \simeq \frac{1}{nbsim} \sum_{i=1}^{nbsim} \theta^r.$$

En reprenant la forme générale (2.5.1) de la loi a posteriori $Be^J(a+y, b+n-y, c; n_1, r; J)$, on obtient la densité prédictive de Y conditionnellement à ϑ par

$$\begin{aligned} Pr(Y=y) &= \int_0^1 p(y|\theta)p(\theta)d\theta \\ &= C \binom{n}{y} \int_0^1 \theta^{a+y-1}(1-\theta)^{b+n-y-1}(1+rCBin(n_1,\theta;J))^c d\theta \end{aligned} \tag{2.5.2}$$

où

$$C = \sum_{y=0}^{n} \binom{n}{y} \int_0^1 \theta^{a+y-1}(1-\theta)^{b+n-y-1}(1+rCBin(n_1,\theta;J))^c d\theta.$$

La nécessité de formuler un a priori sur θ limité à un ensemble de valeurs $\theta \in [dec, dec+e]$ conduit à considérer la loi Bêta décentrée

$$Be_{dec,e}(a,b) = dec + Be(a,b),$$

où dec est l'origine (ou le paramètre de décentrage) et e l'étendue. On caractérise la loi $Be_{dec,e}(a,b)$ par

$$\begin{aligned} E(Be_{dec,e}(a,b)) &= dec + e\frac{a}{a+b} \\ Var(Be_{dec,e}(a,b)) &= e^2 Var(Be(a,b)) \\ &= e^2 \frac{ab}{(a+b)^2(a+b+1)}. \end{aligned}$$

Si $\varphi \sim Be_{dec,e}(a,b)$ et $\vartheta \sim Be(a,b)$, on a alors les relations

$$\begin{aligned} \vartheta = dec + e\varphi &\Longleftrightarrow \varphi = \frac{\varphi - dec}{e} \\ \theta = dec + e\phi &\Longleftrightarrow \phi = \frac{\phi - dec}{e} \\ d\theta = e\, d\phi &\Longleftrightarrow d\phi = \frac{d\theta}{e}. \end{aligned}$$

Un changement de variable permet d'écrire les densités

$$\begin{aligned} p(\theta)d\theta &= \frac{1}{B(a,b)}\, \theta^{a-1}(1-\theta)^{b-1}d\theta & \theta \in [dec, dec+e] \\ p(\phi)d\phi &= \frac{e}{B(a,b)}\, \phi^{a-1}(1-\phi)^{b-1}d\phi & \phi \in [0,1]. \end{aligned}$$

On obtient la densité prédictive de Y conditionnellement à ϑ en remplaçant (2.5.2) par

$$\begin{aligned} Pr(Y=y) &= \int_{dec}^{dec+e} p(y|\theta)p(\theta)d\theta \\ &= \int_0^1 p(y|\phi)p(\phi)\frac{d\phi}{e} \\ = C\binom{n}{y}\int_0^1 \phi^{a-1}(1-\phi)^{b-1}(dec+e\phi)^y(1-dec-e\phi)^{n-y}(1+rCBin(n_1,\phi;J))^c d\phi \end{aligned} \tag{2.5.3}$$

où

$$C = \sum_{y=0}^{n}\binom{n}{y}\int_0^1 \phi^{a-1}(1-\phi)^{b-1}(dec+e\phi)^y(1-dec-e\phi)^{n-y}(1+rCBin(n_1,\phi;J))^c d\phi.$$

Chapitre 3

Application à l'estimation ponctuelle

Ce chapitre présente une nouvelle classe d'estimateurs d'une proportion dans un plan séquentiel comportant un arrêt optionnel. La loi Bêta-J est utilisée comme loi a priori en transposant des paramétrages de la loi Bêta dans l'échantillonnage fixe. Le principe sous-jacent à la détermination d'une forme analytique est exposé en section (3.2). Une méthode de résolution est proposée dans la section (3.3) dans le cas où l'arrêt est lié à la survenue d'au moins un succès. La section (3.4) permet une généralisation dans le contexte d'une étude comparative.

3.1 Procédure de Whitehead et Estimateur VUMSB

Dans l'estimation ponctuelle, l'utilisation de l'estimateur $\hat{\theta}^{MV} = Y/N$ engendre une sous-estimation de θ si l'arrêt en période k est provoqué par une condition $\hat{\theta}^{MV}_{(y)} \leq S_{k,L}$ et une surestimation si $\hat{\theta}^{MV}_{(y)} \geq S_{k,U}$. D'une manière générale, une condition d'arrêt unilatérale de l'expérience par valeurs inférieures (i.e. $S_{k,U} = n_k + 1,\ 1 \leq k \leq K$) ou bien supérieures (i.e. $S_{k,L} = -1,\ 1 \leq k \leq K$) se traduit par un biais plus important qu'une condition d'arrêt bilatérale. Deux méthodes visant à réduire le biais méthodologique lié à un arrêt optionnel de l'expérience sont présentées.

La première approche présentée par Whitehead (1986) [54] consiste à ajuster $\hat{\theta}^{MV}$

en soustrayant une estimation de son biais, obtenue suivant un schéma itératif. L'estimateur idéal est

$$\hat{\theta}^{Whi} = \hat{\theta}^{MV} - b(\theta) \tag{3.1.1}$$

où le biais $b(\theta)$ de $\hat{\theta}^{MV}$ en θ est donné par

$$E_\theta[\hat{\theta}^{MV}] = \sum_{\mathcal{Y}_{(m,y)}} \hat{\theta}^{MV}_{(y)} p(y|\theta) = \theta + b(\theta).$$

Ne connaissant pas la valeur de θ, on peut approcher $b(\theta)$ par $b(\hat{\theta}^{Whi}_{(y)})$ dans

$$E_\theta[\hat{\theta}^{MV}|\hat{\theta}^{Whi}_{(y)}] = \hat{\theta}^{Whi}_{(y)} + b(\hat{\theta}^{Whi}_{(y)}).$$

Après remplacement, l'expression (3.1.1) se réécrit

$$\begin{aligned}\hat{\theta}^{Whi}_{(y)} &\propto \hat{\theta}^{MV}_{(y)} - b(\hat{\theta}^{Whi}_{(y)}) \\ &= \hat{\theta}^{MV}_{(y)} - E_\theta[\hat{\theta}^{MV}|\hat{\theta}^{Whi}_{(y)}] + \hat{\theta}^{Whi}_{(y)}\end{aligned}$$

La détermination de $\hat{\theta}^{Whi}_{(y)}$ passe donc par la résolution de

$$E_\theta[\hat{\theta}^{MV}|\hat{\theta}^{Whi}_{(y)}] = \hat{\theta}^{Whi}_{(y)}.$$

On peut utiliser la méthode des bissectrices qui se fonde sur la détermination de deux limites $L \leq \hat{\theta}^{Whi}_{(y)} \leq U$ suivant un schéma récursif (voir Chang et Coll. (1989) [20]) : après l'initialisation $L = 0$ et $U = 1$, on calcule $E_\theta[\hat{\theta}^{MV}]$ en $\theta = (L+U)/2$. Si $E_\theta[\hat{\theta}^{MV}] < \hat{\theta}^{Whi}_{(y)}$, on remplace L par $\theta = (L+U)/2$; sinon on remplace U par $\theta = (L+U)/2$. La procédure s'arrête selon un niveau de précision min prédéfini (par exemple $U - L \leq min$).

Une deuxième approche consiste à déterminer l'*estimateur sans-biais de variance uniformément minimum* $VUMSB$. Jung et Kim (2004) [39] dérivent $\hat{\theta}^{VUMSB}$ en calculant l'espérance de $\hat{\theta}^{MV}_1 = Y_1/n_1$ sur la première période, qui est un estimateur sans-biais, conditionnellement à la statistique exhaustive complète (M, Y). Selon le théorème de Lehmann-Scheffé (A.3), il en résulte l'estimateur $VUMSB$.

On montre d'abord que (M, Y) est une statistique exhaustive par application du théorème de factorisation. La probabilité associée à une séquence $(y_1, ..., y_m)$ est

$$\begin{aligned} p(y_1, ..., y_m|\theta) &= \sum_{y_1} \cdots \sum_{y_m} Pr(Y_1 = y_1, \ldots, Y_m = y_m|\theta) \\ &= \sum_{y_1} \cdots \sum_{y_m} \binom{n_1}{y_1} \cdots \binom{n_m}{y_m} \theta^{(y_1+\ldots+y_m)} (1-\theta)^{(n_1+\ldots+n_m-y_1-\ldots-y_m)} \end{aligned}$$

dans la restriction

$$\mathcal{R}_{(m,y)} = \{(y_1, .., y_m) : y_1 + .. + y_m = y, S_{k,L} + 1 \leq y_1 + \ldots + y_m \leq S_{k,U} - 1, \forall k = 1, .., m-1\}$$

de dimension $m - 1$.

On obtient une expression fonction de (m, y)

$$\begin{aligned} p(y_1, ..., y_m|\theta) &= Pr(Y_m = y, S_{k,L} + 1 \leq Y_k \leq S_{k,U} - 1,\ k = 1, \ldots, m-1|\theta) \\ &= Pr(M = m, Y_m = y|\theta) \\ &= p(m, y|\theta) \end{aligned}$$

sur le support $\mathcal{S} = \bigcup_{m=1}^{K} \mathcal{S}_m$ où

$$\mathcal{S}_m = \{(m, y) : S_{m-1,L} + 1 \leq y \leq S_{m,L} \text{ ou } S_{m,U} \leq y \leq n_m + S_{m-1,U} - 1\}$$

avec $S_{0,L} = -1$ et $S_{0,U} = 1$, ce qui achève la démonstration.

On prouve ensuite que (M, Y) est une statistique complète en montrant que pour tout $\theta \in [0, 1]$, si on a la condition

$$E_\theta[g(M, Y)] = \sum_{m=1}^{K} \sum_{y=S_{m-1,L}+1}^{S_{m,L}} g(m, y)p(m, y|\theta) + \sum_{m=1}^{K} \sum_{y=S_{m,U}}^{n_m+S_{m-1,U}-1} g(m, y)p(m, y|\theta) = 0,$$

alors $g(M, Y) = 0$ P_θ p.s. pour toute les valeurs (m, y) dans le support $\mathcal{S}$ (voir Annexe A de Jung et Kim (2004) [39]). On note que dans le cas où les observations sont continues, la statistique $(M, S = \sum_i X_i)$ n'est pas complète (voir Liu (1999) [44]).

L'estimation $VUMSB$ est obtenue par $\hat{\theta}_{(m,y)}^{VUMSB} = E[\hat{\theta}_1^{MV}|(m, y)]$. Pour l'observation

(m, y), on a

$$\hat{\theta}^{VUMSB} = \frac{Pr(Y_1 = y_1, M = m, Y = y|\theta)}{Pr(M = m, Y = y|\theta)}$$
$$= \frac{\sum_{\mathcal{Y}(m,y|y_1)} \binom{n_1}{y_1} \theta^{y_1}(1-\theta)^{(n_1-y_1)} \dots \binom{n_m}{y_m} \theta^{y_m}(1-\theta)^{(n_m-y_m)}}{\sum_{\mathcal{Y}(m,y)} \binom{n_1}{y_1} \theta^{y_1}(1-\theta)^{(n_1-y_1)} \dots \binom{n_m}{y_m} \theta^{y_m}(1-\theta)^{(n_m-y_m)}}$$

où

$$\mathcal{Y}_{(m,y|y_1)} = \{(y_2, \dots, y_m) : y_2 + .. + y_m = y - y_1,$$
$$S_{k,L} + 1 \leq y_2 + \dots + y_m \leq S_{k,U} - 1, \forall k \in 2, \dots, m-1\}.$$

Par simplification des coefficients combinatoires, il vient

$$\hat{\theta}^{VUMSB} = \frac{\sum_{\mathcal{Y}_{(M,Y)}} \binom{n_1 - 1}{Y_1 - 1} \binom{n_2}{Y_2} \dots \binom{N}{Y}}{\sum_{\mathcal{Y}_{(M,Y)}} \binom{n_1}{Y_1} \dots \binom{N}{Y}}.$$

3.2 Equation du nouvel estimateur

Le plan expérimental $d_{Seq} = (n_1, S_{1,L}, S_{1,U}, n_2)$ est paramétré par des bornes inférieure et supérieure prédéfinies, $S_{1,L}$ et $S_{1,U}$, pour la réalisation de $Y = Y_1$. L'expérience s'arrête dès la première analyse (i.e. en fin de période 1) si $Y = Y_1 \leq S_{1,L}$ ou si $Y = Y_1 \geq S_{1,U}$; sinon la partie complémentaire $J = [S_{1,L} + 1, S_{1,U} - 1]$ définit la zone de continuation de l'expérience en période 2.

Si k est l'indice de la période $(k = 1, 2)$, n_k et Y_k sont respectivement les nombres fixés d'observations et le nombre observé de réponses pendant la période k. Enfin, si M est la *va* associée au nombre d'analyses effectuées (i.e. l'expérience s'arrête en période M), on note respectivement $N = \sum_{k=1}^{M} n_k$ et $Y = \sum_{k=1}^{M} Y_k$ le nombre d'observations et de succès enregistrées après l'arrêt de l'expérience.

On note l'espace de observations

$$\begin{aligned}\mathcal{Y}_{(m,y)} &= \mathcal{Y}_{(1,y)} \cup \mathcal{Y}_{(2,y)} \\ &= \{y = y_1, y \leq S_{1,L} \text{ ou } y \geq S_{1,U}\} \cup \{y = y_1 + y_2, S_{1,L}+1 \leq y \leq S_{1,U}-1+n_2\}.\end{aligned}$$

Un arrêt unilatéral de l'expérience en période 1, soit par valeur inférieure soit par valeur supérieure, se traduit respectivement par la suppression de la borne $S_{1,U}$ ou $S_{1,L}$. Dans certaines expressions, on utilisera également les formes $S_{1,U} = n_1 + 1$ ou $S_{1,L} = 0$.

Dans la section 2.4, on montre que la correction de l'a priori de Jeffreys offre un mécanisme de compensation du biais expérimental dans le plan séquentiel $d_{Seq} = (n_1, S_{1,L}, S_{1,U}, n_2)$. L'idée développée dans cette section est d'utiliser un a priori du type $\vartheta | d_{Seq} \sim Be^J(a, b, c; n_1, r; J)$ et d'utiliser un paramétrage spécifique a, b, c pour dériver la loi a posteriori. Parmi les estimateurs bayésien fondés sur cette loi, le mode est celui qui offre le plus de perspectives dans la recherche d'une forme analytique.

Dans l'échantillonnage Binomial fixe $Y \sim Bin(n, \theta)$, la moyenne et le mode des lois a posteriori associées respectivement à l'a priori uniforme $\vartheta \sim Be(1, 1)$ et l'a priori de Haldane $\vartheta \sim Be(0, 0)$ coïncident avec l'EMV non biaisé $\hat{\theta} = Y/n$ (les caractéristiques du mode et de la moyenne d'une loi Bêta sont données dans la section (B.2)).

En remarquant que l'a priori de Haldane $\vartheta \propto I(\theta|y)$ dans l'échantillonnage Binomial fixe, on déduit l'a priori de Haldane corrigé $\vartheta | d_{Seq} \propto I(\theta|m, y) \sim Be^J(0, 0, 1; n_1, r; J)$ dans le plan séquentiel d_{Seq}. Cette loi a priori a pour densité

$$p(\theta | d_{Seq}) \propto \theta^{-1}(1 - \theta)^{-1}(1 + rCBin(n_1, \theta; J)).$$

En transposant le valeur $c = 1$, on obtient également l'a priori Uniforme corrigé $\vartheta | d_{Seq} \propto Be^J(1, 1, 1; n_1, r; J)$ avec pour densité

$$p(\theta | d_{Seq}) \propto (1 + rCBin(n_1, \theta; J)).$$

Son profil de densité est proportionnellement identique à celui de la fonction $CBin(n_1, \theta; J)$ décrite dans la figure (2.2) lorsque $J = [0, S_{1,U} - 1]$ pour les valeurs $S_{1,U}$=1,2,3,4,5 et $n_1 = 5$.

La loi a posteriori est une $Be^J(1 + y, 1 + n - y, 1; n_1, r; J)$ dont la densité prend la forme

$$p(\theta|d_{Seq}, y) \propto \theta^y (1-\theta)^{n-y}(1 + rCBin(n_1, \theta; [S_{1,L} + 1, S_{1,U} - 1])).$$

Le mode est la solution de l'équation

$$\frac{d}{d\theta} Log(p(\theta|d_{Seq}, y)) = 0. \tag{3.2.1}$$

La résolution nécessite la réécriture de

$$CBin(n_1, \theta; J) = CBin(n_1, \theta; [0, S_{1,U} - 1]) - CBin(n_1, \theta; [0, S_{1,L}])$$

sous la forme de fonctions Bêta incomplètes (Abramovitz & Stegun (1964) [1]). L'expression développée ci-dessous s'applique à la forme $CBin(n_1, \theta; [0, S_{1,U} - 1])$.

$$\begin{aligned}
CBin(n_1, \theta; [0, S_{1,U} - 1]) &= CBin(n_1, 1-\theta; [S_{1,U}, n_1]) \\
&= Pr(Be(n_1 - S_{1,U} + 1, S_{1,U}) < 1 - \theta) \\
&= \frac{1}{\beta(n_1 - S_{1,U} + 1, S_{1,U})} \int_0^{1-\theta} \xi^{n_1 - S_{1,U}} (1-\xi)^{S_{1,U}-1} d\xi \\
&= I_{1-\theta}(n_1 - S_{1,U} + 1, S_{1,U}).
\end{aligned}$$

La formulation pour $CBin(n_1, \theta; [0, S_{1,L}])$ est obtenue de la même façon.

$$\begin{aligned}
CBin(n_1, \theta; [0, S_{1,L}]) &= CBin(n_1, 1-\theta; [S_{1,L} + 1, n_1]) \\
&= Pr(Be(n_1 - S_{1,L}, S_{1,L} + 1) < 1 - \theta) \\
&= \frac{1}{\beta(n_1 - S_{1,L}, S_{1,L} + 1)} \int_0^{1-\theta} \xi^{n_1 - S_{1,L} - 1} (1-\xi)^{S_{1,L}} d\xi \\
&= I_{1-\theta}(n_1 - S_{1,L}, S_{1,L} + 1).
\end{aligned}$$

En sommant les 2 expressions, il vient

$$CBin(n_1, \theta; [S_{1,L}+1, S_{1,U}-1]) = I_{1-\theta}(n_1 - S_{1,U}+1, S_{1,U}) - I_{1-\theta}(n_1 - S_{1,L}, S_{1,L}+1).$$

On peut maintenant résoudre l'équation (3.2.1) qui se réécrit

$$\begin{aligned}
&\frac{d}{d\theta} Log(p(\theta|d, y)) \\
&\quad = \frac{d}{d\theta}\left[yLog\theta + (n-y)Log(1-\theta) + Log(1+rCBin(n_1, \theta; [S_{1,L}+1, S_{1,U}-1]))\right] \\
&\quad = \frac{y}{\theta} - \frac{n-y}{1-\theta} + \frac{d}{d\theta} Log[1 + rI_{1-\theta}(n_1 - S_{1,U}+1, S_{1,U}) - rI_{1-\theta}(n_1 - S_{1,L}, S_{1,L}+1)] \\
&\quad = \frac{y - n\theta}{\theta(1-\theta)} + \frac{f'(\theta)}{f(\theta)} = 0,
\end{aligned}$$

où

$$\begin{aligned}
f(\theta) &= 1 + rCBin(n_1, \theta; [S_{1,L}+1, S_{1,U}-1]) \\
&= 1 + rI_{1-\theta}(n_1 - S_{1,U}+1, S_{1,U}) - rI_{1-\theta}(n_1 - S_{1,L}, S_{1,L}+1).
\end{aligned}$$

La forme intégrale $I_{1-\theta}$ facilite la dérivation de $f(\theta)$ en θ,

$$\begin{aligned}
f'(\theta) &= \frac{-r}{\beta(n_1 - S_{1,U}+1, S_{1,U})}\, \theta^{S_{1,U}-1}(1-\theta)^{n_1 - S_{1,U}} \\
&\qquad + \frac{r}{\beta(n_1 - S_{1,L}, S_{1,L}+1)}\, \theta^{S_{1,L}}(1-\theta)^{n_1 - S_{1,L}-1} \\
&= \frac{-r\; n_1!}{(n_1 - S_{1,U})!(S_{1,U}-1)!}\, \theta^{S_{1,U}-1}(1-\theta)^{n_1 - S_{1,U}} \\
&\qquad + \frac{r\; n_1!}{(n_1 - S_{1,L}-1)!S_{1,L}!}\, \theta^{S_{1,L}}(1-\theta)^{n_1 - S_{1,L}-1} \\
&= \frac{-r\; S_{1,U}}{\theta}\, Bin(n_1, \theta; S_{1,U}) + \frac{r\;(S_{1,L}+1)}{\theta}\, Bin(n_1, \theta; S_{1,L}+1).
\end{aligned}$$

La forme générale de l'équation à résoudre est maintenant

$$\frac{1-\theta}{y - n\theta} = \frac{1}{r}\, \frac{1 + r\, CBin(n_1, \theta; [S_{1,L}+1, S_{1,U}-1])}{S_{1,U}\, Bin(n_1, \theta; S_{1,U}) - (S_{1,L}+1)\, Bin(n_1, \theta; S_{1,L}+1)}. \tag{3.2.2}$$

A partir de la solution en θ de l'équation (3.2.2), on définit un nouvel estimateur noté $\hat{\theta}^{Mod}$. On note au passage que le terme de droite est une fonction de θ conditionnellement aux paramètres fixés du plan expérimental $n_1, r = n_2/n_1, S_{1,L}$ et $S_{1,U}$.

Le terme de gauche est un quotient dont le dénominateur

$$y - n\theta = n(\frac{y}{n} - \theta)$$

dépend de (m, y) à travers les quantités aléatoires $(n, y/n)$.

3.3 Cas d'un arrêt conditionné à l'observation d'au moins un événement

3.3.1 Principe de résolution graphique

La résolution de l'équation (3.2.2) est d'abord décrite dans le cas où l'observation d'au moins une réponse conduit à l'arrêt de l'expérience en période 1 (i.e $M = 1$ si $y_1 \geq 1$). La zone de continuation définie pour Y_1 se réduit alors à la valeur $J = \{0\}$ et seule la borne $S_{1,U} = 1$ est à considérer. L'équation (3.2.2) se simplifie en

$$\begin{aligned} \frac{1-\theta}{y-n\theta} &= \frac{1}{r\,n_1}\ \frac{1+(1-\theta)^{n_1}}{\theta(1-\theta)^{n_1-1}} \\ \frac{\theta}{y-n\theta} &= \frac{1}{r\,n_1}\ \left(1+\frac{1}{(1-\theta)^{n_1}}\right). \end{aligned} \tag{3.3.1}$$

Les deux termes de la dernière égalité sont l'expression de deux fonctions holomorphes en θ. Leurs courbes sont respectivement décrites dans les figures (3.1) et (3.2) sur les espaces $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 1, 2, .., 5\}$ et $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 5\}$ avec un effectif $n_1 = n_2 = 5$. On visualise l'asymptote verticale en $\theta \to 1$ de la fonction

$$g(\theta) = \frac{1}{r\,n_1}\ \left(1+\frac{1}{(1-\theta)^{n_1}}\right)$$

et les asymptotes verticales en $\theta \to y/n$ $(y = 1, 2, .., 5)$ de la fonction

$$h(\theta) = \frac{\theta}{y - n\,\theta}.$$

Les intersections de $g(\theta)$ et $h(\theta)$ sont les solutions de l'équation (3.3.1) et les valeurs de θ correspondantes définissent les estimations $\hat{\theta}^{Mod}_{(m,y)}$. La fonction $g(\theta)$ est strictement croissante positive sur l'intervalle $\Theta = [0, 1]$ et converge lentement vers son

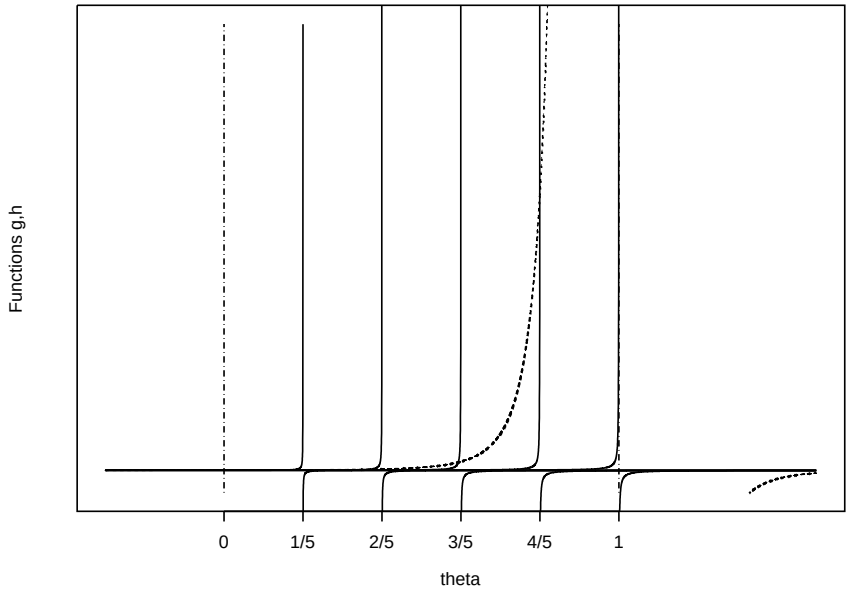

FIG. 3.1 – *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) pour* $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 1, 2, .., 5\}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

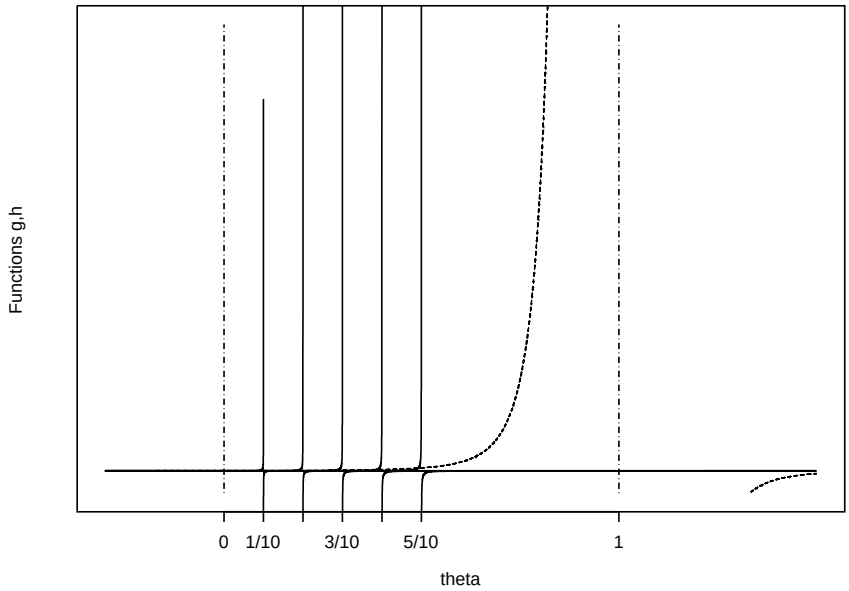

FIG. 3.2 – *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) pour* $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 5\}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

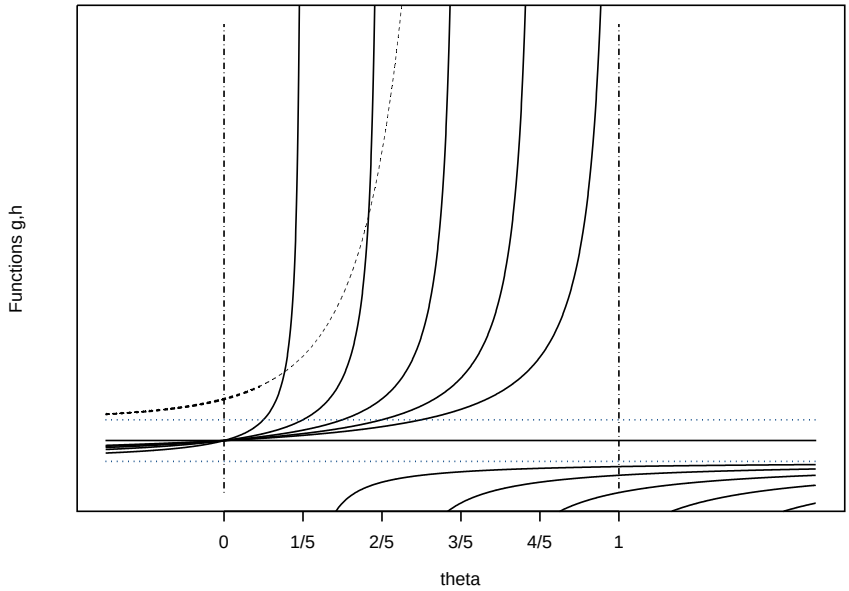

FIG. 3.3 – *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) pour* $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 1, 2, .., 5\}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

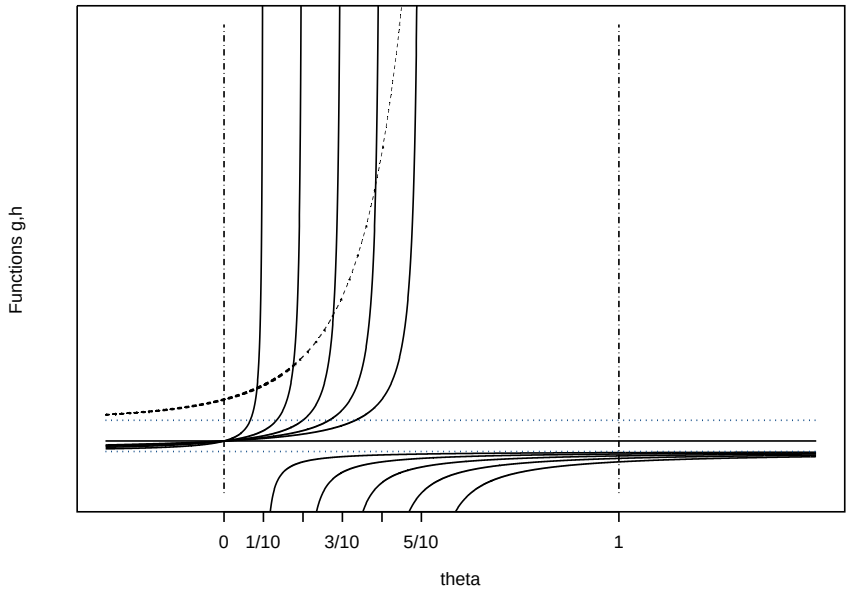

FIG. 3.4 – *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) pour* $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 5\}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

asymptote en $\theta \to 1$. Sur le même intervalle, la partie positive de la fonction $h(\theta)$ est également strictement croissante mais converge rapidement vers son asymptote en $\theta \to y/n$.

Pour la valeur $y = y_1 = n_1$ dans $\mathcal{Y}_{(1,y)}$, les fonctions $g(\theta)$ et $h(\theta)$ partagent le même asymptote en $\theta \to 1$. La figure (3.2) montre que pour θ compris entre 0 et une valeur proche de 1, on a $g(\theta) > h(\theta)$. Cependant, lorsque $\theta \to 1$ on a

$$g(\theta) \to \frac{1}{n_1(1-\theta)} \quad \text{et} \quad h(\theta) \to \frac{1}{rn_1(1-\theta)^{n_1}}.$$

De fait si $n_1 > 2$, les courbes de $g(\theta)$ et de $h(\theta)$ se croisent avant l'asymptote $\theta \to 1$ car

$$\exists\, \theta < 1 \text{ tel que } \frac{1}{1-\theta} < \frac{1}{r(1-\theta)^{n_1}}.$$

Si $n_1 = 1$, les courbes se croisent si $r = n_2/n_1 < 1$. Sinon, si $r > 1$, la solution de l'équation n'existe pas pour $y = n_1 = 1$ dans $\mathcal{Y}_{(1,y)}$; on affecte alors la valeur $\hat{\theta}^{Mod}_{(1,1)} = 1$.

Les deux figures 3.1 et 3.2 sont reprises et présentées avec une échelle verticale plus réduite dans les figures (3.3) et (3.4). Elles permettent de visualiser les asymptotes horizontales de $g(\theta)$ et de $h(\theta)$. Lorsque $\theta \to \pm\infty$, on a

$$g(\theta) \to \frac{1}{r\,n_1} \quad \text{et} \quad h(\theta) \to -\frac{1}{n}.$$

Comme

$$g(0) = \frac{2}{rn_1} \quad \text{et} \quad f(0) = 0,$$

la courbe de $g(\theta)$ croise forcément les courbes de $h(\theta)$, $\forall (m,y) \in \mathcal{Y}_{(m,y)}(n_1 > 2)$.

Quelle que soit la valeur de $y \in \mathcal{Y}_{(m,y)}$, l'estimation $\hat{\theta}^{Mod}_{(m,y)}$ vient corriger la valeur de $\hat{\theta}^{MV}_{(m,y)}$ par une sous-estimation. La figure (3.5) montre les valeurs de $\hat{\theta}^{Mod}_{(m,y)}$ pour $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 1, 2, .., 5\}$ et pour $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 5\}$ en fonction des observations $(m, y/n)$ dans un plan $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$. La correction maximum de $\hat{\theta}^{MV} = Y/n$ est obtenue avec les observations $(m, y) = (1, 5)$ dans $\mathcal{Y}_{(1,y)}$ et $(m, y) = (2, 10)$ dans $\mathcal{Y}_{(2,y)}$.

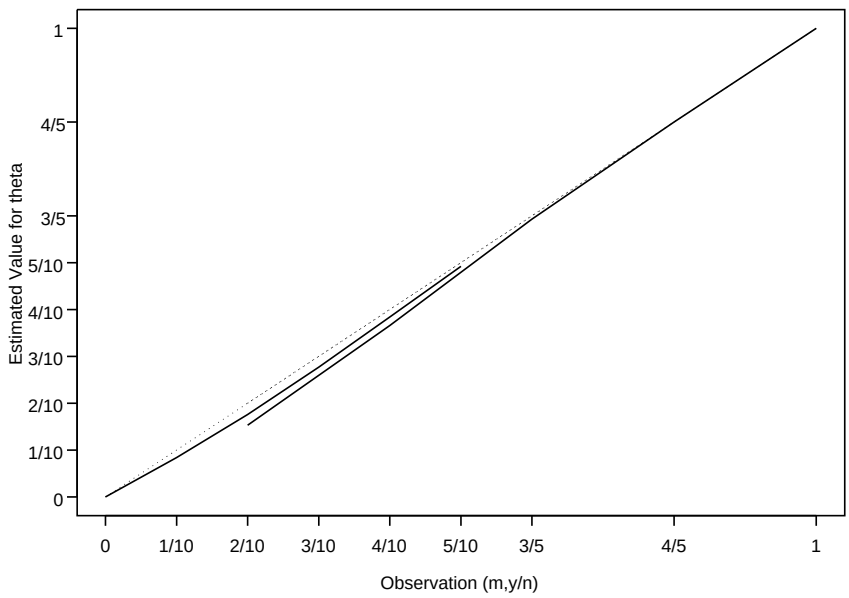

FIG. 3.5 – *Estimateur du Mode en fonction de* $(m, y/n)$ *pour* $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 1, 2, .., 5\}$ *et* $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 5\}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

3.3.2 Dérivation d'une forme analytique non-linéaire

La solution exacte en θ de l'équation (3.2.2) se détermine numériquement. On peut cependant dériver l'expression analytique d'une approximation en utilisant les caractéristiques des fonctions $g(\theta)$ et $h(\theta)$. Les figures (3.6) et (3.7) reprennent les figures (3.3) et (3.4) avec une vue élargie dans l'intervalle $\Theta = [0, 0.5]$.

Quelle que soit la valeur de y, aucun élément ne permet d'identifier précisément l'abscisse ou l'ordonnée du point d'intersection entre $g(\theta)$ et $h(\theta)$. En revanche la courbe de $g(\theta)$ vient couper les asymptotes de $h(\theta)$ au point $(y/n, g(y/n))$. La projection horizontale de ce point sur $h(\theta)$ est proche du point d'intersection avec $g(\theta)$ et son abscisse constitue une bonne approximation de l'abscisse de l'intersection.

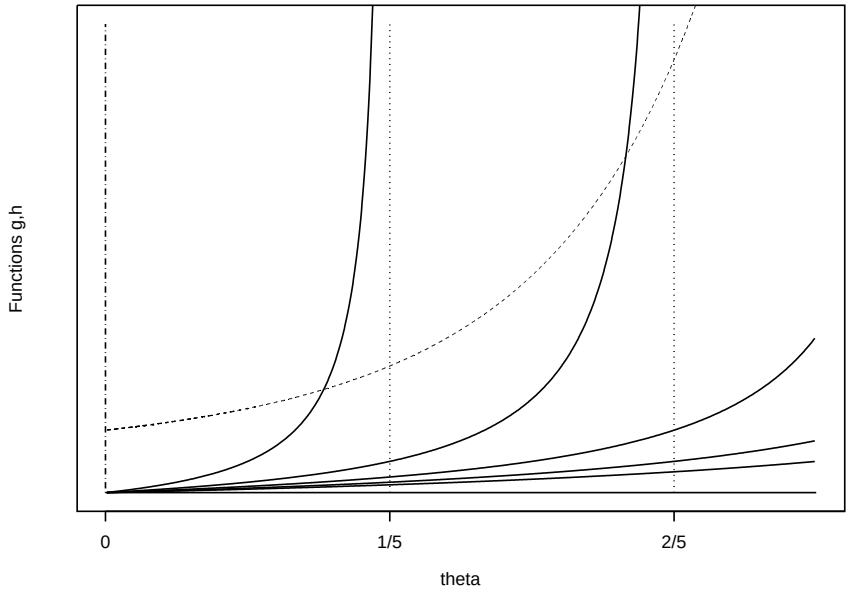

FIG. 3.6 – *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) pour* $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 1, 2, .., 5\}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

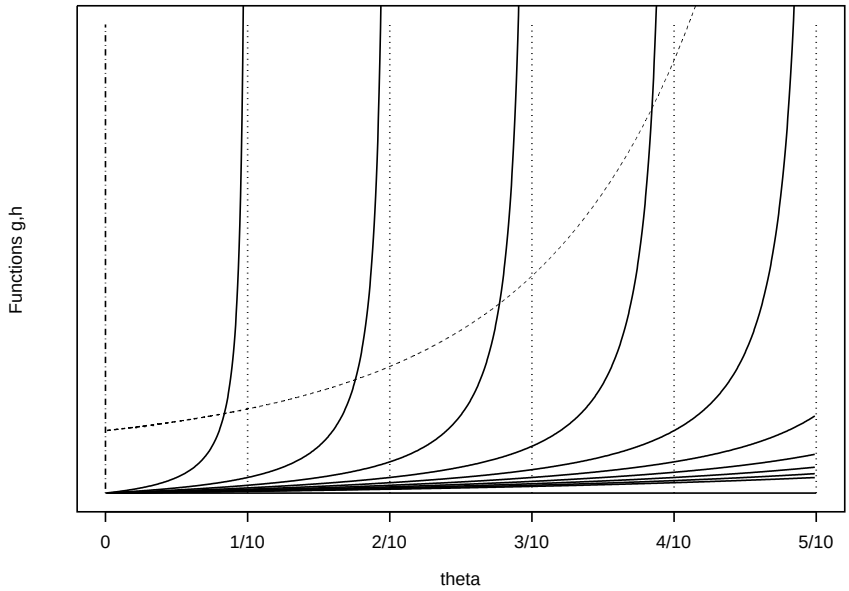

FIG. 3.7 – *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) pour* $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 5\}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

Cette approximation est la solution en θ de l'équation

$$\begin{aligned} g\left(\frac{y}{n}\right) &= h(\theta) \\ \frac{1}{r\,n_1}\left(1+\frac{1}{(1-y/n)^{n_1}}\right) &= \frac{\theta}{y-n\,\theta} \end{aligned} \tag{3.3.2}$$

qui se réécrit

$$\begin{aligned} \theta &= \frac{yg(y/n)}{1+ng(y/n)} \\ &= \frac{y}{n+\frac{1}{g(y/n)}}. \end{aligned}$$

Il en résulte l'estimateur du mode approximé

$$\hat{\theta}^{ModA} = \frac{Y}{n+rn_1\frac{1}{1+\left(\frac{n}{n-Y}\right)^{n_1}}}. \tag{3.3.3}$$

La dérivation de $\hat{\theta}^{ModA}$ peut aisément se généraliser. Cependant, il survient une légère surestimation de $\hat{\theta}^{Mod}$ car le croisement de $h(\theta)$ avec l'asymptote intervient au dessus de l'intersection avec $g(\theta)$. Une approximation encore plus précise de $\hat{\theta}^{Mod}$ peut être dérivée. En reprenant les figures (3.6) et (3.7), l'idée est de déterminer une autre fonction que $g(\theta)$ dont la courbe viendrait croiser les asymptotes de $h(\theta)$ à la hauteur de l'intersection entre $g(\theta)$ et $h(\theta)$.

En notant

$$\begin{aligned} g_{0.5}(\theta) &= \frac{1}{r\,n_1}\left(1+\frac{1}{(1-\theta)^{n_1-0.5}}\right) \\ g_1(\theta) &= \frac{1}{r\,n_1}\left(1+\frac{1}{(1-\theta)^{n_1-1}}\right), \end{aligned}$$

la figure (3.8), permet de visualiser les courbes des trois fonctions

$$g(\theta) \geq g_{0.5}(\theta) \geq g_1(\theta), \qquad \forall\theta \in [0,1].$$

Quelle que soit la valeur de y, la courbe de $g_1(\theta)$ croise les asymptotes de $h(\theta)$ à une ordonnée inférieure à celle du point d'intersection entre $g(\theta)$ et $h(\theta)$. L'abscisse de la projection horizontale sur $h(\theta)$ du point de croisement entre $g_1(\theta)$ et l'asymptote de

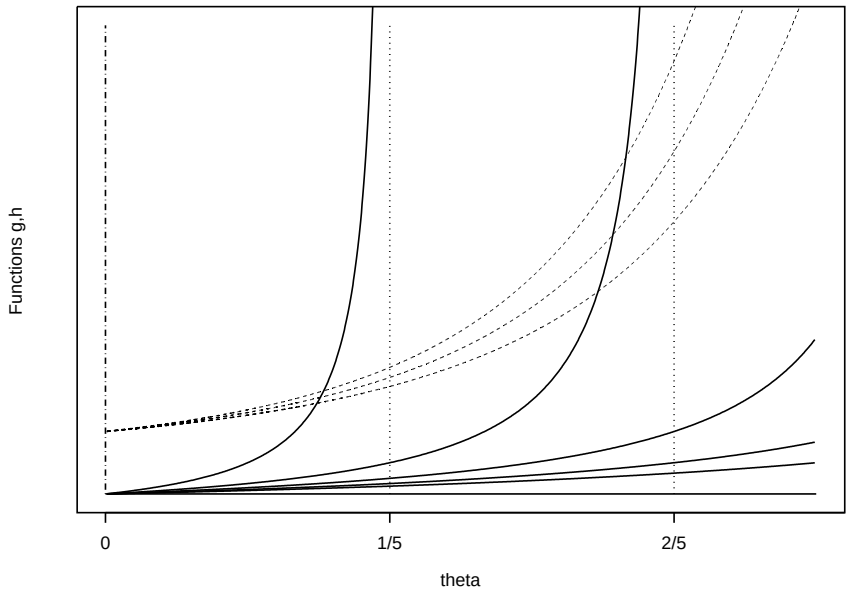

FIG. 3.8 – *Fonctions holomorphes* $g(\theta)$, $g_{0.5}(\theta)$, $g_1(\theta)$ *(- - -) et* $h(\theta)$ *(—) pour* $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 1, 2, .., 5\}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

$h(\theta)$ sous-estime légèrement $\hat{\theta}^{Mod}_{(m,y)}$. La courbe de $g_{0.5}(\theta)$ se situe entre celle de $g(\theta)$ et de $g_1(\theta)$. Elle croise les asymptotes de $h(\theta)$ à une ordonnée très proche de celle du point d'intersection entre $g(\theta)$ et $h(\theta)$. L'abscisse de la projection sur $h(\theta)$ des points de croisement entre $g_{0.5}(\theta)$ et l'asymptote de $h(\theta)$ constitue une très bonne approximation de $\hat{\theta}^{Mod}_{(m,y)}$.

Les valeurs en θ associées au croisement des courbes de $g_{0.5}(\theta)$ et $g_1(\theta)$ avec celles de $h(\theta)$ sont également des estimations du Mode Approximé, que l'on note respectivement $\hat{\theta}^{ModA_{0.5}}$ et $\hat{\theta}^{ModA_{0.5}}$. On en déduit les estimateurs

$$\hat{\theta}^{ModA_{0.5}} = \frac{Y}{n + rn_1 \frac{1}{1+\left(\frac{n}{n-Y}\right)^{n_1-0.5}}}$$

$$\hat{\theta}^{ModA_1} = \frac{Y}{n + rn_1 \frac{1}{1+\left(\frac{n}{n-Y}\right)^{n_1-1}}}.$$

On représente dans la figure (3.9) les valeurs de $\hat{\theta}^{ModA}_{(m,y)} - y/n$, $\hat{\theta}^{ModA_{0.5}}_{(m,y)} - y/n$ et $\hat{\theta}^{ModA_1}_{(m,y)} - y/n$ dans $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 1, 2, .., 5\}$ et dans $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y =$

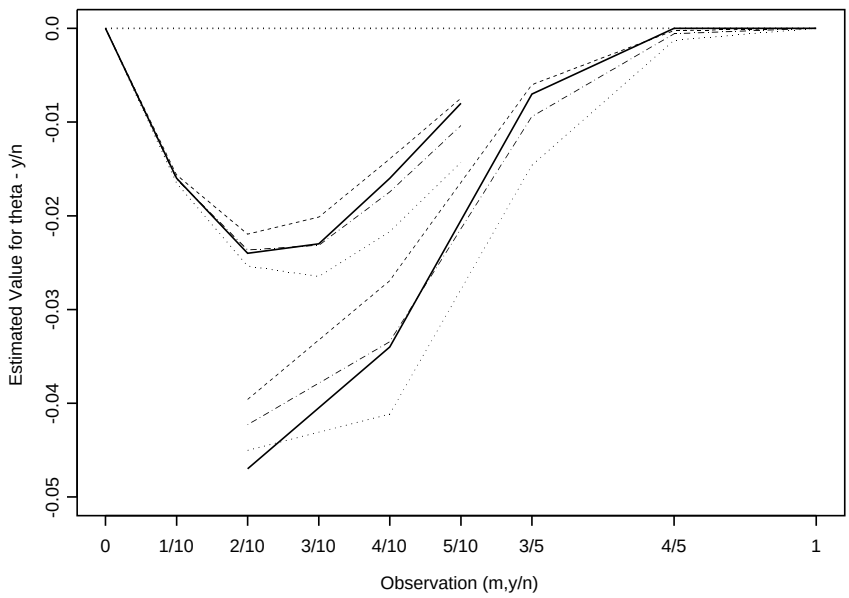

FIG. 3.9 – *Estimations* $\hat{\theta}^{Mod}_{(m,y)}$ *(—),* $\hat{\theta}^{ModA}_{(m,y)}$ *(- - -),* $\hat{\theta}^{ModA_{0.5}}_{(m,y)}$ *(· − ·), et* $\hat{\theta}^{ModA_1}_{(m,y)}$ *(· · ·) soustraites de* y/n *pour* $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 1, 2, .., 5\}$ *et* $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 5\}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

$0, 1, .., 5\}$.

L'estimation $\hat{\theta}^{ModA}_{(m,y)}$ tend à surestimer $\hat{\theta}^{Mod}_{(m,y)}$ alors que $\hat{\theta}^{ModA_1}_{(m,y)}$ la sous-estime. Entre les 2, on vérifie que $\hat{\theta}^{ModA_{0.5}}_{(m,y)}$ constitue une estimation très précise de $\hat{\theta}^{Mod}_{(m,y)}$ dans $\mathcal{Y}_{(m,y)}$.

3.3.3 Dérivation d'une forme analytique linéaire

Les calculs analytiques de la section suivante nécessitent de dériver une forme linéaire de l'estimateur du Mode Approximé. En le réécrivant sous la forme

$$\hat{\theta}^{ModA} = \frac{Y}{n}\left(1 + \frac{rn_1}{n}\frac{1}{1 + \frac{1}{\left(1-\frac{Y}{n}\right)^{n_1}}}\right)^{-1}$$

et en remarquant que la quantité

$$\frac{rn_1}{n}\frac{1}{1 + \frac{1}{\left(1-\frac{Y}{n}\right)^{n_1}}}$$

est proche de 0, un développement limité conduit à une nouvelle expression

$$\begin{aligned}\hat{\theta}^{ModA} &\simeq \frac{Y}{n}\left(1-\frac{rn_1}{n}\frac{1}{1+\frac{1}{\left(1-\frac{Y}{n}\right)^{n_1}}}\right)\\ &= \frac{Y}{n}\left(1-\frac{rn_1}{n}\frac{\left(1-\frac{Y}{n}\right)^{n_1}}{1+\left(1-\frac{Y}{n}\right)^{n_1}}\right).\end{aligned}$$

Cette nouvelle forme permet d'identifier le terme qui compense le biais induit par l'arrêt optionnel de l'expérience. Il dépend du rapport y/n avec une pondération définie par n_1/n.

En remarquant que la quantité

$$\left(1-\frac{Y}{n}\right)^{n_1}$$

est proche de 0, un nouveau développement limité conduit à un estimateur qui présente une forme linéaire en $(1-Y/n)$,

$$\begin{aligned}\hat{\theta}^{ModL} &= \frac{Y}{n}\left(1-\frac{rn_1}{n}\left(1-\frac{Y}{n}\right)^{n_1}\left(1-\left(1-\frac{Y}{n}\right)^{n_1}\right)\right)\\ &= \frac{Y}{n}\left(1-\frac{rn_1}{n}\left(1-\frac{Y}{n}\right)^{n_1}+\frac{rn_1}{n}\left(1-\frac{Y}{n}\right)^{2n_1}\right).\end{aligned}$$

L'estimateur $\hat{\theta}^{ModL}$ est dérivé à partir de $\hat{\theta}^{ModA}$. De la même façon, on peut dériver les estimateurs $\hat{\theta}^{ModL_{0.5}}$ et $\hat{\theta}^{ModL_1}$ à partir de $\hat{\theta}^{ModA_{0.5}}$ et $\hat{\theta}^{ModA_1}$,

$$\begin{aligned}\hat{\theta}^{ModL_{0.5}} &= \frac{Y}{n}\left(1-\frac{rn_1}{n}\left(1-\frac{Y}{n}\right)^{(n_1-0.5)}+\frac{rn_1}{n}\left(1-\frac{Y}{n}\right)^{2(n_1-0.5)}\right)\\ \hat{\theta}^{ModL_1} &= \frac{Y}{n}\left(1-\frac{rn_1}{n}\left(1-\frac{Y}{n}\right)^{(n_1-1)}+\frac{rn_1}{n}\left(1-\frac{Y}{n}\right)^{2(n_1-1)}\right).\end{aligned}$$

On représente dans la figure (3.10) les valeurs de $\hat{\theta}^{ModL}_{(m,y)}-y/n$, $\hat{\theta}^{ModL_{0.5}}_{(m,y)}-y/n$ et $\hat{\theta}^{ModL_1}_{(m,y)}-y/n$ dans $\mathcal{Y}_{(1,y)}=\{y=y_1; y=1,2,..,5\}$ et dans $\mathcal{Y}_{(2,y)}=\{y=y_1+y_2; y=0,1,..,5\}$. Les propriétés décrites pour $\hat{\theta}^{ModA}_{(m,y)}$, $\hat{\theta}^{ModA_{0.5}}_{(m,y)}$ et $\hat{\theta}^{ModA_1}_{(m,y)}$ sont conservées.

Dans le cas où le plan expérimental prévoit un arrêt de l'expérience en période 1 seulement si n_1 réponses sont observées (i.e $M=1$ si $Y_1=n_1$), seule la borne

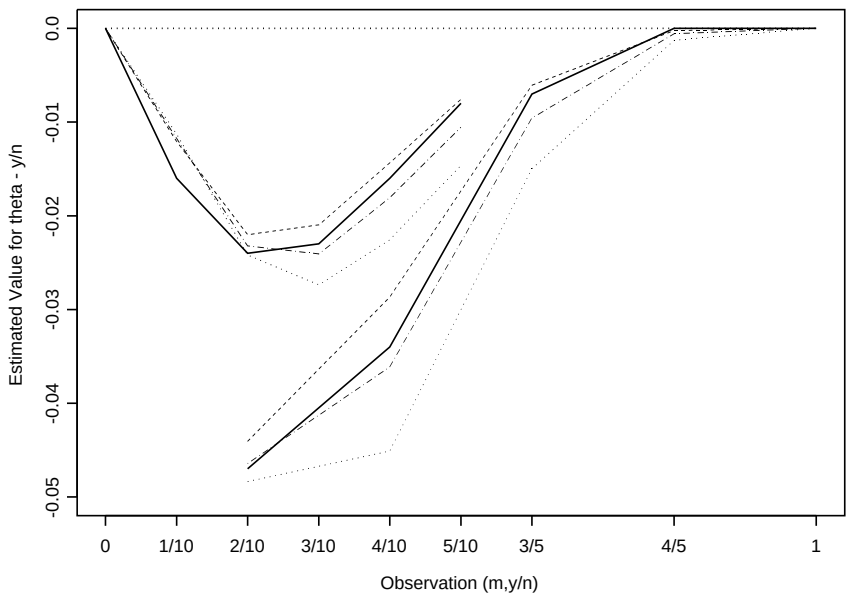

FIG. 3.10 – *Estimations* $\hat{\theta}^{Mod}_{(m,y)}$ *(—),* $\hat{\theta}^{ModL}_{(m,y)}$ *(- - -),* $\hat{\theta}^{ModL_{0.5}}_{(m,y)}$ *(· − ·) et* $\hat{\theta}^{ModL_1}_{(m,y)}$ *(· · ·) soustraites de* y/n *pour* $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 1, 2, .., 5\}$ *et* $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 5\}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

$S_{1,L} = n_1$ est à considérer. L'estimateur du Mode Approximé s'obtient en remplaçant θ par $1 - \theta$ dans l'équation (3.3.1). Il vient

$$\hat{\theta}^{ModA} = \frac{1 + g(1 - Y/n)}{1 + (n - Y)g(1 - Y/n)},$$

où

$$g(1 - Y/n) = \frac{1}{n_1}\left(1 + \left(\frac{n}{Y}\right)^{n_1}\right).$$

Profitant de sa forme linéaire, l'expression de $\hat{\theta}^{ModL}$ est encore plus directe

$$\hat{\theta}^{ModL} = \frac{Y}{n}\left(1 + \frac{n_1}{n}\left(\frac{Y}{n}\right)^{n_1} - \frac{n_1}{n}\left(\frac{Y}{n}\right)^{2n_1}\right).$$

3.3.4 Approximation du biais de l'estimateur du Mode

Une critique majeur formulée à l'encontre de l'EMV dans le contexte d'un plan expérimental séquentiel concerne le biais qu'il induit dans l'estimation ponctuelle.

On montre dans cette section que lorsque $S_{1,U} = 1$ et $n_1 = n_2$, le biais de l'estimateur du Mode est approché par la fonction

$$B_{Seq,\hat{\theta}^{Mod}}(\theta) = B_{Seq,\hat{\theta}^{MV}}(\theta) + \left(1 + \frac{(1-\theta)^{n_1}}{4}\right) \times \left[\frac{\mu_{2n_1-1}(1-\theta)}{n_1^{2n-1}} - \frac{\mu_{2n_1}(1-\theta)}{n_1^{2n_1}} - \frac{\mu_{n_1-0.5}(1-\theta)}{n_1^{n_1-0.5}} + \frac{\mu_{n_1+0.5}(1-\theta)}{n_1^{n_1+0.5}}\right],$$

où μ_q est le moment non-centré d'ordre q de Y.

On exprime le biais d'un estimateur $\hat{\theta}$ par

$$B_{Seq,\hat{\theta}}(\theta) = E_{Seq,\theta}(\hat{\theta}) - \theta,$$

où

$$\begin{aligned} E_{Seq,\theta}(\hat{\theta}) &= \sum_{y \in \mathcal{Y}_{(m,y)}} \hat{\theta}_{(m,y)} Pr((M,Y) = (m,y)) \\ &= \sum_{y \in \mathcal{Y}_{(1,y)}} \hat{\theta}_{(1,y)} Pr((M,Y) = (1,y)) + \sum_{y \in \mathcal{Y}_{(2,y)}} \hat{\theta}_{(2,y)} Pr((M,Y) = (2,y)) \\ &= \sum_{y \in \mathcal{Y}_{(1,y)}} \hat{\theta}_{(1,y)} Pr(Y = y | M = 1) Pr(M = 1) + \sum_{y \in \mathcal{Y}_{(2,y)}} \hat{\theta}_{(2,y)} Pr(Y = y | M = 2) Pr(M = 2) \end{aligned}$$

La probabilité

$$Pr((M,Y) = (1,y)) = Bin(n_1, \theta; y)$$

est associée à l'espace des observations

$$\mathcal{Y}_{(1,y)} = \{0, 1, .., S_{1,L}, S_{1,U}, S_{1,U} + 1, .., n_1\}$$

et la probabilité

$$Pr((M,Y) = (2,y)) = \sum_{i=max}^{y} Bin(n_1, \theta; i) Bin(n_2, \theta; y - i)$$

où $max = max(S_{1,L} + 1, y - n_2)$ à l'espace des observations

$$\mathcal{Y}_{(2,y)} = \{S_{1,L} + 1, S_{1,L} + 2, .., S_{1,U} - 1 + n_2\}.$$

Dans la suite de cette section, les formes analytiques des biais induits par $\hat{\theta}^{Mod}$ et par $\hat{\theta}^{MV}$ sont dérivées dans le contexte d'un plan séquentiel défini par une seule

borne $S_{1,U} = 1$, avec la contrainte supplémentaire $n_1 = n_2$. La généralisation des conclusions et des propriétés de $\hat{\theta}^{ModA}$ est discutée dans la section suivante.

Avec $S_{1,U} = 1$ et $n_1 = n_2$, on a

$$\begin{aligned} E_{Seq,\theta}\left[\hat{\theta}\right] &= \left[\sum_{y=1}^{n_1} \hat{\theta}_{(1,y)} \binom{n_1}{y} \theta^y (1-\theta)^{n_1-y}\right] + \left[\sum_{y=1}^{n_1} \hat{\theta}_{(2,y)} \binom{n_1}{y} \theta^y (1-\theta)^{n_1-y}\right] (1-\theta)^{n_1} \\ &= \sum_{y=1}^{n_1} \binom{n_1}{y} \theta^y (1-\theta)^{n_1-y} \underbrace{\left[\hat{\theta}_{(1,y)} + \hat{\theta}_{(2,y)} (1-\theta)^{n_1}\right]}_{D(\hat{\theta})}. \end{aligned} \tag{3.3.4}$$

Pour $\hat{\theta}^{MV} = Y/N$, on a

$$D(\hat{\theta}^{MV}) = \frac{y}{n_1} + (1-\theta)^{n_1} \frac{y}{2n_1} = \frac{y}{n_1}\left(1 + \frac{(1-\theta)^{n_1}}{2}\right).$$

On déduit le biais

$$\begin{aligned} B_{Seq,\hat{\theta}^{MV}}(\theta) &= \sum_{y=1}^{n_1} \binom{n_1}{y} \theta^y (1-\theta)^{n_1-y} \frac{y}{n_1} \frac{(1-\theta)^{n_1}}{2} \\ &= \frac{1}{2}\theta(1-\theta)^{n_1}. \end{aligned}$$

Le biais $B_{Seq,\hat{\theta}^{MV}}(\theta)$ est une fonction à valeurs positives. Les courbes en fonction de θ sont représentées figure (3.11) pour $n_1 = n_2 = 2, 5, 10$.

La dérivée en θ de $B_{Seq,\hat{\theta}^{MV}}(\theta)$ permet d'obtenir la valeur de θ_{Max} pour laquelle le biais est maximum. De l'expression

$$\frac{d}{d\theta} B_{Seq,\hat{\theta}^{MV}}(\theta) = \frac{1}{2}\left[(1-\theta)^{n_1} - n_1\theta(1-\theta)^{n_1}\right] = 1 - \theta - n_1\theta = 0,$$

il vient

$$\theta_{Max} = \frac{1}{1+n_1}.$$

La valeur du biais maximum est

$$\begin{aligned} B_{Seq,\hat{\theta}^{MV}}(\theta_{Max}) &= \frac{1}{2(1+n_1)}\left(1 - \frac{1}{1+n_1}\right)^{n_1} \\ &= \frac{1}{2}\frac{n_1^{n_1}}{(1+n_1)^{n_1+1}}. \end{aligned}$$

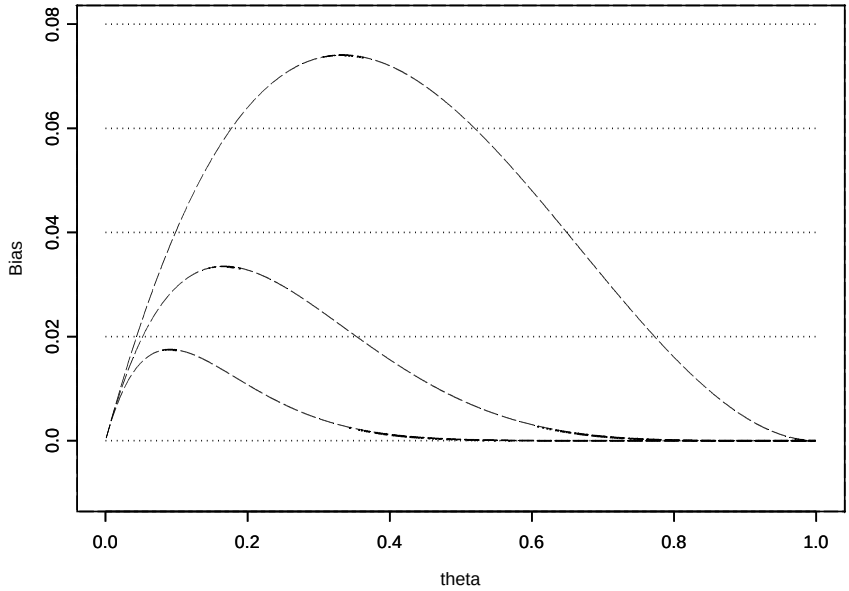

FIG. 3.11 – *Biais de* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1, S_{1,U} = 1, n_2)$ *avec* $n_1 = n_2 = 2, 5, 10$

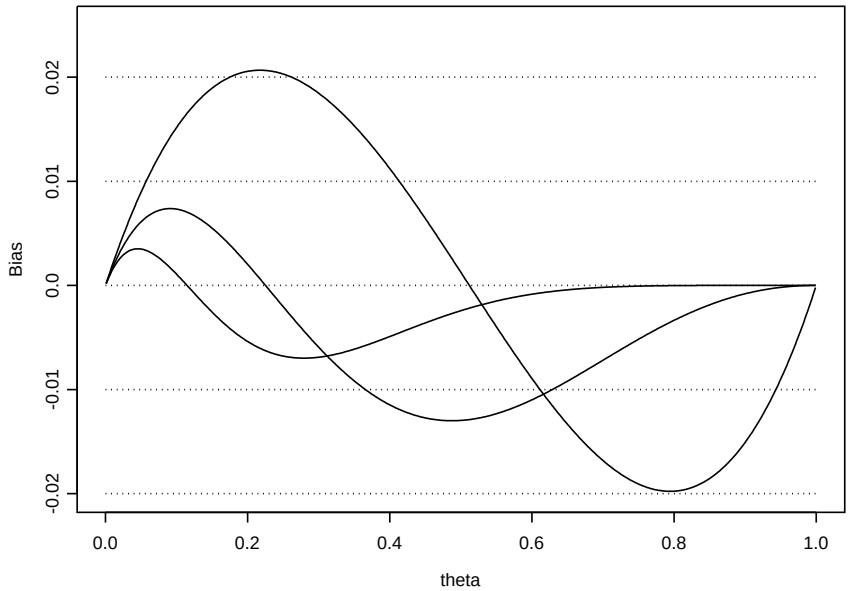

FIG. 3.12 – *Biais de* $\hat{\theta}^{Mod}$ *dans le plan séquentiel* $(n_1, S_{1,U} = 1, n_2)$ *avec* $n_1 = n_2 = 2, 5, 10$

Pour $S_{1,U} = 1$, θ_{Max} et $B_{Seq,\hat{\theta}^{MV}}(\theta_{Max})$ sont 2 fonctions décroissantes en n_1. Plus l'effectif $n_1 = n_2$ augmente plus le maximum du biais se rapproche de $\theta = 0$. On trouve les valeurs numériques pour $n_1 = n_2 = 2, 5, 10$.

$$\begin{array}{lll} n_1 = n_2 = 2 & \theta_{Max} = \dfrac{1}{3} & B_{Seq,\hat{\theta}^{MV}}(\theta_{Max}) = 0.0741 \\ n_1 = n_2 = 5 & \theta_{Max} = \dfrac{1}{6} & B_{Seq,\hat{\theta}^{MV}}(\theta_{Max}) = 0.0335 \\ n_1 = n_2 = 10 & \theta_{Max} = \dfrac{1}{11} & B_{Seq,\hat{\theta}^{MV}}(\theta_{Max}) = 0.0175. \end{array}$$

En comparaison, le biais de l'estimateur du Mode $B_{Seq,\hat{\theta}^{Mod}}(\theta)$ est représenté figure (3.12) pour $n_1 = n_2 = 2, 5, 10$. Empiriquement, la fonction $B_{Seq,\hat{\theta}^{Mod}}(\theta)$ est positive pour θ compris entre 0 et une valeur proche de $1/n_1$, puis devient négative. Deux propriétés majeures de $\hat{\theta}^{Mod}$ sont montrées graphiquement ; le biais maximum induit est en valeur absolue très inférieur à celui associé à $\hat{\theta}^{Mod}$ et ce biais s'annule pour une valeur proche de $\theta = 1/n_1$.

Pour $n_1 = n_2 = 2, 5, 10$, on trouve les valeurs

$$\begin{array}{lll} n_1 = n_2 = 2 & \theta_{Max} = 0.217 & B_{Seq,\hat{\theta}^{Mod}}(\theta_{Max}) = 0.0207 \\ & \theta_0 = 0.511 & B_{Seq,\hat{\theta}^{Mod}}(\theta_0) = 0 \\ & \theta_{Min} = 0.794 & B_{Seq,\hat{\theta}^{Mod}}(\theta_{Min}) = -0.0198 \\ n_1 = n_2 = 5 & \theta_{Max} = 0.091 & B_{Seq,\hat{\theta}^{Mod}}(\theta_{Max}) = 0.0074 \\ & \theta_0 = 0.225 & B_{Seq,\hat{\theta}^{Mod}}(\theta_0) = 0 \\ & \theta_{Min} = 0.487 & B_{Seq,\hat{\theta}^{Mod}}(\theta_{Min}) = -0.0130 \\ n_1 = n_2 = 10 & \theta_{Max} = 0.045 & B_{Seq,\hat{\theta}^{Mod}}(\theta_{Max}) = 0.0035 \\ & \theta_0 = 0.115 & B_{Seq,\hat{\theta}^{Mod}}(\theta_0) = 0 \\ & \theta_{Min} = 0.280 & B_{Seq,\hat{\theta}^{Mod}}(\theta_{Min}) = -0.0070. \end{array}$$

Concernant les approximations de l'estimateur du Mode, la figure (3.13) permet de visualiser le biais de $\hat{\theta}^{Mod}$, $\hat{\theta}^{ModA}$, $\hat{\theta}^{ModA_{0.5}}$ et $\hat{\theta}^{ModA_1}$ dans le plan séquentiel $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$. La courbe de $B_{Seq,\hat{\theta}^{ModA_{0.5}}}(\theta)$ est très proche de celle de

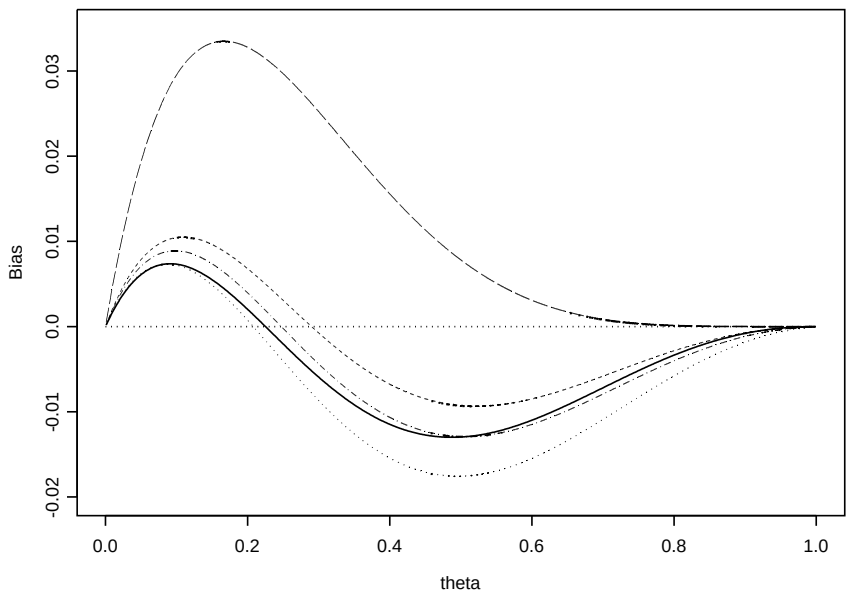

FIG. 3.13 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{Mod}$ *(—),* $\hat{\theta}^{ModA}$ *(- - -),* $\hat{\theta}^{ModA_{0.5}}$ *(· – ·) et* $\hat{\theta}^{ModA_1}$ *(· · ·) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

$B_{Seq,\hat{\theta}^{Mod}}(\theta)$, alors que les courbes de $B_{Seq,\hat{\theta}^{ModA}}(\theta)$ et $B_{Seq,\hat{\theta}^{ModA_1}}(\theta)$ respectivement la surestime et la sous-estime légèrement.

La détermination d'une forme analytique pour approcher la fonction $B_{Seq,\hat{\theta}^{Mod}}(\theta)$ nécessite l'écriture de l'approximation linéaire

$$\hat{\theta}^{ModL} = \frac{Y}{n} - \frac{Yn_1}{n^2}\left(1 - \frac{Y}{n}\right)^{n_1} + \frac{n_1 Y}{n^2}\left(1 - \frac{Y}{n}\right)^{2n_1}$$

pour $m = 1$ et $m = 2$.

On déduit les estimations

$$\hat{\theta}^{ModL}_{(1,y)} = \frac{y}{n_1} - \frac{y}{n_1}\left(1 - \frac{y}{n_1}\right)^{n_1} + \frac{y}{n_1}\left(1 - \frac{y}{n_1}\right)^{2n_1}$$
$$\hat{\theta}^{ModL}_{(2,y)} = \frac{y}{2n_1} - \frac{y}{4n_1}\left(1 - \frac{y}{2n_1}\right)^{n_1} + \frac{y}{4n_1}\left(1 - \frac{y}{2n_1}\right)^{2n_1}.$$

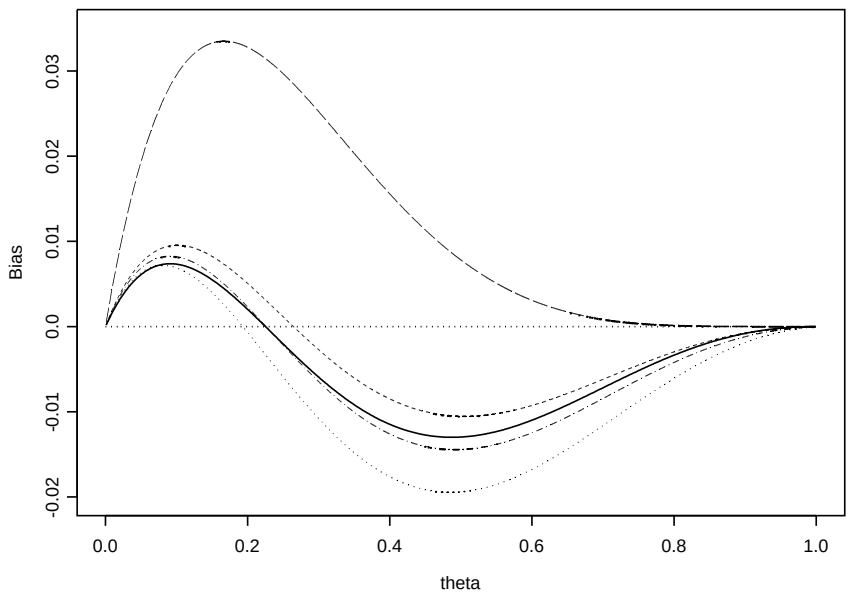

FIG. 3.14 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{Mod}$ *(—),* $\hat{\theta}^{ModL}$ *(- - -),* $\hat{\theta}^{ModL_{0.5}}$ *(· – ·) et* $\hat{\theta}^{ModL_1}$ *(· · ·) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$

En reprenant l'équation (3.3.4) et en remplaçant $\hat{\theta}_{(m,y)}$ par $\hat{\theta}^{ModL}_{(m,y)}$, il vient

$$\begin{aligned} D(\hat{\theta}^{ModL}_{(m,y)}) &= \hat{\theta}^{ModL}_{(1,y)} + \hat{\theta}^{ModL}_{(2,y)}(1-\theta)^{n_1} \\ &= \frac{y}{n_1}[1 + \frac{1}{2}(1-\theta)^{n_1} - \left(1 - \frac{y}{n_1}\right)^{n_1} + \left(1 - \frac{y}{n_1}\right)^{2n_1} \\ &\quad - \frac{(1-\theta)^{n_1}}{4}\left(1 - \frac{y}{2n_1}\right)^{n_1} + \frac{(1-\theta)^{n_1}}{4}\left(1 - \frac{y}{2n_1}\right)^{2n_1}]. \end{aligned}$$

En remplaçant également avec $\hat{\theta}^{ModL_{0.5}}_{(y)}$ et $\hat{\theta}^{ModL_1}_{(y)}$, l'expression (3.3.4) conduit aux biais associés à $\hat{\theta}^{ModL}$, $\hat{\theta}^{ModL_{0.5}}$ et $\hat{\theta}^{ModL_1}$. Les courbes de $B_{Seq,\hat{\theta}^{ModL}}(\theta)$, $B_{Seq,\hat{\theta}^{ModL_{0.5}}}(\theta)$ et $B_{Seq,\hat{\theta}^{ModL_1}}(\theta)$ sont représentées figure (3.14) pour $n_1 = n_2 = 5$ et $S_{1,U} = 1$. Les caractéristiques de $\hat{\theta}^{ModL}$, $\hat{\theta}^{ModL_{0.5}}$ et $\hat{\theta}^{ModL_1}$ dans $\mathcal{Y}_{(m,y)}$ se traduisent par la relation

$$B_{Seq,\hat{\theta}^{ModL_1}}(\theta) < B_{Seq,\hat{\theta}^{ModL_{0.5}}}(\theta) < B_{Seq,\hat{\theta}^{ModL}}(\theta),$$

où le biais de $\hat{\theta}^{ModL_{0.5}}$ est très proche de celui de $\hat{\theta}^{Mod}$.

Dans le but de factoriser la forme précédente, on remplace les termes en

$$\left(1 - \frac{y}{2n_1}\right) \quad \text{par} \quad \left(1 - \frac{y}{n_1}\right).$$

La *va* $D(\hat{\theta}^{ModL})$ s'exprime maintenant par

$$D(\hat{\theta}^{ModL}) \simeq \frac{Y}{n_1} + \frac{Y}{2n_1}(1-\theta)^{n_1} + \left(1+\frac{(1-\theta)^{n_1}}{4}\right)\left(\frac{Y}{n_1}\left(1-\frac{Y}{n_1}\right)^{2n_1} - \frac{Y}{n_1}\left(1-\frac{Y}{n_1}\right)^{n_1}\right),$$

et son espérance s'écrit

$$\begin{aligned}
E_\theta\left[D(\hat{\theta}^{ModL})\right] &\simeq \theta + \frac{\theta}{2}(1-\theta)^{n_1} + \left(1+\frac{(1-\theta)^{n_1}}{4}\right) \\
&\quad \times \left(E_\theta\left[\frac{Y}{n_1}\left(1-\frac{Y}{n_1}\right)^{2n_1}\right] - E_\theta\left[\frac{Y}{n_1}\left(1-\frac{Y}{n_1}\right)^{n_1}\right]\right) \\
&= \theta + \frac{\theta}{2}(1-\theta)^{n_1} + \left(1+\frac{(1-\theta)^{n_1}}{4}\right) \\
&\quad \times \left(\frac{1}{n_1^{2n_1+1}}E_\theta\left[Y(n_1-Y)^{2n_1}\right] - \frac{1}{n_1^{n_1+1}}E_\theta\left[Y(n_1-Y)^{n_1}\right]\right) \\
&= \theta + \frac{\theta}{2}(1-\theta)^{n_1} + \left(1+\frac{(1-\theta)^{n_1}}{4}\right) \\
&\quad \times \left(\frac{1}{n_1^{2n_1+1}}E_{1-\theta}\left[(n_1-Y)Y^{2n_1}\right] - \frac{1}{n_1^{n_1+1}}E_{1-\theta}\left[(n_1-Y)Y^{n_1}\right]\right).
\end{aligned}$$

En notant μ_q le moment non-centré d'ordre q de Y

$$\mu_q(\theta) = E_\theta[Y^q] = \sum_{y=0}^{n_1} y^q \binom{n_1}{y}\theta^y(1-\theta)^{n_1-y},$$

on déduit l'approximation du biais de $\hat{\theta}^{ModL}$

$$\begin{aligned}
B^{App}_{Seq,\hat{\theta}^{ModL}}(\theta) &= E_{Seq,\theta}(\hat{\theta}^{ModL}) - \theta \\
&= \frac{\theta}{2}(1-\theta)^{n_1} + \left(1+\frac{(1-\theta)^{n_1}}{4}\right) \\
&\quad \times [\frac{1}{n_1^{2n_1}}E_{1-\theta}(Y^{2n_1}) - \frac{1}{n_1^{2n_1+1}}E_{1-\theta}(Y^{2n_1+1}) \\
&\quad - \frac{1}{n_1^{n_1}}E_{1-\theta}(Y^{n_1}) + \frac{1}{n_1^{n_1+1}}E_{1-\theta}(Y^{n_1+1})] \\
&= B_{Seq,\hat{\theta}^{MV}}(\theta) + \left(1+\frac{(1-\theta)^{n_1}}{4}\right) \\
&\quad \times \left[\frac{\mu_{2n_1}(1-\theta)}{n_1^{2n_1}} - \frac{\mu_{2n_1+1}(1-\theta)}{n_1^{2n_1+1}} - \frac{\mu_{n_1}(1-\theta)}{n_1^{n_1}} + \frac{\mu_{n_1+1}(1-\theta)}{n_1^{n_1+1}}\right].
\end{aligned}$$

La formule du biais $B^{App}_{Seq,\hat{\theta}^{ModL}}(\theta)$ est dérivée à partir de $\hat{\theta}^{ModL}$. De la même façon, on peut dériver les formules pour $B^{App}_{Seq,\hat{\theta}^{ModL_{0.5}}}(\theta)$ et $B^{App}_{Seq,\hat{\theta}^{ModL_1}}(\theta)$ à partir de $\hat{\theta}^{ModL_{0.5}}$ et

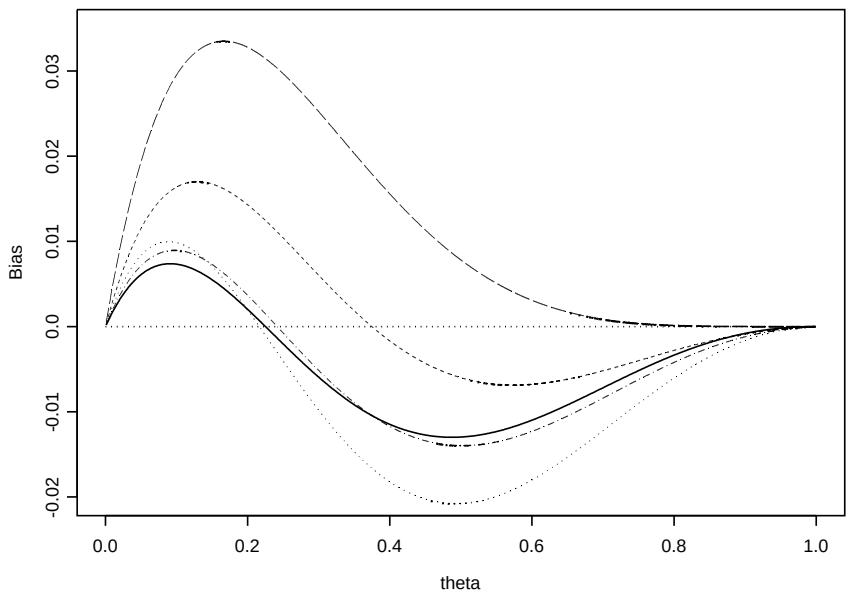

FIG. 3.15 – *Approximation du biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{Mod}$ *(—),* $\hat{\theta}^{ModL}$ *(- - -),* $\hat{\theta}^{ModL_{0.5}}$ *(· – ·) et* $\hat{\theta}^{ModL_1}$ *(· · ·) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$

$\hat{\theta}^{ModL_1}$. On obtient alors

$$\begin{aligned}
B^{App}_{Seq,\hat{\theta}^{ModL_{0.5}}}(\theta) &= B_{Seq,\hat{\theta}^{MV}}(\theta) + \left(1 + \frac{(1-\theta)^{n_1}}{4}\right) \\
&\quad \times \left[\frac{\mu_{2n_1-1}(1-\theta)}{n_1^{2n-1}} - \frac{\mu_{2n_1}(1-\theta)}{n_1^{2n_1}} - \frac{\mu_{n_1-0.5}(1-\theta)}{n_1^{n_1-0.5}} + \frac{\mu_{n_1+0.5}(1-\theta)}{n_1^{n_1+0.5}}\right] \\
B^{App}_{Seq,\hat{\theta}^{ModL_1}}(\theta) &= B_{Seq,\hat{\theta}^{MV}}(\theta) + \left(1 + \frac{(1-\theta)^{n_1}}{4}\right) \\
&\quad \times \left[\frac{\mu_{2n_1-2}(1-\theta)}{n_1^{2n_1-2}} - \frac{\mu_{2n_1-1}(1-\theta)}{n_1^{2n_1-1}} - \frac{\mu_{n_1-1}(1-\theta)}{n_1^{n_1-1}} + \frac{\mu_{n_1}(1-\theta)}{n_1^{n_1}}\right].
\end{aligned}$$

Les courbes de $B^{App}_{Seq,\hat{\theta}^{ModL}}(\theta)$ et $B^{App}_{Seq,\hat{\theta}^{ModL_1}}(\theta)$ dans la figure (3.15) sont plus éloignées de $B_{Seq,\hat{\theta}^{Mod}}(\theta)$ que le sont respectivement $B_{Seq,\hat{\theta}^{ModL}}(\theta)$ et $B_{Seq,\hat{\theta}^{ModL_1}}(\theta)$ dans la figure (3.14). En revanche, $B^{App}_{Seq,\hat{\theta}^{ModL_{0.5}}}(\theta)$ approche $B_{Seq,\hat{\theta}^{Mod}}(\theta)$ avec la même précision que $B_{Seq,\hat{\theta}^{ModL_{0.5}}}(\theta)$.

3.3.5 Etude comparative du biais et de l'efficacité relative lorsque $n_1 = n_2$

Les performances de l'estimateur du Mode peuvent être évaluées relativement à l'estimateur sans-biais de variance uniformément minimum $\hat{\theta}^{VUMSB}$. Pour un plan séquentiel avec deux périodes, l'expression de $\hat{\theta}^{VUMSB}$ décrite dans la section 3.1 se simplifie en

$$\hat{\theta}^{VUMSB} = \begin{cases} \frac{Y}{n_1} & m = 1 \\ \dfrac{\sum_{Y_1=(S_{1,L}+1)\vee(Y-n_2)}^{Y\wedge(S_{1,U}-1)} \binom{n_1-1}{Y_1-1}\binom{n_2}{Y-Y_1}}{\sum_{Y_1=(S_{1,L}+1)\vee(Y-n_2)}^{Y\wedge(S_{1,U}-1)} \binom{n_1}{Y_1}\binom{n_2}{Y-Y_1}} & m = 2, \end{cases}$$

où $a \wedge b = min(a, b)$ et $a \vee b = max(a, b)$ et $\binom{a}{b} = 0$ si $b < 0$.

En présence d'une seule borne supérieure $S_{1,U} = 1$, on déduit

$$\hat{\theta}^{VUMSB} = \begin{cases} \frac{Y}{n_1} & m = 1 \\ 0 & m = 2. \end{cases} \tag{3.3.5}$$

L'estimation $\hat{\theta}^{VUMSB}_{(m,y)}$ n'est pas cohérente avec une observation $(m, y) = (2, y)$ où le nombre de réponses y est non nul.

Par définition, $\hat{\theta}^{VUMSB}$ est un estimateur sans-biais (i.e. $E_{Seq,\theta}[\hat{\theta}^{VUMSB}] = \theta \quad \forall\theta \in [0, 1]$). Cette propriété est à compléter par l'étude de l'erreur quadratique moyenne (EQM), qui pour $\hat{\theta}^{VUMSB}$ se confond avec la variance.

Pour un estimateur $\hat{\theta}$, on a

$$EQM(\hat{\theta}) = \sum_{(m,y)\in\mathcal{S}} (\hat{\theta} - \theta)^2 p(m, y|\theta),$$

où $\mathcal{S}$ est le support de (M, Y).

La figure (3.16) illustre l'erreur quadratique moyenne des estimateurs $\hat{\theta}^{MV}$, $\hat{\theta}^{Mod}$, $\hat{\theta}^{ModA}$ et $\hat{\theta}^{VUMSB}$ dans le plan séquentiel $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$. La forme symétrique de la courbe de $EQM(\hat{\theta}^{VUMSB})$ et la différence avec celle de $EQM(\hat{\theta}^{MV})$ s'explique par la nullité des valeurs de $\hat{\theta}^{VUMSB}_{(2,y)}$.

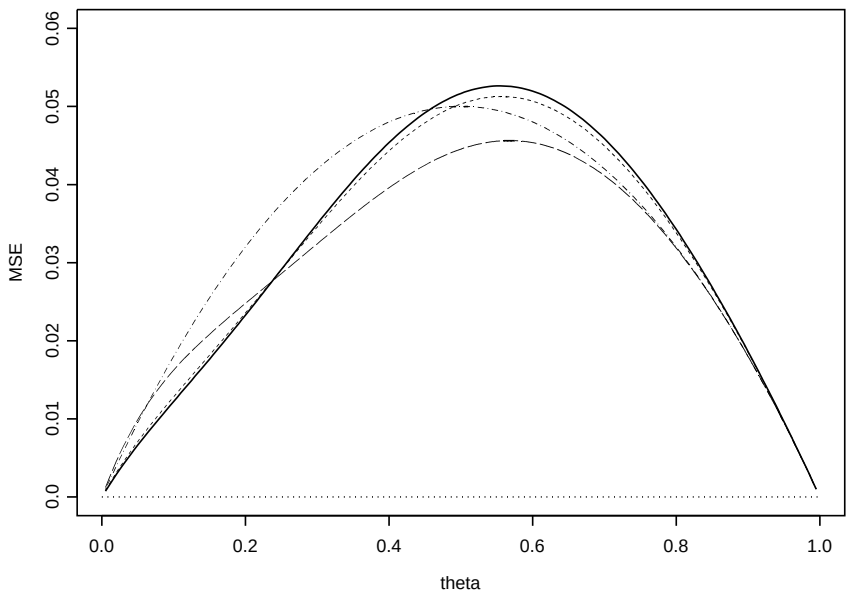

FIG. 3.16 – *EQM de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{Mod}$ *(—),* $\hat{\theta}^{ModA}$ *(- - -) et* $\hat{\theta}^{VUMSB}$ *(· – ·) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

L'efficacité relative qui mesure l'erreur quadratique moyenne d'un estimateur $\hat{\theta}$ relativement à celle de $\hat{\theta}^{MV}$, prise comme référence, est notée

$$ER_{\hat{\theta}} = \frac{EQM(\hat{\theta}^{MV})}{EQM(\hat{\theta})}.$$

Une valeur de $ER_{\hat{\theta}} > 1$ indique une meilleure efficacité de $\hat{\theta}$ en terme d'erreur quadratique moyenne relativement à $\hat{\theta}_{MV}$. La figure (3.17) décrit l'efficacité relative des estimateurs $\hat{\theta}^{Mod}$, $\hat{\theta}^{ModA}$ et $\hat{\theta}^{VUMSB}$ dans le plan séquentiel $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$.

Excepté pour $\theta < 0.1$, on a $ER_{\hat{\theta}^{VUMSB}} < 1$. Jung (2004) [39] mentionne que le surcroît de variance pour $\hat{\theta}^{VUMSB}$ relativement à $EQM(\hat{\theta}^{MV})$ est le prix à payer pour l'estimation sans-biais. L'estimateur $\hat{\theta}^{Mod}$ subit la même tendance mais dans une moindre mesure : la courbe de $ER_{\hat{\theta}^{Mod}}$ reste au dessus de $min(ER_{\hat{\theta}^{VUMSB}})$.

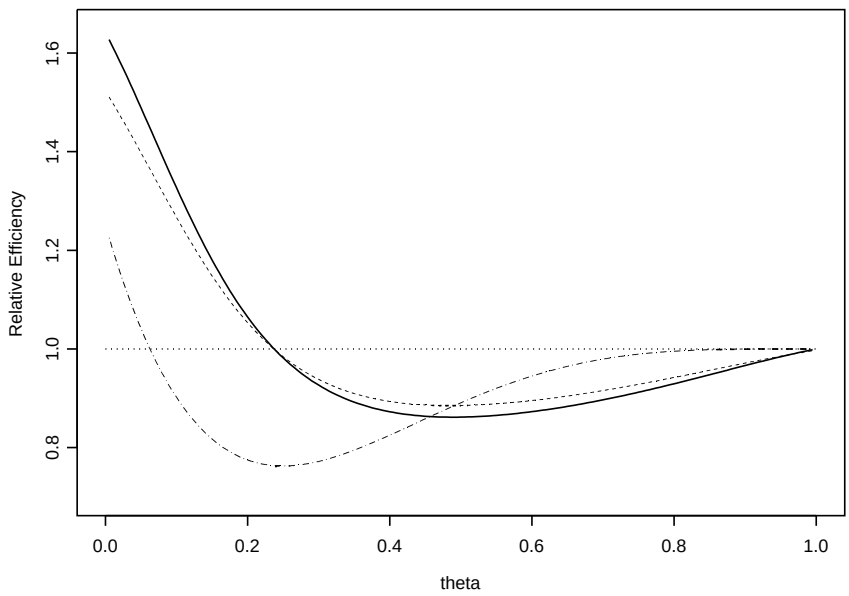

FIG. 3.17 – *Efficacité Relative de* $\hat{\theta}^{Mod}$ *(—),* $\hat{\theta}^{ModA}$ *(- - -) et* $\hat{\theta}^{VUMSB}$ *(· — ·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

3.3.6 Etude comparative du biais lorsque $n_1 \neq n_2$

On termine l'étude des différents estimateurs lorsque $S_{1,U} = 1$ par une analyse du biais et de l'erreur quadratique moyenne pour des valeur $r \neq 1$. La figure (3.18) décrit le biais de $\hat{\theta}^{MV}$ et $\hat{\theta}^{Mod}$ pour des valeurs $n_1 = 5$ associées à $n_2 = 1, 2, .., 10$. Le biais de $\hat{\theta}^{MV}$ est une fonction croissante en n_2 mais l'augmentation de $B_{Seq,\hat{\theta}^{MV}}(\theta)$ diminue à mesure que n_2 augmente.

Le biais de $\hat{\theta}^{Mod}$ est positif pour les valeurs de θ comprises entre 0 et $1/N_1$ approximativement. La courbe de $B_{Seq,\hat{\theta}^{Mod}}(\theta)$ croise l'axe horizontal le plus tardivement lorsque $n_2 = 1$. Le point de croisement se rapproche de $\theta = 0$ à mesure que n_2 augmente, en même temps que le biais négatif augmente. A l'inverse de $\hat{\theta}^{MV}$, l'augmentation du biais négatif semble évoluer linéairement en fonction de $n_2 \leq 10$. Enfin, les courbes de $B_{Seq,\hat{\theta}^{ModA}}(\theta)$ et de $B_{Seq,\hat{\theta}^{Mod}}(\theta)$ sont proches pour $r \leq 1$ (i.e. $n_2 = 1, 2, .., 5$), mais se différencient lorsque $r > 1$.

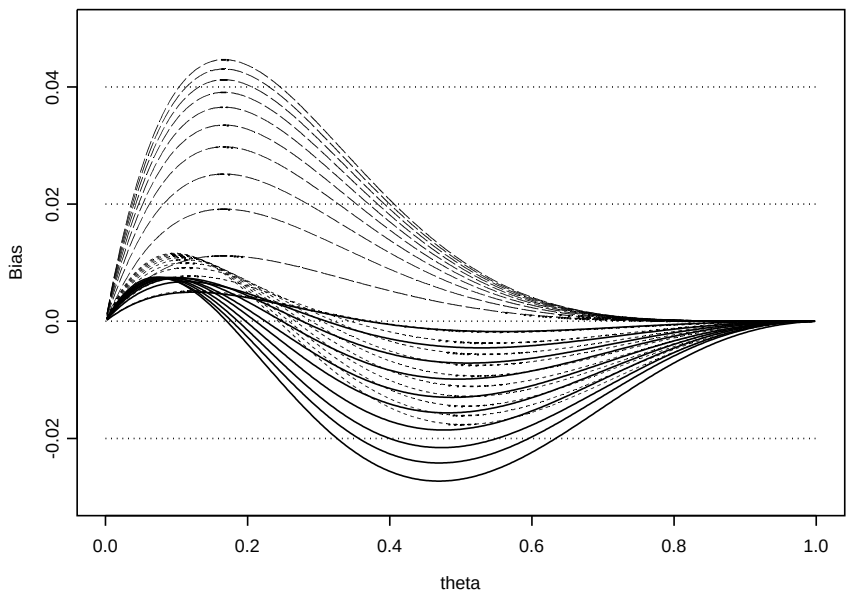

FIG. 3.18 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{Mod}$ *(—) et* $\hat{\theta}^{ModA}$ *(- - -) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 1, 2, .., 10)$

L'efficacité de $\hat{\theta}^{Mod}$ et $\hat{\theta}^{VUMSB}$ relativement à $\hat{\theta}^{MV}$ est décrite dans figure (3.19). Pour les deux estimateurs, les courbes associées à $n_1 = 1$ sont les plus proches de l'axe horizontal. On note la supériorité de $\hat{\theta}^{Mod}$ par rapport à $\hat{\theta}^{VUMSB}$ en terme d'écart quadratique moyen lorsque $\theta < 0.5$. La figure (3.20) montre une efficacité relative supérieure de l'approximation $\hat{\theta}^{ModA}$ par rapport à $\hat{\theta}^{Mod}$ lorsque θ est proche de 0.

Cette première étude a permis de montrer la très bonne performance de l'estimateur du mode en terme de biais, d'écart quadratique et d'approximation de $\hat{\theta}^{Mod}$ par $\hat{\theta}^{ModA}$ pour un rapport $r \leq 1$. Pour $r > 1$, la dégradation du biais constatée dans la figure (3.18) est due au fait que la fonction $B_{Seq,\hat{\theta}^{Mod}}(\theta)$ tarde a trouver son asymptote en n_2. La figure (3.21) présente les courbes de $B_{Seq,\hat{\theta}^{MV}}(\theta)$ et $B_{Seq,\hat{\theta}^{Mod}}(\theta)$ pour des valeurs allant jusqu'à $n_2 = 150$ $(r = 30)$.

Pour des grandes valeurs du rapport $r = n_2/n_1$, le biais négatif maximum de $\hat{\theta}^{Mod}$

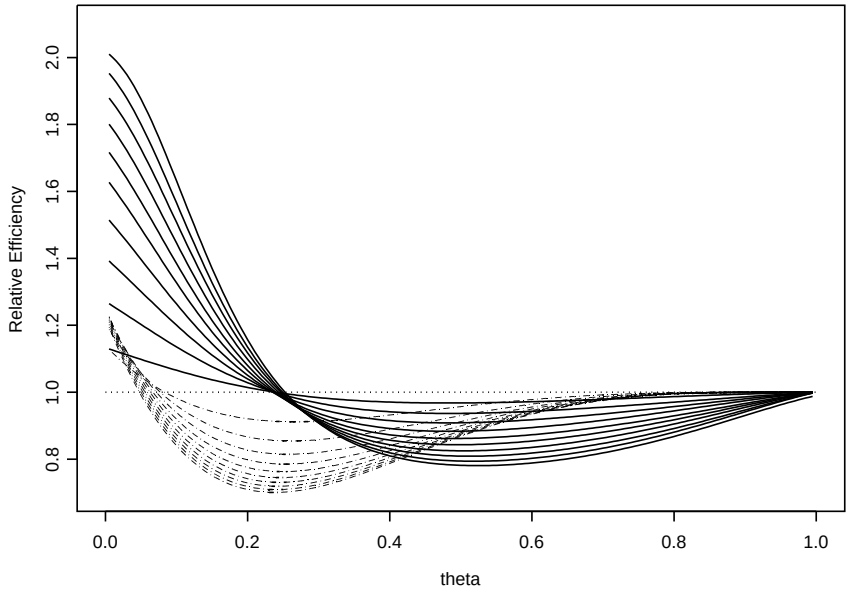

FIG. 3.19 – *Efficacité Relative de* $\hat{\theta}^{Mod}$ *(—),* $\hat{\theta}^{VUMSB}$ *(·−·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 1, 2, .., 10)$

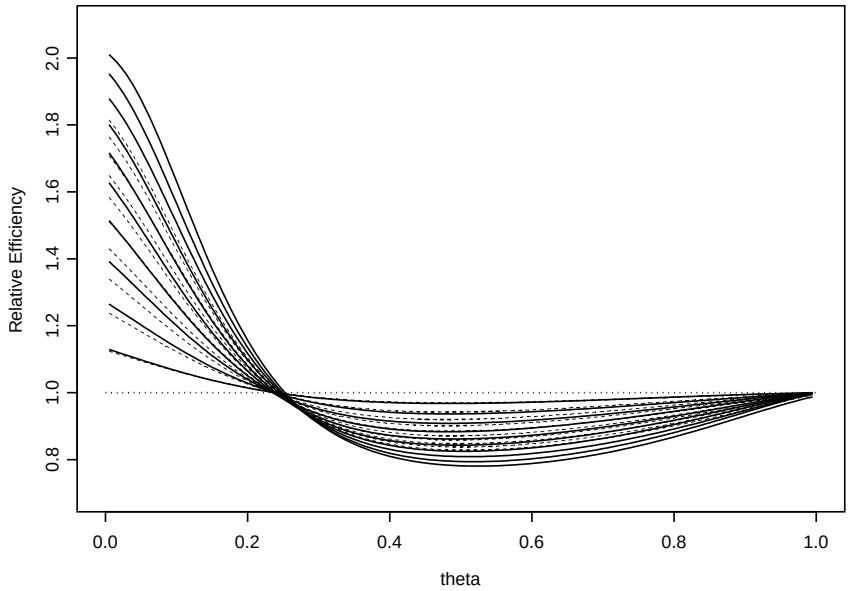

FIG. 3.20 – *Efficacité Relative de* $\hat{\theta}^{Mod}$ *(—) et* $\hat{\theta}^{ModA}$ *(- - -) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 1, 2, .., 10)$

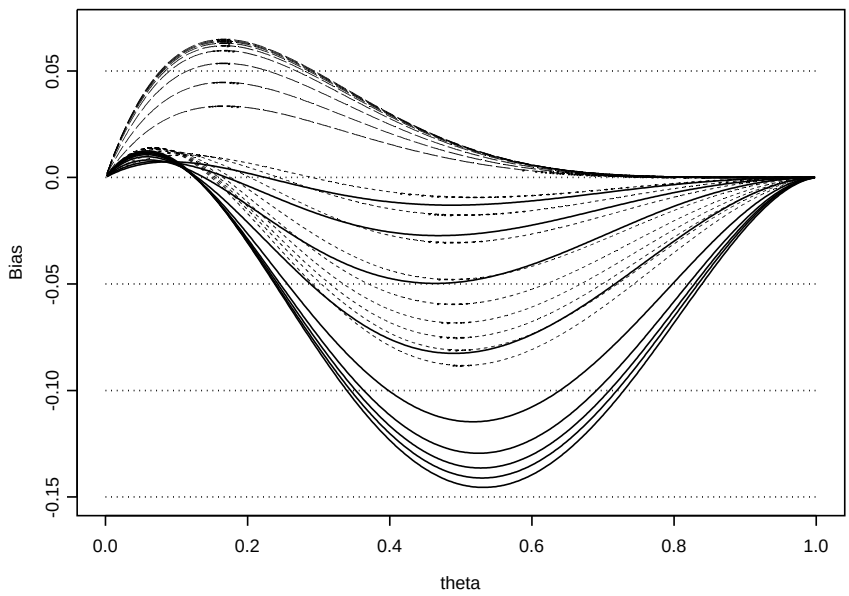

FIG. 3.21 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{Mod}$ *(—) et* $\hat{\theta}^{ModA}$ *(- - -) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5, 10, 20, 40, .., 120, 150)$

va au delà du biais positif maximum de $\hat{\theta}^{MV}$. Dans ce cas, on a la composante $(1 + rCBin(n_1, \theta; J)) \simeq rCBin(n_1, \theta; J)$. De plus, lorsque $S_{1,U} = 1$ (i.e. $J = \{0\}$), la loi a posteriori $Be^J(1 + y, 1 + n - y, 1; n_1, r; \{0\})$ a une densité qui s'écrit

$$\begin{aligned} p(\theta|d_{Seq}, y) &\propto \theta^y (1-\theta)^{n-y}(1 + rCBin(n_1, \theta; \{0\})) \\ &\propto r\theta^y (1-\theta)^{n-y} CBin(n_1, \theta; \{0\}) \\ &\propto \theta^y (1-\theta)^{n-y}(1-\theta)^{n_1} \\ &= \theta^y (1-\theta)^{n+n_1-y}. \end{aligned}$$

Elle tend donc vers le noyau d'une loi $Be(y + 1, n + n_1 - y + 1)$.

En utilisant les caractéristiques standards de la loi Bêta, l'estimateur du mode devient

$$\hat{\theta}^{Mod} = \frac{Y}{N + n_1},$$

ce qui affecte fortement l'estimation $\hat{\theta}^{Mod}_{(m,y)}$ lorsque $m = 1$.

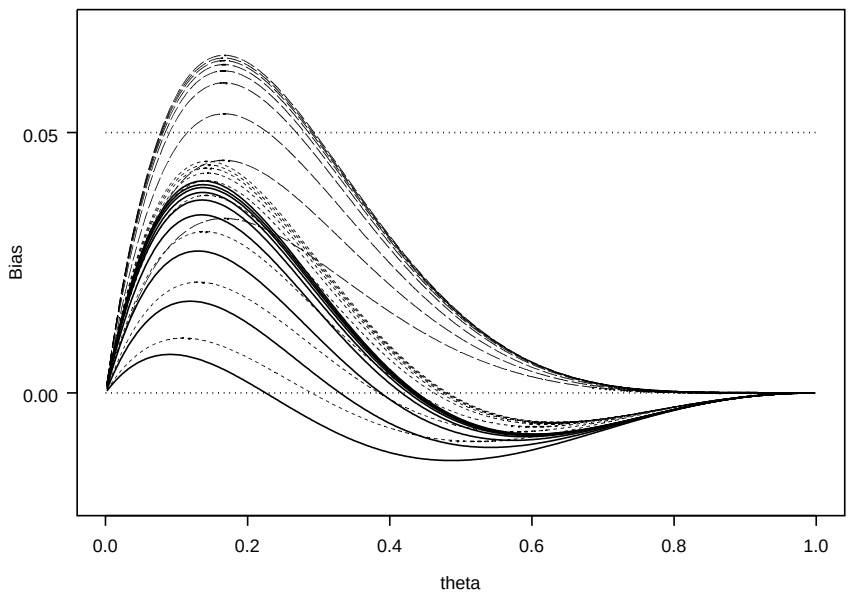

FIG. 3.22 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{Mod_{R=1}}$ *(—) et* $\hat{\theta}^{ModA_{R=1}}$ *(- - -) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5, 10, 20, 40, .., 120, 150)$

3.3.7 Dérivation de l'estimateur du Mode Corrigé

En vue d'améliorer les performances en terme de biais et d'approximation de $\hat{\theta}^{Mod}$ par $\hat{\theta}^{ModA}$, une correction de $\hat{\theta}^{Mod}$ est suggérée lorsque $r > 1$. L'idée est de limiter l'influence de $CBin(n_1, \theta; J)$ dans la composante $(1 + rCBin(n_1, \theta; J))$ en bornant r par valeur supérieure. On note $\hat{\theta}^{Mod_R}$ l'estimateur correspondant à une valeur de $r = R$ fixée et $\hat{\theta}^{ModA_R}$ son approximation que l'on obtient directement de la formule (3.3.3)

$$\hat{\theta}^{ModA_R} = \frac{Y}{n}\left(1 + \frac{Rn_1}{n}\frac{1}{1 + \left(1 - \frac{Y}{n}\right)^{-n_1}}\right)^{-1} \qquad \text{avec R fixé.}$$

Le biais associé à $\hat{\theta}^{Mod_R}$ et $\hat{\theta}^{ModA_R}$ est présenté dans la figure (3.22) pour la valeur $R = 1$ et $n_2 = 5, ..150$. La limitation de l'influence de $CBin(n_1, \theta; J)$ dans $\hat{\theta}^{Mod_{R=1}}$ augmente directement celle de $\theta^y(1-\theta)^{n-y}$. La courbe de la fonction $B_{Seq,\hat{\theta}^{Mod_{R=1}}}(\theta)$ se rapproche ainsi de celle de $B_{Seq,\hat{\theta}^{MV}}(\theta)$ à mesure que n_2 augmente, en présentant toutefois une valeur maximum très inférieure. En note également la superposition

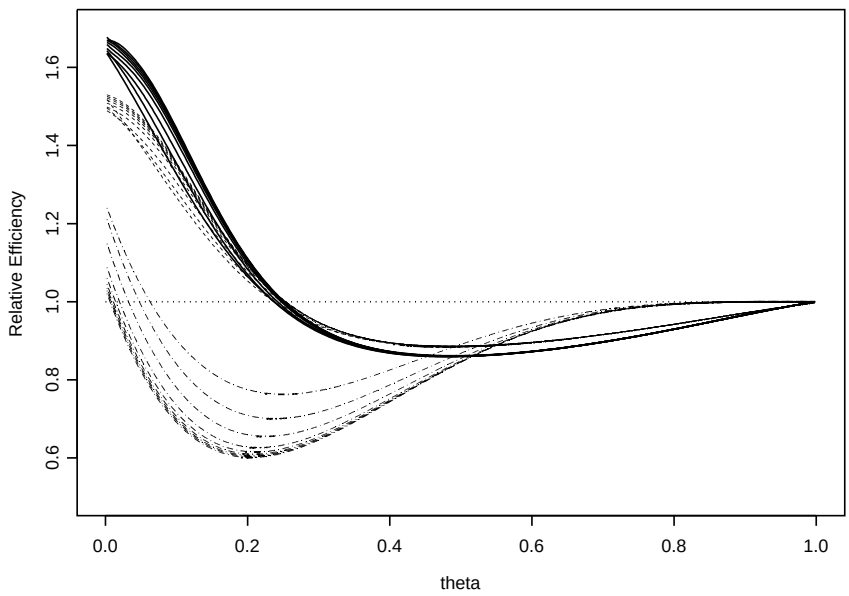

FIG. 3.23 – *Efficacité Relative de* $\hat{\theta}^{Mod_{R=1}}$ *(—),* $\hat{\theta}^{ModA_{R=1}}$ *(- - -) et de* $\hat{\theta}^{VUMSB}$ *(·—·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5, 10, 20, 40, .., 120, 150)$

des profils de courbes de $B_{Seq,\hat{\theta}^{Mod_{R=1}}}(\theta)$ et de $B_{Seq,\hat{\theta}^{ModA_{R=1}}}(\theta)$, soulignant la qualité de l'approximation.

L'efficacité de $\hat{\theta}^{Mod_{R=1}}$ et de son approximation $\hat{\theta}^{ModA_{R=1}}$ relativement à $\hat{\theta}^{MV}$ sont illustrées dans la figure (3.23), qui montre le positionnement favorable en terme d'erreur quadratique moyenne de ces deux estimateurs par rapport à $\hat{\theta}^{VUMSB}$, notamment pour les petites valeurs de θ.

Les estimateurs $\hat{\theta}^{Mod}$ pour $r \leq 1$ et $\hat{\theta}^{ModA_{R=1}}$ pour $r > 1$ peuvent être rassemblés dans un seul et même estimateur défini $\forall\ r > 0$. On note $\hat{\theta}^{ModC}$ l'estimateur du mode corrigé tel que

$$\hat{\theta}^{ModC} = \begin{cases} \hat{\theta}^{Mod} & r \leq 1 \\ \hat{\theta}^{Mod_{R=1}} & r > 1. \end{cases} \tag{3.3.6}$$

L'estimation $\hat{\theta}^{ModC}_{(m,y)}$ dans le cas où $S_{1,U} = 1$ correspond de fait au mode de la loi a posteriori $Be^J(y+1, n-y+1, 1; n_1, min(r,1); \{0\})$.

De la même façon, l'approximation de l'estimateur du mode corrigé se déduit de (3.3.3) par

$$\hat{\theta}^{ModCA} = \frac{Y}{n}\left(1 + \frac{n_1}{n}\frac{min(r,1)}{1+\left(1-\frac{Y}{n}\right)^{-n_1}}\right)^{-1}.$$

L'estimation $\hat{\theta}^{VUMSB}_{(m,y)} = 0$ lorsque $m = 2$ n'est pas satisfaisante et la contre partie de l'estimation non-biaisée en terme d'erreur quadratique moyenne place $\hat{\theta}^{VUMSB}$ loin des performances de $\hat{\theta}^{ModC}$, $\hat{\theta}^{ModCA}$ mais également de $\hat{\theta}^{MV}$. Lorsque $r \leq 1$, le biais induit par $\hat{\theta}^{ModCA}$ s'annule autour des valeurs de θ proches de $1/n_1$; si $r > 1$, il demeure très en deçà du biais induit par $\hat{\theta}^{MV}$. Il est naturel de considérer que la correction effectuée dans $\hat{\theta}^{ModCA}$ s'applique également à $\hat{\theta}^{ModCA_{0.5}}$; il en résulte une estimation encore plus précise de $\hat{\theta}^{ModC}$ par

$$\hat{\theta}^{ModCA_{0.5}} = \frac{Y}{n}\left(1 + \frac{n_1}{n}\frac{min(r,1)}{1+\left(1-\frac{Y}{n}\right)^{0.5-n_1}}\right)^{-1}.$$

L'étude qui précède montre que $\hat{\theta}^{ModCA}$ constitue une alternative avantageuse à $\hat{\theta}^{MV}$ et $\hat{\theta}^{VUMSB}$ dans un plan séquentiel du type $(n_1, S_{1,U} = 1, n_2)$. La section suivante vise à démontrer que ces bonnes propriétés sont généralisables à d'autres plans séquentiels.

3.4 Généralisation à d'autres plans séquentiels

3.4.1 Dérivation de l'estimateur du Mode Corrigé

Les principaux points abordés dans la section précédente sont revus dans un contexte plus général de plans séquentiels à deux périodes $(n_1, S_{1,L}, S_{1,U}, n_2)$. Cette généralisation nécessite de définir l'estimateur du mode corrigé $\hat{\theta}^{ModC}$ ainsi que son approximation $\hat{\theta}^{ModCA}$ pour toute valeurs de $S_{1,L}$ et $S_{1,U}$. On ne trouvera pas de forme dérivée de $\hat{\theta}^{ModCA}$ afin d'optimiser la précision de l'estimation, cette étude pouvant faire l'objet de développements ultérieurs.

Si l'on reprend la même approche, l'équation (3.2.2) doit être dérivée dans une nouvelle forme afin de reproduire les figures (C.1) à (C.3) qui montrent les courbes des fonctions holomorphes $g(\theta)$ et $h(\theta)$. L'idée est de déterminer une égalité qui présente le terme de gauche sous la forme $\theta/(y - n\theta)$. En partant de l'équation (3.2.2)

$$\frac{1-\theta}{y-n\theta} = \frac{1}{r}\,\frac{1 + r\,CBin(n_1, \theta; [S_{1,L}+1, S_{1,U}-1])}{S_{1,U}\,Bin(n_1, \theta; S_{1,U}) - (S_{1,L}+1)\,Bin(n_1, \theta; S_{1,L}+1)},$$

et en multipliant les deux termes par $\theta/(1-\theta)$, il vient

$$\begin{aligned}\frac{\theta}{y-n\theta} &= \frac{\theta}{r(1-\theta)}\,\frac{1 + r\,CBin(n_1, \theta; [S_{1,L}+1, S_{1,U}-1])}{S_{1,U}\,Bin(n_1, \theta; S_{1,U}) - (S_{1,L}+1)\,Bin(n_1, \theta; S_{1,L}+1)} \\ &= \frac{1}{r}\,\frac{1 + r\,CBin(n_1, \theta; [S_{1,L}+1, S_{1,U}-1])}{(n_1 - S_{1,U}+1)Bin(n_1, \theta; S_{1,U}-1) - (n_1 - S_{1,L})Bin(n_1, \theta; S_{1,L})}.\end{aligned}$$

La solution de cette équation est la valeur en θ du point de croisement des courbes des deux fonctions holomorphes $g(\theta)$ et $h(\theta)$ formées par les deux termes de cette égalité. Les figures (3.24) et (3.25) décrivent $g(\theta)$ et $h(\theta)$ dans le plan $(n_1 = 5, S_{1,U} = s_{1,U}, n_2 = 5)$ associé aux valeurs $s_{1,U} = 2, 3$. Les figures (C.1), (C.2) et (C.3) décrivent les mêmes courbes $s_{1,U} = 1, 4, 5$.

Les asymptotes de $h(\theta)$ en $\theta = y/n$ ne sont pas représentées mais l'échelle du graphique permet de visualiser le point $g(y/n)$ projeté horizontalement sur $h(\theta)$. Ce point est proche de l'intersection de $g(\theta)$ avec $h(\theta)$, et sa valeur en θ constitue une approximation de $\hat{\theta}^{ModC}_{(m,y)}$. On l'obtient analytiquement en résolvant l'équation $h(\theta) = g(y/n)$ dérivée de (3.2.2). Il vient

$$\hat{\theta}^{ModCA}_{(m,y)} = \frac{yt}{1+nt},$$

où

$$\begin{aligned} t &= \frac{y}{R(n-y)}\,\frac{1 + R\,CBin(n_1, y/n; [S_{1,L}+1, S_{1,U}-1])}{S_{1,U}\,Bin(n_1, y/n; S_{1,U}) - (S_{1,L}+1)\,Bin(n_1, y/n; S_{1,L}+1)}, \\ R &= max(r, 1). \end{aligned} \tag{3.4.1}$$

Dans le cas d'un plan séquentiel borné unilatéralement par $S_{1,U}$ en période 1, on

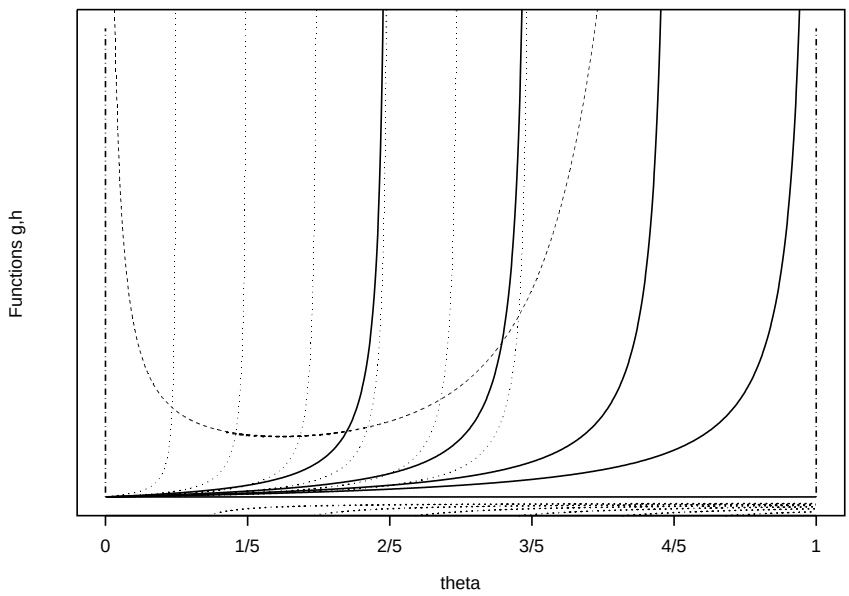

FIG. 3.24 – *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) dans* $\mathcal{Y}_{(1,y)}=\{y=y_1; y=2,3,4,5\}$ *et* $h(\theta)$ *(· · ·) dans* $\mathcal{Y}_{(2,y)}=\{y=y_1+y_2; y=0,1,..,6\}$ *pour le plan séquentiel* $(n_1=5, S_{1,U}=2, n_2=5)$

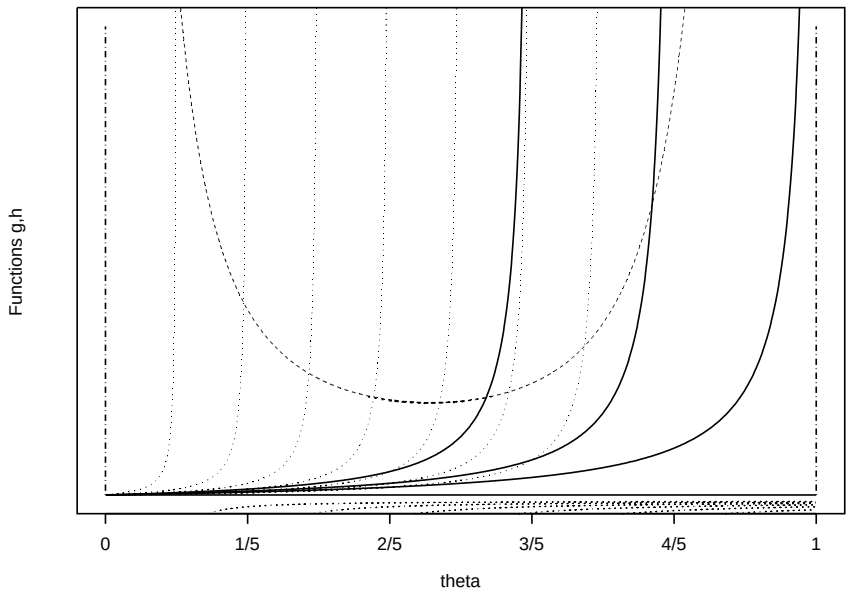

FIG. 3.25 – *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) dans* $\mathcal{Y}_{(1,y)}=\{y=y_1; y=3,4,5\}$ *et* $h(\theta)$ *(· · ·) dans* $\mathcal{Y}_{(2,y)}=\{y=y_1+y_2; y=0,1,..,7\}$ *pour le plan séquentiel* $(n_1=5, S_{1,U}=3, n_2=5)$

remplace (3.4.1) par

$$t = \frac{y}{R(n-y)} \frac{1 + RCBin(n_1, y/n; [0, S_{1,U} - 1])}{S_{1,U}\, Bin(n_1, y/n; S_{1,U})},$$
$$R = min(r, 1).$$

Dans le cas d'un plan séquentiel borné unilatéralement par $S_{1,L}$ en période 1, on remplace (3.4.1) par

$$t = \frac{y}{R(n-y)} \frac{1 + RCBin(n_1, y/n; [S_{1,L} + 1, n_1])}{-(S_{1,L} + 1)\, Bin(n_1, y/n; S_{1,L} + 1)},$$
$$R = min(r, 1).$$

A titre de comparaison, les tableaux dans les figures (3.26) et (3.27) donnent les valeurs de $\hat{\theta}^{MV}_{(m,y)}$, $\hat{\theta}^{ModC}_{(m,y)}$, $\hat{\theta}^{ModCA}_{(m,y)}$ et $\hat{\theta}^{VUMSB}_{(m,y)}$ pour les plans séquentiels ($n_1 = 5, S_{1,U} = s_{1,U}, n_2 = 5$) avec $s_{1,U} = 1, 2$. Les figures (C.4), (C.5) et (C.6) fournissent la même information pour respectivement $s_{1,U} = 3, 4, 5$. Les graphiques associés permettent de visualiser l'écart entre les estimations données par $\hat{\theta}^{ModC}_{(m,y)}$, $\hat{\theta}^{ModCA}_{(m,y)}$ et $\hat{\theta}^{VUMSB}_{(m,y)}$ par rapport aux estimations donnée par $\hat{\theta}^{MV}_{(m,y)} = y/n$ sur $\mathcal{Y}_{(1,y)} = \{y = y_1; y = \{S_{1,U}, .., n_1\}$ et sur $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, .., S_{1,U} + n_2 - 1\}$.

Le graphique (3.26) montre l'écart extrême entre les valeurs de $\hat{\theta}^{VUMSB}_{(1,y)} = y_1/n_1$ et $\hat{\theta}^{VUMSB}_{(2,y)} = 0$ dans un plan séquentiel unilatéral ($n_1, S_{1,U} = 1, n_2$). La contre-partie en terme d'estimation pour l'obtention d'un estimateur sans biais de variance minimum est trop pénalisante pour recommander l'utilisation de $\hat{\theta}^{VUMSB}$ dans ce contexte. Le comportement de $\hat{\theta}^{VUMSB}$ pour les autres plans séquentiels ($n_1, S_{1,U} = i, n_2$) $\forall i \in \{2, 3, .., n_1\}$ reste lié à la condition $m = 1$ (i.e. $\hat{\theta}^{VUMSB}_{(1,y)} = y_1/n_1$) ; l'ajustement du biais ne s'effectue que sur les valeurs de $\hat{\theta}^{VUMSB}_{(2,y)}$. Cette contrainte affecte directement l'interprétation et réduit de fait l'intérêt de cet estimateur.

En revanche l'estimateur du mode corrige la valeur de $\hat{\theta}^{MV} = Y/N$ quelle que soit la période $M = 1, 2$ où l'expérience s'arrête. La correction est maximum dans $\mathcal{Y}_{(1,y)}$ et $\mathcal{Y}_{(2,y)}$ lorsque

$$\frac{y}{n} = \frac{y_1}{n_1} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{S_{1,U}}{n_1},$$

	$\hat{\theta}^{MV}_{(m,y)}$	$\hat{\theta}^{Mod}_{(m,y)}$	$\hat{\theta}^{ModA}_{(m,y)}$	$\hat{\theta}^{VUMSB}_{(m,y)}$
(1,1)	0.200	0.153	0.147	0.200
(1,2)	0.400	0.366	0.370	0.400
(1,3)	0.600	0.593	0.594	0.600
(1,4)	0.800	0.800	0.800	0.800
(1,5)	1	1	1	1
(2,0)	0	0	0	0
(2,1)	0.100	0.084	0.081	0
(2,2)	0.200	0.176	0.175	0
(2,3)	0.300	0.277	0.278	0
(2,4)	0.400	0.384	0.385	0
(2,5)	0.500	0.492	0.492	0

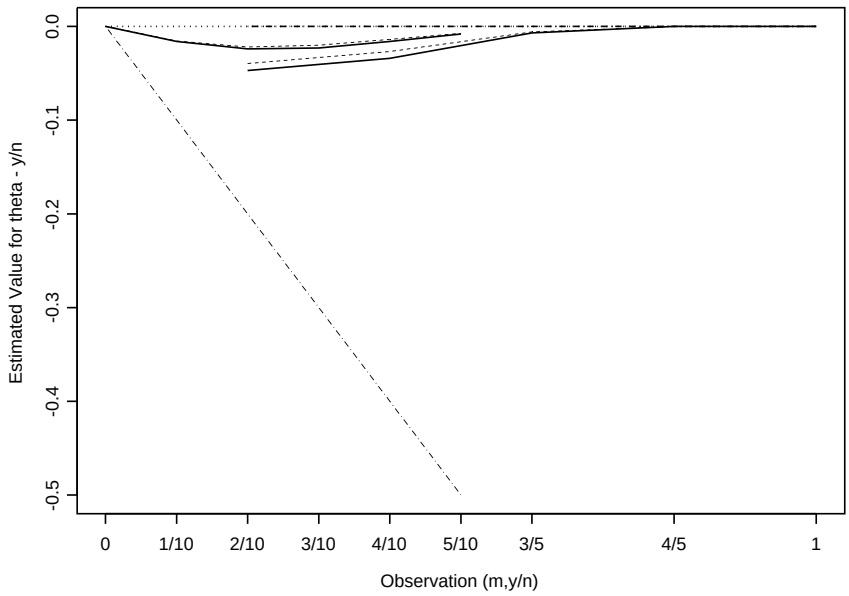

FIG. 3.26 – *Estimation de* $\hat{\theta}^{ModC}_{(m,y)}$ *(—),* $\hat{\theta}^{ModCA}_{(m,y)}$ *(- - -) et* $\hat{\theta}^{VUMSB}_{(m,y)}$ *(· − ·), soustraite de* $\hat{\theta}^{MV}_{(m,y)}$*, dans* $\mathcal{Y}_{(1,y)}=\{y=y_1; y=1,2,..,5\}$ *et* $\mathcal{Y}_{(2,y)}=\{y=y_1+y_2; y=0,1,..,5\}$ *pour le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$

	$\hat{\theta}^{MV}_{(m,y)}$	$\hat{\theta}^{ModC}_{(m,y)}$	$\hat{\theta}^{ModCA}_{(m,y)}$	$\hat{\theta}^{VUMSB}_{(m,y)}$
(1,2)	0.400	0.339	0.331	0.400
(1,3)	0.600	0.558	0.563	0.600
(1,4)	0.800	0.796	0.796	0.800
(1,5)	1	1	1	1
(2,0)	0	0	0	0
(2,1)	0.100	0.094	0.093	0.100
(2,2)	0.200	0.183	0.181	0.143
(2,3)	0.300	0.274	0.271	0.167
(2,4)	0.400	0.369	0.367	0.181
(2,5)	0.500	0.472	0.472	0.192
(2,6)	0.600	0.581	0.582	0.200

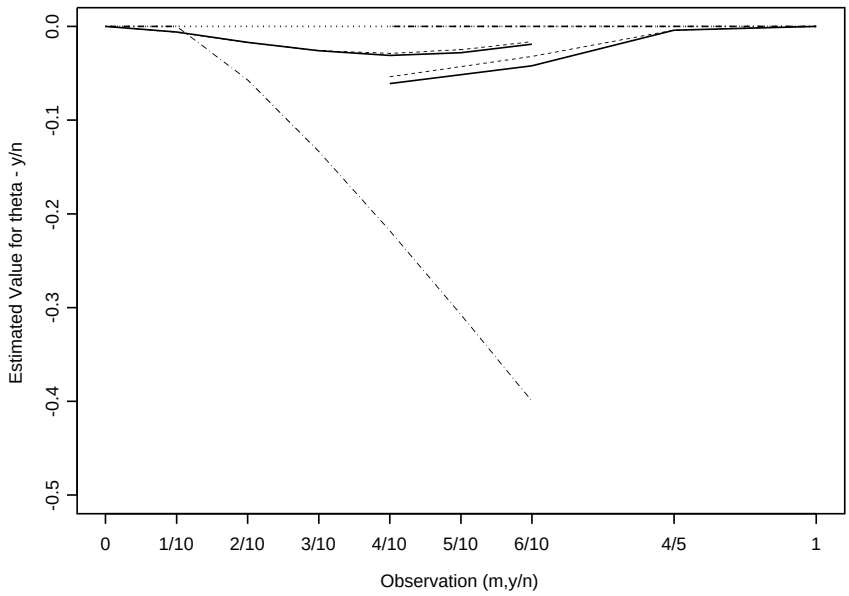

FIG. 3.27 – *Estimation de* $\hat{\theta}^{ModC}_{(m,y)}$ *(—),* $\hat{\theta}^{ModCA}_{(m,y)}$ *(- - -) et* $\hat{\theta}^{VUMSB}_{(m,y)}$*(· − ·), soustraite de* $\hat{\theta}^{MV}_{(m,y)}$*, dans* $\mathcal{Y}_{(1,y)}=\{y=y_1; y=2,3,4,5\}$ *et* $\mathcal{Y}_{(2,y)}=\{y=y_1+y_2; y=0,1,..,6\}$ *pour le plan séquentiel* $(n_1=5, S_{1,U}=2, n_2=5)$

et diminue lorsque y/n s'éloigne de $S_{1,U}/n_1$. L'estimation $\hat{\theta}^{ModC}_{(m,y)}$ sous-estime plus fortement les valeurs de $\hat{\theta}^{MV}_{(m,y)}$ si l'expérience s'arrête en période 1 (i.e. $m = 1$).

La correction associée à $\hat{\theta}^{ModC}$ est cohérente avec la stratégie expérimentale ; les valeurs de l'estimation y/n proches du seuil $S_{1,U}/n_1$ sont plus concernées par le biais expérimental et donc plus sous-estimées. On note enfin que les valeurs obtenues par l'approximation $\hat{\theta}^{ModCA}_{(m,y)}$ sont très proches de l'estimation $\hat{\theta}^{ModC}_{(m,y)}$.

3.4.2 Etude du biais et de l'efficacité relative dans différents plans bornés unilatéralement

La figure (3.28) montre les valeurs du biais induit par les estimateurs $\hat{\theta}^{MV}$, $\hat{\theta}^{ModC}$ et $\hat{\theta}^{ModCA}$ dans les plans séquentiels ($n_1 = 5, S_{1,U} = s_{1,U}, n_2 = 5$) avec $s_{1,U} = 1, 2, .., 5$. La fonction $B_{Seq,\hat{\theta}^{MV}}(\theta)$ est positive $\forall\, \theta \in [0, 1]$ alors que $B_{Seq,\hat{\theta}^{ModC}}(\theta)$ et $B_{Seq,\hat{\theta}^{ModCA}}(\theta)$ croisent l'axe horizontal 2 fois ; une première fois juste après l'origine pour devenir positive et une deuxième fois autour de la valeur $\theta = S_{1,U}/n_1$.

Les cinq courbes associées à $B_{Seq,\hat{\theta}^{MV}}(\theta)$ montrent une valeur maximum autour de $\theta = S_{1,U}/n_1$ alors que celles associées à $B_{Seq,\hat{\theta}^{Mod}}(\theta)$ croisent l'axe horizontal autour de cette valeur. La valeur maximum de $|B_{Seq,\hat{\theta}^{ModC}}(\theta)|$ augmente en fonction de $S_{1,U}$ mais reste largement en deçà de celle de $|B_{Seq,\hat{\theta}^{MV}}(\theta)|$ pour $S_{1,U} = 1, .., n_1$, la valeur $S_{1,U} = n_1$ traduisant le paramétrage le moins avantageux pour $\hat{\theta}^{ModC}$ en terme de biais. Les valeurs de $B_{Seq,\hat{\theta}^{ModC}}(\theta)$ et de son approximation $B_{Seq,\hat{\theta}^{ModCA}}(\theta)$ reste très proches quel que soit θ.

La description de l'efficacité relative de $\hat{\theta}^{ModC}$, $\hat{\theta}^{ModCA}$ et $\hat{\theta}^{VUMSB}$ versus $\hat{\theta}^{MV}$ dans les plans séquentiels ($n_1 = 5, S_{1,U} = s_{1,U}, n_2 = 5$) avec $s_{1,U} = 2, 3$ est décrite dans les figures (3.29) et (3.30). Celles concernant les plans avec $s_{1,U} = 1, 4, 5$ sont respectivement rapportées dans les figures (3.17), (C.7) et (C.8).

La courbe de $ER_{\hat{\theta}^{ModC}}$ est supérieure à 1 sur un intervalle allant de $\theta = 0$ à une valeur

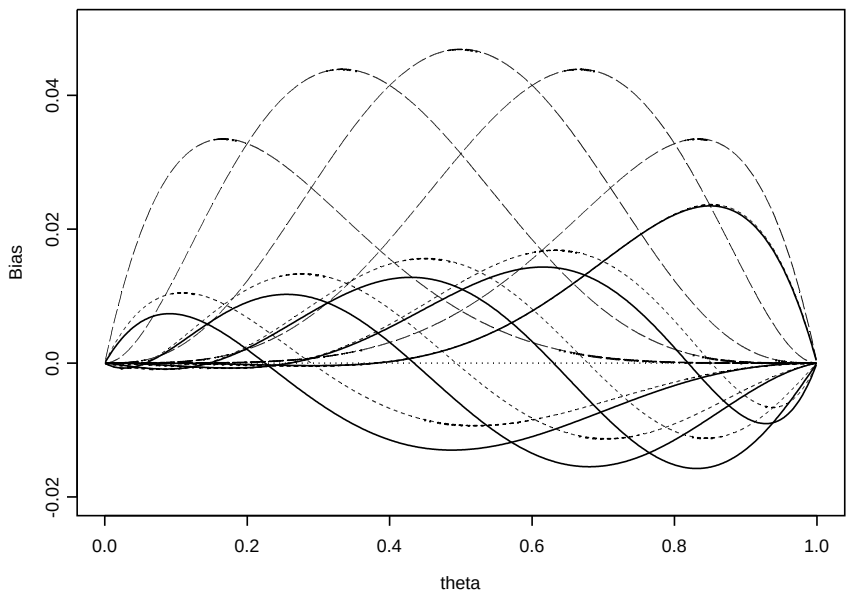

FIG. 3.28 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -) dans le plan séquentiel* $(n_1=5, S_{1,U}=s_{1,U}, n_2=5)$ *avec* $s_{1,U}=1,2,3,4,5$

proche de $S_{1,U}/n_1$, dénotant une erreur quadratique moyenne plus petite pour $\hat{\theta}^{ModC}$ par rapport à $\hat{\theta}^{MV}$. Le profil de $ER_{\hat{\theta}^{ModC}}$ est supérieur à celui de $ER_{\hat{\theta}^{VUMSB}}$ pour $\theta \simeq S_{1,U}/n_1$. On note enfin la superposition des courbes de $ER_{\hat{\theta}^{ModC}}$ et de son approximation $ER_{\hat{\theta}^{ModCA}}$ $\forall\ S_{1,U} = 1, 2, .., 5$.

Les graphiques qui suivent visent à étudier le comportement du biais de $\hat{\theta}^{ModC}$ et $\hat{\theta}^{ModCA}$ en fonction de θ dans le plan séquentiel $(n_1=5, S_{1,U}, n_2)$ lorsque $r \neq 1$.

Les figures (3.31), (3.32) et (3.33) décrivent les courbes de biais associées à l'estimateur $\hat{\theta}^{ModC}$ et $\hat{\theta}^{ModCA}$ dans les plans séquentiels $(n_1 = 5, S_{1,U} = s_{1,U}, n_2)$ avec $s_{1,U} = 1, 2, 3, 4, 5$ lorsque respectivement $n_2=2, 10, 50$.

Relativement à $\hat{\theta}^{MV}(\theta)$, les courbes de biais associées à $\hat{\theta}^{ModC}(\theta)$ et $\hat{\theta}^{ModCA}(\theta)$ montrent des performances optimales lorsque $r < 1$. Lorsque $r > 1$, les courbes de $B_{Seq,\hat{\theta}^{ModC}}(\theta)$ et $B_{Seq,\hat{\theta}^{ModCA}}(\theta)$ se rapprochent de celle de $B_{Seq,\hat{\theta}^{MV}}(\theta)$ à mesure que l'effectif n_2 augmente. Quelle que soit la valeur de r, les courbes de $B_{Seq,\hat{\theta}^{ModC}}(\theta)$

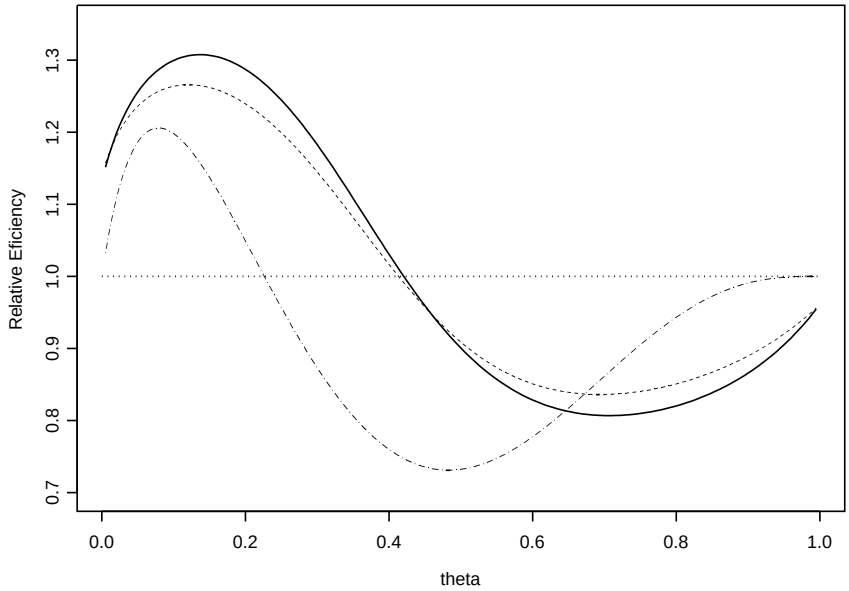

FIG. 3.29 – *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{ModCA}$ *(- - -) et* $\hat{\theta}^{VUMSB}$ *(·—·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=2, n_2=5)$

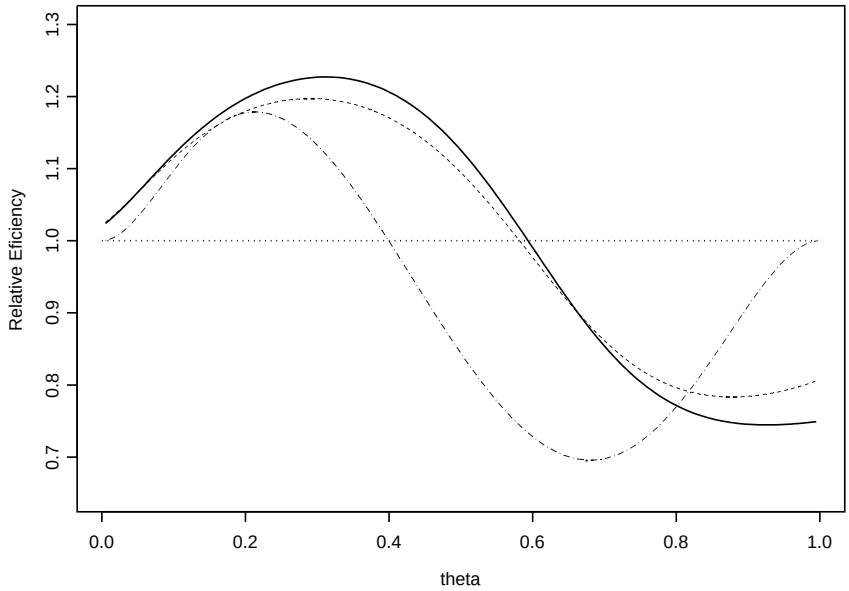

FIG. 3.30 – *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{ModCA}$ *(- - -) et* $\hat{\theta}^{VUMSB}$ *(·—·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=3, n_2=5)$

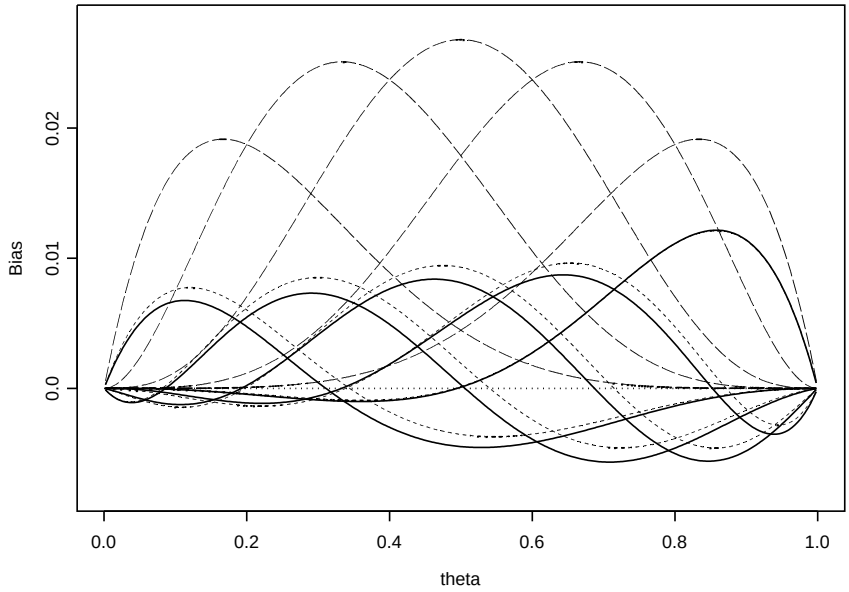

FIG. 3.31 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = s_{1,U}, n_2 = 2)$ *avec* $s_{1,U} = 1, 2, 3, 4, 5$

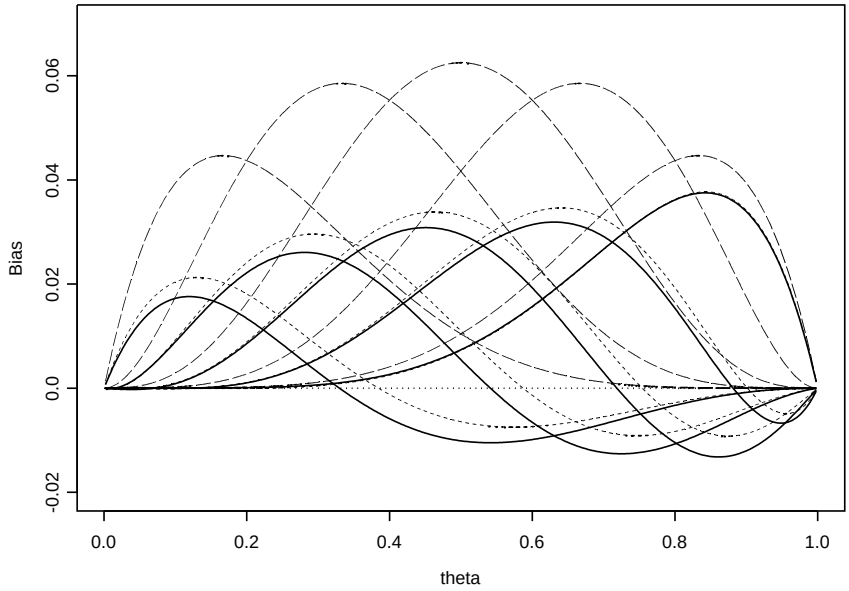

FIG. 3.32 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = s_{1,U}, n_2 = 10)$ *avec* $s_{1,U} = 1, 2, 3, 4, 5$

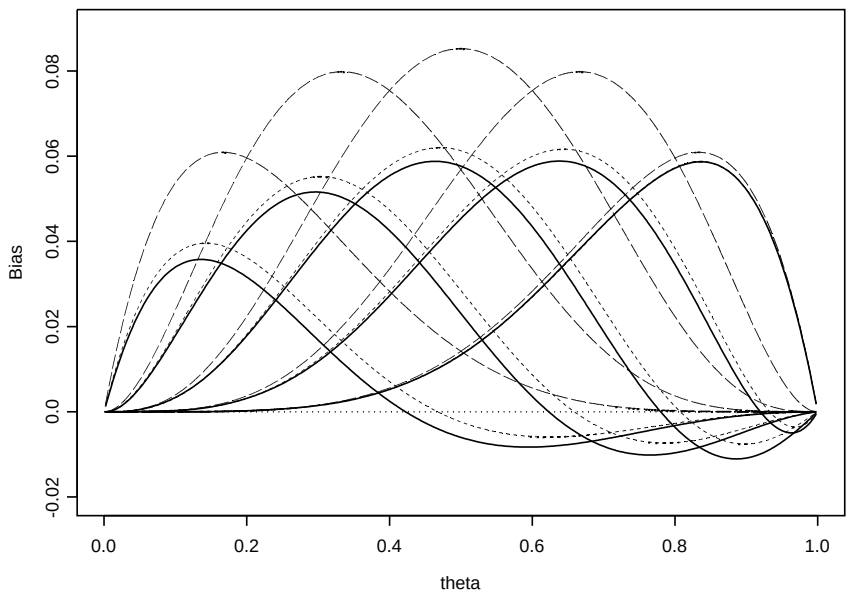

FIG. 3.33 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -) dans le plan séquentiel* $(n_1=5, S_{1,U}=s_{1,U}, n_2=50)$ *avec* $s_{1,U}=1,2,3,4,5$

et $B_{Seq,\hat{\theta}^{ModCA}}(\theta)$ sont les plus proches de celle de $B_{Seq,\hat{\theta}^{MV}}(\theta)$ lorsque $S_{1,U}=n_1$.

3.4.3 Propriété de l'estimateur du Mode Corrigé dans un plan borné bilatéralement

On vérifie dans cette section que l'approche analytique fondée sur le point d'intersection entre les 2 fonctions holomorphes $g(\theta)$ et $h(\theta)$ pour déterminer $\hat{\theta}^{ModC}$ et $\hat{\theta}^{ModCA}$ s'applique également dans un plan séquentiel borné bilatéralement par $S_{1,L}$ et $S_{1,U}$. Le graphique (3.34) montre les courbes de la fonction $g(\theta)$ dans le plan $(n_1=5, S_{1,L}=s_{1,L}, S_{1,U}=s_{1,L}+3, n_2=5)$ avec $s_{1,L}=0,1,2,3$. La superposition de l'ensemble des courbe de la fonction $h(\theta)$ sur les espaces $\mathcal{Y}_{(1,y)}=\{y=y_1; y=\{1,..,n_1\}$ et $\mathcal{Y}_{(2,y)}=\{y=y_1+y_2; y=0,..,n_1+n_2-1\}$ permet de vérifier la consistance de l'approche quelles que soient les valeurs de $s_{1,L}$ et des séquences $(k,y)\in\mathcal{Y}=\mathcal{Y}_{(1,y)}\cup\mathcal{Y}_{(2,y)}$. Lorsque la valeur de $s_{1,L}$ augmente, la courbe de $g(\theta)$ se décale vers les valeurs supérieures de θ en même temps que son minimum se rapproche de l'axe horizontal. On

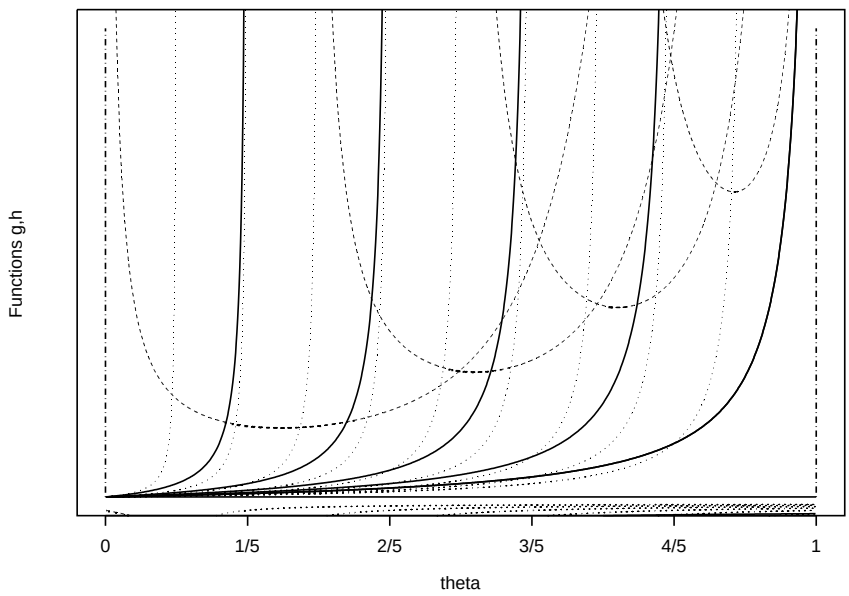

FIG. 3.34 – *Fonctions holomorphes* $g(\theta)$ *(- - -),* $h(\theta)$ *(—) dans* $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 1, 2, .., 5\}$ *et* $h(\theta)$ $(\cdots)$ *dans* $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 10\}$ *pour les plans séquentiels* $(n_1 = 5, S_{1,L} = s_{1,L}, S_{1,U} = s_{1,L} + 3, n_2 = 5)$ *avec* $s_{1,L} = -1, 0, 1, 2, 3$

mesure les performances de $\hat{\theta}^{ModC}$ et $\hat{\theta}^{ModCA}$ en terme de biais. Les figures (3.35), (3.36) montrent les courbes de $B_{Seq,\hat{\theta}^{ModC}}(\theta)$, $B_{Seq,\hat{\theta}^{ModCA}}(\theta)$ et $B_{Seq,\hat{\theta}^{MV}}(\theta)$ dans le plan $(n_1 = 5, S_{1,L} = s_{1,L}, S_{1,U} = s_{1,L} + 3, n_2 = 5)$ avec $s_{1,L} = 0, 1$. Les mêmes courbes pour les valeurs $s_{1,L} = -1, 2, 3$ sont montrées dans les figures (C.9), (C.10) et (C.11). Les valeurs $s_{1,L} = 2$ et $s_{1,L} = 6$ correspondent à des plans bornés unilatéralement respectivement à gauche et droite (voir discussion section (3.4.2)).

Le profil des courbes de $B_{Seq,\hat{\theta}^{ModC}}(\theta)$ et de $B_{Seq,\hat{\theta}^{ModCA}}(\theta)$ suit la variation de $B_{Seq,\hat{\theta}^{MV}}(\theta)$ avec cependant des amplitudes moins fortes et plus centrées sur l'axe horizontal. Le biais maximum de $\hat{\theta}^{MV}$ et $\hat{\theta}^{ModC}$ évoluent peu en fonction de la valeur de $(S_{1,L}, S_{1,U}) = (s_{1,L}, s_{1,L} + 3)$. On trouve respectivement $max(|B_{Seq,\hat{\theta}^{MV}}(\theta)|) \simeq 0.04$ et $max(|B_{Seq,\hat{\theta}^{Mod}}(\theta)|) \simeq 0.02$.

La symétrie des courbes de $B_{Seq,\hat{\theta}^{MV}}(\theta)$ et de $B_{Seq,\hat{\theta}^{Mod}}(\theta)$ autour de la valeur $\theta = 0.5$ dans la figure (3.36) est due au paramétrage $S_{1,L} = 1$ et $S_{1,U} = n_1 - 1$ qui ne favorise

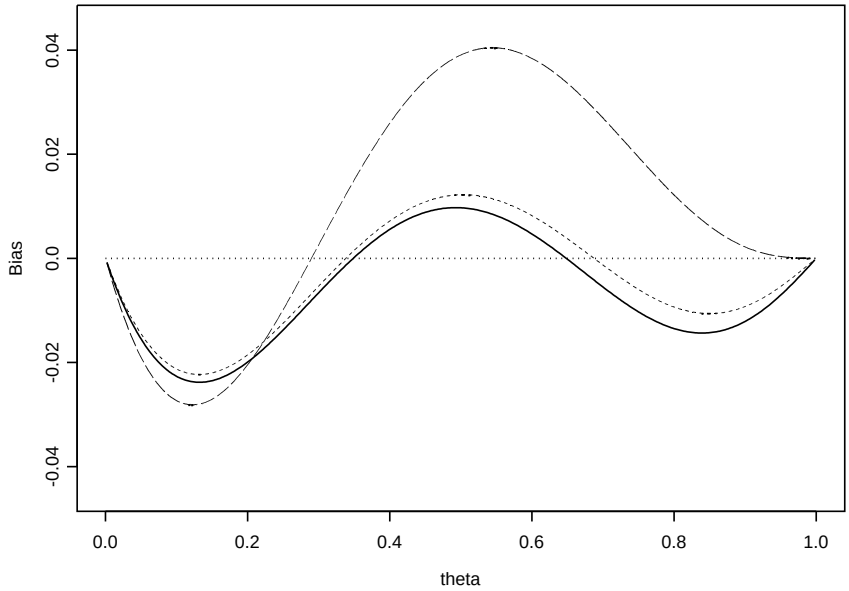

FIG. 3.35 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -) dans le plan séquentiel* $(n_1=5, S_{1,L}=0, S_{1,U}=3, n_2=5)$

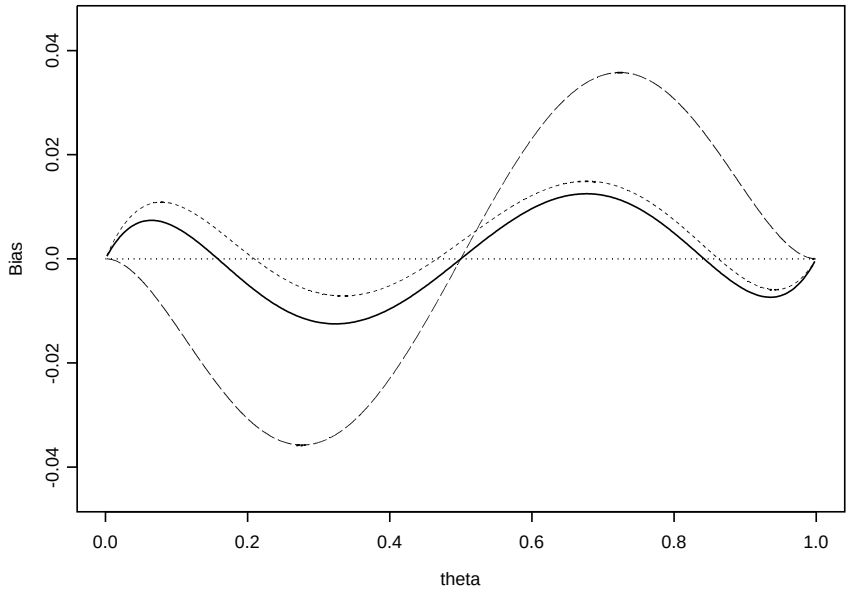

FIG. 3.36 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -) dans le plan séquentiel* $(n_1=5, S_{1,L}=1, S_{1,U}=4, n_2=5)$

pas plus les valeurs faibles que les valeurs fortes de θ. On retrouve également par symétrie les courbes de $B_{Seq,\hat{\theta}^{MV}}(\theta)$ et de $B_{Seq,\hat{\theta}^{Mod}}(\theta)$ présentées dans les figures (C.10) et (C.11) à partir de celles présentées dans les figures (C.9) et (3.35).

3.4.4 Comparaison du biais et de l'efficacité relative avec d'autres estimateurs

L'objectif de cette section est de comparer les performances de $\hat{\theta}^{ModC}$, en terme de biais et d'écart quadratique moyen, avec deux autres estimateurs dont les estimations s'obtiennent numériquement.

Le premier $\hat{\theta}^{Mean}$ est la moyenne a posteriori de $\vartheta|d_{Seq},(m,y)$. L'estimation $\hat{\theta}^{Mean}_{(m,y)}$ s'obtient par l'intégrale

$$\hat{\theta}^{Mean} = \int_0^1 \theta p(\theta|d_{Seq},(m,y))d\theta,$$

où

$$\vartheta|d_{Seq} \sim Be^J(a+y, b+n-y, c; n_1, r; J).$$

La détermination des paramètres a, b, c de la loi a priori $Be^J(a, b, c; n_1, r; J)$ tient d'abord compte d'une caractéristique de la loi $Be(a+y, b+n-y)$ qui est la loi a posteriori correspondant à un a priori $\theta \sim Be(a, b)$. L'espérance de $Be(a+y, b+n-y)$ est $(a+y)/(a+b+n)$ (voir chapitre B.2). Dans un plan experimental non séquentiel, l'a priori $\vartheta \sim Be(0, 0)$ définit une loi a posteriori $\vartheta|y \sim Be(y, n-y)$ dont l'espérance

$$E_\theta[\vartheta|y] = \frac{y}{n}$$

correspond à l'EMV non biaisé. La réduction du biais étant l'objectif recherché pour $\hat{\theta}^{Mean}$ dans le plan séquentiel, les valeurs $a = b = 0$ pour la loi Bêta-J sont retenues.

D'autre part, l'objectif de ce chapitre est de comparer directement les performances de $\hat{\theta}^{Mod}$ et $\hat{\theta}^{Mean}$; le choix de la valeur $c = 1$ traduit une influence similaire dans les 2 cas de la composante $(1 + rCBin(n_1, \theta; J))$.

La loi a posteriori $Be^J(y, n-y, 1; n_1, r; J)$ ainsi définie a une densité qui s'écrit

$$p(\theta|d_{Seq}, (m, y)) = \theta^{y-1}(1-\theta)^{n-y-1}(1 + rCBin(n_1, \theta; J)).$$

Le second estimateur utilisé est $\hat{\theta}^{Whi}$ (voir section 3.1). Cet estimateur est fondé sur la correction de $\hat{\theta}^{MV}$ par $B_{Seq,\hat{\theta}^{MV}}(\hat{\theta}^{Whi})$, qui fournit une approximation de $B_{Seq,\hat{\theta}^{MV}}(\theta)$. L'estimation $\hat{\theta}^{Whi}_{(m,y)}$ est la solution de

$$E_{Seq,\theta}[\hat{\theta}^{MV}|\hat{\theta}^{Whi}_{(m,y)}] = \hat{\theta}^{Whi}_{(m,y)}.$$

Les figures (3.37) et (3.38) décrivent les courbes de biais associé à $\hat{\theta}^{MV}$, $\hat{\theta}^{Mod}$, $\hat{\theta}^{Mean}$ et à $\hat{\theta}^{Whi}$ dans les plans $(n_1 = 5, S_{1,U} = s_{1,U}, n_2 = 5)$ avec respectivement $s_{1,U} = 1, 2$. Les courbes des fonctions $B_{Seq,\hat{\theta}^{MV}}$, $B_{Seq,\hat{\theta}^{Mod}}$, $B_{Seq,\hat{\theta}^{Mean}}$ et $B_{Seq,\hat{\theta}^{Whi}}$ associées aux valeurs $s_{1,U} = 3, 4, 5$ sont respectivement présentées dans les figures (C.12), (C.13) et (C.14).

Les courbes de $B_{Seq,\hat{\theta}^{Mod}}(\theta)$, $B_{Seq,\hat{\theta}^{Mean}}(\theta)$ et $B_{Seq,\hat{\theta}^{Whi}}(\theta)$ présentent un profil similaire. Celle de $B_{Seq,\hat{\theta}^{Whi}}(\theta)$ présente l'amplitude la moins importante autour de l'axe horizontal. La courbe de $B_{Seq,\hat{\theta}^{Mean}}(\theta)$ se situe globalement au dessus de celle de $B_{Seq,\hat{\theta}^{Mod}}(\theta)$. La valeur $S_{1,U} = n_1$ est le paramétrage le moins avantageux en termes d'amélioration relative du biais par rapport à $\hat{\theta}^{MV}$.

Les figures (3.39) et (3.40) montrent l'efficacité relative de $\hat{\theta}^{ModC}$, $\hat{\theta}^{Mean}$ et $\hat{\theta}^{Whi}$ versus $\hat{\theta}^{MV}$ dans les plans séquentiels $(n_1 = 5, S_{1,U} = s_{1,U}, n_2 = 5)$ avec respectivement $s_{1,U} = 1, 2$. Les courbes de $ER_{\hat{\theta}^{ModC}}$, $ER_{\hat{\theta}^{Mean}}$ et $ER_{\hat{\theta}^{Whi}}$ associées aux valeurs $s_{1,U} = 3, 4, 5$ sont respectivement présentées dans les figures (C.15), (C.16) et (C.17).

Les courbes de $ER_{\hat{\theta}^{ModC}}$, $ER_{\hat{\theta}^{Mean}}$ et $ER_{\hat{\theta}^{Whi}}$ ont le même profil, ce qui traduit un comportement similaire de l'écart quadratique moyen de $\hat{\theta}^{ModC}$, $\hat{\theta}^{Mean}$ et $\hat{\theta}^{Whi}$. Cette dernière étude confirme les bonnes propriétés de $\hat{\theta}^{ModC}$ et de fait de son approximation $\hat{\theta}^{ModCA}$ qui présente l'avantage d'offrir une forme analytique simple.

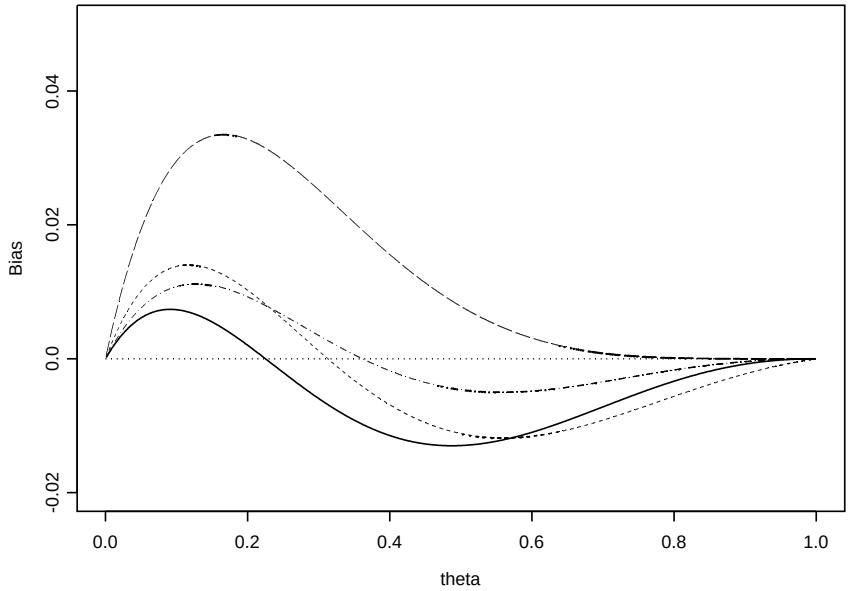

FIG. 3.37 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et de* $\hat{\theta}^{Whi}$ *(· – ·) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

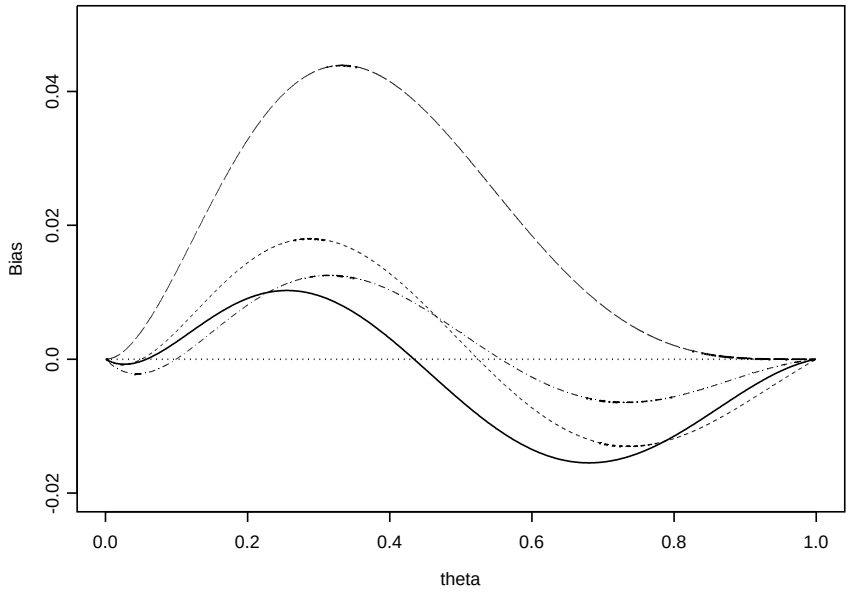

FIG. 3.38 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et de* $\hat{\theta}^{Whi}$ *(· – ·) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 2, n_2 = 5)$

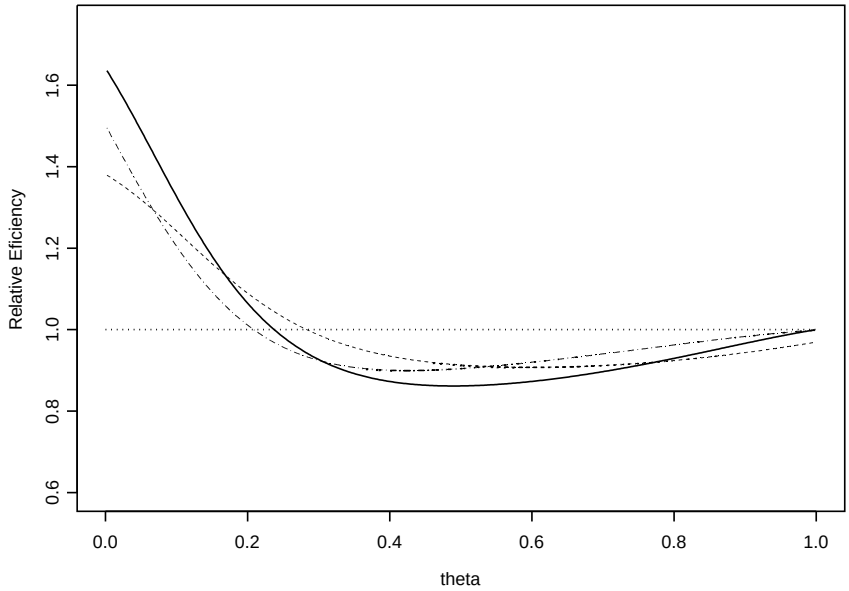

FIG. 3.39 – *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et* $\hat{\theta}^{Whi}$ *(·—·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$

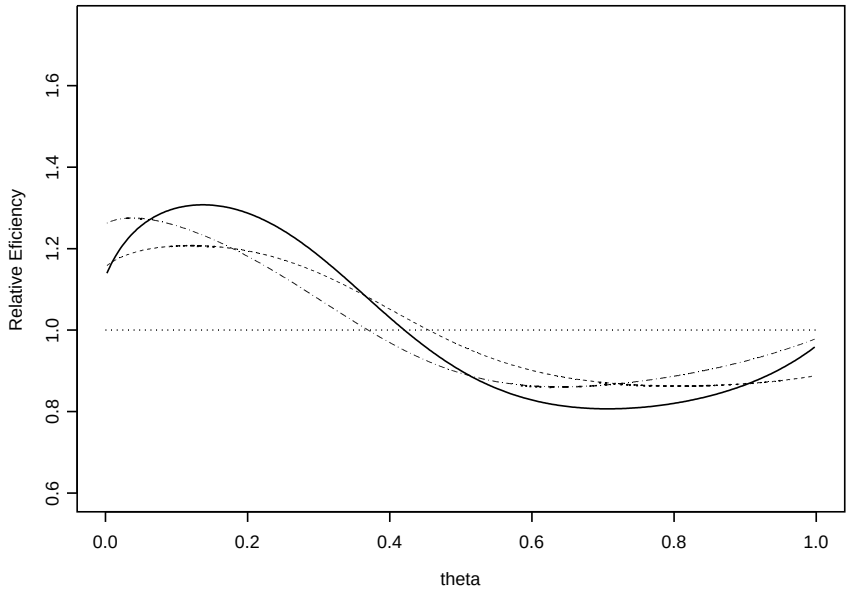

FIG. 3.40 – *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et* $\hat{\theta}^{Whi}$ *(·—·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=2, n_2=5)$

3.5 Conclusion

Dans l'échantillonnage Binomial fixe, le mode et la moyenne des lois a posteriori respectivement basés sur l'a priori Uniforme $Be(1,1)$ et l'a priori de Haldane $Be(0,0)$ conduisent à l'estimateur sans-biais de la proportion observée. Les résultats de ce chapitre ont révélé le lien direct entre le paramétrage de la loi Bêta-J dans un plan séquentiel comprenant un arrêt optionnel et la loi Bêta dans l'échantillonnage fixe. En notant n_1 et n_2 les effectifs prévus en période 1 et 2 et J la zone de continuation de l'expérience en période 2 pour le nombre observé de succès en période 1, on utilise respectivement les paramétrages $a = b = 1$ et $a = b = 0$ dans la loi a priori $Be^J(a, b, 1; n_1, min(1, n_2/n_1); J)$ pour dériver le mode et la moyenne des lois a posteriori résultantes. On a montré que les estimateurs associés ont des propriétés remarquables en terme d'interprétation de l'estimation, de biais et d'écart quadratique moyen. Une forme analytique du biais de l'"estimateur du mode corrigé" a été dérivée lorsque l'observation d'au moins un succès en période 1 conditionne l'arrêt de l'expérience.

Une étude comparative a mis en évidence les déficiences des estimateurs fréquentistes qui tiennent compte de règle d'arrêt, notamment l'estimateur sans-biais de variance minimum où le prix à payer pour annuler le biais se traduit par une grande difficulté d'interprétation. Les biais induits par les nouveaux estimateurs bayésiens sont proches de celui associé à l'"estimateur de Whitehead", qui s'obtient en retranchant à l'estimateur du maximum de vraisemblance son biais estimé.

La dérivation d'une approximation analytique de l'estimateur du mode corrigé devrait concourir à une large diffusion de ces nouveaux outils bayésiens.

Chapitre 4

Application à l'estimation par intervalle

L'objectif de ce chapitre est de proposer une nouvelle méthode d'estimation par intervalle fondée sur l'a priori de Jeffreys corrigé dans un plan expérimental comportant un arrêt optionnel. Les performances en terme de taux de recouvrement et de longueur moyenne de l'intervalle sont comparées avec celles des intervalles de confiance habituellement recommandés. Les notations en général sont celles utilisées dans le chapitre (3).

4.1 Différentes méthodes dans un échantillonnage fixe

Les méthodes les plus usuelles pour déterminer les bornes de l'intervalle de confiance de θ au niveau nominal $1-\alpha$ sont fondées sur l'approximation Normale de la distribution d'échantillonnage de la proportion Y/n. Lorsque l'effectif n d'un échantillonnage fixe est suffisant pour considérer que la loi $Bin(n, \theta)$ est convenablement approchée par la loi $N(n\theta, \theta(1-\theta))$, les bornes de l'intervalle de confiance au niveau nominal $1-\alpha$ sont obtenues par inversion de la zone d'acceptation du test de Wald

$$|(\hat{\theta}^{MV} - \theta)V^{-\frac{1}{2}}(\hat{\theta}^{MV})| \preceq \kappa, \tag{4.1.1}$$

où $\hat{\theta}^{MV} = Y/n$, $V(\theta) = \theta(1-\theta)/\,n$ et $\kappa = z_{1-\alpha/2}$ désigne le quantile $1-\alpha/2$ de la loi Normale centrée réduite. On déduit les bornes de l'*intervalle standard*

$$CI_S = \hat{\theta}^{MV} \pm \kappa V(\hat{\theta}^{MV})^{\frac{1}{2}} = \hat{\theta}^{MV} \pm \kappa\sqrt{\hat{\theta}^{MV}(1-\hat{\theta}^{MV})/n}.$$

Lorsque qu'une limitation de l'effectif observé n ne permet plus de s'inscrire dans une configuration asymptotique, les faibles performances de l'intervalle standard sont mentionnées dans Vollset [51]. En plus de l'absence de solution lorsque $Y/n \in \{0,1\}$, les bornes de CI_S vont au delà de l'intervalle $[0,1]$ lorsque Y/n approchent 0 ou 1 à cause de la symétrie de cet intervalle. Parallèlement, les taux de recouvrement de la borne supérieure $\hat{\theta}_S^{sup}$ et inférieure $\hat{\theta}_S^{inf}$ définis par

$$C_S^{sup}(\theta,n) = P_\theta(\theta \leq \hat{\theta}_S^{sup})$$
$$C_S^{inf}(\theta,n) = P_\theta(\theta \geq \hat{\theta}_S^{inf}),$$

diminuent sensiblement à proximité de $\theta = 1$ pour $C_S^{sup}(\theta,n)$ et de $\theta = 0$ pour $C_S^{inf}(\theta,n)$.

Les bornes de l'*intervalle de Wilson* au niveau nominal $1-\alpha$ sont obtenues par inversion de la zone d'acceptation du test du score où $V(\theta) = \theta(1-\theta)/\,n$ vient remplacer $V(\hat{\theta}^{MV}) = \hat{\theta}^{MV}(1-\hat{\theta}^{MV})/\,n$ dans l'expression (4.1.1). La solution passe alors par la résolution d'une forme quadratique en θ. En notant $\tilde{\theta} = \frac{Y+\kappa^2/2}{n+\kappa^2}$, les bornes sont définies par

$$CI_W = \tilde{\theta} \pm \frac{\kappa n^{1/2}}{n+\kappa^2}\left(\hat{\theta}^{MV}\left(1-\hat{\theta}^{MV}\right) + \frac{\kappa^2}{4n}\right)^{1/2}.$$

Contrairement à l'intervalle de Wilson, la forme de l'intervalle standard est familière et facile à utiliser. L'idée de l'*intervalle d'Agresti-Coull* est de conserver cette forme simple, tout en approchant les performances de l'intervalle de Wilson. En remplaçant $\hat{\theta}^{MV} = Y/n$ par $\tilde{\theta} = \tilde{Y}/\tilde{n}$ avec $\tilde{Y} = Y + \kappa^2/2$ et $\tilde{n} = n + \kappa^2$, les bornes de l'intervalle d'Agresti-Coull sont définies par

$$CI_{AC} = \tilde{\theta} \pm \kappa\sqrt{\tilde{\theta}(1-\tilde{\theta})/\tilde{n}}.$$

L'*intervalle de Clopper-Pearson* se calcule par inversion de la zone d'acceptation du test de la Binomiale au lieu de son approximation Normale dans l'intervalle standard. Il est abordé en détail dans la section B.3 mettant en avant ces propriétés remarquables dans l'échantillonnage fixe en terme de taux de recouvrement $max(C_{CP}^{sup}(\theta, n)) = max(C_{CP}^{inf}(\theta, n)) = 0.975$. Sa réinterprétation Bayésienne permet de définir l'intervalle

$$CI_{CP} = [Be(\alpha/2; y, n-y+1), Be(1-\alpha/2; y+1, n-y)],$$

où $Be(\alpha; a, b)$ est le quantile α associé à la loi $Be(a, b)$.

La conjugaison des lois a priori de type $\vartheta \sim Be(a, b)$ avec la loi d'échantillonnage Binomiale engendre des lois a posteriori $\vartheta|y \sim Be(a+y, b+n-y)$. Les intervalles de crédibilité fondés sur des aires sous la courbes $\alpha/2$ à chaque extrémité de la densité a posteriori présentent des bornes définies par $[Be(\alpha/2; y+a, n-y+b), Be(1-\alpha/2; y+a, n-y+b)]$. L'utilisation de l'a priori non-informatif de Jeffreys $\vartheta \sim Be(1/2, 1/2)$ conduit à l'*intervalle de Jeffreys* qui offre une alternative plus libérale que l'intervalle de Clopper-Pearson. Les bornes de cet intervalle sont définies par

$$CI_J = [Be(\alpha/2; y+1/2, n-y+1/2), Be(1-\alpha/2; y+1/2, n-y+1/2)].$$

Une forme analytique des bornes de CI_J s'obtient par l'approximation générale des quantiles d'une loi *Bêta* donnés dans Abramowitz et Stegun (1964) [1]. L'approximation de la borne supérieure $\hat{\theta}_J^{sup}(y)$ est obtenue par l'expression

$$\hat{\theta}_J^{sup}(y) = \frac{y+1/2}{n+1+(n-y+1/2)(e^{2w}-1)},$$

où $\kappa = z_{1-\alpha/2}$ et

$$\begin{aligned} w = & \frac{\kappa\sqrt{4\frac{y}{n}(1-\frac{y}{n})}/n + (\kappa^2-3)/(6n^2)}{4\frac{y}{n}(1-\frac{y}{n})} \\ & + \frac{(1/2-\frac{y}{n})(\frac{y}{n}(1-\frac{y}{n})(\kappa^2+2)-1/n)}{6n(\frac{y}{n}(1-\frac{y}{n})^2}. \end{aligned}$$

La borne inférieure $\hat{\theta}_J^{inf}(y)$ est obtenue par la même expression en remplaçant κ par $-\kappa$. D'autres formes analytiques des bornes de CI_J obtenues par séries de Edgeworth sont présentées dans [16].

Un autre intervalle de confiance dérivé de l'intervalle de Jeffreys permet d'optimiser le taux de recouvrement en diminuant l'importance des sauts situés sur les extrémités (voir section 4.3, figure (4.2)). Les bornes de l'*intervalle de Jeffreys modifié* (CI_{JM}) sont celles de l'intervalle de Jeffreys excepté pour certaines valeurs extrêmes de Y où une correction des bornes $\hat{\theta}_{JM}^{inf}(y)$ et $\hat{\theta}_{JM}^{sup}(y)$ vient augmenter la largeur de CI_J (voir Brown et Coll. (2001) [13]). Les bornes corrigées sont

$$\hat{\theta}_{JM}^{sup}(0) = p \text{ et } \hat{\theta}_{JM}^{inf}(n) = 1 - p,$$

où p satisfait la relation $(1-p)^n = \alpha/2$ ou encore $p = 1 - (\alpha/2)^{1/n}$, ainsi que

$$\hat{\theta}_{JM}^{inf}(1) = 0 \text{ et } \hat{\theta}_{JM}^{sup}(n-1) = 1.$$

Le dernier intervalle de confiance étudié est l'*intervalle corrigé au second-ordre* qui est fondé sur une optimisation du taux de recouvrement. Les bornes de cet intervalle sont construites pour annuler le biais du premier et du second ordre de l'approximation du taux de recouvrement unilatéral (voir section 4.3). Pour un niveau nominal $1-\alpha$, elles sont définies par

$$CI_2 = \frac{Y+\eta}{n+2\eta} \pm \kappa(V(\hat{\theta}^{MV}) + (\gamma_1 V(\hat{\theta}^{MV}) + \gamma_2)^{\frac{1}{2}})n^{-\frac{1}{2}},$$

où $\eta = \frac{1}{3}\kappa^2 + \frac{1}{6}$, $\gamma_1 = -\frac{13}{18}\kappa^2 - \frac{17}{18}$ et $\gamma_2 = \frac{1}{18}\kappa^2 + \frac{7}{36}$.

Les bornes de ces intervalles de confiance sont données dans les figures (4.19) à (4.26) pour les valeurs $(n = 5; Y = y)$ avec $y = y_1 = 1, 2, 3, 4, 5$ et $(n = 10; Y = y)$ avec $y = y_1 + y_2 = 0, 1, 2, 3, 4, 5$. Par symétrie, on obtient les bornes pour les autres valeurs de (n, y).

4.2 Méthode fréquentistes non-conditionnelles dans un plan séquentiel

L'approche fréquentiste non-conditionnelle pour l'estimation par intervalle et pour les tests dans les plans séquentiels nécessite des procédures d'ajustement qui dépendent des caractéristiques du plan expérimental. Ces ajustements posent une problématique calculatoire liée à la détermination de l'ordre avec lequel les séquences hypothétiques $(y_1, ..., y_M)$ infirment le contenu de H_0. Comme le mentionne Whitehead (1993) [55], ce choix introduit une part de subjectivité.

On peut par exemple définir un intervalle de confiance fréquentiste par l'ensemble des valeurs de $\theta = \theta_0$ contenues dans les hypothèses H_0 et non rejetées par le test. Dans le contexte d'un échantillonnage Binomial, la région d'acceptation $\mathcal{Y}_\theta^{H_0}$ est définie par l'ensemble des valeurs θ satisfaisant la condition $Pr(Y \in \mathcal{Y}_\theta^{H_0}|\theta) = 1-\alpha$. En inversant le test, on construit l'intervalle de confiance exact $CI_{1-\alpha} = \{\theta : Y \in \mathcal{Y}_\theta^{H_0}\}$.

Cet intervalle de confiance reste compliqué à construire dans le cas d'un plan séquentiel car la région d'acceptation $\mathcal{Y}_\theta^{H_0}$ est définie sur les observations hypothétiques non-extrêmes de la statistique exhaustive (M, Y) et non plus seulement de Y. L'intervalle $CI_{1-\alpha} = \{\theta : (M, Y) \in \mathcal{Y}_\theta^{H_0}\}$ est dépendant des conditions imposées sur $\mathcal{Y}_\theta^{H_0} = \{(m, y) : Pr((M, Y) \geq (m, y) \,|\, \theta) > \alpha\}$. De la même façon, on détermine la probabilité de rejet de H_0 par

$P_{\theta=\theta_0}$(d'obtenir des valeurs de (m, y) au moins aussi extrêmes que celles observées).

Il faut au préalable ordonner les observations dans l'espace

$$\mathcal{Y}_{(m,y)} = \{(k, y) : y \in \mathcal{Y}_{(k,y)}^{H_0} \cup \mathcal{Y}_{(k,y)}^{H_1}; k = 1, .., m\}.$$

Il existe des règles simples pour définir un ordre ; dans la suite $(k', y') > (k, y)$ signifie que (k', y') est mieux classé que (k, y). L'ordre du maximum de vraisemblance établi que $(k', y') > (k, y)$ si $\hat{\theta}_{(y')}^{MV} > \hat{\theta}_{(y)}^{MV}$. Chang et O'Brien (1986) [19] ont proposé un ordre basé sur le rapport de vraisemblance. L'espace des observations est ordonné

suivant l'éloignement par rapport à $\theta = \theta_0$ mesuré par la statistique de test :

$$(k', y') > (k, y) \text{ si } \frac{p((k', y')|\hat{\theta}^{MV}_{(y')})}{p((k', y')|\theta_0)} > \frac{p((k, y)|\hat{\theta}^{MV}_{(y)})}{p((k, y)|\theta_0)},$$

où $\hat{\theta}^{MV}_{(y)}$ est l'EMV basé sur (M, Y).

Dans le plan expérimental qui intègre une règle d'arrêt définie par les limites $(S_{1,L}, S_{1,U})$, ...,$(S_{K,L}, S_{K,U})$; $k = 1, ..., K$, Jung et Kim (2004)[39] proposent un ordre des doublets (k, y) dans l'espace $\mathcal{Y}_{(m,y)}$ établi suivant celui des $\hat{\theta}^{VUMSB}_{(m,y)}$ correspondants :

$$\begin{aligned}
\hat{\theta}^{VUMSB}_{(1,0)} < \hat{\theta}^{VUMSB}_{(1,1)} < ... < \hat{\theta}^{VUMSB}_{(1,S_{1,L})} \\
< \hat{\theta}^{VUMSB}_{(2,S_{1,L}+1)} < ... < \hat{\theta}^{VUMSB}_{(2,S_{2,L})} \\
... \\
< \hat{\theta}^{VUMSB}_{(K,S_{K-1,L}+1)} < ... < \hat{\theta}^{VUMSB}_{(K,S_{K,L})} \\
< \hat{\theta}^{VUMSB}_{(K,S_{K,U})} < ... < \hat{\theta}^{VUMSB}_{(K,S_{K-1,U}-1+n_k)} \\
< \hat{\theta}^{VUMSB}_{(K-1,S_{K-1,U})} < ... < \hat{\theta}^{VUMSB}_{(K-1,S_{K-2,U}-1+n_{k-1})} \\
... \\
< \hat{\theta}^{VUMSB}_{(1,S_{1,U})} < ... < \hat{\theta}^{VUMSB}_{(1,n_1)}.
\end{aligned}$$

Cet ordre stochastique est démontré dans l'annexe C de Jung et Kim (2004)[39] pour un plan séquentiel comportant 2 périodes (i.e. $K = 2$).

On démontre également que cet ordre correspond à celui donné dans un cas plus général par Jennison et Turnbull (voir Jennison et Turnbull (1983) [37] et Duffy et Santner (1987) [30]). On a la relation $(k', y') > (k, y)$ si

1. $k' = k$ et $\hat{\theta}_{(k',y')} \geq \hat{\theta}_{(k,y)}$
2. $k' < k$ et $\hat{\theta}_{(k',y')} \geq S_{(k',U)}$
3. $k' > k$ et $\hat{\theta}_{(k',y')} \leq S_{(k,L)}$

Ensuite, en fonction de l'ordre utilisé et conditionnellement à l'observation (M, Y)=(m, y), les bornes $(\hat{\theta}^{inf}, \hat{\theta}^{sup})$ de l'intervalle de confiance contrôlé unilatéralement à gauche

et à droite au niveau $1-\alpha/2$ sont les solutions du système

$$\begin{cases} Pr((M,Y) \geq (m,y)|\theta = \hat{\theta}^{sup}) = \alpha/2, \\ Pr((M,Y) \leq (m,y)|\theta = \hat{\theta}^{inf}) = \alpha/2. \end{cases}$$

4.3 Taux de recouvrement et longueur moyenne

Le taux de recouvrement d'un intervalle de confiance est une fonction discontinue en θ et n, construite sur la base des fonctions de répartition de la loi $Bin(\theta, n)$. Pour les intervalles unilatéraux $[0, \hat{\theta}^{sup}(y)]$ et $[\hat{\theta}^{inf}(y), 1]$, on a les expressions

$$\begin{aligned} C^{sup}(\theta, n) &= \sum_{i=0}^{n} I_{\{\theta \leq \hat{\theta}^{sup}(i)\}} P_\theta(Y = i) \\ &= P_\theta(Y \geq i) I_{\{\theta \leq \hat{\theta}^{sup}(i)\}} \\ C^{inf}(\theta, n) &= \sum_{i=0}^{n} I_{\{\theta \geq \hat{\theta}^{inf}(i)\}} P_\theta(Y = i) \\ &= P_\theta(Y \leq i) I_{\{\theta \geq \hat{\theta}^{inf}(i)\}}. \end{aligned} \tag{4.3.1}$$

Pour la borne supérieure, si $\theta \leq \hat{\theta}^{sup}(0)$ l'ensemble des intervalles unilatéraux $[0, \hat{\theta}^{sup}(i)]$ $i = 0, .., n$ contiennent θ et $C^{sup}(\theta, n) = 1$. Si $\hat{\theta}^{sup}(0) < \theta \leq \hat{\theta}^{sup}(1)$ l'intervalle $[0, \hat{\theta}^{sup}(0)]$ ne contient plus θ et $C^{sup}(\theta, n)$ diminue brusquement dès que θ dépasse la borne $\hat{\theta}^{sup}(0)$. Ensuite, $C^{sup}(\theta, n)$ augmente avec la valeur de θ car la probabilité associée à $Y = 0$ diminue jusqu'à la borne $\hat{\theta}^{sup}(1)$ où un saut semblable se produit lors de son franchissement par θ. Le prolongement de ce raisonnement jusqu'à la borne $\hat{\theta}^{sup}(n)$ permet d'expliquer l'allure hachée des courbes associées à $C^{inf}(\theta, n)$ et à $C^{sup}(\theta, n)$.

L'expression (4.3.1) montre que si $\theta \in]\hat{\theta}^{sup}(i-1), \hat{\theta}^{sup}(i)]$ (resp. $\theta \in [\hat{\theta}^{inf}(i-1), \hat{\theta}^{inf}(i)[$) avec $i = 1, .., n$, les fonctions $P_\theta(Y \geq i)$ (resp. $P_\theta(Y \leq i)$) servent de fonctions supports au taux de recouvrement de la borne supérieure (resp. inférieure). Pour visualiser les taux de recouvrement unilatéraux sur un même graphique, on représente celui de la borne inférieure par $1 - C^{inf}(\theta, n)$, qui utilise également les fonctions $P_\theta(Y \geq i)$ comme fonctions supports.

La figure (4.1) décrit les cinq fonctions supports $P_\theta(Y \geq i)$ avec $n=5$. Elle montre également les six valeurs de $\hat{\theta}_J^{inf}(i)$ et de $\hat{\theta}_J^{sup}(i)$ avec $i = 0, .., 5$ pour l'intervalle de Jeffreys 90% qui délimitent respectivement les cinq intervalles $]\hat{\theta}_J^{sup}(i-1), \hat{\theta}_J^{sup}(i)]$ et $[\hat{\theta}_J^{inf}(i-1), \hat{\theta}_J^{inf}(i)[$, où l'une des cinq fonctions supports se conjugue avec $C_J^{sup}(\theta, n)$ et $1 - C_J^{inf}(\theta, n)$. Les taux de recouvrements unilatéraux pour l'intervalle de Jeffreys sont décrits dans la figure (4.2). Les courbes présentées résultent du découpage des fonctions supports $P_\theta(Y \geq i)$ par les bornes $\hat{\theta}_J^{inf}(i)$ et $\hat{\theta}_J^{sup}(i)$.

Le taux de recouvrement est une fonction qui peut être formulée analytiquement. En suivant l'approche développée par Bhattacharya et Rao (1976) [10], on aboutit à une approximation du premier ordre en $O(n^{-1})$ de

$$P_\theta\left(\frac{n^{1/2}(\hat{\theta}^{MV} - \theta)}{(\theta(1-\theta))^{1/2}} \preceq z\right) = \\ \Phi(z) + [(\frac{1}{2} - g(z,\theta)) + \frac{1}{6}(1-2\theta)(1-z^2)]\phi(z)(n\theta(1-\theta))^{-1/2} + O(n^{-1}), \tag{4.3.2}$$

où $\phi(z)$ est la densité d'une loi Normale centrée réduite, $\Phi(z)$ sa fonction de répartition et $g(z,\theta) = g(z,\theta,n) = n\theta + n^{1/2}V^{1/2}z - [n\theta + n^{1/2}V^{1/2}z]_$, $[]_$ désignant la partie entière.

Suivant la même approche, une approximation jusqu'au second ordre en $O(n^{-3/2})$ est donnée dans Brown et Coll. (2002) [14]. Partant de cette dernière expression, on peut approcher le taux de recouvrement unilatéral $C^{sup}(\theta, n) = P_\theta(\theta < \hat{\theta}^{sup})$, où $\hat{\theta}^{sup}$ reprend l'une des formes analytiques des bornes des intervalles donnés dans la section 4.1. Pour z dans l'équation (4.3.2), on pose une forme linéaire du type $z = z(n) = \lambda_1 + \lambda_2 n^{-1/2} + \lambda_3 n^{-1} + O(n^{-1/2})$. Cette forme permet d'exprimer $\phi(z)$ et $\Phi(z)$ en fonction de λ_1, λ_1, λ_3, $\phi(\lambda_1)$ et $\Phi(\lambda_1)$. Les constantes λ_1, λ_2 et λ_3 sont ensuite déterminées sur la base d'une identification des termes en $n^{-1/2}$ et n^{-1} avec la borne $\hat{\theta}^{sup}$. On mène la même démarche pour le calcul du taux de recouvrement $C^{inf}(\theta, n)$.

Les approximations des taux de recouvrement bilatéraux et unilatéraux par séries

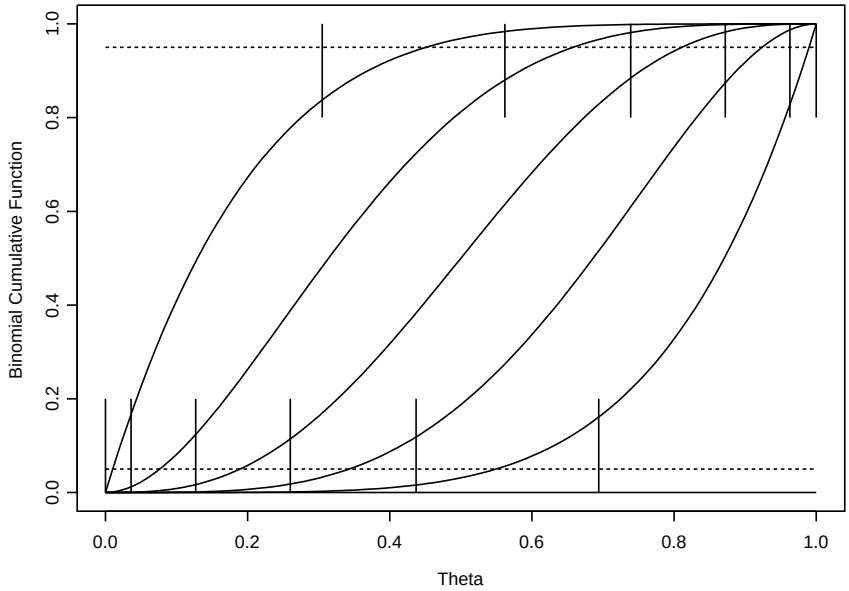

FIG. 4.1 – *Fonctions* $P_\theta(Y \geq i)$ $i = 1, .., 5$ *où* $Y \sim Bin(\theta, n=5)$ *avec un découpage par les bornes* $\theta_J^{inf}(y)$ *et* $\theta_J^{sup}(y)$ *de l'intervalle de Jeffreys 90%*

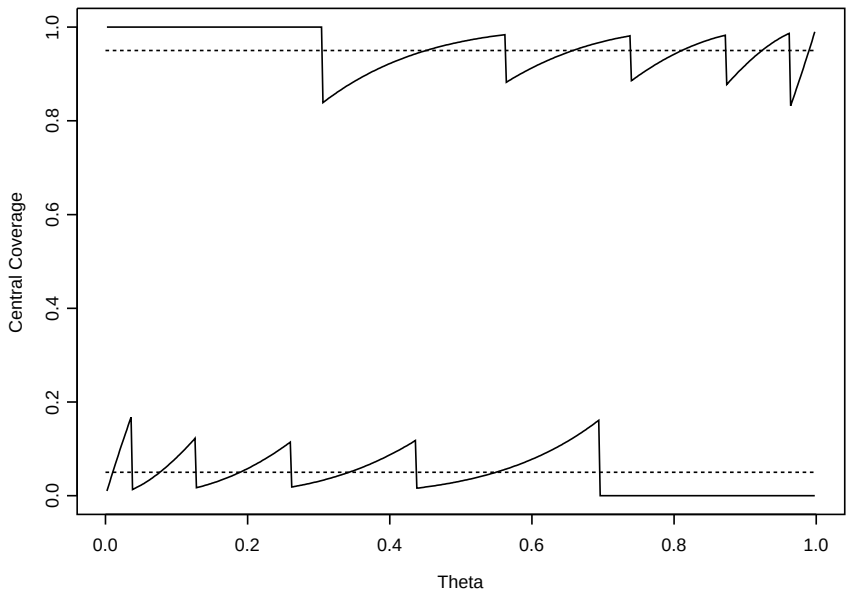

FIG. 4.2 – *Taux de recouvrement de l'intervalle de Jeffreys 90% (n=5)*

de Edgeworth au second ordre des différents intervalles de confiance sont données respectivement dans Brown et Coll. (2002) [14] et Cai (2003) [16]. D'une manière générale, ces approximations présentent une forme du type

$$\begin{aligned} C(\theta, n) =& 1 - \alpha + \text{"}O(n^{-\frac{1}{2}})\, Dviation\text{"} + \text{"}O(n^{-\frac{1}{2}})\, Oscillation\text{"} \\ &+ \text{"}O(n^{-1})\, Dviation\text{"} + \text{"}O(n^{-1})\, Oscillation\text{"} + O(n^{-\frac{3}{2}}), \end{aligned}$$

où les termes "$O(n^{-\frac{1}{2}})\, Dviation$" et "$O(n^{-1})\, Dviation$" décrivent les mouvements non-oscillatoires liés au biais des méthodes d'estimation des bornes.

Les taux de recouvrement pour les intervalles de confiance mentionnés dans la section 4.1 sont décrits dans la suite. Cette étude permet d'évaluer les performances de chacun sur un échantillon de taille très restreinte ($n = 5$).

L'étroitesse de l'intervalle de Jeffreys pour les observations Y non-centrales (i.e. $Y = 0, 1, 4, 5$) est à l'origine des deux sauts situés sur les extrémités des courbes de $C^{sup}(\theta, n)$ et $1 - C^{inf}(\theta, n)$ dans la figure (4.2). Pour la borne supérieure, le taux de recouvrement de $\hat{\theta}_J^{sup}(y)$ subit un saut important suivant que θ est situé en deçà ou au delà de $\hat{\theta}_J^{sup}(0)$ ou de $\hat{\theta}_J^{sup}(n-1)$. L'idée sous jacente à l'intervalle de Jeffreys modifié (voir section (4.1)) est d'élargir l'intervalle de Jeffreys pour les valeurs $Y = 0, n-1$ afin de ramener $C_{JM}^{sup}(\hat{\theta}_{JM}^{sup}(0), n)$ au niveau nominal $1 - \alpha/2$ et d'éliminer le saut correspondants à $C_J^{sup}(\hat{\theta}_J^{sup}(n-1), n)$. En menant la même démarche pour la borne inférieure, on aboutit aux fonctions $C_{JM}^{sup}(\theta, n)$ et $1 - C_{JM}^{inf}(\theta, n)$ décrites dans la figure (4.3) pour $n = 5$.

Le graphique (4.4) permet de visualiser la propriété remarquable de l'intervalle de Clopper-Pearson qui présente des taux de recouvrements unilatéraux qui ne sont jamais inférieurs au niveau nominal $1 - \alpha/2$. Les taux de recouvrement des intervalles de Wilson, Agresti-Coull et de l'intervalle corrigé au second-ordre sont respectivement décrits dans les figures (4.5), (4.6) et (4.7).

En plus du taux de recouvrement, la longueur moyenne d'un intervalle de confiance

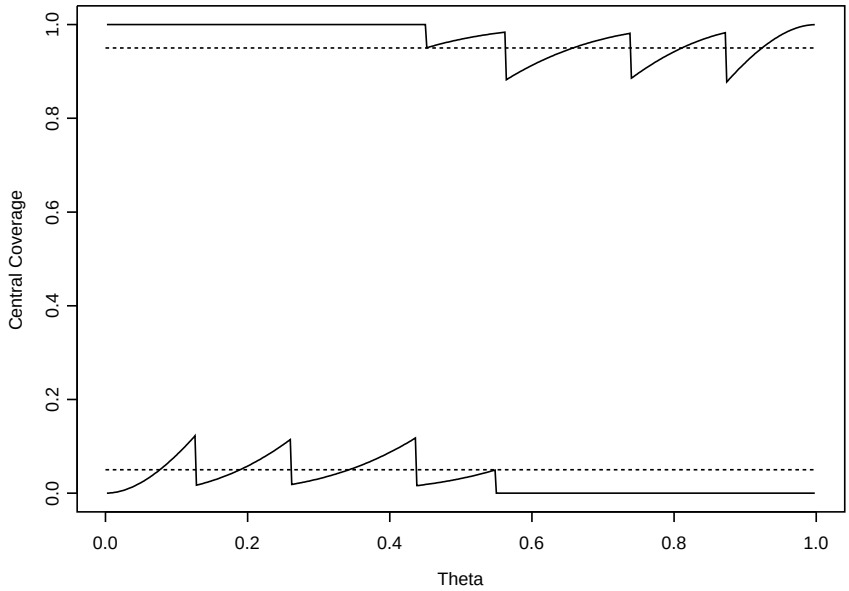

FIG. 4.3 – *Taux de recouvrement de l'intervalle de Jeffreys modifié 90%* ($n=5$)

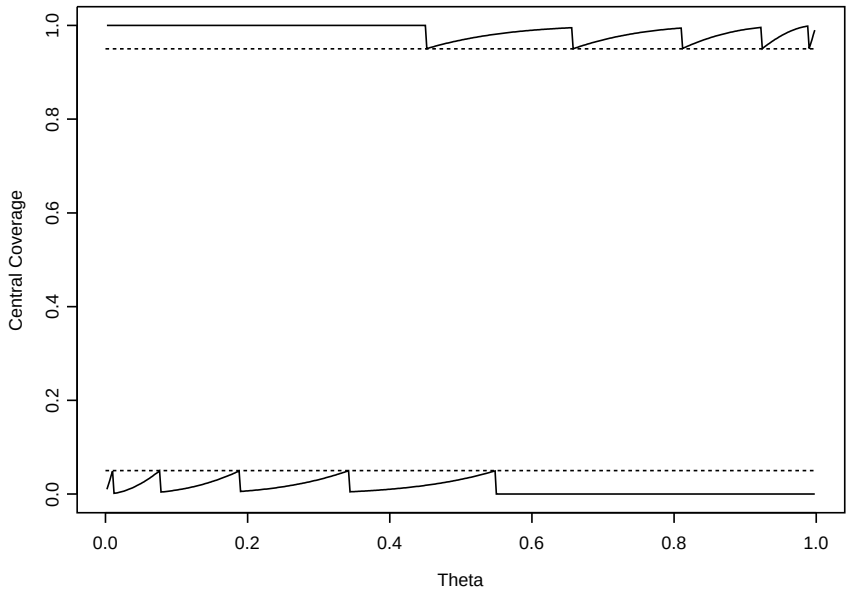

FIG. 4.4 – *Taux de recouvrement de l'intervalle de Clopper-Pearson 90%* ($n=5$)

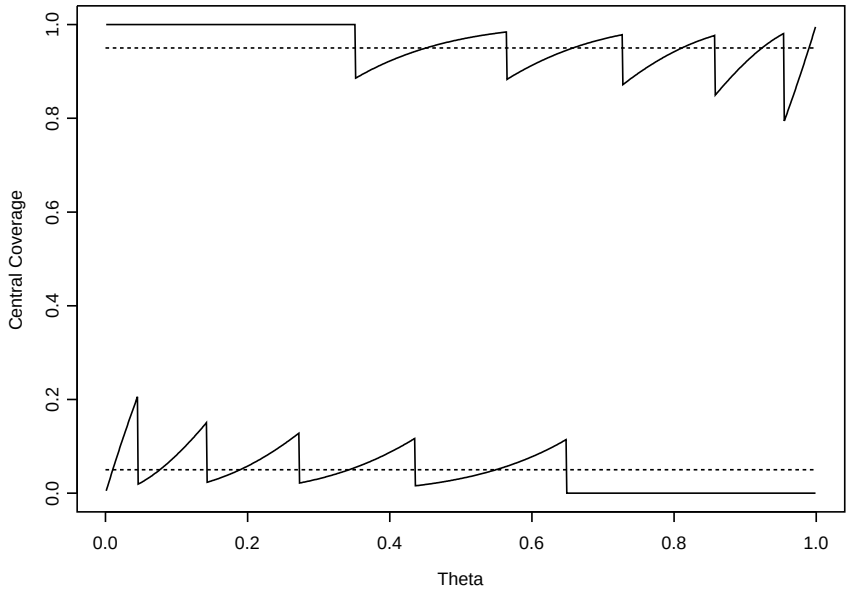

FIG. 4.5 – *Taux de recouvrement de l'intervalle de Wilson 90%* (n=5)

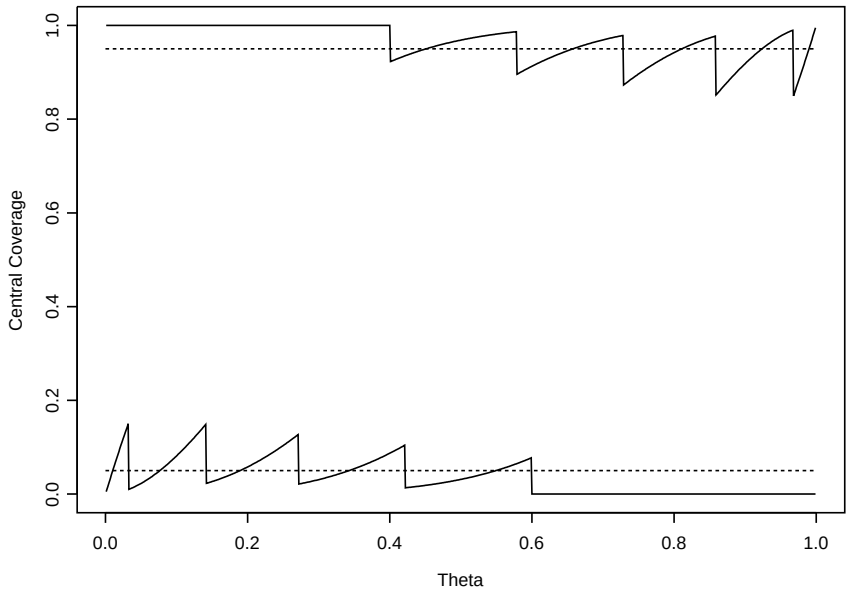

FIG. 4.6 – *Taux de recouvrement de l'intervalle d'Agresti-Coull 90%* (n=5)

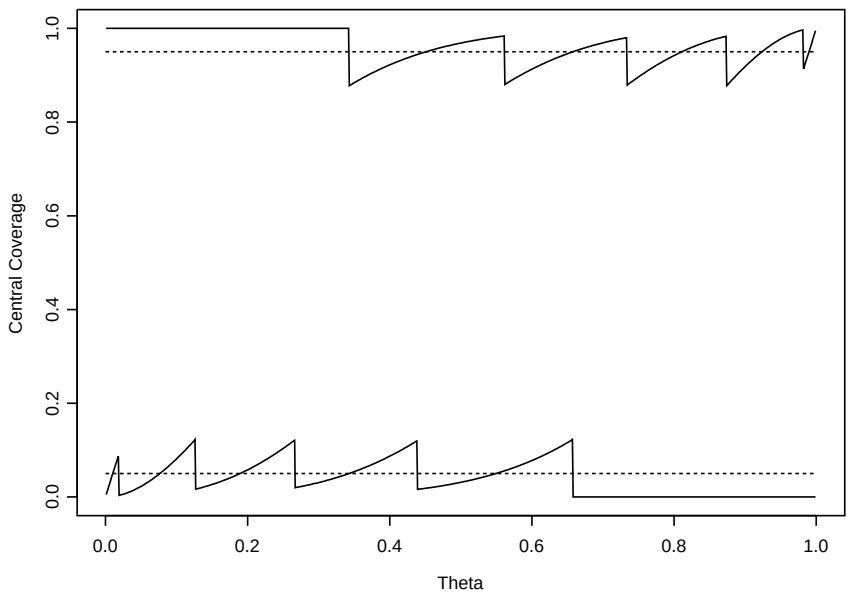

FIG. 4.7 – *Taux de recouvrement de l'intervalle corrigé au second-ordre 90%* (n=5)

est un critère de choix fondamental. On la définie par

$$L(\theta, n) = \sum_{i=0}^{n} (\hat{\theta}^{sup}(i) - \hat{\theta}^{inf}(i)) P_\theta(Y = i).$$

La figure (4.8) montre la longueur moyenne des six intervalles de confiance étudiés. Pour les intervalles ayant une justification bayésienne, le conservatisme de l'intervalle de Clopper-Pearson se traduit par une longueur moyenne uniformément la plus grande. A l'opposé, l'intervalle de Jeffreys est parmi tous les intervalles celui qui présente la longueur moyenne la plus petite. Entre les deux, la courbe associée à l'intervalle de Jeffreys modifié rejoint celle de l'intervalle de Clopper-Pearson lorsque $\theta = 0$ et $\theta = 1$.

L'intervalle corrigé au second-ordre présente une courbe de longueur moyenne proche mais située au dessus de celle de l'intervalle de Jeffreys, alors que la courbe associée à l'intervalle d'Agresti-Coull est proche mais située au dessous de celle de l'intervalle de Jeffreys modifié. L'intervalle de Jeffreys est celui qui présente le profil de longueur

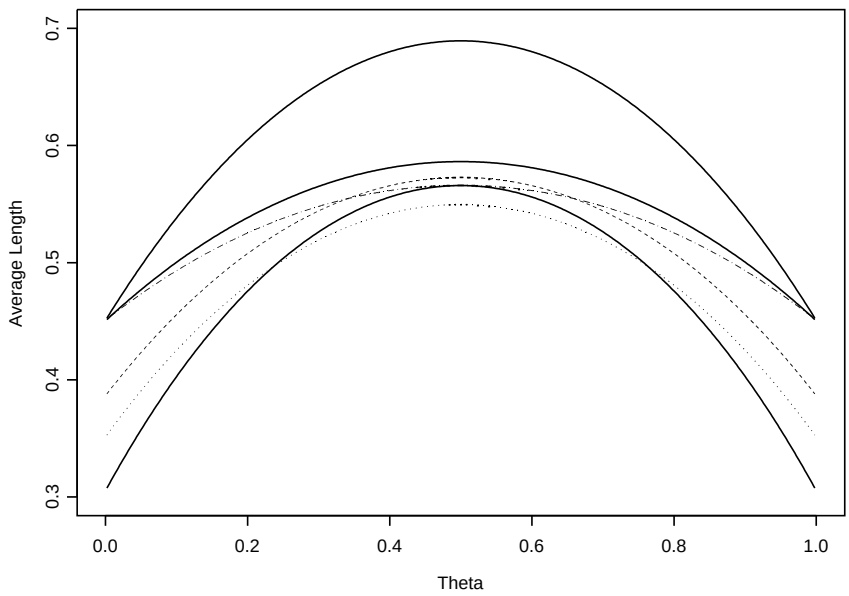

FIG. 4.8 – *Longueur moyenne des intervalles de confiance à 90%* (n=5). *Les intervalles de* CI_J, CI_{JM} *et* CI_{CP} *(—) sont respectivement représentés de bas en haut avec également les intervalles* CI_W *(· · ·),* CI_{AC} *(· − ·) et* CI_2 *(- - -).*

moyenne le plus avantageux.

Dans une perspective d'intervalles de confiance contrôlés unilatéralement dans l'échantillonnage fixe, Cai (2003) [16] recommande l'utilisation de l'intervalle de Jeffreys et de l'intervalle corrigé au second-ordre (l'intervalle de Jeffreys modifié n'est pas étudié dans cet article).

Dans le contexte d'un intervalle de confiance dont la méthode d'estimation des bornes ne tient pas compte de la nature du plan expérimental, le biais induit par la règle d'arrêt se traduit en terme de modification du taux de recouvrement par rapport à un plan comportant un échantillonnage fixe. Par exemple, si la zone de continuation de l'expérience pour Y_1 se réduit à $J = \{0\}$ (i.e. on considère le plan $(n_1, S_{1,U} = 1, n_2)$), alors les fonctions $C^{sup}(\theta; n_1, r; S_{1,U})$ et $C^{inf}(\theta; n_1, r; S_{1,U})$ sont fondées sur le recouvrement des n_1 intervalles de confiance $CI(1,Y)$ associés à $y = y_1 = 1, .., n_1$ (i.e. $y \in \mathcal{Y}_{(1)}$) mais également des $n_2 + 1$ autres intervalles de confiance $CI(2,Y)$

associés à $y = y_2 = 0, .., n_2$ (i.e. $y \in \mathcal{Y}_{(2)}$).

Dans le plan $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$, l'intervalle unilatéral associé à la borne supérieure $CI^{sup}(2,Y) = [0, \hat{\theta}^{sup}(2,Y)]$ vient remplacer l'intervalle $CI^{sup}(0) = [0, \hat{\theta}^{sup}(0)]$ obtenu dans l'échantillonnage fixe $(n = n_1)$. Par exemple, pour les valeurs $\theta \in [\hat{\theta}^{sup}(2,0), \hat{\theta}^{sup}(0)]$ qui était contenues par $CI^{sup}(0)$ et qui ne le sont plus par l'intervalle $CI^{sup}(2,Y)$ pour certaines valeurs $y \in \mathcal{Y}_{(2)}$, il résulte une diminution du taux de recouvrement $C^{sup}(\theta; n_1, r; S_{1,U})$ que l'on quantifie avec la probabilité d'échantillonnage $P_\theta(Y_1 = 0) \sum_{i=0}^{n_2} P_\theta(Y_2 = i) 1_{\theta \leq \hat{\theta}^{sup}(2,i)}$. Les fonctions $C^{sup}(\theta; n_1, r; S_{1,U})$ et $C^{inf}(\theta; n_1, r; S_{1,U})$ sont susceptibles de se différencier de $C^{sup}(\theta; n)$ et $C^{inf}(\theta; n)$ dans l'échantillonnage fixe pour les valeurs de θ respectivement comprises dans les intervalles $[0, \hat{\theta}^{sup}(2, n_2)]$ et $[\hat{\theta}^{inf}(2,0), 1]$.

Plus généralement, dans un plan du type $(n_1, S_{1,U}, S_{1,L}, n_2)$ associé à la zone de continuation $J = [0, S_{1,U}[\cup]S_{1,L}, n_1]$ (i.e. $y \in \bar{\mathcal{Y}}_{(1)}$) et à la zone d'arrêt $\bar{J} = [S_{1,U}, S_{1,L}]$ (i.e. $y \in \mathcal{Y}_{(1)}$), les valeurs de θ peuvent être recouvertes par les intervalles unilatéraux $CI^{sup}(1,Y)$ et/ou $CI^{sup}(2,Y)$ pour la borne supérieure et par $CI^{inf}(1,Y)$ et/ou $CI^{inf}(2,Y)$ pour la borne inférieure, avec $y \in \mathcal{Y}_{(1)} \cup \mathcal{Y}_{(2)}$. Pour les intervalles unilatéraux $CI^{sup}(M,Y)$ et $CI^{inf}(M,Y)$, l'expression des taux de recouvrement se formalise par

$$\begin{aligned}
C^{sup}(\theta; n_1, r; S_{1,U}, S_{1,L}) &= E_{Seq,\theta}[1_{\theta \leq \hat{\theta}^{sup}(M,Y)}] \\
&= \sum_{i \in \mathcal{Y}_{(1)}} P_\theta(Y_1 = i)\, 1_{\{\theta \leq \hat{\theta}^{sup}(1,i)\}} \\
&\quad + \sum_{\substack{i+j=k \\ i \in \bar{\mathcal{Y}}_{(1)},\, k \in \mathcal{Y}_{(2)}}} P_\theta(Y_1 = i) P_\theta(Y_2 = j)\, 1_{\{\theta \leq \hat{\theta}^{sup}(2,k)\}} \\
C^{inf}(\theta; n_1, r; S_{1,U}, S_{1,L}) &= E_{Seq,\theta}[1_{\theta \geq \hat{\theta}^{inf}(M,Y)}] \\
&= \sum_{i \in \mathcal{Y}_{(1)}} P_\theta(Y_1 = i)\, 1_{\{\theta \geq \hat{\theta}^{inf}(1,i)\}} \\
&\quad + \sum_{\substack{i+j=k \\ i \in \bar{\mathcal{Y}}_{(1)},\, k \in \mathcal{Y}_{(2)}}} P_\theta(Y_1 = i) P_\theta(Y_2 = j)\, 1_{\{\theta \geq \hat{\theta}^{inf}(2,k)\}}.
\end{aligned}$$

Cette expression montre que les taux de recouvrement unilatéraux sont construits sur

la base des fonctions $P_\theta(Y_1 \geq i)$ et $P_\theta(Y_1 = i)P_\theta(Y_2 = j)$, où $Y_1 \sim Bin(\theta, n_1)$ et $Y_2 \sim Bin(\theta, n_2)$. Les fonctions supports de $C^{sup}(\theta; n_1, r; S_{1,U}, S_{1,L})$ et $1-C^{inf}(\theta; n_1, r; S_{1,U}, S_{1,L})$ font donc partie d'un ensemble de fonctions définies par

$$P_\theta(Y_1 \geq i)\, 1_{i \in \mathcal{Y}_{(1)}} + \sum_{\substack{i+j=k \\ i \in \bar{\mathcal{Y}}_{(1)},\, k \in \mathcal{Y}_{(2)}}} P_\theta(Y_1 = i)P_\theta(Y_2 = j). \tag{4.3.3}$$

Dans le plan séquentiel $(n_1, S_{1,U} = 1, n_2)$, on déduit l'expression des taux de recouvrement unilatéraux

$$\begin{aligned} C^{sup}(\theta; n_1, r; S_{1,U} = 1) &= \sum_{i=1}^{n_1} P_\theta(Y_1 = i)\, 1_{\{\theta \leq \hat{\theta}^{sup}(1,i)\}} \\ &\quad + P_\theta(Y_1 = 0) \sum_{j=0}^{n_2} P_\theta(Y_2 \geq j)\, 1_{\{\theta \leq \hat{\theta}^{sup}(2,j)\}} \\ C^{inf}(\theta; n_1, r; S_{1,U} = 1) &= \sum_{i=1}^{n_1} P_\theta(Y_1 = i)\, 1_{\{\theta \geq \hat{\theta}^{inf}(1,i)\}} \\ &\quad + P_\theta(Y_1 = 0) \sum_{j=0}^{n_2} P_\theta(Y_2 \leq j)\, 1_{\{\theta \geq \hat{\theta}^{inf}(2,j)\}}. \end{aligned}$$

D'après (4.3.3), l'ensemble des fonctions parmi lesquels on trouve les fonctions supports de $C^{sup}(\theta; n_1, r; S_{1,U} = 1)$ et $1 - C^{inf}(\theta; n_1, r; S_{1,U} = 1)$ sont

$$P_\theta(Y_1 \geq i) \text{ et } P_\theta(Y_1 \geq i) + P_\theta(Y_1 = 0)P_\theta(Y_2 \geq j),$$

avec $i = 1, .., n_1$ et $j = 0, .., n_2$. La figure (4.9) présente les courbes de ces fonctions dans le plan $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$.

Pour la borne supérieure, la sélection des fonctions supports s'effectue en classant au préalable les valeurs des bornes $\hat{\theta}^{sup}(m, y)$ avec $m = 1, 2$ et $y \in \mathcal{Y}$. Pour l'intervalle de Jeffreys dans le plan $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$, les résultats de la figure (4.17) permettent de classer par ordre croissant $\hat{\theta}^{sup}_J(2, 0) = 0.1708 < \hat{\theta}^{sup}_J(2, 1) = 0.3306 < \hat{\theta}^{sup}_J(2, 2) = 0.4525 < \hat{\theta}^{sup}_J(2, 3) = 0.5581 < \hat{\theta}^{sup}_J(1, 1) = 0.5628 < \hat{\theta}^{sup}_J(2, 4) = 0.6527 < \hat{\theta}^{sup}_J(2, 5) = 0.7381 < \hat{\theta}^{sup}_J(1, 2) = 0.7394 < \hat{\theta}^{sup}_J(1, 3) = 0.8722 < \hat{\theta}^{sup}_J(1, 4) = 0.9636 < \hat{\theta}^{sup}_J(1, 5) = 0.9996$. Ces onze valeurs définissent les douze intervalles de $[0, 1]$ dans lesquels les fonctions supports se conjuguent avec $C^{sup}_J(\theta; n_1, r; S_{1,U})$.

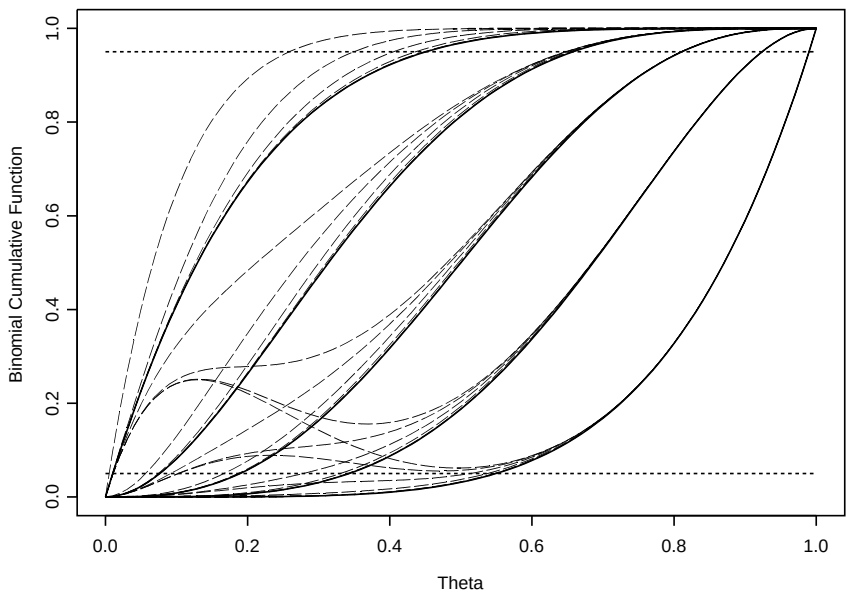

FIG. 4.9 – *Ensemble des fonctions* $P_\theta(Y_1 \geq i)$ $i = 1, .., 5$ *(—) et* $P_\theta(Y_1 \geq i) + P_\theta(Y_1 = 0)P_\theta(Y_2 \geq j)$ $i = 1, .., 5$ *et* $j = 0, .., 5$ *(– –) avec* $n_1 = n_2 = 5$

Si $\theta \in [0, \hat{\theta}_J^{sup}(2,0)]$, tous les intervalles unilatéraux $CI_J^{sup}(M, Y)$ contiennent θ et $C_J^{sup}(\theta; n_1, r; S_{1,U}) = 1$. Si $\theta \in]\hat{\theta}_J^{sup}(2,0), \hat{\theta}_J^{sup}(2,1)]$, l'intervalle $CI_J^{sup}(2,0)$ ne contient plus θ et $C_J^{sup}(\theta; n_1, r; S_{1,U}) = P_\theta(Y_1 \geq 1) + P_\theta(Y_1 = 0)P_\theta(Y_2 \geq 1)$. Ensuite, on a $C_J^{sup}(\theta; n_1, r; S_{1,U}) = P_\theta(Y_1 \geq 1) + P_\theta(Y_1 = 0)P_\theta(Y_2 \geq j)$ si $\theta \in]\hat{\theta}_J^{sup}(2, j-1), \hat{\theta}_J^{sup}(2,j)]$ pour $j = 2, 3$. Si $\theta \in]\hat{\theta}_J^{sup}(2,3), \hat{\theta}_J^{sup}(1,1)]$, on a $C_J^{sup}(\theta; n_1, r; S_{1,U}) = P_\theta(Y_1 \geq 1) + P_\theta(Y_1 = 0)P_\theta(Y_2 \geq 4)$. En revanche, si $\theta \in]\hat{\theta}_J^{sup}(1,1), \hat{\theta}_J^{sup}(2,4)]$, l'intervalle $CI_J^{sup}(1,1)$ ne contient plus θ et $C_J^{sup}(\theta; n_1, r; S_{1,U}) = P_\theta(Y_1 \geq 2) + P_\theta(Y_1 = 0)P_\theta(Y_2 \geq 4)$. Le taux de recouvrement associé aux valeurs $\theta \in]\hat{\theta}_J^{sup}(2,4), \hat{\theta}_J^{sup}(2,5)]$ est $C_J^{sup}(\theta; n_1, r; S_{1,U}) = P_\theta(Y_1 \geq 2) + P_\theta(Y_1 = 0)P_\theta(Y_2 \geq 4)$, mais au delà, plus aucun intervalle $CI_J^{sup}(2, Y_2)$ ne contient θ et pour $\theta \in]\hat{\theta}_J^{sup}(1, i-1), \hat{\theta}_J^{sup}(1,i)]$ avec $i \geq 3$, on a $C_J^{sup}(\theta; n_1, r; S_{1,U}) = P_\theta(Y_1 \geq i)$.

En menant le même raisonnement sur la borne inférieure, on sélectionne les fonctions qui correspondent à $C_J^{sup}(\theta; n_1, r; S_{1,U})$ ou $1 - C_J^{inf}(\theta; n_1, r; S_{1,U})$ sur au moins un intervalle de $[0, 1]$. Les courbes des fonctions supports associées à l'intervalle de

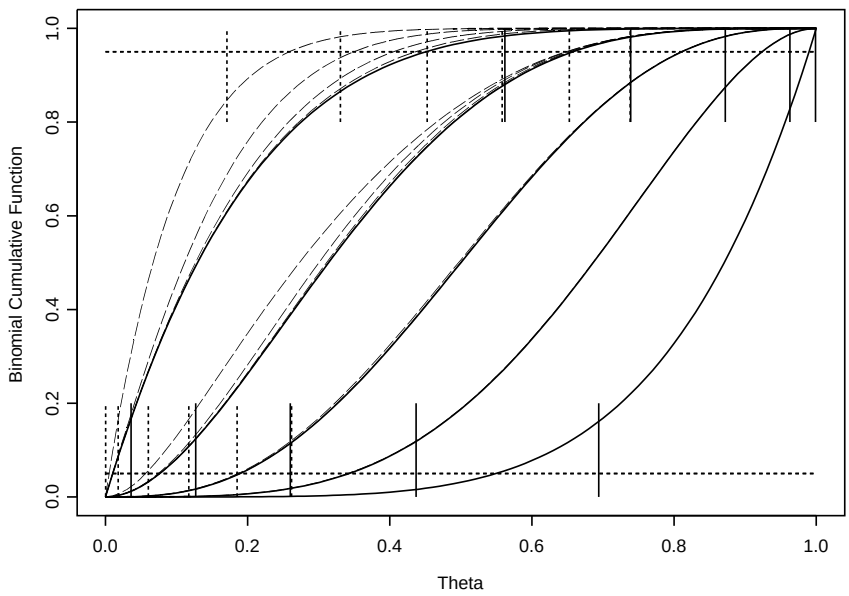

FIG. 4.10 – *Fonctions supports des taux de recouvrement de l'intervalle de Jeffreys 90% dans le plan* $(n_1=5, S_{1,U}=1, n_2=5)$ *avec le découpage par les bornes* $\hat{\theta}_J^{inf}(1,y)$ *et* $\hat{\theta}_J^{sup}(1,y)$ *avec* $y \in \mathcal{Y}_{(1)}$ *(—) et par* $\hat{\theta}_J^{inf}(2,y)$ *et* $\hat{\theta}_J^{sup}(2,y)$ *avec* $y \in \mathcal{Y}_{(2)}$ *(- - -)*

Jeffreys dans le plan $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ sont montrées dans la figure (4.10). On visualise également les bornes $\hat{\theta}_J^{inf}(m,y)$ et $\hat{\theta}_J^{sup}(m,y)$ de l'intervalle de Jeffreys pour $m = 1, 2$ et $y \in \mathcal{Y}_{(1)} \cup \mathcal{Y}_{(2)}$.

Les taux de recouvrements unilatéraux pour l'intervalle de Jeffreys dans le plan $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ sont montrés dans la figure (4.11). Par rapport à l'échantillonnage fixe $(n = 5)$, l'impact de la règle d'arrêt se traduit par une diminution de la zone de couverture à 100% de la borne supérieure (i.e. $C_J^{sup}(\theta; n_1, r; S_{1,U}) < 1$ lorsque $\hat{\theta}_J^{sup}(2,0) \leq \theta \leq \hat{\theta}_J^{sup}(0)$). D'autre part, pour les valeurs $\theta \leq \hat{\theta}_J^{sup}(1,1)$ et $\theta \leq \hat{\theta}_J^{inf}(1,2)$, les fonctions $C_J^{sup}(\theta; n_1, r; S_{1,U}, S_{1,L})$ et $C_J^{inf}(\theta; n_1, r; S_{1,U}, S_{1,L})$ respectivement sous-estiment et surestiment le niveau nominal $1 - \alpha/2$.

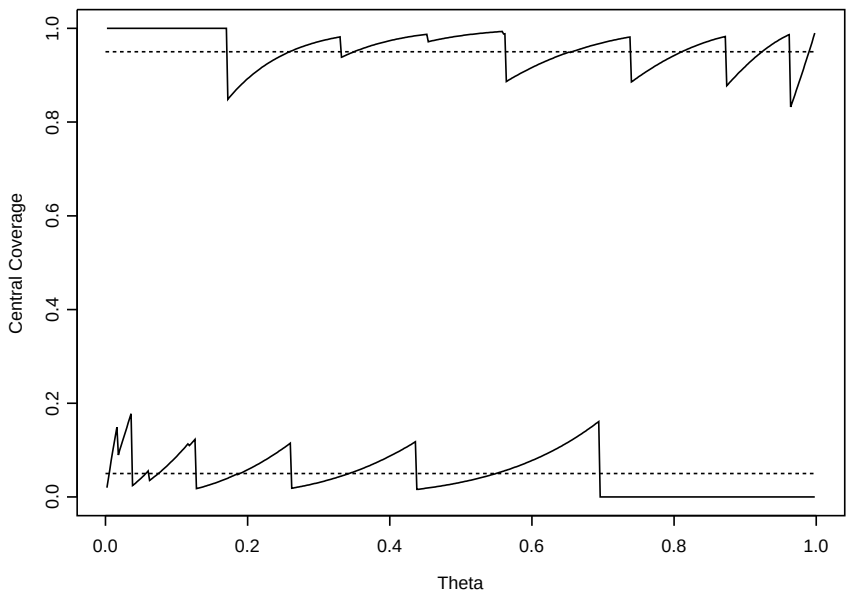

FIG. 4.11 – *Taux de recouvrement des intervalles 90% de Jeffreys (—) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

4.4 Construction de l'intervalle de Jeffreys corrigé

Dans un plan séquentiel d_{Seq} associé à la zone de continuation J pour les Y_1 succès sur les n_1 premières observations, l'a priori de Jeffreys corrigé est distribué suivant une loi $\vartheta | d_{Seq} \sim Be^J(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}; n_1, r; J)$. Dans la section 2.3, on montre que la correction de l'a priori de Jeffreys offre un mécanisme de compensation du biais induit par la règle d'arrêt du plan expérimental. L'objectif de cette section est d'étudier l'impact de cette correction sur la loi a posteriori et sur les bornes de l'intervalle de crédibilité correspondant. L'a priori de Jeffreys classique $\vartheta \sim Be(\frac{1}{2}, \frac{1}{2})$ engendre un intervalle noté intervalle de Jeffreys classique dans la suite.

A partir de la loi a posteriori

$$\vartheta | d_{Seq}, y \sim Be^J(1/2 + y, 1/2 + n - y, 1/2; n_1, r; J),$$

on construit les bornes de l'intervalle de Jeffreys corrigé contrôlé unilatéralement à

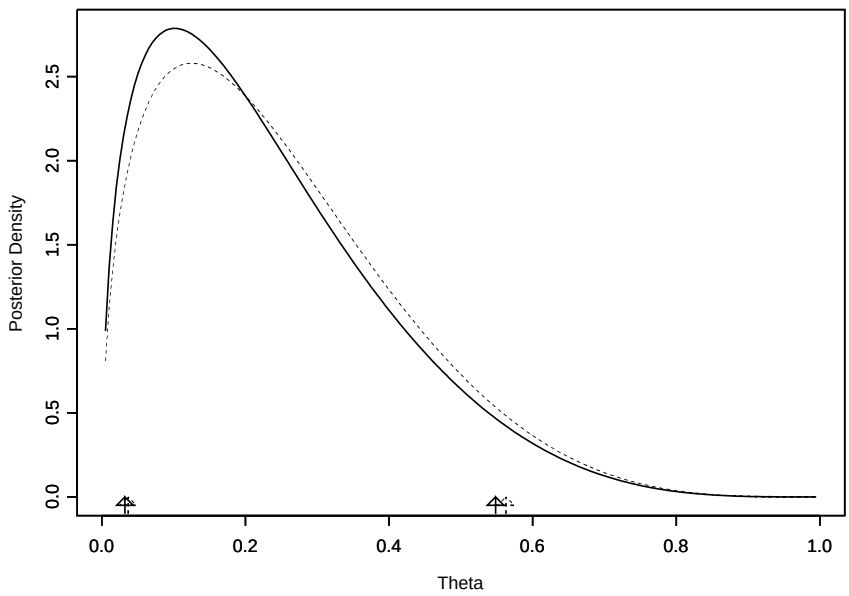

FIG. 4.12 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,1) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$

gauche et à droite au niveau nominal $1-\alpha/2$

$$\hat{\theta}_{JC}^{inf} = Be^J(\alpha/2; 1/2+y, 1/2+n-y, 1/2; n_1, r; J)$$
$$\hat{\theta}_{JC}^{sup} = Be^J(1-\alpha/2; 1/2+y, 1/2+n-y, 1/2; n_1, r; J),$$

où $Be^J(\alpha; a, b, c; n_1, r; J)$ est le α-quantile de la loi $Be^J(a, b, c; n_1, r; J)$.

Dans le plan séquentiel $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$, la zone de continuation de l'expérience en période 2 prend la forme $J = \{0\}$. Les densités a posteriori de $Be^J(1/2+y, 1/2+n-y, 1/2; n_1, r; \{0\})$ sur l'espace des observations $\mathcal{Y}_{(1)} = \{y = y_1; y = 1, 2, .., 5\}$ sont montrées dans les figures (4.12) à (4.16). Les densités a posteriori sur l'espace $\mathcal{Y}_{(2)} = \{y = y_2; y = 0, 1, .., 5\}$ sont montrées dans les figures (C.18) à (C.23). Ces graphiques décrivent également les densités a posteriori $Be(1/2+y, 1/2+n-y)$ obtenues avec l'a priori de Jeffreys classique sur le même espace des observations, ainsi que les bornes des intervalles de Jeffreys classique et corrigé contrôlés unilatéralement à gauche et à droite au niveau nominal $1-\alpha/2 = 0.95$.

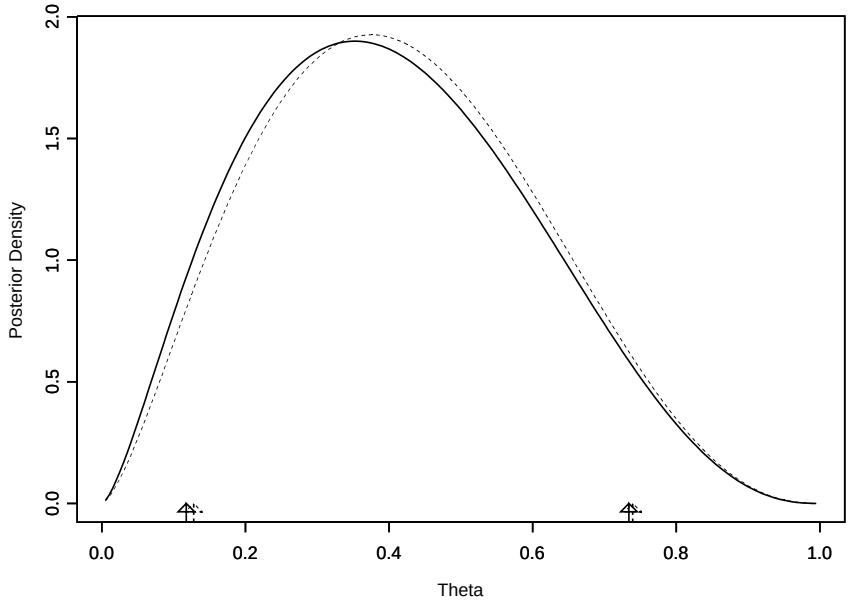

FIG. 4.13 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,2) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$

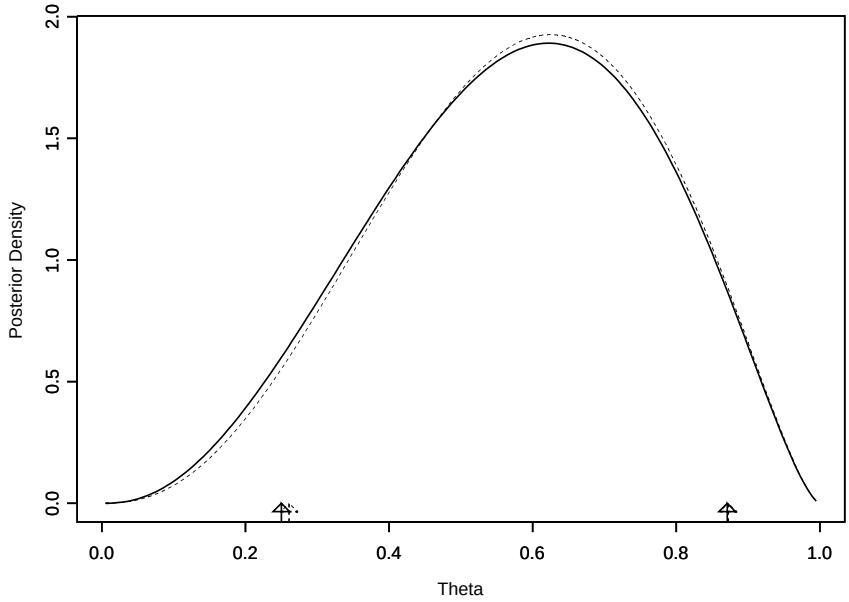

FIG. 4.14 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,3) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$

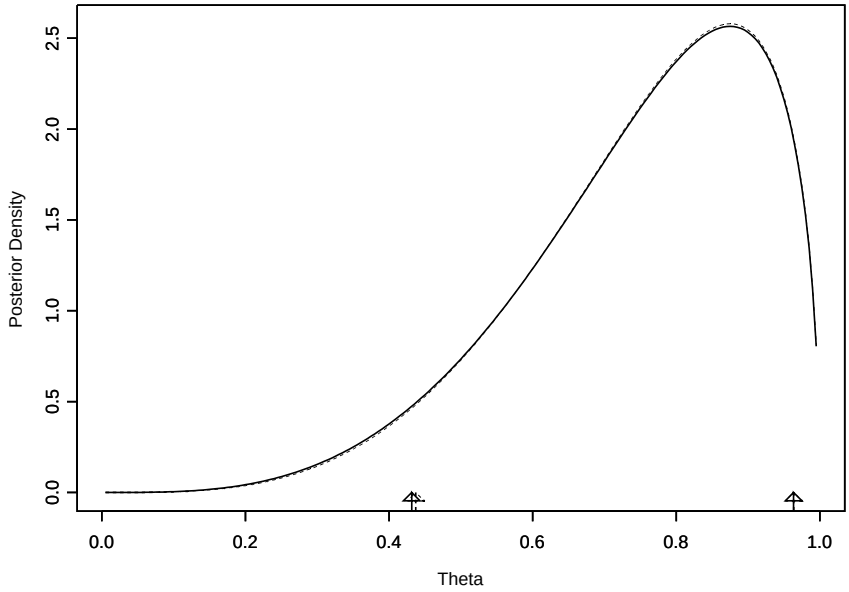

FIG. 4.15 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,4) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$

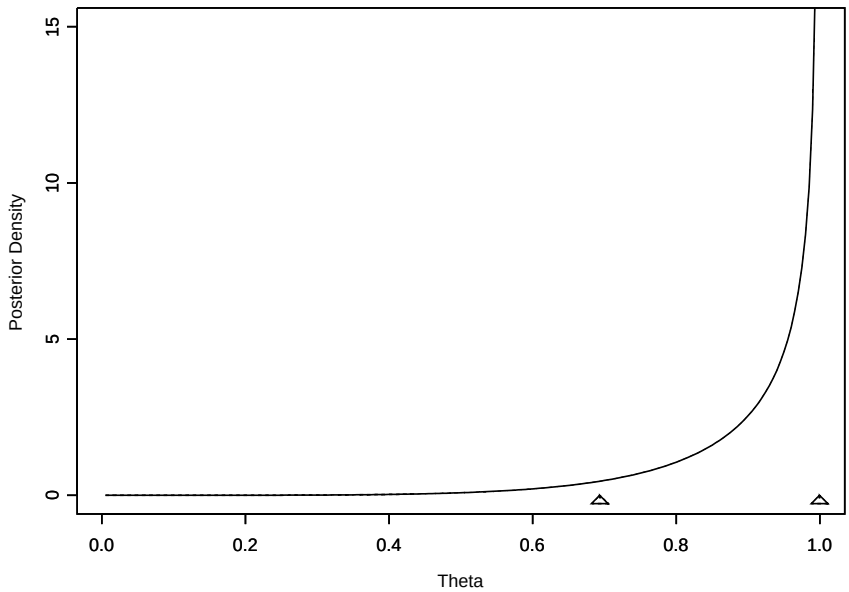

FIG. 4.16 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,5) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$

On observe que la correction de l'a priori de Jeffreys induit un glissement vers zéro des densités $Be^J(1/2+y, 1/2+n-y, 1/2; n_1, r; \{0\})$ par rapport aux densités $Be(1/2+y, 1/2+n-y)$. Les loi a posteriori les plus affectées par la correction sont celles correspondant aux observations (m, y) associées à un rapport y/n proche de $S_{1,U}/n_1 = 0.2$, car elles sont les plus concernées par l'aléas de la *va* M dans le choix de continuer ou non l'expérience en période 2. Les observations (m, y)=$\{(1,1),(1,2),(2,0),(2,1),(2,2),(2,3),(2,4)\}$ révèlent donc les différences les plus grandes. Elles sont également associées aux plus grandes corrections (par sous-estimation) des bornes de l'intervalle de Jeffreys, sauf lorsque $\hat{\theta}_J^{inf}(m, y)$ et $\hat{\theta}_J^{sup}(m, y)$ sont déjà très proches respectivement des valeurs 0 et 1.

Le tableau de la figure (4.17) fournit les valeurs numériques des bornes des intervalles de Jeffreys 90% classique ($\hat{\theta}_J^{inf}(m, y)$, $\hat{\theta}_J^{sup}(m, y)$) et corrigé ($\hat{\theta}_{JC}^{inf}(m, y)$, $\hat{\theta}_{JC}^{sup}(m, y)$) dans le plan ($n_1 = 5, S_{1,U} = 1, n_2 = 5$). Ces valeurs sont représentées dans le graphique attenant en fonction des valeurs de y/n où $n = 5, 10$ et $\mathcal{Y} = \mathcal{Y}_{(1)} \cup \mathcal{Y}_{(2)}$.

On obtient les bornes des intervalles de Jeffreys modifiés classique et corrigé respectivement à partir des bornes des intervalles de Jeffreys classique et corrigé, en posant en plus (voir section (4.1))

$$\hat{\theta}_{JM}^{sup}(2,0) = \hat{\theta}_{JMC}^{sup}(2,0) = 0.2589 \text{ et } \hat{\theta}_{JM}^{inf}(1,5) = \hat{\theta}_{JMC}^{inf}(1,5) = 0.5493,$$
$$\hat{\theta}_{JM}^{inf}(1,1) = \hat{\theta}_{JM}^{inf}(2,1) = \hat{\theta}_{JMC}^{inf}(1,1) = \hat{\theta}_{JMC}^{inf}(2,1) = 0 \text{ et } \hat{\theta}_{JM}^{sup}(4) = \hat{\theta}_{JMC}^{sup}(4) = 1.$$

Ces modifications provoquent un élargissement de l'intervalle de Jeffreys classique 90% que l'on visualise dans la figure (4.18).

Les taux de recouvrement unilatéraux des intervalles de Jeffreys classique et corrigé dans le plan séquentiel ($n_1 = 5, S_{1,U} = 1, n_2 = 5$) sont présentés dans la figure (4.19). La correction vient en compensation de l'effet de la règle d'arrêt. Le glissement vers 0 des bornes de l'intervalle CI_{JC} par rapport à CI_J se traduit par un décalage vers la gauche des sauts des courbes des taux de recouvrement. Les modifications les plus fortes interviennent autour des sauts correspondants aux bornes les plus affectées

	$\hat{\theta}_J^{inf}(m,y)$	$\hat{\theta}_J^{sup}(m,y)$	$\hat{\theta}_{JC}^{inf}(m,y)$	$\hat{\theta}_{JC}^{sup}(m,y)$
(1,1)	0.0364	0.5628	0.0321	0.5484
(1,2)	0.1278	0.7394	0.1175	0.7341
(1,3)	0.2606	0.8722	0.2500	0.8710
(1,4)	0.4372	0.9636	0.4315	0.9634
(1,5)	0.6942	0.9996	0.6933	0.9996
(2,0)	0.0002	0.1708	0.0002	0.1593
(2,1)	0.0179	0.3306	0.0164	0.3176
(2,2)	0.0602	0.4525	0.0560	0.4425
(2,3)	0.1173	0.5581	0.1110	0.5517
(2,4)	0.1851	0.6527	0.1780	0.6491
(2,5)	0.2619	0.7381	0.2553	0.7363

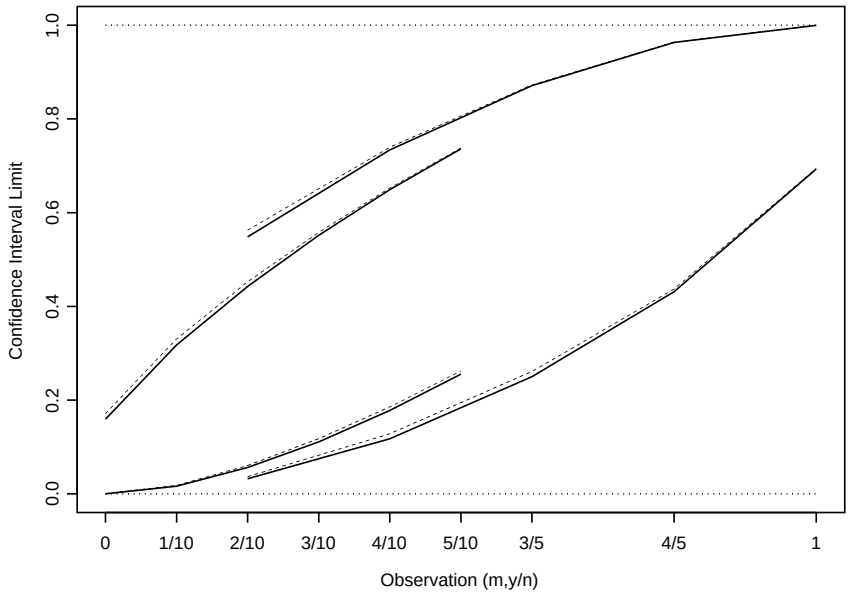

FIG. 4.17 – *Bornes des intervalles de Jeffreys 90% classique (- - -) et corrigé (—) dans le plan séquentiel* ($n_1=5, S_{1,U}=1, n_2=5$)

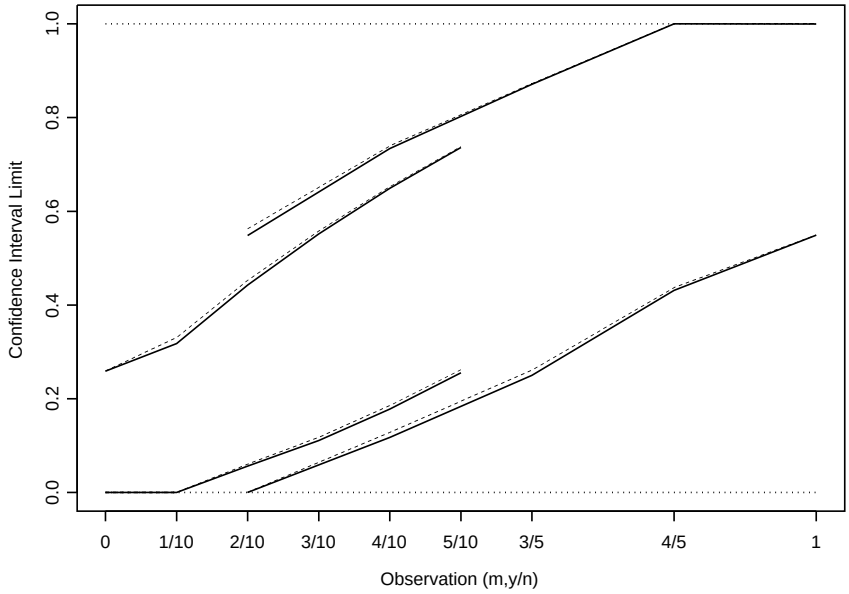

FIG. 4.18 – *Bornes des intervalles de Jeffreys 90% modifié classique (- - -) et Jeffeys modifié corrigé (—) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$

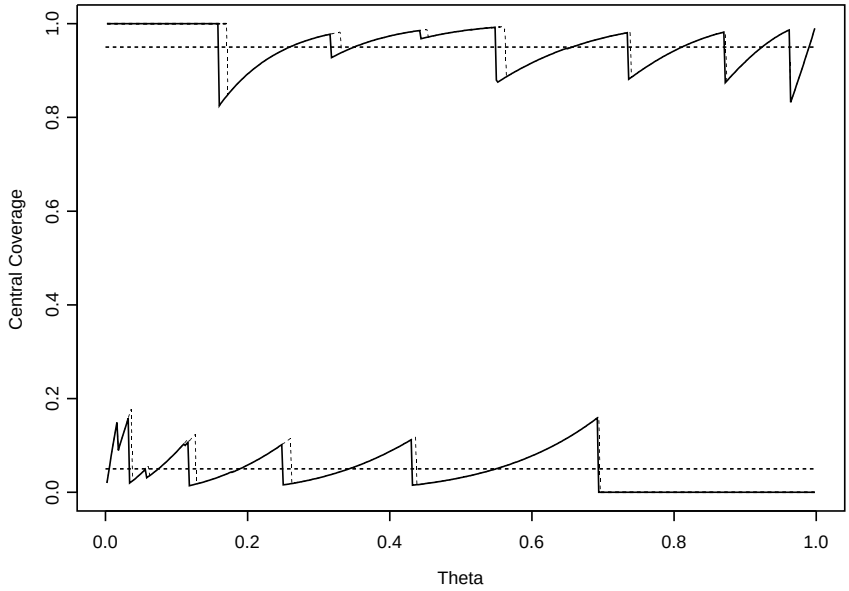

FIG. 4.19 – *Taux de recouvrement des intervalles de Jeffreys 90% classique (- - -) et corrigé (—) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$

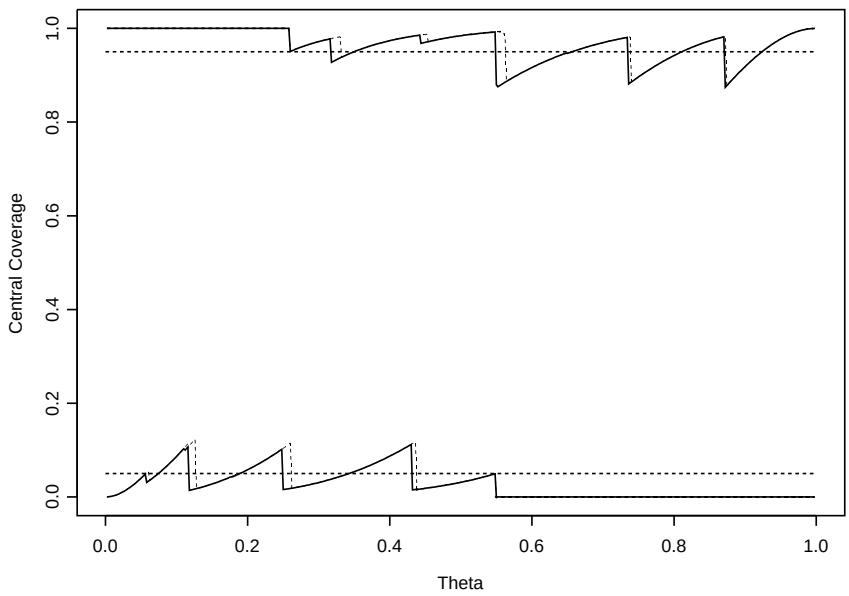

FIG. 4.20 – *Taux de recouvrement des intervalles de Jeffreys modifiés 90% classique (- - -) et corrigé (—) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

par la règle d'arrêt (i.e. $(m, y) = (1, 1)$ et $(m, y) = (2, y_2)$ $\forall y_2 \in \mathcal{Y}_{(2)}$).

Pour les valeurs de θ comprises entre les sauts de $C_J^{sup}(\theta; n_1, r; S_{1,U})$ et de $C_{JC}^{sup}(\theta; n_1, r; S_{1,U})$, la correction induit une diminution du taux de recouvrement de la borne supérieure. Cependant, l'augmentation de l'amplitude des sauts de $C_{JC}^{sup}(\theta; n_1, r; S_{1,U})$ accentue encore la proéminence du premier pic associé à $\hat{\theta}_J^{inf}(1,1)$. Cet effet délétère peut être effacé par l'utilisation des intervalles de Jeffreys modifiés classique et corrigé dont les taux de recouvrement unilatéraux sont montrés dans la figure (4.20).

A contrario, pour les valeurs de θ comprises entre les sauts de $C_J^{inf}(\theta; n_1, r; S_{1,U})$ et de $C_{JC}^{inf}(\theta; n_1, r; S_{1,U})$, la correction induit une augmentation du taux de recouvrement de la borne inférieure, qui compense notamment la diminution induite par la règle d'arrêt pour les faible valeurs de θ. La correction recentre les taux de recouvrement des deux bornes de l'intervalle de Jeffreys autour du niveau nominal unilatéral $1 - \alpha/2$.

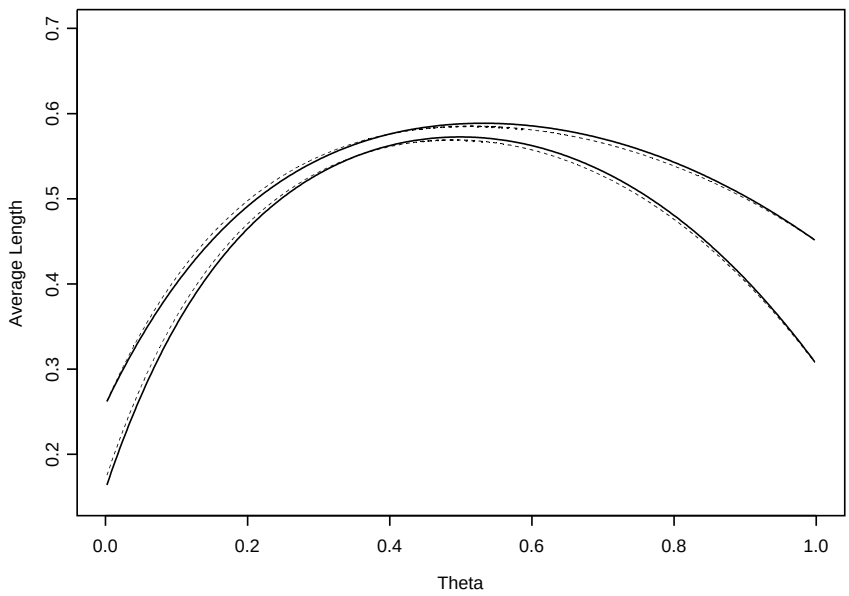

FIG. 4.21 – *Longueur moyenne des intervalles de Jeffreys et Jeffreys modifiés 90% classique (- - -) et corrigé (—) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

Les longueurs moyennes des intervalles de Jeffreys et Jeffreys modifiés 90%, classiques et corrigés, dans le plan $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$ sont montrées dans la figure (4.21). Relativement à l'échantillonnage fixe (voir figure (4.8)), la règle d'arrêt modifie les courbes associées à CI_J et CI_{JM} pour les valeurs $\theta < 0.5$. La longueur moyenne des intervalles diminue sensiblement lorsque θ se rapproche de zéro car la taille des échantillons est plus souvent associée à $n_1 + n_2$ que n_1.

La longueur moyenne de l'intervalle de Jeffreys corrigé est plus petite que celle de l'intervalle de Jeffreys classique lorsque $\theta < 0.4$, et supérieure sinon. Cette même tendance est observée pour l'intervalle de Jeffreys modifié, excepté aux extrémités où les modifications simultanées apportées aux bornes de CI_{JC} et CI_{JMC} font se rencontrer leurs courbes.

4.5 Performance des méthodes usuelles dans le cas d'un arrêt conditionné à l'observation d'au moins un événement

Dans le prolongement de la section précédente, les bornes des intervalles 90% de Wilson, d'Agresti-Coull et de l'intervalle corrigé au second-ordre dans le plan ($n_1 = 5, S_{1,U} = 1, n_2 = 5$) sont présentées dans la figure (4.22) en fonction des valeurs y/n, où $n = 5, 10$ et $\mathcal{Y} = \mathcal{Y}_{(1)} \cup \mathcal{Y}_{(2)}$.

Les trois intervalles se différencient notablement aux extrémités (i.e. y/n égal à 0 ou 1) ; l'intervalle d'Agresti-Coull et l'intervalle corrigé au second-ordre présentent des bornes inférieures négatives lorsque $(m, y) = (2, 0)$ et supérieures à 1 lorsque $(m, y) = (1, 5)$.

L'intervalle de Jennison-Turnbull reprend le principe de détermination des bornes de l'intervalle de Clopper-Pearson dans sa justification fréquentiste. Au lieu d'inverser la zone d'acceptation du test de la Binomiale, on inverse la zone d'acceptation d'un test défini par des valeurs de rejet pour (m, y) qui suivent l'ordre stochastique mentionné en section 4.2. Dans le plan ($n_1 = 5, S_{1,U} = 1, n_2 = 5$), les observations (m, y) sont classées suivant l'ordre $(2, 0) < (2, 1) < (2, 2) < (2, 3) < (2, 4) < (2, 5) < (1, 1) < (1, 2) < (1, 3) < (1, 4) < (1, 5)$.

Dans la figure (4.23), les bornes de l'intervalle de Jennison-Turnbull correspondent à celles de l'intervalle de Clopper-Pearson pour les observations $(m, y) = (1, y)$ avec $y \in \mathcal{Y}_{(1)}$ et se différencient pour les observations $(m, y) = (2, y)$ avec $y \in \mathcal{Y}_{(2)}$. La faiblesse des probabilités d'échantillonnage associées aux observations $(2, Y)$ se traduit par des valeurs de $\hat{\theta}^{sup}_{JT}(2, y)$ et $\hat{\theta}^{inf}_{JT}(2, y)$ respectivement très inférieures à $\hat{\theta}^{sup}_{CP}(2, y)$ et $\hat{\theta}^{inf}_{CP}(2, y)$. On remarque que l'intervalle $CI_{JT}(2, Y)$ ne contient pas $\hat{\theta}^{VUMSB}_{(2,y)} = 0$ lorsque $y = 1, .., 5$, alors que l'ordre suivi pour les observations (m, y) correspond à celui de l'amplitude des estimations $\hat{\theta}^{VUMSB}_{(m,y)}$. L'intervalle $CI_{JT}(2, 5)$ ne contient pas non plus $\hat{\theta}^{MV}_{(2,5)} = 0.5$, ce qui rend son interprétation difficile.

	$\hat{\theta}_W^{inf}(m,y)$	$\hat{\theta}_W^{sup}(m,y)$	$\hat{\theta}_{AC}^{inf}(m,y)$	$\hat{\theta}_{AC}^{sup}(m,y)$	$\hat{\theta}_2^{inf}(m,y)$	$\hat{\theta}_2^{sup}(m,y)$
(1,1)	0.0460	0.5647	0.0324	0.5782	0.0183	0.5613
(1,2)	0.1427	0.7275	0.1413	0.7289	0.1268	0.7331
(1,3)	0.2725	0.8573	0.2711	0.8587	0.2669	0.8732
(1,4)	0.4353	0.9540	0.4217	0.9676	0.4387	0.9816
(1,5)	0.6488	1.0000	0.5990	1.0499	0.6571	1.0434
(2,0)	0.0000	0.2129	-0.0358	0.2488	-0.0085	0.1846
(2,1)	0.0226	0.3477	0.0059	0.3644	0.0073	0.3336
(2,2)	0.0686	0.4592	0.0605	0.4673	0.0526	0.4530
(2,3)	0.1268	0.5583	0.1236	0.5616	0.1123	0.5581
(2,4)	0.1942	0.6484	0.1934	0.6491	0.1821	0.6531
(2,5)	0.2692	0.7307	0.2692	0.7307	0.2605	0.7395

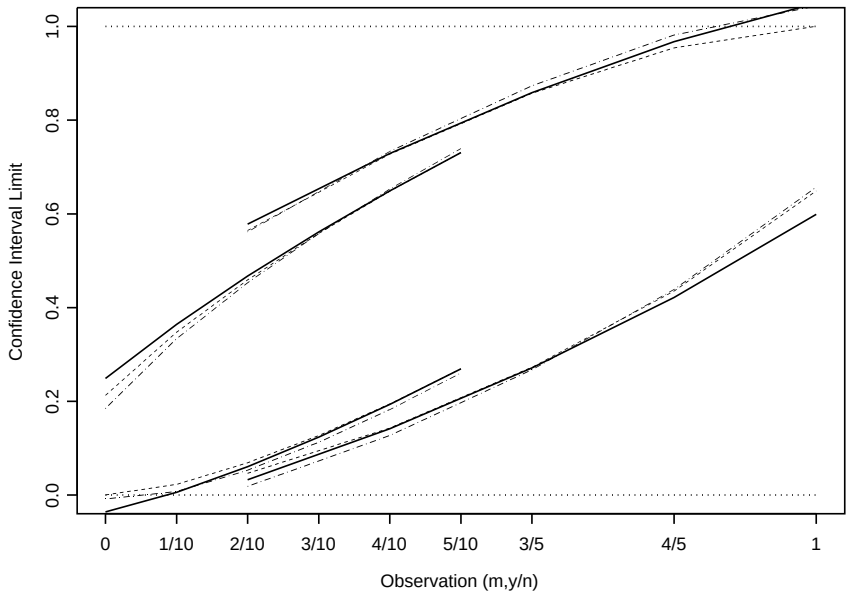

FIG. 4.22 – *Bornes des intervalles 90% de Wilson (- - -), d'Agresti-Coull (—) et de l'intervalle corrigé au second-ordre* $(\cdot - \cdot)$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

	$\hat{\theta}_{CP}^{inf}(m,y)$	$\hat{\theta}_{CP}^{sup}(m,y)$	$\hat{\theta}_{JT}^{inf}(m,y)$	$\hat{\theta}_{JT}^{sup}(m,y)$
(1,1)	0.0102	0.6574	0.0102	0.6574
(1,2)	0.0764	0.8107	0.0764	0.8107
(1,3)	0.1893	0.9236	0.1893	0.9236
(1,4)	0.3426	0.9898	0.3426	0.9898
(1,5)	0.5493	1.0000	0.5493	1.0000
(2,0)	0.0000	0.2589	0.0000	0.2589
(2,1)	0.0051	0.3942	0.0051	0.3495
(2,2)	0.0368	0.5069	0.0100	0.4055
(2,3)	0.0872	0.6066	0.0102	0.4367
(2,4)	0.1500	0.6965	0.0102	0.4487
(2,5)	0.2224	0.7776	0.0102	0.4507

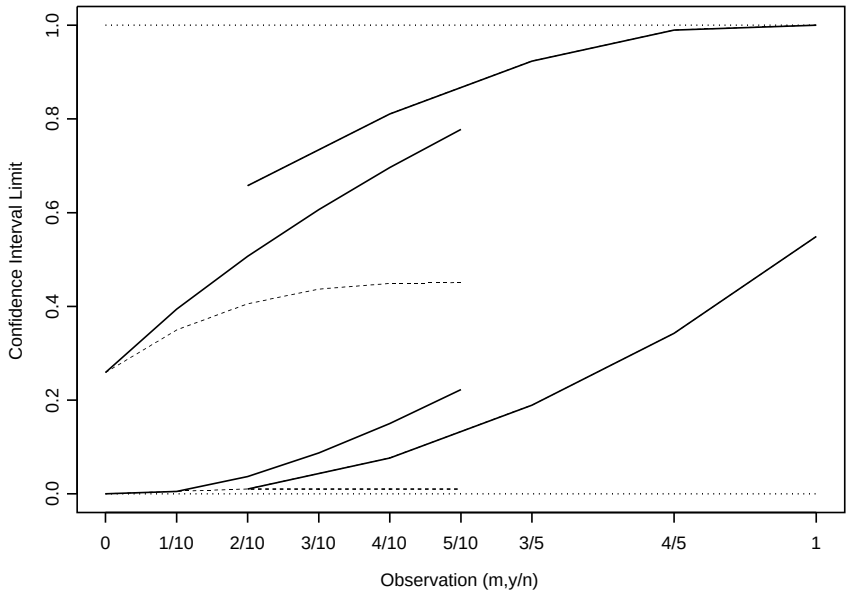

FIG. 4.23 – *Bornes des intervalles 90% de Clopper-Pearson (—) et Jennison-Turnbull (· − ·) dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 1, n_2 = 5$)

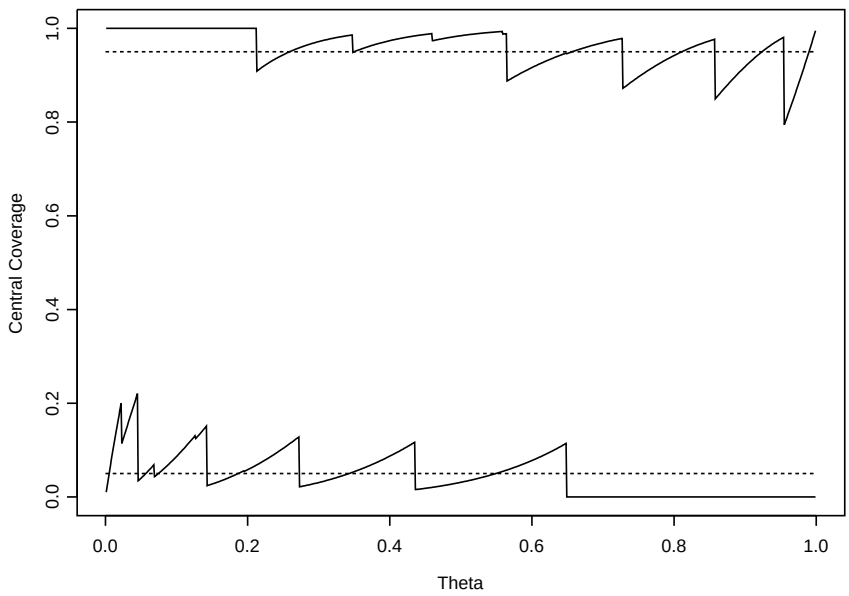

FIG. 4.24 – *Taux de recouvrement de l'intervalle de Wilson 90% dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

Les figures (4.24) et (4.25) révèlent un impact plus fort de la règle d'arrêt sur les intervalles 90% de Wilson et Agresti-Coull par rapport aux intervalles de Jeffreys classique et corrigé. Les taux de recouvrement $C_W^{sup}(\theta; n_1, r; S_{1,U})$ et $C_{AC}^{sup}(\theta; n_1, r; S_{1,U})$ sont en moyenne supérieurs au niveau nominal pour respectivement les valeurs de $\theta < \hat{\theta}_W^{sup}(1,1) = 0.565$ et $\theta < \hat{\theta}_{AC}^{sup}(1,1) = 0.578$, et les taux de recouvrement $C_W^{inf}(\theta; n_1, r; S_{1,U})$ et $C_{AC}^{inf}(\theta; n_1, r; S_{1,U})$ sont en moyenne inférieurs au niveau nominal pour respectivement les valeurs de $\theta < \hat{\theta}_W^{inf}(1,2) = 0.143$ et $\theta < \hat{\theta}_{AC}^{inf}(1,2) = 0.141$.

Les taux de recouvrement associés à l'intervalle corrigé au second-ordre (figure (4.26)) montrent un profile moins perturbé par la règle d'arrêt. Les courbes sont proches de celles correspondant à l'intervalle de Jeffreys corrigé.

La figure (4.27) montre que le taux de recouvrement de la borne inférieure de l'intervalle de Clopper-Pearson dépasse la valeur du niveau nominal unilatéral $1 - \alpha/2$ dans un plan séquentiel du type $(n_1, S_{1,U}, n_2)$. Le mécanisme de correction offert par

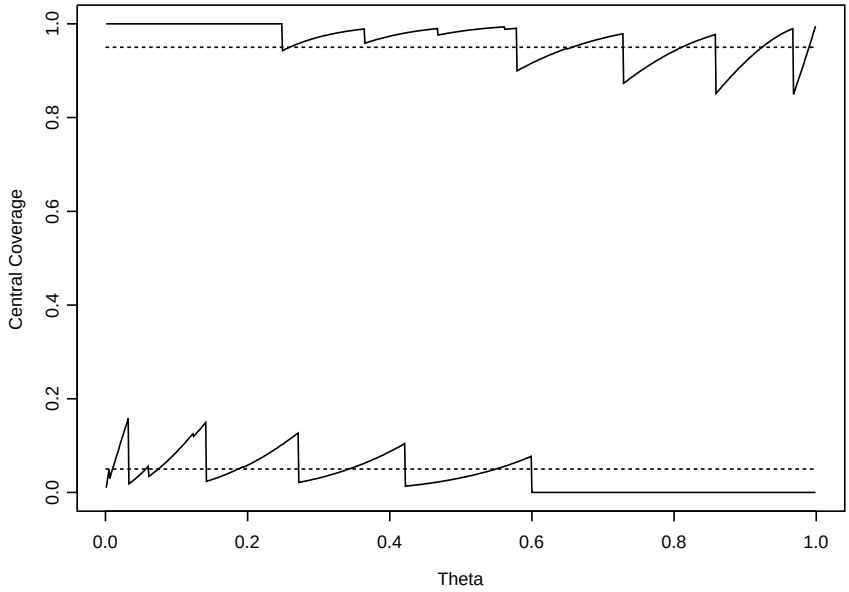

FIG. 4.25 – *Taux de recouvrement de l'intervalle d'Agresti 90% dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

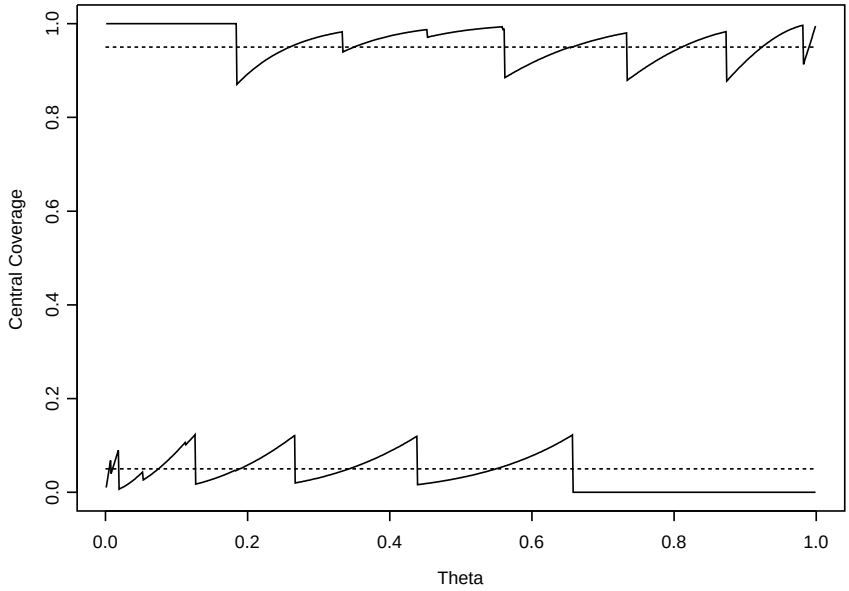

FIG. 4.26 – *Taux de recouvrement de l'intervalle du second-ordre 90% dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

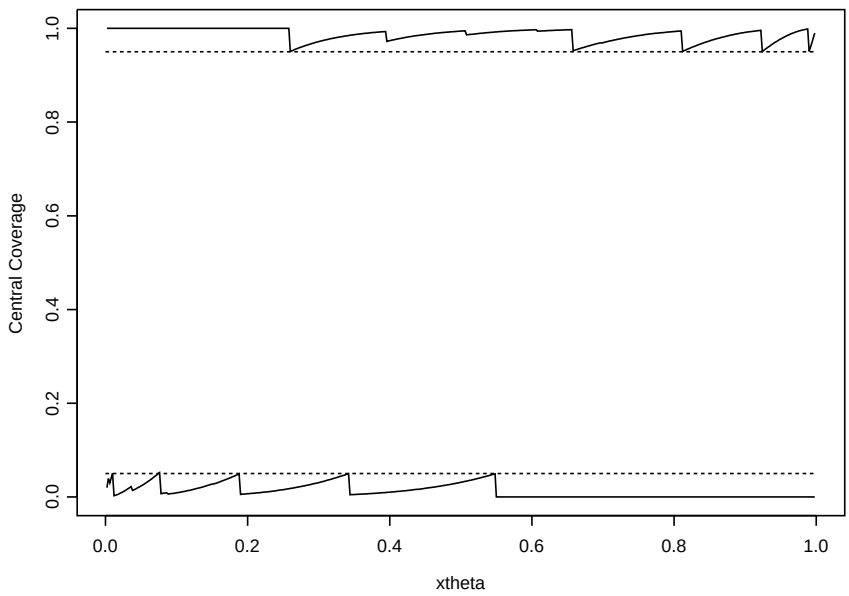

FIG. 4.27 – *Taux de recouvrement de l'intervalle de Clopper-Pearson 90% dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

la loi Bêta-J permet de compenser cet excès ; il suffit alors de remplacer la borne inférieure $\hat{\theta}^{inf}_{CP}(m, y) = Be(\alpha/2; y, n - y + 1)$ par la borne inférieure de l'intervalle de Clopper-Pearson corrigé $\hat{\theta}^{inf}_{CPC}(m, y) = Be^J(\alpha/2; y, n - y + 1, c; n_1, r; \{0\})$, en déterminant la valeur de c telle que $max_\theta(C^{inf}_{CPC}) = 0.05$.

Dans le plan séquentiel $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$, on a $max(C^{inf}_{CP}) = 0.0526$ qui est atteint au deuxième pic pour une valeur $\theta_{max} = \hat{\theta}^{inf}_{CP}(1{,}2) = 0.0764$. Une approche numérique itérative permet d'obtenir c tel que

$$max_\theta(C^{inf}_{CPC}) = max_\theta(E_\theta[Be^J(\theta; y, n - y + 1, c; n_1, r; \{0\})]) = 0.05.$$

On obtient simultanément les valeurs $c = 0.12$ et $\theta_{max} = \hat{\theta}^{inf}_{CPC}(1{,}2) = 0.0744$. Les valeurs de $C^{inf}_{CPC}(\theta; n_1, r; S_{1,U})$ sont montrées dans la figure (4.28).

L'ordre stochastique utilisé pour les observations (m, y) dans la construction de l'intervalle de Jennison-Turnbull engendre des taux de recouvrement unilatéraux conservateurs uniformément supérieurs au niveau nominal $1 - \alpha/2$. Relativement à l'inter-

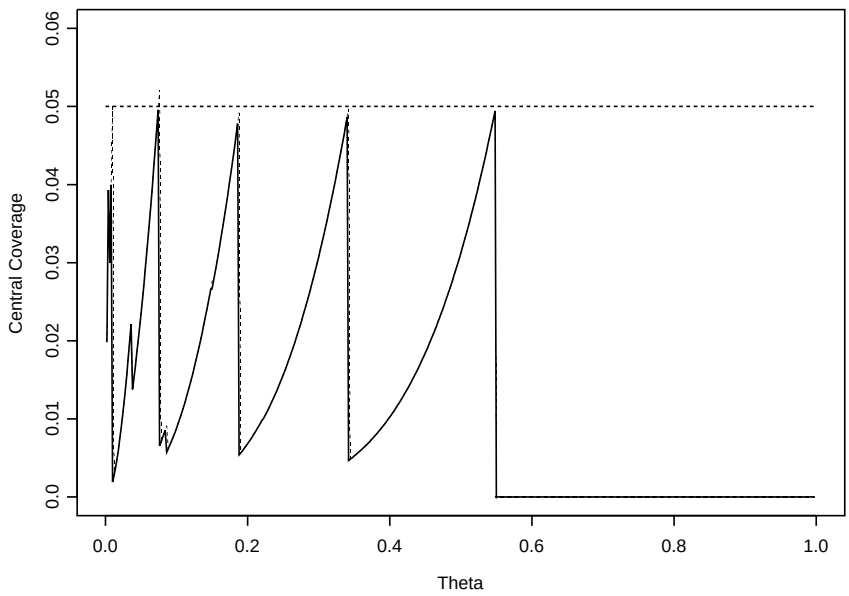

FIG. 4.28 – *Taux de recouvrement de la borne inférieure des intervalles de Clopper-Pearson 90% classique (- - -) et corrigé (—) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

valle de Clopper-Pearson, on observe également que les taux de recouvrement associés à CI_{JT} sont plus proches du niveau nominal, notamment pour $C_{JT}^{sup}(\theta; n_1, r; S_{1,U})$ autour des sauts associés à $\hat{\theta}_{JT}^{sup}(2, y)$ $y \in \mathcal{Y}_{(2)}$.

Les longueurs moyennes des intervalles de confiance usuels sont montrées dans la figure (4.30). Relativement à l'échantillonnage fixe (voir figure (4.8)), la règle d'arrêt modifie les courbes pour les valeurs $\theta < 0.5$. A l'approche de 0, la longueur moyenne des intervalles diminue sensiblement en même temps que l'on observe un resserrement des différentes courbes ; on note également que la longueur moyenne de l'intervalle d'Agresti-Coull devient la plus grande.

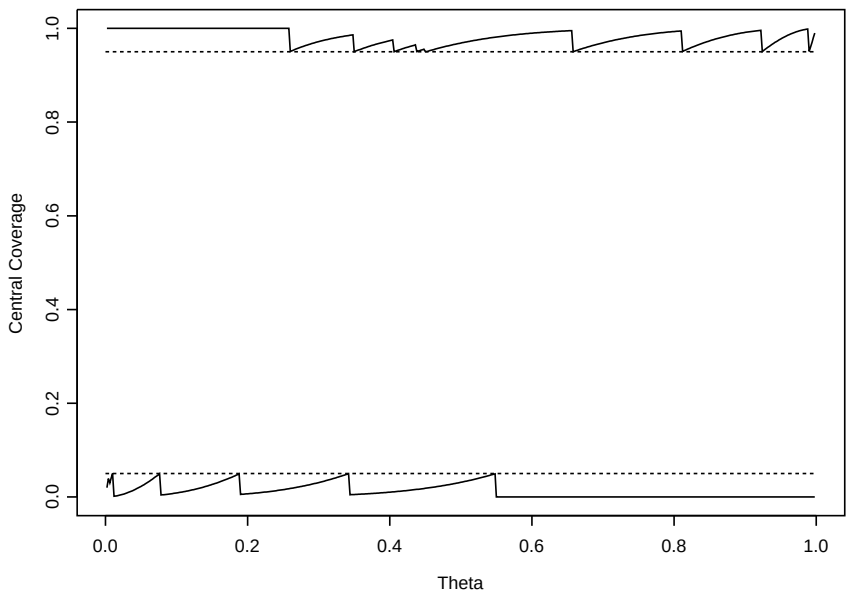

FIG. 4.29 – *Taux de recouvrement de l'intervalle de Jennison-Turnbull 90% dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$

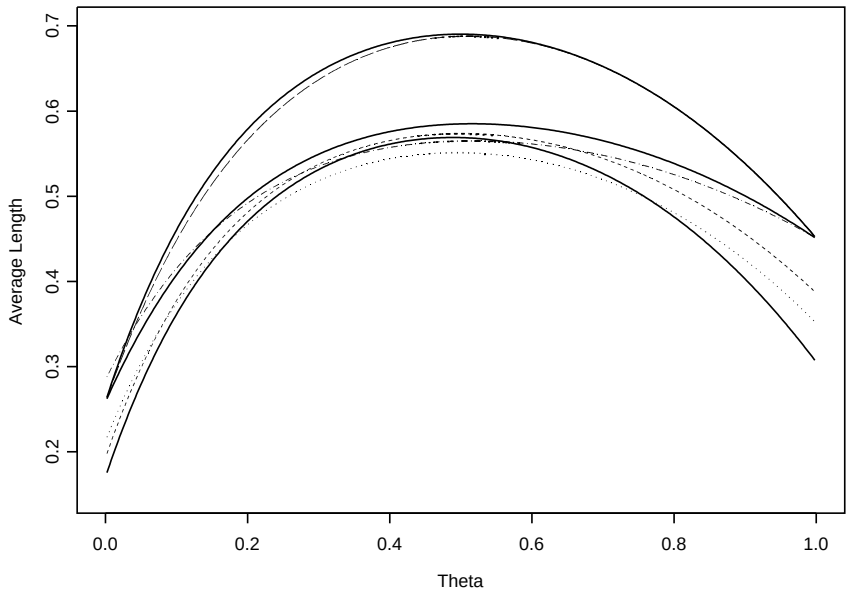

FIG. 4.30 – *Longueur moyenne des intervalles de confiance à 90% dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$. *Les intervalles de* CI_J, CI_{JM} *et* CI_{CP} *(—) sont respectivement représentés de bas en haut, avec également les intervalles* CI_{JT} *(– –),* CI_W *(· · ·),* CI_{AC} *(· – ·) et* CI_2 *(- - -).*

4.6 Généralisation à d'autres plans expérimentaux

L'objectif de cette section est d'étudier les propriétés de l'intervalle de Jeffreys corrigé dans le contexte d'une généralisation à d'autres plans avec un arrêt optionnel.

On fait d'abord varier l'effectif n_1 observé en période 1 tout en conservant la valeur $S_{1,U}/n_1 = 0.2$ et le rapport $r = n_2/n_1 = 1$. Les figures (4.31) et (4.32) montrent l'impact de la correction sur les taux de recouvrement unilatéraux de CI_J dans les plans $(n_1 = 10, S_{1,U} = 2, n_2 = 10)$ et $(n_1 = 20, S_{1,U} = 4, n_2 = 20)$. On observe que l'impact de la règle d'arrêt affecte les mêmes valeurs de θ que dans le plan $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$, et que la correction agit en compensation du biais induit par la règle d'arrêt précisément sur ces valeurs de θ les plus affectés.

On réévalue ensuite les performances des intervalles de confiance dans le plan $(n_1 = 5, S_{1,U} = 2, n_2 = 5)$. Les bornes des intervalles de Jeffreys corrigé et Jennison-Turnbull sont dépendantes de la nature du plan expérimental. Elles sont recalculées dans la figure (4.33). Pour l'intervalle de Jennison-Turnbull, l'ordre stochastique utilisé pour (m, y) est défini par $(2,0) < (2,1) < (2,2) < (2,3) < (2,4) < (2,5) < (2,6) < (1,2) < (1,3) < (1,4) < (1,5)$.

Les bornes des autres intervalles de confiance sont identiques à celles obtenues dans le plan $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$, auxquelles il convient d'ajouter les bornes associées à l'observation $(m, y) = (2, 6) \in \mathcal{Y}_{(2)}$ et de retirer les bornes associées à l'observation $(m, y) = (1, 1) \notin \mathcal{Y}_{(1)}$.

On observe que l'interprétation des bornes de l'intervalle de Jennison-Turnbull est moins incohérente que dans le plan $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$. La faible variation de la borne inférieure $\hat{\theta}^{inf}_{JT}(2, y)$ lorsque $y \geq 3$ est due aux petites probabilités d'échantillonnage associées à ces observations.

Les taux de recouvrement des intervalles de Jeffreys 90% classique et corrigé dans

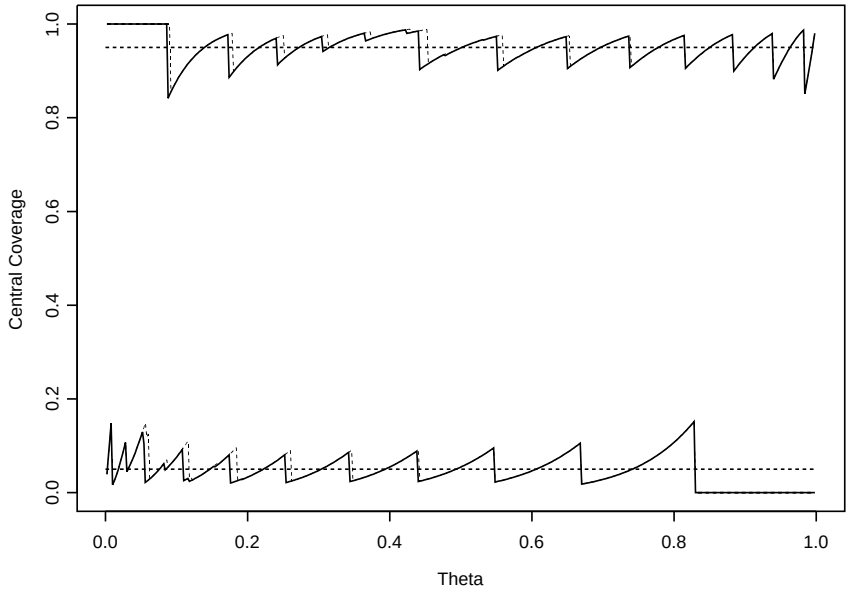

FIG. 4.31 – *Taux de recouvrement des intervalles de Jeffreys 90% classique (- - -) et corrigé (—) dans le plan séquentiel* $(n_1 = 10, S_{1,U} = 2, n_2 = 10)$

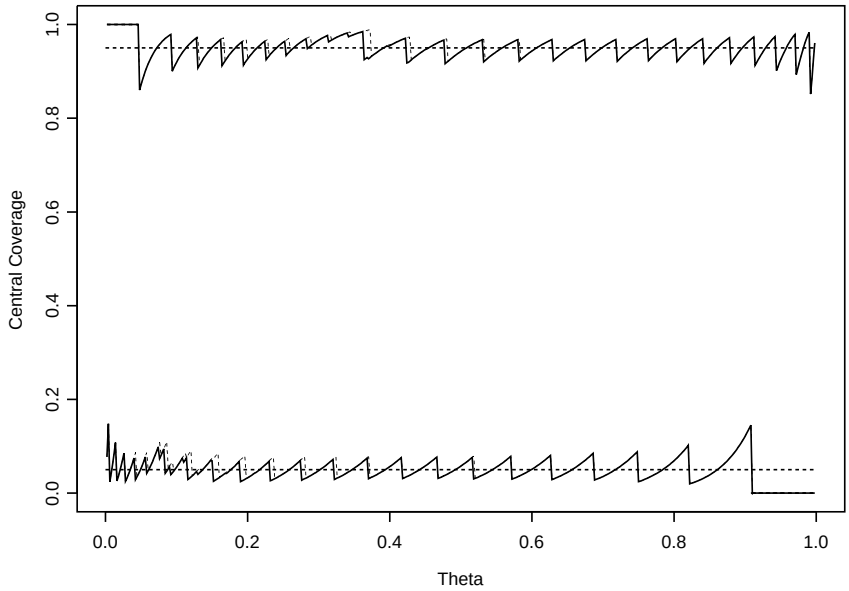

FIG. 4.32 – *Taux de recouvrement des intervalles de Jeffreys 90% classique (- - -) et corrigé (—) dans le plan séquentiel* $(n_1 = 20, S_{1,U} = 4, n_2 = 20)$

	$\hat{\theta}_J^{inf}(m,y)$	$\hat{\theta}_J^{sup}(m,y)$	$\hat{\theta}_{JC}^{inf}(m,y)$	$\hat{\theta}_{JC}^{sup}(m,y)$	$\hat{\theta}_{JT}^{inf}(m,y)$	$\hat{\theta}_{JT}^{sup}(m,y)$
(1,2)	0.1278	0.7394	0.1175	0.7341	0.0764	0.8107
(1,3)	0.2606	0.8722	0.2500	0.8710	0.1893	0.9236
(1,4)	0.4372	0.9636	0.4315	0.9634	0.3426	0.9898
(1,5)	0.6942	0.9996	0.6933	0.9996	0.5493	1.0000
(2,0)	0.0002	0.1708	0.0002	0.1593	0.0000	0.2589
(2,1)	0.0179	0.3306	0.0164	0.3176	0.0051	0.3942
(2,2)	0.0602	0.4525	0.0560	0.4425	0.0368	0.4917
(2,3)	0.1173	0.5581	0.1110	0.5517	0.0663	0.5644
(2,4)	0.1851	0.6527	0.1780	0.6491	0.0755	0.6162
(2,5)	0.2619	0.7381	0.2553	0.7363	0.0764	0.6470
(2,5)	0.2619	0.7381	0.2553	0.7363	0.0764	0.6574

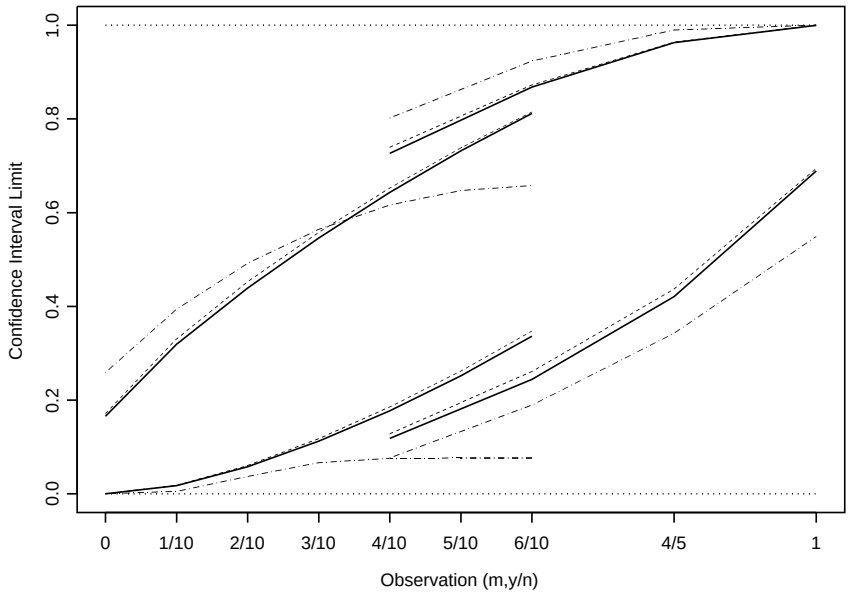

FIG. 4.33 – *Bornes des intervalles de Jeffreys 90% classique (- - -) et corrigé (—), et Jennison-Turnbull (· — ·) dans le plan séquentiel* $(n_1=5, S_{1,U}=2, n_2=5)$

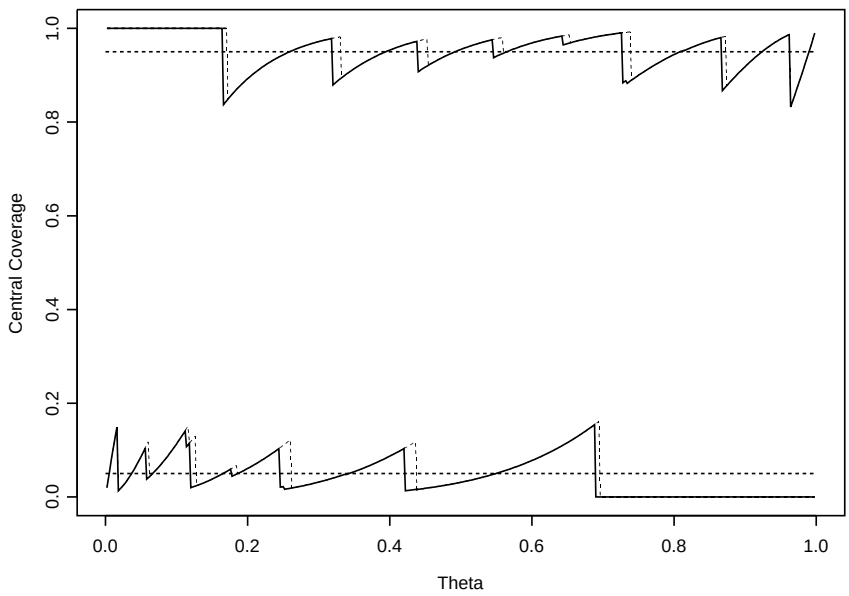

FIG. 4.34 – *Taux de recouvrement des intervalles de Jeffreys 90% classique (- - -) et corrigé (—) dans le plan séquentiel* $(n_1=5, S_{1,U}=2, n_2=5)$

la figure (4.34) montrent que l'impact de la règle d'arrêt et de la correction de l'a priori de Jeffreys s'étendent vers les valeurs supérieures de θ par rapport au plan $(n_1=5, S_{1,U}=1, n_2=5)$. L'intervalle de Jeffreys corrigé présente les meilleurs profiles de taux de recouvrement unilatéraux relativement à l'intervalle de Wilson (figure (C.24)), d'Agresti-Coull (figure (C.25)) et de l'intervalle corrigé au second-ordre (figure (C.26)), qui reste la meilleure alternative parmi ces trois derniers.

Le taux de recouvrement de la borne inférieure de l'intervalle de Clopper-Pearson dans la figure (C.27) laisse apparaître un pic $max_\theta(C_{CP}^{inf}) = 0.0665 > 0.05$ pour la valeur $\theta_{max} = \hat{\theta}_{CP}^{inf}(1,2) = 0.0764$. Suivant l'approche itérative mise en oeuvre dans le plan $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$, on obtient cette fois la valeur $c = 0.94$ dans la loi a priori $Be^J(0, 1, c; n_1, r; J)$ avec $J = \{0, 1\}$ qui garantit $max_\theta(C_{CPC}^{inf}) = 0.05$. La valeur correspondante en θ est $\theta_{max} = \hat{\theta}_{CPC}^{inf}(1,2) = 0.0664$ (figure (4.35)). Enfin, l'intervalle de Jennison-Turnbull (figure (4.36)) confirme ses propriétés en terme de taux de recouvrement unilatéral proche des niveaux nominaux.

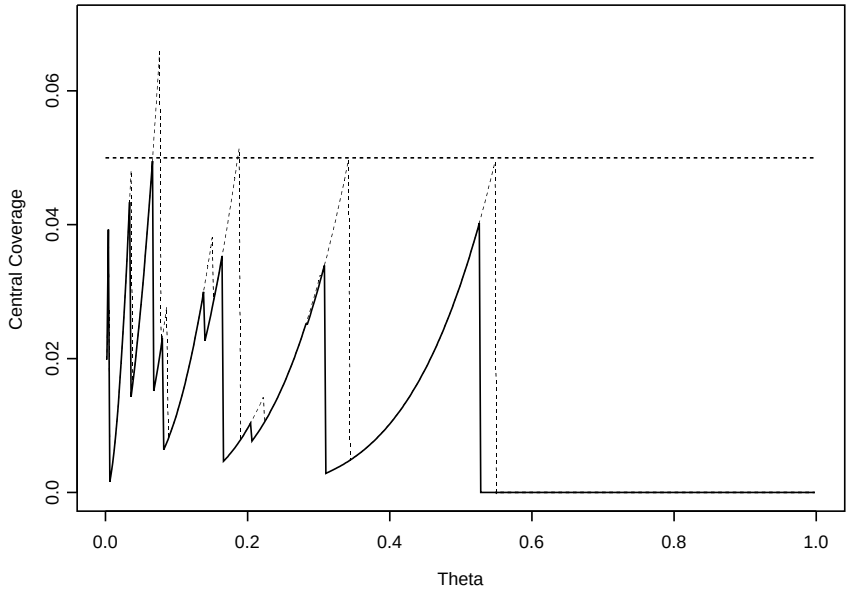

FIG. 4.35 – *Taux de recouvrement de la borne inférieure de l'intervalle de Clopper-Pearson 90% dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 2, n_2 = 5)$

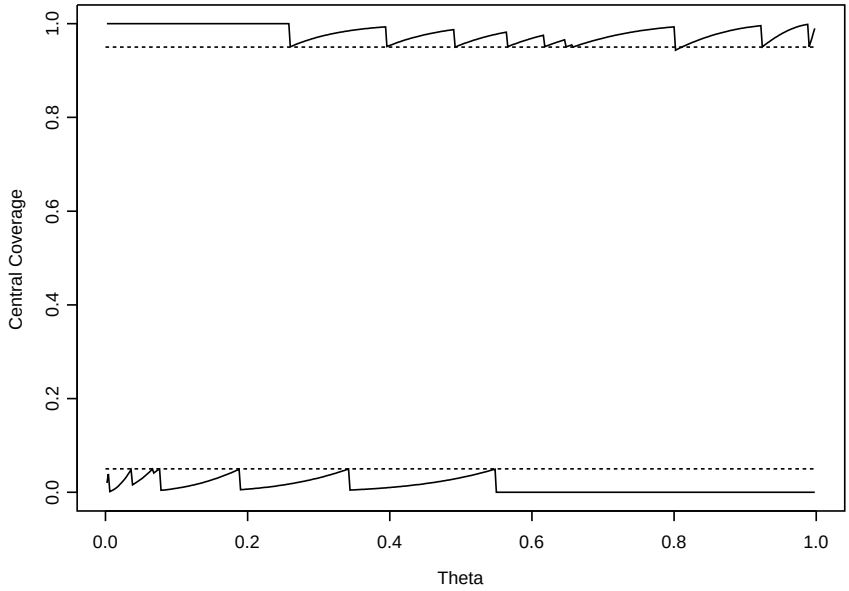

FIG. 4.36 – *Taux de recouvrement de la borne inférieure de l'intervalle de Jennison-Turnbull 90% dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 2, n_2 = 5)$

Les figures (C.28) et (C.29) montrent que les intervalles de Jeffreys classique et corrigé offrent le meilleur profile de courbe de longueur moyenne parmi tous les intervalles.

La dernière étude permet de comparer les performances des intervalles de confiance dans le plan $(n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5)$. Le bornage bilatéral de Y_1 en période 1 définit une zone de continuation du type $J = [0, S_{1,U}[\cup]S_{1,L}, n_1]$ (i.e. l'expérience s'arrête en période 1 si $Y_1 \in \bar{J} = [S_{1,U}, S_{1,L}]$). Dans le plan $(n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5)$, on a $J = \{0, 5\}$.

Les figures (C.30) à (C.44)) montrent les lois a posteriori $\theta|d_{seq}, y \sim Be^J(\frac{1}{2}+y, \frac{1}{2}+n-y, \frac{1}{2}; n_1, r; J)$ pour $y \in \mathcal{Y}$ relativement à $\theta|y \sim Be(\frac{1}{2}+y, \frac{1}{2}+n-y)$ obtenues avec l'a priori de Jeffreys classique. Les courbes des densités corrigées $p(\theta|d_{seq}, y)$ sont tirées à leurs bases à la fois vers les valeurs inférieures et supérieures de θ. Elles sont plus larges lorsqu'elles sont associées à des observations centrales comme $(m, y) = (1, 2)$ dans $y \in \mathcal{Y}_{(1)}$ (figure (C.31)) et $(m, y) = (2, 5)$ dans $y \in \mathcal{Y}_{(2)}$ (figure (C.39)), ou bien tirées vers l'extrémité la plus proche lorsqu'elles sont associées à des observations plus excentrées comme $(m, y) = (1, 1)$ (figure (C.30)) et $(m, y) = (2, 2)$ (figure (C.36)).

L'influence du plan séquentiel $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$ sur l'intervalle de Jeffreys classique ainsi que la compensation induite par la correction de l'a priori sur les taux de recouvrement unilatéraux sont montrés dans la figure (4.37). Pour les valeurs $\theta \leq \hat{\theta}_J^{sup}(1,1)$ et $\theta \leq \hat{\theta}_J^{inf}(1,2)$, les fonctions $C_J^{sup}(\theta; n_1, r; S_{1,U}, S_{1,L})$ et $C_J^{inf}(\theta; n_1, r; S_{1,U}, S_{1,L})$ respectivement sous-estiment et surestiment le niveau nominal $1-\alpha/2$. Sous l'effet de la correction, les sauts des fonctions $C_{JC}^{sup}(\theta; n_1, r; S_{1,U}, S_{1,L})$ et $C_{JC}^{inf}(\theta; n_1, r; S_{1,U}, S_{1,L})$ sont décalés vers 0, ce qui compense de fait le biais expérimental.

A contrario, pour les valeurs $\theta \geq \hat{\theta}_J^{sup}(1,3)$ et $\theta \geq \hat{\theta}_J^{inf}(1,4)$, les taux de recouvrement unilatéraux $C_J^{sup}(\theta; n_1, r; S_{1,U}, S_{1,L})$ et $C_J^{inf}(\theta; n_1, r; S_{1,U}, S_{1,L})$ respectivement

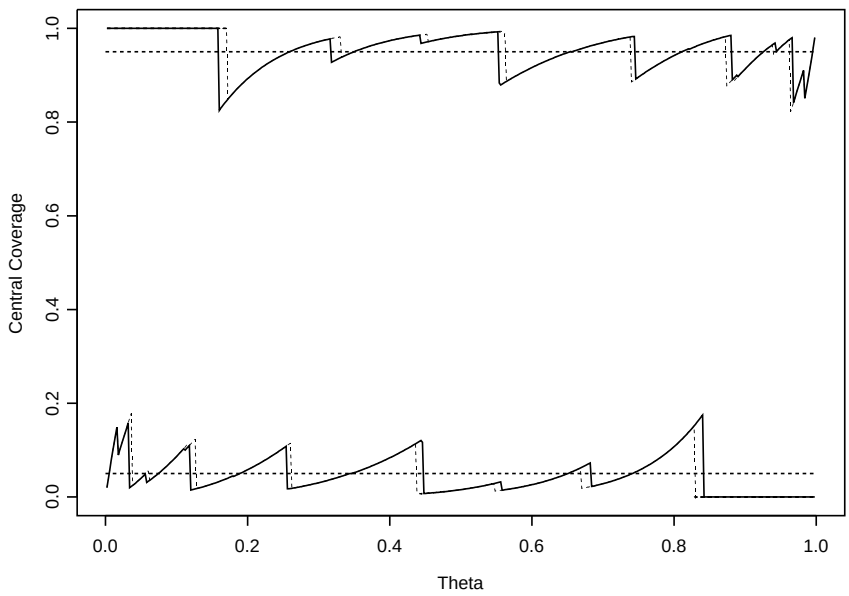

FIG. 4.37 – *Taux de recouvrement des intervalles de Jeffreys 90% classique (- - -) et corrigé (—) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5)$

surestiment et sous-estiment le niveau nominal $1-\alpha/2$. Les sauts des fonctions $C_{JC}^{sup}(\theta; n_1, r; S_{1,U}, S_{1,L})$ et $C_{JC}^{inf}(\theta; n_1, r; S_{1,U}, S_{1,L})$ sont décalés vers 1, compensant de la même manière l'impact du biais expérimental.

A rapports $S_{1,U}/n_1 = 0.2$ et $S_{1,L}/n_1 = 0.8$ constants, les figures (4.38) et (4.39) confirment l'effet de la correction de l'a priori de Jeffreys respectivement dans les plans séquentiels $(n_1 = 10, S_{1,U} = 2, S_{1,L} = 8, n_2 = 10)$ et $(n_1 = 20, S_{1,U} = 4, S_{1,L} = 16, n_2 = 20)$.

La figure (4.40) montre que l'intervalle de Wilson est plus sensible que les intervalles de Jeffreys classique et corrigé au biais induit par la règle d'arrêt sur les taux de recouvrement dans le plan $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$, comme l'est dans une moindre mesure l'intervalle d'Agresti-Coull dans la figure (4.41). Le profil des courbes de l'intervalle corrigé au second-ordre dans la figure (4.42) confirme les bonnes propriétés qui se rapproche de celles de l'intervalle de Jeffreys corrigé. Les

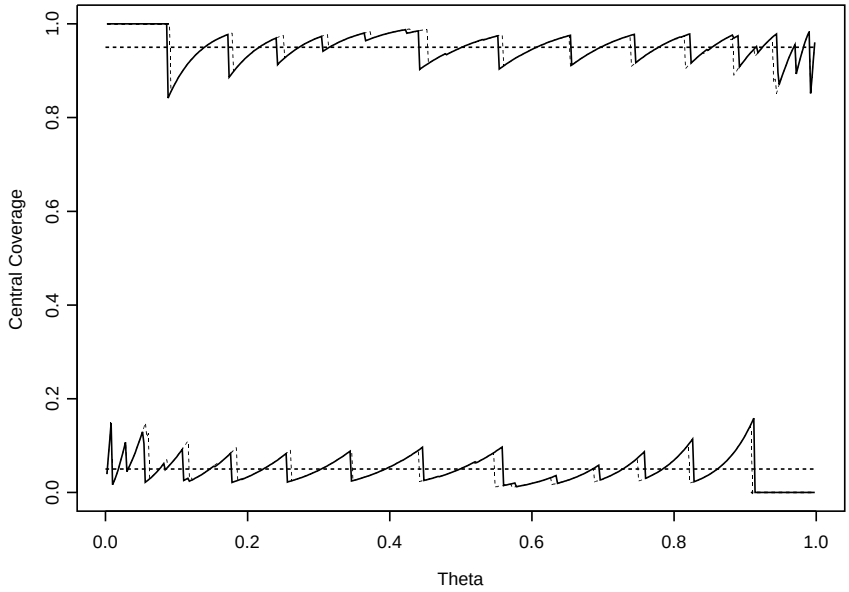

FIG. 4.38 – *Taux de recouvrement des intervalles 90% de Jeffreys classique (- - -) et corrigé (—) dans le plan séquentiel* ($n_1 = 10, S_{1,U} = 2, S_{1,L} = 8, n_2 = 10$)

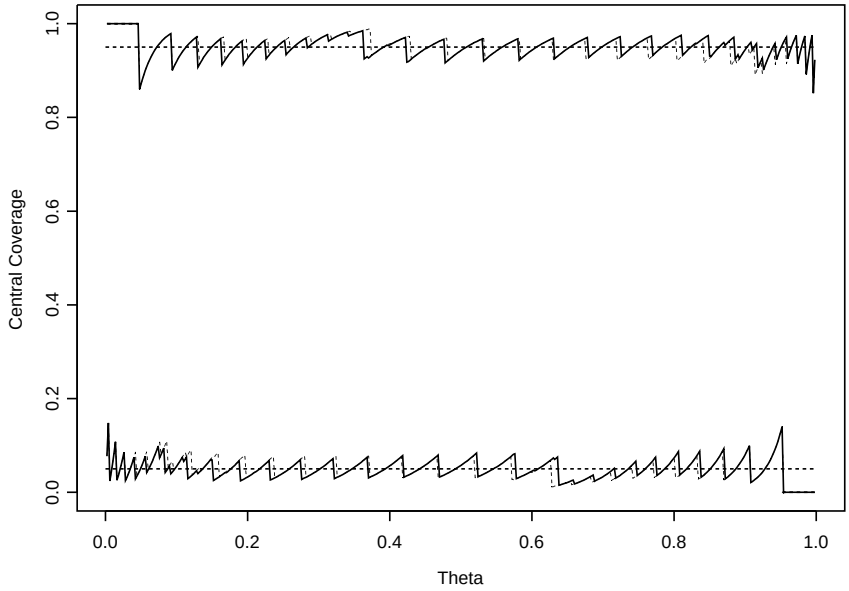

FIG. 4.39 – *Taux de recouvrement des intervalles 90% de Jeffreys classique (- - -) et corrigé (—) dans le plan séquentiel* ($n_1 = 20, S_{1,U} = 4, S_{1,L} = 16, n_2 = 20$)

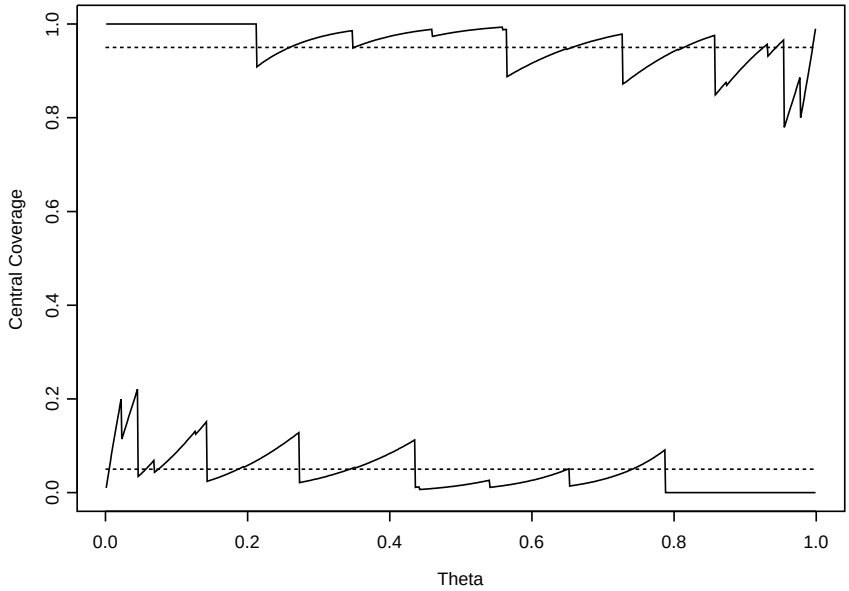

FIG. 4.40 – *Taux de recouvrement de l'intervalle de Wilson 90% dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5$)

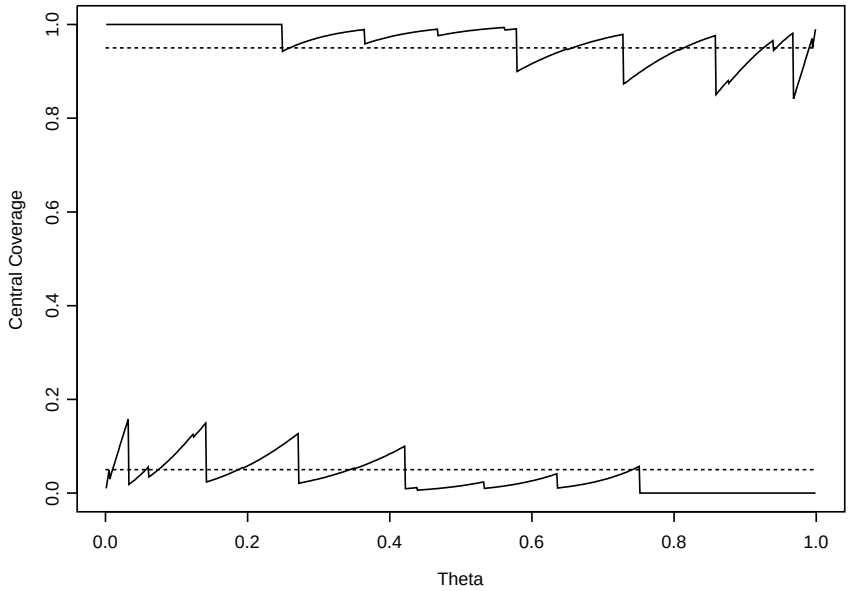

FIG. 4.41 – *Taux de recouvrement de l'intervalle d'Agresti-Coull 90% dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5$)

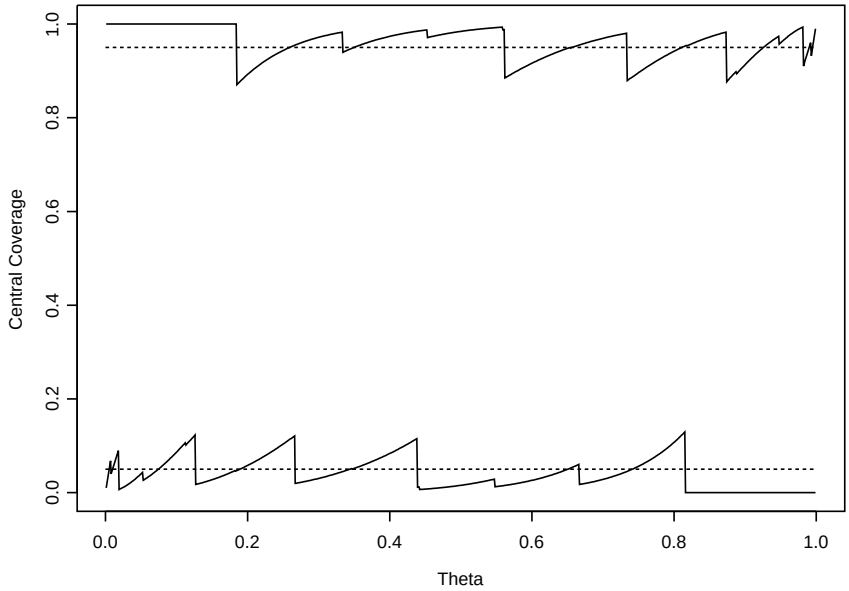

FIG. 4.42 – *Taux de recouvrement de l'intervalle du second-ordre 90% dans le plan séquentiel* ($n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5$)

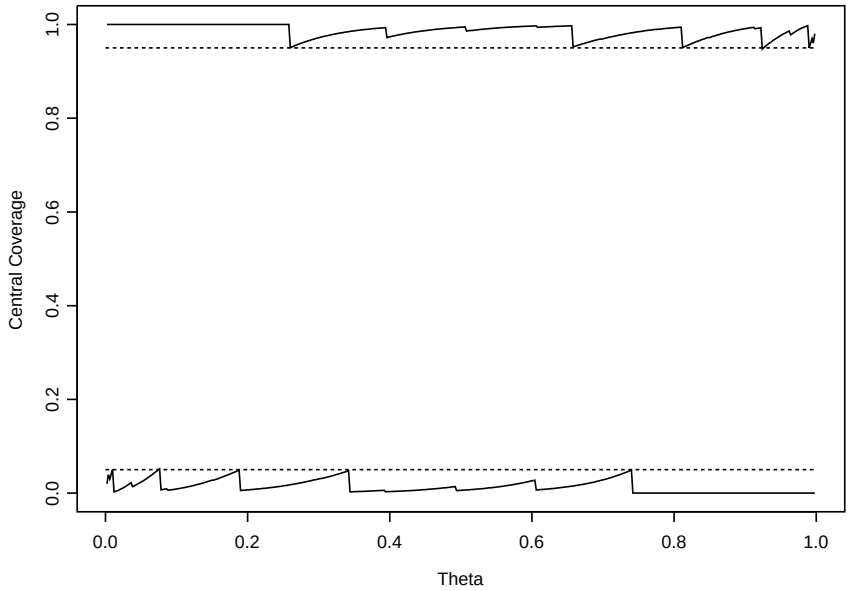

FIG. 4.43 – *Taux de recouvrement de l'intervalle de Clopper-Pearson 90% dans le plan séquentiel* ($n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5$)

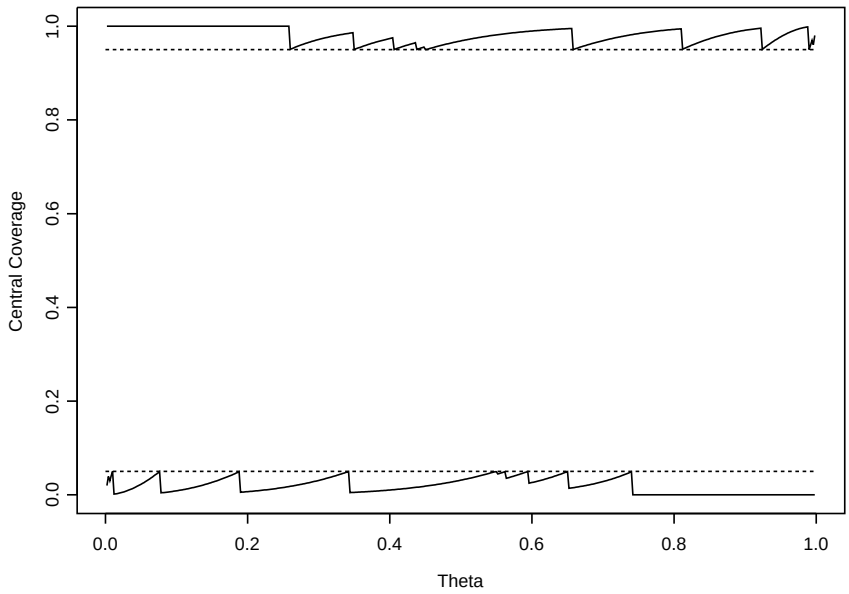

FIG. 4.44 – *Taux de recouvrement de l'intervalle de Jennison-Turnbull 90% dans le plan séquentiel* ($n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5$)

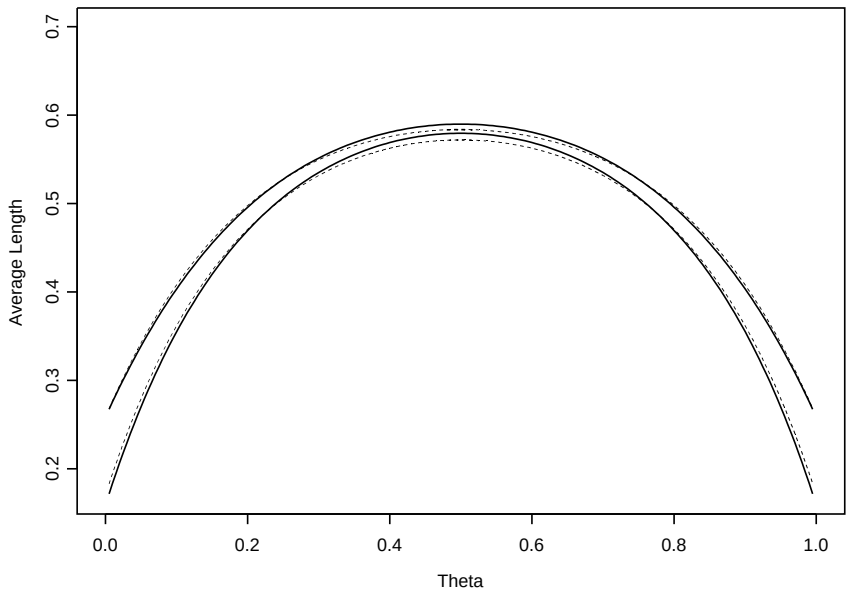

FIG. 4.45 – *Longueur moyenne des intervalles de Jeffreys et Jeffreys modifié 90% classique (- - -) et corrigé (—) dans le plan séquentiel* ($n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5$)

taux de recouvrement unilatéraux des intervalles de Clopper-Pearson et Jennison-Turnbull sont respectivement montrés dans les figures (4.43) et (4.44).

Pour l'intervalle de Jennison-Turnbull, certaines observations $y \in \mathcal{Y}_{(2)}$ sont associées à deux conditions $(Y_1 < S_{1,U})$ et $(Y_1 > S_{1,L})$ en période 1. Par exemple, l'observation $(m, y) = (2, 5)$ correspond à la fois à $y_1 = 0$ et $y_2 = 5$ et à $y_1 = 5$ et $y_2 = 0$. Suivant la règle établie dans la section 4.2, la notation (m, y) appliquée jusqu'à maintenant ne permet plus de définir l'ordre stochastique à utiliser car il est nécessaire de savoir si le prolongement de l'expérience en période 2 est dû à l'une ou l'autre condition sur Y_1. Une notation sous la forme du doublet (y_1, y_2), où $y_2 = .$ indique une valeur manquante, permet une telle différenciation. L'ordre stochastique dans le plan $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$ pour les séquences (y_1, y_2) est alors défini par $(0, 0) < (0, 1) < (0, 2) < (0, 3) < (0, 4) < (0, 5) < (1, .) < (2, .) < (3, .) < (4, .) < (5, 0) < (5, 1) < (5, 2) < (5, 3) < (5, 4) < (5, 5)$. Les courbes des taux de recouvrement de l'intervalle de Jennison-Turnbull dans le plan $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$ sont décrites dans la figure (4.44).

Les longueurs moyennes des intervalles usuels sont montrées dans la figure (C.45). L'allure des courbes confirment les conclusions portées dans les plans bornés unilatéralement. La figure (4.45) montre que la longueur moyenne de l'intervalle de Jeffreys corrigé est inférieure à celle de l'intervalle de Jeffreys classique sur les extrémités.

4.7 Conclusion

La règle de Jeffreys appliquée à la vraisemblance d'un plan séquentiel d_{seq} avec un arrêt optionnel conduit à l'a priori de Jeffreys corrigé $\theta | d_{seq} \sim Be^J(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}; n_1, r; J)$ où $r = n_2/n_1$. Relativement à l'a priori de Jeffreys classique $\theta \sim Be(\frac{1}{2}, \frac{1}{2})$ dans lequel la nature du plan expérimental n'est pas prise en considération, la correction vient compenser l'effet de la règle d'arrêt sur les bornes de l'intervalle de confiance.

L'étude menée dans ce chapitre sur plusieurs plans séquentiels met en évidence les bonnes performances de l'intervalle de Jeffreys corrigé en terme de taux de recouvrement et de longueur moyenne par rapport aux intervalles habituellement recommandés dans les plans à échantillonnage fixe (Wilson, Agresti-Coull, Clopper-Pearson, Jeffreys et Jeffreys modifié).

On montre en plus comment la correction associée à une loi a priori $Be^J(a, b, c; n_1, r; J)$ peut être utilisée dans le contexte de l'intervalle de Clopper-Pearson pour recouvrer la propriété de conservation du niveau nominal par les taux de recouvrement que la règle d'arrêt fait perdre. Cependant, cet interval fondé sur l'inversion de la zone d'acceptation de l'hypothèse nulle dans le test de la loi Binomiale n'a plus sa justification dans l'échantillonnage séquentiel.

On lui substitut l'intervalle de Jennison-Turnbull, construit suivant le même principe, qui est une réponse fréquentiste au problème de la règle d'arrêt. Cependant l'ajustement en fonction du plan expérimental pose une problématique conceptuelle et calculatoire liée à la détermination de l'ordre avec lequel les hypothétiques observations (m, y) infirment le contenu de H_0. Ce choix introduit une part de subjectivité et les bornes associées à cet intervalle sont dans certains cas d'une interprétation peu cohérente.

L'approche par séries de Edgeworth permettrait de mesurer analytiquement l'impact du biais lié à la règle d'arrêt sur les taux de recouvrement et la longueur moyenne, et de formaliser la compensation induite par l'utilisation de l'intervalle de Jeffreys corrigé.

Chapitre 5

Application au test d'hypothèses

L'objectif de ce chapitre est de proposer un nouveau test dans un plan expérimental comportant un arrêt optionnel. Ce test est fondé sur le facteur de Bayes en utilisant l'a priori de Jeffreys. Il reprend le principe de la méthode décrite dans la section (B.5.3) développée par Berger, Boukai et Wang (test BBW) [7] puis par Dass (test de Dass) [22] pour les données discrètes. L'interprétation simultanée dans le contexte des inférences bayésienne et fréquentiste conditionnelle permet d'exprimer directement les probabilités d'erreur liées aux hypothèses en terme de risques conditionnels.

Berger, Boukai et Wang [6] justifient l'inutilité d'un conditionnement du test BBW sur la règle d'arrêt par l'existence d'un autre conditionnement sur une statistique non-ancilaire mesurant la force de l'évidence en faveur des hypothèses. Le contexte de l'inférence fréquentiste non-conditionnelle permet cependant d'établir qu'un arrêt optionnel de l'expérience se traduit par un biais en faveur d'une des hypothèses. La solution proposée dans ce chapitre passe par l'utilisation de l'a priori de Jeffreys corrigé développé dans la section 2.3 pour construire les densités prédictives du facteur de Bayes.

Pour simplifier l'écriture, la notation (k, y) désigne les séquences hypothétiques et observées. La *va* M associée aux nombre d'analyses réalisées n'est plus utilisée.

5.1 Tests fréquentistes séquentiels fondés sur des fonctions de dépense des risques

La succession de tests au seuil de significativité égal au niveau nominal α sur des données qui s'accumulent conduit à une augmentation de l'erreur de type 1 globale au delà de α. Le concept de fonction de dépense du risque α fut introduit par Lan et Demets (1983) [41] sur des variables normalement distribuées. La borne supérieure $\{S_{1,U}, .., S_{K,U}\}$ est définie par une fonction de dépense croissante et prédéfinie $\alpha^* : [0,1] \to [0,\alpha]$ où les valeurs de $S_{k,U}$ sont déterminées par la relation

$$Pr(S_{k,U} \text{ est franchie à la } k^{ieme} \text{ analyse}|H_0) = \alpha^*(t_k).$$

La quantité t_k est la proportion d'information I_k/I_K à la k^{ieme} analyse relativement à I_K l'information maximum possible à l'issue de l'expérience (voir (A.4.1)). La méthode de résolution est fondée sur la propriété d'incrémentations indépendantes de la fonction de score S_k (i.e. la dérivée première de la log-vraisemblance jointe des k échantillonnages successifs) où pour des effectifs suffisants, on a

$$\begin{aligned} &S_k \sim N(\theta I_k, I_k) && k = 1, \ldots, K, \\ &Cov(S_k - S_{k-1}, S_{k'} - S_{k'-1}) = 0 && k \neq k'. \end{aligned}$$

La statistique S_k suit un mouvement Brownien paramétré par θ. La contrainte imposée $I_k = (k/K)I_K$ exprime une équidistance des analyses sur l'échelle I_k.

Pampallona, Tsiatis et Kim (1995) [46] utilise en plus une borne inférieure $\{S_{1,L}, .., S_{K,L}\}$ définie par une fonction de dépense croissante et prédéfinie $\beta^* : [0,1] \to [0,\alpha]$ où les valeurs de $S_{k,L}$ sont déterminées par la relation

$$Pr(S_{k,L} \text{ est franchie à la } k^{ieme} \text{ analyse}|H_0) = \beta^*(t_k).$$

Cette méthode garantit la puissance anticipée mais nécessite un réajustement final de I_K pour satisfaire l'égalité $S_{K,U} = S_{K,L}$. Stallard et Facey (1996) [49] évitent ce réajustement en ne garantissant plus la puissance mais redéfinissant la borne

inférieure par une autre fonction de dépense $\alpha_L^* : [0,1] \to [0, 1-\alpha]$ telle que

$$Pr(S_{k,L} \text{ est franchie à la } k^{ieme} \text{ analyse}|H_0) = \alpha_L^*(t_k).$$

La statistique standardisée

$$Z_k = \frac{\hat{\theta}_k^{MV}}{\sqrt{Var(\hat{\theta}_k^{MV})}} = \frac{S_k}{\sqrt{I_k}} \sim N(\frac{\theta}{\sqrt{I_k}}, 1)$$

peut être utilisée pour décider de continuer ou non l'expérience. Dans le test de l'hypothèse $H_0 : \{\theta = 0\}$, l'expérience est poursuivie en période $k+1$ si à la k^{ieme} analyse on obtient $|Z_k| > S_{k,U}$. La relation entre les termes $S_{k,U}$ $(k = 1, \ldots, K)$ dépend de la nature du plan expérimental. Wang et Tsiatis (1987) [52] formalisent une forme générale où $S_{k,U} = c(k/K)^{\Delta - 1/2}$. On retrouve le plan d'O'Brien-Fleming si $\Delta = 0$ et le plan de Pocock si $\Delta = 1/2$.

Stallard et Todd (2000) [50] reprennent l'approche asymptotique développée dans Stallard et Facey (1996) en dérivant les bornes inférieures et supérieures par un calcul de probabilité exacte. Dans l'échantillonnage Binomial, la relation $t_n = I_k/I_K = \sum_{i=1}^k n_i / \sum_{i=1}^K n_i$ permet d'exprimer les fonctions de dépense en fonction du nombre d'observations. En notant la probabilité de franchir la borne inférieure à la k^{ieme} analyse

$$P^L(k) = Pr(S_{j,L} < Y_j < S_{j,U}, j = 1, \ldots, k-1 \text{ et } Y_k \leq S_{k,L}|H_0)$$

et la probabilité de poursuivre l'expérience jusqu'à la k^{ieme} analyse et d'observer $Y_k = y$

$$P^k(y) = Pr(S_{j,L} < Y_j < S_{j,U}, j = 1, \ldots, k-1 \text{ et } Y_k = y|H_0),$$

les valeurs $S_{k,L}$ et $S_{k,U}$ sont obtenues récursivement par les relations

$$Pr(\text{borne supérieure franchie à la } k^{ieme} \text{ analyse}|H_0) = 1 - \sum_{i=1}^{k-1} P^L(i) - \sum_{y=0}^{S_{k,U}-1} P^k(y)$$

$$Pr(\text{borne inférieure franchie à la } k^{ieme} \text{ analyse}|H_0) = \sum_{i=1}^{k-1} P^L(i) + \sum_{y=0}^{S_{k,L}} P^k(y).$$

5.2 Test fréquentiste conditionnel séquentiel ($TFCS^*$)

5.2.1 Construction du test

Comme dans les chapitres précédents, le plan séquentiel étudié comporte deux échantillonnages binomiaux successifs $Y_1 \sim Bin(\theta, n_1)$ et $Y_2 \sim Bin(\theta, n_2)$ avec un arrêt optionnel entre les deux si $Y = Y_1 \notin J$ où J est la zone de continuation. L'expérience s'arrête dès la première analyse (i.e. $k = 1$) si $Y \in \bar{J}$, sinon elle continue (i.e. $k = 2$).

Le cas général d'une hypothèse nulle et d'une hypothèse alternative composite avec

$$H_0 : \{\theta \in \Theta_0\} \qquad \text{versus} \qquad H_1 : \{\theta \in \Theta_1\}$$

est abordé dans le plan $d_{Seq} = \{n_1, J, n_2\}$.

La solution bayésienne classique ignore la règle d'arrêt de l'expérience. La séquence (k, y) est associée à la densité marginale

$$m_{i,k}(y) = \int_{\Theta_i} p(y|\theta, k)\pi_i(\theta)d\theta \qquad \text{sous } H_i \quad i = 1, 2$$

avec $y = \sum_{i=1}^{k} y_i$ et $n = \sum_{i=1}^{k} n_i$. On considère alors le facteur de Bayes

$$B_k = B_k(y) = m_{0,k}(y)/m_{1,k}(y).$$

L'a priori de Jeffreys classique $\vartheta \sim Be(0.5, 0.5)$ sous H_i confère un statut d'indifférence sur les valeurs du paramètre dans Θ_i.

On a vu dans (2.3) que l'a priori de Jeffreys corrigé $\vartheta|d_{Seq} \sim Be^J(0.5, 0.5, 0.5; n_1, r; J)$ tient compte de la nature séquentielle du plan expérimental. Cet a priori est associé à la densité prédictive

$$m_{i,k}(y|d_{Seq}) = \int_{\Theta_i} p(y|\theta, k)\pi_i(\theta|d_{Seq})d\theta \qquad \text{sous } H_i$$

qui conduit au facteur de Bayes

$$B_{k,Seq} = B_k(y|d_{Seq}) = m_{0,k}(y|d_{Seq})/m_{1,k}(y|d_{Seq}).$$

C'est la solution bayésienne alternative décrite dans les chapitres précédents. La prise en compte de la règle d'arrêt dans l'a priori de Jeffreys traduit l'impartialité du plan expérimental. Les conclusions sont portées par les probabilités d'erreur des hypothèses qui sont identiques aux risques fréquentistes conditionnels (i.e. $\alpha(y|d_{Seq}) = \alpha^*(B_{k,Seq})$ et $\beta(y|d_{Seq}) = \beta^*(B_{k,Seq})$). Dans ce dernier cas, le conditionnement se fait sur un partitionnement de l'espace des observations offrant la même évidence sous H_0 et sous H_1 (voir section (B.5.3)).

La procédure présentée nécessite simplement de définir une valeur (α_r^*, β_a^*), associée à $(\alpha^*(B_{k,Seq}), \beta^*(B_{k,Seq}))$, en deçà de laquelle on considère l'évidence assez forte en faveur de H_0 ou H_1 pour arrêter l'expérience en période 1. Les valeurs de y_1 n'offrant pas l'évidence suffisante pour pouvoir conclure appartiennent à J, la zone de continuation de l'expérience en période 2. Cette formalisation étend la notion de zone de non-décision du test de Dass (voir (B.5.5)).

La relation établie entre $\alpha^*(B_{k,Seq})$, $\beta^*(B_{k,Seq})$ et $B_k(y|d_{Seq})$ permet également de définir la procédure par un bornage (B_r, B_a) des valeurs de $B_k(y|d_{Seq})$.

Test fréquentiste conditionnel Séquentiel *($TFCS^*$). En notant $\mathcal{R}^*_{(i)}$ et $\mathcal{A}^*_{(i)}$ les zones de rejet et d'acceptation de H_0 en période i, on définit le test fréquentiste conditionnel séquentiel ($TFCS^*$) pour les valeurs de y. En période 1,*

si $\alpha^*(B_{1,Seq}) \leq \alpha_r^*$ *ou si* $B_k(y_1|d_{Seq}) \leq B_r$ *ou si* $y_1 \in \mathcal{R}^*_{(1)}$,

H_0 *est rejetée avec un risque conditionnel*

$\alpha^*(B_{1,Seq}) = B_k(y_1|d_{Seq})/(1 + B_k(y_1|d_{Seq}))$.

Si $[\alpha^*(B_{1,Seq}) > \alpha_r^*$ et $\beta^*(B_{1,Seq}) > \beta_a^*]$ *ou si* $B_r < B_k(y_1|d_{Seq}) < B_a$ *ou si*

$y_1 \notin \mathcal{R}^*_{(1)} \cup \mathcal{A}^*_{(1)}$, *l'expérience se poursuit en période 2.*

Si $\beta^*(B_{1,Seq}) \leq \beta_a^*$ *ou si* $B_k(y_1|d_{Seq}) \geq B_a$ *ou si* $y_1 \in \mathcal{A}^*_{(1)}$,

H_0 *est acceptée avec un risque conditionnel* $\beta^*(B_{1,Seq}) = 1/(1 + B_k(y_1|d_{Seq}))$.

Si l'expérience est poursuivie en période 2, on applique le test de Dass à $Y = Y_1 +$

Y_2. Les zones $\mathcal{R}^*_{(2)}$,$\mathcal{A}^*_{(2)}$ et les risques conditionnels associés $\alpha^*(B_{2,Seq})$,$\beta^*(B_{2,Seq})$ se déduisent de la règle (B.5.5)

Une variante du test $TFCS^*$ consiste à conclure en période 2 suivant la règle appliquée sur la période 1.

Variante du test $TFCS^*$. Quel que soit la période $k = 1, 2$,

si $\alpha^*(B_{k,Seq}) \leq \alpha^*_r$ *ou si* $B_k(y|d_{Seq}) \leq B_r$ *ou si* $y \in \mathcal{R}^*_{(k)}$,

H_0 *est rejetée avec un risque conditionnel*

$\alpha^*(B_{k,Seq}) = B_k(y|d_{Seq})/(1 + B_k(y|d_{Seq}))$.

Si $[\alpha^*(B_{k,Seq} > \alpha^*_r$ et $\beta^*(B_{k,Seq}) > \beta^*_a)]$ *ou si* $B_r < B_k(y|d_{Seq}) < B_a$ *ou si*

$y \notin \mathcal{R}^*_{(k)} \cup \mathcal{A}^*_{(k)}$, *pas de conclusion sur* H_0.

Si $\beta^*(B_{k,Seq}) < \beta^*_a$ *ou si* $B_k(y|d_{Seq}) > B_a$ *ou si* $y \in \mathcal{A}^*_{(k)}$,

H_0 *est acceptée avec un risque conditionnel* $\beta^*(B_{k,Seq}) = 1/(1 + B_k(y|d_{Seq}))$.

L'expérience s'arrête en période 1 si une conclusion peut être portée sur H_0. Sinon l'expérience continue en période 2.

Le test $TFCS^*$ sous-entend que J la zone de continuation de l'expérience en période 2 pour y_1 et $\bar{\mathcal{R}}^*_{(2)}$ la zone de non-décision en période 2 pour $y = y_1+y_2$ dans la variante du test $TFCS^*$ recouvrent entièrement les zones de non-décision des tests de Dass en période 1 et 2. La conclusion sur H_0 doit donc être motivée par une évidence suffisamment forte. Dans le cas contraire, les zones J et $\bar{\mathcal{R}}^*_{(2)}$ doivent s'étendre aux zones de non-décision des tests de Dass. Pour les valeurs concernées par l'assertion aléatoire, une solution pratique consiste à confirmer la décision si la probabilité associée est supérieure à 0.5. Par exemple, si il y a une probabilité $p = 0.7 > 0.5$ que H_0 soit rejetée en période 1 alors H_0 est rejetée et l'expérience s'arrête.

La détermination des zones $\mathcal{A}^*_{(1)}$ et $\mathcal{R}^*_{(1)}$ en période 1 nécessite une approche récursive car la définition de la zone de continuation de l'expérience est un paramètre de la loi a priori corrigée. Si le facteur de Bayes $B_1(y|d_{Seq})$ est une fonction monotone

décroissante, la méthode suivante s'applique sur y, sinon il convient de l'appliquer sur les valeurs z qui sont les y réordonnées suivant l'ordre décroissant des valeurs de $B_1(y|d_{Seq})$; la fonction $B_1(z|d_{Seq})$ devient ainsi monotone décroissante (voir fin de la section (B.5.3)).

La zone de rejet et d'acceptation de H_0 présente une forme du type $\mathcal{R}^*_{(1)} = [0, S_{1,L}]$ et $\mathcal{A}^*_{(1)} = [S_{1,U}, n_1]$, pour lesquelles les valeurs de $S_{1,L}$ et $S_{1,U}$ sont à déterminer. On commence avec les valeurs initiales $(S^0_{1,L}, S^0_{1,U})$ déterminée à partir de l'a priori de Jeffreys non-corrigé. Ensuite, les valeurs $(S^1_{1,L}, S^1_{1,U})$ sont calculées en utilisant l'a priori corrigé et défini par la zone de continuation $J =]S^0_{1,L}, S^0_{1,U}[$. Si $S^0_{1,L} = S^1_{1,L}$ et $S^0_{1,U} = S^1_{1,U}$, les bornes correspondent aux valeurs $S_{1,L} = S^1_{1,L}$ et $S_{1,U} = S^1_{1,U}$. Sinon, on doit calculer $(S^{i+1}_{1,L}, S^{i+1}_{1,U})$ en utilisant l'a priori corrigé et défini par la zone de continuation $J =]S^i_{1,L}, S^i_{1,U}[$ jusqu'à l'obtention des égalités $S^i_{1,L} = S^{i+1}_{1,L}$ et $S^i_{1,U} = S^{i+1}_{1,U}$.

5.2.2 Application à un test de supériorité

Une application du test $TFCS^*$ est décrite dans le plan séquentiel $d_{Seq} = \{n_1 = 10, J, n_2 = 10\}$. On considère deux échantillonnages binomiaux successifs $Y_1 \sim Bin(\theta, n_1)$ et $Y_2 \sim Bin(\theta, n_2)$ pour tester les hypothèses

$$H_0 : \{\theta \geq 0.5\} \qquad \text{versus} \qquad H_1 : \{\theta < 0.5\}. \tag{5.2.1}$$

Lorsque l'a priori de Jeffreys corrigé est utilisé, le facteur de Bayes s'écrit

$$B_k(y|d_{Seq}) = m_{0,k}(y|d_{Seq})/m_{1,k}(y|d_{Seq}),$$

où $m_{i,k}(y|d_{Seq})$ désigne la probabilité prédictive marginale de Y sous H_i.

On définit une règle d'arrêt unilatérale en période 1 basée sur la valeur $\alpha^*_r = 0.20$. Si $\alpha^*(B_{1,Seq}) \leq 0.20$ l'expérience s'arrête, sinon elle continue en période 2. La valeur $\alpha^*_r = 0.20$ correspond au facteur de Bayes $B_r = 1/4$ (i.e. l'expérience s'arrête en période 1 si $B_k(y_1|d_{Seq}) < 1/4$). En notant que la fonction B_k est strictement croissante,

y	$B(y)$	Décision*	Risque*
0	0.0006	H_0 rejetée	0.0004
1	0.013	H_0 rejetée	0.013
2	0.073	H_0 rejetée	0.068
3	0.228	H_0 rejetée	0.185
4	0.125	H_0 rejetée	0.339
5	1	non-décision	–
6	1.95	H_0 acceptée	0.339
7	4.39	H_0 acceptée	0.185
8	13.7	H_0 acceptée	0.068
9	77.7	H_0 acceptée	0.013
10	1509	H_0 acceptée	0.0004

TAB. 5.1 – *Test de Dass avec l'a priori de Jeffreys classique pour* $n=10$

les valeurs de y_1 conduisant à l'arrêt de l'expérience en période 1 correspondent à la zone de rejet de H_0 du type $\mathcal{R}^*_{(1)} = [0, S_{1,L}]$.

Le calcul de la borne $S_{1,L}$ s'effectue suivant la procédure itérative décrite dans la section précédente. La valeur initiale $S^0_{1,L}$ est déduite des résultats du test de Dass sur l'échantillon fixe $n = n_1 = 10$. Le tableau (5.1) montre que la plus grande valeur de $\alpha^*(B)$ inférieure à $\alpha^*_r = 0.20$ correspond à $y_1 = 3$ avec $\alpha^*(B) = 0.185$. La zone de continuation $J = [4, 10]$ est ensuite introduite dans l'a priori de Jeffreys corrigé pour la seconde itération. Pour $y = 3$, on obtient la valeur $\alpha^*(B_{1,Seq}) = 0.147$ ce qui traduit une diminution du risque de type 1, mais reste la plus grande valeur en deçà de la limite $\alpha^*_r = 0.20$. On maintient donc la valeur $S_{1,L} = S^0_{1,L} = S^1_{1,L} = 3$.

Le tableau (5.2) montre les densités prédictives et les valeurs déduites du facteur de Bayes pour l'ensemble des séquences (k, y) possibles obtenues avec les a priori de Jeffreys classique et corrigé sur un effectif $n_1 = n_2 = 10$. L'utilisation de l'a priori de Jeffreys corrigé $\vartheta|d_{Seq} \sim Be^J(0.5, 0.5, 0.5; n_1, r; J)$ par rapport à l'a priori de Jeffreys classique $\vartheta \sim Be(0.5, 0.5)$ induit une diminution des probabilités prédictives (i.e. $m_{i,k}(y|d_{Seq}) < m_{i,k}(y)$, $i = 0, 1$, $k = 1, 2$) pour les petites valeurs du rapport y/n. Les lois a priori définies sur $\Theta_1 = [0, 0.5[$ sous H_1 et sur $\Theta_0 = [0.5, 1]$ sous H_0 provoquent un décalage de l'effet de la correction. Sous H_0, la diminution des

	A priori Jeffreys corrigé			A priori Jeffreys classique		
(k,y)	$m_{0,k}(y\|d_{Seq})$	$m_{1,k}(y\|d_{Seq})$	$B_k(y\|d_{Seq})$	$m_{0,k}(y)$	$m_{1,k}(y)$	$B_k(y)$
(1,0)	0.0001	0.213	0.0006	0.0002	0.257	0.0006
(1,1)	0.001	0.137	0.011	0.002	0.148	0.013
(1,2)	0.008	0.136	0.058	0.010	0.132	0.073
(1,3)	0.025	0.143	0.172	0.030	0.132	0.228
(2,3)	0.0001	0.066	$< 10^{-5}$	0.0002	0.069	0.002
(2,4)	0.0006	0.067	$< 10^{-5}$	0.0007	0.066	0.011
(2,5)	0.002	0.071	0.028	0.002	0.067	0.038
(2,6)	0.006	0.077	0.073	0.007	0.070	0.099
(2,7)	0.012	0.082	0.153	0.015	0.073	0.210
(2,8)	0.023	0.084	0.277	0.028	0.075	0.382
(2,9)	0.036	0.079	0.460	0.045	0.070	0.633
(2,10)	0.050	0.068	0.735	0.060	0.060	1
(2,11)	0.060	0.050	1.18	0.070	0.045	1.58
(2,12)	0.065	0.032	2.02	0.075	0.028	2.61
(2,13)	0.067	0.017	3.83	0.073	0.015	4.75
(2,14)	0.067	0.008	8.56	0.070	0.007	10.1
(2,15)	0.068	0.003	23.7	0.067	0.002	26.4
(2,16)	0.070	0.0008	85.0	0.066	0.0007	90.4
(2,17)	0.075	0.0002	417	0.069	0.0002	430
(2,18)	0.085	$< 10^{-4}$	3102	0.076	$< 10^{-4}$	3259
(2,19)	0.107	$< 10^{-5}$	4.1×10^4	0.095	$< 10^{-5}$	4.1×10^4
(2,20)	0.203	$< 10^{-6}$	1.6×10^6	0.180	$< 10^{-6}$	1.6×10^6

TAB. 5.2 – *Densités prédictives et facteur de Bayes de H_0 sur H_1 avec les a priori de Jeffreys classique et corrigé pour $\alpha_r^* = 0.2$ et $n_1 = n_2 = 10$*

probabilités prédictives intervient pour $y_1 \in \mathcal{R}^*_{(1)}$ en période 1 et $y = y_1 + y_2 \in [3, 15]$ en période 2 contre $y_1 \leq 1$ et $y = y_1 + y_2 = 3$ sous l'alternative. Ce décalage sur les valeurs de y se traduit par une diminution du facteur de Bayes (i.e. $B_k(y|d_{Seq}) < B_k(y)$) quel que soit la séquence (k, y).

Dans un plan expérimental où la règle d'arrêt se fonde sur le rapport des évidences entre H_0 et H_1, la correction induit un mécanisme de compensation des effets de la règle d'arrêt. Lors de l'arrêt optionnel, les valeurs du facteur de Bayes sont en moyenne supérieures à celles obtenues si l'expérience ne comportait pas d'arrêt optionnel, favorisant de fait l'hypothèse H_0. En diminuant le facteur de Bayes, la correction vient réajuster le rapport des évidences en faveur de H_1. Elle est la plus

forte pour les valeurs de y/n proches de $S_{1,L}/n_1 = 0.3$ et diminue à mesure que la séquence (k, y) s'en éloigne.

On en déduit l'apport de la correction de l'a priori de Jeffreys sur les décisions et les risques conditionnels dans le test $TFCS^*$. Le réajustement du rapport des évidences en faveur de H_1 provoque une diminution du risque conditionnel d'erreur lorsque H_0 est rejeté (i.e. $\alpha^*(B_{k,Seq}) < \alpha^*(B_k)$) et une augmentation de ce risque lorsque H_0 est acceptée (i.e. $\beta^*(B_{k,Seq}) > \beta^*(B_k)$). L'impact sur les risques conditionnels peut donc conduire à une augmentation de la zone de rejet de H_0 et une diminution de la zone d'acceptation dans le cas d'un bornage unilatéral par une valeur limite α_r^*.

Dans le plan $d_{Seq} = (n_1 = 10, S_{1,L} = 3, n_1 = 10)$, le tableau (5.3) montre les décisions sur H_0 et les risques conditionnels du test $TFCS^*$ pour l'ensemble des séquences (k, y) possibles. En période 2, la symétrie des hypothèses (5.2.1) conjuguée à l'utilisation de l'a priori de Jeffreys classique engendre une valeur $B_k(y) = 1$ pour la séquence centrale $(k, y) = (2, 10)$. Cette valeur du facteur de Bayes caractérise une absence totale d'évidence en faveur du rejet ou de l'acceptation de H_0 qui conduit à une situation de non-décision. La correction vient placer la séquence $(k, y) = (2, 10)$ dans la zone de rejet de H_0 avec cependant une grande incertitude, quantifiée par la composante aléatoire $r = 0.729$.

La figure (5.1) permet de visualiser les résultats du tableau (5.3). La séquence $(k, y) = (2, 10)$ avec l'a priori de Jeffreys corrigé n'y est pas représentée car la probabilité $p = 1 - r = 0.271$ associée au rejet de H_0 est inférieure à 0.5. Ce cas pose un problème d'interprétation qui rejoint les discussions autour des décisions associées à une évidence faible (voir Berger (1997) [5]).

La variante du test $TFCS^*$ présentée en section (5.2.1) permet de contourner ce débat. Associée à la limite $\alpha_r^* = 0.20$ en période 1 et 2, les zones de rejet de la variante sont $\mathcal{R}^*_{(1)} = [0, 3]$ pour $y = y_1$ en période 1 et $\mathcal{R}^*_{(2)} = [3, 7]$ pour $y = y_1 + y_2$ en période 2. La correction vient diminuer l'erreur de type 1 pour la séquence $(k, y) = (2, 8)$ de

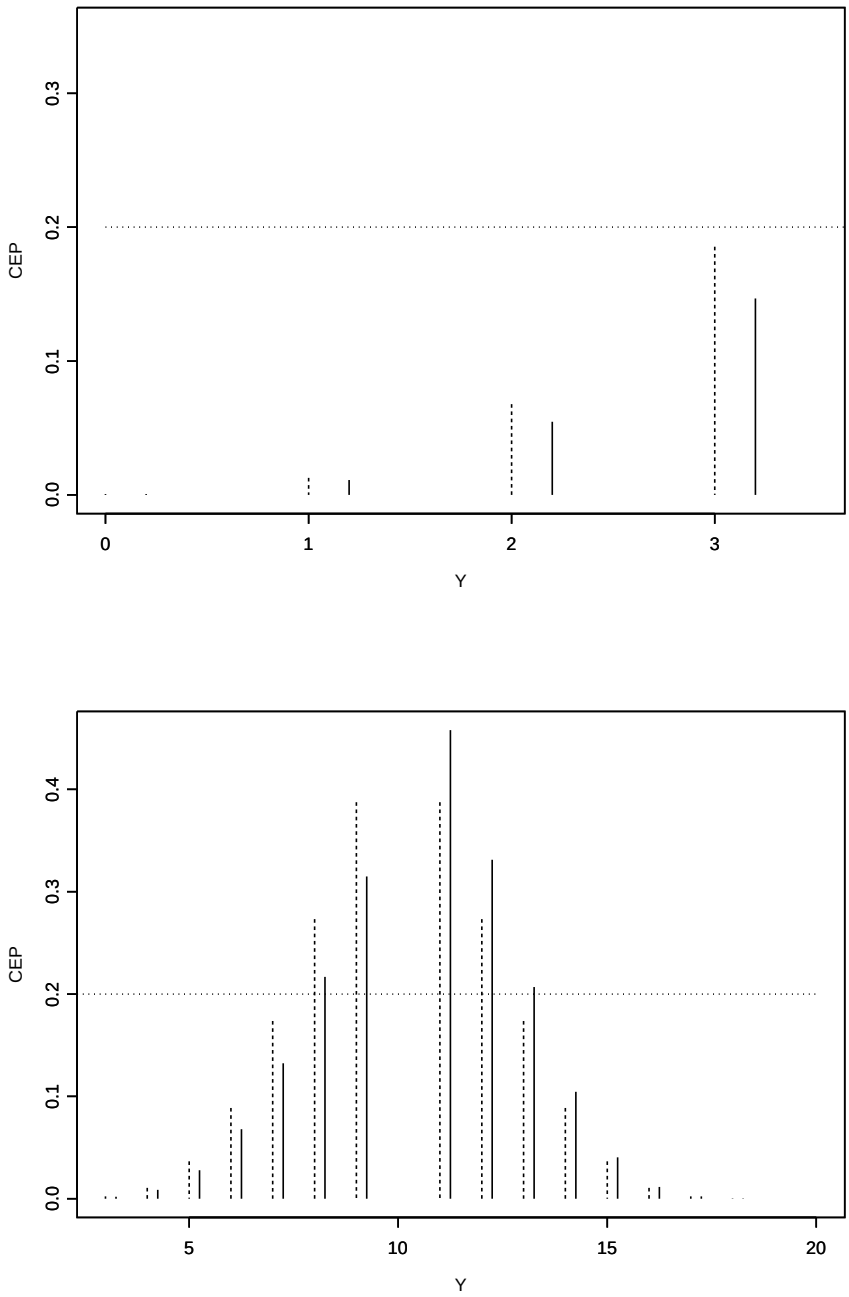

FIG. 5.1 – *Risques fréquentistes conditionnels du test $TFCS^*$ avec les a priori de Jeffreys classique (- - -) et corrigé (—) pour $\alpha_r^* = 0.2$ et $n_1 = n_2 = 10$*

	A priori Jeffreys corrigé			A priori Jeffreys classique		
(k,y)	Décision*	Risque*	Random*	Décision*	Risque*	Random*
(1,0)	H_0 rejetée	0.006	_	H_0 rejetée	0.0006	_
(1,1)	H_0 rejetée	0.011	_	H_0 rejetée	0.013	_
(1,2)	H_0 rejetée	0.055	_	H_0 rejetée	0.068	_
(1,3)	H_0 rejetée	0.147	_	H_0 rejetée	0.185	_
(2,3)	H_0 rejetée	0.002	_	H_0 rejetée	0.002	_
(2,4)	H_0 rejetée	0.008	_	H_0 rejetée	0.011	_
(2,5)	H_0 rejetée	0.028	_	H_0 rejetée	0.036	_
(2,6)	H_0 rejetée	0.068	_	H_0 rejetée	0.090	_
(2,7)	H_0 rejetée	0.132	_	H_0 rejetée	0.174	_
(2,8)	H_0 rejetée	0.216	_	H_0 rejetée	0.277	_
(2,9)	H_0 rejetée	0.315	_	H_0 rejetée	0.387	_
(2,10)	H_0 rejetée	0.424	0.729	non-décision	_	1
(2,11)	H_0 acceptée	0.458	_	H_0 acceptée	0.387	_
(2,12)	H_0 acceptée	0.331	_	H_0 acceptée	0.277	_
(2,13)	H_0 acceptée	0.207	_	H_0 acceptée	0.174	_
(2,14)	H_0 acceptée	0.105	_	H_0 acceptée	0.090	_
(2,15)	H_0 acceptée	0.040	_	H_0 acceptée	0.011	_
(2,16)	H_0 acceptée	0.011	_	H_0 acceptée	0.055	_
(2,17)	H_0 acceptée	0.002	_	H_0 acceptée	0.002	_
(2,18)	H_0 acceptée	0.0003	_	H_0 acceptée	0.0003	_
(2,19)	H_0 acceptée	$< 10^{-4}$	_	H_0 acceptée	$< 10^{-4}$	_
(2,20)	H_0 acceptée	$< 10^{-6}$	_	H_0 acceptée	$< 10^{-6}$	_

TAB. 5.3 – *Décisions et risques fréquentistes conditionnels du test $TFCS^*$ avec les a priori de Jeffreys classique et corrigé pour $\alpha_r^* = 0.2$ et $n_1 = n_2 = 10$*

$\alpha^*(B_2) = 0.277$ à $\alpha^*(B_{2,Seq}) = 0.216$, mais cette dernière valeur reste supérieure à $\alpha_r^* = 0.20$.

Cette première application permet d'aboutir à une règle plus générale. Lorsque l'arrêt optionnel favorise l'évidence d'une hypothèse, la correction introduit un mécanisme de compensation sur le rapport des évidences (facteur de Bayes) et sur les risques conditionnels d'erreur du test $TFCS^*$. La variante du test $TFCS^*$, qui conditionne l'arrêt de l'expérience en période 1 et le rejet de H_0 en période 1 et 2 à un même niveau d'évidence minimum, offre une interprétation simple et cohérente.

5.3 Test fréquentiste non-conditionnel séquentiel ($TFNCS$)

5.3.1 Construction du test

La règle du test décrite dans cette section est fondée sur l'approche fréquentiste conditionnelle (ou bayésienne) développée dans la section précédente. Une procédure simple permet d'ajuster le plan expérimental pour satisfaire des propriétés définies dans le contexte de l'inférence fréquentiste non-conditionnelle. En particulier, la notion de puissance conditionnellement à une erreur de type 1 maintenue en deçà d'une valeur predéfinie du risque α est un des fondements de l'approche décisionnelle du test de Neyman-pearson (voir section (B.4)).

L'assertion "H_0 est rejetée" si $\alpha^*(B_{k,Seq}) < \alpha_r^*$ quel que soit la période $k = 1, 2$ apparaît comme la transposition la plus directe d'une approche qui vise à privilégier l'interprétation dans la règle de décision d'un test. Un prolongement naturel de cette règle consiste à définir la séquentialité (i.e. la poursuite en période 2 de l'expérience) suivant cette même condition. Cette stratégie rejoint le principe de la variante du test $TFCS^*$ exposé dans la section 5.2.1.

On obtient ainsi le test fréquentiste non-conditionnel séquentiel ($TFNCS$) défini par une valeur limite B_r du facteur de Bayes ou indifféremment par une valeur limite α_r^* de l'erreur de type 1 dans le test fréquentiste conditionnel. On en déduit deux zones de rejet de H_0 pour les valeurs de y, $\mathcal{R}_{(1)}$ en période 1 et $\mathcal{R}_{(2)}$ en période 2, respectivement identiques à $\mathcal{R}_{(1)}^*$ et $\mathcal{R}_{(2)}^*$ définies dans la variante du test $TFCS^*$.

Test fréquentiste non-conditionnel séquentiel *($TFNCS$).* Quel que soit la période $k = 1, 2$,

$$\begin{aligned} &\textit{si } \alpha^*(B_{k,Seq}) \leq \alpha_r^* \textit{ ou si } B_k(y|d_{Seq}) \leq B_r \textit{ ou si } y \in \mathcal{R}_{(k)}, \\ &\quad H_0 \textit{ est rejetée avec un risque non-conditionnel prédéfini } \alpha, \\ &\textit{sinon pas de conclusion sur } H_0. \end{aligned}$$

L'expérience s'arrête si H_0 est rejetée en période 1. Sinon, l'expérience continue en

	Jeffreys corrigé		Jeffreys classique	
	$S_{1,L}$	$S_{2,L}$	$S_{1,L}$	$S_{2,L}$
$n_1 = n_2 = 10$	3	7	–	–
$n_1 = n_2 = 20$	7	16	–	–
$n_1 = n_2 = 30$	12	26	11	25
$n_1 = n_2 = 40$	16	35	16	34
$n_1 = n_2 = 50$	21	45	20	43
$n_1 = n_2 = 60$	26	54	25	53
$n_1 = n_2 = 80$	35	73	34	72
$n_1 = n_2 = 100$	45	92	43	91

TAB. 5.4 – *Bornes ($S_{1,L}$,$S_{2,L}$) avec les a priori de Jeffreys classique et corrigé pour* $\alpha_r^* = 0.20$ *(i.e.* $B_r = 1/4$*) et* n_1=n_2=10, 20, 30*, 40*, 50*, 60*, 80*, 100*

	Jeffreys corrigé		Jeffreys classique	
	$S_{1,L}$	$S_{2,L}$	$S_{1,L}$	$S_{2,L}$
$n_1 = n_2 = 10$	2	6	–	–
$n_1 = n_2 = 20$	6	15	6	14
$n_1 = n_2 = 30$	10	23	–	–
$n_1 = n_2 = 40$	15	32	14	32
$n_1 = n_2 = 50$	19	42	18	41
$n_1 = n_2 = 60$	23	51	23	50
$n_1 = n_2 = 80$	32	70	32	68
$n_1 = n_2 = 100$	42	88	41	87

TAB. 5.5 – *Bornes ($S_{1,L}$,$S_{2,L}$) avec les a priori de Jeffreys classique et corrigé pour* $\alpha_r^* = 0.10$ *(i.e.* $B_r = 1/9$*) et* n_1=n_2=10, 20*, 30, 40*, 50*,60*, 80*, 100*

période 2.

Si le facteur de Bayes est une fonction monotone décroissante, les régions de rejet de H_0 sont du type $\mathcal{R}_{(1)} = \mathcal{R}^*_{(1)} = [0, S_{1,L}]$ et $\mathcal{R}_{(2)} = \mathcal{R}^*_{(2)} = [0, S_{2,L}]$. Sinon, ce bornage par limite supérieure s'applique aux valeurs de z qui sont les valeurs de y ordonnées suivant les valeurs décroissantes du facteur de Bayes.

Des valeurs de $S_{1,L}$ et $S_{2,L}$ associées aux hypothèses (5.2.1) sont fournies dans les tableaux (5.4) et (5.5). Ces tableaux correspondent respectivement à des valeurs $\alpha_r^* = 0.20$ (i.e. $B_r = 1/4$) et $\alpha_r^* = 0.10$ (i.e. $B_r = 1/9$) avec les effectifs $n_1 = n_2 = 10, 20, 30, 40, 50, 60, 80, 100$. Les bornes $S_{1,L}$ et $S_{2,L}$ calculées avec l'a priori de

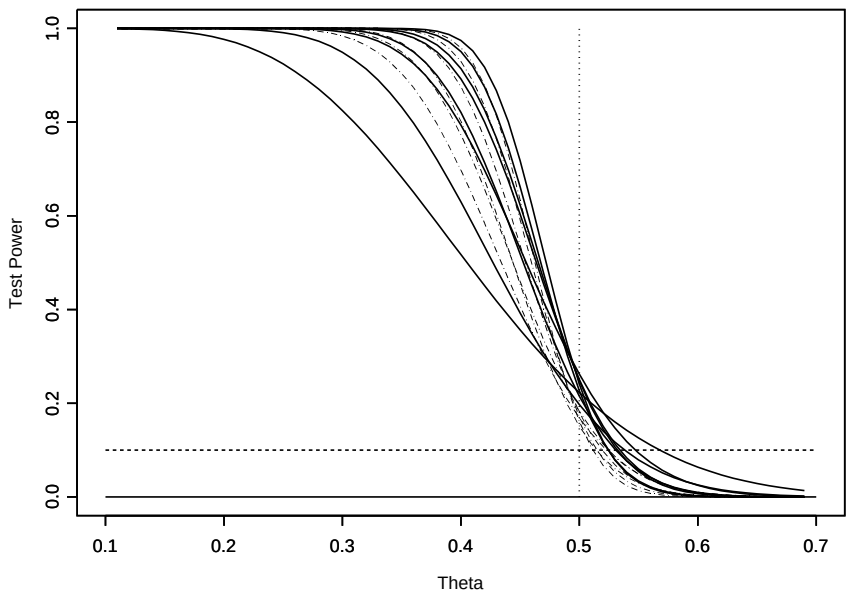

FIG. 5.2 – *Courbes de puissance du test $TFNCS$ avec les a priori de Jeffreys classique ($\cdot - \cdot$) et corrigé (—) pour $\alpha_r^* = 0.20$ et n_1=n_2=10, 20, 30*, 40*, 50*, 60*, 80*, 100**

Jeffreys classique sont données en cas de modification par rapport à l'a priori de Jeffreys corrigé. Dans ce cas, on note que les valeurs de $S_{1,L}$ et/ou $S_{2,L}$ calculées avec la correction augmentent, ce qui étend les zones $\mathcal{R}_{(1)}$ et/ou $\mathcal{R}_{(2)}$.

La performance d'un test fréquentiste non-conditionnel se mesure en terme de puissance. En notant que la zone de continuation de l'expérience en période 2 pour y_1 est le complémentaire de la zone de rejet de H_0 (i.e. $J = \bar{\mathcal{R}}_{(1)}$), on obtient la puissance par

$$\begin{aligned} Pow(\theta|d_{Seq}) &= E_{Seq,\theta}[1_{y_1 \in \mathcal{R}_{(1)}}] + E_{Seq,\theta}[1_{y_1+y_2 \in \mathcal{R}_{(2)}}] \\ &= \sum_{i \in \mathcal{R}_{(1)}} P_\theta(Y_1 = i) + \sum_{\substack{i+j=l \\ i \in J,\, l \in \mathcal{R}_{(2)}}} P_\theta(Y_1 = i) P_\theta(Y_2 = j). \end{aligned}$$

Bien que la nature discrète des données ne permet pas l'établissement d'une règle générale pour $\theta \in \Theta_1$, l'augmentation des effectifs $n_1 = n_2$ coincide avec un accroissement de la puissance lorsque θ s'éloigne suffisamment de Θ_0. Les courbes des

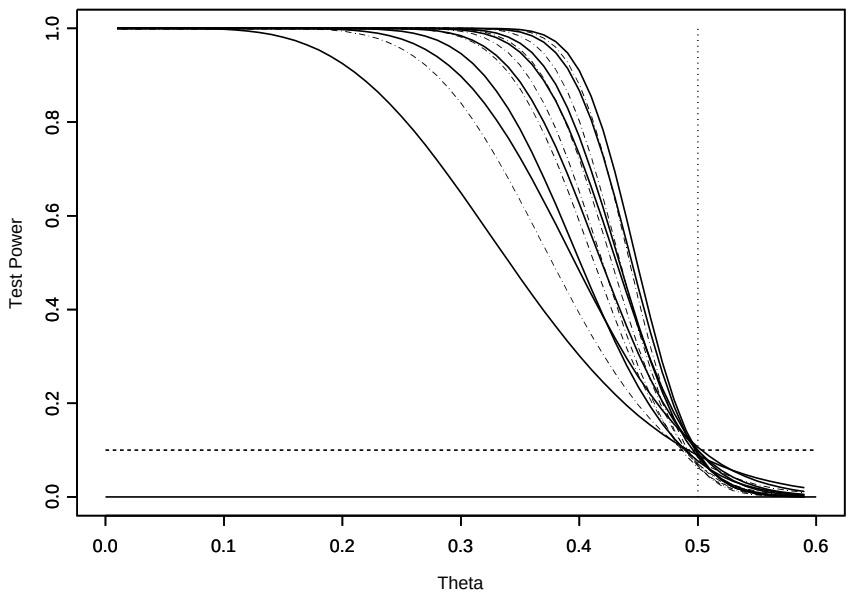

FIG. 5.3 – *Courbes de puissance du test $TFNCS$ avec les a priori de Jeffreys classique $(\cdot - \cdot)$ et corrigé (—) pour $\alpha_r^* = 0.10$ et n_1=n_2= 10, 20*, 30, 40*, 50*, 60*, 80*, 100**

fonctions puissance sont décrites avec $\alpha_r^* = 0.20$ dans la figure (5.2) et avec $\alpha_r^* = 0.10$ dans la figure (5.3) pour les effectifs $n_1 = n_2 = 10, 20, 30, 40, 50, 60, 80, 100$.

Ces courbes permettent un ajustement de l'effectif en fonction d'une probabilité minimum (i.e. puissance) pour détecter une valeur particulière de $\theta = \theta_1$ sous H_1. Les pentes augmentent avec l'effectif. Elles reflètent un pouvoir discriminant croissant entre deux valeurs de $\theta \in \Theta_1$. La correction de l'a priori de Jeffreys induit une augmentation de la puissance (i.e. $Pow(\theta|d_{Seq}) \geq Pow(\theta)$) qui se traduit par un décalage des courbes vers les valeurs supérieures de θ.

Lorsque l'hypothèse nulle est composite (i.e. $H_0 : \{\theta \in \Theta_0\}$), l'erreur de type 1 du test $TFNCS$ correspond à $Max_{\theta \in \Theta_0} Pow(\theta|d_{Seq})$. Avec les hypothèses (5.2.1), ce maximum est atteint pour $\theta = 0.5$. En définissant une valeur de α_r^*, il convient de vérifier que l'effectif envisagé garantit le risque nominal α. Le cas échéant, on diminue l'erreur de type 1 du test $TFNCS$ en diminuant la valeur de α_r^*.

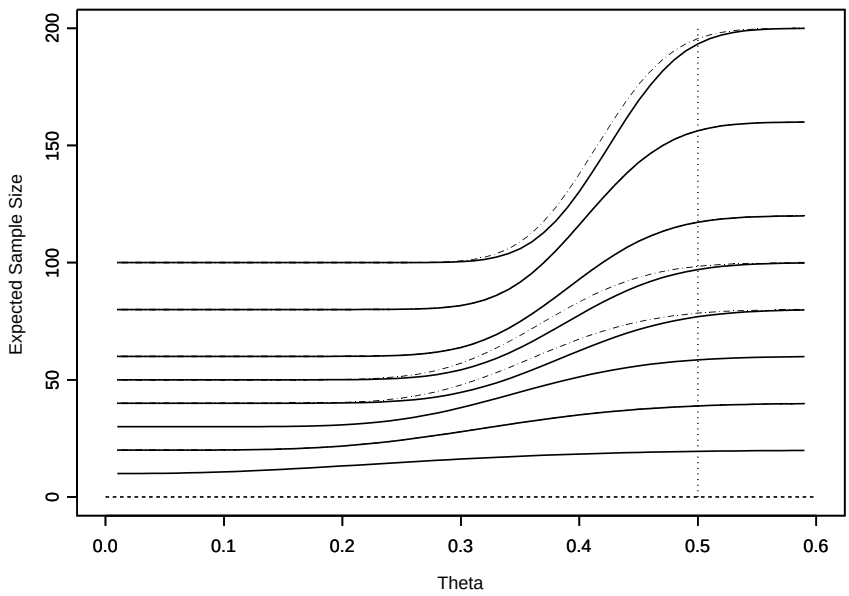

FIG. 5.4 – *Courbes des effectifs moyens du test $TFNCS$ avec les a priori de Jeffreys classique ($\cdot - \cdot$) et corrigé (—) pour $\alpha_r^* = 0.10$ et n_1=n_2=10, 20*, 30, 40*, 50*,60*, 80*, 100**

Pour $\alpha = 0.10$, la figure (5.2) montre que la valeur $\alpha_r^* = 0.20$ (i.e. $B_r = 1/4$) ne garantit dans aucun des cas $n_1 = n_2 = \{10, .., 100\}$ le risque nominal prédéfini. En diminuant la limite à $\alpha_r^* = 0.10$ (i.e. $B_r = 1/9$), la figure (5.3) montre que les erreurs de type 1 du test $TFNCS$ sont très proches de la valeur nominale $\alpha = 0.10$. Seul l'effectif $n_1 = n_2 = 20$ aboutit à une erreur de type 1 supérieure (i.e. $0.106 > 0.10$).

La performance fréquentiste d'un test séquentiel se mesure également en terme d'effectif moyen en fonction des valeurs de $\theta \in \Theta_0 \cup \Theta_1$. On visualise les courbes associées aux cas $n_1 = n_2 = \{10, 20, 30, 40, 50, 60, 80, 100\}$ dans la figure (5.4) pour $\alpha_r^* = 0.10$. Lorsque la correction de l'a priori de Jeffreys modifie la borne $S_{1,L}$ en période 1, l'augmentation de la puissance se traduit par une diminution de l'espérance des effectifs.

	Jeffreys corrigé		Jeffreys classique	
	$S_{1,L}$	$S_{2,L}$	$S_{1,L}$	$S_{2,L}$
$n_1 = n_2 = 10$	1	5	–	–
$n_1 = n_2 = 20$	5	13	–	–
$n_1 = n_2 = 30$	9	22	9	21
$n_1 = n_2 = 40$	13	31	13	30
$n_1 = n_2 = 50$	18	40	17	39
$n_1 = n_2 = 60$	22	49	21	48
$n_1 = n_2 = 80$	31	67	30	66
$n_1 = n_2 = 100$	40	87	39	84

TAB. 5.6 – *Bornes ($S_{1,L}$,$S_{2,L}$) avec les a priori de Jeffreys classique et corrigé pour $\alpha_r^* = 0.05$ (i.e. $B_r = 1/19$) et n_1=n_2=10, 20, 30*, 40*, 50*, 60*, 80*, 100**

5.3.2 Exemple

Un exemple pratique permet d'exposer les propriétés du test $TFNCS$. Sur la base des hypothèses (5.2.1), un expérimentateur souhaite construire un plan expérimental autorisant un arrêt optionnel à la moitié des observations. L'arrêt en période 1 et le rejet de H_0 doivent reposer sur une même probabilité d'erreur (ou un même risque conditionnel) au plus égale à $\alpha_r^* = 0.05$, soit une évidence 19 fois plus forte en faveur de l'hypothèse H_1 qu'en faveur de H_0. Relativement au risque non-conditionnel prédéfini $\alpha = 0.05$, une puissance du test proche de 90% est souhaitée pour la valeur $\theta_1 = 0.35$.

La prise en considération du biais expérimental lié à l'arrêt optionnel de l'expérience est un prérequis méthodologique pour l'analyse. La conclusion sera fondée sur le rejet ou non de H_0 par la procédure fréquentiste non-conditionnelle. Cependant l'interprétation nécessitera également une évaluation quantitative en terme de rapport d'évidence des deux hypothèses et de probabilité d'erreur relative à la séquence observée (k, y).

L'approche du test $TFNCS$ avec prise en compte de la règle d'arrêt du plan expérimental est clairement indiquée dans l'énoncé. Une première étape consiste à calculer les bornes $(S_{1,L}, S_{2,L})$ associées à plusieurs valeurs de $n_1 = n_2$ (tableau (5.6)) pour

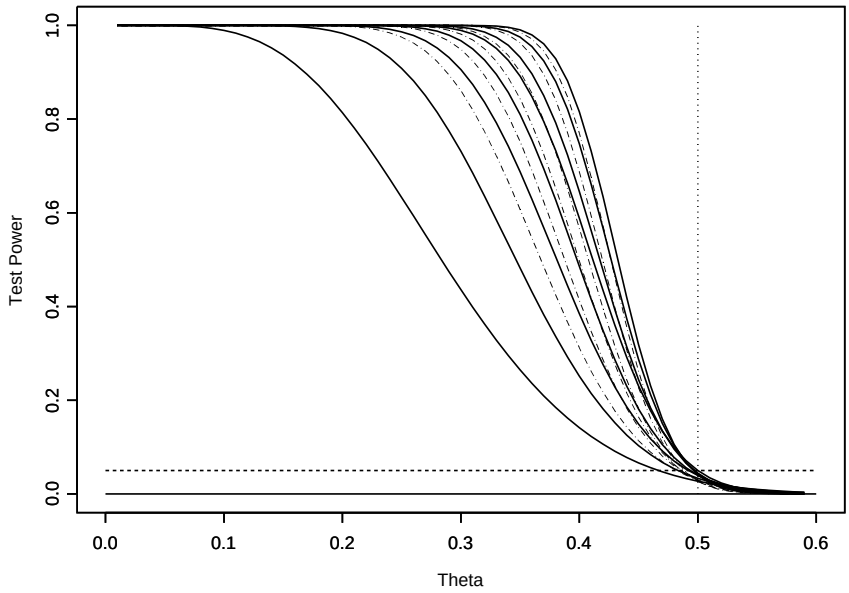

FIG. 5.5 – *Courbes de puissance du test $TFNCS$ avec les a priori de Jeffreys classique ($\cdot - \cdot$) et corrigé (—) pour $\alpha_r^* = 0.05$ et n_1=n_2=10, 20, 30*, 40*, 50*, 60*, 80*, 100**

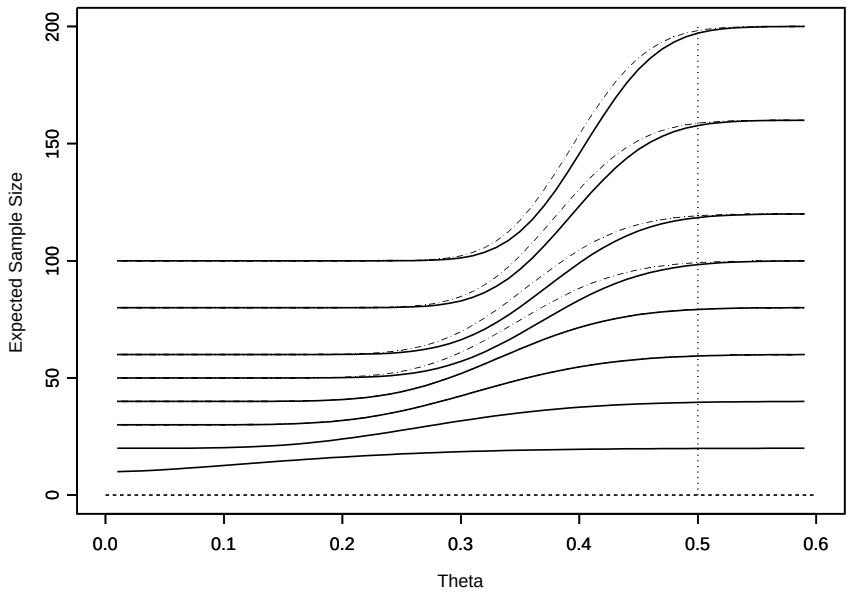

FIG. 5.6 – *Courbes des effectifs moyens du test $TFNCS$ avec les a priori de Jeffreys classique ($\cdot - \cdot$) et corrigé (—) pour $\alpha_r^* = 0.05$ et n_1=n_2=10, 20, 30, 40, 50*, 60*, 80*, 100**

ensuite produire les courbes de puissances. La figure (5.5) montre que l'ensemble des effectifs considérés $n_1 = n_2 = \{10, 20, 30, 40, 50, 60, 80, 100\}$ satisfont la condition de conservation du risque $\alpha = 0.05$. Une diminution de la valeur de α_r^* pour diminuer l'erreur de type 1 du test $TFNCS$ n'est donc pas nécessaire.

L'effectif $n_1 = n_2 = 30$ associé à la borne $S_{1,L} = 9$ fournit une puissance de 90.6% pour la valeur $\theta_1 = 0.35$, avec une erreur de type 1 égale à 0.039. L'effectif moyen correspondant est $E_{\theta_1=0.35}(N) = 42.3$. D'après le tableau (5.6), l'utilisation de l'a priori de Jeffreys corrigé réduit à $\mathcal{R}_{(2)} = [0, 28]$ la zone de rejet de H_0 du test $TFNCS$ en période 2 au lieu de $\mathcal{R}_{(2)} = [0, 29]$ avec l'a priori de Jeffreys classique. Les intervalles $\mathcal{R}_{(1)}$ et $\mathcal{R}_{(2)}$ correspondent également aux zones de rejet de H_0 dans la variante du test $TFCS^*$ avec une règle définie par la même limite $\alpha_r^* = 0.05$. L'utilisation de la variante du test $TFCS^*$ permet d'aboutir à une conclusion cohérente avec le test $TFNCS$.

Le tableau (5.7) montre l'information fournie par la procédure construite avec l'a priori de Jeffreys corrigé pour l'ensemble des séquences (k, y) possibles dans le plan $(n_1 = 30, S_{1,L} = 9, n_2 = 30)$. Le facteur de Bayes ajusté sur la règle d'arrêt $B_k(y|d_{Seq})$ quantifie le rapport des évidences de H_0 sur H_1. Excepté pour les séquences $(k, y) = \{(2, 29), (2, 30), (2, 31)\}$, les décisions sur H_0 du test $TFCS^*$ sont associées à des probabilités bayésiennes d'erreurs réinterprétables en terme de risques fréquentistes conditionnels. L'assertion aléatoire porte sur la séquence $(k, y) = (2, 29)$ où $r = 0.427$ est la composante aléatoire et $p = 1 - r = 0.573$ est la probabilité que H_0 soit rejetée avec un risque conditionnel associé $\alpha^*(B_{2,Seq}) = 0.354$. Les résultats du test $TFCS^*$ et de sa variante sont identiques pour $y_1 \in [0, S_{1,L} = 9]$ et $y = y_1 + y_2 \in [9, S_{2,L} = 22]$. Dans la variante du test $TFCS^*$, l'intervalle $]S_{2,L}, n_1 + n_2]$ devient zone de non-décision pour $y = y_1 + y_2$ car l'évidence était jugée trop faible pour pouvoir rejeter H_0.

Les conclusions du test $TFNCS$ et de la variante du test $TFCS^*$ sont identiques mais le risque non-conditionnel est fixé au niveau nominal $\alpha = 0.05$ alors que le

		Test $TFCS^*$			Test $TFNCS$	
(k,y)	$B_k(y\|d_{Seq})$	Décision*	Risque*	Random*	Décision	Risque
(1,0)	$< 10^{-9}$	H_0 rejetée	$< 10^{-9}$	_	H_0 rejetée	0.05
(1,1)	$< 10^{-7}$	H_0 rejetée	$< 10^{-7}$	_	H_0 rejetée	0.05
(1,2)	$< 10^{-6}$	H_0 rejetée	$< 10^{-6}$	_	H_0 rejetée	0.05
(1,3)	$< 10^{-5}$	H_0 rejetée	$< 10^{-5}$	_	H_0 rejetée	0.05
(1,4)	$< 10^{-4}$	H_0 rejetée	$< 10^{-4}$	_	H_0 rejetée	0.05
(1,5)	0.0002	H_0 rejetée	0.0003	_	H_0 rejetée	0.05
(1,6)	0.001	H_0 rejetée	0.001	_	H_0 rejetée	0.05
(1,7)	0.004	H_0 rejetée	0.004	_	H_0 rejetée	0.05
(1,8)	0.013	H_0 rejetée	0.013	_	H_0 rejetée	0.05
(1,9)	0.032	H_0 rejetée	0.031	_	H_0 rejetée	0.05
(2,9)	$< 10^{-7}$	H_0 rejetée	$< 10^{-7}$	_	H_0 rejetée	0.05
(2,10)	$< 10^{-6}$	H_0 rejetée	$< 10^{-6}$	_	H_0 rejetée	0.05
(2,11)	$< 10^{-6}$	H_0 rejetée	$< 10^{-6}$	_	H_0 rejetée	0.05
(2,12)	$< 10^{-5}$	H_0 rejetée	$< 10^{-5}$	_	H_0 rejetée	0.05
(2,13)	$< 10^{-4}$	H_0 rejetée	$< 10^{-4}$	_	H_0 rejetée	0.05
(2,14)	$< 10^{-4}$	H_0 rejetée	$< 10^{-4}$	_	H_0 rejetée	0.05
(2,15)	0.0001	H_0 rejetée	$< 10^{-3}$	_	H_0 rejetée	0.05
(2,16)	0.0004	H_0 rejetée	$< 10^{-3}$	_	H_0 rejetée	0.05
(2,17)	0.001	H_0 rejetée	0.0001	_	H_0 rejetée	0.05
(2,18)	0.002	H_0 rejetée	0.0003	_	H_0 rejetée	0.05
(2,19)	0.006	H_0 rejetée	0.006	_	H_0 rejetée	0.05
(2,20)	0.013	H_0 rejetée	0.013	_	H_0 rejetée	0.05
(2,21)	0.025	H_0 rejetée	0.024	_	H_0 rejetée	0.05
(2,22)	0.044	H_0 rejetée	0.042	_	H_0 rejetée	0.05
(2,23)	0.072	H_0 rejetée	0.068	_	_	_
(2,24)	0.113	H_0 rejetée	0.102	_	_	_
(2,25)	0.166	H_0 rejetée	0.142	_	_	_
(2,26)	0.234	H_0 rejetée	0.189	_	_	_
(2,27)	0.318	H_0 rejetée	0.241	_	_	_
(2,28)	0.421	H_0 rejetée	0.296	_	_	_
(2,29)	0.549	H_0 rejetée	0.354	0.427	_	_
(2,30)	0.711	Non-Décision	_	1	_	_
(2,31)	0.922	Non-Décision	_	1	_	_
(2,32)	1.21	H_0 rejetée	0.453	_	_	_
(2,33)	1.61	H_0 rejetée	0.383	_	_	_
(2,34)	2.21	H_0 rejetée	0.311	_	_	_
(2,35)	3.15	H_0 rejetée	0.241	_	_	_
(2,36)	4.72	H_0 acceptée	0.175	_	_	_
(2,37)	7.50	H_0 acceptée	0.118	_	_	_
(2,38)	12.7	H_0 acceptée	0.073	_	_	_
(2,39)	23.3	H_0 acceptée	0.041	_	_	_
(2,40)	45.9	H_0 acceptée	0.021	_	_	_
(2,41)	98.0	H_0 acceptée	0.010	_	_	_
(2,42)	227	H_0 acceptée	0.004	_	_	_
(2,43)	572	H_0 acceptée	0.002	_	_	_
(2,44)	1.7×10^{3}	H_0 acceptée	0.0006	_	_	_
(2,45)	4.9×10^{3}	H_0 acceptée	0.0002	_	_	_
(2,46)	1.6×10^{4}	H_0 acceptée	$< 10^{-4}$	_	_	_
(2,47)	5.5×10^{4}	H_0 acceptée	$< 10^{-4}$	_	_	_
(2,48)	2.1×10^{5}	H_0 acceptée	$< 10^{-5}$	_	_	_
(2,49)	9.2×10^{5}	H_0 acceptée	$< 10^{-5}$	_	_	_
(2,50)	4.4×10^{6}	H_0 acceptée	$< 10^{-6}$	_	_	_
(2,51)	2.3×10^{7}	H_0 acceptée	$< 10^{-7}$	_	_	_
(2,52)	1.4×10^{8}	H_0 acceptée	$< 10^{-8}$	_	_	_
(2,53)	1.0×10^{9}	H_0 acceptée	$< 10^{-9}$	_	_	_
(2,54)	8.5×10^{9}	H_0 acceptée	$< 10^{-9}$	_	_	_
(2,55)	8.5×10^{10}	H_0 acceptée	$< 10^{-10}$	_	_	_
(2,56)	1.0×10^{12}	H_0 acceptée	$< 10^{-12}$	_	_	_
(2,57)	1.7×10^{13}	H_0 acceptée	$< 10^{-13}$	_	_	_
(2,58)	4.0×10^{14}	H_0 acceptée	$< 10^{-14}$	_	_	_
(2,59)	1.6×10^{16}	H_0 acceptée	$< 10^{-16}$	_	_	_
(2,60)	1.9×10^{18}	H_0 acceptée	$< 10^{-18}$	_	_	_

TAB. 5.7 – *Décisions et risques associés au test $TFCS^*$ et au test $TFNCS$ avec l'a priori de Jeffreys corrigé pour $\alpha_r^* = 0.05$ et $n_1 = n_2 = 30$*

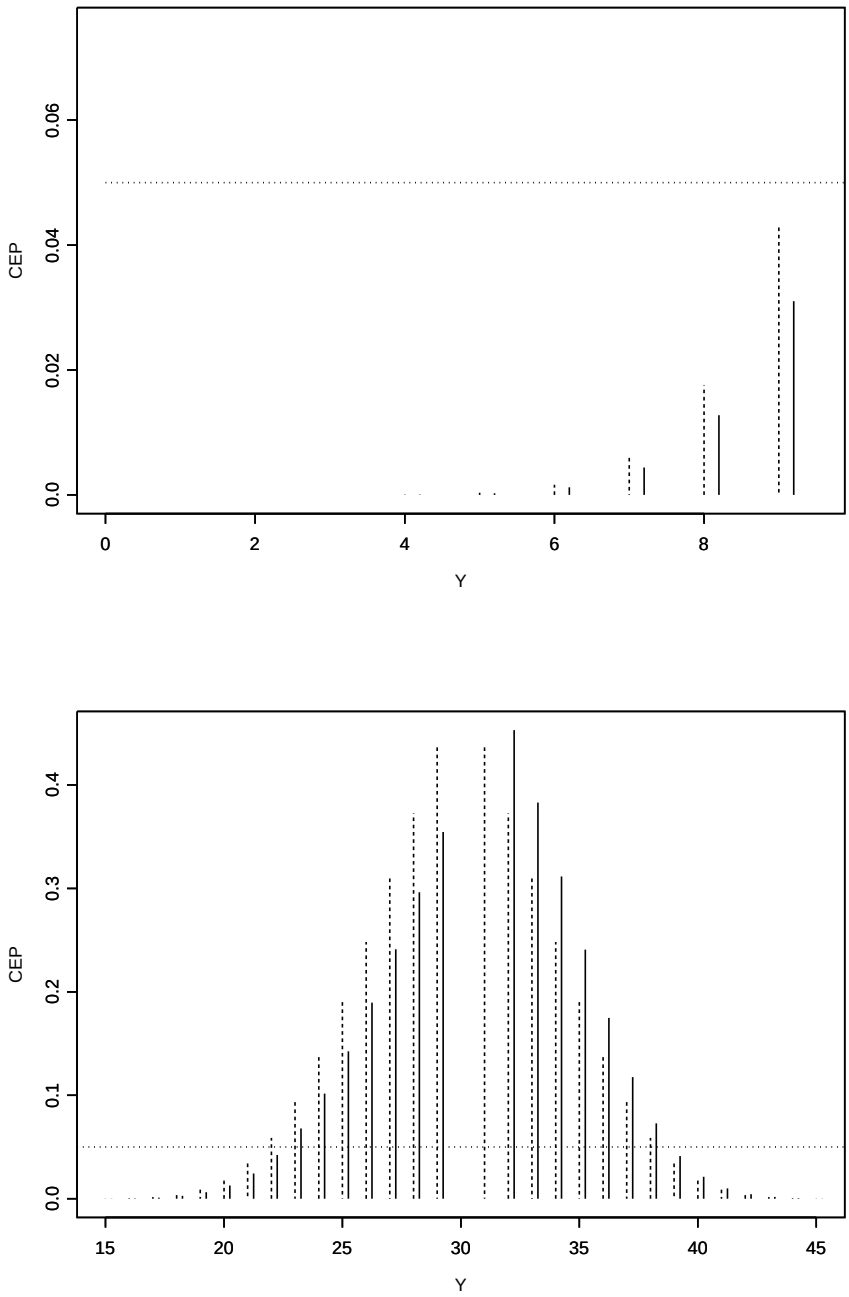

FIG. 5.7 – *Risques fréquentistes conditionnels associés au test $TFCS^*$ avec les a priori de Jeffreys classique (- - -) et corrigé (—) pour $\alpha_r^* = 0.05$ et $n_1 = n_2 = 30$*

risque conditionnel dépend de l'évidence fournie par la séquence (k, y) en faveur de l'une ou l'autre des hypothèses.

La figure (5.7) permet de visualiser les risques conditionnels du test $TFCS^*$ associés aux a priori de Jeffreys corrigé (tableau (5.7)) et classique (tableau (C.1)). Les risques conditionnels avec l'a priori de Jeffreys classique sont répartis symétriquement autour de la séquence centrale $(k, y) = (2, 30)$ en période 2. La correction provoque un décalage de la répartition des risques vers les valeurs supérieures de y, ainsi que l'introduction de l'assertion aléatoire décrite précédemment pour la séquence $(k, y) = (2, 29)$. Comme la probabilité $p = 0.573$ associée au rejet de H_0 est supérieure à 0.5 (tableau (5.7)), la conclusion est confirmée pour cette séquence dans la figure (5.7).

5.4 Application à une problématique d'équivalence

5.4.1 Détermination des tests $TFCS^*$ et $TFNCS$

Après avoir conclu dans la première expérience que l'hypothèse $H_0 : \{\theta \geq 0.5\}$ ne pouvait être tenue pour vraie, l'expérimentateur souhaite montrer que la valeur du paramètre θ est comprise entre 0.2 et 0.4. Une règle de rejet de H_0 fondée sur une valeur limite $\alpha_r^* \leq 0.10$ (i.e. $B_{k,Seq}(y) \leq 1/9$) dans la variante du test $TFCS^*$ est souhaitée afin de garantir un risque $\alpha = 0.10$ du test $TFNCS$. En anticipant l'égalité des effectifs $n_1 = n_2$, le plan expérimental devra également assurer une puissance proche de 90% pour $\theta_1 = 0.30$.

Les hypothèses associées à cette problématique sont des hypothèses d'équivalence définies par

$$H_0 : \{\theta \in [0, 0.2] \cup [0.4, 1]\} \qquad \text{versus} \qquad H_1 : \{\theta \in]0.2, 0.4[\}. \tag{5.4.1}$$

La démarche du test $TFNCS$ nécessite d'abord de déterminer une valeur limite α_r^* maximum pour laquelle le niveau nominal $\alpha = 0.10$ est respecté pour les effectifs

	Jeffreys corrigé				Jeffreys classique			
	$S_{1,L}$	$S_{1,U}$	$S_{2,L}$	$S_{2,U}$	$S_{1,L}$	$S_{1,U}$	$S_{2,L}$	$S_{2,U}$
$n_1 = n_2 = 30$	-1	31	16	20	_	_	_	_
$n_1 = n_2 = 40$	12	12	21	28	_	_	_	_
$n_1 = n_2 = 50$	14	16	26	35	14	16	25	35
$n_1 = n_2 = 60$	16	20	30	43	_	_	_	_
$n_1 = n_2 = 80$	21	28	39	58	_	_	_	_
$n_1 = n_2 = 100$	26	35	48	73	25	35	48	73
$n_1 = n_2 = 120$	30	43	57	89	_	_	_	_
$n_1 = n_2 = 150$	37	54	70	112	_	_	_	_

TAB. 5.8 – *Bornes ($S_{1,L}$,$S_{1,U}$,$S_{2,L}$,$S_{2,U}$) avec les a priori de Jeffreys classique et corrigé pour $\alpha_r^* = 0.10$ (i.e. $B_r = 1/9$) et $n_1 = n_2$*=30, 40, 50*, 60, 80, 100*, 120, 150

	Jeffreys classique et corrigé			
	$S_{1,L}$	$S_{1,U}$	$S_{2,L}$	$S_{2,U}$
$n_1 = n_2 = 60$	-1	61	32	40
$n_1 = n_2 = 80$	23	25	41	55
$n_1 = n_2 = 100$	28	33	51	70
$n_1 = n_2 = 120$	32	40	60	85
$n_1 = n_2 = 150$	39	51	73	107

TAB. 5.9 – *Bornes ($S_{1,L}$,$S_{1,U}$,$S_{2,L}$,$S_{2,U}$) avec les a priori de Jeffreys classique et corrigé pour $\alpha_r^* = 0.05$ (i.e. $B_r = 1/19$) et $n_1 = n_2$*=60, 80, 100, 120, 150

envisagés. On en déduit les zones de rejet de H_0 en période 1 et 2 qui permettront de calculer l'erreur de type 1 (i.e. $\max_{\theta=0.2,0.4}(\Pr(\text{Rejet } H_0|\theta))$).

L'a priori de Jeffreys corrigé dans le plan $d_{Seq} = \{n_1, J, n_2\}$ s'écrit

$$\theta|d_{Seq} \sim w_a Be^J_{0,0.2}(0.5, 0.5, 0.5; n_1, r; J) + w_b Be^J_{0.4,0.6}(0.5, 0.5, 0.5; n_1, r; J),$$

où $w_a + w_b = 1$ et $Be^J_{d,e}(0.5, 0.5, 0.5; n_1, r; J)$ désigne la loi Bêta-J décentrée avec un paramètre de décentrage d et une étendue e (voir section (2.5)). On suppose dans un premier temps que les poids w_a et w_b sont égaux (i.e. $w_a = w_b = 0.5$).

La bilatéralité des hypothèses se traduit par des zones de rejet de H_0 bornées par valeurs inférieures et supérieures du type $\mathcal{R}_{(1)}$=$\mathcal{R}^*_{(1)}$=$\bar{J}$=[$S_{1,L}$,$S_{1,U}$] et $\mathcal{R}_{(2)}$=$\mathcal{R}^*_{(2)}$= [$S_{2,L}$,$S_{2,U}$] (le rôle de $S_{k,L}$ et $S_{k,U}$ est inversé par rapport à la section précédente).

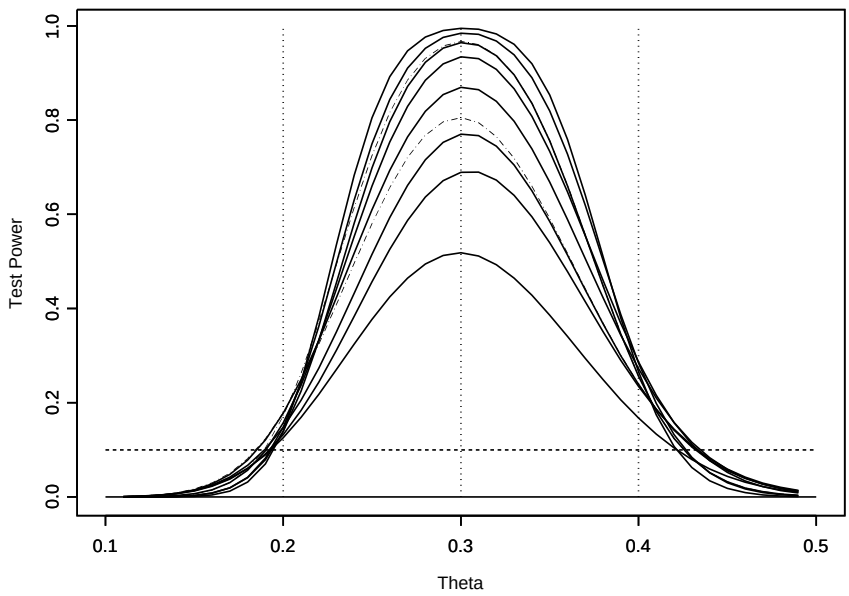

FIG. 5.8 – *Courbes de puissance du test $TFNCS$ avec les a priori de Jeffreys classique ($\cdot - \cdot$) et corrigé (—) pour $\alpha_r^* = 0.10$ et $n_1 = n_2$=30, 40, 50*, 60, 80, 100*, 120, 150*

Les valeurs des bornes $S_{1,L}$, $S_{1,U}$, $S_{2,L}$ et $S_{2,U}$ sont données dans le tableau (5.8) pour $\alpha_r^* = 0.10$ (i.e. $B_r = 1/9$) et les effectifs $n_1 = n_2 = 30, 40, 50, 60, 80, 100, 120, 150$.

La figure (5.8) montre les courbes de puissance correspondantes. Pour l'effectif $n_1 = n_2 = 30$, aucune séquence (k, y) ne présente une valeur de $\alpha(B_{1,Seq})$ inférieure à $\alpha_r^* = 0.10$ en période 1 (i.e. $\mathcal{R}_{(1)} = \varnothing$). Le plan expérimental associé est un échantillonnage fixe (n=60).

Aucun effectif ne garantissant le risque nominal $\alpha = 0.10$ avec la limite $\alpha_r^* = 0.10$, une nouvelle valeur $\alpha_r^* = 0.05$ (i.e. $B_r = 1/19$) est étudiée. Les bornes associées $S_{1,L}$, $S_{1,U}$, $S_{2,L}$ et $S_{2,U}$ sont données dans le tableau (5.9). Les courbes de puissance et les courbes des effectifs moyens sont respectivement décrites dans les figures (5.9) et (5.10) pour les effectifs $n_1 = n_2 = 60, 80, 100, 120, 150$. L'effectif $n_1 = n_2 = 60$ génère une zone de rejet de H_0 vide en période 1 (i.e. $\mathcal{R}_{(1)} = \varnothing$), qui se traduit par une courbe constante de l'effectif moyen ($n = 120$) dans la figure (5.10).

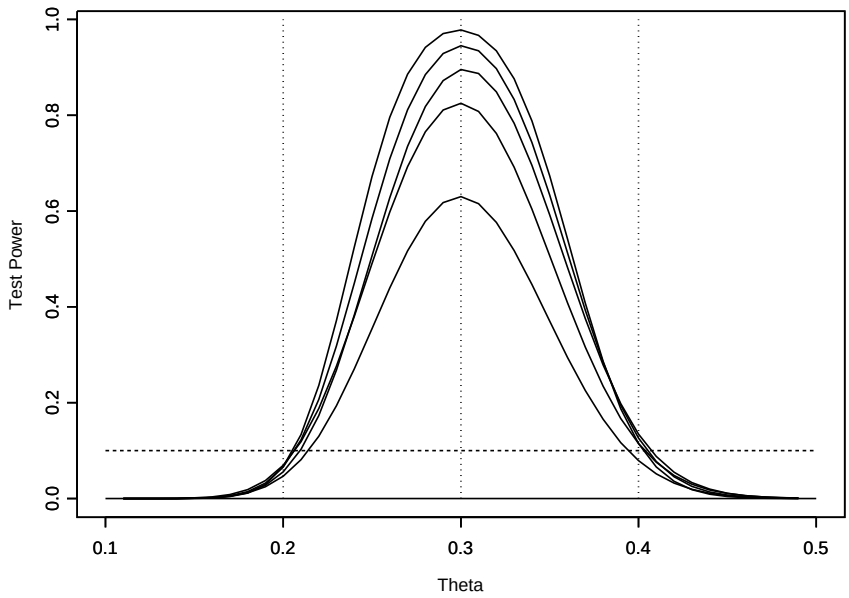

FIG. 5.9 – *Courbes de puissance du test* $TFNCS$ *avec les a priori de Jeffreys classique et corrigé (—) pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2$=60, 80, 100, 120, 150

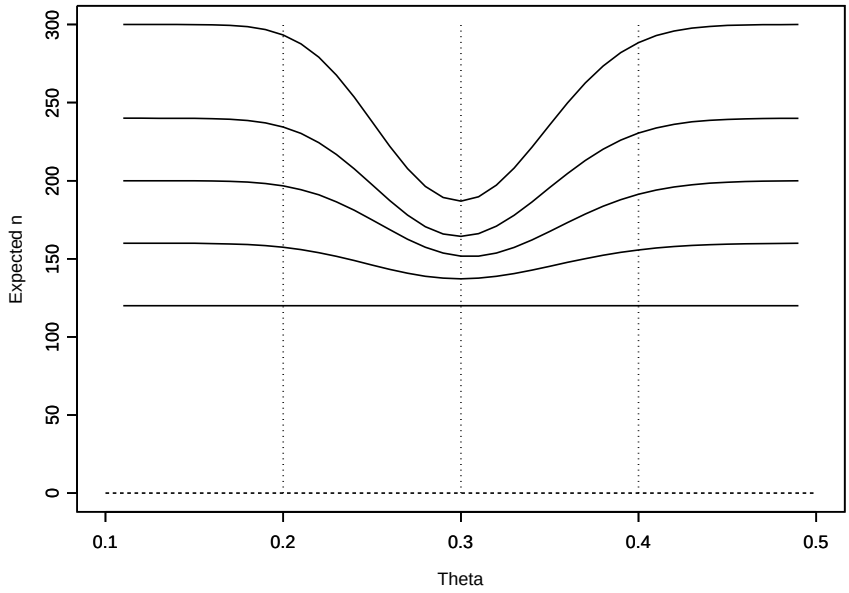

FIG. 5.10 – *Courbes des effectifs moyens du test* $TFNCS$ *avec les a priori de Jeffreys classique et corrigé (—) pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2$=60, 80, 100, 120, 150

L'assymétrie des hypothèses conjuguée à une pondération $(w_a, w_b) = (0.5, 0.5)$ de l'a priori de Jeffreys entre les intervalles $H_0^a = [0, 0.2]$ et $H_0^b = [0.4, 1]$ se traduit par des probabilités de rejet de H_0 inégale aux bornes $\theta = 0.2, 0.4$ de l'espace Θ_0. En particulier, on a $Pr(\text{Rejet } H_0|\theta = 0.2) < 0.10$ et $Pr(\text{Rejet } H_0|\theta = 0.4) > 0.10$ pour les effectifs étudiés (excepté pour $n_1 = n_2 = 60$ associé au plan non-séquentiel). On note d'autre part que l'objectif fixé en terme de puissance est atteint avec $Pow(\theta_1 = 0.3) = 0.895$ pour $n_1 = n_2 = 100$.

5.4.2 Test symétrique aux bornes ($TFNCS^{SB}$)

L'idée développée dans la suite est de déterminer une pondération (w_a, w_b), telle que $w_a + w_b = 1$, qui rapproche au maximum les probabilités de rejeter H_0 aux deux bornes $\theta_0 = 0.2, 0.4$. Le principe de symétrie aux bornes est l'élément majeur de cette stratégie.

Principe de symétrie aux bornes (PSB). *Dans le cas d'une hypothèse nulle composite où l'espace Θ_0 est délimité par 2 bornes (i.e. $H_0 : \{\theta \in [0, \theta_0^a] \cup [\theta_0^b, 1]\}$), le test TFNCS associé à la pondération $(\tilde{w}_a, \tilde{w}_b)$ est dit symétrique aux bornes ($TFNCS^{SB}$) si il n'existe pas une autre pondération (w_a, w_b) telle que*

$$\begin{aligned}&|Pr(Rejet\ H_0|\theta = \theta_0^a; w_a, w_b) - Pr(Rejet\ H_0|\theta = \theta_0^b; w_a, w_b)| \\ &\qquad < |Pr(Rejet\ H_0|\theta = \theta_0^a; \tilde{w}_a, \tilde{w}_b) - Pr(Rejet\ H_0|\theta = \theta_0^b; \tilde{w}_a, \tilde{w}_b)|.\end{aligned}$$

Le PSB met en avant l'importance égale des probabilités de rejeter H_0 aux deux bornes $\{\theta_0^a, \theta_0^b\}$ de Θ_0 pour le calcul de l'erreur de type 1 dans le test $TFNCS^{SB}$. La détermination de la pondération (w_a, w_b) nécessite une approche récursive. Pour les hypothèses (5.4.1) et utilisant l'a priori de Jeffreys corrigé, on aboutit aux valeurs limites $(w_a, w_b) = (0.35, 0.65)$ avec $Pr(\text{Rejet } H_0|\theta = 0.2) < Pr(\text{Rejet } H_0|\theta = 0.4)$ et $(w_a, w_b) = (0.36, 0.64)$ avec $Pr(\text{Rejet } H_0|\theta = 0.2) > Pr(\text{Rejet } H_0|\theta = 0.4)$. Les bornes $S_{1,L}$, $S_{1,U}$, $S_{2,L}$ et $S_{2,U}$ pour ces pondérations sont données dans le tableau (5.10). Les probabilités de rejeter H_0 pour les valeurs $\theta = 0.2, 0.3$ et 0.4 se trouvent

	Jeffreys corrigé				Jeffreys classique			
	$S_{1,L}$	$S_{1,U}$	$S_{2,L}$	$S_{2,U}$	$S_{1,L}$	$S_{1,U}$	$S_{2,L}$	$S_{2,U}$
$w_a = 0.35; w_a = 0.65$	27	32	49	68	27	32	49	71
$w_a = 0.36; w_a = 0.64$	27	32	50	69	27	32	49	71
$w_a = 0.5; w_a = 0.5$	28	33	51	70	_	_	_	_

TAB. 5.10 – *Bornes* $(S_{1,L}, S_{1,U}, S_{2,L}, S_{2,U})$ *pour* $\alpha_r^* = 0.05$ *(i.e.* $B_r = 1/19$*) et* $n_1 = n_2 = 100$

	Pr(Rejet $H_0 \mid \theta$)		
	$\theta = 0.2$	$\theta = 0.3$	$\theta = 0.4$
$w_a = 0.35; w_a = 0.65$	0.098	0.890	0.087
$w_a = 0.36; w_a = 0.64$	0.083	0.896	0.099
$w_a = 0.5; w_a = 0.5$	0.056	0.895	0.135

TAB. 5.11 – *Probabilités de rejeter* H_0 *avec l'a priori de Jeffreys corrigé pour* $\alpha_r^* = 0.05$ *(i.e.* $B_r = 1/19$*) et* $n_1 = n_2 = 100$

dans le tableau (5.11).

Par rapport à $(w_a, w_b) = (0.5, 0.5)$, ces deux pondérations diminuent les probabilités de rejeter H_0 et font revenir les erreurs de type 1 en deçà du niveau nominal $\alpha = 0.10$. D'après le PSB, le test $TFNCS^{SB}$ est associé à la pondération $(\tilde{w}_a, \tilde{w}_b) = (0.35, 0.65)$ car elle procure l'écart minimum

$$|Pr(\text{Rejet } H_0 | \theta = 0.2; \tilde{w}_a, \tilde{w}_b) - Pr(\text{Rejet } H_0 | \theta = 0.4; \tilde{w}_a, \tilde{w}_b)| = 0.011.$$

L'erreur de type 1 est $max_{\theta=0.2,0.4}(Pr(\text{Rejet } H_0 | \theta; \tilde{w}_a, \tilde{w}_b)) = 0.098$ associée à une puissance $Pow(\theta_1 = 0.3) = 0.89$ et un effectif moyen $E_{\theta_1=0.3}(N) = 151.4$. Les figures (5.11) et (5.12) permettent de mesurer l'effet de la pondération en superposant respectivement les courbes de puissance et des effectifs moyens des tests $TFNCS^{SB}$ et $TFNCS$ avec $(w_a, w_b) = (0.5, 0.5)$, associés à l'a priori de Jeffreys corrigé.

Pour chaque séquence (k, y) de la région de rejet $\mathcal{R}_{(k)}$ de H_0 en période k du test $TFNCS^{SB}$, le tableau (5.12) fournit également les valeurs de $B_k(y|d_{Seq})$ et $\alpha^*(B_{k,Seq})$ associées à la variante du test $TFCS^*$ correspondant. La présentation des valeurs de $B_k(y)$ et $\alpha^*(B_k)$ obtenues avec l'a priori de Jeffreys classique permet

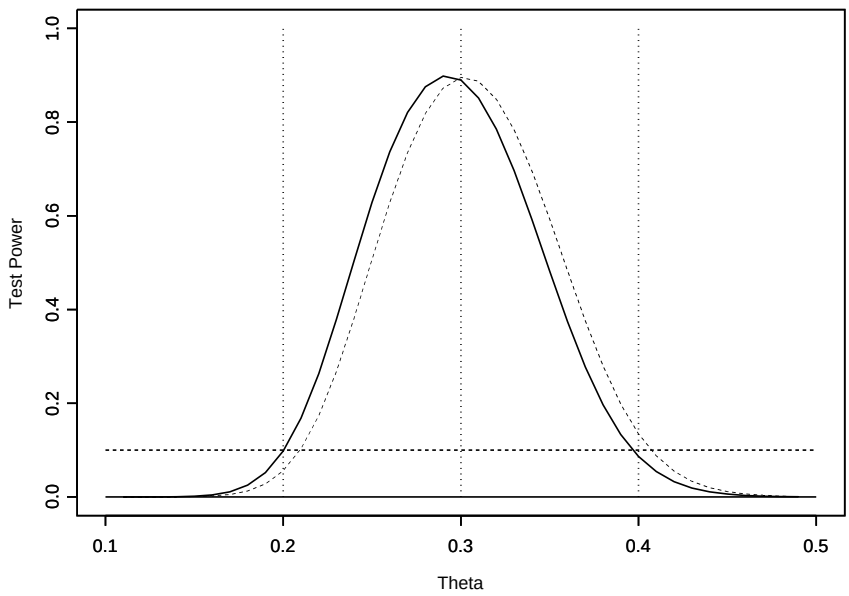

FIG. 5.11 – *Courbes de puissance du test* $TFNCS^{SB}$ *(—) et du test* $TFNCS$ *avec la pondération* $(w_a, w_b) = (0.5, 0.5)$ *de l'a priori de Jeffreys (- - -) pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2 = 100$

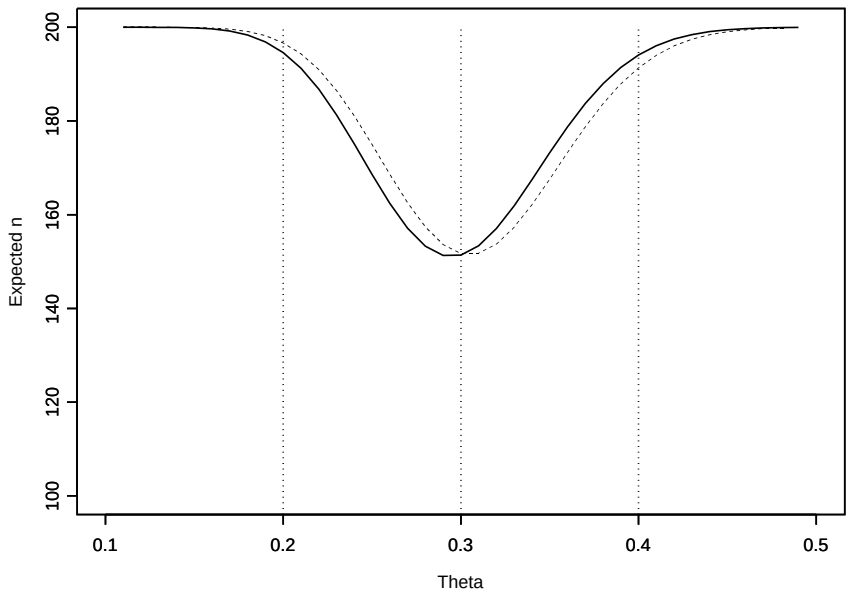

FIG. 5.12 – *Courbes des effectifs moyens du test* $TFNCS^{SB}$ *(—) et du test* $TFNCS$ *avec la pondération* $(w_a, w_b) = (0.5, 0.5)$ *de l'a priori de Jeffreys (- - -) pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2 = 100$

	Jeffreys corrigé			Jeffreys classique		
(k, y)	$B_k(y\|d_{Seq})$	Risque*	Risque	$B_k(y)$	Risque*	Risque
(1,27)	3.99×10^{-2}	3.84×10^{-2}	0.05	3.89×10^{-2}	3.75×10^{-2}	0.05
(1,28)	2.96×10^{-2}	2.88×10^{-2}	0.05	2.90×10^{-2}	2.81×10^{-2}	0.05
(1,29)	2.55×10^{-2}	2.49×10^{-2}	0.05	2.50×10^{-2}	2.44×10^{-2}	0.05
(1,30)	2.72×10^{-2}	2.65×10^{-2}	0.05	2.68×10^{-2}	2.61×10^{-2}	0.05
(1,31)	3.46×10^{-2}	3.34×10^{-2}	0.05	3.41×10^{-2}	3.30×10^{-2}	0.05
(1,32)	4.73×10^{-2}	4.52×10^{-2}	0.05	4.68×10^{-2}	4.47×10^{-2}	0.05
(2,49)	5.16×10^{-2}	4.90×10^{-2}	0.05	5.00×10^{-2}	4.76×10^{-2}	0.05
(2,50)	3.88×10^{-2}	3.73×10^{-2}	0.05	3.75×10^{-2}	3.61×10^{-2}	0.05
(2,51)	2.84×10^{-2}	2.76×10^{-2}	0.05	2.74×10^{-2}	2.67×10^{-2}	0.05
(2,52)	2.03×10^{-2}	1.99×10^{-2}	0.05	1.96×10^{-2}	1.92×10^{-2}	0.05
(2,53)	1.42×10^{-2}	1.40×10^{-2}	0.05	1.36×10^{-2}	1.34×10^{-2}	0.05
(2,54)	9.62×10^{-3}	9.53×10^{-2}	0.05	9.24×10^{-3}	9.16×10^{-3}	0.05
(2,55)	6.42×10^{-3}	6.37×10^{-3}	0.05	6.12×10^{-3}	6.09×10^{-3}	0.05
(2,56)	4.27×10^{-3}	4.25×10^{-3}	0.05	4.00×10^{-3}	3.99×10^{-3}	0.05
(2,57)	2.95×10^{-3}	2.94×10^{-3}	0.05	2.65×10^{-3}	2.64×10^{-3}	0.05
(2,58)	2.38×10^{-3}	2.37×10^{-3}	0.05	1.86×10^{-3}	1.85×10^{-3}	0.05
(2,59)	2.37×10^{-3}	2.36×10^{-3}	0.05	1.62×10^{-3}	1.62×10^{-3}	0.05
(2,60)	2.90×10^{-3}	2.89×10^{-3}	0.05	1.72×10^{-3}	1.72×10^{-3}	0.05
(2,61)	4.03×10^{-3}	4.02×10^{-3}	0.05	2.25×10^{-3}	2.24×10^{-3}	0.05
(2,62)	5.87×10^{-3}	5.83×10^{-3}	0.05	3.18×10^{-3}	3.17×10^{-3}	0.05
(2,63)	8.54×10^{-3}	8.47×10^{-3}	0.05	4.58×10^{-3}	4.56×10^{-3}	0.05
(2,64)	1.22×10^{-2}	1.21×10^{-2}	0.05	6.55×10^{-3}	6.51×10^{-2}	0.05
(2,65)	1.71×10^{-2}	1.69×10^{-2}	0.05	9.19×10^{-3}	9.10×10^{-3}	0.05
(2,66)	2.35×10^{-2}	2.30×10^{-2}	0.05	1.26×10^{-2}	1.24×10^{-2}	0.05
(2,67)	3.15×10^{-2}	3.05×10^{-2}	0.05	1.69×10^{-2}	1.66×10^{-2}	0.05
(2,68)	4.13×10^{-2}	3.97×10^{-2}	0.05	2.22×10^{-2}	2.17×10^{-2}	0.05
(2,69)	_	_	_	2.86×10^{-2}	2.78×10^{-2}	0.05
(2,70)	_	_	_	3.61×10^{-2}	3.49×10^{-2}	0.05
(2,71)	_	_	_	4.49×10^{-2}	4.30×10^{-2}	0.05

TAB. 5.12 – *Facteurs de Bayes et risques associés au test* $TFNCS^{SB}$ *et à la variante correspondante du test* $TFCS^*$ *pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2 = 100$

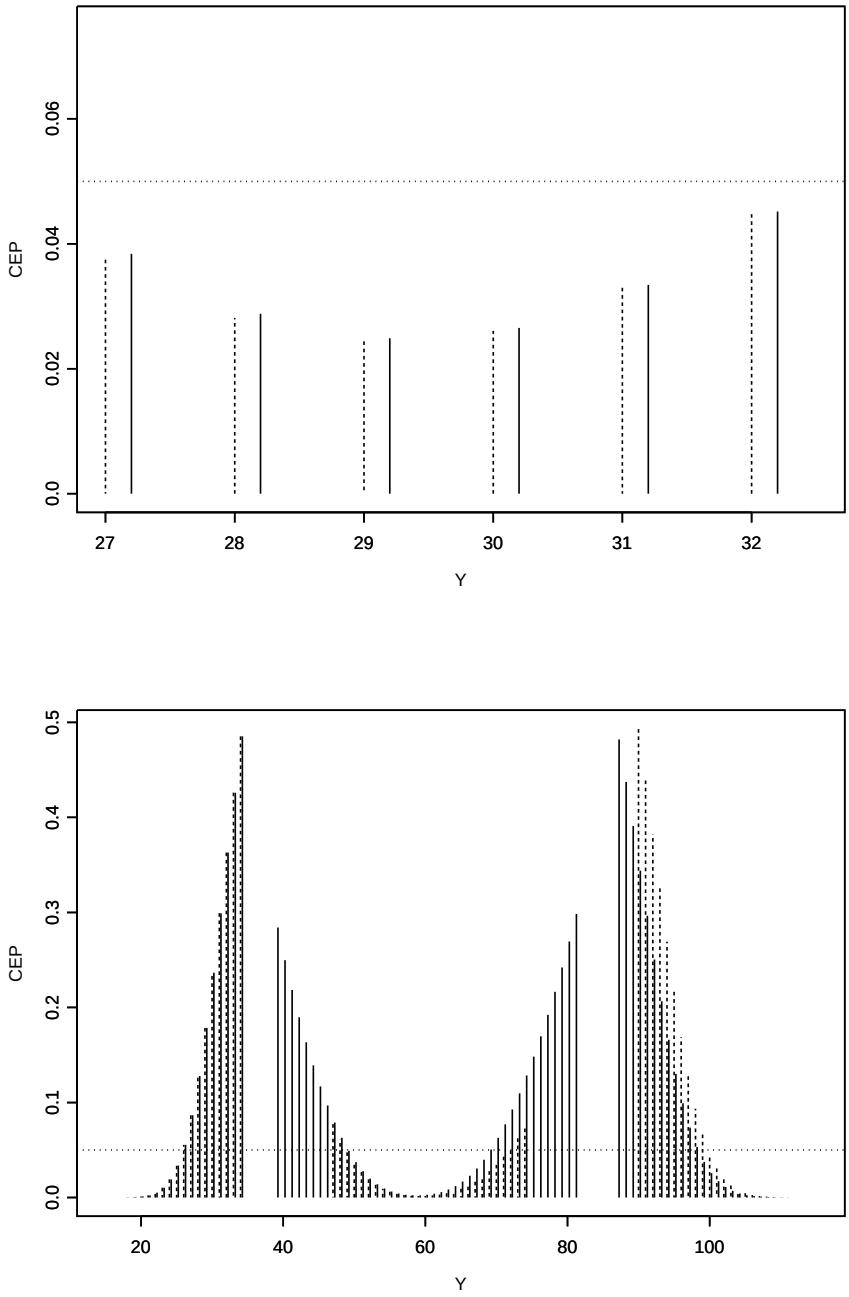

FIG. 5.13 – *Risques fréquentistes conditionnels du test* $TFCS^*$ *associé au test* $TFNCS^{SB}$ *obtenus avec les a priori de Jeffreys classique (- - -) et corrigé (—) pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2 = 100$

de mesurer l'impact de la correction dans le contexte des hypothèses (5.4.1).

Dans le plan expérimental $d_{Seq} = \{n_1, J, n_2\}$ associé aux hypothèses d'équivalence (5.4.1), l'utilisation de l'a priori de Jeffreys corrigé par rapport à l'a priori de Jeffreys classique induit une augmentation du facteur de Bayes $B_k(y|d_{Seq}) > B_k(y)$. Ce réajustement du rapport des évidences en faveur de H_0 vient compenser les valeurs du facteur de Bayes, qui sont en moyenne inférieures lors de l'arrêt optionnel en période 1 à celles obtenues en période 2 si l'expérience s'était poursuivie. L'effet de la correction sur le test $TFCS^*$ et sa variante se traduit par une augmentation des risques conditionnels de rejeter H_0 (et une diminution des risques conditionnels du test $TFCS^*$ de l'accepter). De fait la zone de rejet de H_0 du test $TFNCS^{SB}$ et de la variante du test $TFCS^*$ correspondant diminue en période 2 de $\mathcal{R}_{(2)} = \mathcal{R}^*_{(2)} = [49, 71]$ à $\mathcal{R}_{(2)} = \mathcal{R}^*_{(2)} = [49, 68]$.

La figure (5.13) décrit les risques conditionnels du test $TFCS^*$ correspondants au test $TFCS^{SB}$. La zone de non-décision subit un décalage vers les valeurs supérieures de y ainsi qu'un accroissement important en période 2. Le décalage est un effet direct de la correction de l'a priori de Jeffreys alors que l'accroissement est dû à un changement de condition dans la règle (B.5.3) pour déterminer la zone de non-décision.

5.4.3 Etude de l'erreur de type 1 du test $TFNCS^{SB}$ en fonction du décentrage de Θ_0

Cette dernière section a pour objectif d'étudier le comportement du test $TFNCS^{SB}$ en fonction du décentrage dec du sous-espace $\Theta_0 = [0, dec] \cup [dec + 0.2, 1]$ dans [0,1] pour la limite $\alpha^*_r = 0.05$ et l'effectif $n_1 = n_2 = 100$. La pondération (w_a, w_b) de l'a priori de Jeffreys corrigé assure un centrage de la fonction puissance et des valeurs $p_d = Pr(\text{Rejet } H_0|\theta = dec)$ et $p_{dec+0.2} = Pr(\text{Rejet } H_0|\theta = dec + 0.2)$ telles que $|p_{dec} - p_{dec+0.2}|$ soit minimum.

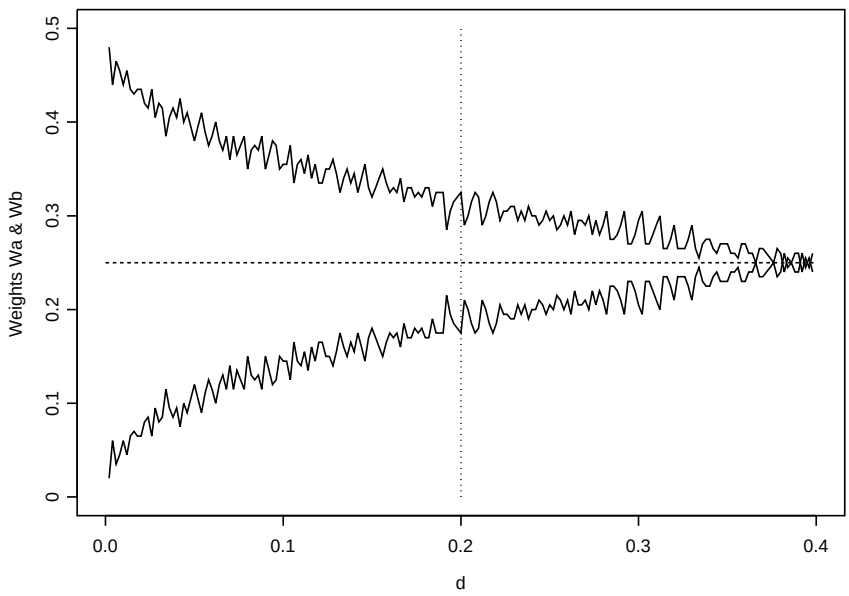

FIG. 5.14 – *Pondérations* $(\tilde{w}_a, \tilde{w}_b)$ *du test* $TFNC^{SB}$ *en fonction du paramètre de décentrage dec dans l'hypothèse* $H_0 : \{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ *pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2 = 100$

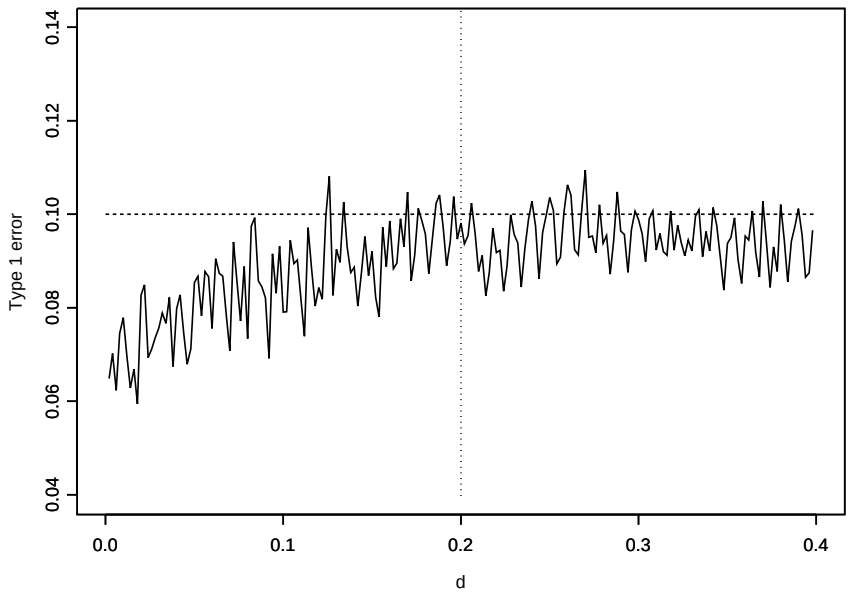

FIG. 5.15 – *Erreurs de type1 du test* $TFNC^{SB}$ *en fonction du paramètre de décentrage dec dans l'hypothèse* $H_0 : \{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ *pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2 = 100$

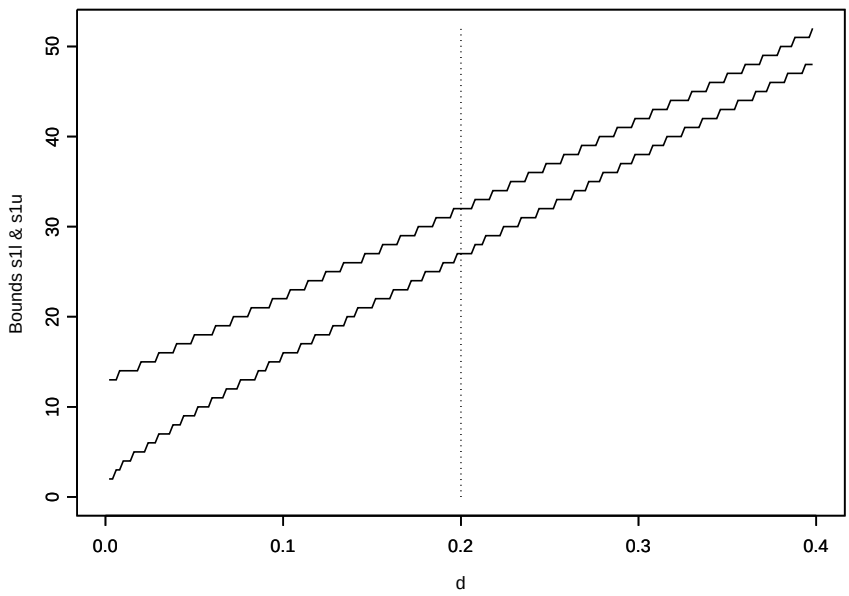

FIG. 5.16 – *Bornes* $(S_{1,L}, S_{1,U})$ *de la région de rejet de* H_0 *du test* $TFNC^{SB}$ *en fonction du paramètre de décentrage dec dans l'hypothèse* $H_0 : \{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ *pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2 = 100$

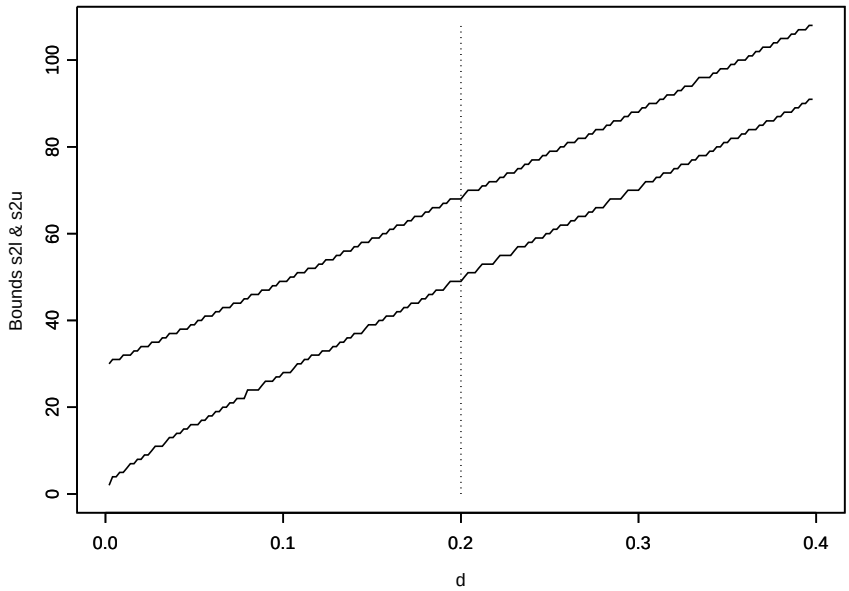

FIG. 5.17 – *Bornes* $(S_{2,L}, S_{2,U})$ *de la région de rejet de* H_0 *du test* $TFNC^{SB}$ *en fonction du paramètre de décentrage dec dans l'hypothèse* $H_0 : \{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ *pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2 = 100$

Les valeurs de p_{dec} et $p_{dec+0.2}$ en fonction de θ telles que $p_{dec} > p_{dec+0.2}$ et $p_{dec} < p_{dec+0.2}$ sont respectivement montrées dans les figures (C.46) et (C.47) pour dec compris entre 0 et 0.4. Les valeurs de p_{dec} et $p_{dec+0.2}$ pour $dec \in [0.4, 0.8]$ se déduisent par symétrie. Les pondération (w_a, w_b) correspondantes sont décrites dans la figure (C.48) et on trouve les bornes associées $(S_{1,L}, S_{1,U})$ en période 1 et $(S_{2,L}, S_{2,U})$ en période 2 respectivement dans les figures (C.49) et (C.50).

Pour chaque valeur de dec, le choix de la pondération du test $TFNCS^{SB}$ se porte sur la valeur $(\tilde{w}_a, \tilde{w}_b)$=$(w_a, w_b)$ qui correspond à la plus petite différence $|p_{dec} - p_{dec+0.2}|$ lorsque $p_{dec} > p_{dec+0.2}$ ou $p_{dec} < p_{dec+0.2}$ (voir figure (5.14)). L'erreur de type 1 du test $TFNCS^{SB}$ est ensuite le maximum des deux valeurs p_{dec} et $p_{dec+0.2}$ retenues. La figure (5.14) montre qu'il existe peu de valeurs de dec pour lesquelles cette erreur dépasse le niveau nominal $\alpha = 0.10$. Les bornes correspondantes $(S_{1,L}, S_{1,U})$ en période 1 et $(S_{2,L}, S_{2,U})$ en période 2 sont respectivement décrites dans les figures (5.16) et (5.17).

5.5 Conclusion

La procédure séquentielle décrite dans cette section repose sur une règle d'arrêt et de rejet de H_0 fondée sur une valeur limite unique du rapport des évidences des deux hypothèses testées. On utilise simultanément un test fréquentiste non-conditionnel (test $TFNCS$) et un test fréquentiste conditionnel (variante du test $TFCS^*$) admettant la même règle de rejet de H_0. Dans le test $TFNCS$, le risque d'erreur est fixé à un niveau nominal α. Dans le test $TFNS^*$, les erreurs dépendent de la séquence (k, y) et s'interprètent à la fois en terme de probabilité bayésienne sur l'hypothèse et de risque fréquentiste conditionné à une mesure de l'évidence.

Cette dernière application de l'a priori de Jeffreys corrigé montre le mécanisme de compensation des effets de la séquencialité sur le rapport des évidences. La correction réajuste le facteur de Bayes dans le sens de sa valeur moyenne si le plan expérimental

ne comportait pas d'arrêt optionnel. Elle modifie de fait les risques conditionnels du test $TFCS^*$ et la zone de rejet de H_0 commune aux deux tests. Suivant le sens du réajustement du facteur de Bayes, la correction est conservatrice ou libérale pour le test $TFNCS$.

Dans le cas où le sous-espace Θ_0 sous H_0 est un intervalle non disjoint, le conditionnement de l'a priori de Jeffreys sur les paramètres du plan expérimental garantit l'unicité des solutions. Pour un test d'équivalence avec le sous-espace disjoint $\Theta_0 = \Theta_0^a \cup \Theta_0^b = [0, \theta_0^a] \cup [\theta_0^b, 1]$, une règle générale sur la répartition des poids associés à chaque composante Θ_0^a est nécessaire pour garantir cette unicité. La stratégie proposée dans ce chapitre s'appuie sur le principe de symétrie aux bornes qui préconise des valeurs de $Pr(\text{Rejet } H_0 | \theta = \theta_0^a, \theta_0^b)$ les plus proches possibles.

L'utilisation simultanée des principes sur lesquels se fondent les procédures fréquentistes et bayésiennes aboutit à une approche optimale. Par rapport aux procédures fréquentistes décrites dans la section (B.5), le test $TFNCS$ est construit sur une règle de décision directement interprétable et constante quel que soit la période où l'expérience s'arrête. Par rapport aux procédures bayésiennes classiques, la correction de l'a priori de Jeffreys définit un état d'ignorance conditionnellement au paramètres du plan expérimental, en garantissant en plus l'unicité des solutions.

La mise à disposition d'un outil quantitatif puissant comme le facteur de Bayes corrigé et la possibilité d'exprimer l'erreur en termes de risques conditionnels et non-conditionnels en conjuguant l'inférence des tests $TFCS^*$ et $TFNCS$ permet d'interpréter les résultats à un niveau très supérieur à celui proposé dans les procédures existantes pour plans séquentiels.

5.6 Discussion

R.Fisher, J.Neyman et H.Jeffreys ont chacun argumenté en faveur d'approches du test d'hypothèse conduisant à des interprétations différentes. R.Fisher conçoit la valeur de $p = p(X)$ comme une mesure de l'évidence en faveur de H_0. L'absence de conditionnement sur l'observation $X = x$ et l'utilisation d'une hypothèse alternative lui fait rejeter l'approche de Neyman qui fonde l'inférence sur deux types d'erreurs liées au rejet ou non de H_0. L'idée d'une valeur prédéterminée de l'erreur de type 1 ($= \alpha$) s'oppose à la nature aléatoire de $p = p(X)$, qui est uniformément distribuée sur $[0, 1]$ sous H_0 (voir Hubbard et Bayarri (2003) [35]). H.Jeffreys rejoint Neyman dans la nécessité de poser une alternative mais s'oppose à la démarche fréquentiste d'utiliser une aire-sous-la courbe comme outil d'inférence.

Le test BBW (décrit section (B.5.2)) peut être interprété comme une tentative de réconcilier ces trois points de vue (voir Berger (2003) [4]). L'idée est de suivre l'approche de Neyman-Pearson en définissant formellement les risques fréquentistes de type 1 et 2. On conditionne ensuite sur la valeur de $p = p(X)$ (ou sur une fonction strictement monotone de p comme le facteur de Bayes) sous H_0 et H_1 qui mesure la force de l'évidence, comme le suggérait R.Fisher. Les risques d'erreurs fréquentistes obtenus sont égaux aux probabilités a posteriori de chaque hypothèse, suivant la démarche suggérée par H.Jeffreys.

Comme pour l'estimation par intervalle, l'approche fréquentiste non-conditionnelle des procédures de test séquentiel nécessitent de définir a priori un ordre stochastique pour les séquences d'observations. La difficulté est d'abord calculatoire pour définir les bornes d'arrêt de l'expérience. La méthode peut être basée sur une optimisation de la puissance ou du nombre d'observations attendu sous H_0 ou H_1. Le manque d'unicité des solutions est une critique partagée avec l'approche bayésienne qui dépend de la définition d'un a priori sous les deux hypothèses. Dans l'inférence fréquentiste conditionnelle, Berger et Coll. (1994) [7] met en avant le conditionnement sur

la statistique censée quantifier l'évidence potentielle en faveur des hypothèses pour argumenter l'inutilité d'un deuxième conditionnement sur le nombre d'observations réalisées. Comme dans l'approche bayésienne classique, les procédures fréquentistes conditionnelles sont donc sujettes au biais expérimental induit par la règle d'arrêt.

Bayarri et Berger (2004) [2] suggèrent que toute démarche inférentielle devrait simultanément s'appuyer sur les approches bayésienne et fréquentiste. Cependant cette position consensuelle visait à argumenter le test BBW. La procédure présentée dans ce chapitre va plus loin en apportant une réponse justifiée à cette assertion.

Conclusion

Le principe de la règle d'arrêt est une conséquence du principe de vraisemblance. Il sous-entend que 1) la décision d'arrêter l'expérience ne doit dépendre que des données et que 2) les résultats ne dépendent en aucun cas du moment où l'expérience s'arrête. La formulation classique du théorème de Bayes ne remet pas en cause ces principes mais le biais expérimental qu'induit un arrêt prématuré de l'expérience n'est pas pris en compte dans l'inférence bayésienne. En acceptant une telle transgression, une formulation alternative du théorème de Bayes et les outils performants développés dans cette thèse apportent une réponse cohérente et justifiée au problème de la règle d'arrêt dans l'inférence bayésienne.

Le cadre théorique se fonde sur la reformulation par R. de Cristofaro (1996) [24] du théorème de Bayes qui permet l'intégration de l'information sur le plan expérimental dans la loi a priori. Cette réécriture offre une légitimité nouvelle à l'a priori de Jeffreys qui garantit la propriété d'impartialité du plan expérimental. En considérant que l'information sur le plan expérimental perdue dans la vraisemblance peut être récupérée dans l'information de Fisher, on exploite ensuite l'idée simple d'utiliser une vraisemblance directement associée au plan d'échantillonnage pour dériver une famille de lois a priori dont fait partie l'"a priori de Jeffreys corrigé".

Le cas de l'évaluation d'une proportion dans le contexte de K échantillonnages Binomiaux successifs conduit à considérer la loi Bêta-J conjuguée avec la loi Binomiale. Soient n_k et Y_k respectivement le nombre prédéfini d'observations et le nombre de succès observés pendant la période k $(1 \leq k \leq K)$ et soit M la *va* associée au

nombre total d'analyses effectuées. En notant $r_k = n_k/n_1$ et $J_k \ni Y = \sum_{i=1}^{k} Y_i$ la zone de continuation de l'expérience en période $k+1$, une loi a priori conditionnelle au plan expérimental d_{Seq} est distribuée suivant une loi Bêta-J, i.e.

$$\varphi | d_{Seq} \sim Be^J(a, b, c; n_1, r_2, r_3, .., r_K; J_1, J_2, .., J_{K-1}),$$

si elle a pour densité

$$p(\phi | d_{Seq}) = C\ \phi^{a-1}(1-\phi)^{b-1}[1 + r_2 Pr(M \geq 2) + \ldots + r_K Pr(M = K)]^c.$$

La règle d'arrêt de l'expérience est définie par $Pr(M \geq k) = Pr(Y \in J_k | M > k-1)$.

L'a priori de Jeffreys corrigé est associée au paramétrage $a = b = c = 1/2$. Il permet de compenser les effets de la règle d'arrêt proportionnellement au biais induit sur l'estimateur du maximum de vraisemblance (i.e. la proportion observée). Ce résultat étend les conclusions de R. de Cristofaro portant sur les a priori de Jeffreys dans l'échantillonnage Binomial fixé et Binomial Négatif, respectivement distribués suivant les lois $Be(\frac{1}{2}, \frac{1}{2})$ et $Be(0, \frac{1}{2})$. Lorsque $K = 2$, $n_1 = 1$, $n_2 \to +\infty$ et que l'observation d'un succès conditionne l'arrêt de l'expérience en période 1, l'a priori de Jeffreys corrigé tend vers une loi $Be(\frac{1}{2}, 1)$. Cette compensation intermédiaire entre $Be(\frac{1}{2}, \frac{1}{2})$ et $Be(0, \frac{1}{2})$ souligne la cohérence dans l'interprétation.

Ce rapport de thèse contient des applications dans l'estimation ponctuelle, l'estimation par intervalle et dans le test d'hypothèses lorsque $K = 2$. D'autres cas lorsque $K > 2$ sont étudiés dans les propositions d'articles en annexe. L'ensemble des méthodes bayésiennes développées présentent des propriétés fréquentistes remarquables.

L'étude dans l'estimation ponctuelle révèle le lien entre le paramétrage des lois Bêta dans l'échantillonnage fixe et Bêta-J dans l'échantillonnage séquentiel. Le mode et la moyenne des lois a posteriori respectivement fondées sur l'a priori Uniforme $Be(1, 1)$ et l'a priori de Haldane $Be(0, 0)$ conduisent à l'estimateur du maximum de vraisemblance dans l'échantillonnage fixe. Dans l'échantillonnage séquentiel, les estimateurs

associés au mode et à la moyenne dérivés de distributions a priori de type Bêta-J respectivement paramétrées par $a = b = c = 1$ et $a = b = 0, c = 1$ présentent des propriétés remarquables en terme d'interprétation, de biais et d'écart quadratique moyen. La dérivation d'une forme analytique simple du mode constitue une alternative avantageuse dans la majorité des plans séquentiels.

Dans l'échantillonnage Binomial fixe, l'intervalle de crédibilité construit à partir de l'a priori de Jeffreys classique présente des propriétés optimales en terme de taux de recouvrement et de longueur moyenne. On a montré que l'intervalle construit à partir de l'a priori de Jeffreys corrigé conserve ces bonnes propriétés en venant compenser l'effet de la règle d'arrêt sur les bornes. La supériorité de cet intervalle est établie en terme d'interprétation par rapport à l'intervalle de Jennison-Turnbull qui reste la seule réponse fréquentiste au problème de la règle d'arrêt.

Bayarri et Berger (2004) [2] suggèrent que toute démarche inférentielle devrait simultanément s'appuyer sur les approches bayésienne et fréquentiste. Cependant cette position consensuelle visait à promouvoir le test fréquentiste conditionnel BBW. La procédure de test présentée dans le chapitre (5) va plus loin en apportant une réelle justification à cette assertion. On utilise simultanément un test fréquentiste non-conditionnel et un test bayésien admettant comme règle d'arrêt et de rejet de H_0 une même valeur limite du facteur de Bayes. L'a priori de Jeffreys corrigé offre un mécanisme de compensation des effets de la règle d'arrêt sur le rapport des évidences des deux hypothèses testées. Le réajustement du facteur de Bayes va dans le sens de sa valeur moyenne si le plan expérimental ne comportait pas d'arrêt optionnel. Dans le test fréquentiste non-conditionnel, les risques d'erreur sont fixés à des niveaux nominaux α et β. Dans le test bayésien, les erreurs dépendent de l'observation et s'interprètent en terme de probabilités d'erreurs des hypothèses. Pour la plupart des séquences observées, ces probabilités d'erreurs se réinterprètent en terme de risques fréquentistes conditionnels à une mesure de l'évidence.

L'utilisation simultanée des principes sur lesquels se fondent les procédures fréquen-

tistes et bayésiennes aboutit à une approche optimale. Par rapport aux tests fréquentistes non-conditionnels, la règle de décision est facilement interprétable et reste constante quelle que soit la période. Par rapport aux procédures bayésiennes classiques, la correction de l'a priori de Jeffreys définit un état d'ignorance conditionnellement aux spécificités du plan expérimental et garantit l'unicité des solutions. L'énoncé d'un principe sur la répartition des poids de la loi a priori sous H_0 offre en plus l'opportunité de minimiser les risques de type 1 associés à chaque composante d'une hypothèse nulle disjointe.

Discussion et perspectives

En acceptant plus ou moins dogmatiquement le principe de vraisemblance et son lien implicite avec la formulation classique du théorème de Bayes, les statisticiens bayésiens n'ont pas su proposer de solutions au problème de la règle d'arrêt. Dans le sens inverse les procédures bayésiennes ont peu d'écho dans les sciences expérimentales, dès lors que le plan expérimental peut avoir un effet sur le résultat.

L'utilisation de la règle de Jeffreys pour déterminer la loi a priori est d'abord justifiée par la propriété d'invariance (voir Jeffreys (1961)[36]). Box et Tiao (1992) [12] argumentent ensuite sur le principe de translation de la vraisemblance. Enfin R. de Cristofaro (1996) [24] apporte une nouvelle justification avec le principe d'impartialité du plan expérimental. L'idée nouvelle de formaliser l'a priori de Jeffreys à partir de l'information de Fisher d'une vraisemblance d'échantillonnage Binomial séquentiel offre pour la première fois une perspective bayésienne formelle et cohérente au problème de la règle d'arrêt.

Cette thèse a montré le potentiel des procédures construites à partir d'un a priori intégrant les spécificités du plan expérimental. Par rapport aux procédures bayésiennes classiques la correction apportée permet de compenser le biais induit sur la distribution a posteriori ainsi que sur les indicateurs bayésiens qui en sont dérivés dans l'estimation ponctuelle, l'estimation par intervalle et le test d'hypothèses.

La supériorité par rapport aux procédures fréquentistes, y compris celles tenant compte de la règle d'arrêt, est plus surprenante. Dans celles-ci, le prix à payer pour

le respect des propriétés fondamentales se traduit par une difficulté d'interprétation. On a montré les incohérences de l'estimateur sans biais de variance minimum et des intervalles de confiance fondés sur un ordre stochastique des séquences hypothétiques. De plus cet ordre présente la contrainte d'introduire un élément de subjectivité dans l'inférence sur le paramètre, ce qui est habituellement reproché aux procédures bayésiennes qui nécessitent la détermination d'une loi a priori. Pour le test d'hypothèses, le défaut d'interprétation et l'absence d'unicité des bornes de rejet de l'hypothèse nulle dans les procédures fréquentistes sont des contraintes entièrement supprimées dans la procédure bayésienne présentée dans cette thèse. De plus, la définition d'une distribution a priori satisfaisant le principe d'égalité au bornes garantie une répartition juste du risque α entre les différentes composantes d'une hypothèse nulle disjointe.

La reformulation du théorème de Bayes par R. de Cristofaro justifie l'utilisation de lois a priori intégrant la règle d'arrêt. Le travail novateur réalisé dans cette thèse donne un contenu formalisé dans l'échantillonnage Binomial. L'apport pour l'inférence bayésienne est fondamental. Il ouvre la voie à la détermination de lois a priori dérivés de l'information de Fisher visant à optimiser les propriétés fréquentistes.

Le choix délibéré de proposer une solution globale - concept, estimation, intervalle de confiance, test - dans l'étude d'une proportion s'est fait au détriment d'un travail plus large concernant d'autres types d'échantillonnages (Normal, Poisson,..) ou d'autres problématiques. Un effort de recherche reste donc a accomplir en direction des éléments restants. Des développements sont également envisageables dans toutes les disciplines utilisant l'information de Fisher.

Enfin, une remise en cause sérieuse des principes de vraisemblance et de la règle d'arrêt peut maintenant être argumentée. Dans une formulation plus directe, une procédure inférentielle qui ignore l'information sur le plan expérimental ne peut plus être qualifiée d'optimale.

Bibliographie

[1] M. Abramovitz and I.A. Stegun, *Handbook of mathematical functions*, Dover, New York, 1964.

[2] M.J. Bayarri and J.O. Berger, *The interplay of bayesian and frequentist analysis*, Statistical Science **19** (2004), no. 1, 58–80.

[3] J.O. Berger, *Statistical decision theory and bayesian analysis*, Springer-Verlag, 1988.

[4] ———, *Could fisher, jeffreys and neyman have agreed on testing ?*, Statistical Science **18** (2003), no. 1, 1–32.

[5] J.O. Berger, B. Boukai, and Y. Wang, *Unified frequentist and bayesian testing of a precise hypothesis*, Statistical Science **12** (1997), no. 3, 133–160.

[6] ———, *Simultaneous bayesian-frequentist sequential testing of nested hypotheses*, Biometrika **86** (1999), no. 1, 79–92.

[7] J.O. Berger, L.D. Brown, and R.L. Wolpert, *Unified conditional frequentist and bayesian test for fixed and sequential hypothesis testing*, Ann. Statist. **22** (1994), 317–352.

[8] J.O. Berger and T. Sellke, *Testing a point null hypothesis : the irreconcilability of p values and evidence*, JASA **82** (1987), no. 397, 112–139.

[9] J.O. Berger and R.L. Wolpert, *The likelihood principle (2nd edition)*, Hayward, CA : Institute of Mathematical Statistics Monograph Series, 1988.

[10] R.N. Bhattacharya and R.R. Rao, *Normal approximation and asymptotic expansions*, Wiley, 1976.

[11] A. Birnbaum, *On the foundations of statistical inference (with discussion)*, JASA **57** (1962), no. 298, 269–326.

[12] G.E. Box and G.C. Tiao, *Bayesian inference in statistical analysis*, Wiley, 1992.

[13] L.D. Brown, T.T. Cai, and A. DasGupta, *Interval estimation for a binomial proportion (with discussion)*, Statistical Science **16** (2001), 101–133.

[14] ———, *Confidence interval for a binomial proportion and edgeworth expansions*, Annal of Statistics **30** (2002), 160–201.

[15] ———, *Interval estimation in exponential families*, Statistica Sinica **13** (2003), 19–49.

[16] T.T. Cai, *One-sided confidence intervals in discrete distributions*, Technical Report, University of Pennsylvania (2003).

[17] B.P. Carlin and T.A. Louis, *Bayes and empirical bayes methods for data analysis*, Chapman & Hall, 2000.

[18] R. Carnap, *Logical fondations of probability*, The University of Chicago Press, Chicago, 1962.

[19] M.N Chang and P.C O'Brien, *Confidence interval following group sequential tests*, Controlled Clinical Trials **7** (1986), 18–26.

[20] M.N Chang, H.S. Wieand, and V.T. Chang, *The bias of the sample proportion following a group sequential phase-two clinical trial*, Statistics in Medicine **8** (1989), 563–570.

[21] C.J Clopper and E.S Pearson, *The use of confidence or fiducial limits illustrated in the case of the binomial*, Biometrika **26** (1934), 404–413.

[22] S. Dass, *Unified bayesian and conditional frequentist testing for discrete distributions*, Sankhya **63b** (2001), 251–269.

[23] S. Dass and J.O. Berger, *Unified bayesian and conditional frequentist testing of composite hypotheses*, Scandinavian Journal of Statistics **30** (2003), no. 1, 193–210.

[24] R. de Cristofaro, *L'influence du plan d'échantillonnage dans l'inférence statistique*, Journal de la Société statistique de Paris **137** (1996), no. 4, 23–34.

[25] _____, *The inductive reasoning in statistical inference*, Comm. Stat. Theory Methods, 7, vol. 31, 2002, pp. 1079–1089.

[26] _____, *The inductive reasoning in statistical inference*, (2005).

[27] _____, *On the fondations of likelihood principle*, Journal of Statistical Planning and Inference **in Press** (2005).

[28] _____, *On the relevance of sampling rule in bayesian inference*, (2005).

[29] _____, *The problem of statistical induction and some of its solutions*, (2005).

[30] D. Duffy and T. J. Santner, *Confidence intervals for a binomial parameter based on multistage tests*, Biometrics **43** (1987), 81–94.

[31] R.A. Fisher, *Statistical methods and scientific inference, 2nd edition*, Oliver and Boyd, Edinburgh, 1956.

[32] D. Fourdrinier, *Statistique inférentielle*, Dunod, 2002.

[33] L.S. Freedman and D.J. Spiegelhalter, *The assesment of subjective opinion and its use in relation to stopping rules for clinical trials*, The Statistician **33** (1983), 153–160.

[34] C. Howson and P. Urbach, *Scientific reasoning : the bayesian approach, 2nd edition*, La salle, IL : Open Court, 1993.

[35] R. Hubbard and M.J. Bayarri, *Confusion over measure of evidence (p's) versus error (α's) in classical statistical testing*, The American Statistician **57** (2003), 171–183.

[36] H. Jeffreys, *Theory of probability*, Oxford University Press (3ème édition), 1961.

[37] C. Jennison and B.W. Turnbull, *Confidence intervals for a binomial parameter following a multistage test with an application to mil-std 105d and medical trials*, Technometrics **25** (1983), 49–58.

[38] _____, *Group sequential methods with applications to clinical trials*, Chapman & Hall, 2000.

[39] S. Jung and K.M. Kim, *On the estimation of the binomial probability in multistage clinical trials*, Statistics in Medicine **23** (2004), 881–896.

[40] J. Kiefer, *Conditional confidence statements and confidence estimators*, JASA **72** (1977), no. 360, 789–827.

[41] K. Lan and D.L. DeMets, *Discrete sequential boundaries for clinical trials*, Biometrika **70** (1983), no. 3, 659–663.

[42] B. Lecoutre, *Distribution bêta-j - cas particuliers*, Rapport Technique - LMRS, Université de Rouen (2004).

[43] D.V. Lindley, *The philosophy of statistics*, The Statistician **49** (2000), 293–337.

[44] A. Liu and W.J. Hall, *Unbiased estimation following a group sequential test*, Biometrika **86** (1999), 71–78.

[45] R.G. Newcombe, *Two-sided confidence intervals for the single proportion : Comparison of seven methods*, Statistics in Medicine **17** (1998), 857–872.

[46] S. Pampallona, A.A. Tsiatis, and K. Kim, *Interim monitoring of group sequential trials using spending functions for the type 1 and type 2 error probabilities*, Drug Information Journal **35** (2001), 1113–1121.

[47] C. Robert, *L'analyse statistique bayésienne*, Economica, 1990.

[48] T. Sellke, M.J. Bayarri, and J.O. Berger, *Calibration of p-values for testing precise null hypotheses*, The American Statistician **55** (2001), 62–71.

[49] N. Stallard and K.M. Facey, *Comparison of the spending function method and the christmas tree correction for group sequential trials*, Journal of Biopharmaceutical Statistics **6** (1996), 361–373.

[50] N. Stallard and S. Todd, *Exact sequential tests for single samples of discrete responses using spending functions*, Statistics in Medicine **19** (2000), 3051–3064.

[51] S.E. Vollset, *Confidence intervals for a binomial proportion*, Statistics in Medicine **12** (1993), 809–824.

[52] S.K. Wang and A.A. Tsiatis, *Approximately optimal one-parameter boundaries for group sequential trials*, Biometrics **43** (1987), 193–199.

[53] Y. Wang, B. Boukai, and J.O. Berger, *A bayesian testing procedure with valid conditional frequentist interpretation*, (1994), 187–195.

[54] J. Whitehead, *On the bias of maximum likelihood estimation following a sequential test*, Biometrika **73** (1986), no. 3, 573–581.

[55] ———, *The case for frequentism in clinical trials*, Statistics in Medicine **12** (1993), 1405–1413.

[56] G. Zech, *Frequentist and bayesian confidence intervals*, (2002).

[]

Annexe A

Notions de base

L'objectif de ce chapitre est de définir le cadre théorique et de rappeler les concepts généraux qui sont utilisés dans la partie princeps du document. Les énoncés sont illustrés dans la perspective d'une étude sur les proportions.

A.1 Éléments de probabilité

Une expérience, pouvant donner lieu à des résultats différents lorsqu'elle est répétée dans des conditions identiques, est qualifiée d'*aléatoire* si son résultat ne peut être prévu à l'avance. On considère le résultat de cette expérience comme un élément w de l'*ensemble fondamental* Ω des résultats de toutes les expériences susceptibles d'être menées. Le résultat peut prendre la forme d'une simple observation ou bien d'une quantité fonction de cette observation.

Un *événement* est une assertion logique associée au résultat de l'expérience. Il est réalisé si la proposition correspondante est vérifiée par l'expérience. Dans une démarche probabiliste, on associe un événement à la partie des éléments de Ω qui lui correspond.

On note $\mathcal{P}(\Omega)$ l'*ensemble des parties* de Ω dans lequel on peut définir une classe

de parties appelée *tribu* $\mathcal{T} \in \mathcal{P}(\Omega)$ dont chaque élément est considéré comme un événement. Pour cela $\mathcal{T}$ doit satisfaire aux trois axiomes suivants :

1. $\forall A$ un événement tel que $A \in \mathcal{T}$ et soit $\overline{A}$ son complémentaire dans Ω alors $\overline{A} \in \mathcal{T}$,
2. $\forall A_1, A_2, ..., A_n \in \mathcal{T}$ un ensemble dénombrable d'événements alors $\bigcup A_i \in \mathcal{T}$,
3. $\Omega \in \mathcal{T}$.

Le couple $(\Omega, \mathcal{T})$ constitue un *espace probabilisable* (mesurable) sur lequel on définit une *probabilité* (mesure) $P_w : \mathcal{T} \mapsto [0, 1]$ telle que soient vérifiées les conditions

1. $P_w(\Omega) = 1$,
2. $\forall A_1, A_2, ..., A_n \in \mathcal{T}$ un ensemble dénombrable d'événements incompatibles alors $P_w(\bigcup A_i) = \sum_i P_w(A_i)$.

Le triplet $(\Omega, \mathcal{T}, P_w)$ constitue un *espace probabilisé*.

La probabilité conditionnelle tient une place fondamentale parmi tous les concepts probabilistes. Pour $A, B \in \mathcal{T}$, on appelle *probabilité conditionnelle* de A sachant B, que l'on note $P_c(A|B)$ dans ce paragraphe, la probabilité définie par

$$P_c(A|B) = P_w(A \cap B)/P_w(B).$$

On vérifie facilement que P_c satisfait les deux conditions ci-dessus. C'est donc une nouvelle probabilité définie sur l'espace probabilisable $(\Omega, \mathcal{T})$.

Le *théorème de Bayes* découle d'une application directe des probabilités conditionnelles. Dans l'énoncé, P désigne toute probabilité définie sur $(\Omega, \mathcal{T})$.

Théorème de Bayes. *On considère A un événement et $B_1, B_2, .., B_J$ une partition complète de Ω. En supposant connue la probabilité conditionnelle $P(A|B_j)$ ainsi que les probabilités marginales $P(B_j)$, on dérive la probabilité associée à B_j conditionnellement à une observation de A par l'équation*

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{j=1}^{J} P(A|B_j)P(B_j)} \tag{A.1.1}$$

Le concept de *variable aléatoire* formalise la notion de grandeur variant suivant le résultat d'une expérience aléatoire.

Définition. *Une variable aléatoire (va) X est une application mesurable d'un espace probabilisé $(\Omega, \mathcal{T}, P_w)$ dans un espace probabilisable $(E, \mathcal{E})$. La mesure P^X est une probabilité associée à X si quel que soit $C \in \mathcal{E}$, P^X est la mesure image de P_w par X, c.a.d*

$$P^X(C) = P_w(\{w; X(w) \in C\}) = P_w(X^{-1}(C)).$$

Un résultat $w \in \Omega$ peut être associé à un couple de *va* (X, Y). On écrira alors la fonction de distribution jointe

$$p(x, y) = Pr(X = x, Y = y) = Pr(\{w; X(w) = x\}, \{w; Y(w) = y\}),$$

et la fonction de distribution de X conditionnellement à une valeur $Y = y$

$$p(x|y) = Pr(X = x|Y = y) = Pr(\{w; X(w) = x\}|\{w; Y(w) = y\}) = p(x, y)/p(y).$$

Si X est une *va* discrète à valeur dans l'ensemble E, on notera l'espérance mathématique

$$E[X] = \int_{\Omega} X(w) dP_w = \int_{E} x \, dP^X(x) = \sum_{x_i \in E} x_i \, Pr(X = x_i).$$

Au delà de ce chapitre, la mention au résultat $w \in \Omega$ n'apparaîtra plus. Par exemple, si $x \in \mathcal{E}$ est une valeur possible d'une *va* discrète $X : (\Omega, \mathcal{T}) \to (E, \mathcal{E})$, la probabilité associée à l'événement $\{w; X(w) = x\}$ sera notée $Pr(X = x)$ ou encore $p(x)$ qui exprime la fonction de distribution de X.

A.2 Échantillonnage

Dans la démarche statistique, on suppose que l'*observation* x est la *réalisation* sur l'*échantillon* observé d'une *va* X définie sur l'espace des observations $\mathcal{X}$. Par opposition à la démarche probabiliste où la loi de probabilité suivie par le phénomène que l'on cherche à modéliser est supposée connue, la démarche statistique sous-entend

que X est distribuée suivant une loi de probabilité P_θ indexée par un paramètre inconnu θ contenu dans l'espace des paramètres Θ. P_θ est appelée *loi d'échantillonnage* et sert à caractériser l'ensemble des observations que l'on aurait pu obtenir par le même procédé en répétant l'expérience un grand nombre de fois.

Définition. *On appelle modèle statistique le triplet* $(\mathcal{X}, \mathcal{T}, P_\theta; \theta \in \Theta)$ *où* P_θ *est une famille de lois de probabilité définie sur* $\mathcal{T}$ *une tribu fixée de* $\mathcal{X}$ *l'espace des échantillons. On appelle statistique sur ce modèle toute application* S*, indépendante de* θ*, à valeur dans un autre espace probabilisable.*

Pour définir la probabilité P_θ, il est nécessaire de caractériser un espace probabilisable et donc de fixer une tribu. On supposera dans la suite que P_θ est toujours définie, avec une forme analytique connue, et que l'ensemble des parties de $\mathcal{X}$ est mesurable. La mention à la tribu sur $\mathcal{X}$ sera implicite et on utilisera le doublet $(\mathcal{X}, P_\theta; \theta \in \Theta)$ pour définir le modèle statistique correspondant.

On considère une série de n *observations* $\mathbf{x} = (x_1, x_2, ..., x_n)$ de n *va* binaires $\mathbf{X} = (X_1, X_2, ..., X_n)$, avec $X_i = 0$ ou 1 indépendantes conditionnellement à une valeur fixée de θ et ayant une même probabilité θ de *succés* (i.e. $Pr(X_i = 1) = \theta$). La *va* $\mathbf{X}$ est alors une séquence $\{0, 1\}^n$ de n tirages de la loi de *Bernouilli* ($X_i \sim Ber(\theta)$) avec chacun une fonction de distribution

$$p(x_i|\theta) = \theta^{x_i}(1-\theta)^{x_i} = \begin{cases} \theta & \text{si } x_i = 1 \\ 1-\theta & \text{sinon.} \end{cases}$$

Le modèle d'échantillonnage associé est le modèle de Bernouilli $(\{0, 1\}^n, Ber(\theta)^{\otimes n}; \theta \in [0, 1])$ où $Ber(\theta)^{\otimes n}$ est la loi puissance n de la loi $Ber(\theta)$. La fonction de distribution jointe de $\mathbf{X}$ s'écrit

$$p(\mathbf{x}|\theta) = \prod_{i=1}^{n} \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

pour les 2^n valeurs possibles de $\mathbf{x}$.

Le modèle Binomial est dérivé du modèle d'échantillonnage de Bernouilli par la statistique $S : (x_1, x_2, ..., x_n) \mapsto y = \sum_{i=1}^{n} x_i$. La *va* associée Y est distribuée suivant

la loi image de $Ber(\theta)^{\otimes n}$ par S ; c'est une loi Binomiale $Bin(n, \theta)$ définie dans l'espace $\{0, 1, .., n\}$ avec une fonction de distribution

$$p(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}.$$

La fonction de distribution de $Bin(n, \theta)$ est la fonction de distribution de $Ber(\theta)^{\otimes n}$ multiplié par un facteur combinatoire C_n^y qui est la probabilité d'obtenir la séquence $\mathbf{x} = (x_1, x_2, ..., x_n)$ pour une valeur y fixée. Les propriétés de la loi de Bernouilli en terme d'espérance $E[X_i|\theta] = \theta$ et de variance $V[X_i|\theta] = \theta(1-\theta)$ conduisent à celles de la loi Binomiale $E[Y|\theta] = n\theta$ et $V[Y|\theta] = n\theta(1-\theta)$.

De plus l'égalité

$$\frac{p(y+1|\theta)}{p(y|\theta)} = \frac{(n-y)\theta}{(y+1)(1-\theta)}$$

permet de déduire que la fonction de distribution de $Bin(n, \theta)$ est croissante en y (i.e. $p(y+1|\theta) > p(y|\theta)$) tant que

$$y < (n+1)\theta - 1$$

et décroissante sinon et que le mode est atteint lorsque

$$mode(p(y|\theta)) = [(n+1)\theta]_,$$

où $[]_$ désigne la partie entière.

La *va* $T = Y/n$ (taux de succès) possède la même fonction de distribution que Y pour les valeurs $t = 0, 1/n, 2/n, .., 1$. En posant $y = nt$,

$$p(t|\theta) = \binom{n}{nt} \theta^t (1-\theta)^{n(1-t)}.$$

On en déduit que $E[T|\theta] = \theta$ et $V[T|\theta] = \theta(1-\theta)/n$.

Dans la suite, $E[Y|\theta]$ sera noté $E_\theta[Y]$.

A.3 Réduction des données

Le principe de réduction des données consiste à utiliser une statistique fonction de l'observation afin d'en réduire la dimension. Si en plus la statistique satisfait la propriété d'exhaustivité, on ne perd aucune information sur le paramètre étudié.

Définition. *Une statistique S définie sur un modèle $(\mathcal{X}, P_\theta; \theta \in \Theta)$ à valeurs dans E est exhaustive pour θ si $p(X|S=s)$, la fonction de distribution de X conditionnellement à une valeur $S=s$, ne dépend pas de θ $\forall s \in E$.*

Par exemple, si $\mathbf{X} = (X_1, X_2, ..., X_n)$ est un n-échantillon dans le modèle de Bernouilli $(\{0,1\}^n, Ber(\theta)^{\otimes n}; \theta \in [0,1])$, l'information sur θ contenue dans $Y = \sum_{i=1}^n X_i$ peut complètement se substituer à celle contenue dans $\mathbf{X}$ car Y satisfait la propriété d'exhaustivité : l'approche par probabilités conditionnelles permet de démontrer que la distribution de $\mathbf{X}$ sachant $Y = y$ ne dépend pas de θ.

$$\begin{aligned} Pr(\mathbf{X} = \mathbf{x} | Y = y) &= \frac{Pr(\mathbf{X} = \mathbf{x} \cap Y = y)}{Pr(Y = y)} \\ &= \begin{cases} \frac{\prod_{i=1}^n \theta^{x_i}(1-\theta)^{n-x_i}}{\binom{n}{y}\theta^y(1-\theta)^{n-y}} & \text{si } \sum_{i=1}^n x_i = y \\ 0 & \text{sinon} \end{cases} \\ &= \begin{cases} 1/\binom{n}{y} & \text{si } \sum_{i=1}^n x_i = y \\ 0 & \text{sinon} \end{cases} \end{aligned}$$

On peut également chercher à mesurer l'information sur θ contenue dans $\mathbf{X}$ pour déterminer si cette information est conservée lorsque les données sont résumées par une statistique. Dans cette perspective, le théorème de factorisation permet de savoir si une statistique est exhaustive pour θ.

Si on note $L(\theta; x_1, x_2, .., x_n)$ la probabilité conjointe de l'échantillon $p(x_1, x_2, .., x_n \,|\, \theta) = Pr(X_1 = x_1 \cap X_2 = x_2 \cap \ldots \cap X_n = x_n \,|\, \theta)$, $L(\theta; \mathbf{x})$ comme fonction de θ seul est appelée *vraisemblance* de θ. On remarque que $L(\theta; \mathbf{X})$ est une *va*, mais également une statistique exhaustive par définition.

Théorème de Factorisation. *Une statistique S définie sur le modèle $(\mathcal{X}, P_\theta; \theta \in \Theta)$*

est exhaustive pour θ si

$$L(\theta; \boldsymbol{x}) = g(s, \theta)h(\boldsymbol{x})$$

où $h(\boldsymbol{x})$ ne dépend pas de θ et $g(s, \theta)$ dépend de $\boldsymbol{x}$ uniquement à travers s, la valeur observée de S.

En d'autres termes, une fois s observée, aucune autre information en provenance de l'échantillon n'apporte d'information supplémentaire sur θ.

Dans le modèle $(\{0,1\}^n, Ber(\theta)^{\otimes n}; \theta \in [0,1])$, la vraisemblance s'écrit

$$L(\theta; \mathbf{x}) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{n-x_i} = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}.$$

L'identification $g(\mathbf{x}, \theta) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{n-x_i}$ et $h(\mathbf{x}) = 1$ démontre que $\mathbf{X} = (X_1, X_2, .., X_n)$ est une statistique exhaustive pour θ. En posant $y = \sum_{i=1}^n x_i$, la vraisemblance s'écrit également

$$L(\theta; \mathbf{x}) = \theta^y(1-\theta)^{(n-y)},$$

ce qui permet une deuxième identification $g(y, \theta) = \theta^y(1-\theta)^{(n-y)}$ et $h(y) = 1$. La *va* associée Y est donc une autre statistique exhaustive pour θ.

Le concept de minimalité intervient pour définir une statistique qui puisse réduire au maximum les donnée sans perte d'information.

Définition. *Soient $h(\mathbf{x}, \mathbf{z}; \theta) = \prod_{i=1}^n L(\theta; x_i) / \prod_{i=1}^n L(\theta; z_i)$ le rapport des vraisemblances en $\mathbf{x}, \mathbf{z} \in \mathcal{X}^n$, la statistique S est exhaustive minimale pour θ si $h(\mathbf{x}, \mathbf{z}; \theta)$ est indépendante de θ dans l'unique cas $S(\boldsymbol{X}) = S(\boldsymbol{Z})$.*

Dans le modèle $(\{0,1\}^n, Ber(\theta)^{\otimes n}; \theta \in [0,1])$ avec $\mathbf{x}, \mathbf{z} \in \{0,1\}^n$, on a

$$\frac{\prod_{i=1}^n L(\theta; x_i)}{\prod_{i=1}^n L(\theta; z_i)} = \frac{\prod_{i=1}^n \theta^{x_i}(1-\theta)^{(n-x_i)}}{\prod_{i=1}^n \theta^{z_i}(1-\theta)^{(n-z_i)}} = \left(\frac{\theta}{1-\theta}\right)^{\sum_{i=1}^n x_i - \sum_{i=1}^n z_i}$$

On note que l'expression $(\frac{\theta}{1-\theta})^{\sum_{i=1}^n x_i - \sum_{i=1}^n z_i}$ est indépendante de θ uniquement si $\sum_{i=1}^n x_i = \sum_{i=1}^n z_i$. La statistique $Y = S(\mathbf{X}) = \sum_{i=1}^n X_i$ est donc exhaustive minimale pour θ.

Une statistique exhaustive S atteint un certain degré d'optimalité dans la réduction des données si elle est minimale mais également si aucune fonction g de S n'est libre. Dans ce cas, S est une statistique complète.

Définition. *Soit S une statistique définie dans le modèle $(\mathcal{X}, P_\theta; \theta \in \Theta)$ et F une famille de fonctions définie de $\mathcal{X}$ dans R, S est F-complète si quelle que soit la fonction g telle que $g \circ S \in F$ et $\forall \theta \in \Theta$, on a*

$$\int_{\mathcal{X}} g \circ S dP_\theta = 0 \Longrightarrow g \circ S = 0 \ P_\theta - p.s.$$

Dans le modèle $(\{0, 1, \ldots, n\}, Bin(n, \theta); \theta \in [0, 1])$, la *va* $Y \sim Bin(n, \theta)$ est une statistique complète car l'équation

$$\begin{aligned} 0 &= \sum_{y=0}^{n} g(y) \binom{n}{y} \theta^y (1-\theta)^{(n-y)} \\ &= (1-\theta)^n \sum_{y=0}^{n} g(y) \binom{n}{y} \left(\frac{\theta}{1-\theta}\right)^y \end{aligned}$$

aboutit à la résolution d'un polynôme de degré n dont la seule solution est $g(y) = 0$.

Si l'on dispose d'une statistique exhaustive, le théorème de Rao-Blackwell permet d'améliorer au sens de la variance les performance d'un estimateur sans biais de θ (i.e. $E_\theta[\hat{\theta}^{SB}] = \theta$).

Théorème de Rao-Blackwell. *Soit $\hat{\theta}$ un estimateur et S est une statistique exhaustive pour θ, on a l'inégalité*

$$E_\theta[E[\hat{\theta}|S] - \theta]^2 \leq E_\theta[\hat{\theta} - \theta]^2.$$

Le théorème de Lehman-Scheffé va plus loin en utilisant la statistique exhaustive complète.

Théorème de Lehman-Scheffé. *Si $\hat{\theta}^{SB}$ est un estimateur sans biais de θ et S une statistique exhaustive complète alors l'estimateur $\hat{\theta}^{VUMSB} = E[\hat{\theta}^{SB}|S]$ satisfait l'inégalité $V_\theta[\hat{\theta}^{VUMSB}] \leq V_\theta[\hat{\theta}^{SB}]$. L'estimateur $\hat{\theta}^{VUMSB}$ est sans biais de variance uniformément minimum (VUMSB) pour toute valeur de θ*

Démonstration. D'après le théorème de Lehman-Scheffé, si S est une statistique exhaustive complète dans un modèle statistique et la fonction $h(S)$ un estimateur sans biais de θ alors $h(S)$ est l'estimateur $VUMSB$ de θ.

On suppose maintenant que T est un autre estimateur sans biais de θ. D'après le théorème de Rao-Blackwell, T est amélioré par $E[T|S]$. Ainsi, si $h(S)$ n'est pas $VUMSB$ alors il doit exister une autre fonction $h^*(S)$ qui est sans biais et dont la variance est inférieure à celle de $h(S)$ $\forall\theta \in \Theta$.

La statistique S étant exhaustive complète, l'égalité

$$E_\theta[h^*(S) - h(S)] \equiv 0$$

implique que $h^*(S) = h(S)$.

On démontre ainsi l'unicité d'un estimateur $VUMSB$. □

A.4 Information de Fisher

Ce qui précède nous montre tout l'avantage de travailler avec une statistique qui préserve l'information en provenance des données ; l'information de Fisher nous donne un moyen de la quantifier.

Définition. *On appelle information de Fisher $I_n(\theta)$ apportée sur θ par un n-échantillon la quantité positive ou nulle :*

$$I_n(\theta) = E\left[\left(\frac{\partial}{\partial\theta} LogL(\theta;\mathbf{x})\right)^2\right] = -E\left[\frac{\partial^2}{\partial\theta^2} LogL(\theta;\mathbf{x})\right] \tag{A.4.1}$$

La pente en θ de la courbe de vraisemblance est associée à la valeur de $\frac{\partial}{\partial\theta} LogL(\theta;\mathbf{X})$. On note l'importance de cette *va* dans la démonstration de (A.4.1).

Démonstration. La vraisemblance $L(\theta;\mathbf{x})$ est une fonction de distribution car

$$\sum_{\mathbf{x}\in\mathcal{X}^n} L(\theta;\mathbf{x}) = 1.$$

En dérivant les deux membres par rapport à θ et en utilisant l'égalité

$$\frac{\partial}{\partial\theta}L(\theta;\mathbf{x}) = L(\theta;\mathbf{x})\frac{\partial}{\partial\theta}LogL(\theta;\mathbf{x}), \tag{A.4.2}$$

on trouve

$$\sum_{\mathbf{x}\in\mathcal{X}^n}\left(\frac{\partial}{\partial\theta}LogL(\theta;\mathbf{x})\right)L(\theta;\mathbf{x}) = 0$$

ce qui montre que $\frac{\partial}{\partial\theta}LogL(\theta;\mathbf{X})$ est une *va* centrée de variance $I_n(\theta)$.

En dérivant une seconde fois et en utilisant de nouveau (A.4.2), on trouve

$$\sum_{\mathbf{x}\in\mathcal{X}^n}\left(\frac{\partial^2}{\partial\theta^2}LogL(\theta;\mathbf{x})\right)L(\theta;\mathbf{x})dx + \sum_{\mathbf{x}\in\mathcal{X}^n}\left(\frac{\partial}{\partial\theta}LogL(\theta;\mathbf{x})\right)^2 L(\theta;\mathbf{x}) = 0$$

ce qui finit de démontrer (A.4.1). □

Lorsque l'espace $\Theta \ni \theta$ ne dépend pas de la valeur de θ, on a la propriété d'additivité $I_n(\theta) = nI_1(\theta)$ qui signifie que chaque observation indépendante à la même importance. Dans le modèle Binomial avec $Y \sim Bin(n,\theta)$, on a

$$\begin{aligned}
-E_\theta\left[\frac{\partial^2}{\theta^2}LogL(\theta;Y)\right] &= -E_\theta\left[\frac{\partial^2}{\theta^2}(YLog(\theta) + (n-Y)Log(1-\theta))\right] \\
&= -E_\theta\left[-\frac{Y}{\theta^2} - \frac{n-Y}{(1-\theta)^2}\right] \\
&= -\left[-\frac{n\theta}{\theta^2} - \frac{n-n\theta}{(1-\theta)^2}\right] \\
&= \frac{n}{\theta(1-\theta)}.
\end{aligned}$$

L'information de Fisher dépend de θ. Elle est maximum pour les valeurs extrêmes de θ (i.e. $\theta \to 0$ ou $\theta \to 1$), ce qui indique que l'information apportée par le modèle est maximum pour ces valeurs. L'*estimateur du maximum de vraisemblance* permet une interprétation plus intuitive de l'information de Fisher.

Soit l'observation $x \in \mathcal{X}$, la valeur de θ pour laquelle $L(\theta;x)$ est maximum est celle qui maximise la probabilité d'observer de nouveau x si l'expérience était répétée.

Définition. *L'estimation du maximum de vraisemblance $\hat{\theta}^{MV}_{(x)}$ est la valeur de θ pour laquelle la vraisemblance $L(\theta, x)$ est maximum. La va associée $\hat{\theta}^{MV}$ est l'estimateur du maximum de vraisemblance (EMV).*

Dans un modèle d'échantillonnage, si $LogL(\theta, \mathbf{x})$ est dérivable en θ, l'estimation $\hat{\theta}^{MV}_{(\mathbf{x})}$ est la solution de l'équation $\frac{\partial}{\partial\theta} LogL(\theta; \mathbf{x}) = 0$. L'invariance est une propriété importante de l'EMV (i.e. $g(\hat{\theta}^{MV})$ est l'EMV de $g(\theta)$).

L'information de Fisher permet de caractériser la précision de l'EMV en apportant une mesure du pic de la courbe de la log-vraisemblance autour de la valeur de $\hat{\theta}^{MV}_{(\mathbf{x})}$. Un pic mal défini est caractérisé par une évolution lente de la pente $\frac{\partial}{\partial\theta} LogL$. Ce cas correspond à des valeurs peu élevées de $\frac{\partial^2}{\partial\theta^2} Log\, L$ <u>et de $I_1(\theta)$</u> aux alentours de $\hat{\theta}^{MV}_{(\mathbf{x})}$.

Démonstration. L'égalité pour la quantité dérivée

$$-\frac{1}{n}\frac{\partial^2}{\partial\theta^2} Log\, L = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2}{\partial\theta^2} Log\, p(x_i|\theta)$$

permet de l'interpréter comme la moyenne de n fonctions identiques de $(x_1, \ldots, x_n)$.

En supposant que θ_0 est la valeurs vraie de θ telle que les x_i sont tirés d'une distribution $p(x_i|\theta)$, alors pour n suffisamment grand, cette moyenne converge en probabilité vers son espérance

$$-E_{x_i|\theta_0}\left[\frac{\partial^2}{\theta^2} Log\, p(x_i|\theta)\right] = -\int \frac{\partial^2}{\theta^2} Log\, p(x_i|\theta) dx_i = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2}{\theta^2} Log\, p(x_i|\theta) = a(\theta, \theta_0).$$

D'autre part, toujours pour n suffisamment grand, $\hat{\theta}^{MV}$ converge en probabilité vers θ_0. En regroupant, il vient les égalités

$$\left(-\frac{1}{n}\frac{\partial^2}{\partial\theta^2} Log\, L\right)_{\hat{\theta}^{MV}_{(x)}} \simeq a(\hat{\theta}^{MV}_{(x)}, \theta_0) \simeq a(\hat{\theta}^{MV}_{(x)}, \hat{\theta}^{MV}_{(x)}) \simeq I_1(\hat{\theta}^{MV}_{(x)}).$$

□

A contrario, un pic bien défini de la courbe de la log-vraisemblance est associé à des valeurs élevées de $I_1(\theta)$ pour les valeurs de θ proches de $\hat{\theta}^{MV}_{(\mathbf{x})}$, caractérisant un

EMV plus précis. Cette interprétation est conforme à l'inégalité de Cramer-Rao : si $\hat{\theta}^{MV}$ est sans biais, la variance de $\hat{\theta}^{MV}$ tend asymptotiquement vers $I_n^{-1}(\theta)$.

On peut enfin formaliser le lien existant localement entre la vraisemblance, l'EMV et $I_n(\theta)$. Un développement limité d'ordre 2 en $\hat{\theta}^{MV}_{(x)}$ permet d'approcher la log-vraisemblance par

$$\begin{aligned} Log(L(\theta;x)) &= Log(\prod_{i=1}^{n} L(\theta;x_i)) \\ &\simeq Log(L(\hat{\theta}^{MV}_{(x)};x)) + \frac{1}{2}(\theta - \hat{\theta}^{MV}_{(x)})^2 \left(\frac{\partial^2}{\partial\theta^2} Log\, L\right)_{\hat{\theta}^{MV}_{(x)}} . \end{aligned}$$

On déduit de la forme exponentielle

$$\begin{aligned} L(\theta;x) &= exp\left(Log(L(\hat{\theta}^{MV}_{(x)};x)) + \frac{1}{2}(\theta - \hat{\theta}^{MV}_{(x)})^2 \left(\frac{\partial^2}{\partial\theta^2} Log\, L\right)_{\hat{\theta}^{MV}_{(x)}}\right) \\ &\simeq exp\left(\frac{1}{2}(\theta - \hat{\theta}^{MV}_{(x)})^2 \left(\frac{\partial^2}{\partial\theta^2} Log\, L\right)_{\hat{\theta}^{MV}_{(x)}}\right). \end{aligned}$$

Ainsi la vraisemblance est approximativement distribuée en θ suivant

$$L(\theta;x) \sim N\left(\hat{\theta}^{MV}_{(x)}, -1/\left(\frac{\partial^2}{\partial\theta^2} Log\, L\right)_{\hat{\theta}^{MV}_{(x)}}\right).$$

Après une nouvelle approximation, on obtient enfin la relation

$$L(\theta;x) \sim N\left(\hat{\theta}^{MV}_{(x)}, I_n^{-1}(\hat{\theta}^{MV}_{(x)})\right).$$

Par opposition au principe d'exhaustivité, une statistique ancillaire n'apporte, par elle même, pas d'information sur le paramètre.

Définition. *Une statistique $A = g(X)$ définie sur le modèle $(\mathcal{X}, P_\theta; \theta \in \Theta)$ est ancillaire pour θ si*

$$P(A|\theta) = P(A)$$

ne dépend pas de θ, $\forall \theta \in \Theta$.

Dans le modèle Binomial, on déduit aisément l'estimateur $\hat{\theta}^{MV} = Y/n$. D'autre part, si N la *va* associée à la taille de l'échantillon est rattaché à un processus aléatoire dans une expérience où la réalisation n est indépendante du résultat, N est une statistique ancillaire pour θ dans ce modèle. La valeur de n n'apporte aucune information sur θ bien qu'elle soit un élément majeur du *plan expérimental*. Elle permet en outre de caractériser la précision de l'EMV puisque $\lim_{n\to\infty} V_\theta[\hat{\theta}^{MV}] = \theta(1-\theta)/n$.

Corollaire. *La vraisemblance en θ basée sur une va X est égale à la vraisemblance en θ basée sur X conditionnellement à la valeur d'une statistique ancillaire. Soit A la statistique ancillaire, on a*

$$L(\theta; x) = p(x|\theta) = p(x|a, \theta).$$

Démonstration.

$$p(x|\theta) = p(x, a|\theta) = p(x|a, \theta)p(a|\theta).$$

En notant que la vraisemblance en θ basée sur une statistique ancillaire est constante (i.e. $p(a|\theta) = c$), on trouve finalement que

$$p(x|\theta) \propto p(x|a, \theta)$$

□

Annexe B

Inférence statistique

Dans ce document, on considère que l'observation provient d'un modèle statistique $(\mathcal{X}, P_\theta; \theta \in \Theta)$ paramétré par un scalaire θ, qui résume l'information du phénomène que l'on cherche à étudier. D'autres approches (Jeffreys (1964) [36]) sous entendent que l'ensemble de ces phénomènes sont par nature aléatoires.

Les méthodes d'inférence répondent à deux problématiques : l'estimation ponctuelle ou par intervalle de la valeur de θ et l'acceptation ou le rejet d'hypothèses portant sur une valeur ou un ensemble de valeurs de θ. Les inférences bayésienne et fréquentiste sont décrites dans la première section. Des méthodes de référence pour l'estimation et le test sont ensuite exposées pour l'évaluation d'une proportion dans l'échantillonnage fixe. La dernière section est dédiée à l'inférence fréquentiste conditionnelle.

B.1 Approches bayésienne et fréquentiste

L'approche bayésienne a été développée par Pierre-Simon Laplace (1749-1827) et Thomas Bayes (1702-1761). Elle est cohérente avec la théorie des probabilités et résulte d'une simple application du théorème de Bayes.

Dans l'équation (A.1.1), si l'on remplace la probabilité $P(A)$ par une distribution

prédéfinie $\Pi(d\theta) = p(\theta)$ et $P(A|B_j)$ par la loi d'échantillonnage de probabilité $p(x|\theta)$, on dérive la distribution a posteriori $\Pi(d\theta|x) = p(\theta|x)$

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int\limits_{\theta\in\Theta} p(x|\theta)p(\theta)d\theta} \sim p(x|\theta)p(\theta) \tag{B.1.1}$$

Les assertions probabilistes sur θ, portées par la *va* associée ϑ, sont censées quantifier le degré d'incertitude sur sa valeur. Dans l'approche bayésienne, on accepte de fait l'idée d'une information a priori sur θ que l'on formalise dans la distribution a priori de probabilité $p(\theta)$.

Dans le modèle Binomial où $Y \sim Bin(n, \theta)$, une loi a priori sur θ du type $\vartheta \sim Be(a, b)$ est associée à une loi a posteriori $\vartheta/y \sim Be(a + y, b + n - y)$.

Démonstration. Pour tout $\theta \in [0, 1]$, la loi Bêta $Be(a, b)$ admet pour densité

$$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}.$$

De plus pour tout $y \in \{0, 1, ..., n\}$, la densité jointe $p(y, \theta)$ est définie par

$$p(y, \theta) = \binom{n}{y}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{y+a-1}(1-\theta)^{n-y+b-1}.$$

En intégrant sur θ, on dispose de la densité marginale pour y d'une loi connue sous le nom de Bêta-Binomiale de paramètre n, a et b :

$$\begin{aligned} m(y) &= \int_0^1 p(y, \theta)d\theta \\ &= \binom{n}{y}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(y+a)\Gamma(n-y+b)}{\Gamma(n+a+b)}. \end{aligned}$$

La loi a posteriori de ϑ sachant y est une $Be(y + a, n - y + b)$. Sa densité s'obtient par

$$p(\theta|y) = \frac{p(y, \theta)}{p(y)} = \frac{\Gamma(n+a+b)}{\Gamma(y+a)\Gamma(n-y+b)}\theta^{y+a-1}(1-\theta)^{n-y+b-1}.$$

□

L'écriture du modèle bayésien intègre les assertions probabilistes portées sur le paramètre.

Définition. *Le modèle statistique bayésien est un modèle statistique* $(\mathcal{X}, P_\theta; \theta \in \Theta)$ *tel que l'espace du paramètre* θ *soit muni d'une loi de probabilité appelée loi a priori.*

L'idée d'inverser le rôle de l'observation et du paramètre dans $p(x|\theta)$ pour déterminer $p(\theta|x)$ est à l'origine de l'inférence fiduciaire (Fisher (1956) [31]). Cette démarche s'appuie sur le principe du pivot et ne fait pas intervenir d'autre quantité que la loi d'échantillonnage. Elle consiste à déterminer une loi de distribution sur le paramètre probabilisé en normalisant la fonction de vraisemblance. La densité de la loi fiduciaire se dérive par

$$p(\theta|x) = \frac{L(\theta; x)}{\int_{\theta \in \Theta} L(\theta; x) d\theta}.$$

En s'affranchissant d'une loi a priori sur θ, la loi fiduciaire pour θ n'exprime que l'information apportée par les données. Cependant, les difficultés liées à la résolution analytique n'ont pas permis le développement de l'inférence fiduciaire au delà des cas triviaux. Le développement des lois a priori non informatives dans l'approche bayésienne permet toutefois d'en conserver le cadre conceptuel.

L'approche fréquentiste a été développé au début du XXème siècle par Ronald Fisher (1890-1962), et Jerzy Neyman (1894-1981) pour le test d'hypothèse dans son approche décisionnelle. Par opposition à l'inférence bayésienne, aucune assertion probabiliste n'est portée sur $\theta \in \Theta$. L'inférence fréquentiste se fonde sur la distribution d'échantillonnage et le principe de répétition de l'expérience. Elle s'interprète sur l'hypothétique espace d'observations $\mathcal{X}$ dont fait partie celui que l'on a observé. La performance des procédures fréquentistes se mesurent en moyennant les résultats sur l'espace $\mathcal{X}$.

B.2 Estimation ponctuelle

Dans le modèle statistique $(\mathcal{X}, P_\theta; \theta \in \Theta)$, où $x \in \mathcal{X}$ désigne la réalisation de la *va* X sur le n-échantillon observé et P_θ la loi d'échantillonnage de probabilité $p(x|\theta)$,

l'estimation ponctuelle consiste à approcher les valeurs de θ à partir de x. Un estimateur T est donc une statistique fonction de la *va* X à valeurs dans un domaine acceptable pour le paramètre $\theta \in \Theta$.

L'inférence fréquentiste permet d'établir que $\hat{\theta}$ est un estimateur non-biaisé de θ (i.e. $E_\theta(\hat{\theta}) = \theta$). Si l'expérience était reconduite une infinité de fois dans les mêmes conditions, la valeur moyenne de l'estimation $\hat{\theta}_{(x)}$ serait θ (loi des grands nombres).

Dans l'approche bayésienne, la loi a posteriori $p(\theta|x)$ est considérée comme le meilleur résumé de l'information sur θ, une fois l'expérience réalisée. Par analogie avec l'EMV, on peut en donner le mode, qui est également le maximum de $p(x|\theta)p(\theta)$, ou encore la moyenne et la médiane. Le mode conserve certaines propriétés d'optimalité de l'EMV, en particulier celle de l'invariance lorsque la loi a priori est uniforme.

Dans le modèle Binomial, les estimateurs bayésiens découlent des caractéristiques de la loi bêta $Be(\alpha, \beta)$ dont la moyenne s'écrit

$$\begin{aligned}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \theta\theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \\ &= \frac{\alpha}{\alpha+\beta}\end{aligned}$$

Le paramétrage de la loi a posteriori de ϑ, à savoir $\alpha = a + y$ et $\beta = b + n - y$, conduit à l'estimation

$$\begin{aligned}Moy(p(\theta|y)) &= \frac{a+y}{a+b+n} \\ &= \frac{a+b}{a+b+n}\,\frac{a}{a+b} + \frac{n}{a+b+n}\,\frac{y}{n}\end{aligned}$$

$Moy(p(\theta|y))$ s'écrit donc comme une combinaison linéaire de la moyenne a priori $Moy(p(\theta)) = a/(a+b)$ et de l'estimation $\hat{\theta}^{MV}_{(y)} = y/n$. Lorsque n devient grand, il reste la composante asymptotique y/n, effaçant ainsi l'apport de l'information a priori.

Le mode de la loi $Be(\alpha, \beta)$ est solution de l'équation

$$\begin{aligned}\frac{\partial}{\partial\theta} L(\theta; \alpha, \beta) &= \frac{\partial}{\partial\theta}[(\alpha - 1)Log(\theta) + (\beta - 1)Log(1 - \theta)] \\ &= \frac{\alpha - 1}{\theta} - \frac{\beta - 1}{1 - \theta} = 0\end{aligned}$$

Cette expression s'annule pour une valeur

$$Mode(L(\theta; \alpha, \beta)) = \frac{\alpha - 1}{\alpha + \beta - 2}.$$

En reparamétrant selon le loi a posteriori $Be(a + y, b + n - y)$, on trouve que

$$\begin{aligned}Mode(p(\theta|y)) &= \frac{a + y - 1}{n + a + b + -2} \\ &= k\,\frac{y}{n} + (1 - k)\,\frac{a - 1}{a + b - 2} \qquad \text{avec } k = \frac{n}{n + a + b - 2}.\end{aligned}$$

Comme $k \in [0, 1]$, le mode de $p(\theta|y)$ est une valeur comprise entre le mode de la loi a priori $Mode(p(\theta)) = (a - 1)/(a + b - 2)$ et le mode de la vraisemblance $Mode(L(\theta; y)) = Mode(p(\theta|y)) = y/n$. Lorsque n devient grand, il reste la composante asymptotique y/n.

B.3 Estimation par intervalle

L'intervalle de confiance consiste à déterminer un ensemble de valeurs de θ considérées comme compatibles avec l'observation $x \in \mathcal{X}$. Une valeur θ est supportée par x, si sa vraisemblance $L(\theta; x)$ est grande. Sur la base de cette seule information, on peut retenir un ensemble de valeurs de θ pour lesquelles la vraisemblance est supérieure à une limite prédéfinie. Cependant l'amplitude de $L(\theta; x)$ n'étant pas contrôlable à l'avance, la division par $L(\hat{\theta}^{MV}_{(x)}; x)$, où $\hat{\theta}^{MV}_{(x)}$ est l'estimation de l'EMV de θ, permet d'obtenir la fonction de vraisemblance relative

$$r(\theta) = \frac{L(\theta; x)}{L(\hat{\theta}^{MV}_{(x)}; x)} = \frac{p(x|\theta)}{max_\theta\, p(x|\theta)},$$

à valeur dans [0,1].

Dans le modèle Binomial avec $Y \sim Bin(n, \theta)$, la vraisemblance relative conditionnellement à la valeur observée y s'interprète comme une mesure du support apporté à θ par $Bin(n, \theta)$ relativement au support maximum en $\theta = \hat{\theta}^{MV}_{(y)}$ apporté par $Bin(n, \hat{\theta}^{MV}_{(y)})$. L'intervalle de confiance qui en découle est défini par les valeurs de θ pour lesquelles $r(\theta)$ excède une valeur prédéfinie comprise entre 0 et 1.

La normalisation de la vraisemblance induit un autre changement d'échelle qui permet de construire un intervalle de confiance sur la base d'un raisonnement probabiliste. La vraisemblance sous la forme

$$L(\theta; y) = p(\theta|y) = c\,\theta^y (1-\theta)^{n-y}$$

exprime la densité d'une loi fiduciaire pour ϑ conditionnellement à une valeur $Y = y$.

Avec la constante de normalisation $c = (n+1)\binom{n}{y}$, on reconnaît la densité d'une loi $Be(y+1, n-y+1)$ qui se trouve être la loi bayésienne a posteriori de ϑ/y, associée à la loi a priori $\vartheta \sim U[0,1](\sim Be(1,1))$, dite loi a priori non informative de Laplace. L'intervalle de confiance dont les bornes $(\hat{\theta}^{inf}, \hat{\theta}^{sup})$ sont définies par

$$\begin{cases} \int_{\hat{\theta}^{inf}}^{\hat{\theta}^{sup}} p(\theta|y)d\theta = 1-\alpha \\ + \text{ une contrainte entre } \hat{\theta}^{inf} \text{ et } \hat{\theta}^{sup} \end{cases}$$

trouve donc une justification dans l'inférence fiduciaire et l'inférence bayésienne. Dans ce dernier cas, on l'appelle intervalle de crédibilité contrôlé bilatéralement avec un degré d'incertitude $1-\alpha$.

Lorsque le degré d'incertitude est contrôlé unilatéralement, les bornes sont définies par

$$\int_0^{\hat{\theta}^{sup}} p(\theta|y)d\theta = \int_{\hat{\theta}^{inf}}^1 p(\theta|y)d\theta = 1-\alpha/2.$$

L'intervalle de crédibilité déterminé sur la base de la loi a priori uniforme $Be(1,1)$ présente l'avantage de quantifier l'incertitude autour de l'EMV, qui est la réduction en un point de cet intervalle (i.e. intervalle d'incertitude 0%).

Dans le modèle statistique $(\mathcal{X}, P_\theta; \theta \in \Theta)$, l'inférence fréquentiste permet d'établir qu'un intervalle de confiance $[\hat{\theta}^{inf}, \hat{\theta}^{sup}]$ est contrôlé au niveau $1-\alpha$. Si l'expérience était reconduite une infinité de fois dans les mêmes conditions, $(1-\alpha)(\times 100)\%$ des intervalles de confiance calculés sur les échantillons ainsi générés contiendraient θ (ou avec une fréquence $(1-\alpha)(\times 100)\%$ d'où le terme fréquentiste).

Les valeurs de θ retenues dans un intervalle de confiance fréquentiste sont déterminées sur la base de la loi d'échantillonnage paramétrée par θ, que l'on confronte à x^*, la valeur observée de x. Si x^* n'est pas trop loin dans la queue de distribution, la valeur de θ est considérée comme suffisamment supportée par l'observation x^*. Dans un intervalle de confiance unilatéral fondé sur l'inversion du test de la Binomiale, la borne inférieure $\hat{\theta}^{inf}$ correspond à la valeur de θ telle que $P_\theta(X \geq x^*) = \alpha/2$.

Dans le modèle Binomial où y est la valeur observée, les bornes $\hat{\theta}^{inf}$ et $\hat{\theta}^{sup}$ de l'intervalle bilatéral sont les valeurs extrêmes de θ qui satisfont les égalités

$$Pr(Y \geq y \,|\, Y \sim Bin(n, \hat{\theta}^{inf})) = \alpha/2 \tag{B.3.1a}$$

$$\text{et } Pr(Y \leq y \,|\, Y \sim Bin(n, \hat{\theta}^{sup})) = \alpha/2. \tag{B.3.1b}$$

Ces bornes sont celles de l'intervalle de Clopper-Pearson (1934) [21] qui se calculent analytiquement.

On note la fonction Bêta incomplète

$$\begin{aligned} I_\xi(a,b) &= \frac{1}{\beta(a,b)} \int_0^\xi t^{a-1}(1-t)^{b-1} dt & \xi \in [0,1] \\ &= 1 - I_{1-\xi}(b,a) \\ \text{avec } \beta(a,b) &= \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} & a, b > 0 \end{aligned}$$

En utilisant le lien existant entre la loi Binomiale et la fonction Bêta incomplète (Abramowitz, M. et Stegun, I.A. (1964)[1]), l'équation (B.3.1a) se récrit

$$Pr(Y \geq y \,|\, Y \sim Bin(n, \hat{\theta}^{inf})) = I_{\hat{\theta}^{inf}}(y, n-y+1) \qquad \text{si } y \neq 0.$$

De la même façon, l'équation (B.3.1b) se récrit

$$\begin{aligned} Pr(Y \leq y \,|\, Y \sim Bin(n, \hat{\theta}^{sup})) &= 1 - P(Y \geq y+1 | Y \sim Bin(n, \hat{\theta}^{sup})) \\ &= 1 - I_{\hat{\theta}^{sup}}(y+1, n-y) \qquad \text{si } y \neq n \\ &= I_{\hat{\theta}^{sup}}(n-y, y+1) \end{aligned}$$

On en déduit que les bornes $(\hat{\theta}^{inf}_{CP}, \hat{\theta}^{sup}_{CP})$ de l'intervalle de Clopper-Pearson (CI_{CP}) sont déterminées par les quantiles $\alpha/2$ des lois $Be(y, n-y+1)$ et $Be(n-y, y+1)$ qui sont les lois bayésiennes a posteriori associées respectivement aux lois a priori $\vartheta \sim Be(0,1)$ et $\vartheta \sim Be(1,0)$. L'intervalle CI_{CP} trouve donc une interprétation bayésienne en terme d'intervalles de crédibilité unilatéraux.

On définit le taux de recouvrement unilatéraux pour les bornes inférieure et la borne supérieure respectivement par

$$C^{inf}(\theta, n) = P_\theta(\theta \geq \hat{\theta}^{inf}) \tag{B.3.2a}$$

$$C^{sup}(\theta, n) = P_\theta(\theta \leq \hat{\theta}^{sup}). \tag{B.3.2b}$$

Une propriété fréquentiste remarquable de l'intervalle de Clopper-Pearson tient dans le taux de recouvrement de chacune des bornes qui n'est jamais inférieur au niveau nominal $1 - \alpha/2$.

Démonstration. Si l'on considère $\hat{\theta}^{inf}$ la borne inférieure de l'intervalle de Clopper-Pearson unilatéral au niveau $1 - \alpha/2$ et y_{max} la dernière valeur de $Y = y$ pour laquelle $\theta \geq \hat{\theta}^{inf} (= \hat{\theta}^{inf}(y))$, on a la relation

$$\begin{aligned} C^{inf}_{CP}(\theta, n) &= \sum_{i=0}^{n} I_{\{\theta \geq \hat{\theta}^{inf}\}} Pr(Y = i \,|\, \theta) \\ &= \sum_{i=0}^{y_{max}} P_\theta(Y = i) \\ &= P_\theta(Y \leq y_{max}) \\ &= 1 - P_\theta(Y > y_{max}) \geq 1 - \alpha/2. \end{aligned}$$

La dernière inégalité est toujours vérifiée car d'après (B.3.1a) toute valeur de θ comprise dans l'intervalle de Clopper-Pearson associé à la valeur $Y = y_{max}$ satisfait la relation

$$Pr(Y > y_{max} \,|\, Y \sim Bin(n, \theta)) < \alpha/2.$$

Le même raisonnement mené sur la borne supérieure conduit à la même conclusion.

□

B.4 Test d'hypothèses

Les règles d'inférence sont celles qui conduisent à conclure à la vérité d'une proposition à partir d'une autre proposition prise comme hypothèse.

Dans le contexte d'une procédure de test, cette définition formalise le processus inductif consistant à définir une hypothèse tenue pour vraie sur un ensemble de données, qui est ensuite confirmée ou infirmée par l'observation d'un échantillon aléatoire supposé représentatif.

B.4.1 Tests dans l'inférence fréquentiste

Supposons que le sous-ensemble $\Theta_0 \in \Theta$ corresponde à des valeurs d'intérêt pour θ dans un contexte scientifique donné. Dans le modèle statistique $(\mathcal{X}, P_\theta; \theta \in \Theta)$, on note x^* l'observation sur le n-échantillon. On associe à l'hypothèse $\{\theta \in \Theta_0\}$, notée H_0, le niveau de significativité

$$p = Max_\theta \text{ Pr(d'observer des données infirmant } H_0 \text{ au moins autant que } x^* | H_0).$$

Comme pour l'estimation, le contexte de l'inférence fréquentiste est celui d'une répétition à l'infini et dans les mêmes conditions, de l'expérience réalisée. Pour toute valeur de $\theta \in \Theta_0$, on peut ordonner les hypothétiques valeurs de $x \in \mathcal{X}$ en fonction de la force avec laquelle elles infirment H_0 et calculer la fréquence des x qui

conduisent à infirmer le contenu de H_0 au moins aussi fortement que x^*. Le niveau de significativité p s'interprète comme le maximum de cette fréquence, en faisant varier θ dans Θ_0.

La prise en compte de la seule hypothèse H_0 s'inscrit dans la démarche du test de l'hypothèse nulle de Fisher. Le niveau de significativité p informe sur la nature probable ou improbable de H_0 mais ne conduit pas à une décision conclusive. Cette interprétation rapproche Fisher d'une position quasi-bayésienne.

Le test d'hypothèses de Neyman-Pearson sous-entend l'existence de deux sous-ensembles disjoints Θ_0 et Θ_1 tels que $\Theta_0 \cup \Theta_1 = \Theta$. On cherche à déterminer de manière conclusive si l'une des hypothèses

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

est supportée par les observations.

Les caractéristiques du test sont déterminées avant l'expérience par les risques d'erreur en choisissant l'une ou l'autre hypothèse. En définissant la zone de rejet de H_0

$$R = \{\text{x ; l'observation de x conduit au rejet de } H_0\},$$

on définit les erreurs

$$\begin{aligned} &\text{Erreur type 1} = Sup_\theta(P_\theta(R|\theta)) && \forall\theta \in \Theta_0 \\ &\text{Erreur type 2} = P_{\theta_1}(\bar{R}|\theta)) && \theta_1 \in \Theta_1, \end{aligned}$$

où $\overline{R}$ est le complémentaire de R.

Dans la démarche de Neyman-Pearson, l'erreur de type 2 est déterminée en maintenant l'erreur de type 1 inférieure à un niveau nominal α prédéfini. Cette démarche s'inscrit dans celle plus large de la théorie de la décision avec des coûts inconnus et une règle ϕ à valeur dans $\{0, 1\}$. Pour chaque $x \in \mathcal{X}$, ϕ est la probabilité avec

laquelle l'hypothèse H_1 est acceptée. La règle de décision est définie par

$$\begin{aligned} \phi(x) =& 1 && \text{si } x \in R \\ \phi(x) =& 0 && \text{si } x \in \bar{R}. \end{aligned}$$

Les erreurs sont réexprimées en terme d'espérance de ϕ avec d'abord

$$\text{Erreur type 1} = Sup_{\theta \in \Theta_0} E_\theta(\phi(X)).$$

L'erreur de type 2 (ou 1-Puissance) ne peut être formulée que si H_1 est une hypothèse ponctuelle. Si H_1 est une hypothèse composite, le test fondé sur la règle ϕ_0 est qualifié d'uniformément le plus puissant si l'erreur type 1 est au plus égale à α et si pour tout autre test associé à une règle ϕ, on a

$$E_{\theta_0}(\phi(X)) \geq E_\theta(\phi(X)) \qquad \forall \theta \in \Theta_1.$$

L'interprétation des risques s'inscrit dans le cadre fréquentiste d'une répétition à l'infini de l'expérience.

Le lemne de Neyman-Pearson garantit l'existence, l'unicité et l'optimalité d'un test sur une famille de paramètres associée à une fonction rapport de vraisemblance monotone. Dans ce cas, on a $\forall x \in \mathcal{X}$

$$\frac{L(\theta_2, x)}{L(\theta_1, x)} > 1$$

si $\theta_2 > \theta_1$ ou bien si $\theta_2 < \theta_1$.

B.4.2 Tests dans l'inférence bayésienne

L'inférence bayésienne permet des approches plus directes en se fondant sur les lois a posteriori pour le calcul des risques d'erreur. L'approche préconisée par Jeffreys (1964) [36] s'accorde avec celle de Neyman sur la nécessité de comparer deux hypothèses. En définissant les probabilités a priori

$$\pi_0 = Pr(H_0) = Pr(\theta \in \Theta_0) \qquad \pi_1 = Pr(H_1) = Pr(\theta \in \Theta_1)$$

avec $\pi_0 + \pi_1 = 1$, on considère le rapport a priori

$$\pi_0/\pi_1.$$

De la même façon, les probabilités a posteriori

$$p_0 = Pr(H_0|x) = Pr(\theta \in \Theta_0|x) \qquad p_1 = Pr(H_1|x) = Pr(\theta \in \Theta_1|x)$$

avec $p_0 + p_1 = 1$ conduisent au rapport a posteriori

$$p_0/p_1.$$

Le *facteur de Bayes* défini par

$$B(x) = \frac{(p_0/p_1)}{(\pi_0/\pi_1)} = \frac{p_0\pi_1}{p_1\pi_0}$$

peut être interprété comme le rapport de H_0 sur H_1, conditionnellement aux données observées mais aussi relativement aux deux densité $\rho_0(\theta)$ et $\rho_1(\theta)$ qui sont les restrictions de la loi a priori $p(\theta)$ à respectivement Θ_0 et Θ_1.

Une fois renormalisées

$$\begin{aligned} \rho_0(\theta) =& p(\theta)/\pi_0 & \theta \in \Theta_0 \\ \rho_1(\theta) =& p(\theta)/\pi_1 & \theta \in \Theta_1, \end{aligned}$$

on explicite p_0 par

$$\begin{aligned} p_0 =& Pr(\theta \in \Theta_0|x) \\ =& \int_{\theta\in\Theta_0} p(\theta|x)d\theta \\ \propto& \int_{\theta\in\Theta_0} p(\theta)p(x|\theta)d\theta \\ \propto& \int_{\theta\in\Theta_0} p(x|\theta)\rho_0(\theta)d\theta. \end{aligned}$$

En procédant de la même façon avec $\rho_1(\theta)$, on obtient

$$p_1 \propto \int_{\theta\in\Theta_1} p(x|\theta)\rho_1(\theta)d\theta.$$

La réécriture du facteur de Bayes par

$$B(x) = \frac{(p_0/p_1)}{(\pi_0/\pi_1)} = \frac{\int_{\theta\in\Theta_0} p(x|\theta)\rho_0(\theta)d\theta}{\int_{\theta\in\Theta_1} p(x|\theta)\rho_1(\theta)d\theta}$$

permet de l'interpréter comme un rapport de vraisemblance pondérées par respectivement ρ_0 et ρ_1. Une valeur de $B(x)$ proche de 1 exprime l'absence de support des données en faveur de l'une ou l'autre des hypothèses ; une valeur faible (resp. forte) exprime un support de x en faveur de l'hypothèse H_1 (resp. H_0). Dans la section 2.3, on montre que les densités ρ_0 et ρ_1 peuvent jouer un rôle majeur dans le contexte d'un test séquentiel.

La fonction $B(x)$ est directement reliée à p_0 et p_1. Comme $p_0/p_1 = B(x)\pi_0/\pi_1$ et $p_1 = 1 - p_0$, p_0 et p_1 se réécrivent respectivement

$$\begin{aligned} p_0 =& \frac{B(x)}{[(1-\pi_0)/\pi_0] + B(x)} \\ p_1 =& \frac{[(1-\pi_0)/\pi_0]}{[(1-\pi_0)/\pi_0] + B(x)}. \end{aligned}$$

Le test fondé sur le facteur de Bayes est défini par la règle

$$\begin{aligned} &\text{Si } B(x) \leq 1, H_0 \text{ est rejetée avec une erreur } \alpha(B) = Pr(\theta \in \Theta_0|x) = p_0 \\ &\text{Si } B(x) > 1, H_0 \text{ est acceptée avec une erreur } \beta(B) = Pr(\theta \in \Theta_1|x) = p_1. \end{aligned} \tag{B.4.1}$$

Si H_0 et H_1 sont deux hypothèses ponctuelles

$$H_0 : \{\theta = \theta_0\} \qquad \text{versus} \qquad H_1 : \{\theta = \theta_1\},$$

on applique la même procédure avec un facteur de Bayes correspondant au rapport des vraisemblances

$$B(x) = \frac{p(x|\theta_0)}{p(x|\theta_1)}.$$

Les a priori non-informatifs $\pi_0 = \pi_1 = 1/2$ sont utilisés dans la suite.

B.5 Inférence fréquentiste conditionnelle

B.5.1 Conditionnement et partitionnement

Le conditionnement dans un test d'hypothèse fut argumenté par R. Fisher (1956) [31], qui préconisait l'utilisation d'une statistique ancilaire lorsqu'elle existe (voir section A.4), afin de ne pas déformer l'information sur le paramètre. La première étude formelle sur le sujet a été publiée dans Kieffer (1977) [40]. Kieffer étend le conditionnement à une statistique non-ancilaire $S(X)$, justifiant ainsi un partitionnement de l'espace des observations $\mathcal{X}$ en fonction des niveaux de $S(X)$. Dans le test des hypothèses H_0 versus H_1, l'espace $\mathcal{X}$ est partitionné en une famille de sous-échantillons

$$C = \{C^b_{H_i}, b \in B, i = 0, 1\}$$

dans laquelle b indice une partition dans l'ensemble des partitions B.

En conservant l'approche de Neyman-Pearson, l'hypothèse H_i est acceptée si X tombe dans $C^b_{H_i}$, avec un risque d'erreur défini par $P_{H_i}(\text{mauvaise décision}|C^b)$ dans la partition $C^b = C^b_{H_0} \cup C^b_{H_1}$.

Par exemple, B peut contenir deux éléments du type $B = \{\text{évidence faible, évidence forte}\}$. Chaque élément indice une partition de sous-échantillons offrant un faible ou un fort pouvoir de conclusion en faveur des hypothèses. Les familles de sous-échantillons sont alors

$$\mathcal{X} = \{C^{faible}_{H_0}, C^{faible}_{H_1}, C^{fort}_{H_0}, C^{fort}_{H_1}\},$$

avec des risques conditionnels associés au rejet de H_0 définis par

$$\begin{aligned}
\alpha^{faible} =& P_{H_0}(\text{rejet de } H_0|C^{faible}) \\
=& P_{H_0}(C^{faible}_{H_1}|C^{faible}) \\
\alpha^{fort} =& P_{H_0}(\text{rejet de } H_0|C^{fort}) \\
=& P_{H_0}(C^{fort}_{H_1}|C^{fort}).
\end{aligned}$$

Les risques β^{faible} et β^{fort} associés aux rejet de H_1 s'obtiennent de la même façon avec

$$\begin{aligned}\beta^{faible} =& P_{H_1}(C_{H_0}^{faible}|C^{faible})\\ \beta^{fort} =& P_{H_1}(C_{H_0}^{fort}|C^{fort}).\end{aligned}$$

B.5.2 Test Conditionnel BBW

Dans le contexte d'un test de deux hypothèses ponctuelles

$$H_0 : \{\theta = \theta_0\} \qquad \text{versus} \qquad H_1 : \{\theta = \theta_1\},$$

Berger, Brown et Wolpert (1994) [7] suggèrent un conditionnement sur la valeur d'une statistique $S(X) = s$ qui mesure la force de l'évidence dans X. Les risques conditionnels dépendent de l'observation x et sont définis par

$$\begin{aligned}\alpha(s) =& P_{H_0}(\text{rejet } H_0|S(X) = s)\\ \beta(s) =& P_{H_1}(\text{rejet } H_1|S(X) = s).\end{aligned}$$

Dans la suite de leur démarche, ils rejoignent Birnbaum (1962) [11] et Barnard (dans la discussion de Kieffer) (1977) [40] qui argumentent intuitivement en faveur d'un conditionnement par la valeur du rapport des vraisemblances (ou du facteur de Bayes) $B = B(x) = p(x|\theta_0)/p(x|\theta_1)$. Dans [7], BBW produisent une argumentation plus formelle. Le raisonnement suivant suppose que la distribution de X est continue et admet une densité ; il est étendu au cas discret dans la sous-section suivante.

En notant $F_0^{B(X)}$ et $F_1^{B(X)}$ les fonctions de répartition (supposées inversibles) de $B(X)$ respectivement sous H_0 et H_1, le test non-conditionnel le plus puissant avec des risques d'erreur type 1 et type 2 égaux est défini par la valeur critique c qui satisfait

$$F_0^{B(X)}(c) = 1 - F_1^{B(X)}(c). \tag{B.5.1}$$

Ce test rejette H_0 si $B(x) \leq c$ et l'accepte sinon, avec des risques d'erreurs déterminés avant l'expérience $\alpha = \beta = F_0^{B(X)}(c) = 1 - F_1^{B(X)}(c)$ (voir figure (B.1)).

D'autre part, on a vu dans (B.4.1) que les erreurs du test fondé sur le facteur de Bayes sont exprimées en terme de probabilité a posteriori de H_0 et de H_1 avec $\alpha(B) = B(x)/(1+B(x))$ et $\beta(B) = 1/(1+B(x))$. BBW suggèrent une modification de ce test bayésien afin d'exclure les situations contradictoires où H_0 est rejetée ou acceptée avec une probabilité d'erreur supérieure à 0.5. La solution proposée passe par la création d'une zone de non-décision qui rassemble les valeurs de x conduisant à cette situation. En supposant que la fonction $B(x)$ est croissante et en notant r la limite supérieure de la zone de rejet de H_0 (i.e. H_0 est rejetée si $B(x) \leq r$) et a la limite inférieure de la zone d'acceptation de H_0 (i.e. H_0 est acceptée si $B(x) \geq a$), l'intervalle $r < B(x) < a$ définit la zone de non-décision telle que

$$\begin{aligned} r =& 1 \text{ et } a = \psi(1) \qquad \text{si } F_0^{B(X)}(1) \leq 1 - F_1^{B(X)}(1) \\ r =& \psi^{-1}(1) \text{ et } a = 1 \quad \text{si } F_0^{B(X)}(1) > 1 - F_1^{B(X)}(1). \end{aligned} \tag{B.5.2}$$

où $\psi(B) = F_0^{B(X)(-1)}(1 - F_1^{B(X)}(B))$ et $\psi^{-1}(B) = F_1^{B(X)(-1)}(1 - F_0^{B(X)}(B))$. Les bornes $\{r, a\}$ conduisent de fait à une situation de symétrie des aires sous la courbe des densités de $B(X)$ sous H_0 et sous H_1, semblable à celle décrite dans (B.5.1)(voir figure (B.1)).

L'étude de Berger et Coll. montre que les probabilités d'erreur de ce test bayésien modifié sont égales aux risques d'erreur du même test dans l'inférence fréquentiste conditionnelle

$$\alpha^*(s) = \alpha(B) \qquad \text{et} \qquad \beta^*(s) = \beta(B)$$

lorsque le conditionnement se fait sur la *va*

$$S(X) = min\{B(X), \psi^{-1}(B(X))\}$$

dans le domaine $\mathcal{X}^* = \{X : 0 \leq s(X) \leq r\}$. Ce conditionnement sous-entend qu'une valeur de $B(x)$ identique sous H_0 et H_1 produit la même évidence statistique en faveur ou contre H_0.

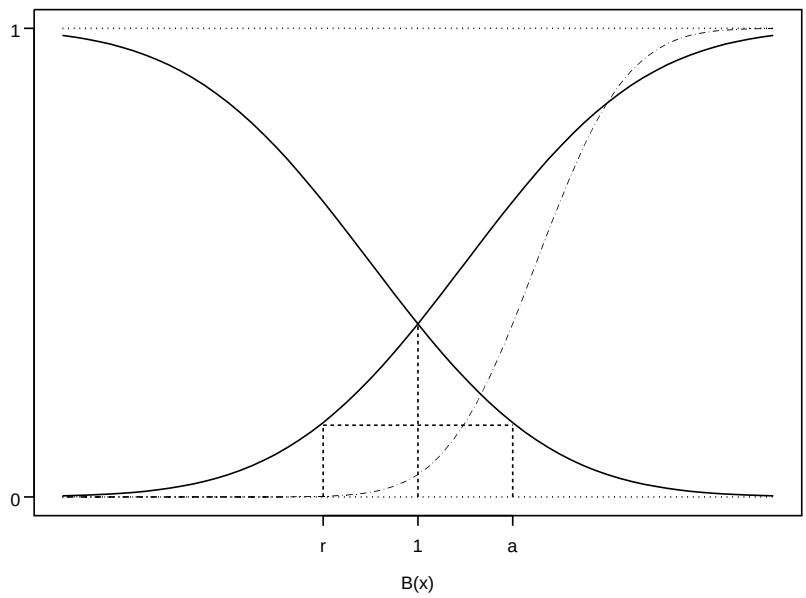

FIG. B.1 – *Valeur de r sur la fonction $F_0(B)$ (—) si H_0 et H_1 sont symétriques (i.e. $c = 1$) et $(\cdot - \cdot)$ sinon (ex. $c > 1$), et, valeur de a sur la fonction $1 - F_1(B)$*

Démonstration. La démonstration utilise le partitionnement de l'espace des observations par les valeurs de $S(X) = s$

$$\mathcal{X}^* = \bigcup_{s \in S} \mathcal{X}_s^* \quad \text{avec} \quad \mathcal{X}_s^* \cap \mathcal{X}_{s'}^* = \emptyset.$$

Pour $0 < s < r$, les partitions engendrées par $S(X)$ s'écrivent

$$\begin{aligned} \mathcal{X}_s^* &= \{x \in \mathcal{X}^* : S(x) = s\} \\ &= \{x \in \mathcal{X}^* : B(x) = s \text{ ou } B(x) = \psi(s)\}. \end{aligned}$$

En notant $f_0^{B(X)}$ et $f_1^{B(X)}$ les densités de $B(X)$ respectivement sous H_0 et H_1, on établit que

$$\begin{aligned} \frac{d}{ds}\psi(s) &= \frac{d}{ds} F_0^{B(X)(-1)}(1 - F_1^{B(X)}(s)) \\ &= \frac{-\frac{d}{ds} F_1^{B(X)}(s)}{\frac{d}{ds} F_0^{B(X)}(F_0^{B(X)(-1)}(1 - F_1^{B(X)}(s)))} \\ &= \frac{-f_1^{B(X)}(s)}{f_0^{B(X)}(\psi(s))}. \end{aligned}$$

Cette équation permet d'aboutir à

$$\begin{aligned}
\alpha^*(s) &= P(\text{rejet } H_0|\mathcal{X}_s^*) \\
&= \frac{P_{H_0}(B(x)=s)}{P_{H_0}(B(x)=s)+P_{H_0}(B(x)=\psi(s))} \\
&= \frac{f_0^{B(X)}(s)}{f_0^{B(X)}(s)+f_0^{B(X)}(\psi(s))|\frac{d}{ds}\psi(s)|} \\
&= \frac{f_0^{B(X)}(s)}{f_0^{B(X)}(s)+f_1^{B(X)}(s)} \\
&= \frac{B(s)}{B(s)+1} = \alpha(B).
\end{aligned}$$

La démonstration pour $\beta^*(s)$ est similaire ; elle s'étend aisément au cas $\pi_0 \neq \pi_1$. $\square$

La statistique $S(X)$ est ancilaire uniquement si $F_0(1) = 1 - F_1(1)$; dans ce cas, la zone de non-décision n'existe plus.

Les relations

$$\begin{aligned}
E_{H_0}[\alpha^*] &= E_{H_0}[P_{\theta_0}(\text{Erreur type } 1|S(X)=s)] \\
&= P_{\theta_0}(\text{Erreur type } 1) = \alpha \\
E_{H_1}[\alpha^*] &= E_{H_1}[P_{\theta_1}(\text{Erreur type } 2|S(X)=s)] \\
&= P_{\theta_1}(\text{Erreur type } 2) = \beta
\end{aligned}$$

montrent que les risques conditionnels $\alpha^*(s)$ et $\beta^*(s)$ sont obtenus en redistribuant les risques non-conditionnels α et β dans les partitions générées par les niveaux de $S(X)$ respectivement sous H_0 et H_1.

Le cas d'une alternative composite $H_1 : \{\theta \in \Theta_1\}$ est étudié dans Wang, Berger et Boukai (1994) [53] et Berger, Boukai et Wang (1997) [5] avec l'a priori $\pi_0 = \frac{1}{2}$ pour $H_0 : \{\theta = \theta_0\}$ et l'a priori $\pi_1(\theta) = p(\theta)/2$ définie sur Θ_1. On considère alors le facteur de Bayes

$$B(x) = \frac{p(x|\theta)}{m_1(x)} \qquad\qquad y = 0, 1, ..n$$

où $m_1(x) = \int_{\Theta_1} p(x|\theta)\pi_1(\theta)d\theta$ est la probabilité marginal de X sous H_1.

Comme dans l'inférence fréquentiste non-conditionnelle, le risque d'erreur conditionnel de type 2 est une fonction de θ exprimé par

$$\begin{aligned}\beta(\theta|s) &= P_\theta(B(x) > \psi(1)|S(X) = s) \\ &= \frac{f^*_\theta(\psi(s))|\frac{d}{ds}\psi(s)|}{f^*_\theta(s) + f^*_\theta(\psi(s))|\frac{d}{ds}\psi(s)|.}\end{aligned}$$

Soit $p_1(\theta|s)$ la densité a posteriori de ϑ conditionnellement à H_1 et $S(X) = s$, Berger et Coll. (1997) montrent que le risque conditionnel de type 2 moyenné par $p_1(\theta|s)$ sur Θ_1 s'écrit

$$\begin{aligned}\beta^*(s) &= E^{p_1(\theta|s)}[\beta(\theta|s)] \\ &= \int_{\Theta_1} \beta(\theta|s)p_1(\theta|s)d\theta \\ &= \frac{1}{1 + B(s)}.\end{aligned}$$

L'égalité $\beta^*(s) = \beta(B)$ est donc maintenue lorsque l'hypothèse H_1 est composite.

Dass et Berger (2003) [23] montrent que cette caractéristique est également vérifiée dans le cas général d'un test de deux hypothèses composites disjointes

$$H_0 : \{\theta \in \Theta_0\} \qquad \text{versus} \qquad H_1 : \{\theta \in \Theta_1\} \qquad\qquad \Theta_0 \cup \Theta_1 = \Theta,$$

lorsqu'il existe une structure d'invariance telle que le risque d'erreur conditionnel de type 1 est constant sur Θ_0. Le facteur de Bayes considéré est alors

$$B(x) = \frac{m_0(x)}{m_1(x)}$$

où $m_i(x) = \int_{\Theta_i} p(x|\theta)\pi_i(\theta)d\theta$ est la probabilité marginal de X sous H_i.

Le cas spécifique d'un plan expérimental séquentiel est étudié dans Berger, Boukai et Wang (1999) [6]. En associant la densité marginale $m_{i,k}((k, x))$ à la séquence (k, x) sous H_i, le facteur de Bayes correspondant s'écrit

$$B_k = B_k((k, x)) = m_{0,k}((k, x))/m_{1,k}((k, x)),$$

où $m_{i,k}((k, x)) = \int_{\Theta_i} p((k, x)|\theta)\pi_i(\theta)d\theta$. En notant M la période où l'expérience s'arrête, si $F_0^{B_M}(r) = 1 - F_1^{B_M}(a)$ alors F_i est une fonction inversible et le test BBW peut se généraliser en remplaçant $B(x)$ par B_M.

Berger et Coll. (1999) présentent un test tronqué en $k = K$ fondé sur deux bornes prédéfinies $\{R, A\}$ telle que

$$M = \{\text{premier } k < K \text{ tel que } B_k \notin [R, A]\,;\, \text{sinon } k = K\}.$$

Tant que $M < K$, l'expérience s'arrête si les bornes R ou A sont atteintes par B_k $k \in [1, K-1]$. Les risques conditionnels associés au rejet ou à l'acceptation de H_0 sont respectivement $\alpha^*(B_M) = B_M/(B_M + 1)$ et $\beta^*(B_M) = 1/(B_M + 1)$. Si $M = K$, l'expérience s'arrête et les bornes de rejet ou d'acceptation de H_0 sont les valeurs $r, a \in [R, A]$ définie dans le test BBW (voir (B.5.2)).

Outre la possibilité d'interpréter directement la règle d'arrêt de l'expérience en terme de facteur de Bayes, Berger et Coll. (1999) montrent que ce test fréquentiste conditionnel est optimal et produit une inférence indépendante de la façon dont l'expérience s'arrête. Cette dernière assertion vient infirmer les fondements de l'inférence fréquentiste.

B.5.3 Test Conditionnel d'une proportion

Le test BBW précédemment mentionné a été développé en supposant que les distributions sous H_0 et H_1 sont continues et admettent une densité. Dans le cas discret, Dass (2001) [22] introduit une $v.a.$

$$\tilde{Y} = Y + U,$$

où U est une $v.a.$ uniforme indépendante de Y.

La $v.a.$ $\tilde{Y}$ admet une densité sur l'intervalle $[0, n+1[$ sous H_0 et H_1. La prise en considération des valeurs de $\tilde{Y}$ nécessite l'extension de la définition de $B(y)$ à

$$B(\tilde{y}) = B(y) \qquad \text{si } y = [\tilde{y}]_,$$

où $[]_$ désigne la partie entière.

En supposant que $Y \sim Bin(n, \theta)$, le test des hypothèses ponctuelles

$$H_0 : \{\theta = \theta_0\} \qquad \text{versus} \qquad H_1 : \{\theta = \theta_1\} \qquad \theta_0 < \theta_1,$$

conduit à considérer le facteur de Bayes

$$\begin{aligned} B(y) &= \frac{Bin(y; n, \theta_0)}{Bin(y; n, \theta_1)} \\ &= \left(\frac{1-\theta_0}{1-\theta_1}\right)^n \left(\frac{\theta_0/(1-\theta_0)}{\theta_1/(1-\theta_1)}\right)^y \qquad\qquad y = 0, 1, .., n. \end{aligned}$$

Les fonctions de répartition de $B(Y)$ et Y s'écrivent

$$F_0^{B(Y)}(B(y)) = F_0^Y(y) = \sum_{k=0}^{y-1} Bin(k; n, \theta_0) \qquad \text{sous } H_0$$

$$1 - F_1^{B(Y)}(B(y)) = 1 - F_1^Y(y) = \sum_{k=y}^{n} Bin(k; n, \theta_1) \qquad \text{sous } H_1.$$

La fonction B est décroissante en $\tilde{y}$. La règle du test BBW établie dans (B.5.2) s'exprime indifféremment en terme de $B(\tilde{y})$ ou $\tilde{y}$.

Si $B(\tilde{y}) \leq r$ ou $\tilde{y} \geq y_r$,

H_0 est rejetée avec un risque conditionnel $\alpha^*(B) = B(\tilde{y})/(1 + B(\tilde{y})) = \alpha(\tilde{Y})$.

Si $r > B(\tilde{y}) > a$ ou $y_a < \tilde{y} < y_r$, pas de décision.

Si $B(\tilde{y}) \geq a$ ou $\tilde{y} \leq y_a$,

H_0 est acceptée avec un risque conditionnel $\beta^*(B) = B(\tilde{y})/(1 + B(\tilde{y})) = \beta(\tilde{Y})$.

En définissant

$$\begin{aligned} y^* &= min\{y : B(y) < 1, y \in \{0, 1, .., n\}\} \\ y_r &= min\{y : F_1^Y(y+1) \leq F_0^Y(y^*), y \in \{0, 1, .., n\}\} \\ q &= (F_0^Y(y^*) - F_1^Y(y_r + 1))/Bin(y_r; \theta_1, n) \\ y_a &= max\{y : F_0^Y(y) \leq F_1^Y(y^*), y \in \{0, 1, .., n\}\} \\ p &= (F_1^Y(y^*) - F_0^Y(y_a))/Bin(y_a; \theta_0, n), \end{aligned}$$

les valeurs de $\tilde{y}_r$ et $\tilde{y}_a$ sont déterminées par

$$\tilde{y}_a = y_a = y^* \text{ et } \tilde{y}_r = \psi^{-1}(\tilde{y}_a) = y_r + 1 - q \quad \text{si } F_0^Y(y^*) \leq F_1^Y(y^*), \quad \text{(B.5.3)}$$

$$\text{et} \qquad \tilde{y}_a = \psi(\tilde{y}_r) = y_a + p \text{ et } \tilde{y}_r = y_r = y^* \qquad \text{si } F_0^Y(y^*) > F_1^Y(y^*). \quad \text{(B.5.4)}$$

Il en résulte une valeur non-entière pour $\tilde{y}_a$ ou $\tilde{y}_r$. La transposition de la règle du test BBW à $Y = \{0, 1, .., n\}$ nécessite l'introduction d'une assertion probabiliste dans l'interprétation du ***test de Dass*** qui s'exprime par

Si $y < y_r$, H_0 est rejetée avec un risque conditionnel $\alpha^*(B)$,

si $y = y_r$ et $\tilde{y}_r \neq y_r$

. il y a une probabilité $1 - q$ de rejeter H_0 avec un risque conditionnel $\alpha^*(B)$

. il y a une probabilité q que y tombe dans la zone de non-décision,

si $y_r < y < y_a$, pas de décision,

si $y = y_a$ et $\tilde{y}_a \neq y_a$

. il y a une probabilité p d'accepter H_0 avec un risque conditionnel $\beta^*(B)$

. il y a une probabilité $1 - p$ que y tombe dans la zone de non-décision,

Si $y > y_a$, H_0 est acceptée avec un risque conditionnel $\beta^*(B)$.

(B.5.5)

Du fait de la nature discrète de Y, $B(y)$ n'est pas toujours une fonction monotone en y. On définit alors l'ensemble $\mathcal{B} = \{.., b_{-2}, b_{-1}, b_0, b_1, b_2, ..\}$ des valeurs de $B(y)$ classées par ordre décroissant. Soit la *va* Z telle que $Z = k$ si et seulement si $b = b_k$ ($b \in \mathcal{B}$), on déduit les densités prédictives associées f_0^Z et f_1^Z des valeurs de $m_i(y)$ respectivement sous H_0 et H_1. La *va* $\tilde{Z} = Z + U$ admet une densité sous H_0 et H_1 et permet d'utiliser la procédure définie dans le cas de deux hypothèses simples.

Cette transformation de variable permet de considérer le cas des hypothèses composites. Comme pour les données continues, on remplace la probabilité d'échantillonnage $p(z|\theta_i)$ dans le cas d'une hypothèse ponctuelle $H_i : \{\theta = \theta_i\}$ par la probabilité marginale $m_i(z)$ dans le cas d'une hypothèse composite $H_i : \{\theta \in \Theta_i\}$. L'utilisation

d'un a priori du type $\pi_i(\theta) \sim Be(a,b)$ sur le sous-espace $\Theta_i = [dec, dec+e]$ conduit à l'écriture

$$\begin{aligned} m_i(y) &= \int_{dec}^{dec+e} p(y|\theta)\pi_i(\theta)d\theta \\ &= C\binom{n}{y}\int_0^1 \phi^{a-1}(1-\phi)^{b-1}(dec+e\phi)^y(1-dec-e\phi)^{n-y}d\phi. \end{aligned}$$

où C est la constante de normalisation (voir (2.5.3)).

On note que si $dec = 0$, cette expression se simplifie en

$$\begin{aligned} m_i(y) &= \int_0^e p(y|\theta)\pi_i(\theta)d\theta \\ &= \binom{n}{y}\frac{e^y}{Be(a,b)}\int_0^1 \phi^{a-1}(1-\phi)^{b-1}(e\phi)^y(1-e\phi)^{n-y}d\phi \\ &= \binom{n}{y}\frac{e^y}{Be(a,b)}\int_0^1 \phi^{y+a-1}(1-\phi)^{b-1}\sum_{j=0}^{n-y}\binom{n-y}{j}(-e\phi)^j d\phi \\ &= \binom{n}{y}\frac{1}{Be(a,b)}\sum_{j=0}^{n-y} e^{y+j}(-1)^j\binom{n-y}{j}\int_0^1 \phi^{y+a+j-1}(1-\phi)^{b-1}d\phi \\ &= \binom{n}{y}\sum_{j=0}^{n-y}(-1)^j\binom{n-y}{j}e^{y+j}\frac{\Gamma(a+b)\Gamma(y+a+j)}{\Gamma(a)\Gamma(a+b+y+j)}. \end{aligned}$$

Ainsi, l'a priori de Jeffreys classique $\pi_i(\theta) \sim Be(0.5, 0.5)$ défini sur $\Theta_i = [0, e[$ sous H_i conduit à la densité marginale

$$m_i(y) = \binom{n}{y}\sum_{j=0}^{n-y}(-1)^j\binom{n-y}{j}e^{y+j}\frac{\Gamma(y+0.5+j)}{\sqrt{\pi}\,\Gamma(1+y+j)}.$$

B.5.4 Calibration de la fonction 'p' non-conditionnelle

En écho aux critiques sur l'utilisation de la valeur de 'p' comme une mesure quantitative en faveur ou contre l'hypothèse H_0, Sellke, Bayarri et Berger (2001) [48] proposent une calibration sur une échelle qui la rend directement interprétables. Si l'on note $T(X)$ la statistique d'un test bilatéral, la *va* $p = p(X)$ est définie par

$$p(X) = 2[1 - \Phi(T(X))],$$

où Φ est la fonction quantile inverse pour la loi Normale.

En notant $f(p|\xi)$ sa densité sous l'alternative (ξ est un paramètre inconnu), les hypothèses s'expriment par

$$H_0 : \{p \sim U_{[0,1]}\} \qquad \text{versus} \qquad H_1 : \{p \sim f(p|\xi)\}.$$

La famille de loi $Be(\xi, 1)$ pour $p(X)$ sous l'alternative permet également de reformuler l'hypothèse $H_0 : \{p \sim Be(\xi, 1)\,; \xi = 1\}$. En notant $\eta(\xi)$ la densité de ξ sous H_1, le facteur de Bayes de H_0 sur H_1 est donné par

$$B_\eta(p) = \frac{f(p\,|\,\xi = 1)}{\int_0^1 f(p\,|\,\xi)\eta(\xi)d\xi}.$$

On obtient la valeur minimum de $B_\eta(p)$ par

$$\begin{aligned} \underline{B} = inf_{\forall\eta} B_\eta(p) &= \frac{f(p\,|\,\xi = 1)}{Sup_\xi \xi p^{\xi-1}} \\ &= -epLog(p) && \text{si } p < e^{-1} \\ &= 1 && \text{sinon.} \end{aligned}$$

On déduit la borne inférieure de $\alpha^*(B)$ et la borne supérieure de $\beta^*(B)$ en utilisant la relation établie avec le facteur de Bayes

$$\underline{\alpha}^*(B) = \frac{\underline{B}}{1 + \underline{B}} \qquad \text{et} \qquad \overline{\beta}^*(B) = \frac{1}{1 + \underline{B}}.$$

La fonction $-epLog(p)$ s'interprète comme l'évaluation du facteur de Bayes la moins favorable à H_0. C'est une calibration biaisée par nature qui peut être utilisée lors d'une première exploration. Les procédures BBW ou de Dass sont à privilégier si l'on veut tenir compte d'une alternative spécifique.

L'utilisation de cette famille d'alternative permet de montrer que le conditionnement du test BBW peut se faire sur les valeurs de 'p' respectivement sous H_0 et H_1. Dans le cas d'un test de deux hypothèses ponctuelles, l'hypothèse alternative peut être formulée par $H_1 : \{p \sim Be(\xi, 1)\,; \xi = \xi_1\}$. Le facteur de Bayes s'écrit

$$B(p) = \xi^{-1} p^{1-\xi}$$

qui est une fonction strictement croissante de p. Le conditionnement du test BBW sur

$$min\{B(X), \Psi(B(X))\}$$

équivaut donc au conditionnement sur

$$min\{p_{H_0}, p_{H_1}\}$$

où p_{H_0} et p_{H_1} sont les valeurs de 'p' respectivement sous H_0 et H_1.

Annexe C

Figures et Tableaux

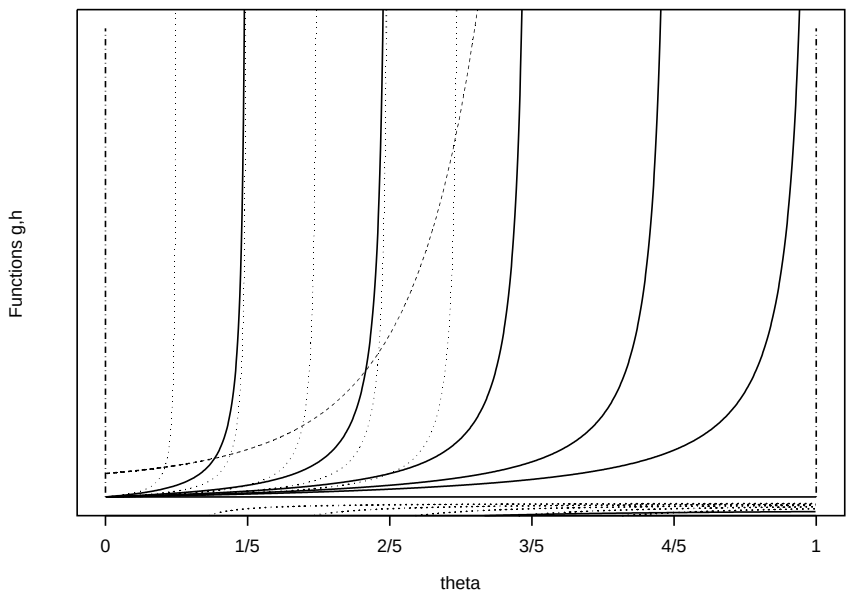

FIG. C.1 – *Fonctions holomorphes* $g(\theta)$ *(- - -),* $h(\theta)$ *(—) dans* $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 1, 2, .., 5\}$ *et* $h(\theta)$ *(· · ·) dans* $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 5\}$ *pour le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

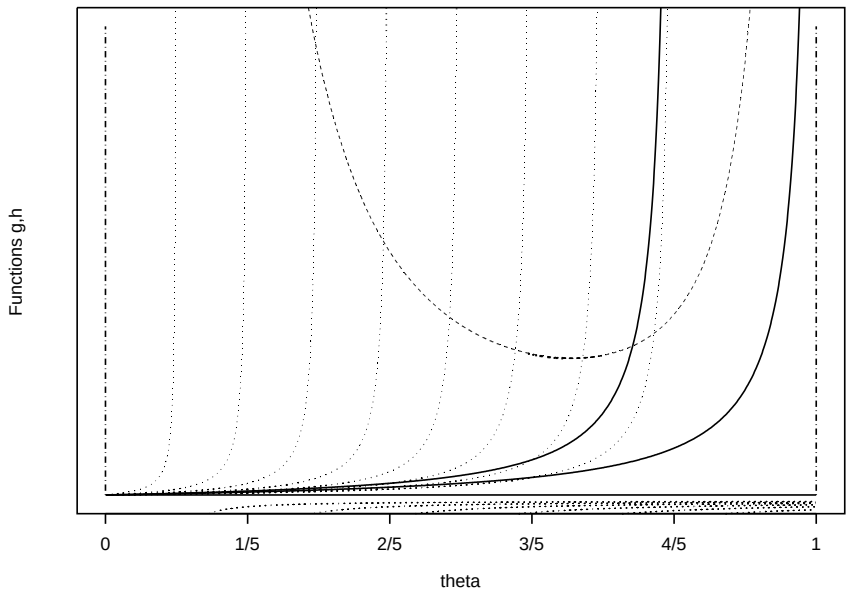

FIG. C.2 – *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) dans* $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 4, 5\}$ *et* $h(\theta)$ *(· · ·) dans* $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 8\}$ *pour le plan séquentiel* $(n_1 = 5, S_{1,U} = 4, n_2 = 5)$

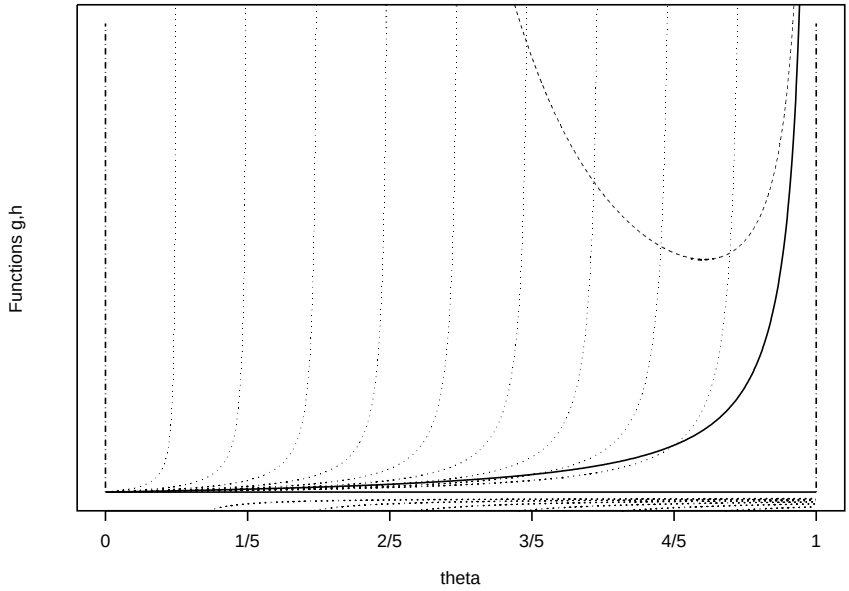

FIG. C.3 – *Fonctions holomorphes* $g(\theta)$ *(- - -) et* $h(\theta)$ *(—) dans* $\mathcal{Y}_{(1,y)}=\{y=y_1; y=5\}$ *et* $h(\theta)$ $(\cdots)$ *dans* $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 9\}$ *pour le plan séquentiel* $(n_1=5, S_{1,U}=5, n_2=5)$

	$\hat{\theta}^{MV}_{(m,y)}$	$\hat{\theta}^{ModC}_{(m,y)}$	$\hat{\theta}^{ModCA}_{(m,y)}$	$\hat{\theta}^{VUMSB}_{(m,y)}$
(1,3)	0.600	0.536	0.525	0.600
(1,4)	0.800	0.769	0.774	0.800
(1,5)	1	1	1	1
(2,0)	0	0	0	0
(2,1)	0.100	0.099	0.099	0.100
(2,2)	0.200	0.194	0.194	0.200
(2,3)	0.300	0.286	0.285	0.273
(2,4)	0.400	0.377	0.374	0.323
(2,5)	0.500	0.470	0.467	0.357
(2,6)	0.600	0.568	0.566	0.382
(2,7)	0.700	0.673	0.674	0.400

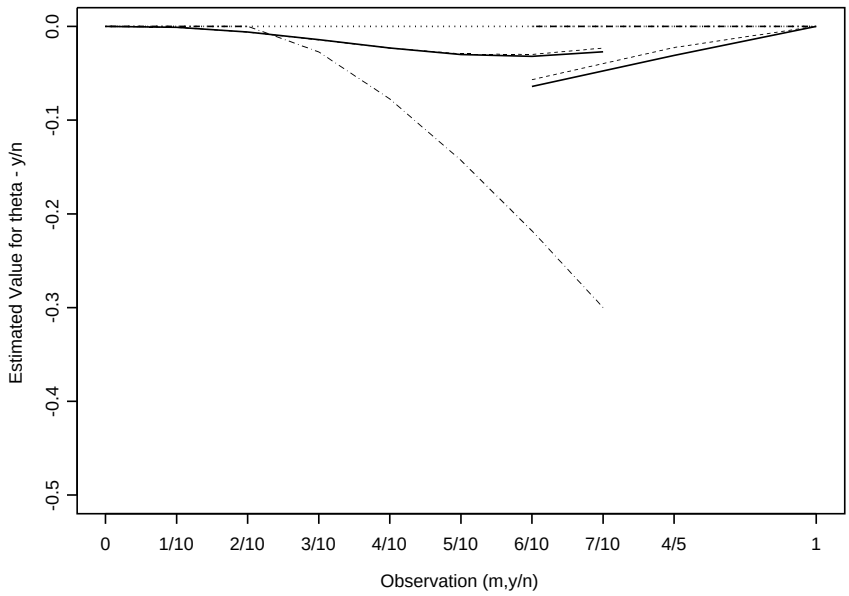

FIG. C.4 – *Estimation de* $\hat{\theta}^{ModC}_{(m,y)}$ *(—),* $\hat{\theta}^{ModCA}_{(m,y)}$ *(- - -) et* $\hat{\theta}^{VUMSB}_{(m,y)}$ *(· – ·), soustraite de* $\hat{\theta}^{MV}_{(m,y)}$*, dans* $\mathcal{Y}_{(1,y)}=\{y=y_1; y=3,4,5\}$ *et* $\mathcal{Y}_{(2,y)}=\{y=y_1+y_2; y=0,1,..,7\}$ *pour le plan séquentiel* $(n_1=5, S_{1,U}=3, n_2=5)$

	$\hat{\theta}^{MV}_{(m,y)}$	$\hat{\theta}^{ModC}_{(m,y)}$	$\hat{\theta}^{ModCA}_{(m,y)}$	$\hat{\theta}^{VUMSB}_{(m,y)}$
(1,4)	0.800	0.742	0.730	0.800
(1,5)	1	1	1	1
(2,0)	0	0	0	0
(2,1)	0.100	0.100	0.099	0.100
(2,2)	0.200	0.200	0.199	0.200
(2,3)	0.300	0.300	0.296	0.300
(2,4)	0.400	0.390	0.390	0.390
(2,5)	0.500	0.465	0.482	0.465
(2,6)	0.600	0.523	0.573	0.523
(2,7)	0.700	0.567	0.667	0.567
(2,8)	0.800	0.600	0.770	0.600

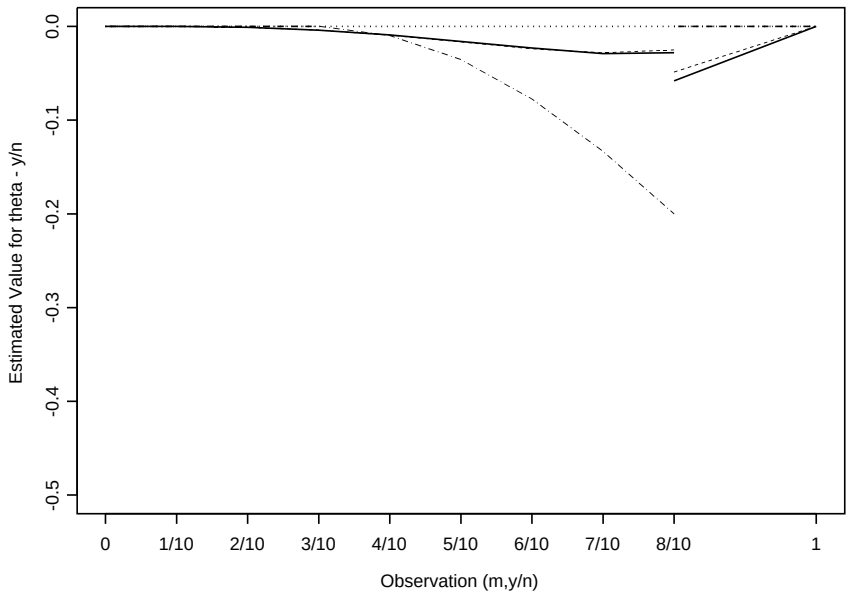

FIG. C.5 – *Estimation de* $\hat{\theta}^{ModC}_{(m,y)}$ *(—),* $\hat{\theta}^{ModCA}_{(m,y)}$ *(- - -) et* $\hat{\theta}^{VUMSB}_{(m,y)}$ *(· – ·), soustraite de* $\hat{\theta}^{MV}_{(m,y)}$*, dans* $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 4, 5\}$ *et* $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 8\}$ *pour le plan séquentiel* $(n_1 = 5, S_{1,U} = 4, n_2 = 5)$

	$\hat{\theta}^{MV}_{(m,y)}$	$\hat{\theta}^{ModC}_{(m,y)}$	$\hat{\theta}^{ModCA}_{(m,y)}$	$\hat{\theta}^{VUMSB}_{(m,y)}$
(1,5)	1	1	1	1
(2,0)	0	0	0	0
(2,1)	0.100	0.100	0.100	0.100
(2,2)	0.200	0.200	0.200	0.200
(2,3)	0.300	0.300	0.300	0.300
(2,4)	0.400	0.398	0.399	0.400
(2,5)	0.500	0.496	0.496	0.498
(2,6)	0.600	0.592	0.592	0.590
(2,7)	0.700	0.687	0.686	0.673
(2,8)	0.800	0.781	0.778	0.743
(2,9)	0.900	0.878	0.874	0.800

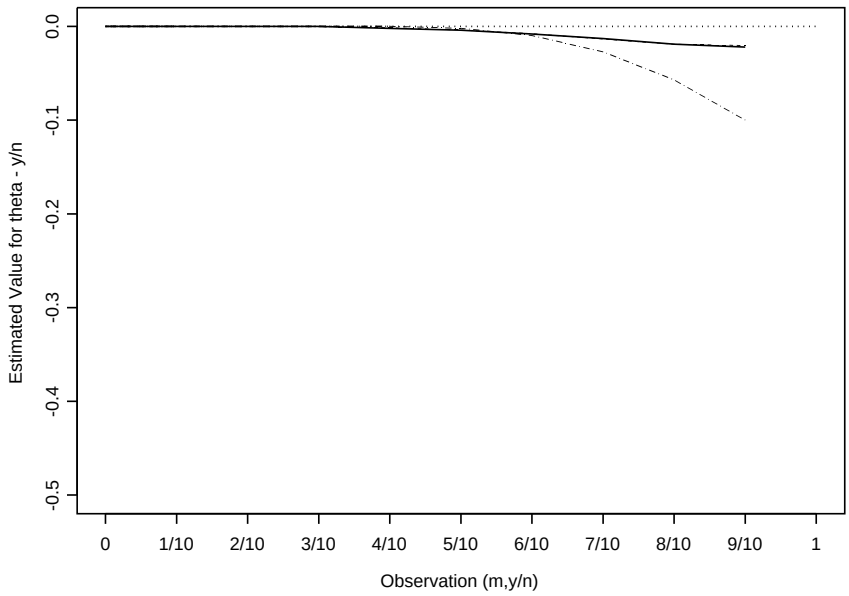

FIG. C.6 – *Estimation de* $\hat{\theta}^{ModC}_{(m,y)}$ *(—),* $\hat{\theta}^{ModCA}_{(m,y)}$ *(- - -) et* $\hat{\theta}^{VUMSB}_{(m,y)}$ *(·−·), soustraite de* $\hat{\theta}^{MV}_{(m,y)}$*, dans* $\mathcal{Y}_{(1,y)} = \{y = y_1; y = 5\}$ *et* $\mathcal{Y}_{(2,y)} = \{y = y_1 + y_2; y = 0, 1, .., 9\}$ *pour le plan séquentiel* $(n_1 = 5, S_{1,U} = 5, n_2 = 5)$

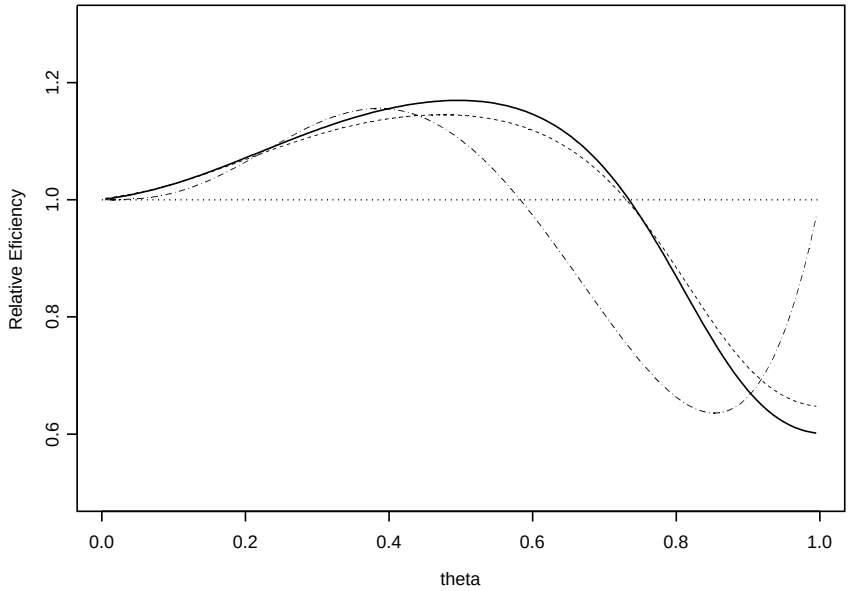

FIG. C.7 – *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{ModCA}$ *(- - -) et* $\hat{\theta}^{VUMSB}$ *(· — ·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 4, n_2 = 5)$

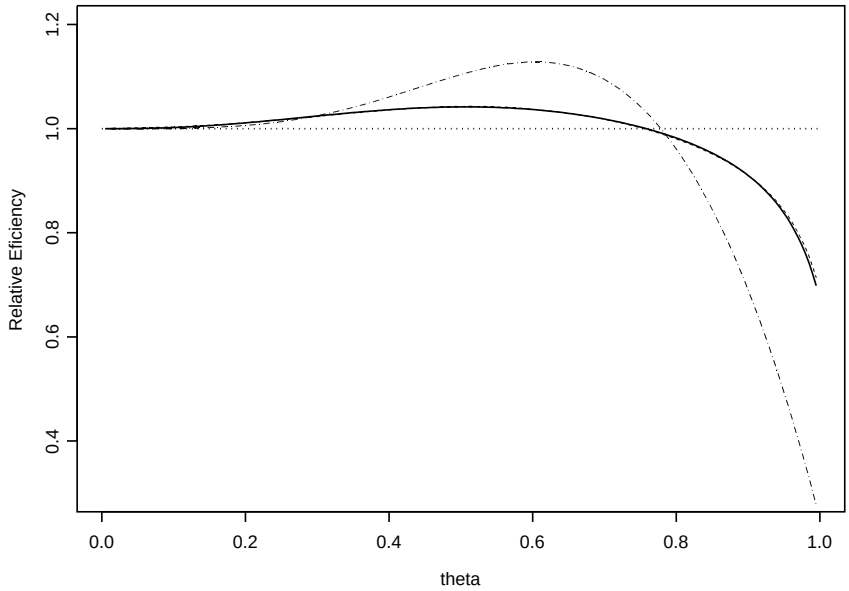

FIG. C.8 – *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{ModCA}$ *(- - -) et* $\hat{\theta}^{VUMSB}$ *(· — ·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 5, n_2 = 5)$

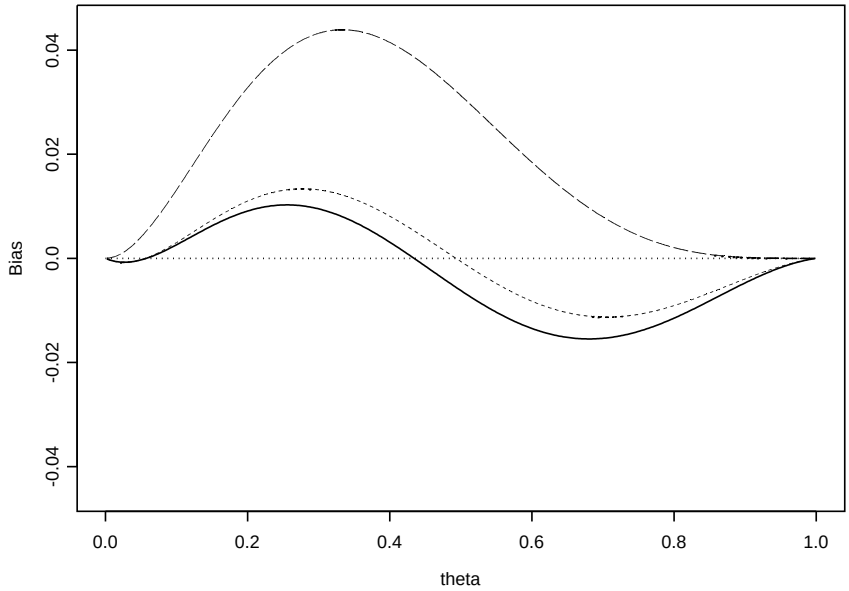

FIG. C.9 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -)dans le plan séquentiel* $(n_1 = 5, S_{1,L} = -1, S_{1,U} = 2, n_2 = 5)$

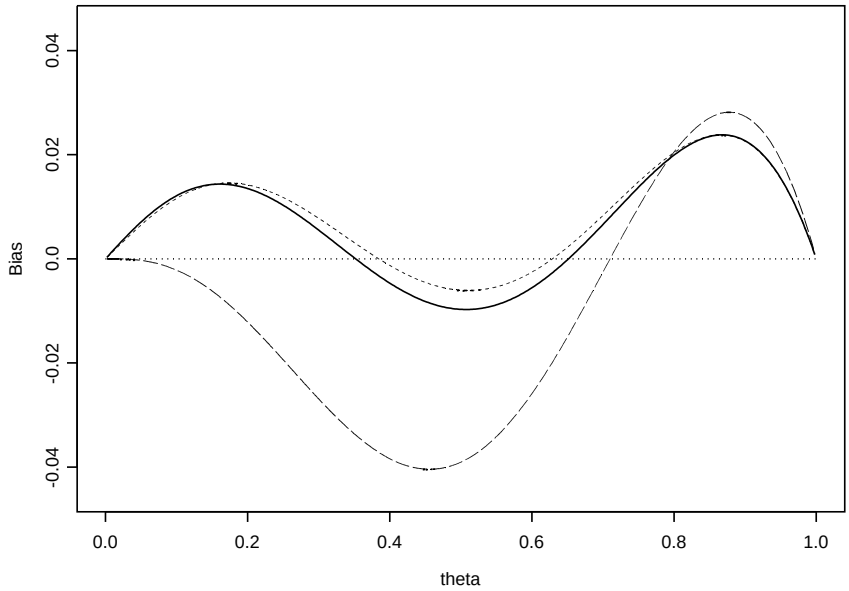

FIG. C.10 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -) dans le plan séquentiel* $(n_1 = 5, S_{1,L} = 2, S_{1,U} = 5, n_2 = 5)$

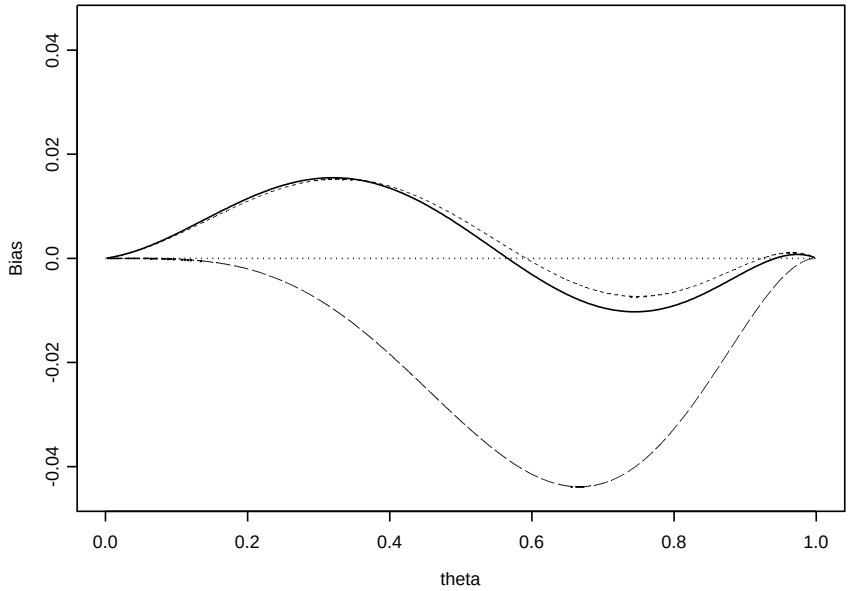

FIG. C.11 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—) et de* $\hat{\theta}^{ModCA}$ *(- - -) dans le plan séquentiel* $(n_1=5, S_{1,L}=3, S_{1,U}=6, n_2=5)$

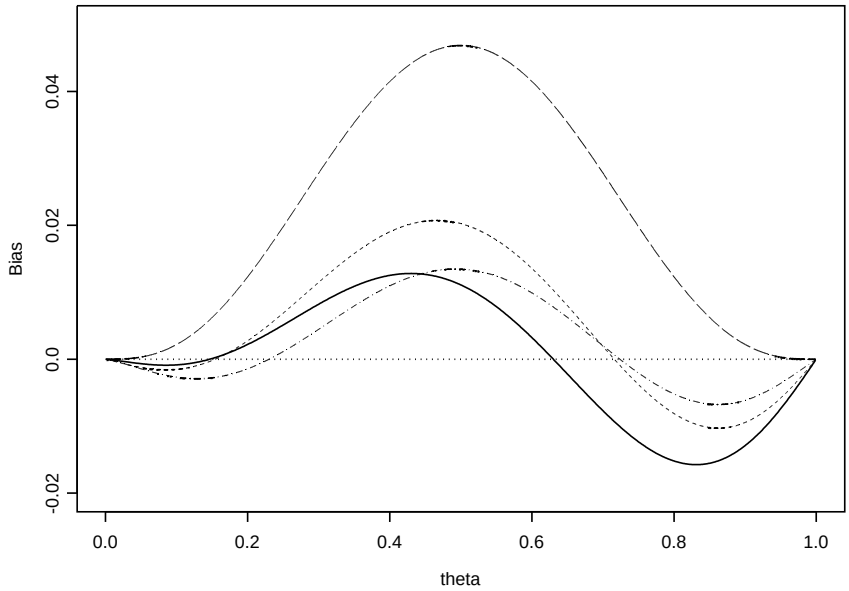

FIG. C.12 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et de* $\hat{\theta}^{Whi}$ *(· – ·) dans le plan séquentiel* $(n_1=5, S_{1,U}=3, n_2=5)$

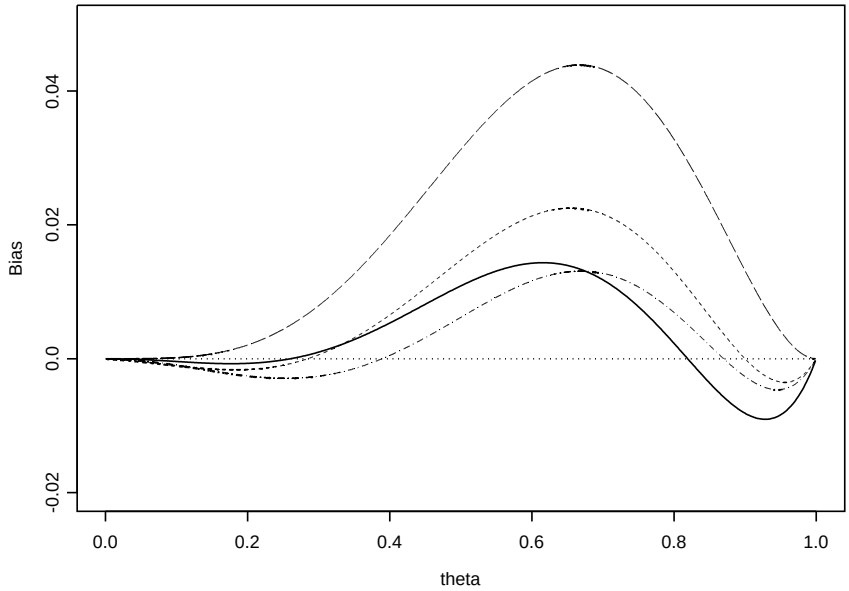

FIG. C.13 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et de* $\hat{\theta}^{Whi}$ *(· – ·) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 4, n_2 = 5)$

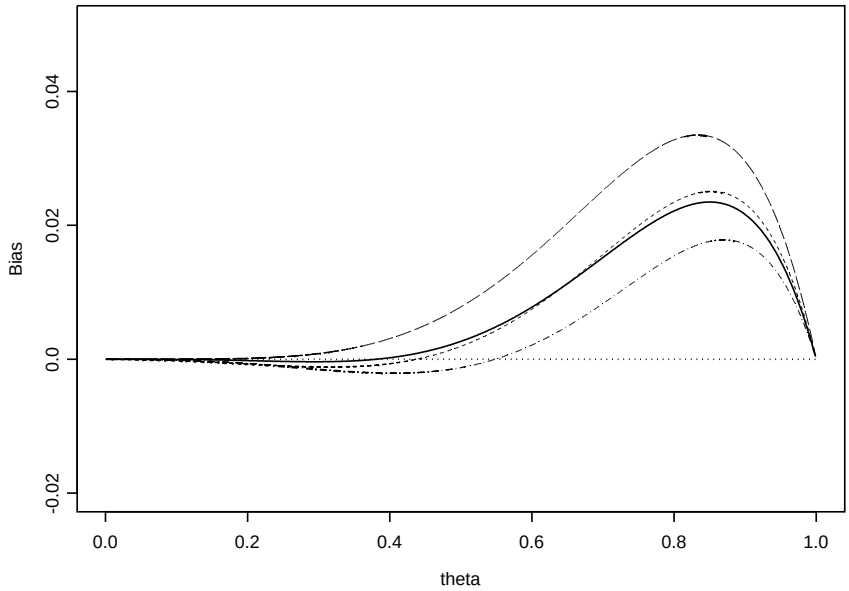

FIG. C.14 – *Biais de* $\hat{\theta}^{MV}$ *(– –),* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et de* $\hat{\theta}^{Whi}$ *(· – ·) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 5, n_2 = 5)$

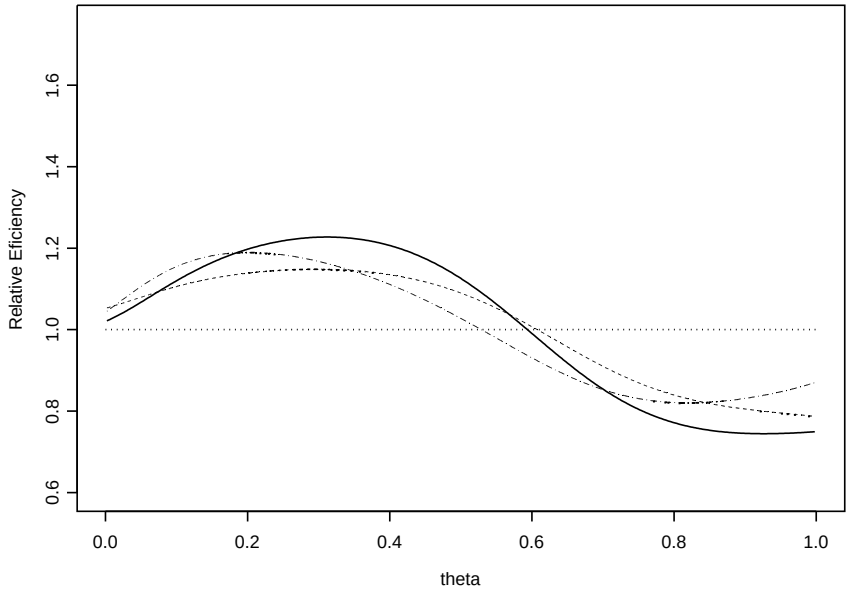

FIG. C.15 – *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et* $\hat{\theta}^{Whi}$ *(·—·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=3, n_2=5)$

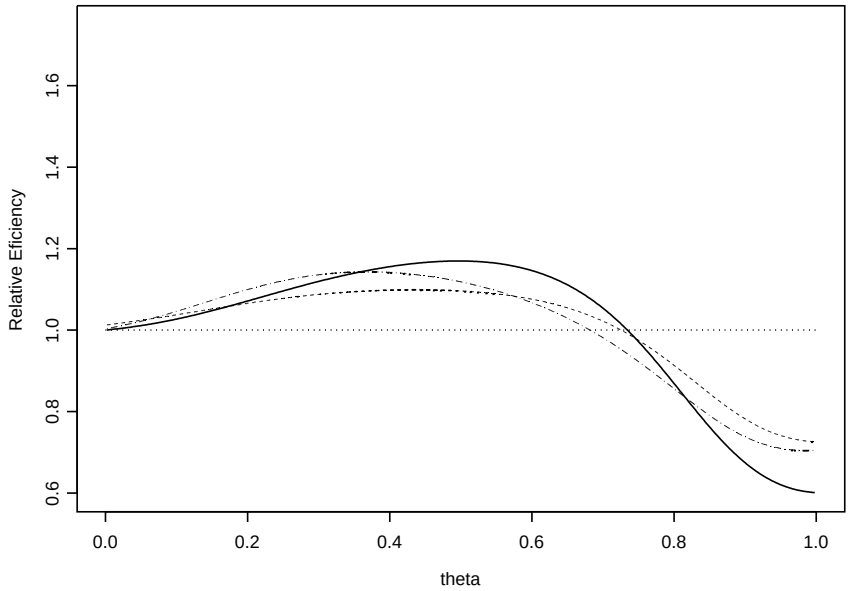

FIG. C.16 – *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et* $\hat{\theta}^{Whi}$ *(·—·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=4, n_2=5)$

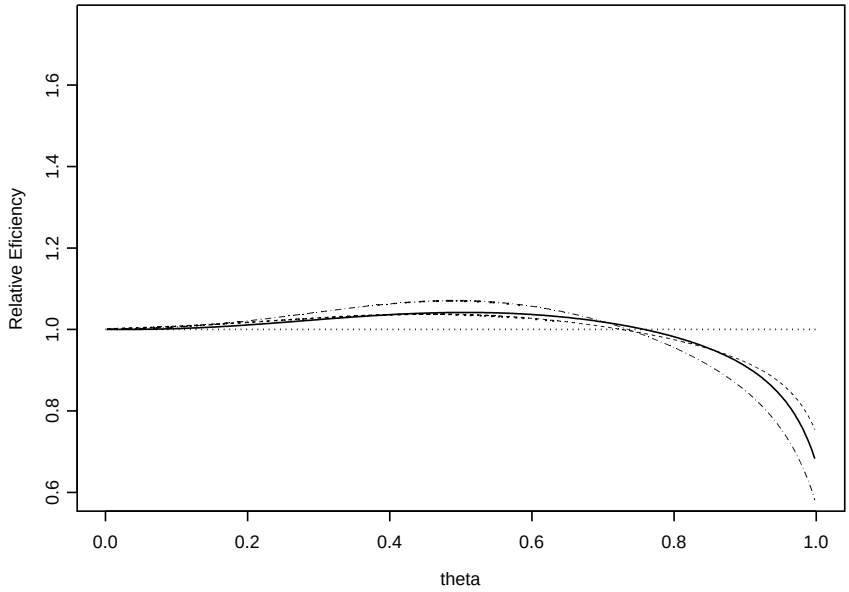

FIG. C.17 – *Efficacité relative de* $\hat{\theta}^{ModC}$ *(—),* $\hat{\theta}^{Mean}$ *(- - -) et* $\hat{\theta}^{Whi}$ *(·—·) versus* $\hat{\theta}^{MV}$ *dans le plan séquentiel* $(n_1=5, S_{1,U}=5, n_2=5)$

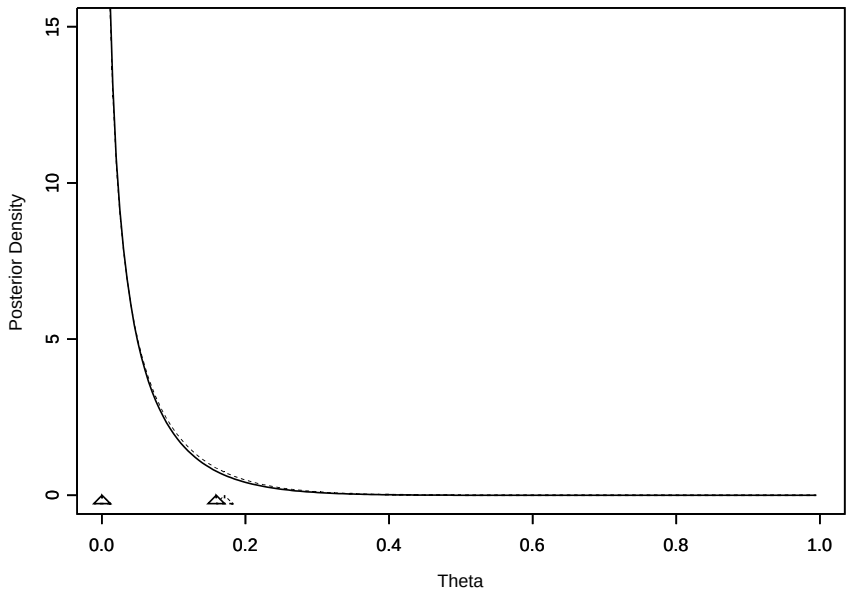

FIG. C.18 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,0) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$

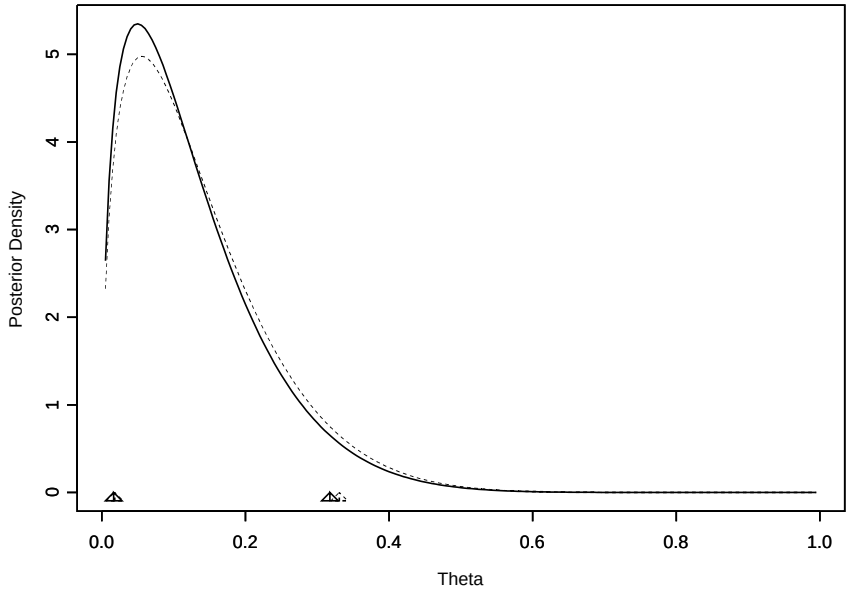

FIG. C.19 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,1) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$

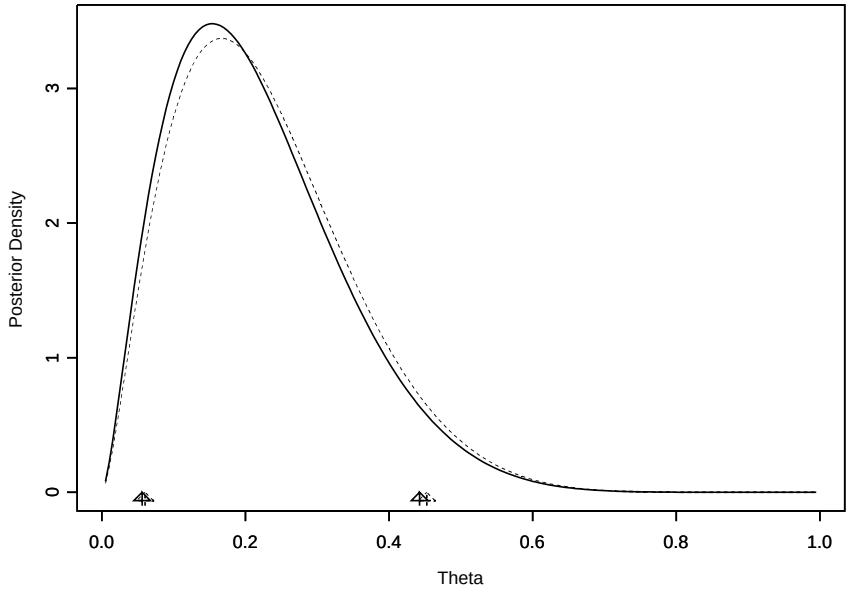

FIG. C.20 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,2) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$

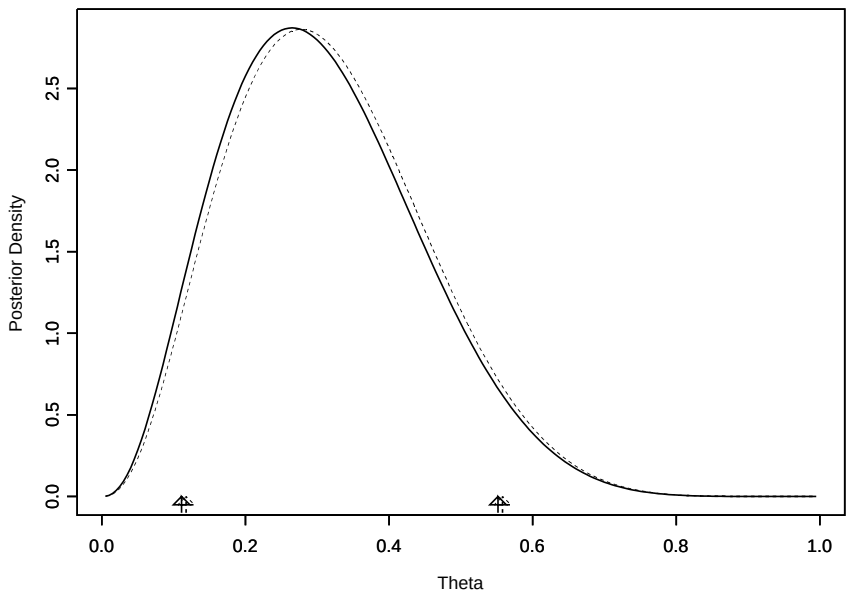

FIG. C.21 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,3) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$

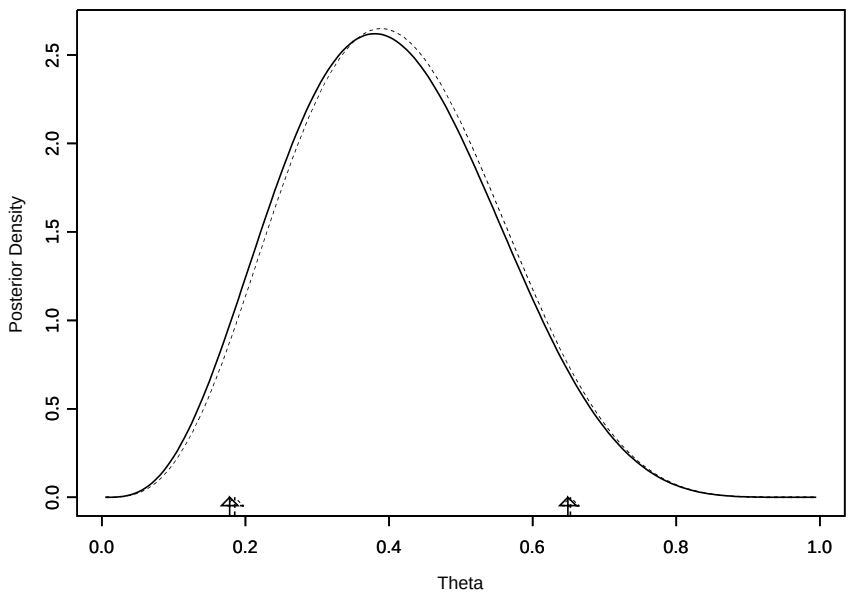

FIG. C.22 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,4) dans le plan séquentiel* $(n_1=5, S_{1,U}=1, n_2=5)$

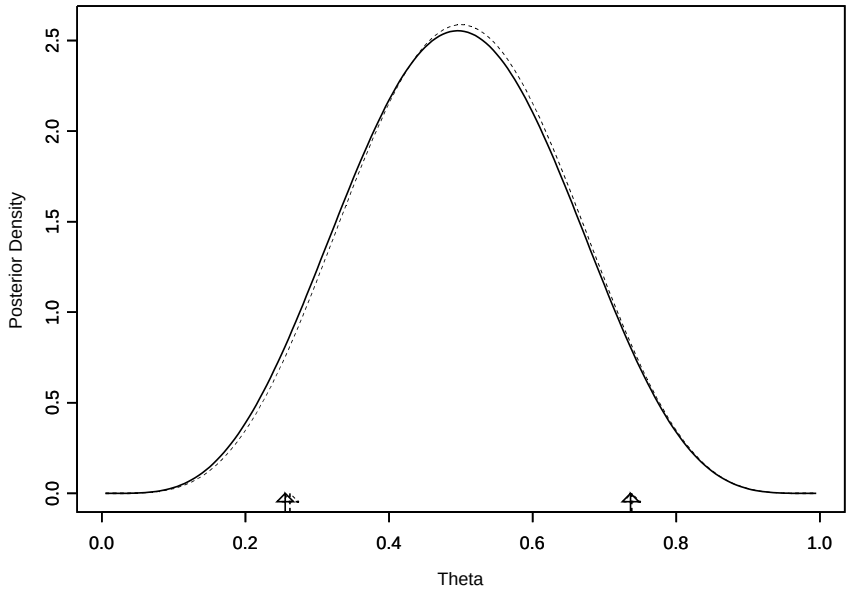

FIG. C.23 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,5) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, n_2 = 5)$

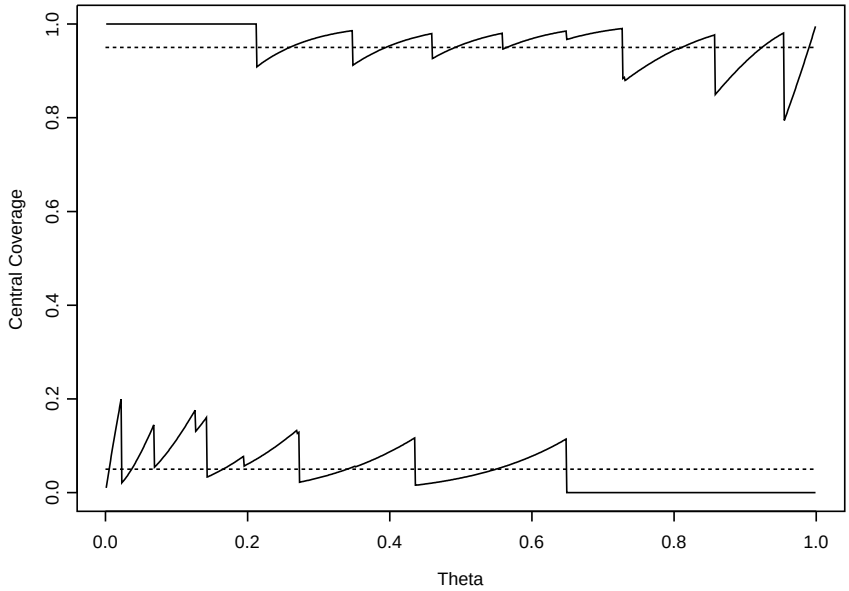

FIG. C.24 – *Taux de recouvrement de l'intervalle de Wilson 90% dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 2, n_2 = 5$)

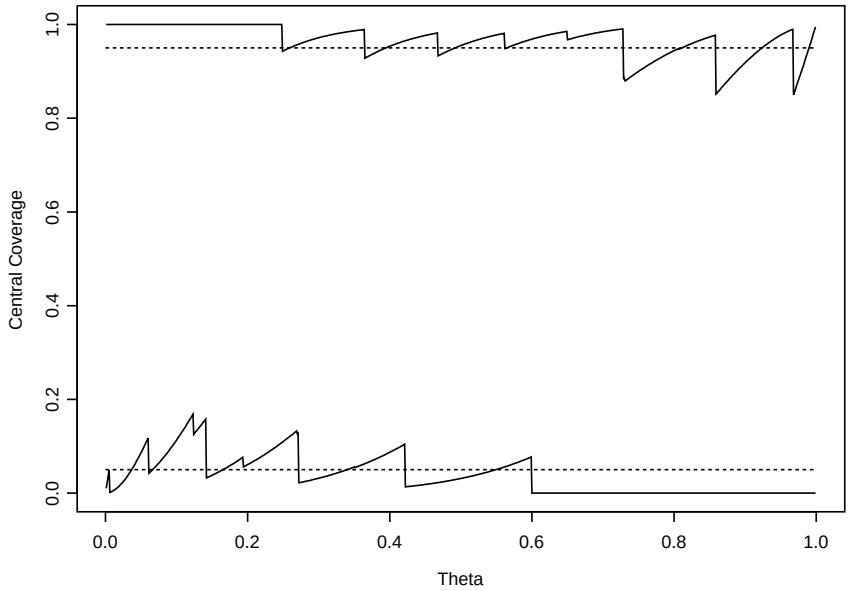

FIG. C.25 – *Taux de recouvrement de l'intervalle d'Agresti 90% dans le plan séquentiel* ($n_1 = 5, S_{1,U} = 2, n_2 = 5$)

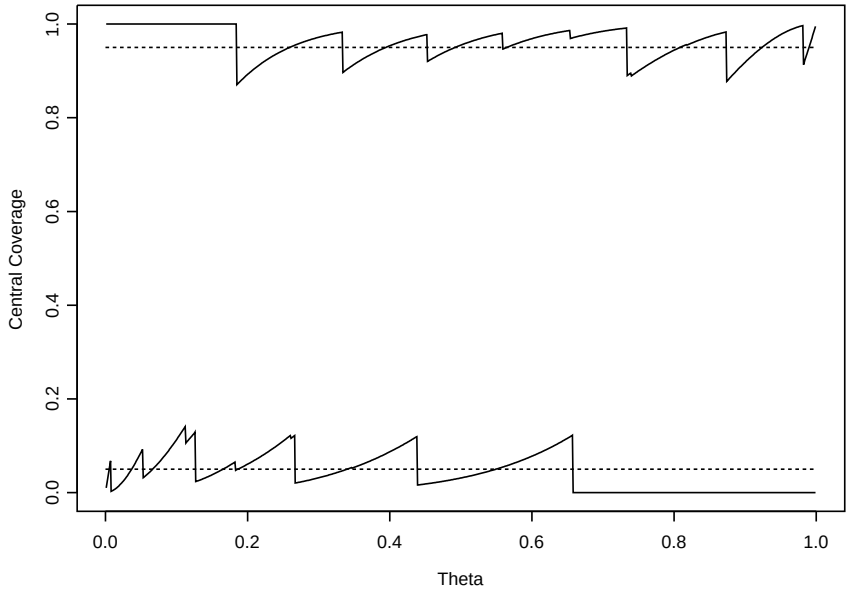

FIG. C.26 – *Taux de recouvrement de l'intervalle du second-ordre 90% dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 2, n_2 = 5)$

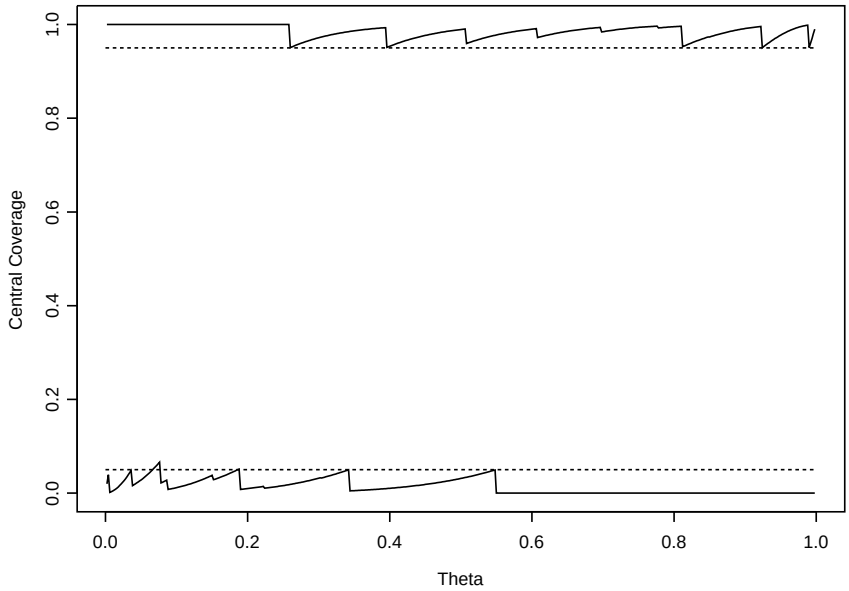

FIG. C.27 – *Taux de recouvrement de l'intervalle de Clopper-Pearson 90% dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 2, n_2 = 5)$

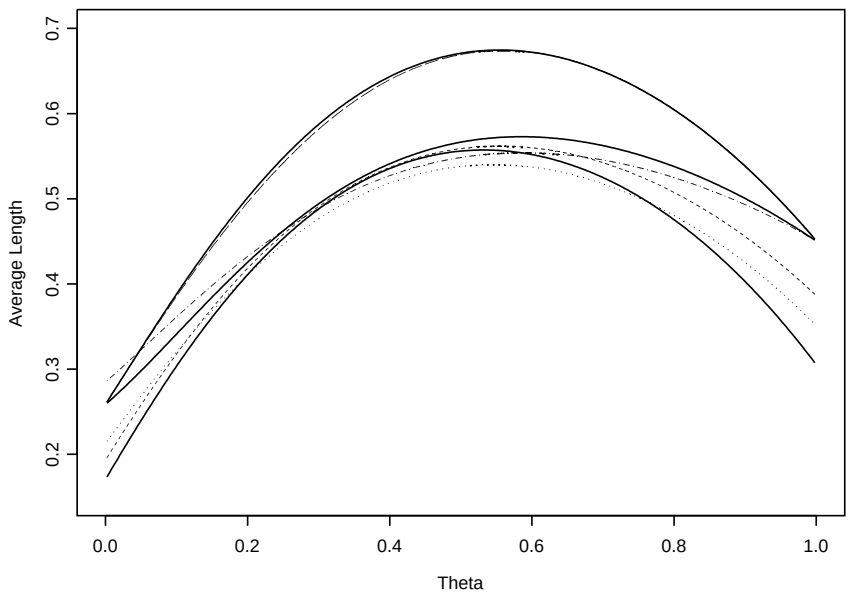

FIG. C.28 – *Longueur moyenne des intervalles de confiance 90% dans le plan séquentiel* $(n_1=5, S_{1,U}=2, n_2=5)$. *Les intervalles* CI_J, CI_{JM} *et* CI_{CP} *(—) sont respectivement représentés de bas en haut, avec également les intervalles* CI_{JT} *(– – –),* CI_W *(· · ·),* CI_{AC} *(· – ·) et* CI_2 *(- - -).*

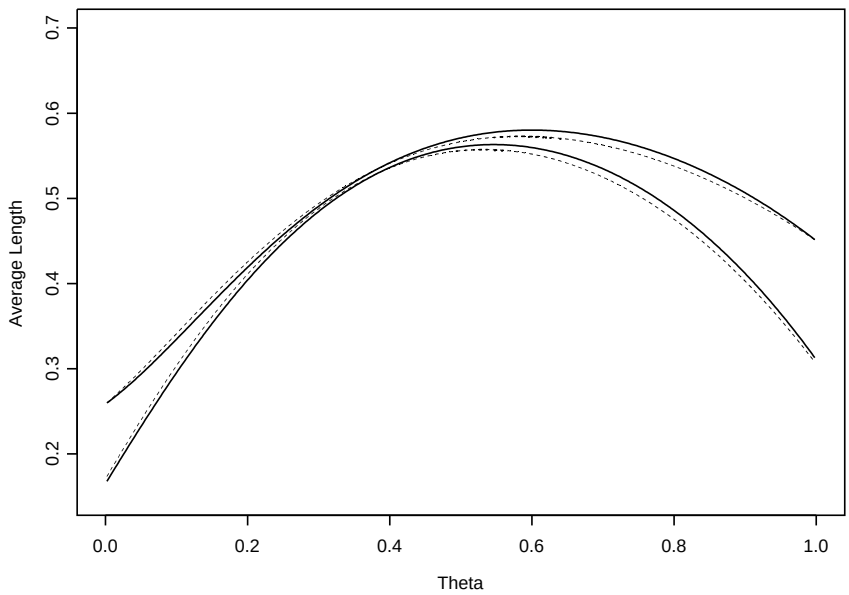

FIG. C.29 – *Longueur moyenne des intervalles de Jeffreys modifiés 90% classique (- - -) et corrigé (—) dans le plan séquentiel* $(n_1=5, S_{1,U}=2, n_2=5)$

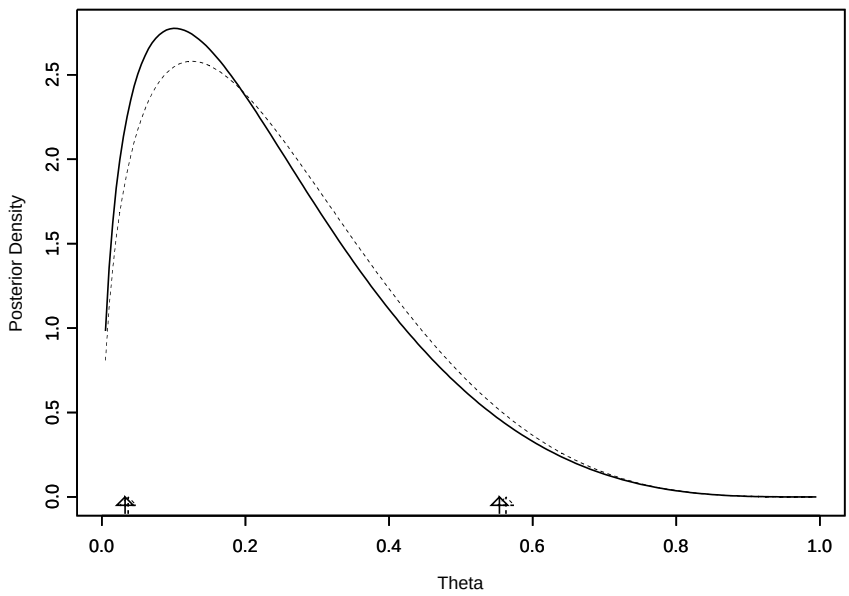

FIG. C.30 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,1) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$

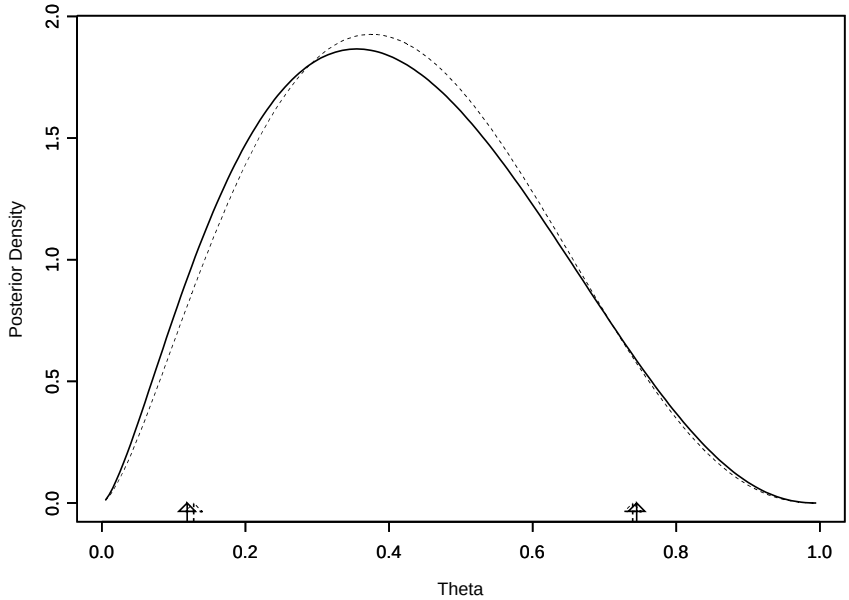

FIG. C.31 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,2) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$

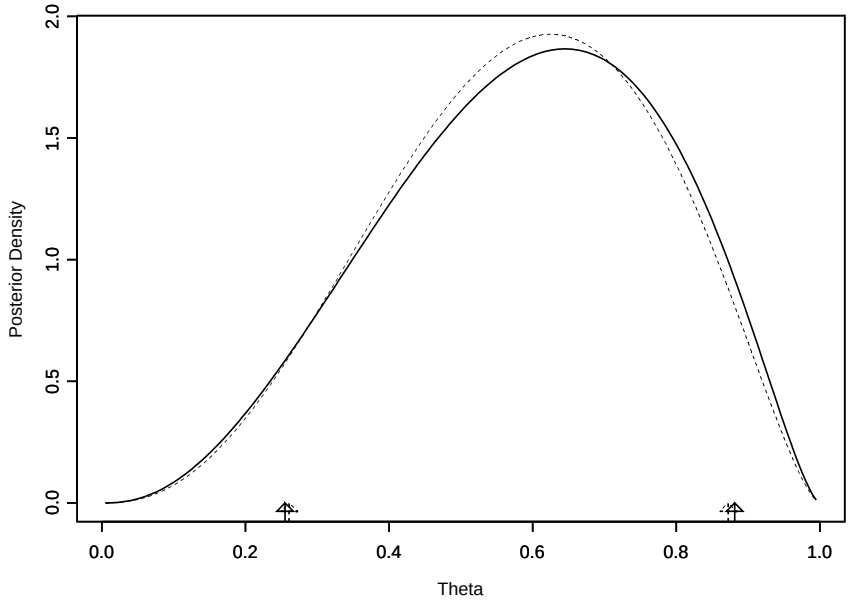

FIG. C.32 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,3) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$

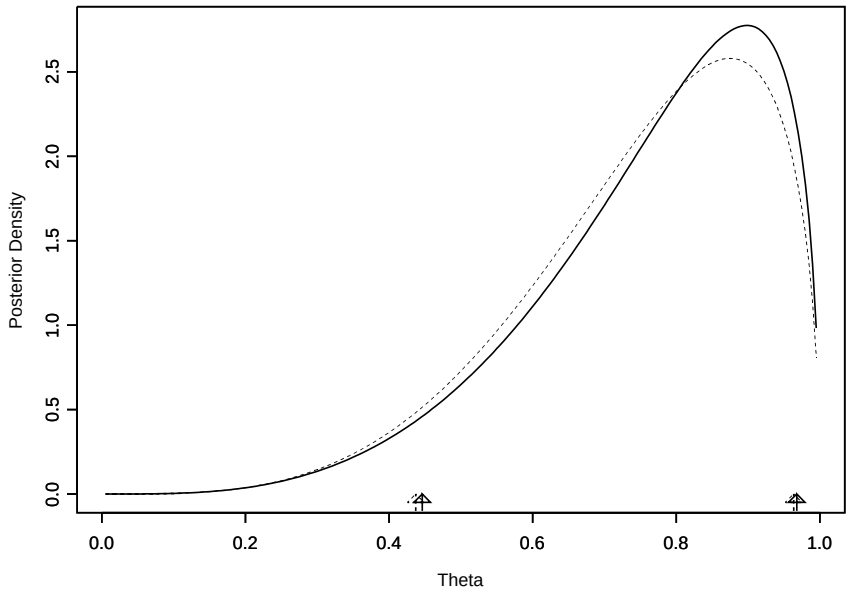

FIG. C.33 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(1,4) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$

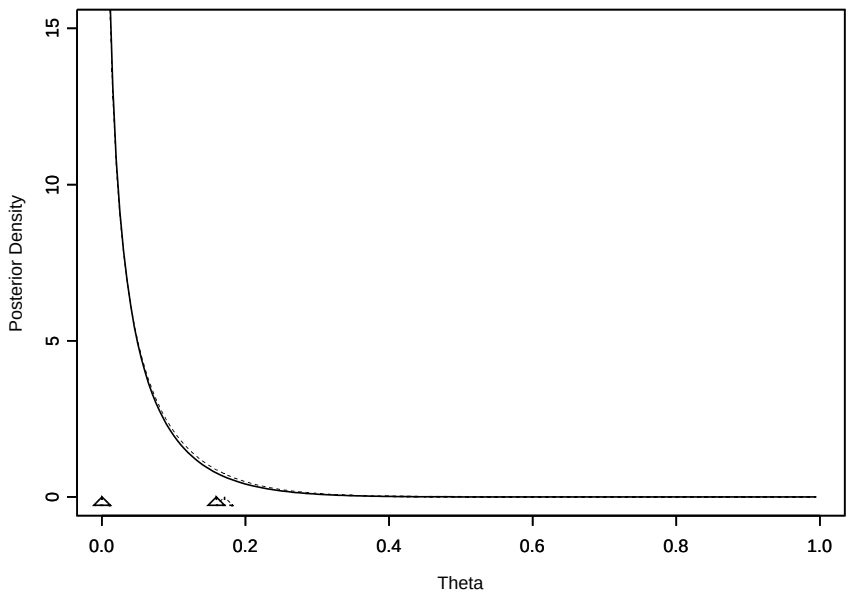

FIG. C.34 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,0) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$

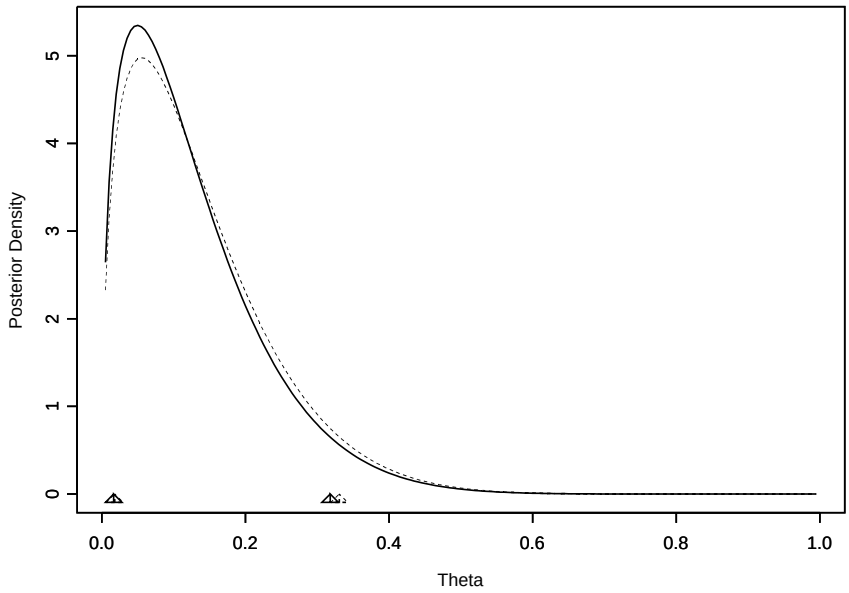

FIG. C.35 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,1) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$

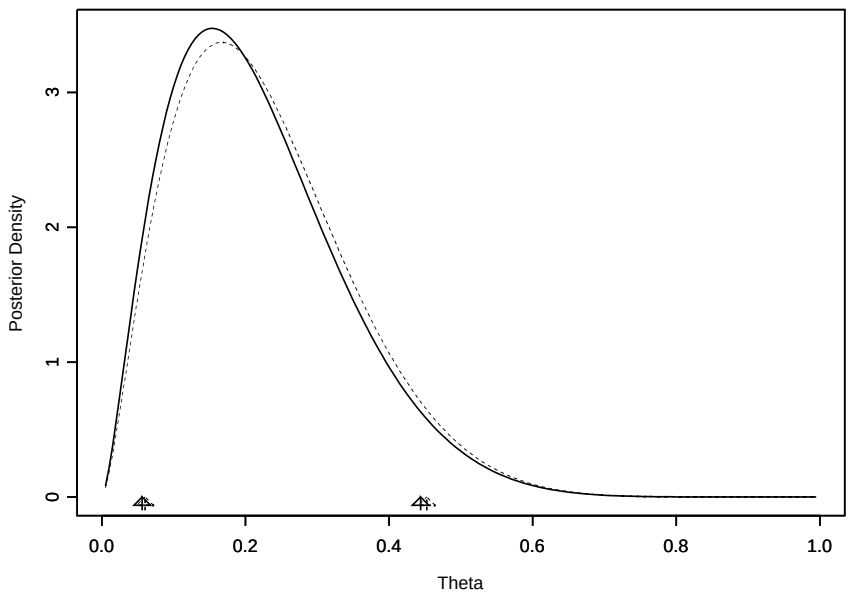

FIG. C.36 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,2) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$

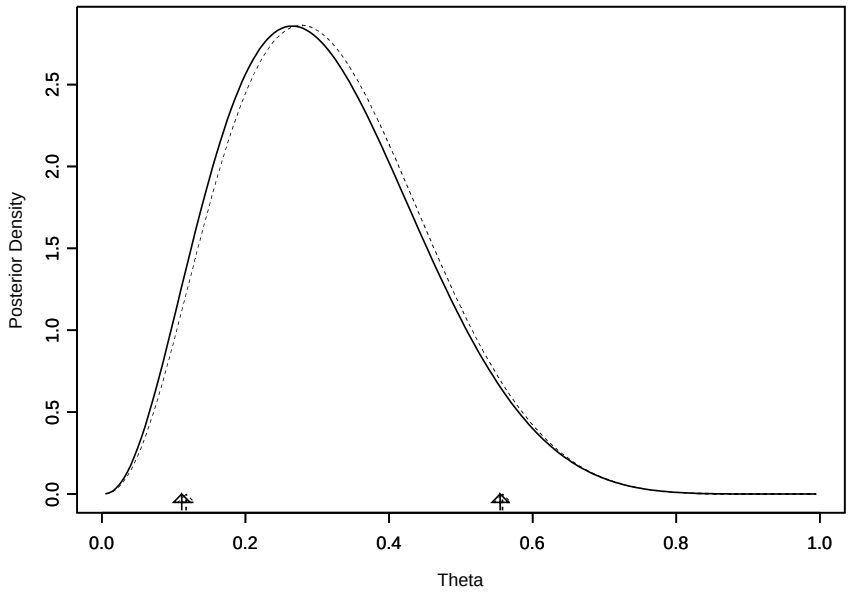

FIG. C.37 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,3) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$

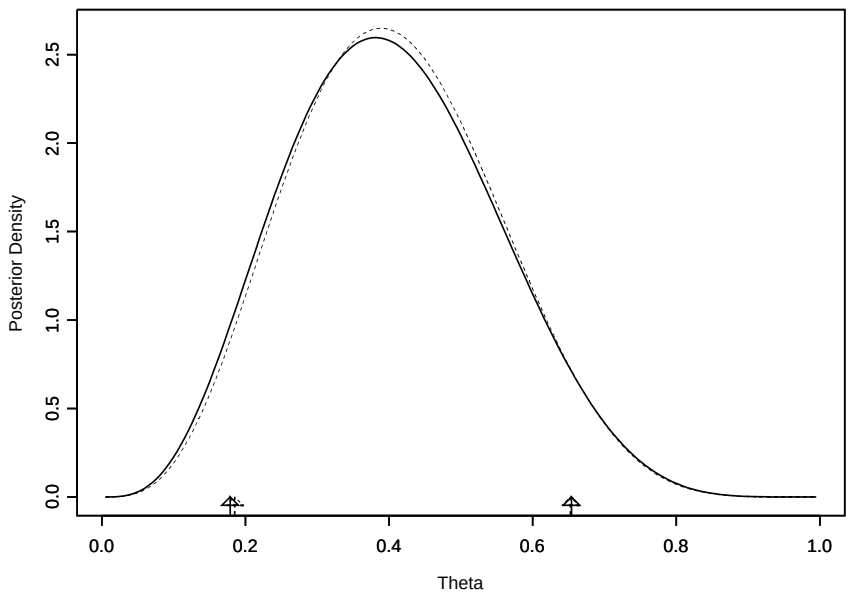

FIG. C.38 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,4) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$

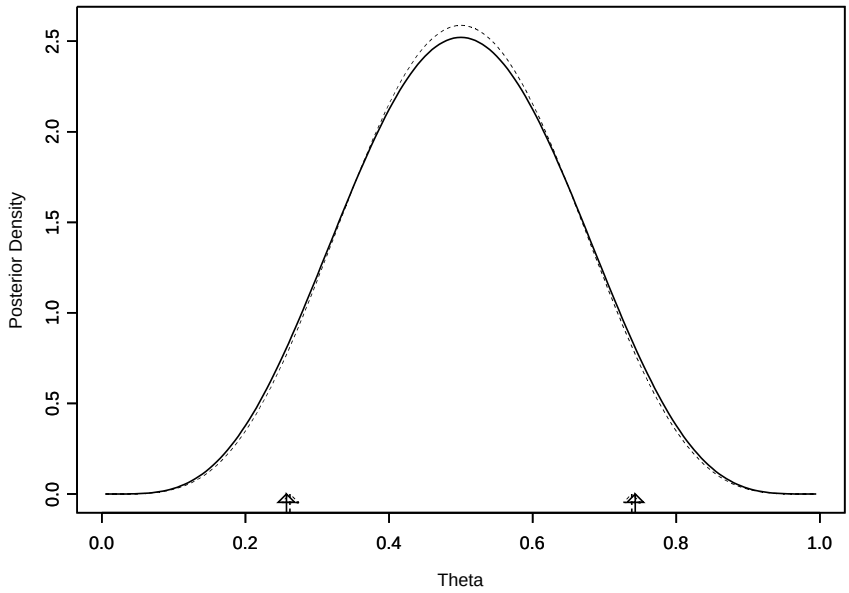

FIG. C.39 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,5) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$

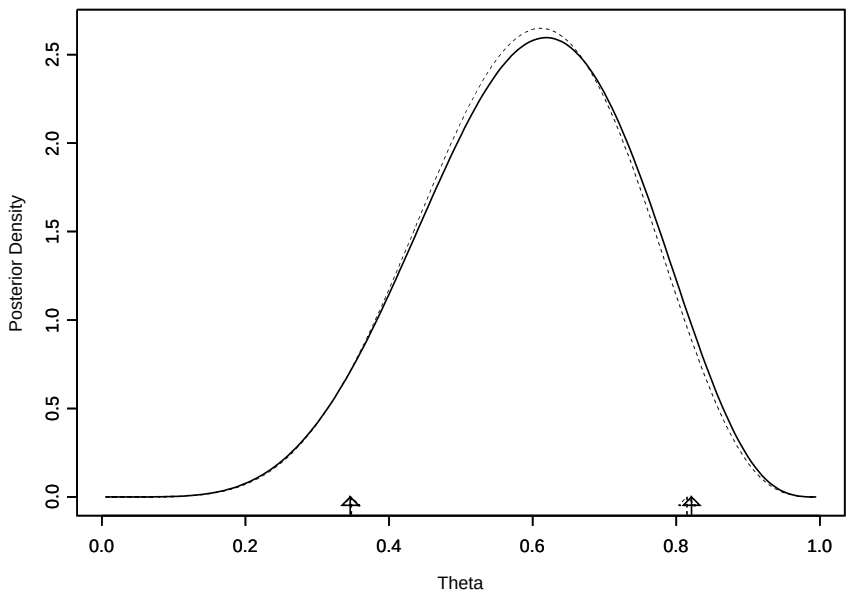

FIG. C.40 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,6) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$

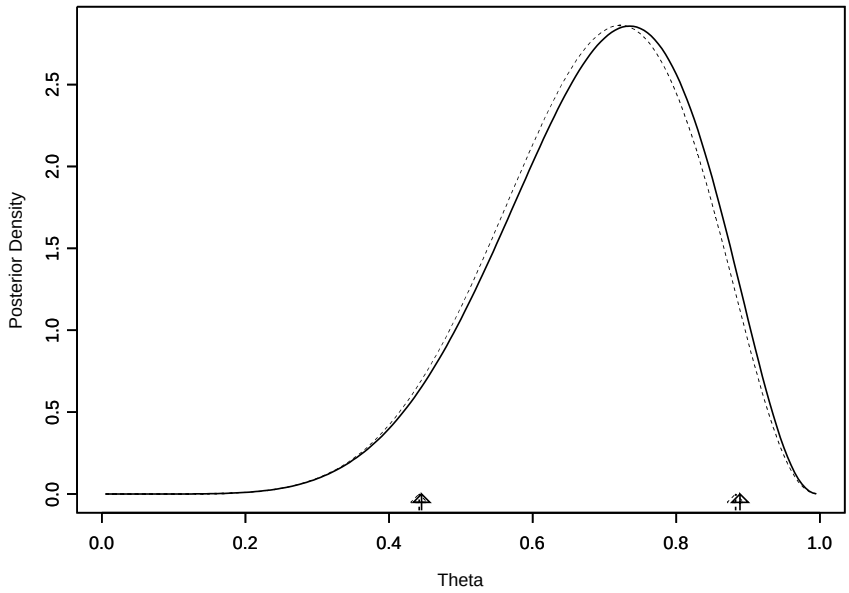

FIG. C.41 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,7) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$

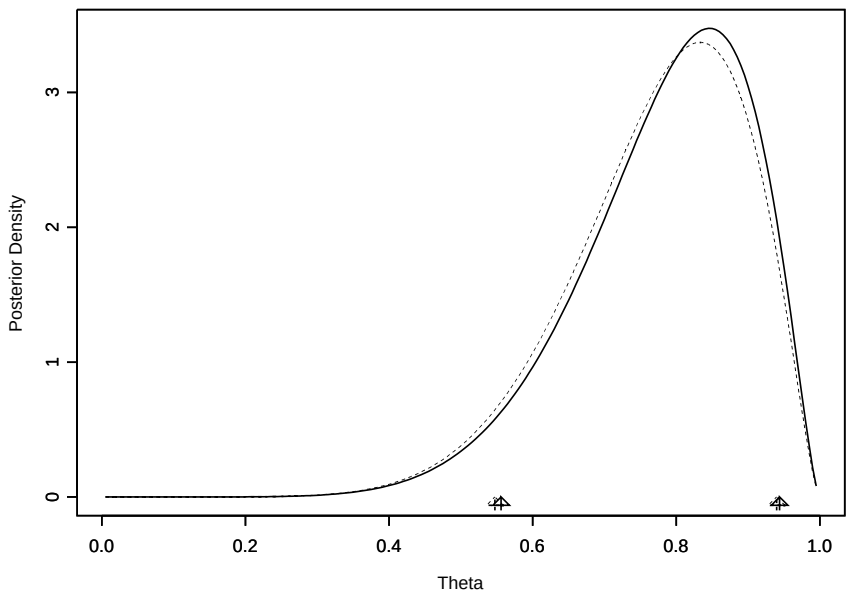

FIG. C.42 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,8) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$

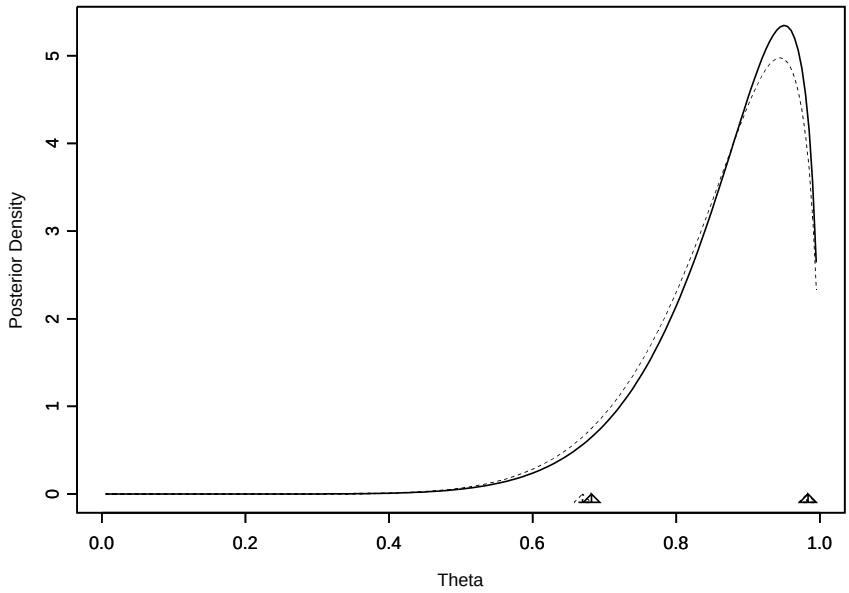

FIG. C.43 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,9) dans le plan séquentiel* $(n_1 = 5, S_{1,U} = 1, S_{1,L} = 4, n_2 = 5)$

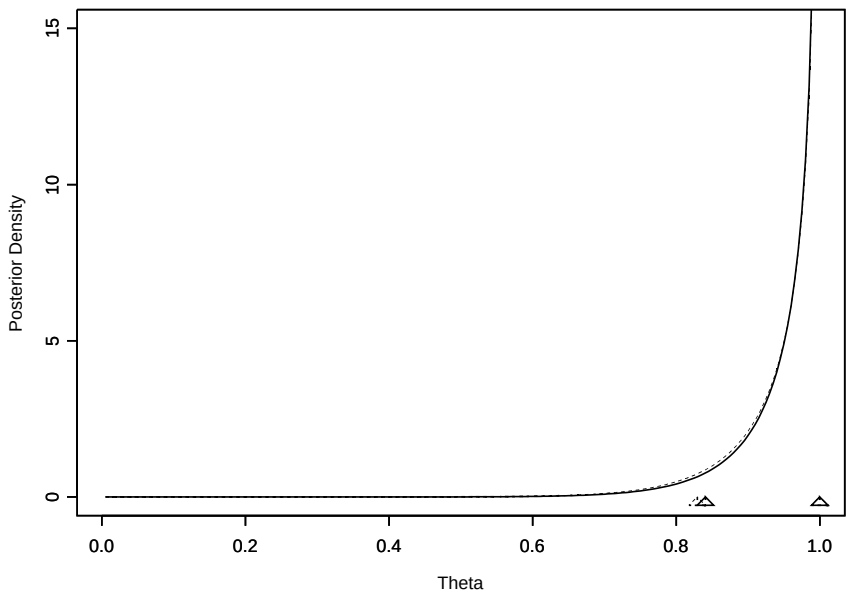

FIG. C.44 – *Loi a posteriori avec l'a priori de Jeffreys classique (- - -) et corrigé (—) pour l'observation (m,y)=(2,10) dans le plan séquentiel* ($n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5$)

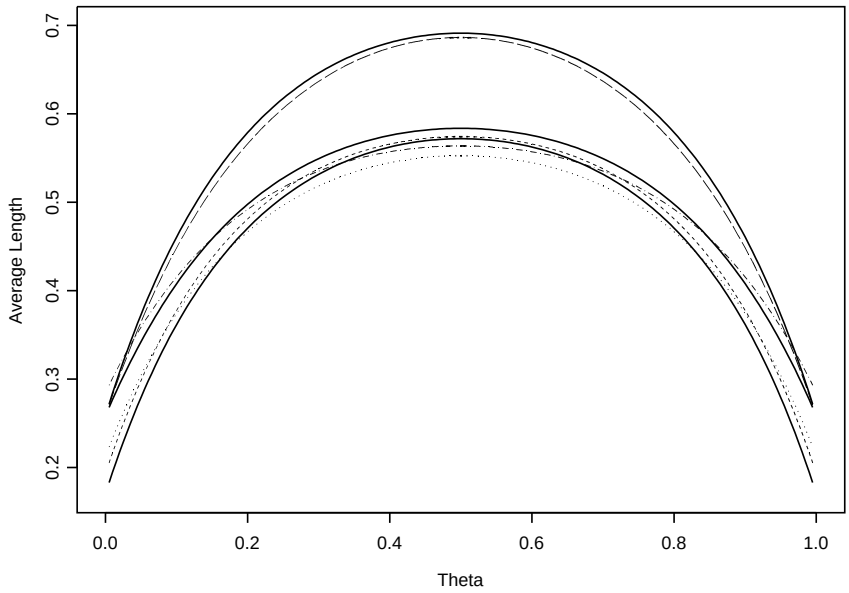

FIG. C.45 – *Longueur moyenne des intervalles de confiance 90% dans le plan séquentiel* ($n_1=5, S_{1,U}=1, S_{1,L}=4, n_2=5$). *Les intervalles* CI_J, CI_{JM} *et* CI_{CP} *(—) sont respectivement représentés de bas en haut avec également les intervalles* CI_{JT} *(– – –),* CI_W *(· · ·),* CI_{AC} *(· – ·) et* CI_2 *(- - -).*

		Test $TFCS^*$			Test $TFNCS$	
(k,y)	$B_k(y\|d_{Seq})$	Décision*	Risque*	Random*	Décision	Risque
(1,0)	$< 10^{-9}$	H_0 rejetée	$< 10^{-9}$	_	H_0 rejetée	0.05
(1,1)	$< 10^{-7}$	H_0 rejetée	$< 10^{-7}$	_	H_0 rejetée	0.05
(1,2)	$< 10^{-6}$	H_0 rejetée	$< 10^{-6}$	_	H_0 rejetée	0.05
(1,3)	$< 10^{-5}$	H_0 rejetée	$< 10^{-5}$	_	H_0 rejetée	0.05
(1,4)	$< 10^{-4}$	H_0 rejetée	$< 10^{-4}$	_	H_0 rejetée	0.05
(1,5)	0.0003	H_0 rejetée	0.0003	_	H_0 rejetée	0.05
(1,6)	0.002	H_0 rejetée	0.002	_	H_0 rejetée	0.05
(1,7)	0.006	H_0 rejetée	0.006	_	H_0 rejetée	0.05
(1,8)	0.018	H_0 rejetée	0.017	_	H_0 rejetée	0.05
(1,9)	0.045	H_0 rejetée	0.043	_	H_0 rejetée	0.05
(2,9)	$< 10^{-7}$	H_0 rejetée	$< 10^{-7}$	_	H_0 rejetée	0.05
(2,10)	$< 10^{-6}$	H_0 rejetée	$< 10^{-6}$	_	H_0 rejetée	0.05
(2,11)	$< 10^{-5}$	H_0 rejetée	$< 10^{-5}$	_	H_0 rejetée	0.05
(2,12)	$< 10^{-5}$	H_0 rejetée	$< 10^{-5}$	_	H_0 rejetée	0.05
(2,13)	$< 10^{-4}$	H_0 rejetée	$< 10^{-4}$	_	H_0 rejetée	0.05
(2,14)	$< 10^{-4}$	H_0 rejetée	$< 10^{-4}$	_	H_0 rejetée	0.05
(2,15)	0.0002	H_0 rejetée	0.0002	_	H_0 rejetée	0.05
(2,16)	0.0006	H_0 rejetée	0.0006	_	H_0 rejetée	0.05
(2,17)	0.002	H_0 rejetée	0.002	_	H_0 rejetée	0.05
(2,18)	0.004	H_0 rejetée	0.004	_	H_0 rejetée	0.05
(2,19)	0.009	H_0 rejetée	0.009	_	H_0 rejetée	0.05
(2,20)	0.018	H_0 rejetée	0.018	_	H_0 rejetée	0.05
(2,21)	0.035	H_0 rejetée	0.034	_	H_0 rejetée	0.05
(2,22)	0.062	H_0 rejetée	0.059	_	_	_
(2,23)	0.103	H_0 rejetée	0.093	_	_	_
(2,24)	0.160	H_0 rejetée	0.138	_	_	_
(2,25)	0.235	H_0 rejetée	0.190	_	_	_
(2,26)	0.330	H_0 rejetée	0.248	_	_	_
(2,27)	0.448	H_0 rejetée	0.310	_	_	_
(2,28)	0.594	H_0 rejetée	0.372	_	_	_
(2,29)	0.774	H_0 rejetée	0.436	_	_	_
(2,30)	1	Non-Décision	_	1	_	_
(2,31)	1.29	H_0 rejetée	0.436	_	_	_
(2,32)	1.68	H_0 rejetée	0.372	_	_	_
(2,33)	2.23	H_0 rejetée	0.310	_	_	_
(2,34)	3.03	H_0 rejetée	0.248	_	_	_
(2,35)	4.26	H_0 rejetée	0.190	_	_	_
(2,36)	6.26	H_0 acceptée	0.138	_	_	_
(2,37)	9.71	H_0 acceptée	0.093	_	_	_
(2,38)	16.0	H_0 acceptée	0.059	_	_	_
(2,39)	28.4	H_0 acceptée	0.034	_	_	_
(2,40)	54.1	H_0 acceptée	0.018	_	_	_
(2,41)	112	H_0 acceptée	0.009	_	_	_
(2,42)	252	H_0 acceptée	0.004	_	_	_
(2,43)	632	H_0 acceptée	0.002	_	_	_
(2,44)	1.8×10^{3}	H_0 acceptée	0.0006	_	_	_
(2,45)	5.1×10^{3}	H_0 acceptée	0.0002	_	_	_
(2,46)	1.6×10^{4}	H_0 acceptée	$< 10^{-4}$	_	_	_
(2,47)	5.6×10^{4}	H_0 acceptée	$< 10^{-4}$	_	_	_
(2,48)	2.2×10^{5}	H_0 acceptée	$< 10^{-5}$	_	_	_
(2,49)	9.2×10^{5}	H_0 acceptée	$< 10^{-5}$	_	_	_
(2,50)	4.4×10^{6}	H_0 acceptée	$< 10^{-6}$	_	_	_
(2,51)	2.4×10^{7}	H_0 acceptée	$< 10^{-7}$	_	_	_
(2,52)	1.4×10^{8}	H_0 acceptée	$< 10^{-8}$	_	_	_
(2,53)	1.0×10^{9}	H_0 acceptée	$< 10^{-9}$	_	_	_
(2,54)	8.5×10^{9}	H_0 acceptée	$< 10^{-9}$	_	_	_
(2,55)	8.5×10^{10}	H_0 acceptée	$< 10^{-10}$	_	_	_
(2,56)	1.1×10^{12}	H_0 acceptée	$< 10^{-12}$	_	_	_
(2,57)	1.7×10^{13}	H_0 acceptée	$< 10^{-13}$	_	_	_
(2,58)	4.0×10^{14}	H_0 acceptée	$< 10^{-14}$	_	_	_
(2,59)	1.6×10^{16}	H_0 acceptée	$< 10^{-16}$	_	_	_
(2,60)	1.9×10^{18}	H_0 acceptée	$< 10^{-18}$	_	_	_

TAB. C.1 – *Décisions et risques associés au test $TFCS^*$ et au test $TFNCS$ avec l'a priori de Jeffreys classique pour $\alpha_r^* = 0.05$ et $n_1 = n_2 = 30$*

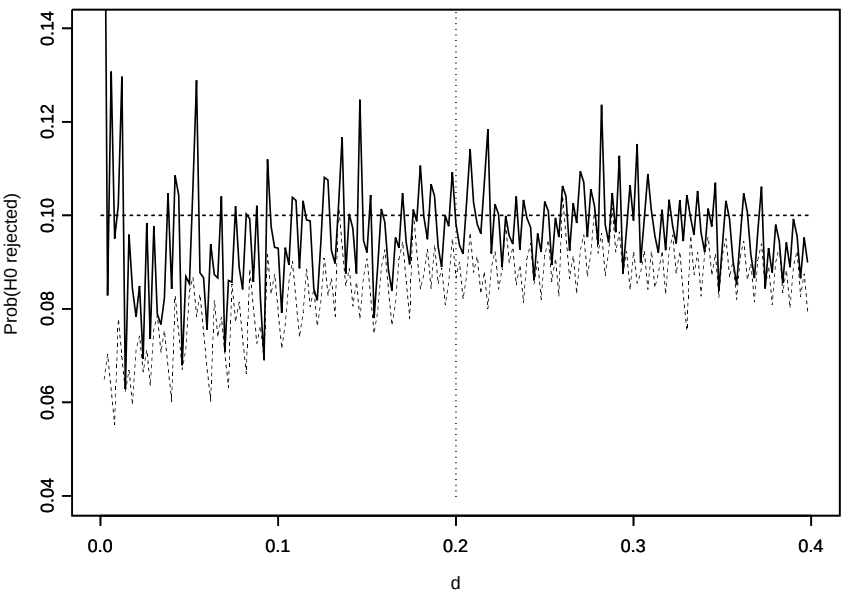

FIG. C.46 – *Valeurs limites de* $Pr(RejetH_0|\theta_0 = dec)$ *(—) et* $Pr(RejetH_0|\theta_0 = dec + 0.2)$ *(- - -) telles que* $Pr(RejetH_0|\theta_0 = dec) < Pr(RejetH_0|\theta_0 = dec + 0.2)$ *en fonction du paramètre de décentrage dec dans l'hypothèse* $H_0 : \{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ *pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2 = 100$

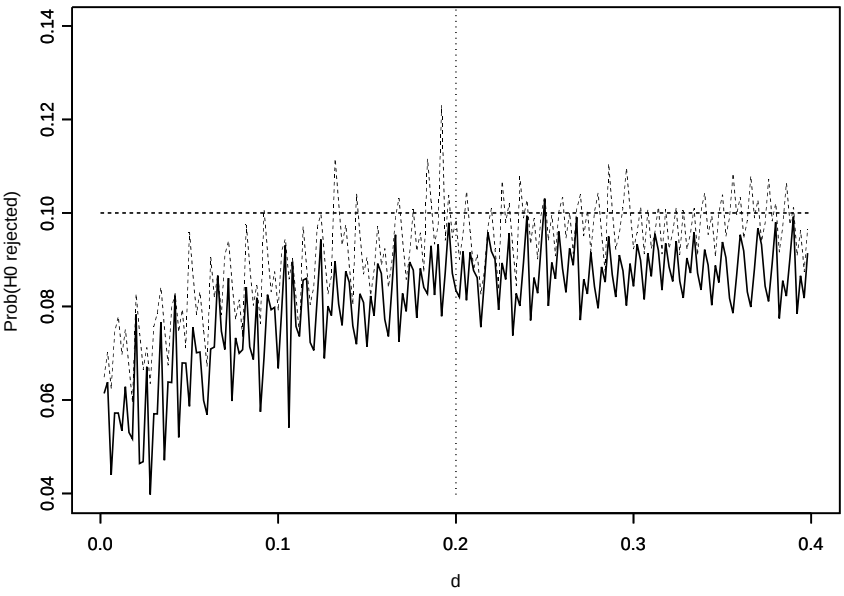

FIG. C.47 – *Valeurs limites de* $Pr(RejetH_0|\theta_0 = dec)$ *(—) et* $Pr(RejetH_0|\theta_0 = dec + 0.2)$ *(- - -) telles que* $Pr(RejetH_0|\theta_0 = dec) > Pr(RejetH_0|\theta_0 = dec + 0.2)$ *en fonction du paramètre de décentrage dec dans l'hypothèse* $H_0 : \{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ *pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2 = 100$

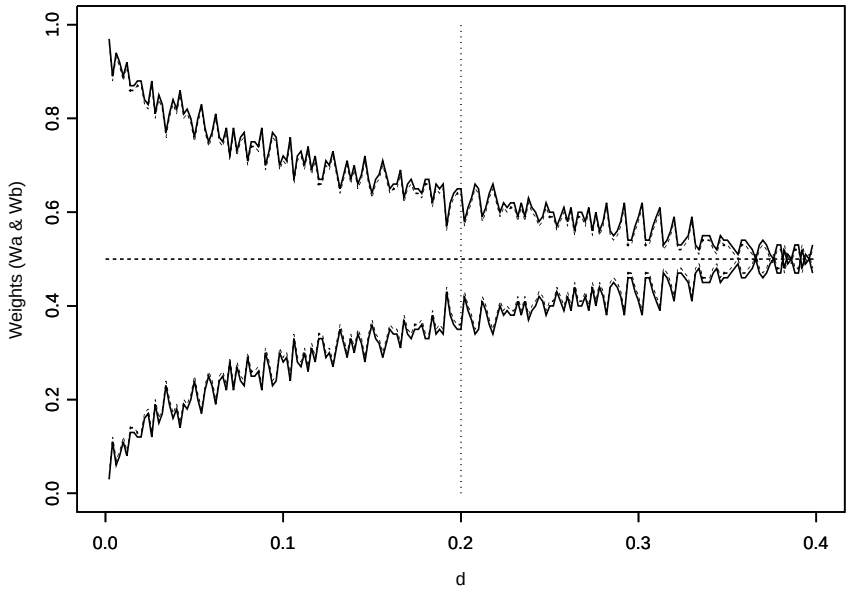

FIG. C.48 – *Pondérations limites* (w_a, w_b) *en fonction du paramètre de décentrage dec dans l'hypothèse* $H_0 : \{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ *avec* $Pr(RejetH_0|\theta_0 = dec) > Pr(RejetH_0|\theta_0 = dec + 0.2)$ *(—) et* $Pr(RejetH_0|\theta_0 = dec) < Pr(RejetH_0|\theta_0 = dec + 0.2)$ *(- - -) pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2 = 100$

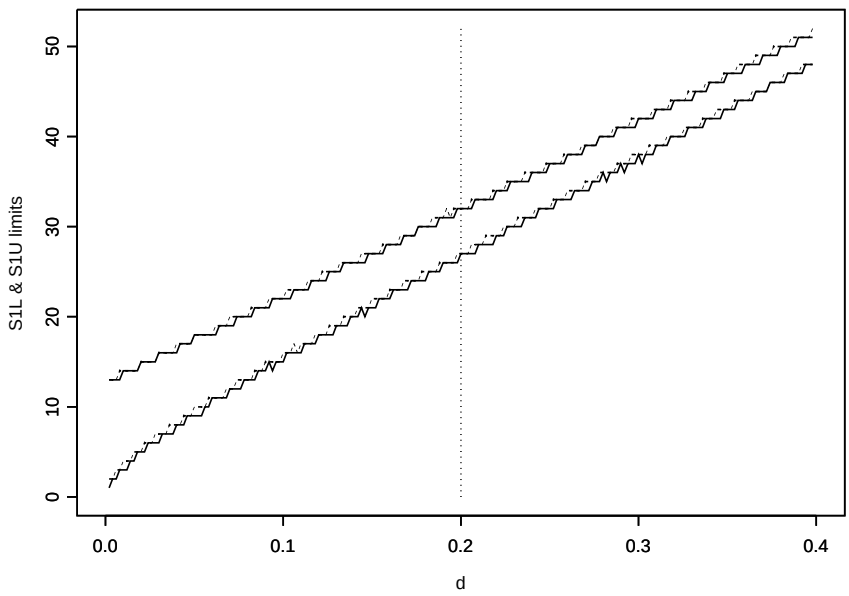

FIG. C.49 – *Bornes* $(S_{1,L}, S_{1,U})$ *de la région de rejet de* H_0 *en fonction du paramètre de décentrage dec dans l'hypothèse* H_0 : $\{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ *avec* $Pr(RejetH_0|\theta_0 = dec) > Pr(RejetH_0|\theta_0 = dec + 0.2)$ *(—) et* $Pr(RejetH_0|\theta_0 = dec) < Pr(RejetH_0|\theta_0 = dec + 0.2)$ *(- - -) pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2 = 100$

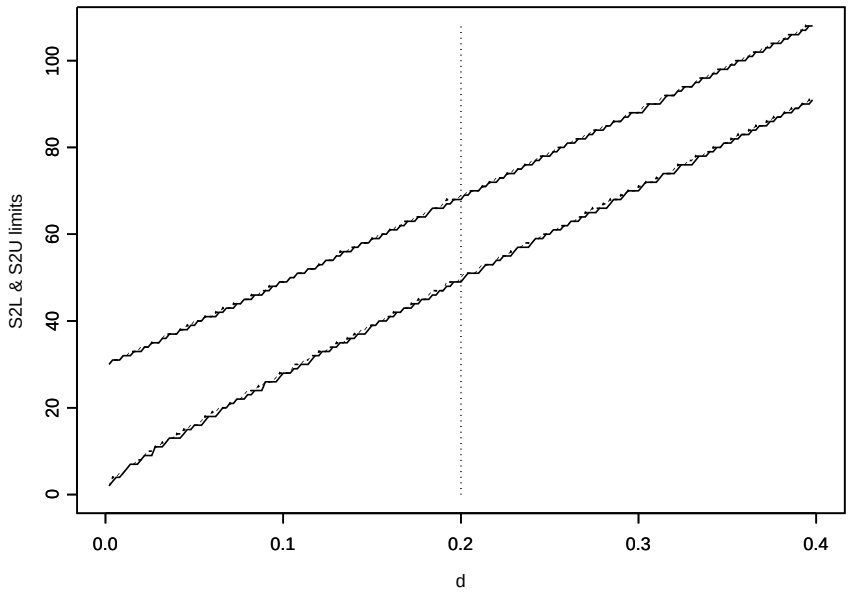

FIG. C.50 – *Bornes* $(S_{2,L}, S_{2,U})$ *de la région de rejet de* H_0 *en fonction du paramètre de décentrage dec dans l'hypothèse* H_0 : $\{\theta \in [0, dec] \cup [dec + 0.2, 1]\}$ *avec* $Pr(RejetH_0|\theta_0 = dec) > Pr(RejetH_0|\theta_0 = dec + 0.2)$ *(—) et* $Pr(RejetH_0|\theta_0 = dec) < Pr(RejetH_0|\theta_0 = dec + 0.2)$ *(- - -) pour* $\alpha_r^* = 0.05$ *et* $n_1 = n_2 = 100$