

Université de Montréal

Approche bayésienne pour estimer le rapport de

cotes dans un tableau de contingence 2×2

par

Marc Fredette

Département de mathématiques et de statistique
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de

Maître ès sciences (M.Sc.)
en statistique

août 1999

© Marc Fredette, 1999



QA
3
U54
1999
V.019

Université de Montréal

Approche bayésienne pour estimer le rapport de

coûts dans un tableau de contingence 2 X 2

par

et

Marc Frenette

Département de mathématiques et de statistiques
Faculté des sciences

Mention: maîtrise à la fin de la maîtrise
en vue de l'obtention du grade de

Maîtrise en mathématiques
en statistique

1999

1999



Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

**Approche bayésienne pour estimer le rapport de
cotes dans un tableau de contingence 2×2**

présenté par

Marc Fredette

a été évalué par un jury composé des personnes suivantes :

Urs Maag

(président-rapporteur)

Jean-François Angers

(directeur de recherche)

Robert Cléroux

(membre du jury)

Mémoire accepté le :

26 Octobre 1999

SOMMAIRE

Dans ce mémoire, nous étudions les tests d'indépendance de variables aléatoires catégorielles à deux modalités. Ces différents tests sont répartis dans deux catégories, ceux basés sur des modélisations fréquentistes, ainsi que ceux déterminés à partir de modèles bayésiens. En ce qui a trait aux modèles bayésiens, nous présentons ceux où l'information disponible *a priori* concerne les probabilités conjointes de ces variables ou une mesure de leur association.

Notre recherche est surtout consacrée à l'étude de la distribution *a posteriori* du logarithme du rapport de cotes sous la modélisation usuelle, modélisation où les observations admettent une loi multinomiale et les probabilités conjointes, une loi de Dirichlet. Nous voyons comment estimer les moments de la distribution *a posteriori*, ainsi que la façon de les calculer exactement lorsque les paramètres de la loi de Dirichlet prennent des valeurs entières. Finalement, nous présentons trois approximations de cette distribution : l'approximation normale, l'expansion de Latorre, ainsi qu'une nouvelle approximation déterminée à partir du théorème d'inversion des fonctions caractéristiques.

REMERCIEMENTS

Je voudrais tout d'abord remercier le Conseil de recherches en sciences naturelles et en génie pour le support financier qu'il m'a accordé. Je remercie également André Montpetit pour son aide à la mise en pages ainsi que Philippe Larocque et Chantal Labbé pour leur précieuse collaboration. Merci à Jonathan Dumas pour son aide et ses judicieux conseils. Également, un gros merci à Aurélie pour son aide, sa patience, sa compréhension et tout le reste...

Pour ce qui est des remerciements à mon directeur de recherche, Jean-François Angers, il suffit de parcourir ceux de ses nombreux étudiants précédents pour réaliser à quel point il est difficile d'être à la fois original et élogieux. Je le remercie donc très sincèrement pour sa disponibilité, sa patience, sa compréhension, ses judicieux conseils et ses brillantes idées. Je risque fort de ne jamais retrouver meilleure relation de travail, merci Jean-François.

Table des matières

Sommaire	iii
Remerciements	iv
Table des figures	viii
Liste des tableaux	ix
Introduction	1
Chapitre 1. Approche classique pour tester l'indépendance des variables catégorielles	4
1.1. Les variables catégorielles	4
1.1.1. Les tableaux de contingence	5
1.1.2. Le rapport de cotes	7
1.1.3. Indépendance des variables catégorielles	9
1.2. Fonctions de vraisemblance des variables catégorielles	11
1.2.1. Distribution de Poisson	12
1.2.2. Distribution multinomiale	12
1.2.3. Distributions multinomiales indépendantes	14
1.2.4. Estimateurs du maximum de vraisemblance	15
1.3. Test d'indépendance avec statistiques quadratiques	17
1.3.1. La statistique X^2 de Pearson	18

1.3.2.	Test d'indépendance avec la statistique X^2	19
1.3.3.	Test du rapport de vraisemblance de Wilks	20
1.4.	Test par le rapport de cotes	22
1.5.	Test exact de Fisher	25
Chapitre 2. Approche bayésienne pour tester l'indépendance des variables catégorielles		27
2.1.	Concepts de base de la théorie bayésienne	27
2.1.1.	Le modèle statistique bayésien	28
2.1.2.	Lois <i>a priori</i> conjuguées et non informatives	29
2.1.3.	Estimation bayésienne ponctuelle	30
2.1.4.	Région de crédibilité	32
2.2.	Modélisation bayésienne usuelle pour les variables catégorielles	33
2.2.1.	Définition de la loi de Dirichlet	34
2.2.2.	Propriétés de la loi de Dirichlet	38
2.2.3.	L'approche bayésienne empirique	39
2.2.4.	Relation avec le test exact de Fisher	40
2.3.	Modélisation bayésienne particulière	42
2.3.1.	La loi logistique-normale	42
2.3.2.	Loi <i>a priori</i> sur une mesure d'association	43
2.4.	Modélisation bayésienne hiérarchique	45
2.4.1.	Hiérarchisation des modèles usuels	46
2.4.2.	Loi <i>a priori</i> hiérarchique sur une mesure d'association	48
2.4.3.	Modélisation hiérarchique particulière	50

Chapitre 3. Distribution <i>a posteriori</i> du logarithme du rapport de cotes.....	52
3.1. Normalité asymptotique du logarithme du rapport de cotes	53
3.2. Les moments exacts.....	56
3.3. Une nouvelle approximation du logarithme du rapport de cotes....	63
3.3.1. Expansion de Latorre.....	63
3.3.2. Détermination de la nouvelle approximation	65
3.3.3. Étude de l'approximation	70
3.4. Calcul de régions de crédibilité.....	81
3.5. Comparaison des principales approches.....	83
Conclusion	96
Annexe A: Les programmes.....	98
Bibliographie	107

Table des figures

3.5.1	Approximations de la distribution de ϕ à partir du premier jeu de données.....	92
3.5.2	Approximations de la distribution de ϕ à partir du deuxième jeu de données.....	93
3.5.3	Approximations de la distribution de ϕ à partir du troisième jeu de données.....	94
3.5.4	Convergence de l'expansion de Latorre.....	95

Liste des tableaux

1.1.1	Exemple d'un tableau de contingence 2×3	6
1.1.2	Notations des probabilités conjointes, marginales et conditionnelles dans un tableau de probabilités 2×3	7
3.3.1	Les 11 premières fonctions $M_k(z)$	71
3.3.2	Les 11 premiers $D_k(a)$	73
3.5.1	Tableau de contingence du premier jeu de données.....	84
3.5.2	Tableau de contingence du deuxième jeu de données.....	85
3.5.3	Tableau de contingence du troisième jeu de données.....	86
3.5.4	Comparaison des méthodes fréquentistes.....	87
3.5.5	Comparaison des méthodes bayésiennes pour le premier jeu de données	91
3.5.6	Comparaison des méthodes bayésiennes pour le deuxième jeu de données.....	91
3.5.7	Comparaison des méthodes bayésiennes pour le troisième jeu de données.....	91

INTRODUCTION

L'étude de l'indépendance entre deux variables aléatoires catégorielles à deux modalités est un problème statistique fréquemment abordé. Plusieurs domaines dont l'épidémiologie, la psychiatrie et la recherche clinique requièrent l'utilisation de ces variables se regroupant dans un tableau de contingence 2×2 . Par contre, plusieurs modèles, et donc plusieurs tests, différents furent suggérés à ce jour, tests ne menant pas toujours aux mêmes conclusions. Parmi ces tests, les plus connus sont sans aucun doute le test basé sur la statistique quadratique de Pearson (1900), celui basé sur la statistique quadratique de Wilks (1935), ainsi que le test exact de Fisher (1934). Depuis quelques décennies, plusieurs articles traitent d'un test d'indépendance issu d'une modélisation bayésienne, il consiste à étudier le comportement *a posteriori* du logarithme du rapport de cotes lorsque le vecteur des probabilités admet *a priori* une loi de Dirichlet. Les principales découvertes sur ce sujet se retrouvent dans les articles de Lindley (1964), Bloch et Watson (1967), Altham (1969) et Aitchison et Shen (1980), ainsi que dans le livre de O'Hagan (1994, chapitre 10).

Dans ce mémoire, nous présenterons plusieurs de ces modèles concernant les variables catégorielles, ainsi que les différents tests qui s'y rattachent, en les répartissant en deux catégories, les modèles classiques, ou fréquentistes, ainsi que les modèles bayésiens. En ce qui a trait aux modélisations bayésiennes, nous présentons ceux où l'information disponible *a priori* concerne les probabilités conjointes de ces variables ou une mesure de leur association.

Dans le premier chapitre, nous débuterons en définissant différents concepts de base tels que les variables aléatoires catégorielles, les tableaux de contingence et le rapport de cotes. Par la suite, nous présenterons différents modèles classiques et les estimateurs du maximum de vraisemblance des paramètres inconnus incorporés dans ces modèles. Nous y présenterons quatre différents modèles. Le premier utilise la loi de Poisson et suppose que la taille échantillonnale est aléatoire. Le second utilise la loi multinomiale en supposant que la taille de l'échantillon est fixe. Le troisième requiert que la répartition des effectifs marginaux d'une des deux variables soit connue et utilise des lois multinomiales indépendantes. Pour ce qui est du dernier modèle, il suppose que tous les effectifs marginaux soient connus et dans ce contexte la probabilité d'une répartition donnée est obtenue par la loi hypergéométrique. Par la suite, nous présenterons quatre formes de tests classiques s'appliquant à un ou plusieurs des modèles mentionnés. Nous verrons les tests basés sur les statistiques quadratiques X^2 et G^2 , un autre test utilisant la normalité asymptotique de l'estimateur du maximum de vraisemblance du logarithme du rapport de cotes et finalement, le test exact de Fisher.

Le second chapitre se veut une introduction à l'approche bayésienne pour tester l'indépendance des variables aléatoires catégorielles. C'est dans ce chapitre que nous présenterons les différents modèles bayésiens. Nous débuterons en définissant les concepts de base de la théorie bayésienne, concepts tels que les lois *a priori*, les estimations bayésiennes ponctuelles et les régions de crédibilité. Par la suite, nous présenterons le modèle bayésien le plus fréquemment utilisé lorsque nous étudions des variables catégorielles, modèle où nous supposons que les variables suivent une loi multinomiale et que les paramètres inconnus de cette loi suivent *a priori* la loi conjuguée de Dirichlet. Cette loi *a priori* incorpore donc de l'information concernant les probabilités conjointes de ces variables aléatoires.

Dans les deux dernières sections de ce chapitre, nous présenterons divers modèles bayésiens que nous diviserons en deux catégories : ceux utilisant une autre loi *a priori* que la loi de Dirichlet et les modèles bayésiens hiérarchiques.

Le dernier chapitre est principalement consacré à l'étude de l'indépendance des variables catégorielles lorsque nous utilisons le modèle bayésien usuel défini au chapitre précédent. Pour ce faire, nous étudierons le comportement *a posteriori* du logarithme du rapport de cotes. Nous débiterons en proposant différentes estimations des moments de cette distribution *a posteriori*, par la suite, nous verrons comment calculer exactement ces moments lorsque les paramètres de la loi de Dirichlet prennent des valeurs entières. Nous présenterons ensuite trois différentes approximations de cette distribution : l'approximation normale, l'expansion de Latorre (1982), ainsi qu'une nouvelle approximation déterminée à partir de la fonction caractéristique du logarithme du rapport de cotes. Nous terminerons en comparant les tests basés sur les modèles fréquentistes à ceux déterminés à partir de la modélisation bayésienne usuelle. Ces derniers sont basés sur les régions α -crédibles, l'alternative bayésienne aux intervalles de confiance.

Chapitre 1

APPROCHE CLASSIQUE POUR TESTER L'INDÉPENDANCE DES VARIABLES CATÉGORIELLES

Dans ce premier chapitre, nous définissons tout d'abord en quoi consistent les données catégorielles, comment représenter conjointement leur distribution et comment définir leur indépendance. Par la suite, nous voyons différentes manières fréquentistes de tester l'indépendance de ces données à partir de leur fonction de vraisemblance. Hormis le test exact de Fisher présenté à la section 1.5, toutes les méthodes d'inférence présentées dans ce chapitre nécessitent l'utilisation de distributions asymptotiques. Ce chapitre est tiré des trois premiers chapitres du livre d'Agresti (1990).

1.1. LES VARIABLES CATÉGORIELLES

Tout au long de ce mémoire, les différentes inférences proposées ne sont utilisées que pour des variables aléatoires catégorielles. Nous débutons donc ce mémoire avec une définition précise de ces variables, ensuite nous énumérerons leurs différents types.

Définition 1.1.1. *Une variable catégorielle est une variable aléatoire dont l'échelle de mesure consiste en un nombre fini de modalités.*

Il y a trois différents types de variables catégorielles et les différentes méthodes statistiques utilisées dans ce mémoire seront toujours valables pour ces trois types. Il y a les variables nominales, les variables ordinales et celles dites à intervalle. Les variables nominales sont celles où les différentes modalités n'ont pas d'ordre naturel. Ces variables sont discrètes et qualitatives, comme par exemple la couleur des yeux d'un individu. Les variables dites ordinales possèdent par contre un ordre naturel entre les différentes modalités. Par exemple, elles peuvent mesurer un taux de satisfaction (favorable, indifférent, défavorable). Ces variables sont également discrètes et qualitatives. Finalement, les variables à intervalle sont celles où nous pouvons établir une distance numérique entre deux modalités. Elles sont toujours quantitatives, mais peuvent être discrètes ou bien continues. Lorsqu'elles sont continues, ces variables doivent être regroupées en un certain nombre de classes. Par exemple, nous pouvons regrouper en différentes classes la taille et la masse des individus.

1.1.1. Les tableaux de contingence

Soit (X_i, Y_i) , où i varie de 1 à n , un vecteur aléatoire de dimension $2 \times n$, où X et Y sont deux variables catégorielles admettant respectivement I et J modalités. Nous sommes intéressés à trouver une manière de représenter ces résultats afin d'analyser la distribution conjointe de ces deux variables. Nous utilisons un tableau de contingence, terme introduit par Karl Pearson (1904).

Définition 1.1.2. *Un tableau de contingence de dimension $I \times J$ représente conjointement la fréquence de la répartition des effectifs. L'élément de la cellule (i, j) dans ce tableau est noté n_{ij} , il représente le nombre d'observations où les deux variables catégorielles ont respectivement i et j comme modalité.*

TABLEAU 1.1.1. Exemple d'un tableau de contingence 2×3 .

$X \backslash Y$	1	2	3	Total
1	n_{11}	n_{12}	n_{13}	n_{1+}
2	n_{21}	n_{22}	n_{23}	n_{2+}
Total	n_{+1}	n_{+2}	n_{+3}	n

Le tableau 1.1.1 représente un tableau de contingence de dimension 2×3 . Notons que l'indice "+" indique une sommation sur l'indice qu'il remplace, c'est-à-dire :

$$n_{+j} = \sum_{i=1}^I n_{ij} \quad \text{et} \quad n_{i+} = \sum_{j=1}^J n_{ij}.$$

Nous avons donc les égalités suivantes,

$$\sum_{i=1}^I n_{i+} = \sum_{j=1}^J n_{+j} = n_{++} \equiv n.$$

Soit θ_{ij} la probabilité que le vecteur (X, Y) se retrouve dans la cellule (i, j) du tableau de contingence. L'ensemble des θ_{ij} détermine alors la distribution conjointe de X et Y . Leur distribution marginale est déterminée en additionnant ces probabilités conjointes. De plus, dans plusieurs tableaux de contingence, nous considérons parfois qu'une des deux variables est explicative et que l'autre est une variable de réponse. Nous pouvons donc obtenir la distribution conditionnelle de la variable de réponse quand l'autre a une valeur fixe. Cette probabilité conditionnelle est notée $\theta_{j|i}$. Nous reviendrons sur ce sujet dans la sous-section 1.2.3. Le tableau 1.1.2 présente ces différentes notations.

TABLEAU 1.1.2. *Notations des probabilités conjointes, marginales et conditionnelles dans un tableau de probabilités 2×3 .*

$X \setminus Y$	1	2	3	Total
1	θ_{11}	θ_{12}	θ_{13}	θ_{1+}
	$(\theta_{1 1})$	$(\theta_{2 1})$	$(\theta_{3 1})$	(1.0)
2	θ_{21}	θ_{22}	θ_{23}	θ_{2+}
	$(\theta_{1 2})$	$(\theta_{2 2})$	$(\theta_{3 2})$	(1.0)
Total	θ_{+1}	θ_{+2}	θ_{+3}	1.0

1.1.2. Le rapport de cotes

Avant de parler de l'indépendance des variables aléatoires catégorielles, il est important de définir ce qu'est un rapport de cotes, ce concept étant très lié à celui de l'indépendance. Prenons, par exemple, un tableau de probabilité 2×2 , pour une rangée i donnée, la cote que l'observation soit dans la première colonne plutôt que dans la deuxième est définie comme étant le rapport de ses probabilités conjointes, c'est-à-dire :

$$\Omega_i = \theta_{1|i} / \theta_{2|i} = \theta_{i1} / \theta_{i2} \quad i = 1, 2.$$

Chaque Ω_i est non négatif et sa valeur nous indique, pour une rangée i donnée, combien de fois il est plus probable qu'une observation se trouve dans la colonne 1 plutôt que la colonne 2. Cela nous amène à définir le rapport de cotes.

Définition 1.1.3. *Dans un tableau 2×2 , le rapport de cotes ou ratio des produits croisés, noté δ , correspond au rapport des cotes Ω_1 et Ω_2 , c'est-à-dire*

$$\delta = \frac{\Omega_1}{\Omega_2} = \frac{\theta_{11}\theta_{22}}{\theta_{12}\theta_{21}}. \quad (1.1.1)$$

Tout comme une cote, δ est un réel non négatif. Notons que lorsqu'une de ces quatre probabilités vaut zéro, nous posons que δ vaut 0 ou ∞ . Nous remarquons également que le rapport de cotes est invariant si nous changeons les lignes avec les colonnes.

La définition du rapport de cotes peut être généralisée pour les tableaux de probabilités $I \times J$. Il est alors défini, pour chaque paire de rangées (i, k) et de colonnes (j, l) , de la manière suivante :

$$\delta_{ik,jl} = \frac{\theta_{ij}\theta_{kl}}{\theta_{il}\theta_{kj}}.$$

Cet ensemble de rapport de cotes nous amène beaucoup d'informations redondantes. Par contre, certains sous-ensembles de $(I - 1)(J - 1)$ rapports de cotes peuvent nous permettre de reconstituer tous les $\binom{I}{2}\binom{J}{2}$ rapports de cotes de notre tableau de probabilités. Considérons, par exemple, le sous-ensemble suivant :

$$\{\alpha_{ij} \mid 1 \leq i \leq (I - 1), 1 \leq j \leq (J - 1)\} \quad (1.1.2)$$

où

$$\alpha_{ij} = \frac{\theta_{ij}\theta_{IJ}}{\theta_{Ij}\theta_{iJ}} = \delta_{iI,jJ}. \quad (1.1.3)$$

Nous remarquons que ces α_{ij} prennent 1 comme valeur dès que i vaut I ou j vaut J . Voici maintenant la proposition démontrant qu'il est suffisant de connaître les α_{ij} pour retrouver tous les rapports de cotes.

Proposition 1.1.1. *Toute valeur $\delta_{ik,jl}$ peut être connue si nous connaissons tout l'ensemble $\{\alpha_{ij}\}$ défini dans les équations (1.1.2) et (1.1.3).*

PREUVE.

Comme $\alpha_{ij} = 1$ si $i = I$ ou $j = J$. Nous avons alors que

$$\begin{aligned} \frac{\alpha_{ij}\alpha_{kl}}{\alpha_{il}\alpha_{kj}} &= \frac{\left(\frac{\theta_{ij}\theta_{IJ}}{\theta_{Ij}\theta_{iJ}}\right) \left(\frac{\theta_{kl}\theta_{IJ}}{\theta_{Il}\theta_{kJ}}\right)}{\left(\frac{\theta_{il}\theta_{IJ}}{\theta_{Il}\theta_{iJ}}\right) \left(\frac{\theta_{kj}\theta_{IJ}}{\theta_{Ij}\theta_{kJ}}\right)} \\ &= \frac{\theta_{ij}\theta_{kl}}{\theta_{il}\theta_{kj}} \\ &= \delta_{ik,jl}, \end{aligned}$$

pour tout $i, k = 1, \dots, I$ et pour tout $j, l = 1, \dots, J$. □

Nous verrons dans la prochaine sous-section comment nous pouvons vérifier l'indépendance de nos variables aléatoires catégorielles en n'utilisant que les α_{ij} .

1.1.3. Indépendance des variables catégorielles

En pratique, nous utilisons surtout les tableaux de contingence pour tester si la probabilité de chaque cellule correspond à une valeur donnée, ou bien pour savoir si les variables aléatoires catégorielles, contenues dans ce tableau, sont indépendantes. Comme l'aspect qui nous intéressera le plus dans ce mémoire sera de tester l'indépendance des variables catégorielles, voici comment elle est définie.

Définition 1.1.4. *Deux variables aléatoires catégorielles, admettant respectivement I et J modalités, sont indépendantes si et seulement si toutes les probabilités conjointes correspondent au produit de leurs probabilités marginales, c'est-à-dire que*

$$\theta_{ij} = \theta_{i+}\theta_{+j} \text{ où } i = 1, \dots, I \text{ et } j = 1, \dots, J.$$

Lorsque Y est considérée comme une variable réponse de la variable explicative X , il est plus logique de définir l'indépendance en disant que toutes les

probabilités conditionnelles de Y sont identiques à leur probabilité marginale respective, c'est-à-dire :

$$\theta_{j|i} = \theta_{ij}/\theta_{i+} = (\theta_{i+}\theta_{+j})/\theta_{i+} = \theta_{+j} \text{ où } i = 1, \dots, I. \quad (1.1.4)$$

Plutôt que de vérifier si nos variables catégorielles satisfont la définition précédente, il nous sera plus facile pour certains tests d'étudier les valeurs des α_{ij} définis dans l'équation (1.1.3). La relation entre l'indépendance et les rapports de cotes étant clairement établie avec la proposition suivante.

Proposition 1.1.2. *Deux variables aléatoires catégorielles sont indépendantes si et seulement si tous les α_{ij} valent 1.*

PREUVE.

(\implies)

Si les deux variables aléatoires sont indépendantes, les α_{ij} prennent nécessairement la valeur 1, puisque

$$\begin{aligned} \alpha_{ij} &= \frac{\theta_{ij}\theta_{IJ}}{\theta_{Ij}\theta_{iJ}} \\ &= \frac{(\theta_{i+}\theta_{+j})(\theta_{I+}\theta_{+J})}{(\theta_{I+}\theta_{+j})(\theta_{i+}\theta_{+J})} \\ &= 1. \end{aligned}$$

(\impliedby)

À partir de la proposition 1.1.1, nous voyons que $\theta_{ij}\theta_{kl} = \theta_{kj}\theta_{il}$ lorsque les α_{ij} valent tous 1. Ainsi,

$$\begin{aligned} \theta_{+j}\theta_{i+} &= \left(\sum_{k=1}^I \theta_{kj} \right) \left(\sum_{l=1}^J \theta_{il} \right) \\ &= \sum_{k=1}^I \sum_{l=1}^J \theta_{kj}\theta_{il} \end{aligned}$$

$$\begin{aligned}
&= \sum_{k=1}^I \sum_{l=1}^J \theta_{ij} \theta_{kl} \\
&= \theta_{ij} \left(\sum_{k=1}^I \sum_{l=1}^J \theta_{kl} \right) \\
&= \theta_{ij}.
\end{aligned}$$

□

Dans un tableau 2×2 , le seul α_{ij} qui ne vaut pas nécessairement 1 est α_{11} , c'est-à-dire le rapport de cotes δ défini par l'équation (1.1.1). Nous avons donc le corollaire suivant.

Corollaire 1.1.1. *Dans un tableau de contingence 2×2 , deux variables aléatoires catégorielles sont indépendantes si et seulement si δ vaut 1.*

Corollaire 1.1.2. *Dans un tableau de contingence 2×2 , deux variables aléatoires catégorielles sont indépendantes si et seulement si ϕ vaut 0, où ϕ est tel que*

$$\phi = \log(\delta) = \sum_{i=1}^2 \sum_{j=1}^2 c_{ij} \log(\theta_{ij}), \quad (1.1.5)$$

où $C = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$.

Nous verrons à la section 1.4 un test d'indépendance basé sur le rapport de cotes. Par ailleurs, dans le chapitre 3 nous verrons des tests bayésiens basés sur différentes approximations de la distribution *a posteriori* du rapport de cotes.

1.2. FONCTIONS DE VRAISEMBLANCE DES VARIABLES CATÉGORIELLES

Avant de présenter différents tests d'indépendance, nous énumérons dans cette section les fonctions de vraisemblance les plus fréquemment utilisées pour des

variables aléatoires catégorielles. Nous présentons la distribution de Poisson, la distribution multinomiale et les distributions multinomiales indépendantes pour chaque modalité de la variable explicative. Par la suite, nous montrons comment estimer leurs différents paramètres en maximisant leur fonction de vraisemblance.

1.2.1. Distribution de Poisson

Soit N_{ij} l'effectif de la cellule (i, j) . Cette variable suit une loi de Poisson de paramètre m_{ij} si sa fonction de masse est définie de la façon suivante :

$$P(n_{ij}) = \begin{cases} \frac{\exp(-m_{ij})m_{ij}^{n_{ij}}}{n_{ij}!} & \text{si } n_{ij} \in \mathbb{N}, \\ 0 & \text{sinon.} \end{cases} \quad (1.2.1)$$

Soient $\mathbb{V}[N]$ et $\mathbb{E}[N]$ la variance et l'espérance de la variable aléatoire N . Nous avons alors $\mathbb{V}[N_{ij}] = \mathbb{E}[N_{ij}] = m_{ij}$.

Sous cette modélisation, les variables aléatoires N_{ij} sont indépendantes. Nous pouvons donc montrer que $N = \sum_{i=1}^I \sum_{j=1}^J N_{ij}$ suit également une loi de Poisson de paramètre $\sum_{i=1}^I \sum_{j=1}^J m_{ij}$. Sous cette modélisation, nous avons donc une taille échantillonnale aléatoire, d'espérance et de variance $\sum_{i=1}^I \sum_{j=1}^J m_{ij}$.

1.2.2. Distribution multinomiale

Soit (N_{11}, \dots, N_{IJ}) un vecteur aléatoire tel que $n = \sum_{i=1}^I \sum_{j=1}^J N_{ij}$ est fixe. Ce vecteur admet une distribution multinomiale de paramètre $\underline{\theta} = (\theta_{11}, \dots, \theta_{IJ})$ tel

que $\sum_{i=1}^I \sum_{j=1}^J \theta_{ij} = 1$ si sa fonction de masse s'écrit de la façon suivante :

$$P(n_{11}, \dots, n_{IJ}) = \begin{cases} \binom{n}{n_{11}, \dots, n_{IJ}} \prod_{i=1}^I \prod_{j=1}^J (\theta_{ij}^{n_{ij}}) & \text{si } \sum_{i=1}^I \sum_{j=1}^J n_{ij} = n \text{ et } \theta_{ij} > 0, \\ 0 & \text{sinon,} \end{cases} \quad (1.2.2)$$

où

$$\binom{n}{n_{11}, \dots, n_{IJ}} = \frac{n!}{\prod_{i=1}^I \prod_{j=1}^J n_{ij}!}.$$

Nous avons alors que $\mathbb{E}[n_{ij}] = n\theta_{ij}$ et $\mathbb{V}[n_{ij}] = n\theta_{ij}(1 - \theta_{ij})$. Cette distribution est une généralisation de la loi binomiale avec cette fois $(IJ - 1)$ paramètres. Elle sera la plus utilisée par la suite.

La proposition suivante énonce que si nous modélisons la répartition des effectifs à l'aide d'une loi de Poisson, tout en fixant à n la taille de l'échantillon, les effectifs seront alors répartis selon une distribution multinomiale.

Proposition 1.2.1. *Soit N_{ij} l'effectif de la cellule (i, j) . Si N_{ij} admet une loi de Poisson de paramètre m_{ij} pour tout i et pour tout j , alors le vecteur aléatoire (N_{11}, \dots, N_{IJ}) admet une loi multinomiale de paramètres $\theta_{ij} = m_{ij} \left(\sum_{i=1}^I \sum_{j=1}^J m_{ij} \right)^{-1}$ si la taille échantillonnale est fixée à n .*

PREUVE.

$$P \left(N_{ij} = n_{ij} \forall i, j \mid \sum_{i=1}^I \sum_{j=1}^J N_{ij} = n \right)$$

$$\begin{aligned}
&= \frac{P\left(N_{ij} = n_{ij} \forall i, j, \sum_{i=1}^I \sum_{j=1}^J N_{ij} = n\right)}{P\left(\sum_{i=1}^I \sum_{j=1}^J N_{ij} = n\right)} \\
&= \frac{\prod_{i=1}^I \prod_{j=1}^J \left(\frac{m_{ij}^{n_{ij}}}{n_{ij}!}\right) \exp(-m_{ij})}{\left(\frac{1}{n!}\right) \left(\sum_{i=1}^I \sum_{j=1}^J m_{ij}\right)^n \exp\left(\sum_{i=1}^I \sum_{j=1}^J m_{ij}\right)} \\
&= \binom{n}{n_{11}, \dots, n_{IJ}} \prod_{i=1}^I \prod_{j=1}^J (\theta_{ij}^{n_{ij}}).
\end{aligned}$$

□

1.2.3. Distributions multinomiales indépendantes

Supposons qu'une des deux variables soit explicative et que pour chacune de ses I modalités l'effectif n_{i+} soit connu. Nous pouvons alors modéliser la répartition des effectifs au sein d'un tableau de contingence $I \times J$ avec I distributions multinomiales indépendantes. La fonction de masse pour une modalité i fixée s'écrivant ainsi :

$$P(n_{i1}, \dots, n_{iJ}) = \binom{n_{i+}}{n_{i1}, \dots, n_{iJ}} \prod_{j=1}^J (\theta_{j|i}^{n_{ij}}).$$

Dans ce contexte, $\theta_{j|i}$ peut être écrit sous une autre forme puisque les effectifs de chaque modalité sont fixés :

$$\begin{aligned}
\theta_{i+} &= P[\text{l'observation soit dans la modalité } i] \\
&= \frac{\text{Card}\{\text{la modalité } i\}}{\text{Card}\{\text{l'effectif total}\}} \\
&= \frac{n_{i+}}{n}.
\end{aligned}$$

Nous avons donc :

$$\begin{aligned}\theta_{j|i} &= \frac{\theta_{ij}}{\theta_{i+}} \\ &= \frac{n\theta_{ij}}{n_{i+}}.\end{aligned}\tag{1.2.3}$$

Pour ce qui est de la distribution conjointe, nous l'obtenons en multipliant chacune des I distributions multinomiales :

$$P(n_{11}, \dots, n_{IJ}) = \prod_{i=1}^J \binom{n_{i+}}{n_{i1}, \dots, n_{iJ}} \prod_{j=1}^J \left(\frac{n\theta_{ij}}{n_{i+}}\right)^{n_{ij}}.\tag{1.2.4}$$

1.2.4. Estimateurs du maximum de vraisemblance

Afin de pouvoir effectuer les inférences voulues, il nous faut trouver une façon d'estimer les paramètres de ces trois différentes modélisations. Nous montrons ici comment estimer les paramètres de ces distributions à l'aide de la méthode du maximum de vraisemblance (MV).

Proposition 1.2.2. *Si la répartition des effectifs d'un tableau de contingence suit une distribution de Poisson telle que définie en (1.2.1), alors l'estimation des paramètres selon la méthode du MV est donnée par :*

$$\hat{m}_{ij} = n_{ij}.$$

Proposition 1.2.3. *Si la répartition des effectifs d'un tableau de contingence suit une distribution multinomiale, ou I distributions multinomiales indépendantes telles que définies en (1.2.2) et en (1.2.4), alors l'estimation des paramètres selon la méthode du MV est donnée par :*

$$p_{ij} = \hat{\theta}_{ij} = \frac{n_{ij}}{n}.$$

Pour une preuve complète de ces deux propositions, voir Birch (1963). Ici, nous nous contenterons de démontrer que le vecteur $\frac{1}{n}(n_{11}, \dots, n_{IJ})$ est un point critique de la fonction de vraisemblance d'une loi multinomiale. Birch (1963) démontre que s'il existe un tel point critique à cette fonction de vraisemblance, ce point est unique et il correspond à l'estimateur du MV.

PREUVE. (de la proposition 1.2.3)

Soit L le logarithme de la fonction de vraisemblance donnée par l'équation (1.2.2),

$$\begin{aligned} L &= \log[l(\theta_{11}, \dots, \theta_{IJ})] \\ &= \log[P(n_1, \dots, n_{IJ})] \\ &= \log(n!) - \sum_{i=1}^I \sum_{j=1}^J n_{ij}! + \sum_{i=1}^I \sum_{j=1}^J n_{ij} \log(\theta_{ij}). \end{aligned}$$

Comme la somme des θ_{ij} vaut 1, θ_{IJ} peut s'écrire comme une fonction de tous les autres. Nous avons donc

$$\begin{aligned} \frac{\partial L}{\partial \theta_{ij}} &= n_{ij} \left(\frac{\partial \log(\theta_{ij})}{\partial \theta_{ij}} \right) + n_{IJ} \left(\frac{\partial \log(\theta_{IJ})}{\partial \theta_{ij}} \right) \\ &= \frac{n_{ij}}{\theta_{ij}} - \frac{n_{IJ}}{\theta_{IJ}}. \end{aligned}$$

Les solutions des dérivées partielles de L satisfont donc

$$\frac{\hat{\theta}_{ij}}{\hat{\theta}_{IJ}} = \frac{n_{ij}}{n_{IJ}}$$

Comme nous sommes sous la contrainte que la somme des $\hat{\theta}_{ij}$ vaut 1,

$$1 = \sum_{i=1}^I \sum_{j=1}^J \hat{\theta}_{ij}$$

$$\begin{aligned}
&= \frac{\hat{\theta}_{IJ} \left(\sum_{i=1}^I \sum_{j=1}^J n_{ij} \right)}{n_{IJ}} \\
&= \frac{n \hat{\theta}_{IJ}}{n_{IJ}}.
\end{aligned}$$

Ce qui implique que $\hat{\theta}_{IJ} = \frac{n_{IJ}}{n}$ et $\hat{\theta}_{ij} = \frac{n_{ij}}{n}$. □

L'estimateur MV pour θ_{ij} correspond donc à la proportion des observations dans la cellule (i, j) sur le nombre total d'observations. Nous notons donc p_{ij} l'estimateur du maximum de vraisemblance de θ_{ij} .

Sous la contrainte où nos deux variables aléatoires catégorielles sont indépendantes, l'estimateur du MV pour une distribution multinomiale ne correspond plus à p_{ij} . En remplaçant les θ_{ij} par $\theta_{i+}\theta_{+j}$ dans la fonction de vraisemblance, nous obtenons comme estimateur

$$p_{ij}^* = \hat{\theta}_{i+}\hat{\theta}_{+j} = \left(\frac{n_{i+}}{n} \right) \left(\frac{n_{+j}}{n} \right).$$

La démarche étant similaire à celle utilisée précédemment, nous ne la présentons pas ici.

1.3. TEST D'INDÉPENDANCE AVEC STATISTIQUES QUADRATIQUES

Dans cette section, nous étudions deux tests d'indépendance utilisant des statistiques quadratiques qui convergent en loi vers des variables aléatoires χ^2 . Ces deux tests, valides pour des observations provenant d'une distribution multinomiale, furent introduits par Pearson (1900) et Wilks (1935). Nous verrons également que ces deux tests, tout comme celui basé sur la loi asymptotique du rapport de cotes (voir la section 1.4), ne sont adéquats que si la taille échantillonnale est très grande, ce qui n'est pas le cas du test exact de Fisher que nous verrons à la section 1.5.

1.3.1. La statistique X^2 de Pearson

Pearson (1900) introduit un des premiers tests d'adéquation, il visait à vérifier s'il était justifié de modéliser nos observations par une loi multinomiale avec un vecteur de probabilité fixé. Nous décrivons ici ce test d'adéquation et nous montrons, à la sous-section suivante, comment l'appliquer afin de tester l'indépendance de variables aléatoires catégorielles.

Soit H_0 l'hypothèse nulle stipulant que la répartition des IJ effectifs suit une distribution multinomiale de paramètre $\underline{\theta}^0$ telle que $\sum_{i=1}^I \sum_{j=1}^J \theta_{ij}^0 = 1$. Considérons la statistique

$$X^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - m_{ij})^2}{m_{ij}}, \quad (1.3.1)$$

où $m_{ij} = E_{H_0}[N_{ij}] = n\theta_{ij}^0$. En vertu de la proposition suivante, nous rejetons l'hypothèse nulle lorsque cette statistique prend de grandes valeurs. Pour ce qui est de la valeur-p, elle correspond à la probabilité qu'une variable aléatoire χ^2 avec $(IJ - 1)$ degrés de liberté soit supérieure à la valeur observée de X^2 .

Proposition 1.3.1. *Si la répartition des effectifs d'un tableau de contingence $I \times J$ suit une loi multinomiale de paramètres $\underline{\theta}^0$, alors la statistique X^2 définie en (1.3.1) converge en loi vers une χ^2 à $(IJ - 1)$ degrés de liberté.*

PREUVE.

Soit $\underline{p} = (p_{11}, \dots, p_{IJ})$ et $\underline{\theta}^0 = (\theta_{11}^0, \dots, \theta_{I(J-1)}^0)$. Nous savons par le théorème limite central que

$$\sqrt{n}(\underline{p} - \underline{\theta}^0)' \xrightarrow{\mathcal{L}} \mathcal{N}_{IJ-1}(0, \Sigma_0),$$

où Σ_0 est la matrice de covariance de $\sqrt{n}\underline{p}$, de dimension $(IJ - 1) \times (IJ - 1)$. Ses éléments $\sigma_{ij,kl}$ sont tels que :

$$\sigma_{ij,kl} = \frac{Cov(n_{ij}, n_{kl})}{n} = \begin{cases} \theta_{ij}^0(1 - \theta_{ij}^0) & \text{si } (i, j) = (k, l), \\ -\theta_{ij}^0\theta_{kl}^0 & \text{sinon.} \end{cases} \quad (1.3.2)$$

Nous pouvons alors montrer que l'équation (1.3.1) peut s'écrire sous la forme

$$X^2 = n(\underline{p} - \underline{\theta}^0)' \Sigma_0^{-1} (\underline{p} - \underline{\theta}^0)$$

(voir Agresti, 1990 page 44). Par conséquent, $X^2 \xrightarrow{\mathcal{L}} \chi_{(IJ-1)}^2$. \square

1.3.2. Test d'indépendance avec la statistique X^2

Dans un tableau de contingence $I \times J$ où les observations suivent une loi multinomiale, l'hypothèse d'indépendance

$$H_0 : \theta_{ij} = \theta_{i+}\theta_{+j} \text{ pour tout } i \text{ et tout } j, \quad (1.3.3)$$

n'est pas simple comme celle vue à la sous-section précédente, les θ_{i+} et les θ_{+j} étant inconnus. Afin de pallier ce problème, Pearson (1900, 1922) suggèrent d'utiliser la statistique X^2 en remplaçant les m_{ij} dans l'équation (1.3.1) par

$$\hat{m}_{ij}^* = n \hat{p}_{ij}^* = \frac{n_{i+}n_{+j}}{n}. \quad (1.3.4)$$

Fisher (1922) montre que cette modification de la statistique X^2 , en estimant les $I + J$ probabilités marginales, ne change pas sa convergence en loi vers une χ^2 , mais par contre les degrés de libertés sont réduits du nombre de paramètres estimés, passant de $(IJ - 1)$ à $(I - 1)(J - 1)$. Nous avons donc notre premier test d'indépendance, il utilise la statistique X^2 telle que

$$X^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - \hat{m}_{ij}^*)^2}{\hat{m}_{ij}^*},$$

avec comme valeur-p

$$P[\chi_{(I-1)(J-1)}^2 \geq X^2].$$

Comme nous utilisons la distribution asymptotique de la statistique, il n'est pas approprié d'effectuer ce test à partir d'un échantillon de petite taille. Cochran (1954) suggère de ne pas utiliser ce test si plus de 20% des \hat{m}_{ij}^* sont inférieurs à 5.

Nous remarquons que cette statistique est calculée à partir d'une distribution discrète mais est approximée par une distribution continue. Yates (1934) suggère donc d'utiliser une correction de continuité dans le cas des tableaux de contingence 2×2 . La statistique devient donc

$$X_c^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(|n_{ij} - \hat{m}_{ij}^*| - 0.5)^2}{\hat{m}_{ij}^*}, \quad (1.3.5)$$

où X_c^2 a la même distribution asymptotique que X^2 .

1.3.3. Test du rapport de vraisemblance de Wilks

Nous pouvons également tester l'indépendance de nos variables aléatoires en utilisant le test du rapport de vraisemblance (RV). Soit H_0 l'hypothèse d'indépendance formulée en (1.3.3) pour une distribution multinomiale. Le rapport de vraisemblance Λ est défini comme suit :

$$\Lambda = \frac{\max_{\underline{\theta} \in \omega_0} [l(\underline{\theta})]}{\max_{\underline{\theta} \in \Omega} [l(\underline{\theta})]},$$

où Ω représente l'ensemble des valeurs de $\underline{\theta}$ et ω_0 est un sous-ensemble de Ω qui représente l'ensemble des valeurs que $\underline{\theta}$ peut prendre sous H_0 . Nous remarquons

que les termes au numérateur et au dénominateur seront maximisés si nous remplaçons $\underline{\theta}$ par les estimateurs du MV contraints et non contraints énoncés à la section 1.2.4.

Wilks (1935, 1938) démontra que $-2 \log(\Lambda)$ converge en loi vers une variable aléatoire χ^2 dont le nombre de degrés de liberté correspond à $\dim(\Omega) - \dim(\omega_0)$. Cette statistique est notée G^2 . Nous avons :

$$\begin{aligned}
 G^2 &= -2 \log(\Lambda) \\
 &= -2 \log \left[\frac{l(\underline{p}^*)}{l(\underline{p})} \right] \\
 &= -2 \log \left[\frac{\prod_{i=1}^I \prod_{j=1}^J \left(\frac{n_{i+} n_{+j}}{n^2} \right)^{n_{ij}}}{\prod_{i=1}^I \prod_{j=1}^J \left(\frac{n_{ij}}{n} \right)^{n_{ij}}} \right] \\
 &= 2 \sum_{i=1}^I \sum_{j=1}^J n_{ij} \log \left(\frac{n_{ij} n}{n_{i+} n_{+j}} \right). \tag{1.3.6}
 \end{aligned}$$

Comme $\dim(\Omega) = IJ - 1$ et $\dim(\omega_0) = (I - 1) + (J - 1)$, le nombre de degrés de liberté est $(I - 1)(J - 1)$.

Nous remarquons que sous l'hypothèse nulle, X^2 et G^2 convergent asymptotiquement vers la même distribution. Il est donc intéressant d'étudier si l'une de ces deux statistiques converge plus rapidement que l'autre vers une variable aléatoire de densité $\chi_{(I-1)(J-1)}^2$. Larntz (1978), Koehler et Larntz (1980) et Koehler (1986) énoncent tous qu'il est préférable d'utiliser la statistique X^2 pour tester l'indépendance à partir d'un petit échantillon. Ils montrent aussi que la distribution de G^2 n'est pas bien approximée par une χ^2 quand $\frac{n}{IJ}$ est inférieur à 5.

1.4. TEST PAR LE RAPPORT DE COTES

À partir de la proposition 1.1.2, nous savons que tester l'indépendance de deux variables aléatoires catégorielles équivaut à tester si les rapports de cotes α_{ij} prennent des valeurs différentes de 1. Dans cette section, nous présentons comment tester l'indépendance à l'aide d'intervalles de confiance pour les rapports de cotes.

Si nos observations proviennent d'une loi multinomiale ou de I lois multinomiales indépendantes, nous pouvons estimer les rapports de cotes à l'aide des estimateurs du MV de $\delta_{ij,kl}$,

$$\hat{\delta}_{ij,kl} = \frac{\hat{\theta}_{ij}\hat{\theta}_{kl}}{\hat{\theta}_{il}\hat{\theta}_{kj}} = \frac{n_{ij}n_{kl}}{n_{il}n_{kj}}. \quad (1.4.1)$$

Comme les n_{ij} ont une probabilité non nulle de valoir zéro, il est impossible de calculer l'espérance et la variance de $\hat{\delta}_{ij,kl}$. Gart et Zweifl (1967) ainsi que Haldane (1955) proposent l'estimateur

$$\bar{\delta}_{ij,kl} = \frac{(n_{ij} + 0.5)(n_{kl} + 0.5)}{(n_{il} + 0.5)(n_{kj} + 0.5)} \quad (1.4.2)$$

afin de corriger ce problème.

Considérons maintenant le logarithme de l'estimation d'un rapport de cotes

$$\log(\hat{\delta}_{ij,kl}) = \log(n_{ij}) + \log(n_{kl}) - \log(n_{il}) - \log(n_{kj}).$$

Sa structure étant additive plutôt que multiplicative, sa distribution converge plus rapidement vers une loi normale que celle de $\hat{\delta}_{ij,kl}$. Notre approche, afin de tester l'indépendance, consiste à vérifier si les intervalles de confiance pour $\log(\alpha_{ij})$ contiennent simultanément zéro. À partir de la proposition suivante, nous voyons qu'un intervalle de confiance de niveau $(1 - \alpha)$ pour $\log(\delta_{ij,kl})$ est donné par :

$$\log(\hat{\delta}_{ij,kl}) \pm z_{\alpha/2} \sqrt{\frac{1}{n_{ij}} + \frac{1}{n_{kl}} + \frac{1}{n_{il}} + \frac{1}{n_{kj}}}.$$

Proposition 1.4.1. Soit $\hat{\delta}_{ij,kl}$ l'estimateur de $\delta_{ij,kl}$ tel que spécifié dans l'équation (1.4.1). Alors,

$$\hat{\phi}_{ij,kl} = \log(\hat{\delta}_{ij,kl}) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\phi_{ij,kl}, \hat{\sigma}^2(\hat{\phi}_{ij,kl})\right),$$

$$\text{où } \hat{\sigma}^2(\hat{\phi}_{ij,kl}) = \frac{1}{n_{ij}} + \frac{1}{n_{kl}} + \frac{1}{n_{il}} + \frac{1}{n_{kj}}.$$

PREUVE.

Par le théorème limite central, nous avons que

$$\underline{p} \xrightarrow{\mathcal{L}} \mathcal{N}(\underline{\theta}, n^{-1}\Sigma_0),$$

où les éléments de Σ_0 sont définis à l'équation (1.3.2).

Posons que

$$g(\underline{p}) = \log(\hat{\delta}_{ij,kl}) = \log(p_{a_1}) + \log(p_{a_2}) - \log(p_{a_3}) - \log(p_{a_4}),$$

où $a_1 = ij$, $a_2 = kl$, $a_3 = il$ et $a_4 = kj$.

Si $\theta_{a_i} \neq 0 \forall i$, alors

$$g(\underline{p}) \xrightarrow{\mathcal{L}} \mathcal{N}(g(\underline{\theta}), \sigma_g^2),$$

où

$$g(\underline{\theta}) = \log\left(\frac{\theta_{a_1}\theta_{a_2}}{\theta_{a_3}\theta_{a_4}}\right)$$

et

$$\begin{aligned} \sigma_g^2 &= \frac{1}{n} \sum_{(i,j)} \sum_{(k,l)} \sigma_{ij,kl} \left. \frac{\partial g}{\partial \theta_{ij}} \right|_{\underline{p}=\underline{\theta}} \left. \frac{\partial g}{\partial \theta_{kl}} \right|_{\underline{p}=\underline{\theta}} \\ &= \frac{1}{n} \sum_{i=1}^4 \sum_{j=1}^4 \sigma_{a_i, a_j} \left. \frac{\partial g}{\partial \theta_{a_i}} \right|_{\underline{p}=\underline{\theta}} \left. \frac{\partial g}{\partial \theta_{a_j}} \right|_{\underline{p}=\underline{\theta}} \\ &= \frac{1}{n} \sum_{i=1}^4 \left(\frac{1 - \theta_{a_i}}{\theta_{a_i}} \right) - 1 \end{aligned}$$

$$= \frac{1}{n} \sum_{i=1}^4 \frac{1}{\theta_{a_i}}.$$

Comme

$$\hat{\sigma}^2(\hat{\phi}_{a_1, a_2}) = \sum_{i=1}^4 \frac{1}{n_{a_i}} \xrightarrow{\mathcal{P}} \sigma_g^2,$$

nous obtenons le résultat en appliquant le théorème de Slutsky. \square

Corollaire 1.4.1. *Soit $\tilde{\delta}_{ij,kl}$ l'estimateur de $\delta_{ij,kl}$ tel que spécifié dans l'équation (1.4.2). Alors,*

$$\tilde{\phi}_{ij,kl} = \log(\tilde{\delta}_{ij,kl}) \xrightarrow{\mathcal{L}} \mathcal{N} \left(\phi_{ij,kl}, \frac{1}{n_{ij} + 0.5} + \frac{1}{n_{kl} + 0.5} + \frac{1}{n_{il} + 0.5} + \frac{1}{n_{kj} + 0.5} \right).$$

Dans le cas d'un tableau de contingence 2×2 , H_0 sera rejetée, au niveau α , si zéro ne fait pas partie de l'intervalle de confiance $\log(\hat{\delta}) \pm z_{\alpha/2} \sigma(\phi)$. Pour ce qui est des tableaux de contingence $I \times J$, il est possible de tester si tous les $\log(\alpha_{ij})$ prennent 0 comme valeur. Nous voyons à partir de la proposition 1.1.2 que cela équivaut à tester l'indépendance de deux variables aléatoires catégorielles. Ce test se déduit de Goodman (1964).

À partir des résultats de Plackett (1962) et Goodman (1963), nous pouvons supposer que $\log(N_{ij})$ suit asymptotiquement une loi normale de moyenne $\log(n\theta_{ij})$ et de variance $1/n_{ij}$, nous pouvons de plus supposer qu'ils sont tous asymptotiquement indépendants. Cela implique que

$$\log(\hat{\alpha}_{ij}) \xrightarrow{\mathcal{L}} \mathcal{N} \left(\log(\alpha_{ij}), \frac{1}{n_{ij}} + \frac{1}{n_{IJ}} + \frac{1}{n_{iJ}} + \frac{1}{n_{Ij}} \right),$$

où $\hat{\alpha}_{ij} = \frac{n_{ij}n_{IJ}}{n_{iJ}n_{Ij}}$. Comme $\hat{\alpha}_{ij} \equiv 0$ si $i = I$ ou $j = J$, nous ne sommes intéressés qu'aux $(I-1)(J-1)$ autres $\log(\hat{\alpha}_{ij})$. Posons $\hat{\underline{\alpha}} = (\hat{\alpha}_{11}, \dots, \hat{\alpha}_{(I-1)(J-1)})$. Nous

obtenons alors que

$$\log(\hat{\mathcal{Q}}) \xrightarrow{\mathcal{L}} \mathcal{N}_{(I-1)(J-1)}(\mathbf{0}, \mathcal{V})$$

lorsque les variables aléatoires catégorielles sont indépendantes. La matrice \mathcal{V} a comme éléments

$$\mathcal{V}_{(i,j),(k,l)} = \begin{cases} \frac{1}{n_{ij}} + \frac{1}{n_{IJ}} + \frac{1}{n_{iJ}} + \frac{1}{n_{Ij}} & \text{si } (i, j) = (k, l), \\ \frac{1}{n_{IJ}} + \frac{1}{n_{iJ}} & \text{si } i = k \text{ et } j \neq l, \\ \frac{1}{n_{IJ}} + \frac{1}{n_{Ij}} & \text{si } i \neq k \text{ et } j = l, \\ \frac{1}{n_{IJ}} & \text{si } i \neq k \text{ et } j \neq l, \end{cases}$$

où $i, k = 1, \dots, (I - 1)$ et $j, l = 1, \dots, (J - 1)$. Nous pouvons alors utiliser la statistique

$$(\log(\hat{\mathcal{Q}}))\mathcal{V}^{-1}(\log(\hat{\mathcal{Q}}))'$$

qui converge en loi vers une variable aléatoire de densité $\chi_{(I-1)(J-1)}^2$.

1.5. TEST EXACT DE FISHER

Dans cette section, nous présentons un test d'indépendance n'utilisant pas la distribution échantillonnale asymptotique, mais plutôt sa distribution exacte. Nous pouvons alors tester l'indépendance de variables aléatoires catégorielles même avec de très petits échantillons. Par contre, ce test requiert que tous les effectifs marginaux soient fixes. Ce test, appelé test exact de Fisher (voir Fisher, 1934) est présenté ici dans le cas d'un tableau de contingence 2×2 , Agresti (1990, sous-section 3.5.3) en présente une généralisation aux tableaux de contingence $I \times J$.

Soit un tableau 2×2 où les effectifs sont répartis selon une des trois distributions vues à la section 1.2. Sous l'hypothèse d'indépendance, la fonction de masse

$p(n_{11}, \dots, n_{22})$ ne dépend d'aucun paramètre si nous connaissons les valeurs de n_{i+} et n_{+j} . Ainsi, le fait de connaître n_{11} nous permet de déterminer la valeur des trois autres effectifs. De plus, nous savons alors que n_{11} suit une loi hypergéométrique (voir Agresti, 1990, sous-section 3.5.1) telle que

$$p(n_{11}, \dots, n_{22}) = p(n_{11}) = \frac{\binom{n_{1+}}{n_{11}} \binom{n_{2+}}{n_{+1} - n_{11}}}{\binom{n}{n_{+1}}},$$

où $n_{11} \in \{m_-, \dots, m^-\}$, avec $m_- = \max(0, n_{1+} + n_{+1} - n)$ et $m^- = \min(n_{1+}, n_{+1})$.

Dans ce contexte, le rapport de cotes est uniquement déterminé par la valeur de n_{11} et leur variation va dans le même sens. Donc, si nous testons $H_0 : \delta = 1$ versus $H_1 : \delta < 1$, la valeur-p sera déterminée par

$$\alpha = \sum_{i=m_-}^{n_{11}} \frac{\binom{n_{1+}}{i} \binom{n_{2+}}{n_{+1} - i}}{\binom{n}{n_{+1}}}. \quad (1.5.1)$$

Tocher (1950) montre que ce test, connu sous le nom de test exact de Fisher, est le test uniformément le plus puissant lorsqu'il est randomisé. Notons aussi que si le test est bilatéral, la valeur-p est souvent définie comme la somme des $p(i)$ tels que $p(i) \leq p(n_{11})$.

Toutes les méthodes fréquentistes présentées dans ce chapitre seront utilisées avec quelques jeux de données à la fin du chapitre 3. Nous les comparerons alors avec différentes méthodes basées sur l'approche bayésienne, approche que nous présentons dans le chapitre suivant.

Chapitre 2

APPROCHE BAYÉSIENNE POUR TESTER L'INDÉPENDANCE DES VARIABLES CATÉGORIELLES

Ce chapitre présente différentes modélisations, pour des variables aléatoires catégorielles, basées sur la théorie bayésienne. Après avoir introduit les concepts de base de la théorie bayésienne, nous présentons, à la section 2.2, le modèle bayésien le plus souvent utilisé avec des variables aléatoires catégorielles. Par la suite, nous présentons divers modèles bayésiens en privilégiant ceux qui permettent l'incorporation d'une idée *a priori* d'indépendance.

2.1. CONCEPTS DE BASE DE LA THÉORIE BAYÉSIENNE

Les approches classiques que nous avons présentées dans le chapitre précédent étaient essentiellement basées sur l'information échantillonnale. La théorie bayésienne consiste plutôt à combiner l'information sur les paramètres apportée par l'échantillon à celle étant disponible *a priori*. Nous voyons dans cette section comment modéliser cette information *a priori* et comment la combiner à l'information échantillonnale. Dans cette section, toutes les définitions, propositions et théorèmes énoncés sont issus de Robert (1992).

2.1.1. Le modèle statistique bayésien

Soit le vecteur des paramètres inconnus $\underline{\theta}$, appartenant à un sous-ensemble Θ d'un espace vectoriel de dimension finie. Afin de pouvoir définir ce qu'est un modèle statistique bayésien, il faut présenter ici la loi *a priori* $\pi(\underline{\theta})$, une loi de probabilité sur les paramètres inconnus. Nous étudierons ce type de loi dans la sous-section suivante.

Définition 2.1.1. *Nous appelons modèle statistique bayésien la donnée d'un modèle statistique paramétré $f(\underline{x} | \underline{\theta})$ et d'une loi a priori sur les paramètres $\pi(\underline{\theta})$.*

L'innovation principale par rapport aux modèles classiques est donc de proposer en sus une loi de probabilité sur les paramètres.

Ce modèle statistique bayésien nous permet de déterminer la loi *a posteriori* $\pi(\underline{\theta} | \underline{x})$, qui combine l'information contenue dans la loi *a priori* et dans l'échantillon. Elle est obtenue en appliquant le théorème de Bayes.

Théorème 2.1.1. *Soit $\pi(\underline{\theta})$ une loi a priori sur le vecteur des paramètres $\underline{\theta}$ et $f(\underline{x} | \underline{\theta})$ la densité de \underline{x} . Alors, la distribution de $\underline{\theta}$ conditionnellement à \underline{x} , ou la distribution a posteriori, a pour densité*

$$\pi(\underline{\theta} | \underline{x}) = \frac{f(\underline{x} | \underline{\theta})\pi(\underline{\theta})}{m(\underline{x})},$$

où $m(\underline{x}) = \int_{\Theta} f(\underline{x} | \underline{\theta})\pi(\underline{\theta})d\underline{\theta}$.

Cette distribution *a priori* constitue la base de toute inférence bayésienne. Nous verrons d'ailleurs à la fin de cette section quelques inférences sur $\underline{\theta}$ basées sur cette distribution.

2.1.2. Lois *a priori* conjuguées et non informatives

Les difficultés majeures reliées à l'application de la théorie bayésienne sont la détermination et la paramétrisation adéquate de la loi *a priori*. Plusieurs techniques ont été développées afin de pallier ces difficultés, nous présentons ici les deux techniques que nous utiliserons par la suite : l'utilisation d'une loi *a priori* provenant d'une famille de lois conjuguée ainsi que les lois *a priori* non informatives de Jeffreys.

Définition 2.1.2. Une famille \mathcal{F} de lois sur Θ est dite conjuguée si, pour tout $\pi \in \mathcal{F}$, la loi *a posteriori* $\pi(\theta | \underline{x})$ appartient également à \mathcal{F} .

Si nous nous restreignons au cas où \mathcal{F} est paramétrée, le passage de la loi *a priori* à la loi *a posteriori* se réduit à un changement de paramètre. L'attrait de l'utilisation des lois *a priori* conjuguées est donc que les lois *a posteriori* sont toujours calculables. De plus, Raiffa et Schlaiffer (1961) justifient cette approche par le fait que l'information apportée par \underline{x} sur θ étant limitée, elle ne devrait donc pas conduire à une remise en cause de la détermination de π mais seulement sur sa paramétrisation.

Quant aux lois *a priori* non informatives, nous les utilisons lorsqu'il est impossible de bâtir une distribution *a priori* basée sur des considérations subjectives. Dans ce contexte, nous pouvons entre autres utiliser la loi *a priori* non informative introduite par Jeffreys (1961), basée sur *l'information de Fisher* qui est définie par :

$$\mathcal{I}(\theta) = \mathbb{E}_{\underline{x}} \left[\frac{\partial}{\partial \theta} \left(\log [f(\underline{X} | \theta)] \right) \right]^2,$$

où $\mathbb{E}_{\underline{x}}$ représente l'espérance prise par rapport à $f(\underline{x} | \theta)$ et où θ est unidimensionnel. La loi *a priori* est alors définie de façon à être proportionnelle à la racine

carrée de l'information sur θ apportée par $f(\underline{x} | \theta)$, c'est-à-dire

$$\pi(\theta) \propto \sqrt{\mathcal{I}(\theta)}.$$

Les lois de Jeffreys ont la propriété très intéressante d'être invariantes pour une reparamétrisation $\eta = h(\theta)$ où h est bijective, cela étant dû au fait que

$$\mathcal{I}(\theta) = \mathcal{I}(h(\theta)) \left(h'(\theta) \right)^2.$$

Si $\underline{\theta} \in \mathbb{R}^p$ où $p > 1$, la loi *a priori* non informative de Jeffreys est alors donnée par

$$\pi(\underline{\theta}) \propto \sqrt{\det(\mathcal{I}(\underline{\theta}))},$$

où $\mathcal{I}(\underline{\theta})$ est une matrice de dimension $p \times p$. Les éléments de cette matrice sont notés $\mathcal{I}_{ij}(\underline{\theta})$ et sont égaux à

$$-\mathbb{E}_{\underline{x}} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \left(\log [f(\underline{X} | \underline{\theta})] \right) \right].$$

2.1.3. Estimation bayésienne ponctuelle

Nous voyons dans cette sous-section comment utiliser la loi *a posteriori* pour estimer ponctuellement les paramètres inconnus. Nous procédons en trouvant la règle de Bayes pour un modèle bayésien et une fonction de perte donnés. Il nous faut donc tout d'abord définir ce qu'est une fonction de perte.

Définition 2.1.3. Soit Θ un sous-ensemble de l'espace vectoriel des paramètres et soit \mathcal{A} , l'ensemble des valeurs possibles pour les estimateurs ponctuels de Θ . Nous appelons alors fonction de perte toute fonction $L(\underline{\theta}, \delta)$ telle que $L : \underline{\theta} \times \mathcal{A} \rightarrow (-k^2, \infty)$ où $k \in \mathbb{R}$.

Une telle fonction mesure en quelque sorte le coût engendré par l'erreur que nous faisons lors de notre inférence. Nous présenterons à la fin de cette sous-section une fonction de perte que nous utiliserons beaucoup par la suite, la perte quadratique. Voyons auparavant le rôle que la fonction de perte joue dans la détermination d'un estimateur ponctuel.

La véritable perte encourue étant fonction des paramètres inconnus ainsi que de la fonction de vraisemblance du modèle, une façon de procéder consiste à choisir un estimateur qui minimise des pertes espérées. Voici trois différents types de perte espérée.

Définition 2.1.4. *Soit \mathcal{D} l'ensemble des règles de décision possibles.*

i) Le risque fréquentiste, $R(\underline{\theta}, \delta)$ où $\delta \in \mathcal{D}$, est l'espérance de la fonction de perte sous la fonction de densité de l'échantillon,

$$R(\underline{\theta}, \delta) = \mathbb{E}_{\underline{x}} [L(\underline{\theta}, \delta)] = \int_{\mathcal{X}} L(\underline{\theta}, \delta(\underline{x})) f(\underline{x} | \underline{\theta}) d\underline{x}.$$

ii) Le risque de Bayes, $r(\pi, \delta)$, est défini comme l'espérance de $R(\underline{\theta}, \delta)$ sous la loi a priori de $\underline{\theta}$,

$$r(\pi, \delta) = \mathbb{E}_{\underline{\theta}} [R(\underline{\theta}, \delta)] = \int_{\Theta} \int_{\mathcal{X}} L(\underline{\theta}, \delta(\underline{x})) f(\underline{x} | \underline{\theta}) \pi(\underline{\theta}) d\underline{x} d\underline{\theta}.$$

iii) La perte espérée a posteriori, $\rho(\pi, \delta | \underline{x})$, est définie comme étant l'espérance de la fonction de perte sous la loi a posteriori ,

$$\rho(\pi, \delta | \underline{x}) = \mathbb{E}_{\underline{\theta} | \underline{x}} [L(\underline{\theta}, \delta)] = \int_{\Theta} L(\underline{\theta}, \delta(\underline{x})) \pi(\underline{\theta} | \underline{x}) d\underline{\theta}.$$

L'estimateur que nous utiliserons par la suite sera la règle de Bayes. Elle peut être calculée à partir des définitions 2.1.4 *ii*) ou 2.1.4 *iii*).

Définition 2.1.5. *Pour une loi a priori π et une fonction de perte $L(\theta, \delta)$ données, la règle de Bayes δ^π est la règle de décision telle que*

$$r(\pi, \delta^\pi) = \min_{\delta \in \mathcal{D}} r(\pi, \delta).$$

Théorème 2.1.2. *Si la règle de Bayes existe, elle peut être obtenue en choisissant, pour chaque $\underline{x} \in \mathcal{X}$, la valeur de a qui minimise $\rho(\pi, a | \underline{x})$.*

Nous terminons cette sous-section avec une proposition donnant la règle de Bayes sous la fonction de perte quadratique $L(\theta, \delta) = (\theta - \delta)^2$.

Proposition 2.1.1. *La règle de Bayes sous la fonction de perte quadratique correspond à la moyenne de la loi a posteriori, c'est-à-dire*

$$\delta^\pi(x) = E_{\theta|\underline{x}}[\theta | \underline{x}] = \int_{\Theta} \theta \pi(\theta | \underline{x}) d\theta.$$

2.1.4. Région de crédibilité

L'approche bayésienne propose la détermination de régions α -crédibles comme alternative aux intervalles de confiance de niveau $1 - \alpha$.

Définition 2.1.6. *Un ensemble $\mathcal{C}_{\underline{x}}$ est une région α -crédible si*

$$P_{\theta|\underline{x}}(\theta \in \mathcal{C}_{\underline{x}} | \underline{x}) \geq 1 - \alpha.$$

Parmi les différentes régions α -crédibles, notons la *région α -crédible PHDP* (plus haute densité *a posteriori*), région qui a l'avantage d'avoir un volume

minimal parmi les régions α -crédibles. Nous reviendrons sur ces concepts à la fin du chapitre 3.

Définition 2.1.7. *Un ensemble C_x^π est une région α -crédible PHDP s'il s'écrit sous la forme*

$$C_x^\pi = \{\underline{\theta} \mid \pi(\underline{\theta} \mid \underline{x}) \geq k_\alpha\},$$

où k_α est la plus grande valeur telle que

$$P_{\underline{\theta} \mid \underline{x}}(\underline{\theta} \in C_x^\pi \mid \underline{x}) \geq 1 - \alpha.$$

Ces régions de crédibilité seront utilisées comme tests d'hypothèses par la suite. Par exemple, si nous désirons tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ au niveau α , nous rejettons H_0 si $\theta_0 \notin C_x^\pi$.

2.2. MODÉLISATION BAYÉSIENNE USUELLE POUR LES VARIABLES CATÉGORIELLES

Ayant énuméré les concepts de base de la modélisation bayésienne, nous pouvons maintenant appliquer cette théorie dans notre étude des variables aléatoires catégorielles. Dans cette section, nous étudions la loi de Dirichlet, une loi *a priori* souvent utilisée lorsque les observations proviennent d'une distribution multinomiale. Nous voyons d'ailleurs que dans ce contexte, cette distribution forme une famille conjuguée et que la loi non informative de Jeffreys appartient à cette famille. Après avoir vu quelques propriétés et quelques façons de déterminer les paramètres de cette loi, nous montrerons la relation entre cette modélisation bayésienne et celle menant au test exact de Fisher. Il est important de spécifier que jusqu'à la fin de ce chapitre, ainsi que dans le chapitre suivant, les paramètres $\underline{\theta} = (\theta_{11}, \dots, \theta_{IJ})$, et les composantes du vecteur aléatoire $\underline{n} = (n_{11}, \dots, n_{IJ})$ sont

notés avec un seul indice lorsqu' il n'y aura pas d'ambiguïté. Ainsi, pour les données regroupées dans un tableau de contingence 2×2 , l'indice 11 devienra 1, 12 devienra 2, 21 devienra 3 et 22 devienra 4.

2.2.1. Définition de la loi de Dirichlet

Supposons que le vecteur aléatoire \underline{N} suit une distribution multinomiale de paramètres $(n, \underline{\theta})$ où $n = \sum_{j=1}^p N_j$ est connu, mais $\underline{\theta}$ est inconnu et défini sur le simplexe $\mathcal{S}^p = \{\theta_j \geq 0, i = 1, \dots, p; \sum_{j=1}^p \theta_j = 1\}$. Dans ce contexte, une loi *a priori* fréquemment utilisée est la loi de Dirichlet.

Définition 2.2.1. *Le vecteur $\underline{\theta} \in \mathcal{S}^p$ admet une loi de Dirichlet de paramètres a_j (notée $\mathcal{D}(\underline{a})$) si et seulement si*

$$\pi(\underline{\theta}) = B(\underline{a})^{-1} \prod_{j=1}^p \theta_j^{a_j-1},$$

où $\underline{a} = (a_1, \dots, a_p) \in (\mathbb{R}^+)^p$ et $B(\underline{a}) = \prod_{j=1}^p \Gamma(a_j) / \Gamma(\sum_{j=1}^p a_j)$.

La proposition suivante nous permet de calculer les différents moments d'un vecteur aléatoire admettant une loi de Dirichlet.

Proposition 2.2.1. *Si $\underline{\theta}$ admet comme distribution une loi de Dirichlet de paramètres a_j , alors*

$$\mathbb{E}\left[\prod_{j=1}^p \theta_j^{m_j}\right] = \frac{B(\underline{a} + \underline{m})}{B(\underline{a})}.$$

PREUVE.

$$\mathbb{E}\left[\prod_{j=1}^p \theta_j^{m_j}\right] = B(\underline{a})^{-1} \int_{\mathcal{S}^p} \prod_{j=1}^p \theta_j^{a_j+m_j-1} d\underline{\theta}$$

$$= \frac{B(\underline{a} + \underline{m})}{B(\underline{a})}.$$

□

Corollaire 2.2.1. *Si $\underline{\theta}$ admet comme distribution une loi de Dirichlet de paramètres a_j , alors*

$$\begin{aligned} \mathbb{E}[\theta_j] &= \frac{a_j}{a_+}; \\ \mathbb{V}[\theta_j] &= \frac{a_j(a_+ - a_j)}{a_+^2(a_+ + 1)}; \\ \text{Cov}[\theta_j, \theta_i] &= \frac{-a_j a_i}{a_+^2(a_+ + 1)}; \\ \mathbb{E}\left[\frac{\theta_1 \theta_4}{\theta_2 \theta_3}\right] &= \frac{a_1 a_4}{(a_2 - 1)(a_3 - 1)}; \end{aligned}$$

où $a_+ = \sum_{j=1}^p a_j$.

Nous remarquons que multiplier le vecteur \underline{a} par un scalaire positif n'affecte pas l'espérance d'un paramètre, mais cela modifie sa variance et sa covariance.

Notons qu'il arrive aussi que la loi de Dirichlet de paramètres a_j soit écrite sous la forme $\mathcal{D}(a_+, \underline{\eta})$. Sous cette nouvelle paramétrisation, la densité s'écrit

$$\pi(\underline{\theta}) = B(a_+, \underline{\eta})^{-1} \prod_{j=1}^p \theta_j^{a_j + \eta_j - 1}, \quad (2.2.1)$$

où $\underline{\eta} \in \mathcal{S}^p$.

Nous remarquons alors que

$$\mathbb{E}[\theta_j] = \eta_j, \quad (2.2.2)$$

$$\mathbb{V}[\theta_j] = \frac{\eta_j(1 - \eta_j)}{(a_+ + 1)}. \quad (2.2.3)$$

Sous cette paramétrisation, nous remarquons mieux la signification des différents paramètres de cette loi, $\underline{\eta}$ correspondant à notre idée *a priori* de la valeur de $\underline{\theta}$

et a_+ représentant en quelque sorte l'importance que nous accordons *a priori* à cette idée. Plusieurs valeurs de a_+ sont suggérées et utilisées dans la littérature, les principales étant 0, 1, $p/2$, p ainsi que \sqrt{n} (voir Walley, 1997). Notons que lorsque nous posons $a_+ = \infty$, cela veut dire que nous sommes certain que $\underline{\theta} = \underline{\eta}$.

Comme nous l'avons mentionné à la section précédente, la base de toute inférence bayésienne est la loi *a posteriori* des paramètres inconnus. Lorsque la loi *a priori* est celle de Dirichlet et que les observations proviennent d'une loi multinomiale, nous savons en vertu du théorème 2.1.1 que

$$\pi(\underline{\theta} \mid \underline{n}) \propto \left(\prod_{j=1}^p \theta_j^{a_j-1} \right) \left(\prod_{j=1}^p \theta_j^{n_j} \right) = \prod_{j=1}^p \theta_j^{a_j+n_j-1}.$$

La loi *a posteriori* de $\underline{\theta}$ est donc une Dirichlet de paramètres $a_j + n_j$. Cette loi forme donc une famille de lois conjuguée lorsque les observations ont une distribution multinomiale. Nous pouvons donc facilement calculer la règle de Bayes pour estimer θ_j sous le coût quadratique, car l'espérance *a posteriori* de θ_j est

$$\frac{a_+}{a_+ + n} \left(\frac{a_j}{a_+} \right) + \frac{n}{a_+ + n} \left(\frac{n_j}{n} \right). \quad (2.2.4)$$

Cela correspond à une moyenne pondérée de l'espérance *a priori* de θ_j et de l'estimateur du maximum de vraisemblance de θ_j , avec des poids respectifs de $\frac{a_+}{a_+ + n}$ et $\frac{n}{a_+ + n}$. En quelque sorte, nous pouvons donc considérer a_+ comme étant la taille de notre échantillon *a priori*. Ainsi, choisir p_{ij} comme estimateur de θ_{ij} correspond à choisir $a_+ = 0$ (voir la proposition 1.2.3).

Pour ce qui est de la détermination de la loi *a priori* non informative de Jeffreys, la proposition suivante énonce que cette loi *a priori* est également une Dirichlet.

Proposition 2.2.2. *Soit le vecteur \underline{n} admettant une loi multinomiale de paramètres $(n, \underline{\theta})$ où n est connu. La loi a priori non informative de Jeffreys pour le vecteur des paramètres inconnus $\underline{\theta} \in \mathcal{S}^p$ sera une Dirichlet de paramètres $a_j = \frac{1}{2}$.*

Nous effectuons ici la preuve pour $p = 4$. Rappelons que p vaut 4 lorsque nous utilisons la loi multinomiale pour décrire le comportement aléatoire des variables regroupées dans un tableau de contingence 2×2 .

PREUVE. Soit $\mathcal{I}(\underline{\theta})$ la matrice 3×3 représentant l'information de Fisher. Ses éléments $\mathcal{I}_{ij}(\underline{\theta})$ sont tels que

$$\begin{aligned} \mathcal{I}_{ij}(\underline{\theta}) &= -\mathbb{E}_{\underline{x}} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \left(\log [f(\underline{X} | \underline{\theta})] \right) \right] \\ &= -\mathbb{E}_{\underline{x}} \left[\frac{\partial}{\partial \theta_i} \left(\frac{n_j}{\theta_j} - \frac{n_p}{\theta_p} \right) \right] \\ &= \begin{cases} \mathbb{E}_{\underline{x}} \left[\frac{n_j}{\theta_j^2} + \frac{n_p}{\theta_p^2} \right] & \text{si } j = i, \\ \mathbb{E}_{\underline{x}} \left[\frac{n_p}{\theta_p^2} \right] & \text{si } j \neq i, \end{cases} \\ &= \begin{cases} n \left(\frac{1}{\theta_j} + \frac{1}{\theta_p} \right) & \text{si } j = i, \\ n \left(\frac{1}{\theta_p} \right) & \text{si } j \neq i. \end{cases} \end{aligned}$$

Nous pouvons alors montrer que

$$\det[\mathcal{I}(\underline{\theta})] = n^3 (\prod_{j=1}^4 \theta_j)^{-1}.$$

La loi non informative de Jeffreys sera telle que

$$\pi(\underline{\theta}) \propto \sqrt{\left(\prod_{j=1}^p \theta_j \right)^{-1}},$$

ce qui correspond à une loi de Dirichlet où les paramètres valent tous $\frac{1}{2}$. \square

La loi non informative de Jeffreys est alors telle que $a_+ = \frac{p}{2}$, une des valeurs de a_+ fréquemment suggérées dans la littérature.

2.2.2. Propriétés de la loi de Dirichlet

La loi de Dirichlet possède plusieurs propriétés très intéressantes, en voici d'ailleurs trois que nous utiliserons par la suite. Les preuves des deux premières propositions se retrouvent dans Wilks (1962, section 7.7) et celle de la dernière dans Aitchison et Shen (1980).

Propriété 2.2.1. *Soit $\theta_1, \dots, \theta_p$ des variables aléatoires indépendantes de lois respectives $\text{Gamma}(a_1, a_+), \dots, \text{Gamma}(a_p, a_+)$ et soit $C(\underline{\theta})$ un vecteur de dimension p où chaque composante correspond à $\theta_i / \sum_{j=1}^p \theta_j$, alors le vecteur $C(\underline{\theta})$ suit une loi de Dirichlet de paramètres a_j .*

Propriété 2.2.2. *Soit $\underline{\theta}$ un vecteur de dimension p suivant une loi de Dirichlet de paramètres a_j , alors le vecteur $(\sum_{j=1}^{p_1} \theta_i, \sum_{j=p_1+1}^{p_2} \theta_i, \dots, \sum_{j=p_{s-1}+1}^p \theta_i)$ de dimension $s < p$ suit une $\mathcal{D}(\sum_{j=1}^{p_1} a_i, \sum_{j=p_1+1}^{p_2} a_i, \dots, \sum_{j=p_{s-1}+1}^p a_i)$.*

Propriété 2.2.3. *Soit $\underline{\theta}$ un vecteur de dimension p suivant une loi de Dirichlet de paramètres a_j , alors pour tout c tel que $1 \leq c < p$, $C(\theta_1, \dots, \theta_c)$ est indépendant de $C(\theta_{c+1}, \dots, \theta_p)$.*

Nous remarquons également que lorsque $p = 2$, la loi $\mathcal{D}(\underline{a})$ correspond à une loi Beta(a_1, a_2). Ainsi, dans un tableau de contingence 2×2 où nous supposons que $(n_1 + n_2)$ et $(n_3 + n_4)$ sont connus, la modélisation usuelle *a priori* serait (voir l'équation (1.2.3)) :

$$\frac{n\theta_1}{n_1 + n_2} \sim \text{Beta}(\alpha_1, \beta_1),$$

$$\frac{n\theta_3}{n_3 + n_4} \sim \text{Beta}(\alpha_2, \beta_2).$$

Notons que Hashemi, Nandram et Goldberg (1997), suggèrent les paramètres

$$\alpha_i = \max\left(1, \frac{\mu_i\{\mu_i(1 - \mu_i) - v_i\}}{v_i}\right),$$

$$\beta_i = \max\left(1, \frac{(1 - \mu_i)\{\mu_i(1 - \mu_i) - v_i\}}{v_i}\right),$$

où (μ_1, v_1) et (μ_2, v_2) représentent respectivement notre idée *a priori* sur l'espérance et la variance de $\frac{n\theta_1}{n_1 + n_2}$ et $\frac{n\theta_3}{n_3 + n_4}$.

2.2.3. L'approche bayésienne empirique

Il existe une approche, ne faisant pas à proprement parler de la modélisation bayésienne, qui consiste à faire appel à des approximations fréquentistes afin de pallier les insuffisances de l'information *a priori*. Nous voyons ici deux cas où cette approche, nommée bayésienne empirique, est appliquée aux variables aléatoires catégorielles.

Lorsque $(\theta_1, \dots, \theta_4) \sim \mathcal{D}(a_1, \dots, a_4)$, Maritz (1989) suggère d'utiliser la méthode des moments pour donner au vecteur \underline{a} une valeur dépendant de nos observations. Comme $\mathbb{E}\left[\frac{n_j}{n}\right] = \theta_j$ et $\mathbb{E}\left[\frac{n_j(n_j - 1)}{n(n - 1)}\right] = \theta_j^2$, où θ_j est tel que $\mathbb{E}[\theta_j] = \frac{a_j}{a_+}$ et $\mathbb{E}[\theta_j^2] = \frac{a_j(a_j + 1)}{a_+(a_+ + 1)}$, cette méthode nous permet d'obtenir le vecteur \underline{a}^{BE} tel que

$$\frac{a_j^{BE}}{a_+^{BE}} = \frac{n_j}{n} \quad \text{avec} \quad a_+^{BE} = \frac{1 - \widehat{Q}}{\widehat{Q} - \widehat{B}},$$

où

$$\widehat{B} = \sum_{j=1}^4 \left(\frac{n_j}{n}\right)^2 \quad \text{et} \quad \widehat{Q} = \sum_{j=1}^4 \frac{n_j(n_j - 1)}{n(n - 1)}.$$

Fienberg et Holland (1970, 1973) tentèrent quant à eux de trouver une approche bayésienne empirique afin que la règle de Bayes pour θ_j , écrite en (2.2.4),

ait la plus petite erreur quadratique moyenne. Ils montrèrent que le paramètre a_+ devrait alors s'écrire sous la forme :

$$a_+(\underline{\eta}, \underline{\theta}) = \frac{1 - \sum_{j=1}^{IJ} \theta_j^2}{\sum_{j=1}^{IJ} (\eta_j - \theta_j)^2}.$$

Pour estimer de façon empirique les deux vecteurs inconnus, ils suggèrent de remplacer les θ_j par p_j et les η_j par p_j^* . Cette approche a donc la particularité de donner une idée *a priori* d'indépendance uniquement à partir des valeurs prises par nos observations (voir la sous-section 1.2.4).

2.2.4. Relation avec le test exact de Fisher

Nous terminons cette section en explicitant la relation qui existe entre la modélisation bayésienne et le test exact de Fisher. Ces résultats, tous issus de Altham (1969) montrent que si nous désirons tester la présence d'une association négative (*i.e.* $\delta < 1$) dans un tableau de contingence 2×2 , un choix particulier des paramètres de la loi de Dirichlet nous mène exactement aux mêmes conclusions que le test de Fisher. De plus, ces résultats montrent qu'un choix de lois *a priori* non informatives nous mène à des conclusions similaires aux conclusions fréquentistes.

Supposons que \underline{n} suit une loi multinomiale de paramètres n et $\underline{\theta}$, où $\underline{\theta} = (\theta_1, \dots, \theta_4)$ suit une distribution de Dirichlet de paramètres a_j . Altham (1969) montre qu'il est possible de calculer exactement la probabilité *a posteriori* que l'association entre les variables catégorielles soit négative,

$$P(\theta_1\theta_4 < \theta_2\theta_3 \mid \underline{n}) = \sum_{s=m_-}^{m_3-1} \left[\frac{\binom{m_1 + m_3 - 1}{s} \binom{m_2 + m_4 - 1}{m_3 + m_4 - 1 - s}}{\binom{m_+ - 2}{m_1 + m_2 - 1}} \right] \quad (2.2.5)$$

où $\underline{m} = (a_1 + n_1, \dots, a_4 + n_4)$ et $m_- = \max(0, m_3 - m_2)$. Nous remarquons que cette probabilité est identique à la valeur-p du test exact de Fisher si nous posons $a_1 = a_4 = 0$ et $a_2 = a_3 = 1$, c'est-à-dire une croyance *a priori* que l'association est négative.

Posons maintenant $a_j = t$ où $0 \leq t \leq 1$. Assigner une telle valeur à ces paramètres semble plausible lorsque nous n'avons pas d'idée *a priori* sur la distribution de $\underline{\theta}$, car nous retrouvons ainsi la loi de Jeffreys ou la distribution uniforme sur \mathcal{S}^4 , selon que t vaille $\frac{1}{2}$ ou 1. Dans ce contexte, Altham (1969) montre que

$$0 < \alpha - \beta < \frac{\binom{n_1 + n_2}{n_1} \binom{n_3 + n_4}{n_3}}{\binom{n}{n_1 + n_3}},$$

où $\alpha =$ valeur-p du test exact de Fisher (voir l'équation (1.5.1)) et $\beta = P(\theta_1\theta_4 < \theta_2\theta_3 \mid \underline{n})$. Ainsi, le test exact de Fisher est plus conservateur, mais la différence entre ces deux niveaux de signification est inférieure à la probabilité que nous ayons effectivement observé le vecteur (n_1, \dots, n_4) , une probabilité qui décroît lorsque la taille de l'échantillon augmente.

Il faut par contre noter que lorsque nous utilisons le test exact de Fisher, nous supposons que toutes les marges sont connues à l'avance et dans ce cas le résultat (2.2.5) n'est plus exact. Altham (1969) justifie tout de même son approche par le fait que le test exact de Fisher est souvent utilisé même lorsque les distributions marginales ne sont pas connues.

2.3. MODÉLISATION BAYÉSIENNE PARTICULIÈRE

Nous présentons ici des modèles bayésiens qui n'utilisent pas la loi de Dirichlet comme loi *a priori*. Nous discutons d'abord de l'utilisation de la loi logistique-normale comme loi *a priori*, puis un modèle où l'information *a priori* est disponible pour une mesure d'association des composantes de $\underline{\theta}$. Notons que nous nous restreignons ici aux modèles non hiérarchiques, cette approche étant présentée à la section suivante.

2.3.1. La loi logistique-normale

La loi logistique-normale est elle aussi une loi définie sur le simplexe \mathcal{S}^p . Elle est facile à utiliser à cause de son étroite relation avec une loi normale de dimension $(p-1)$. Le contenu de cette sous-section est tiré de Aitchison et Shen (1980).

Définition 2.3.1. Soit $\underline{\nu} \sim \mathcal{N}_{p-1}(\underline{\mu}, \Sigma)$, alors le vecteur $\underline{U} \in \mathcal{S}^p$ suit une logistique-normale de paramètres $\underline{\mu}$ et Σ , notée $\mathcal{L}_p(\underline{\mu}, \Sigma)$, si et seulement si il s'écrit sous la forme suivante :

$$\underline{U} = \left(\frac{e^{\nu_1}}{1 + \sum_{j=1}^{p-1} e^{\nu_j}}, \dots, \frac{e^{\nu_{p-1}}}{1 + \sum_{j=1}^{p-1} e^{\nu_j}}, 1 - \sum_{j=1}^{p-1} U_j \right).$$

Nous remarquons que cette loi est plus riche avec ses $\frac{(p-1)(p+2)}{2}$ paramètres que la Dirichlet avec p . Par contre, l'utilisation de la logistique-normale comme loi *a priori* nous amène à une loi *a posteriori* dont les moments ne sont pas calculables analytiquement. Cette loi *a posteriori* a la forme suivante :

$$\pi(\underline{u} | \underline{n}) \propto \det(2\pi\Sigma)^{-\frac{1}{2}} \left(\prod_{j=1}^{p-1} u_j \right)^{n_j-1} \exp \left[\frac{-1}{2} \left\{ \log \left(\frac{u^{(p)}}{u_p} \right) - \underline{\mu} \right\}^T \Sigma^{-1} \right]$$

$$\times \left\{ \log \left(\frac{\underline{u}^{(p)}}{u_p} \right) - \underline{\mu} \right\},$$

où $\underline{u}^{(p)}$ est le vecteur des $(p - 1)$ premières composantes de \underline{u} .

Maintenant, énonçons deux propriétés de cette loi et comparons les aux propriétés 2.2.1 et 2.2.3 de la loi de Dirichlet.

Propriété 2.3.1. *Considérons $C(\underline{\omega})$ tel que défini dans la propriété 2.2.1. Si $\underline{\nu} \sim \mathcal{N}_{p-1}(\underline{\mu}, \Sigma)$ alors $C(\exp(\underline{\nu})) \sim \mathcal{L}_p(A\underline{\mu}, A\Sigma A^T)$ où A est une matrice de dimension $(p - 1) \times p$ telle que les $(p - 1)$ premiers vecteurs-colonnes, tous de dimension $p - 1$, forment la matrice identité $I_{(p-1)}$ et où le dernier vecteur-colonne est tel que toutes ses composantes valent -1 .*

Propriété 2.3.2. *Soit \underline{u} un vecteur de dimension p admettant une loi logistique-normale comme distribution, alors pour tout $1 \leq c < p$,*

$$\text{Cov}(C(u_1, \dots, u_c), C(u_{c+1}, \dots, u_p)) \neq 0.$$

Rappelons que lorsque nous utilisons la loi de Dirichlet comme loi *a priori*, nous supposons que notre vecteur est issu de p variables aléatoires de lois gammas indépendantes de même paramètre d'échelle. De plus nous supposons que deux groupes de composition sont indépendants s'ils n'ont pas de paramètres en commun, hypothèses qui ne sont pas nécessairement vraies. La loi logistique-normale, quant à elle, ne nécessite pas la formulation de telles hypothèses. Par contre, cette loi ne possédant pas de propriété équivalente à la propriété 2.2.2 et n'ayant pas de loi *a posteriori* se réduisant à une forme simple, son utilisation peut entraîner de lourds problèmes de calculs.

2.3.2. Loi *a priori* sur une mesure d'association

Jusqu'à la fin de ce chapitre nous revenons avec une notation à deux indices, car cela simplifie beaucoup la notation. Dans un tableau de contingence 2×2 ,

notre modélisation bayésienne usuelle reflète une ignorance sur les trois valeurs $(\theta_{11}, \theta_{12}, \theta_{21})$ que nous modélisons à partir des quatre paramètres $(a_{11}, a_{12}, a_{21}, a_{22})$. L'utilisation de la Dirichlet comme loi *a priori* est souvent critiquée car elle ne contient pas assez de paramètres pour incorporer de la connaissance sur l'association de $\underline{\theta}$. Albert et Gupta (1983a) proposent une modélisation non hiérarchique permettant d'incorporer une idée *a priori* sur une reparamétrisation des valeurs inconnues, valeurs nous apportant une connaissance sur l'interaction des composantes de $\underline{\theta}$.

Albert et Gupta (1983a) montrent que nous pouvons reparamétriser la densité d'une loi multinomiale de paramètres n et $\underline{\theta}$ par

$$f(\underline{n} \mid \theta_{1+}, \theta_{+1}, \rho) = \binom{n}{n_{11}, \dots, n_{22}} \prod_{i=1}^2 (\theta_{i+}\theta_{+i} + \rho)^{n_{ii}} (\theta_{i+}(1 - \theta_{+i}) - \rho)^{n_{i\bar{i}(2,i)}},$$

où

$$\mathbb{I}(i, j) = \begin{cases} 1 & \text{si } i = j, \\ 2 & \text{si } i \neq j, \end{cases}$$

$$\rho = \theta_{11} - \theta_{1+}\theta_{+1}.$$

Le paramètre ρ est donc une mesure d'association prenant zéro comme valeur si et seulement si les variables sont indépendantes.

Albert et Gupta (1983a) suggèrent une loi *a priori* de la forme

$$\pi(\theta_{1+}, \theta_{+1}, \rho) = \pi_1(\theta_{1+})\pi_2(\theta_{+1})\pi_3(\rho \mid \theta_{1+}, \theta_{+1}),$$

où θ_{1+} et θ_{+1} suivent respectivement une loi *a priori* Beta de paramètres (a_1, a_2) et (b_1, b_2) . Pour ce qui est de $\rho \mid \theta_{1+}\theta_{+1}$, la loi *a priori* à donc la forme suivante

$$\pi_3(\rho \mid \theta_{1+}, \theta_{+1}) \propto \prod_{i=1}^2 (\theta_{i+}\theta_{+i} + \rho)^{v_{ii}} (\theta_{i+}(1 - \theta_{+i}) - \rho)^{v_{i\bar{i}(2,i)}}.$$

Nous avons donc huit paramètres inconnus plutôt que quatre avec la loi de Dirichlet. Pour incorporer à ce modèle une idée *a priori* d'indépendance, les auteurs suggèrent de prendre une valeur de (v_{11}, \dots, v_{22}) afin que le mode de $\pi_3(\rho \mid \mathbb{E}[\theta_{1+}], \mathbb{E}[\theta_{+1}])$ soit zéro.

La difficulté de cette approche survient lorsque nous devons estimer, *a posteriori*, θ_{1+} , θ_{+1} et ρ , la loi *a posteriori* de ces trois paramètres étant

$$\prod_{i=1}^2 (\theta_{i+} \theta_{+i} + \rho)^{n_{ii} + v_{ii}} (\theta_{i+} (1 - \theta_{+i}) - \rho)^{n_{i\bar{1}(2,i)} + v_{i\bar{1}(2,i)}} \theta_{i+}^{a_i - 1} \theta_{+i}^{b_i - 1}.$$

Comme Albert et Gupta (1983a) n'étaient pas en mesure de trouver les moments ou le mode des paramètres dans cette équation, ils proposèrent la méthode d'estimation suivante. Nous estimons d'abord θ_{1+} par $\mathbb{E}[\theta_{1+} \mid n_{1+}]$, c'est-à-dire $(a_1 + n_{1+}) / (a_1 + a_2 + n)$. Après avoir estimé θ_{+1} de la même façon, nous remplaçons ces deux paramètres dans l'équation *a posteriori*. Il est alors possible de trouver le mode de ρ en solutionnant

$$\sum_{i=1}^2 \left[\frac{v_{ii} + n_{ii}}{\theta_{i+} \theta_{+i} + \rho} - \frac{n_{i\bar{1}(2,i)} + v_{i\bar{1}(2,i)}}{\theta_{i+} (1 - \theta_{+i}) - \rho} \right] = 0.$$

2.4. MODÉLISATION BAYÉSIENNE HIÉRARCHIQUE

Lorsque nous sommes incapable d'attribuer adéquatement des valeurs spécifiques aux paramètres de notre loi *a priori*, nous pouvons utiliser une approche bayésienne empirique en leur attribuant des valeurs en fonction de \underline{n} (voir la

sous-section 2.2.3). Une autre alternative consiste à utiliser un modèle bayésien hiérarchique faisant intervenir plusieurs niveaux de distributions conditionnelles.

Définition 2.4.1. *Nous appelons modèle bayésien hiérarchique un modèle statistique bayésien où la loi a priori $\pi(\underline{\theta})$ est décomposée en distributions conditionnelles et en une distribution marginale, c'est-à-dire :*

$$\pi(\underline{\theta}) = \int_{\Theta_1} \dots \int_{\Theta_n} \pi_1(\underline{\theta} | \underline{\theta}_1) \dots \pi_n(\underline{\theta}_{n-1} | \underline{\theta}_n) \pi_{n+1}(\underline{\theta}_n) d\underline{\theta}_1 \dots d\underline{\theta}_n,$$

où les composantes de $\underline{\theta}_{\sim_i}$ sont nommées hyperparamètres.

Nous voyons donc différents modèles hiérarchiques appliqués à l'étude des tableaux de contingence. Nous apportons une attention particulière aux modèles utilisant cette hiérarchie afin d'amener une idée *a priori* d'indépendance entre les variables aléatoires.

2.4.1. Hiérarchisation des modèles usuels

Afin d'incorporer une idée *a priori* d'égalité des paramètres θ_i , Good (1976) suggère une loi $\mathcal{D}(\underline{a})$ symétrique, *i.e.* pour laquelle $a_1 = \dots a_p = \frac{a_+}{p}$. Cette loi n'a donc qu'un seul paramètre inconnu et tous les θ_i sont interchangeables. Le fait d'avoir un seul paramètre étant assez contraignant, Good (1976) propose d'incorporer une distribution marginale sur celui-ci, la loi retenue est la loi log-Cauchy, une loi dont la fonction de densité est décroissante en a_+ . La loi *a priori* s'écrit donc

$$\pi(\underline{\theta}) = \int_0^\infty \mathcal{D}^*(\underline{a}) \pi_2(a_+) da_+,$$

où $\mathcal{D}^*(\underline{a})$ représente la fonction de densité de cette loi symétrique et $\pi_2(a_+) = (a_+ \pi[1 + (\log(a_+))^2])^{-1}$.

Good et Crook (1987) suggèrent d'utiliser plutôt une distribution marginale ayant trois paramètres, la loi log-Student généralisée

$$\pi_2(a_+ | v, \lambda, u) = \frac{\lambda^v}{a_+ \sqrt{v} B(1/2, v/2)} \{ \lambda^2 + v^{-1} (\log(a_+/u))^2 \}^{-(v+1)/2},$$

où $v, \lambda, u \in \mathbb{R}^+$. Nous retrouvons d'ailleurs la distribution log-Cauchy lorsque les trois paramètres valent 1.

Si la première approche de cette sous-section considère a_+ comme étant un hyperparamètre, la seconde (voir Albert et Gupta, 1982) traite plutôt \underline{a} comme un vecteur de p hyperparamètres. Soit $\underline{\theta}$ de loi $\mathcal{D}(a_+, \underline{\eta})$ (voir l'équation 2.2.1), incorporons une idée *a priori* d'indépendance en posant $\eta_{ij} = \eta_i \eta_j$. Nous pouvons maintenant hiérarchiser ce modèle en attribuant une distribution aux vecteurs $\underline{\eta}_1$ et $\underline{\eta}_2$:

$$\begin{aligned} \underline{\eta}_1 &\sim \mathcal{D}(L_1, \lambda_1), \\ \underline{\eta}_2 &\sim \mathcal{D}(L_2, \lambda_2). \end{aligned}$$

De cette façon, nous avons une distribution marginale sur $\underline{\theta}$ avec $(I + J + 3)$ paramètres. À partir des équations (2.2.2) et (2.2.3), nous voyons que la valeur a_+ représente en quelque sorte notre conviction en l'indépendance, tandis que L_1 et L_2 représentent notre conviction que les valeurs de λ_1 et λ_2 estiment bien les vecteurs $\underline{\eta}_1$ et $\underline{\eta}_2$.

Sous cette modélisation, nous obtenons que

$$\mathbb{E}[\theta_{ij} | \underline{n}] = \frac{k}{a_+ + n} \mathbb{E}[\eta_{1i} | \underline{n}] \mathbb{E}[\eta_{2j} | \underline{n}] + \frac{n}{a_+ + n} \left(\frac{n_{ij}}{n} \right),$$

où $\pi_2(\eta_1, \eta_2 \mid \underline{n})$ est proportionnel à

$$\prod_{i=1}^I \eta_{1i}^{n_{i+} + L_1 \lambda_{1i}} - J \prod_{j=1}^J \eta_{2j}^{n_{+j} + L_2 \lambda_{2j}} - I.$$

Il est intéressant de regarder l'influence du paramètre a_+ sur la règle de Bayes sus mentionnée. Si a_+ vaut zéro, notre règle de Bayes vaut n_{ij}/n , l'estimateur du maximum de vraisemblance de θ_{ij} , tandis que si a_+ tend vers l'infini, nous obtenons $\mathbb{E}[\theta_{i+} \mid \underline{n}] \mathbb{E}[\theta_{+j} \mid \underline{n}]$.

2.4.2. Loi *a priori* hiérarchique sur une mesure d'association

Nous avons vu, à la fin de la sous-section 2.2.2, comment utiliser les lois bêta afin de modéliser des variables aléatoires catégorielles représentées dans un tableau de contingence 2×2 , où n_{1+} et n_{+1} sont connus. Nous présentons ici deux extensions hiérarchiques à cette modélisation. Elles furent introduites respectivement par Albert et Gutpa (1983a) ainsi que Albert (1987). Nous remarquons que ces deux extensions permettent d'incorporer une idée *a priori* d'indépendance.

Deux variables aléatoires catégorielles sont indépendantes si et seulement si $\theta_{1|i} = \theta_{+1} \forall i$ (voir l'équation (1.1.4)), si nous supposons que $\theta_{1|1}$ et $\theta_{1|2}$ sont deux variables aléatoires indépendantes et identiquement distribuées, selon Albert et Gutpa (1983a) cela équivaut à incorporer une idée *a priori* d'indépendance sur ces deux variables aléatoires catégorielles. Alors posons que $\theta_{1|1}$ et $\theta_{1|2}$ sont deux variables aléatoires indépendantes de loi conjuguée $\text{Beta}(B, K)$ où $B, K \in \mathbb{N}$, supposons que K est connu et attribuons à B une loi hiérarchique, uniforme et discrète, prenant des valeurs allant de 1 à $(K - 1)$. Dans ce cas, nous pouvons montrer que $\theta_{1|1}$ et $\theta_{1|2}$ ont une loi *a priori* marginale uniforme sur $(0, 1)$ et que leur corrélation est de $(1 - \frac{2}{K})$. Donc si $K = 2$, nous supposons que leur corrélation est nulle, tandis que si $K = \infty$, nous supposons que la corrélation

vaut un et donc que nos deux variables aléatoires catégorielles sont identiques. Sous cette modélisation, il est possible d'obtenir analytiquement l'espérance *a posteriori* des deux paramètres :

$$\mathbb{E}[\theta_{1|i}] = \left(\frac{\sum_{B=1}^{K-1} Bd(B)}{\frac{K-1}{K} \sum_{B=1} d(B)} \right) + \left(1 - \frac{K}{n_{i+} + K} \right) \left(\frac{n_{i1}}{n_{i+}} - \frac{\sum_{B=1}^{K-1} Bd(B)}{K \sum_{B=1} d(B)} \right),$$

$$\text{où } d(B) = \prod_{i=1}^2 \binom{n_{i1} + B - 1}{n_{i1}} \binom{n_{i+} - n_{i1} + K - B - 1}{n_{i+} - n_{i1}}.$$

Albert et Gupta (1983a) montrent que si $K = 2$, nous obtenons $\frac{n_{1i} + 1}{n_{i+} + 2}$, un léger ajustement de l'estimateur du MV de $\theta_{1|i}$ lorsque la taille échantillonnale est relativement grande. Par contre, si K tend vers l'infini nous obtenons $\frac{n_{+i} + 1}{n + 2}$, un léger ajustement de l'estimateur du MV de $\theta_{1|i}$ lorsque les variables aléatoires catégorielles sont indépendantes.

Dans la seconde modélisation, nous allons hiérarchiser une reparamétrisation des valeurs inconnues. Tout d'abord, nous pouvons montrer que $f(\underline{n} | \underline{\theta})$ peut être écrit sous la forme $f(\underline{n} | \alpha, \beta)$, où

$$\alpha = \log \left(\frac{\theta_{1|1}\theta_{1|2}}{\theta_{2|1}\theta_{2|2}} \right),$$

$$\beta = \frac{1}{2} \log \left(\frac{\theta_{1|1}\theta_{2|2}}{\theta_{2|1}\theta_{1|2}} \right) = \frac{\phi}{2}.$$

Nous hiérarchisons ce modèle en supposant que ces deux paramètres sont indépendants, nous introduisons une loi non informative sur α et nous supposons que β suit une loi normale de moyenne 0 et de variance σ^2 . L'indépendance des variables aléatoires catégorielles est donc suggérée en attribuant 0 comme moyenne à β . Le seul hyperparamètre auquel nous devons attribuer une valeur est donc

σ^2 . Albert (1987) suggère d'utiliser une loi *a priori* pour σ^2 ,

$$v\sigma^{-2} \sim \mathcal{X}_v^2,$$

où v est une constante entière. Nous pouvons alors trouver de façon analytique la loi *a posteriori* $\pi(\alpha, \beta \mid \underline{n})$, mais nous devons approximer numériquement la loi marginale $\pi(\beta \mid \underline{n})$, β étant le seul paramètre intéressant à estimer *a posteriori* lorsque nous nous intéressons surtout à l'indépendance des variables.

2.4.3. Modélisation hiérarchique particulière

À la sous-section 2.3.2, nous avons vu deux modèles où nous reparamétrisons les trois paramètres inconnus $(\theta_{11}, \theta_{12}, \theta_{21})$ dans un tableau de contingence 2×2 . De façon semblable, nous voyons ici une approche hiérarchique qui reparamétrise les 4 paramètres inconnus $(a_+, \eta_{11}, \eta_{12}, \eta_{21})$ de notre loi *a priori*.

Soit $(\theta \mid a_+, \underline{\eta}) \sim \mathcal{D}(a_+, \underline{\eta})$, Albert et Gupta (1983b, section 2.1) montrent que nous pouvons écrire tous les $\underline{\eta}$ sous la forme $\eta_{ij} = f_{ij}(\eta_{1+}, \eta_{+1}, \delta_0)$ où $\delta_0 = \eta_{11}\eta_{22}(\eta_{12}\eta_{21})^{-1}$. Par le corollaire 2.2.1 nous avons que

$$\mathbb{E}[\delta] = \eta_{11}\eta_{22} \left(\eta_{12} - \frac{1}{a_+} \right)^{-1} \left(\eta_{21} - \frac{1}{a_+} \right)^{-1} \cong \delta_0.$$

Il est alors clair que nous attribuons à δ_0 la valeur *a priori* que nous désirons attribuer à δ . Quant au vecteur (η_{1+}, η_{+1}) , nous lui attribuons une loi uniforme sur le carré unitaire. La loi *a priori* marginale sur $\underline{\theta}$ dépend donc uniquement des deux paramètres δ_0 et a_+ ,

$$\pi(\underline{\theta} \mid \delta_0, a_+) = \int_0^1 \int_0^1 \Gamma(a_+) \prod_{i=1}^2 \prod_{j=1}^2 \frac{\theta_{ij}^{a_+ f_{ij}(\eta_{1+}, \eta_{+1}, \delta_0) - 1}}{\Gamma(a_+ f_{ij}(\eta_{1+}, \eta_{+1}, \delta_0))} d\eta_{1+} d\eta_{+1}.$$

Nous obtenons alors

$$\mathbb{E}[\theta_{ij} \mid \underline{n}] = (n + a_+)^{-1} \left(n_{ij} + a_+ \mathbb{E}[f_{ij}(\eta_{1+}, \eta_{+1}, \delta_0) \mid \underline{n}] \right),$$

où $\mathbb{E}[f_{ij}(\eta_{1+}, \eta_{+1}, \delta_0) \mid \underline{n}]$ doit être approximée numériquement.

Nous terminons cette section en énonçant brièvement une idée, proposée par Walley (1997), afin de ne pas recourir aux extensions bayésiennes empiriques et hiérarchiques lorsque notre connaissance *a priori* est insuffisante. Soit $\underline{\theta}$ de loi $\mathcal{D}(a_+, \underline{\eta})$ où a_+ est connu. Si nous n'avons pas d'idée *a priori* de la valeur de $\underline{\eta}$, Walley(1997) suggère alors de toujours calculer les probabilités *a posteriori* minimales et maximales des événements. Prenons, par exemple, l'événement A_{ij} stipulant que la prochaine réalisation est élément de la cellule (i, j) , les probabilités minimales et maximales valent respectivement :

$$P_{max}(A_{ij}) = \frac{n_{ij} + a_+}{n + a_+} \quad (\text{atteint lorsque } \eta_{ij} \rightarrow 1),$$

$$P_{min}(A_{ij}) = \frac{n_{ij}}{n + a_+} \quad (\text{atteint lorsque } \eta_{ij} \rightarrow 0).$$

Ainsi, toute analyse devrait tenir compte de ces deux probabilités limites. Si par exemple nous cherchons une région α -crédible pour un paramètre θ_{ij} , Walley (1997) suggère une région (a, b) telle que :

$$P_{min}(\theta_{ij} \in (a, b) \mid \underline{n}) \geq 0.95.$$

Nous avons donc présenté dans ce chapitre différents modèles bayésiens utilisés avec des variables aléatoires catégorielles. Le prochain chapitre sera consacré exclusivement à l'étude de la distribution *a posteriori* du logarithme du rapport de cotes sous la modélisation bayésienne usuelle présentée à la section 2.2.

Chapitre 3

DISTRIBUTION *A POSTERIORI* DU LOGARITHME DU RAPPORT DE COTES

Nous avons vu avec le corollaire 1.1.2 l'étroite relation entre ϕ , le logarithme du rapport de cotes, et l'indépendance de variables aléatoires catégorielles regroupées dans un tableau de contingence 2×2 . À partir de cette relation, nous pouvons effectuer directement des inférences bayésiennes sur l'indépendance de ces variables aléatoires en étudiant la distribution *a posteriori* de ϕ . Alors que les différentes méthodes vues dans le chapitre précédent portaient principalement sur le comportement *a posteriori* des paramètres $\underline{\theta}$, ce chapitre portera uniquement sur l'étude de la distribution *a posteriori* de ϕ , lorsque $\underline{\theta}$ admet une loi conjuguée de Dirichlet comme loi *a priori*.

Nous débutons en montrant la normalité asymptotique de ϕ , distribution asymptotique couramment énoncée dans la littérature. Nous trouvons par la suite les moments exacts de ϕ ainsi qu'une meilleure approximation que l'approximation normale. Finalement, nous comparons les principales méthodes présentées dans ce mémoire à l'aide de différents tableaux de contingence 2×2 . Afin d'alléger la notation, nous supposons tout au cours de ce chapitre que $\underline{\theta}$ est distribué selon une loi de Dirichlet de paramètres a_j . Ainsi, lorsque nous désirerons utiliser *a posteriori* les divers résultats et formules de ce chapitre, il nous suffira de remplacer les paramètres a_j par ceux de la loi *a posteriori*, c'est-à-dire $a_j + n_j$.

3.1. NORMALITÉ ASYMPTOTIQUE DU LOGARITHME DU RAPPORT DE COTES

Nous voyons dans cette section que lorsque tous les paramètres a_j sont grands, la loi normale approxime très bien la distribution de ϕ . Cette approximation est fréquemment mentionnée dans la littérature, voir entre autres Lindley (1964), Bloch et Watson (1967), Altham (1969), Aitchison et Shen (1980) ou bien O'Hagan (1994, chapitre 10). Par contre, le choix des paramètres de cette approximation ne fait pas l'unanimité.

Afin de trouver une distribution qui approxime bien ϕ lorsque nos paramètres sont tous grands, nous étudierons la fonction génératrice des moments de ϕ . Cette approche est issue de O'Hagan(1994, chapitre 10). Si $\underline{c} = (1, -1, -1, 1)$, alors

$$\begin{aligned}
 \mathcal{M}_\phi(t) &= \mathbb{E}[e^{t\phi}] \\
 &= \mathbb{E}[e^{\sum_{j=1}^4 tc_j \log(\theta_j)}] \\
 &= \mathbb{E}\left[\prod_{j=1}^4 \theta_j^{tc_j}\right] \\
 &= \prod_{j=1}^4 \Gamma(a_j + tc_j) / \Gamma(a_j). \tag{3.1.1}
 \end{aligned}$$

En utilisant la série de Stirling pour le logarithme de la fonction gamma, nous sommes en mesure de montrer que la fonction génératrice des moments de ϕ est telle que la différence entre son logarithme et celui de la fonction génératrice des moments d'une loi normale est d'ordre a_j^{-2} .

Comme

$$\log \Gamma(x) = \left(x - \frac{1}{2}\right) \log(x) - x + \frac{\log(2\pi)}{2} + \frac{1}{12x} - \mathcal{O}(x^{-3}), \tag{3.1.2}$$

nous avons

$$\begin{aligned}
\log \mathcal{M}_\phi(t) &= \sum_{j=1}^4 \left[\left(a_j + tc_j - \frac{1}{2} \right) \log(a_j + tc_j) - \left(a_j - \frac{1}{2} \right) \log(a_j) \right. \\
&\quad \left. - tc_j - \frac{tc_j}{12a_j(a_j + tc_j)} + \mathcal{O}(a_j^{-3}) \right] \\
&= \sum_{j=1}^4 \left[tc_j \log(a_j) + \left(a_j + tc_j - \frac{1}{2} \right) \log\left(1 + \frac{tc_j}{a_j}\right) - \mathcal{O}(a_j^{-2}) \right] \\
&= t \left[\sum_{j=1}^4 c_j \left(\log(a_j) - \frac{1}{2a_j} \right) \right] + \frac{t^2}{2} \left[\sum_{j=1}^4 \frac{1}{a_j} \right] - \mathcal{O}(a_j^{-2}).
\end{aligned}$$

Ainsi, lorsque tous les a_j^2 sont grands, ϕ est bien approximé par une $\mathcal{N}(\mu, \sigma^2)$ avec

$$\mu = \sum_{j=1}^4 c_j \left(\log(a_j) - \frac{1}{2a_j} \right), \quad (3.1.3)$$

$$\sigma^2 = \sum_{j=1}^4 \frac{1}{a_j}. \quad (3.1.4)$$

À partir de (3.1.1) et de (3.1.2), nous pouvons écrire l'espérance et la variance de ϕ sous la forme d'une série de puissance,

$$\begin{aligned}
\mathbb{E}[\phi] &= \mathcal{M}'_\phi(0) = \sum_{j=1}^4 c_j \left(\log(a_j) - \frac{1}{2a_j} - \frac{1}{12a_j^2} + \mathcal{O}(a_j^{-4}) \right), \\
\mathbb{V}[\phi] &= \mathcal{M}''_\phi(0) - (\mathcal{M}'_\phi(0))^2 = \sum_{j=1}^4 \left(\frac{1}{a_j} + \frac{1}{2a_j^2} + \frac{1}{6a_j^3} - \mathcal{O}(a_j^{-5}) \right).
\end{aligned}$$

Remarquons que la différence entre les moments exacts de ϕ et les moments approximatifs (3.1.3) et (3.1.4) est d'ordre a_j^{-2} dans les deux cas.

D'autres approximations pour les moments de ϕ furent suggérées par Lindley(1964) ainsi que Bloch et Watson(1967). Leurs deux approches étant similaires, nous pouvons écrire les approximations qu'ils proposent sous la même forme,

$$\mu = \sum_{j=1}^4 c_j \log(a_j - k_j) \quad (3.1.5)$$

$$\sigma^2 = \sum_{j=1}^4 \frac{1}{a_j - k_j}, \quad (3.1.6)$$

où $k_j = \frac{1}{2}$ pour les approximations de Lindley (1964) et $k_j = \frac{1}{2} - \frac{1}{24a_j}$ pour celles de Bloch et Watson (1967). En utilisant l'expansion en série de Taylor de $\log(a_j + k_j)$, nous pouvons montrer que la précision des approximations de Lindley (1964) est du même ordre que celles de O'Hagan (1994, chapitre 10). En ce qui a trait aux approximations de Bloch et Watson (1967), elles s'écrivent sous une forme moins simple que pour les deux approximations précédentes, mais sont plus justes puisque la différence entre (3.1.5) et (3.1.6) et les moments exacts est d'ordre a_j^{-3} lorsque k_j est remplacé par $\frac{1}{2} - \frac{1}{24a_j}$. Dans la prochaine section, nous voyons une façon de calculer exactement les moments de ϕ lorsque nous imposons certaines conditions sur la valeur des paramètres a_j .

Si nous désirons uniquement calculer la probabilité d'une association négative (c'est-à-dire $\phi < 0$), Altham (1969) suggère plutôt d'utiliser la normalité asymptotique basée cette fois sur les moments d'une distribution hypergéométrique, étant donnée son étroite relation avec ϕ (voir la sous-section 2.2.4). En utilisant la correction de continuité, nous approximos la probabilité d'une association négative par :

$$1 - \Phi \left(\frac{a_1 - \frac{1}{2} - \mu}{\sigma} \right),$$

où

$$\begin{aligned}\mu &= \frac{(a_1 + a_2 - 1)(a_1 + a_3 - 1)}{a_+} \\ \sigma^2 &= \frac{(a_1 + a_2 - 1)(a_3 + a_4 - 1)(a_1 + a_3 - 1)(a_2 + a_4 - 1)}{(a_+ - 2)(a_+ - 1)^2}.\end{aligned}$$

De nombreuses simulations (voir Altham, 1969) suggèrent que cette approximation est plus précise qu'une approximation normale avec les paramètres de Bloch et Watson (1967), lorsque la probabilité d'association négative est faible (inférieure à $\frac{1}{10}$).

3.2. LES MOMENTS EXACTS

La littérature traitant de l'approximation des moments de ϕ est beaucoup plus complète que celle traitant de la détermination de ses moments exacts. Aitchison et Shen(1980) ainsi que Maritz(1989), constituent essentiellement la littérature à ce sujet.

Afin de déterminer les moments exacts de ϕ , nous utilisons sa fonction caractéristique, qui est définie pour tout t réel (voir Bloch et Watson, 1967),

$$\varphi_\phi(t) = \mathbb{E}[e^{it\phi}] = \prod_{j=1}^4 \frac{\Gamma(a_j + itc_j)}{\Gamma(a_j)},$$

où $i = \sqrt{-1}$.

Les fonctions caractéristiques, l'alternative complexe des fonctions génératrices

des moments, nous permettent elles aussi de trouver tous les moments entiers d'une variable aléatoire (voir Billingsley, 1979 section 26).

Proposition 3.2.1. *Soit $\varphi_X(t) = \mathbb{E}[e^{itX}]$ où X est une variable aléatoire et $t \in \mathbb{R}$. Alors,*

$$\mathbb{E}[X^n] = (-i)^n \left(\frac{\partial^n}{\partial t^n} [\varphi_X(t)] \Big|_{t=0} \right).$$

Cette proposition nous amène au théorème suivant qui énonce tous les moments exacts de ϕ .

Théorème 3.2.1. *Soit ϕ le logarithme du rapport de cotes issu d'un tableau de contingence 2×2 . Si le vecteur des paramètres inconnus $\underline{\theta}$ admet comme distribution une loi de Dirichlet de paramètres a_j , alors*

$$\mathbb{E}[\phi^n] = \sum_{k=1}^n \binom{n-1}{k-1} \left\{ \sum_{j=1}^4 (c_j)^k \psi^{(k-1)}(a_j) \right\} \mathbb{E}[\phi^{n-k}],$$

où $\underline{c} = (1, -1, -1, 1)$, $\psi^{(k)}(x) = \frac{\partial^k}{\partial x^k} [\psi(x)]$ et $\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$.

PREUVE.

Posons $f(t) = \sum_{j=1}^4 \log(\Gamma(a_j + itc_j))$. Alors, $f^{(n)}(t) = \sum_{j=1}^4 (ic_j)^n \psi^{(n-1)}(a_j + itc_j)$.

Si nous montrons par induction que

$$\varphi_{\phi}^{(n)}(t) = \sum_{k=1}^n \binom{n-1}{k-1} f^{(k)}(t) \varphi_{\phi}^{(n-k)}(t), \quad (3.2.1)$$

le résultat s'en suit.

Vérifions que l'équation (3.2.1) est vraie lorsque n vaut 1.

$$\varphi'_{\phi}(t) = \sum_{j=1}^4 \left(\prod_{k \neq j} \frac{\Gamma(a_k + itc_k)}{\Gamma(a_k)} \right) \left(\frac{\Gamma(a_j + itc_j)}{\Gamma(a_j)} \right) ic_j \psi(a_j + itc_j)$$

$$\begin{aligned}
&= \sum_{j=1}^4 ic_j \psi(a_j + itc_j) \left(\prod_{k=1}^4 \frac{\Gamma(a_k + itc_k)}{\Gamma(a_k)} \right) \\
&= \sum_{j=1}^4 ic_j \psi(a_j + itc_j) \varphi_\phi(t).
\end{aligned}$$

Supposons maintenant que cette équation est vraie pour la $(n-1)$ ^{ième} dérivée de $\varphi_\phi(t)$. Alors, montrons que cela implique que cette formule est nécessairement vraie pour la n ^{ième} dérivée de $\varphi_\phi(t)$.

$$\begin{aligned}
\varphi_\phi^{(n)}(t) &= \frac{\partial(\varphi_\phi^{(n-1)}(t))}{\partial t} \\
&= \sum_{k=1}^{n-1} \binom{n-2}{k-1} f^{(k)}(t) \varphi_\phi^{(n-k)}(t) + \sum_{k=2}^n \binom{n-2}{k-2} f^{(k)}(t) \varphi_\phi^{(n-k)}(t) \\
&= f'(t) \varphi_\phi^{(n-1)}(t) + \sum_{k=2}^{n-1} \left\{ \binom{n-2}{k-1} + \binom{n-2}{k-2} \right\} f^{(k)}(t) \varphi_\phi^{(n-k)}(t) \\
&\quad + f^{(n)}(t) \varphi_\phi(t) \\
&= \sum_{k=1}^n \binom{n-1}{k-1} f^{(k)}(t) \varphi_\phi^{(n-k)}(t).
\end{aligned}$$

□

Corollaire 3.2.1. *Soit ϕ le logarithme du rapport de cotes issu d'un tableau de contingence 2×2 . Si le vecteur des paramètres inconnus $\underline{\theta}$ admet comme distribution une loi de Dirichlet de paramètres a_j , alors*

$$\begin{aligned}
\mathbb{E}[\phi] &= \sum_{j=1}^4 c_j \psi(a_j); \\
\mathbb{V}[\phi] &= \sum_{j=1}^4 \psi'(a_j);
\end{aligned}$$

$$\begin{aligned} \mathbb{E}[\phi^3] &= \left(\sum_{j=1}^4 c_j \psi(a_j) \right)^3 + 3 \left(\sum_{j=1}^4 c_j \psi(a_j) \right) \left(\sum_{j=1}^4 \psi'(a_j) \right) \\ &\quad + \sum_{j=1}^4 c_j \psi^{(2)}(a_j). \end{aligned}$$

Il est plus simple d'écrire les moments de ϕ de façon récursive plutôt que de n'utiliser que les dérivées d'ordres multiples de la fonction $f(t)$, cela nous permettant de les représenter sous forme algébrique. Si nous voulions écrire $\mathbb{E}[\phi^{10}]$ de façon non récursive, nous obtiendrions une équation avec 41 termes.

Nous remarquons que les moments approximatifs de O'Hagan (1994, chapitre 10) correspondent aux premiers termes de l'expansion asymptotique de $\sum_{j=1}^4 c_j \psi(a_j)$ et $\sum_{j=1}^4 \psi'(a_j)$ (voir Abramovitz et Stegun, 1965 section 6.4). C'est là que réside la difficulté entourant l'utilisation des moments exacts ; nous devons utiliser les expansions asymptotiques de fonctions spéciales afin de les évaluer et ceci peut nécessiter le calcul de plusieurs termes lorsque les a_j sont petits.

Maintenant, nous voyons une nouvelle façon d'évaluer les moments de ϕ , sans avoir à recourir aux expansions asymptotiques de diverses fonctions spéciales. Afin d'y parvenir, nous devons toutefois nous restreindre aux cas où les a_j sont des entiers positifs. Notons que dans les différents exemples contenus dans la littérature, il est très fréquent que les a_j soient des entiers positifs. Cette restriction n'est donc pas trop contraignante et simplifie beaucoup les calculs des différents moments, spécialement lorsque les a_j sont petits. Cependant, ceci ne peut pas être utilisé si la densité *a priori* de $\tilde{\theta}$ correspond à la loi non informative de Jeffreys (voir la proposition 2.2.2).

Notre démarche est basée sur la formule suivante (voir Gradshteyn et Ryzhik, 1980, formule 8.362).

Lemme 3.2.1. Soit $\psi(x) = \frac{\partial}{\partial x} \log \Gamma(x)$. Nous avons alors pour tout $x \in \mathbb{R}^+$ et pour tout $n \in \mathbb{N}^*$,

$$\begin{aligned}\psi(x) &= \psi(1) - \sum_{k=0}^{\infty} \left(\frac{1}{x+k} - \frac{1}{k+1} \right), \\ \psi^{(n)}(x) &= (-1)^{n+1} n! \sum_{k=0}^{\infty} \frac{1}{(x+k)^{n+1}}.\end{aligned}$$

Voyons maintenant que les moments de ϕ , peuvent également être écrits en utilisant les fonctions zeta de Rieman.

Définition 3.2.1. Nous appelons fonctions zeta de Rieman, les fonctions $\zeta(a, k)$ et $\zeta(k)$ telles que :

$$\begin{aligned}\zeta(k) &= \sum_{n=1}^{\infty} \frac{1}{n^k}, \\ \zeta(k, a) &= \sum_{n=0}^{\infty} \frac{1}{(a+n)^k} = \zeta(k) - \sum_{n=1}^{a-1} \frac{1}{n^k},\end{aligned}$$

nous remarquons que $\zeta(k) = \zeta(k, 0)$.

Corollaire 3.2.2. Soit ϕ le logarithme du rapport de cotes d'un tableau de contingence 2×2 . Si le vecteur des paramètres inconnus $\underline{\theta}$ admet comme distribution une loi de Dirichlet de paramètres a_j , alors

$$\mathbb{E}[\phi^n] = \sum_{k=1}^n \binom{n-1}{k-1} b_k(\underline{a}) \mathbb{E}[\phi^{n-k}],$$

où

$$b_k(\underline{a}) = (k-1)! \sum_{j=1}^4 (-c_j)^k \zeta(k, a_j). \quad (3.2.2)$$

Corollaire 3.2.3. *Soit ϕ le logarithme du rapport de cotes. Nous avons alors que*

$$\mathbb{E}[\phi] = b_1(\underline{a});$$

$$\mathbb{V}[\phi] = b_2(\underline{a});$$

$$\mathbb{E}[\phi^3] = (b_1(\underline{a}))^3 + 3b_1(\underline{a})b_2(\underline{a}) + b_3(\underline{a}).$$

Grâce au théorème suivant, nous innovons dans l'étude des moments de ϕ . Nous pouvons maintenant évaluer exactement les moments de tout ordre de ϕ et cela sans avoir à recourir aux expansions asymptotiques de $\psi^{(n)}(a_j)$. Ces expansions peuvent parfois donner des résultats inexacts lorsque les a_j sont petits et que nous ne prenons pas beaucoup de termes. Rappelons par contre que nous devons nous restreindre au cas où les a_j sont des entiers positifs.

Théorème 3.2.2. *Soit $b_k(\underline{a})$ défini dans l'équation (3.2.2). Si $\underline{a} \in (\mathbb{N}^*)^4$, alors, pour tout entier k , $b_k(\underline{a})$ peut s'écrire sous forme d'une puissance de π et d'une somme finie de nombres rationnels.*

PREUVE. Posons

$$\gamma(a, b) = \begin{cases} \sum_{n=a}^{b-1} \frac{1}{n^k} & \text{si } a < b, \\ -\sum_{n=b}^{a-1} \frac{1}{n^k} & \text{si } b < a, \\ 0 & \text{si } a = b. \end{cases}$$

Nous pouvons alors montrer que

$$\begin{aligned}
b_k(\underline{a}) &= (k-1)! \sum_{j=1}^4 (-c_j)^k \zeta(k, a_j) \\
&= (k-1)! \sum_{j=1}^4 (-c_j)^k [\zeta(k) - \gamma(1, a_j)] \\
&= (k-1)! \left[\zeta(k) \sum_{j=1}^4 (-c_j)^k - \sum_{j=1}^4 (-c_j)^k \gamma(1, a_j) \right] \\
&= \begin{cases} (k-1)! [\gamma(a_2, a_1) + \gamma(a_3, a_4)] & \text{si } k \text{ est impair,} \\ (k-1)! \left[\frac{2(2\pi)^k (-1)^{\frac{k+2}{2}} B_k}{k!} - \sum_{j=1}^4 \sum_{n=1}^{a_j-1} \frac{1}{n^k} \right] & \text{si } k \text{ est pair,} \end{cases}
\end{aligned}$$

à partir de l'identité de Parseval où B_k représente le $k^{\text{ième}}$ nombre de Bernoulli (voir Edwards, 1974 chapitre 1). \square

Nous remarquons que lorsque les a_j sont petits, les valeurs de $b_k(\underline{a})$ sont alors faciles à calculer et nous avons très peu de fractions à évaluer. Nous retrouvons au début de l'annexe A le programme nous permettant d'évaluer exactement ces $b_k(\underline{a})$.

À partir du moment où nous pouvons évaluer exactement les deux premiers moments de ϕ , il est très intéressant de comparer si l'approximation normale de ϕ , avec ces paramètres, est plus adéquate que celles présentées à la section précédente. Aitchison et Shen (1980) montrent que les paramètres μ et σ^2 qui minimisent la distance de Kullback entre la densité de ϕ et celle d'une $\mathcal{N}(\mu, \sigma^2)$ sont effectivement l'espérance et la variance de ϕ . Ainsi, il est clair que lorsque les a_j sont petits et entiers, les moments exacts de ϕ devraient être utilisés si nous désirons approximer sa densité par celle d'une loi normale. Notons en terminant cette section que selon Aitchison et Shen (1980), si Lindley (1964) ainsi que Bloch

et Watson (1967) n'annoncèrent pas les moments exacts de ϕ , ce n'était pas dû au fait qu'ils étaient incapables de les trouver, mais plutôt parce qu'ils désiraient éviter l'utilisation des fonctions spéciales telle que $\psi(x)$. Nous voyons maintenant que cette restriction n'a plus sa raison d'être lorsque les a_j sont entiers.

3.3. UNE NOUVELLE APPROXIMATION DU LOGARITHME DU RAPPORT DE COTES

Nous regardons maintenant s'il n'y a pas d'autres alternatives que l'approximation normale pour estimer la distribution de ϕ . Cette approximation étant bonne lorsque les paramètres a_j sont élevés, nous nous intéresserons à des alternatives étant plus adéquates lorsque ces paramètres prennent des valeurs modérées. Nous présentons d'abord l'expansion de Latorre, qui représente la distribution de ϕ sous forme d'une somme infinie de fonctions hypergéométriques, ensuite une nouvelle approximation qui est plus précise que la distribution normale lorsque les paramètres a_j sont petits.

3.3.1. Expansion de Latorre

Les travaux de Latorre (1982, 1984 et 1985) sur la distribution exacte de ϕ se veulent une généralisation de Altham (1969) qui donnait la valeur exacte de $P[\exp(\phi) < 1]$. Principalement, Latorre (1982) montre comment écrire $P[\exp(\phi) < t]$ sous la forme d'une somme infinie et ce, pour tout t réel positif. Il est ensuite facile d'écrire la loi de ϕ sous cette même forme.

Théorème 3.3.1. *Soit ϕ le logarithme du rapport de cotes admettant la densité $\pi(x)$. Si le vecteur des paramètres inconnus $\underline{\theta}$ admet comme distribution une loi*

de Dirichlet de paramètres a_j , alors

$$\pi_\phi(x) = \begin{cases} Ke^{xa_1} F(a_1 + a_3, a_1 + a_2, a_+; 1 - e^x) & \text{si } x \leq 0, \\ Ke^{-xa_2} F(a_1 + a_2, a_2 + a_4, a_+; 1 - e^{-x}) & \text{si } x > 0, \end{cases}$$

où $K = \frac{B(a_1 + a_2, a_3 + a_4)}{B(a_1, a_3)B(a_2, a_4)}$ et

$$F(a, b, c; x) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{i=0}^{\infty} \frac{\Gamma(a+i)\Gamma(b+i)x^i}{\Gamma(c+i)i!}.$$

Notons que lorsque nous programmons cette expansion dans l'annexe A, la fonction hypergéométrique $F(a, b, c; x)$ est écrite différemment afin d'éviter certains problèmes de convergence, c'est-à-dire que nous l'écrivons sous la forme

$$F(a, b, c; x) = 1 + \sum_{i=1}^{\infty} \left[\prod_{j=0}^{i-1} \frac{(a+j)(b+j)}{(c+j)(j+1)} \right] x^i.$$

À partir de ce théorème, nous pouvons également trouver une représentation pour la fonction de répartition de ϕ . Par contre, tout comme $\pi_\phi(x)$, la fonction de répartition s'écrit sous la forme d'une somme infinie de quotients de fonctions gamma, une représentation difficile à manipuler. Cette fonction de répartition nous sera utile pour calculer des régions de crédibilité pour ϕ à la section suivante.

Corollaire 3.3.1. *Soit ϕ le logarithme du rapport de cotes, si θ admet une distribution de Dirichlet de paramètres a_j , alors la fonction de répartition $\Pi_\phi(x)$,*

correspondant à $P[\phi \leq x]$, peut s'écrire de la façon suivante :

$$\Pi_\phi(x) = \begin{cases} K \left[e^{xa_1} + \frac{\Gamma(a_+)}{\Gamma(a_1 + a_3)\Gamma(a_1 + a_2)} \sum_{j=1}^{\infty} \frac{\Gamma(a_1 + a_3 + j)\Gamma(a_1 + a_2 + j)}{\Gamma(a_+ + j)j!} \right. \\ \quad \times B(a_1, j + 1)I_{e^x}(a_1, j + 1) \Big] \text{ si } x \leq 0, \\ 1 - K \left[e^{-xa_2} + \frac{\Gamma(a_+)}{\Gamma(a_1 + a_2)\Gamma(a_2 + a_4)} \sum_{j=1}^{\infty} \frac{\Gamma(a_1 + a_2 + j)\Gamma(a_2 + a_4 + j)}{\Gamma(a_+ + j)j!} \right. \\ \quad \times B(a_2, j + 1)I_{e^{-x}}(a_2, j + 1) \Big] \text{ si } x > 0. \end{cases}$$

Notons que la fonction $I_x(p, q)$ correspond à la fonction bêta incomplète,

$$I_x(p, q) = B(p, q)^{-1} \int_0^x u^{p-1}(1-u)^{q-1} du.$$

Nous pouvons déceler deux lacunes aux expansions de Latorre. Tout d'abord, elles sont très difficiles à manipuler, principalement pour $\Pi_\phi(x)$ qui nécessite l'utilisation de fonctions bêtas incomplètes. Nous verrons également à la section 3.5 que nous devons calculer un grand nombre de termes avant d'obtenir une approximation suffisamment juste de la densité de ϕ ou de sa fonction de répartition.

3.3.2. Détermination de la nouvelle approximation

Nous présentons maintenant une nouvelle approximation de la distribution de ϕ . La motivation première de cette démarche consistait à trouver une approximation convergeant plus rapidement vers la vraie distribution de ϕ que l'approximation normale. Nous verrons à la section 3.5 que cette approximation donne souvent des résultats aussi satisfaisants que ceux de Latorre, mais elle s'écrit sous une forme beaucoup plus simple et nécessite le calcul de moins de termes.

Afin de bien comprendre la preuve du théorème menant à la détermination de cette nouvelle approximation, nous devons énoncer d'abord trois autres théorèmes en plus de présenter la formule suivante (voir Gradshteyn et Ryzhik, 1980 équation

8.326),

$$\frac{\Gamma(x+iy)}{\Gamma(x)} = \frac{x e^{i\psi(1)y}}{x+y} \prod_{n=1}^{\infty} \frac{e^{iy/n}(x+n)}{x+n+iy},$$

$\forall x \in \mathbb{R}$ et $\forall y \in \mathbb{R}^*$. Les preuves de ces trois théorèmes se retrouvent respectivement dans Spivak(1980, chapitre 19), Walker(1991, section 5.2) et Billingsley(1979, chapitre 26), notons que le troisième théorème est fréquemment appelé le théorème d'inversion.

Théorème 3.3.2. *Soit $f(x)$ une fonction telle que $f', f'', \dots, f^{(m+1)}$ sont définies sur $[a, x]$, alors*

$$f(x) = \sum_{j=0}^m \frac{f^{(j)}(a)}{j!} (x-a)^j + \frac{f^{(m+1)}(a+p(x-a))}{(m+1)!} (x-a)^{m+1},$$

où $0 < p < 1$.

Théorème 3.3.3. *Soit $\hat{f}(z)$ la transformée de Fourier de $f(t)$, c'est-à-dire*

$$\hat{f}(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-itz} dt.$$

Si $\int_{-\infty}^{\infty} f(t) dt, \int_{-\infty}^{\infty} t f(t) dt, \dots, \int_{-\infty}^{\infty} t^k f(t) dt$ sont finis, alors

$$\widehat{z^k f}(z) = i^k \left(\frac{\partial^k}{\partial z^k} \hat{f}(z) \right),$$

où $k \in \mathbb{N}$.

Théorème 3.3.4. *Soit $\varphi_X(t)$ la fonction caractéristique d'une variable aléatoire X et soit $f_X(x)$ sa fonction de densité, alors*

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \varphi_X(t) dt.$$

Nous pouvons maintenant énoncer le théorème nous menant à une nouvelle approximation de la densité de ϕ .

Théorème 3.3.5. *Soit ϕ le logarithme du rapport de cotes d'un tableau de contingence 2×2 lorsque le vecteur θ des paramètres inconnus admet une distribution de Dirichlet de paramètres a_j et soit Z la variable centrée et réduite de ϕ , c'est-à-dire $Z = [\phi - b_1(\underline{a})]b_2(\underline{a})^{-1/2}$. Posons également que*

$$g(t) = \exp \left\{ \sum_{k=3}^m \frac{b_k(\underline{a})}{k!} \left(\frac{it}{\sqrt{b_2(\underline{a})}} \right)^k + \frac{b_{m+1}(\underline{a} + itp_1\underline{c})}{(m+1)!} \left(\frac{it}{\sqrt{b_2(\underline{a})}} \right)^{m+1} \right\} \quad (3.3.1)$$

La fonction de densité de Z peut alors s'écrire sous la forme

$$\pi_Z(z) = \sum_{k=0}^m D_k(\underline{a}) M_k(z) + E_m(z),$$

où

$$\begin{aligned} D_k(\underline{a}) &= \frac{i^k g^{(k)}(0)}{\sqrt{2\pi k!}}, \\ M_k &= \frac{\partial^k}{\partial z^k} (e^{-z^2/2}), \\ E_m(z) &= \frac{1}{\sqrt{2\pi(m+1)!}} \int_{-\infty}^{\infty} e^{-itz} f(t) g^{(m+1)}(p_2 t) t^{m+1} dt, \end{aligned}$$

où $0 < p_1, p_2 < b_2(\underline{a})^{-1/2}$ et $f(t)$ correspond à la fonction de densité d'une loi normale centrée et réduite.

PREUVE. La première étape consiste à écrire le logarithme de la fonction caractéristique sous la forme d'une somme finie.

$$\log(\varphi_\phi(t)) = \log \left[\prod_{j=1}^4 \frac{\Gamma(a_j + itc_j)}{\Gamma(a_j)} \right]$$

$$\begin{aligned}
&= \sum_{j=1}^4 \log \left[\frac{\Gamma(a_j + itc_j)}{\Gamma(a_j)} \right] \\
&= \sum_{j=1}^4 \log \left[\frac{a_j e^{i\psi(1)c_j t}}{a_j + ic_j t} \prod_{n=1}^{\infty} \left(\frac{e^{ic_j t/n} (a_j + n)}{a_j + n + ic_j t} \right) \right] \\
&= \sum_{j=1}^4 \left[\sum_{n=0}^{\infty} \log \left(\frac{a_j + n}{a_j + n + ic_j t} \right) + c_j \left(\sum_{n=1}^{\infty} \frac{it}{n} + it\psi(1) \right) \right] \\
&= - \sum_{j=1}^4 \sum_{n=0}^{\infty} \log \left(1 + \frac{ic_j t}{a_j + n} \right) \\
&= \sum_{j=1}^4 \sum_{n=0}^{\infty} \left[\sum_{k=1}^m \frac{(-1)^k}{k} \left(\frac{ic_j t}{a_j + n} \right)^k + \frac{(-1)^k}{(m+1)} \left(\frac{ic_j t}{a_j + ic_j p_1 t} \right)^{m+1} \right] \\
&= \sum_{k=1}^m \frac{(it)^k}{k!} b_k(\underline{a}) + \frac{(it)^{m+1}}{(m+1)!} b_{m+1}(\underline{a} + itp_1 \underline{c}).
\end{aligned}$$

Nous pouvons maintenant exprimer la fonction caractéristique de Z sous une forme exponentielle,

$$\begin{aligned}
\varphi_Z(t) &= \exp \left\{ \sum_{k=2}^m \frac{(it)^k}{k!} b_k(\underline{a}) + \frac{(it)^{m+1}}{(m+1)!} b_{m+1}(\underline{a} + itp_1 \underline{c}) \right\} \\
&= g(t) \exp \left\{ \frac{-t^2 b_2(\underline{a})}{2} \right\}.
\end{aligned}$$

La densité de Z peut maintenant être déterminée à partir du théorème d'inversion,

$$\begin{aligned}
\pi_Z(z) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itz} \varphi_Z(t) dt \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-itz} g(t) f(t) dt \\
&= \int_{-\infty}^{\infty} e^{-itz} \left(\sum_{k=0}^m i^{-k} D_k(\underline{a}) t^k + \frac{g^{(m+1)}(p_2 t) t^{m+1}}{(m+1)!} \right) f(t) dt \\
&= \sum_{k=0}^m i^{-k} D_k(\underline{a}) \int_{-\infty}^{\infty} e^{-itz} t^k f(t) dt + E_m(z)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{k=0}^m D_k(\underline{a}) \left(\frac{\partial^k}{\partial z^k} (e^{-z^2/2}) \right) + E_m(z) \\
&= \sum_{k=0}^m D_k(\underline{a}) M_k(z) + E_m(z).
\end{aligned}$$

□

Notre approximation de la distribution de ϕ sera alors obtenue en négligeant le terme $E_m(z)$, elle aura donc la forme suivante :

$$\hat{\pi}_{(\phi, m)}(\phi) = \frac{\hat{\pi}_{(Z, m)} \left([\phi - b_1(\underline{a})] b_2(\underline{a})^{-1/2} \right)}{b_2(\underline{a})^{1/2}},$$

où

$$\hat{\pi}_{(Z, m)}(z) = \sum_{k=0}^m D_k(\underline{a}) M_k(z).$$

Cette façon de procéder pour approximer une fonction de densité est semblable à celle utilisant l'expansion d'Edgeworth (voir Hall, 1992 chapitre 2). D'ailleurs, Bloch et Watson (1967) mentionnent très brièvement cette approche et l'utilisent avec les moments approximatifs de ϕ .

Nous pouvons maintenant approximer la fonction de répartition de ϕ à partir du théorème précédent.

Corollaire 3.3.2. *Soit ϕ le logarithme du rapport de cotes lorsque le vecteur $\underline{\theta}$ des paramètres inconnus admet une distribution de Dirichlet de paramètres a_j . Soit Z , la variable centrée et réduite de ϕ , c'est-à-dire $Z = [\phi - b_1(\underline{a})] b_2(\underline{a})^{-1/2}$. La fonction de répartition de Z peut alors s'écrire sous la forme*

$$\Pi_Z(z) = \Phi(z) + \sum_{k=1}^m D_k(\underline{a}) M_{k-1}(z) + \int_{-\infty}^z E_m(z) dz,$$

où $\Phi(z)$ est la fonction de répartition d'une loi normale centrée et réduite et les termes D_k , M_k et $E_m(z)$ sont définis dans l'énoncé du théorème 3.3.5.

Notre approximation de la fonction de répartition de ϕ sera alors obtenue en négligeant l'intégrale du terme d'erreur $E_m(z)$, elle aura donc la forme suivante :

$$\widehat{\Pi}_{(\phi,m)}(z) = \widehat{\Pi}_{(z,m)} \left([z - b_1(\underline{a})] b_2(\underline{a})^{-1/2} \right),$$

où

$$\widehat{\Pi}_{(z,m)}(z) = \Phi(z) + \sum_{k=1}^m D_k(\underline{a}) M_{k-1}(z).$$

3.3.3. Étude de l'approximation

Nous étudions ici les trois principales composantes de notre nouvelle approximation, c'est-à-dire les coefficients $D_k(\underline{a})$, les fonctions $M_k(z)$ ainsi que le terme d'erreur $E_m(z)$, tous définis dans le théorème 3.3.5. Nous nous intéressons principalement à trouver des façons simples de calculer les $D_k(\underline{a})$ et les $M_k(z)$. De plus, nous trouverons la vitesse de convergence vers 0 des $D_k(\underline{a})$ et de $E_m(z)$ lorsque les paramètres a_j prennent de grandes valeurs.

Comme la fonction $M_k(z)$ correspond à la dérivée d'ordre k de la fonction $e^{-z^2/2}$, elle se représente sous la forme de cette fonction multipliée par un polynôme d'ordre k , où les puissances ont la même parité que k . La fonction $M_k(z)$ converge donc vers 0 lorsque $|z|$ tend vers l'infini. Cette convergence est par contre plus lente plus k est élevé. Le tableau 3.3.1 représente les fonctions de $M_0(z)$ à $M_{10}(z)$. Même s'il nous est impossible d'écrire la fonction $M_k(z)$ de façon explicite, il nous est possible de déterminer toutes les fonctions $M_k(z)$, sans avoir à

TABLEAU 3.3.1. Les 11 premières fonctions $M_k(z)$

k	$M_k(z)$
0	$e^{-z^2/2}$
1	$-ze^{-z^2/2}$
2	$(z^2 - 1)e^{-z^2/2}$
3	$(-z^3 + 3z)e^{-z^2/2}$
4	$(z^4 - 6z^2 + 3)e^{-z^2/2}$
5	$(-z^5 + 10z^3 - 15z)e^{-z^2/2}$
6	$(z^6 - 15z^4 + 45z^2 - 15)e^{-z^2/2}$
7	$(-z^7 + 21z^5 - 105z^3 + 105z)e^{-z^2/2}$
8	$(z^8 - 28z^6 + 210z^4 - 420z^2 + 105)e^{-z^2/2}$
9	$(-z^9 + 36z^7 - 378z^5 + 1260z^3 - 945z)e^{-z^2/2}$
10	$(z^{10} - 45z^8 + 630z^6 - 3150z^4 + 4725z^2 - 945)e^{-z^2/2}$

effectuer une seule dérivée, lorsque nous connaissons $M_0(z)$ et $M_1(z)$. Cette propriété des $M_k(z)$ provient de son étroite relation avec les polynômes d'Hermite et leur formule de récurrence (voir Gradshteyn et Ryzhik, 1980 équation 8.952).

Définition 3.3.1. Nous appelons polynôme d'Hermite d'ordre n la fonction

$$H_n(x) = (-1)^n e^{x^2} \left[\frac{\partial^n}{\partial x^n} (e^{-x^2}) \right].$$

Proposition 3.3.1. Lorsque $n \geq 2$,

$$H_n(x) = 2xH_{n-1}(x) - 2(n-1)H_{n-2}(x).$$

À partir de la définition et de la proposition précédente, nous pouvons maintenant donner la formule de récurrence pour $M_k(z)$ lorsque $k \geq 2$. Cette formule se retrouve dans la proposition suivante.

Proposition 3.3.2. *Posons*

$$M_k(z) = \frac{\partial^k}{\partial z^k}(e^{-z^2/2}).$$

Lorsque $|z| < \infty$, cette fonction peut également s'écrire sous la forme

$$M_k(z) = \begin{cases} e^{-z^2/2} & \text{si } k = 0, \\ -ze^{-z^2/2} & \text{si } k = 1, \\ -zM_{k-1}(z) - (k-1)M_{k-2}(z) & \text{si } k \geq 2. \end{cases}$$

PREUVE.

La preuve est triviale lorsque k vaut 0 ou 1, effectuons la preuve dans le cas où $k \geq 2$.

$$\begin{aligned} M_k(z) &= \frac{\partial^k}{\partial z^k}(e^{-z^2/2}) \\ &= \left(\frac{1}{\sqrt{2}}\right)^k \frac{\partial^k}{\partial x^k} e^{-x^2} \Big|_{x = \frac{z}{\sqrt{2}}} \\ &= H_k(z/\sqrt{2})(-\sqrt{2})^{-k} e^{-z^2/2} \\ &= (-\sqrt{2})^{-k} e^{-z^2/2} [\sqrt{2}zH_{k-1}(z/\sqrt{2}) - 2(k-1)H_{k-2}(z/\sqrt{2})] \\ &= -zM_{k-1}(z) - (k-1)M_{k-2}(z). \end{aligned}$$

□

Pour ce qui est des coefficients $D_k(\underline{a})$, il faut également dériver k fois la fonction exponentielle d'un certain polynôme afin de trouver leurs valeurs respectives. Les coefficients de ce polynôme n'étant pas des constantes mais plutôt des $b_l(\underline{a})$ où l varie de 2 à $(m+1)$, les termes $D_k(\underline{a})$ sont plus difficiles à calculer que les $M_k(z)$. Le tableau 3.3.2 contient d'ailleurs la liste des 11 premiers coefficients $D_k(\underline{a})$. Notons qu'afin d'alléger la notation de ce tableau, les coefficients b_l correspondent en fait aux coefficients $b_l(\underline{a})$.

TABLEAU 3.3.2. Les 11 premiers $D_k(\underline{a})$

k	$D_k(z)$
0	$1/\sqrt{2\pi}$
1	0
2	0
3	$(b_3)b_2^{-3/2}/(\sqrt{2\pi}3!)$
4	$(b_4)b_2^{-2}/(\sqrt{2\pi}4!)$
5	$(b_5)b_2^{-5/2}/(\sqrt{2\pi}5!)$
6	$(10b_3^2 + b_6)b_2^{-3}/(\sqrt{2\pi}6!)$
7	$(35b_3b_4 + b_7)b_2^{-7/2}/(\sqrt{2\pi}7!)$
8	$(35b_4^2 + 56b_3b_5 + b_8)b_2^{-4}/(\sqrt{2\pi}8!)$
9	$(280b_3^3 + 126b_4b_5 + 84b_3b_6 + b_9)b_2^{-9/2}/(\sqrt{2\pi}9!)$
10	$(2100b_3^2b_4 + 126b_5^2 + 210b_4b_6 + 120b_3b_7 + b_{10})b_2^{-5}/(\sqrt{2\pi}10!)$

Nous remarquons que le nombre de termes augmente plus rapidement pour $D_k(\underline{a})$ que pour $M_k(z)$. Par exemple, $M_{20}(z)$ est représenté par $e^{-z^2/2}$ multiplié par un polynôme de degré 20 comprenant 11 termes, alors que $D_{20}(\underline{a})$ se décompose en 49 termes qui sont fonction d'un produit de différents $b_l(\underline{a})$.

Il est toutefois possible d'écrire les coefficients $D_k(\underline{a})$ de façon récursive. La proposition suivante démontre en effet que tout $D_k(\underline{a})$ peut s'écrire en fonction des coefficients $D_0(\underline{a})$ à $D_{k-3}(\underline{a})$. Cette formule de récurrence est plus complexe que celle donnée pour $M_k(z)$, mais elle simplifie beaucoup les calculs. Cette proposition ne nous permet pas de trouver la valeur de $D_k(\underline{a})$ lorsque k est supérieur à m , mais nous n'utilisons jamais de tels coefficients dans notre approximation.

Proposition 3.3.3. *Soit $D_k(\underline{a})$ tel que*

$$D_k(\underline{a}) = \frac{i^k g^{(k)}(0)}{\sqrt{2\pi} k!},$$

où $g(t)$ est donné à l'équation (3.3.1).

Lorsque $k \leq m$ ces coefficients peuvent également être écrits de la façon suivante :

$$D_k(\underline{a}) = \begin{cases} \frac{1}{\sqrt{2\pi}} & \text{si } k = 0, \\ 0 & \text{si } k = 1, 2, \\ \sum_{l=3}^k \frac{D_{k-l}(\underline{a}) b_l(\underline{a}) (-1)^l}{k(l-1)! \sigma^l} & \text{si } 3 \leq k \leq m, \end{cases}$$

où $\sigma^2 = b_2(\underline{a})$.

PREUVE.

Posons que $f(t) = \log[g(t)]$, il est alors clair que

$$f^{(l)}(0) = \begin{cases} 0 & \text{si } l \leq 2, \\ b_l(\underline{a}) \left(\frac{i}{\sigma}\right)^l & \text{si } 3 \leq l \leq m. \end{cases}$$

et nous pouvons montrer que

$$g^{(k)}(t) = \sum_{l=1}^k \binom{k-1}{l-1} g^{(k-l)}(t) f^{(l)}(t),$$

par une induction semblable à celle effectuée dans la preuve du théorème 3.2.1.

Alors,

$$g^{(k)}(0) = \begin{cases} 1 & \text{si } k = 0, \\ 0 & \text{si } 0 < k < 3, \\ \sum_{l=3}^k \binom{k-1}{l-1} g^{(k-l)}(0) b_l(\underline{a}) \left(\frac{i}{\sigma}\right)^l & \text{si } 3 \leq k \leq m. \end{cases}$$

Le résultat est alors obtenu en remplaçant les $g^{(k)}(0)$ dans la définition de $D_k(\underline{a})$.

□

Dans notre nouvelle approximation, ce sont seulement les coefficients $D_k(\underline{a})$ qui dépendent des paramètres a_j . Il nous serait intéressant de montrer que tous ces $D_k(\underline{a})$, à l'exception de $D_0(\underline{a})$, convergent vers 0 lorsque les paramètres a_j tendent vers l'infini. Nous aurions alors que

$$\lim_{a_j \rightarrow \infty} \widehat{\pi}_{(Z,m)}(z) = D_0 M_0(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}},$$

ce qui n'entre pas en contradiction avec nos résultats sur la normalité asymptotique de ϕ . Comme le comportement asymptotique des $D_k(\underline{a})$ est relié à celui des $b_l(\underline{a})$, où l varie de 2 à k , nous devons donc étudier tout d'abord celui des $b_l(\underline{a})$. Mais pour étudier le comportement asymptotique des $b_l(\underline{a})$, nous devons d'abord trouver celui des fonctions $\zeta(l, a_j)$ puisque

$$b_l(\underline{a}) = (l-1)! \sum_{j=1}^4 (-c_j)^l \zeta(l, a_j).$$

Le théorème suivant, qui montre que la fonction $\zeta(l, a)$ se comporte comme $a^{-(l-1)}$ lorsque a est grand, nous permettra alors de déterminer complètement le comportement asymptotique des $b_l(\underline{a})$ et donc des $D_k(\underline{a})$.

Théorème 3.3.6. *Lorsque $l \geq 2$,*

$$\lim_{a \rightarrow \infty} a^{l-1} \zeta(l, a) = \frac{1}{l-1}.$$

Ainsi la fonction $\zeta(l, a)$ est $\mathcal{O}(a^{-(l-1)})$

PREUVE.

Soit $a \geq 1$ et $I = [a, \infty]$. Définissons une fonction $s(x)$ telle que

$$s(x) = \frac{1}{(a+n)^l} \text{ si } x \in [a+n, a+n+1], \text{ où } n = 0, 1, 2, \dots$$

Cette fonction est telle que

$$\int_I s(x) dx = \sum_{n=0}^{\infty} \frac{1}{(a+n)^l} = \zeta(l, a).$$

Soient deux fonctions $f(x)$ et $g(x)$ telles que

$$\begin{aligned} f(x) &= \frac{1}{x^l}, \\ g(x) &= \frac{1}{(x-1)^l}. \end{aligned}$$

Il est alors clair que

$$f(x) \leq s(x) \leq g(x) \forall x \in I.$$

En intégrant ces trois fonctions sur I puis en multipliant le résultat par a^{l-1} nous obtenons

$$\frac{1}{l-1} \leq a^{l-1} \zeta(l, a) \leq \left(\frac{a}{a-1} \right)^{l-1} \frac{1}{l-1}.$$

Nous obtenons alors le résultat puisque la borne supérieure tend vers $\frac{1}{l-1}$ lorsque a tend vers l'infini. \square

Nous pouvons maintenant établir le comportement asymptotique de $b_i(\underline{a})$, une conséquence directe du théorème précédent.

Corollaire 3.3.3. *Soit $b_l(\underline{a})$ tel que*

$$b_l(\underline{a}) = (l-1)! \sum_{j=1}^4 (-c_j)^l \zeta(l, a_j),$$

où $l \geq 2$. Alors $b_l(\underline{a}) = \mathcal{O}(a_{\min}^{l-1})$ où $a_{\min} = \min(a_1, a_2, a_3, a_4)$.

Nous pouvons maintenant étudier le comportement asymptotique des $D_k(\underline{a})$, ce qui est plus complexe. Nous pouvons déduire à partir de la définition de $D_k(\underline{a})$ qu'il s'écrit comme une somme de termes fonction des coefficients $b_l(\underline{a})$, où l varie entre 3 et k , somme étant par la suite divisée par $b_2(\underline{a})^{k/2}$. La forme de l'ensemble de ces termes est un peu plus complexe à décrire, elle se lit comme suit :

$$\left\{ b_{l_1}(\underline{a}) b_{l_2}(\underline{a}) \dots b_{l_p}(\underline{a}) \mid 3 \leq l_1 \leq l_2 \leq \dots \leq l_p \leq k \text{ et } \sum_{i=1}^p l_i = k \right\},$$

où p peut varier d'un terme à l'autre. Par exemple, si $k = 10$, nous voyons par le tableau 3.3.2 que $D_{10}(\underline{a})$ contient cinq termes. Les vecteurs (l_1, l_2, \dots, l_p) décrivant ces cinq termes sont $(3,3,4)$, $(3,7)$, $(4,6)$, $(5,5)$ et (10) . Lorsque nous écrivons l'entier k sous la forme d'un multiple de 3 et de son reste, c'est-à-dire sous la forme $3q + r$ où $r \leq 2$, nous pouvons alors montrer que la valeur maximale de p sera toujours de q , le résultat de la division entière de k par 3. Lorsque p atteint sa valeur maximale et que r vaut 0 ou 1, le seul vecteur de dimension q aura la forme $(3, \dots, 3, 3+r)$. Lorsque $r = 2$ et $q > 1$, les deux seuls vecteurs de dimension q seront $(3, \dots, 3, 3, 5)$ et $(3, \dots, 3, 4, 4)$.

Ces remarques sur la forme des différents termes des $D_k(\underline{a})$ est nécessaire afin d'établir leur comportement asymptotique. Chaque $b_i(\underline{a})$ étant $\mathcal{O}(a_{min}^{-(i-1)})$, il est clair qu'un terme de la forme $b_{l_1}(\underline{a})b_{l_2}(\underline{a})\dots b_{l_p}(\underline{a})$ sera $\mathcal{O}(a_{min}^{-(k-p)})$. Ainsi, pour tout $D_k(\underline{a})$, les termes qui convergent le moins rapidement lorsque a_{min} est grand sont ceux où p atteint sa valeur maximale q et seront donc $\mathcal{O}(a_{min}^{-(k-q)})$. Finalement, comme chaque terme des $D_k(\underline{a})$ est multiplié par $b_2(\underline{a})^{-k/2}$, nous avons la proposition suivante :

Proposition 3.3.4. *Soit $k \geq 3$. Si nous écrivons l'entier k sous la forme $3q + r$ où $r = \{0, 1, 2\}$, alors $D_k(\underline{a})$ est $\mathcal{O}(a_{min}^{-(q+r)/2})$.*

PREUVE.

Comme $b_2(\underline{a})^{-k/2}D_k(\underline{a}) = \mathcal{O}(a_{min}^{-(k-m)})$ et que $k = 3q + r$, nous avons donc que

$$\begin{aligned} D_k(\underline{a}) &= \mathcal{O}(a_{min}^{-(k-q)})\mathcal{O}(a_{min}^{k/2}) \\ &= \mathcal{O}(a_{min}^{-(k/2-q)}) \\ &= \mathcal{O}(a_{min}^{-(q+r)/2}). \end{aligned}$$

□

Ainsi, lorsque nous utilisons l'approximation normale, nous remarquons que des termes tels que $D_3(\underline{a})$ et $D_4(\underline{a})$ respectivement $\mathcal{O}(a_{min}^{-1/2})$ et $\mathcal{O}(a_{min}^{-1})$ sont négligés. Notons ici que nous parlons de l'approximation de la densité de ϕ et non pas de ces moments.

Nous terminons cette section avec l'étude du terme d'erreur $E_m(z)$. Tout comme les $D_k(\underline{a})$, nous étudions le comportement asymptotique de $E_m(z)$ lorsque

les paramètres a_j tendent vers l'infini. La détermination du comportement asymptotique de $E_m(z)$ est beaucoup plus ardue que pour $D_k(\underline{a})$. Nous nous restreignons donc au cas où m est congru à 2 modulo 4, la preuve étant plus simple dans ce contexte.

Proposition 3.3.5. *Soit $E_m(z)$ tel que*

$$E_m(z) = \frac{\int_{-\infty}^{\infty} e^{-itz} f(t) g^{(m+1)}(p_2 t) t^{m+1} dt}{\sqrt{2\pi}(m+1)!},$$

où $f(t)$ correspond à la fonction de densité d'une loi normale centrée et réduite, $0 < p_2 < 1/\sigma$, $\sigma = b_2(\underline{a})^{1/2}$ et où $g(t)$ est donné par l'équation (3.3.1). Si nous supposons que $m \bmod 4 = 2$, alors

$$|E_m(z)| \leq \mathcal{O}(a_{\min}^{-(m-1)/2}).$$

PREUVE.

Comme

$$|E_m(z)| \leq \frac{\int_{-\infty}^{\infty} f(t) |g^{(m+1)}(p_2 t)| |t|^{m+1} dt}{\sqrt{2\pi}(m+1)!},$$

si nous montrons que la fonction $|g^{(m+1)}(p_2 t)|$ est bornée par une fonction $h(\underline{a})$ pour toute valeur de t , nous obtiendrons alors

$$\begin{aligned} |E_m(z)| &\leq h(\underline{a}) \frac{\int_{-\infty}^{\infty} f(t) |t|^{m+1} dt}{\sqrt{2\pi}(m+1)!} \\ &= h(\underline{a}) \frac{2^{(m-1)/2} \Gamma(\frac{m}{2} + 1)}{\sqrt{\pi}(m+1)!}. \end{aligned}$$

Il nous suffira alors de montrer que $h(\underline{a})$ est $\mathcal{O}(a_{\min}^{-(m+1)/2})$ pour obtenir le résultat.

La fonction $g(t)$ étant l'exponentielle d'une fonction se bornant par un polynôme, nous pouvons montrer que

$$|g^{(m+1)}(t)| \leq |g(t)| \sum_{l=0}^{(m+1)^{m+1}} k_l(\underline{a}) |t|^l,$$

où les $k_l(\underline{a})$ sont des termes bornant des fonctions de t , fonctions obtenues en dérivant un nombre fini de fois la fonction $b_{m+1}(\underline{a} + itp_1\underline{c})$. Si nous montrons que la fonction $|g(t)|$ est bornée, nous saurons qu'il existe un point t_0 maximisant le côté droit de cette inégalité. C'est cette fonction évaluée en t_0 qui constituera notre fonction $h(\underline{a})$.

Calculons maintenant $|g(t)|$, notons que $\Re(z)$ et $\Im(z)$ sont les fonctions nous donnant respectivement la partie réelle et la partie imaginaire du nombre complexe z .

$$|g(t)| = \exp \left\{ \sum_{k=2}^{m/2} \frac{b_{2k}(\underline{a})(-t^2)^k}{\sigma^{2k} 2k!} + \Re(\Omega_{m+1}) \right\}, \quad (3.3.2)$$

où

$$\begin{aligned} \Re(\Omega_{m+1}) &= \Re \left[\left(\frac{b_{m+1}(\underline{a} + ip_1\underline{c})}{(m+1)!} \right) \left(\frac{it}{\sigma} \right)^{m+1} \right] \\ &= \Re \left[\sum_{j=1}^4 \sum_{n=0}^{\infty} \frac{(-c_j it \sigma^{-1})^{m+1}}{(m+1)} (a_j + n + ic_j p_1 t)^{-(m+1)} \right] \\ &= \sum_{j=1}^4 \sum_{n=0}^{\infty} \left(\frac{1}{m+1} \right) \left(\frac{c_j t \sigma^{-1}}{(a_j + n)^2 + (p_1 t)^2} \right)^{m+1} \Im [(a_j + n - ic_j p_1 t)^{m+1}] \\ &= \sum_{j=1}^4 \sum_{n=0}^{\infty} \left(\frac{1}{m+1} \right) \left(\frac{c_j t \sigma^{-1} (a_j + n)}{(a_j + n)^2 + (p_1 t)^2} \right)^{m+1} \\ &\quad \times \sum_{l=0}^{m/2} \binom{m+1}{2l} \left(\frac{p_1 t}{a_j + n} \right)^{2l} (-1)^l. \end{aligned}$$

Nous remarquons que

$$\lim_{|t| \rightarrow \infty} \Re \left[\left(\frac{b_{m+1}(\underline{a} + ip_1 t \underline{c})}{(m+1)!} \right) \left(\frac{it}{\sigma} \right)^{m+1} \right] = 0.$$

Puisque $m/2$ est impair, nous voyons dans l'équation (3.3.2) que le terme dominant de $|g(t)|$ est

$$\exp \left\{ \frac{-b_m(\underline{a})t^m}{m!\sigma^m} \right\}$$

et puisque $b_m(\underline{a})$ est positif, nous pouvons montrer que $|g(t)|$ atteint son maximum en un certain point et que cette fonction est donc bornée.

Pour ce qui est du comportement asymptotique de $h(\underline{a})$, en examinant la forme de $g^{m+1}(t)$, il est évident que le terme dominant de $h(\underline{a})$, c'est-à-dire le terme convergeant le moins rapidement, se retrouve en $k_0(\underline{a})$ et correspond à $k|b_{m+1}(\underline{a})|/b_2(\underline{a})^{(m+1)/2}$. Ainsi, la fonction $h(\underline{a})$ est $\mathcal{O}(a_{\min}^{-(m-1)/2})$. \square

3.4. CALCUL DE RÉGIONS DE CRÉDIBILITE

À partir de la proposition 1.4.1, il est facile de déterminer un intervalle de confiance de niveau $1 - \alpha$ pour ϕ . Lorsque a_{\min} est grand, il en est de même pour sa région α -crédible PHDP lorsque nous approximations la densité de ϕ par une loi normale. Afin de bien comparer, à la section suivante, ces deux estimations par intervalle à celles obtenues à l'aide de $\widehat{\pi}_{(\phi, m)}(z)$ ou de l'expansion de Latorre, nous montrons ici comment déterminer une région α -crédible de ces deux dernières approximations. Notons que les régions α -crédibles décrites ici ne seront pas PHDP, mais plutôt sous la forme $[a, b]$ où

$$\int_{-\infty}^a \pi_{\phi}(x) dx = \int_b^{\infty} \pi_{\phi}(x) dx = \frac{\alpha}{2}.$$

En procédant ainsi, il nous est plus facile de comparer les extrémités de ces deux régions de crédibilité avec celles calculées à partir de la loi normale. De plus, de par la forme de ces deux approximations, il nous est très difficile de trouver la région α -crédible PHDP.

Débutons avec l'expansion de Latorre, notons respectivement $\widehat{\pi}_{(L,m)}(x)$ et $\widehat{\Pi}_{(L,m)}(x)$ la fonction de densité et de répartition de ϕ lorsque nous nous limitons aux m premiers termes de l'expansion, nous cherchons donc les valeurs de a et b telles que, pour un m fixé,

$$\widehat{\Pi}_{(L,m)}(a) = \frac{\alpha}{2} \text{ et } \widehat{\Pi}_{(L,m)}(b) = 1 - \frac{\alpha}{2}.$$

Nous étudierons à la section suivante l'effet de la valeur de m sur les valeurs de a et b .

Afin de trouver la solution de chacune de ces deux équations, nous utiliserons la méthode itérative de Newton (voir Scheid, 1988 chapitre 25). À partir de deux valeurs initiales a_0 et b_0 , nous itérons de la façon suivante :

$$\begin{aligned} a_{n+1} &= a_n - \frac{\widehat{\Pi}_{(L,m)}(a_n) - \frac{\alpha}{2}}{\pi_{(L,m)}(a_n)}, \\ b_{n+1} &= b_n - \frac{\widehat{\Pi}_{(L,m)}(b_n) + \frac{\alpha}{2} - 1}{\pi_{(L,m)}(b_n)}, \end{aligned}$$

avec comme critère d'arrêt

$$\begin{aligned} |\widehat{\Pi}_{(L,m)}(a_{n_1}) - \frac{\alpha}{2}| &< 10^{-10}, \\ |\widehat{\Pi}_{(L,m)}(b_{n_2}) + \frac{\alpha}{2} - 1| &< 10^{-10}. \end{aligned}$$

Notre région α -crédible sera alors l'intervalle $[a_{n_1}, b_{n_2}]$.

L'approche est la même pour ce qui est de $\hat{\pi}_{(\phi,m)}$ et $\hat{\Pi}_{(\phi,m)}$. Nous itérons alors de la façon suivante :

$$\begin{aligned} a_{n+1} &= a_n - \frac{\hat{\Pi}_{(\phi,m)}(a_n) - \frac{\alpha}{2}}{\hat{\pi}_{(\phi,m)}(a_n)}, \\ b_{n+1} &= b_n - \frac{\hat{\Pi}_{(\phi,m)}(b_n) + \frac{\alpha}{2} - 1}{\hat{\pi}_{(\phi,m)}(b_n)}, \end{aligned}$$

avec le même critère d'arrêt que pour l'expansion de Latorre.

Pour ce qui est du choix des valeurs initiales a_0 et b_0 dans ces deux cas, une excellente façon de procéder consiste à prendre les extrémités de la région α -crédible de l'approximation normale de ϕ utilisée avec ses moments exacts. Les valeurs initiales sont donc

$$\begin{aligned} a_0 &= \sqrt{b_2(\underline{a})} z_{\alpha/2} + b_1(\underline{a}), \\ b_0 &= \sqrt{b_2(\underline{a})} z_{1-\alpha/2} + b_1(\underline{a}), \end{aligned}$$

où z_α correspond au quantile d'ordre α d'une distribution normale centrée et réduite.

3.5. COMPARAISON DES PRINCIPALES APPROCHES

Nous terminons ce chapitre en comparant différentes approches, bayésiennes et fréquentistes, pour tester l'indépendance de variables aléatoires catégorielles. Cette étude sera faite à partir de trois différents tableaux de contingence 2×2 . Il est à noter que la taille échantillonnale n'est jamais supérieure à 30 dans ces trois tableaux, les résultats étant très similaires d'une approche à l'autre lorsque n prend de grande valeur.

Le premier jeu de données provient d'une étude sur la criminalité des frères jumeaux, ces données furent publiées dans Lange (1931). Parmi 30 criminels ayant tous un frère jumeau, on observa les deux variables catégorielles suivantes : une variable ordinale visant à connaître si l'autre frère jumeau avait déjà été condamné pour un crime, ainsi qu'une variable nominale déterminant si ces jumeaux étaient de type génétique monozygote ou dizygote. Le tableau de contingence concernant ces données est représenté par le tableau 3.5.1. Notons que dans chacun des tableaux de contingence présentant les données de cette section, le chiffre entre parenthèses représente \hat{m}_j^* , la valeur de n_j espérée sous l'hypothèse d'indépendance (voir l'équation (1.3.4)). Tester l'indépendance de ces variables consiste à vérifier si le type génétique des frères jumeaux a une influence sur la criminalité du frère jumeau d'un criminel.

TABLEAU 3.5.1. *Tableau de contingence du premier jeu de données*

	Coupables	Innocents	Total
Dizygotes	2 (6.8)	15 (10.2)	17
Monozygotes	10 (5.2)	3 (7.8)	13
Total	12	18	30

Le second jeu de données provient de Fisher (1935). En buvant une tasse de thé, une femme déclare être en mesure de détecter ce qui a été versé en premier entre le thé et le lait. On lui fit boire 8 tasses de thé et on nota à chaque fois si sa prédiction était la bonne. Cette expérience étudie donc deux variables nominales, la prédiction et la vraie réponse. L'indépendance de ces deux variables signifierait

que cette dame est incapable de détecter ce qui est versé en premier. Les résultats sont regroupés dans le tableau de contingence 3.5.2.

TABLEAU 3.5.2. *Tableau de contingence du deuxième jeu de données*

Vérité\Prédiction	Thé	Lait	Total
Lait	1 (2)	3 (2)	4
Thé	3 (2)	1 (2)	4
Total	4	4	8

Pour ce qui est du dernier jeu de données, il provient de l'étude de Chin *et al.* (1961). Parmi les 6 habitants de Des Moines ayant plus de 40 ans et souffrant de la poliomyélite, on s'intéressa à savoir si le fait d'avoir reçu le vaccin Salk avait eu une influence sur la probabilité d'être paralysé. Ces résultats sont regroupés dans le tableau 3.5.3. Nous remarquons qu'une des cellules ce de tableau a 0 comme valeur, ce qui nous empêchera d'utiliser deux des approches fréquentistes, vues au chapitre 1, car elles ne sont valables que lorsque tous les n_j sont positifs. Ces deux approches sont celles utilisant la statistique G^2 de Wilks (voir l'équation (1.3.6)) et la distribution asymptotique de $\hat{\phi}$ (voir la proposition 1.4.1).

Ces trois expériences seront modélisées de la même façon, nous supposons que le vecteur (n_1, n_2, n_3, n_4) représentant la répartition des effectifs est distribué selon une loi multinomiale (voir l'équation (1.2.2)) de paramètres $\underline{\theta}$ et n , où n est connu. Lorsque nous effectuons des inférences bayésiennes, nous supposons que le vecteur des paramètres inconnus $\underline{\theta}$ admet une loi de Dirichlet de paramètres $a_j = 1$ pour $j = 1, \dots, 4$. Étant donnée l'absence d'information *a priori* dans ces expériences, nous modélisons le vecteur $\underline{\theta}$ avec une loi *a priori* uniforme sur le

TABLEAU 3.5.3. *Tableau de contingence du troisième jeu de données*

	Paralysés	Non-paralysés	Total
Vaccinés	0 (0.33)	1 (0.67)	1
Non-vaccinés	2 (1.67)	3 (3.33)	5
Total	2	4	6

simplexe \mathcal{S}^4 . Notre loi *a posteriori* sera donc une loi de Dirichlet de paramètres $n_j + 1$ pour $j = 1, \dots, 4$. Incorporer une information *a priori* aussi faible nous permettra de mieux comparer les résultats entre les approches bayésiennes et fréquentistes. Cela pourrait montrer les avantages à utiliser une approche bayésienne même lorsqu'il n'y a pas d'information *a priori* disponible, surtout lorsque la taille échantillonnale est très petite où lorsque certains n_j sont nuls.

Clairement, il n'est pas logique de modéliser le deuxième jeu de données avec une loi multinomiale, certains effectifs marginaux étant connus. Le but de cette section étant de comparer différentes méthodes, nous modélisons nos données de façon à ce que plusieurs méthodes s'appliquent. Notons entre autres que pour chacun de ces jeux de données nous effectuerons le test exact de Fisher (voir section 1.5), c'est donc dire que nous supposerons à ce moment que les effectifs marginaux donnés sont connus avant l'expérimentation.

Débutons maintenant avec la comparaison des différentes méthodes fréquentistes. Pour chaque jeu de données, nous avons testé l'indépendance de ces variables aléatoires catégorielles avec l'un des cinq tests suivants : le test d'indépendance utilisant la statistique X^2 de Pearson (voir l'équation (1.3.1)), celui avec la statistique X_c^2 qui utilise la correction de Yates pour les tableaux de contingence 2×2 (voir l'équation (1.3.5)), le test du rapport de vraisemblance avec la

statistique G^2 de Wilks (voir l'équation (1.3.6)), le test bilatéral exact de Fisher et finalement, le test d'indépendance basé sur la distribution asymptotique de $\hat{\phi}$ (voir la proposition 1.4.1). Les résultats se trouvent tous dans le tableau 3.5.4, notons que les intervalles de confiance que nous y retrouvons ont un niveau de signification de 95%. En ce qui a trait au troisième jeu de données, le test d'indépendance basé sur la distribution de $\hat{\phi}$ est en fait celui basé sur la distribution de $\bar{\phi}$ (voir le corollaire 1.4.1).

TABLEAU 3.5.4. *Comparaison des méthodes fréquentistes*

Méthodes	JEU1	JEU2	JEU3
X^2 (valeur-p)	13.033 (0.0003)	2.000 (0.1573)	0.600 (0.4386)
X_c^2 (valeur-p)	10.458 (0.0012)	0.500 (0.4795)	0.150 (0.6985)
G^2 (valeur-p)	7.010 (0.0008)	1.047 (0.3063)	Impossible
Fisher	(0.0005)	(0.4857)	(1.0000)
I.C. ($\hat{\phi}$) $\hat{\phi}$	(-5.179,-1.259) -3.219	(-5.398,1.003) -2.197	(-4.351,2.827) -0.762

La principale constatation de cette analyse est que pour chaque jeu de données, il n'y a aucune contradiction entre les différentes méthodes d'inférence. Analysons chacun de ces jeux de données à un niveau de signification de 5%. Dans le premier jeu de données, les cinq méthodes rejettent clairement l'hypothèse d'indépendance. De plus, l'intervalle de confiance pour $\hat{\phi}$ ne prenant que des valeurs négatives, cela suggère que cette hypothèse d'indépendance a été rejetée au détriment d'une association négative entre les variables aléatoires. Pour ce

qui est des deux autres jeux de données, l'hypothèse d'indépendance n'est jamais rejetée à 5%, et les valeurs-p sont très élevées. Notons par contre que si le test exact bilatéral de Fisher a une valeur-p de 1 dans le troisième cas, c'est qu'il en est toujours ainsi dès qu'un des n_j vaut 0. Rappelons de plus qu'à l'exception du test de Fisher, toutes les autres méthodes utilisent la distribution asymptotique de leur statistique respective, distributions rarement appropriées lorsque la taille échantillonnale est inférieure à 10. De plus, nous rappelons qu'il est déconseillé d'utiliser les statistiques X^2 et X_c^2 lorsque certaines fréquences espérées sont inférieures à cinq, ce qui est pourtant le cas pour tous les \hat{m}_j^* de ces deux tableaux de contingence. C'est pourquoi il serait préférable d'utiliser une approche bayésienne, qui n'impose aucune contrainte sur la taille échantillonnale.

Pour chacun des 3 jeux de données, nous avons comparé sept différentes méthodes. Tout d'abord, nous avons retenu quatre approximations normales en utilisant les moments approximatifs de Lindley (1964), Bloch et Watson (1967) et O'Hagan (1994, chapitre 10) ainsi que l'approximation normale à l'aide des moments exacts. De plus, nous avons utilisé la nouvelle approximation avec $m = 5$ ainsi que l'expansion de Latorre avec $m = 1000$. Finalement, nous avons simulé 10 000 variables aléatoires provenant d'une loi de Dirichlet de paramètres $n_j + 1$ et pour chacune de ces réalisations nous avons calculé le logarithme du rapport de cotes. Pour chacun de ces modèles, nous présentons leurs estimés respectifs quant à l'espérance de ϕ , sa variance, sa région de crédibilité ainsi qu'un graphique estimant sa densité sous chacune de ces méthodes. Les résultats sont présentés dans les tableaux 3.5.4 à 3.5.6 ainsi que les figures 3.5.1 à 3.5.3.

Notons que en ce qui a trait à la simulation des ϕ , nous avons utilisé la moyenne et la variance échantillonnale de ces 10 000 données afin d'estimer l'espérance et la variance de ϕ . Pour ce qui est de la région de crédibilité, au niveau

de signification de 5%, nous avons utilisé les quantiles empiriques d'ordre 2.5% et 97.5%. Finalement, nous présentons l'histogramme de ces résultats afin d'avoir une idée sur le comportement de la fonction de densité de ϕ .

Nous remarquons tout d'abord que toutes ces méthodes nous mènent aux mêmes conclusions à propos de l'indépendance de ces variables aléatoires. Les données du premier jeu nous permettent de rejeter l'indépendance, ce qui n'est pas le cas pour les deux autres jeu de données. Ces conclusions sont les mêmes que celles obtenues par les méthodes fréquentistes. Pour ce qui est des 4 approximations normales, nous remarquons tout d'abord que parmi les 3 méthodes n'utilisant pas les moments exacts, les moments approximatifs de Bloch et Watson (1967) sont toujours les plus près des vrais moments. Pour le premier jeu de données, nous remarquons que les estimations sont très similaires entre elles, cela est probablement dû au fait que la taille échantillonnale était relativement grande avec 30 observations. Par contre, quand la taille échantillonnale est réduite, les résultats diffèrent. Nous remarquons que l'approximation avec les moments de O'Hagan (1994, chapitre 10) nous procure des résultats très différents des autres, surtout en ce qui a trait à l'estimation de la variance. En effet, cette approximation semble systématiquement sous-estimer la variance lorsque la taille échantillonnale est petite, elle nous procure donc des régions de crédibilité beaucoup moins larges que les autres. De plus, il semble que plus la taille échantillonnale est petite, plus les estimés de l'espérance et la variance sont éloignés des vraies valeurs. Par contre, l'approximation normale avec les moments exacts ne produit pas nécessairement de meilleure régions de crédibilité que les trois autres avec ces données.

Nous remarquons également que la nouvelle approximation et l'expansion de Latorre donnent tous deux des résultats très près de ceux produits par la simulation. Pour chaque jeu de données, nous voyons d'ailleurs que ces deux densités pourraient bien représenter celle suggérée par la simulation. Nous remarquons par contre que dans le dernier jeu de données, la région de notre approximation possède une borne inférieure qui diffère des deux autres. Cela semble dû au fait que les paramètres y sont très petits et que cette approximation requiert plus de termes. En prenant un terme de plus, notre région critique passe d'ailleurs de $(-4.519, 2.618)$ à $(-4.09, 2.407)$.

Finalement, nous remarquons à quel point il faut prendre beaucoup de termes dans l'expansion de Latorre afin d'obtenir des résultats convenables. La figure 3.5.4 nous montre qu'il faut prendre un minimum de 200 termes de cette expansion pour obtenir un graphique adéquat de la fonction de densité. Ces problèmes sont encore plus flagrants lorsque nous estimons la fonction de répartition. En effet, nous avons dû prendre 1000 termes avant d'obtenir une région α -crédible qui ne variait pas trop lorsque le nombre de termes augmente. Ces constatations nous portent à privilégier notre nouvelle approximation qui donne des résultats semblables aux simulations mais avec seulement 5 termes.

Nous avons donc présenté et comparé plusieurs approximations de la distribution *a posteriori* du logarithme du rapport de cotes. Il en ressort que notre nouvelle approximation et l'expansion de Latorre, estiment très bien cette distribution et que, en ce qui concerne les approximations normales, c'est celle utilisant les moments exacts qui performe le mieux. Nous reviendrons sur ces méthodes dans la conclusion, où nous présenterons les avantages et inconvénients de chaque approximation.

TABLEAU 3.5.5. *Comparaison des méthodes bayésiennes pour le premier jeu de données*

Méthodes	$\hat{E}[\phi]$	$\hat{V}[\phi]$	Régions de crédibilité
O'Hagan	-2.901	0.737	(-4.583,-1.218)
Lindley	-2.923	0.846	(-4.725,-1.121)
Bloch et Watson	-2.915	0.842	(-4.714,-1.116)
Moments exacts	-2.914	0.838	(-4.709,-1.119)
$\hat{\pi}_{(\phi,5)}$	-2.914	0.838	(-4.856,-1.216)
$\hat{\pi}_{(L,1000)}$	N.D.	N.D.	(-4.830,-1.235)
Simulation	-2.924	0.850	(-4.864,-1.258)

TABLEAU 3.5.6. *Comparaison des méthodes bayésiennes pour le deuxième jeu de données*

Méthodes	$\hat{E}[\phi]$	$\hat{V}[\phi]$	Régions de crédibilité
O'Hagan	-1.636	1.500	(-4.037,0.764)
Lindley	-1.695	1.9048	(-4.400,1.010)
Bloch et Watson	-1.673	1.8848	(-4.364,1.018)
Moments exacts	-1.667	1.858	(-4.338,1.005)
$\hat{\pi}_{(\phi,5)}$	-1.667	1.858	(-4.542,0.890)
$\hat{\pi}_{(L,1000)}$	N.D.	N.D.	(-4.502,0.863)
Simulation	-1.677	1.891	(-4.534,0.839)

TABLEAU 3.5.7. *Comparaison des méthodes bayésiennes pour le troisième jeu de données*

Méthodes	$\hat{E}[\phi]$	$\hat{V}[\phi]$	Régions de crédibilité
O'Hagan	-0.614	2.0833	(-3.443,2.215)
Lindley	-0.762	3.3524	(-4.351,2.827)
Bloch et Watson	-0.699	3.1863	(-4.197,2.800)
Moments exacts	-0.667	2.969	(-4.044,2.710)
$\hat{\pi}_{(\phi,5)}$	-0.667	2.969	(-4.519,2.618)
$\hat{\pi}_{(L,1000)}$	N.D.	N.D.	(-4.123,2.496)
Simulation	-0.665	2.844	(-4.249,2.471)

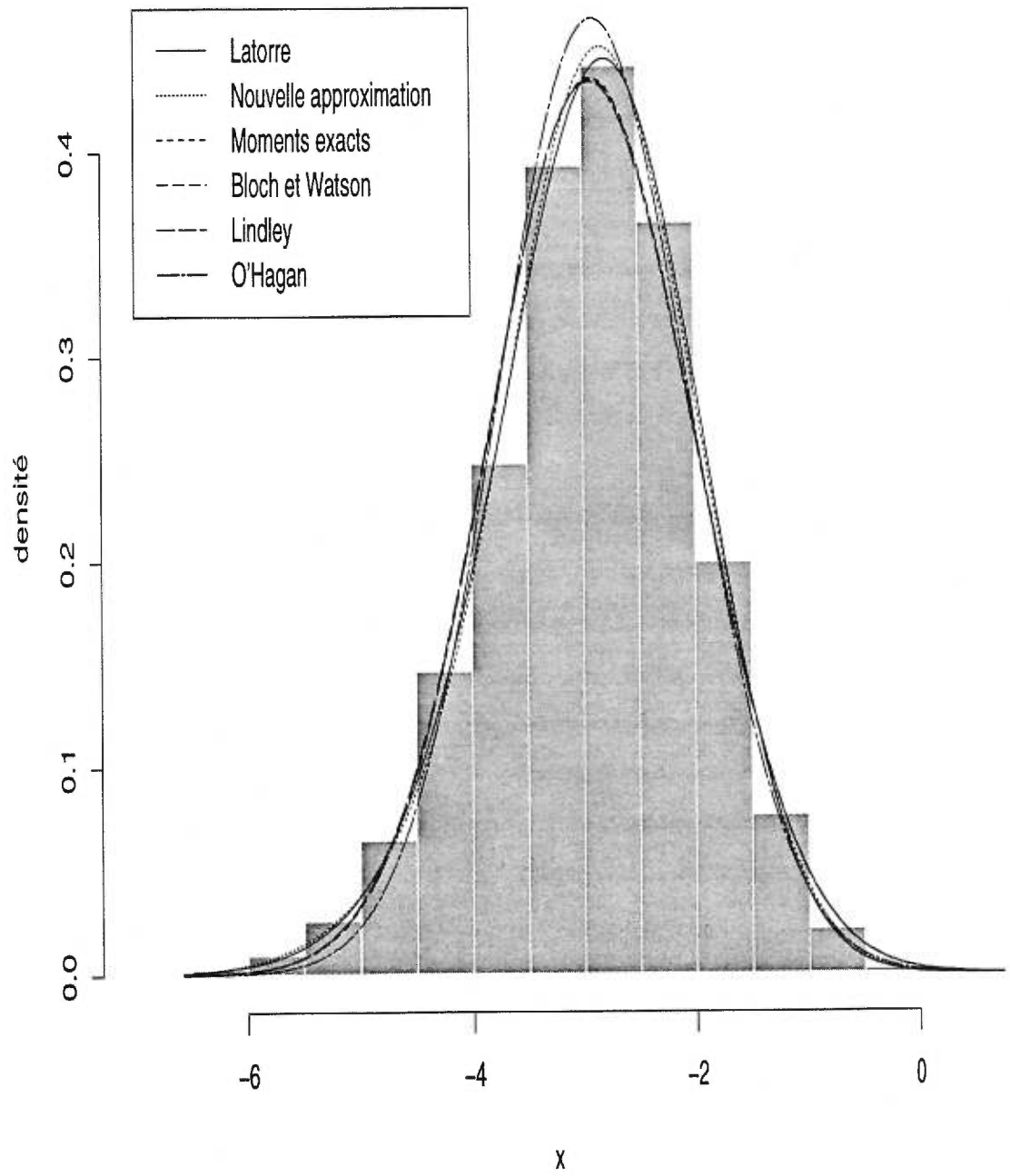


FIGURE 3.5.1. *Approximations de la distribution de ϕ à partir du premier jeu de données.*

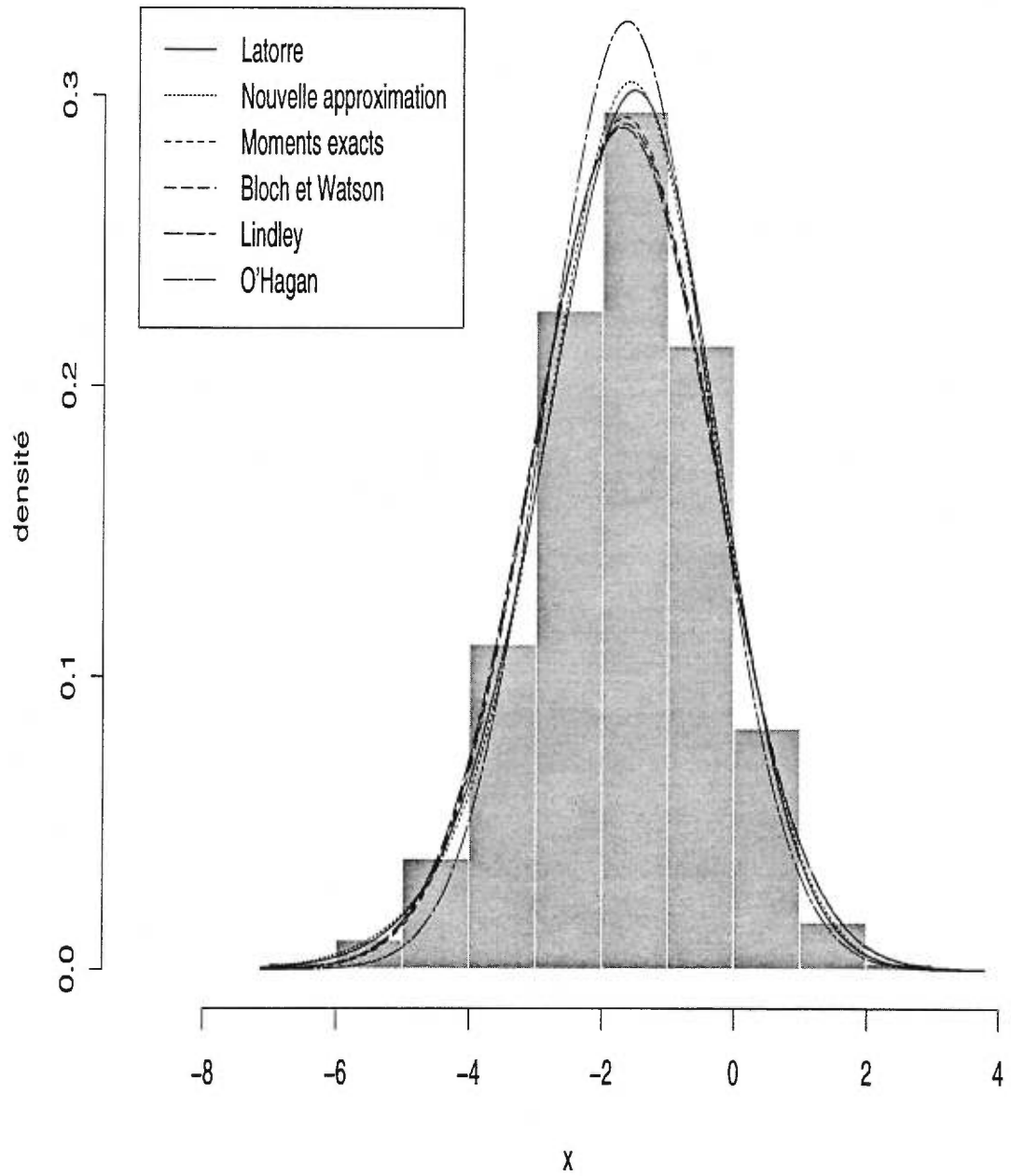


FIGURE 3.5.2. *Approximations de la distribution de ϕ à partir du deuxième jeu de données.*

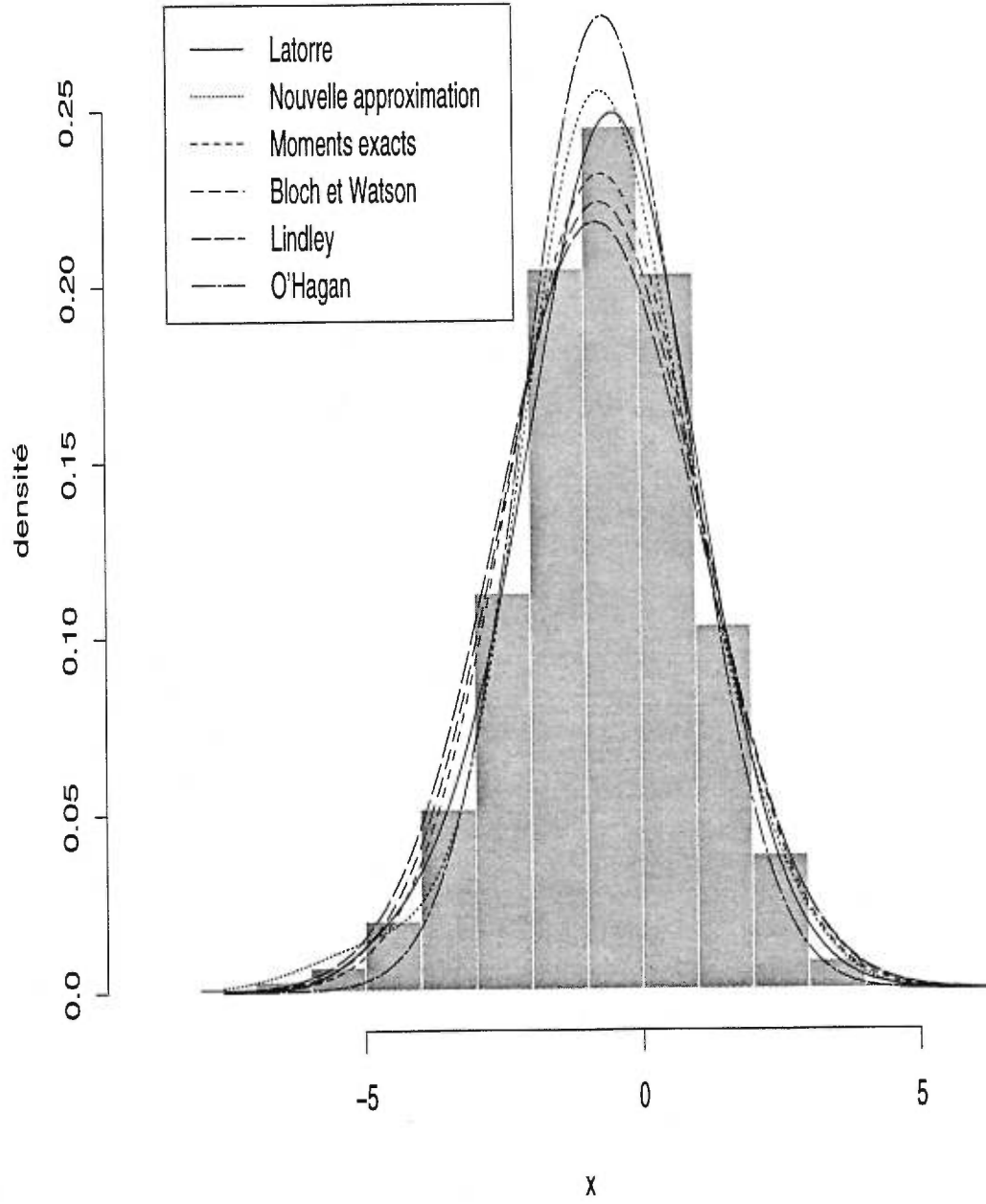


FIGURE 3.5.3. *Approximations de la distribution de ϕ à partir du troisième jeu de données.*

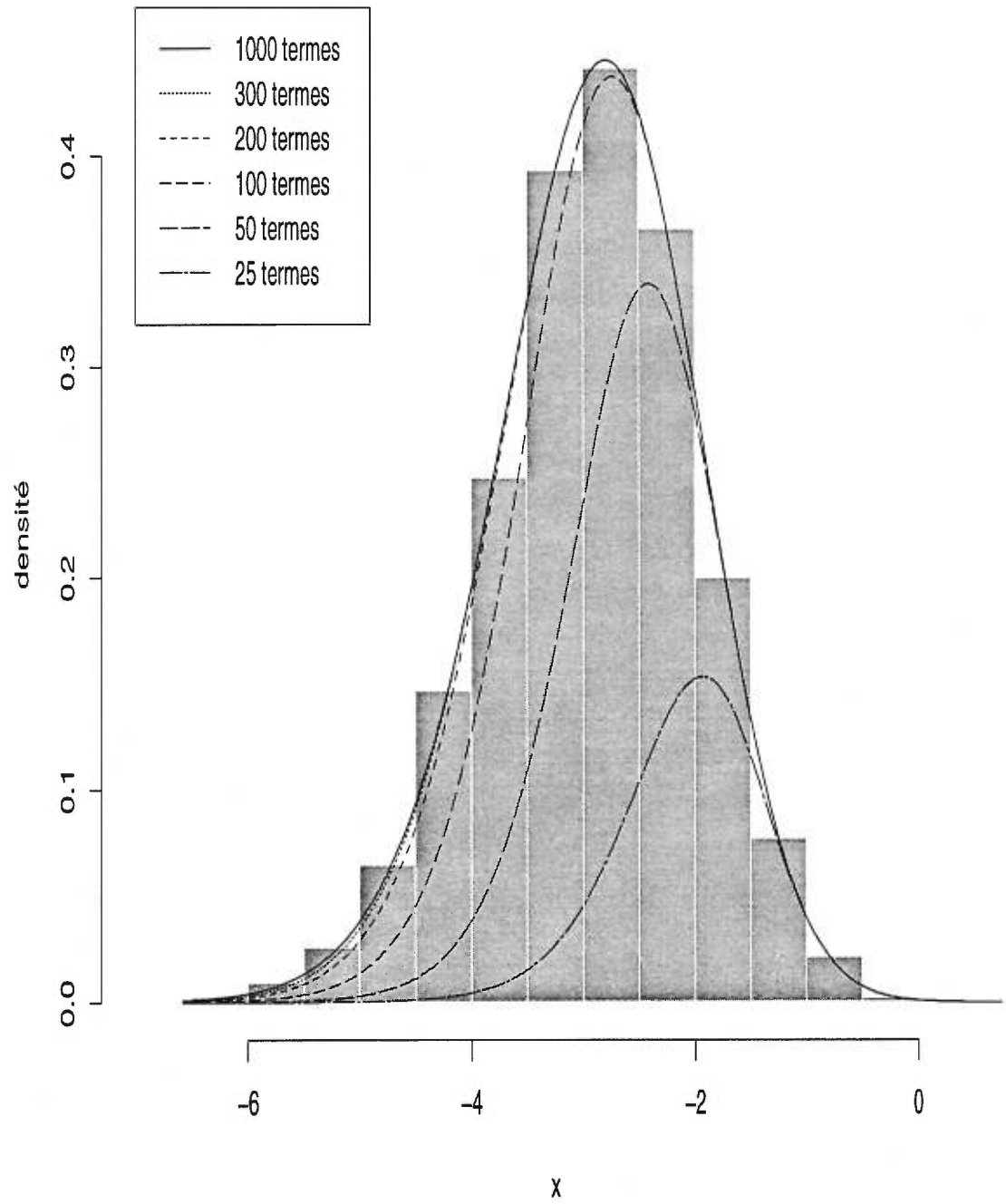


FIGURE 3.5.4. *Convergence de l'expansion de Latorre.*

CONCLUSION

Dans ce mémoire, nous avons étudié les tests d'indépendance de variables aléatoires à deux modalités. Ces différents tests étaient issus de modèles que nous avons répartis en deux grandes catégories, les modèles fréquentistes et les modèles bayésiens. Notre recherche était surtout consacrée à l'étude de la distribution *a posteriori* du logarithme du rapport des cotes sous la modélisation bayésienne usuelle, c'est-à-dire lorsque les observations admettent une loi multinomiale et les probabilités conjointes suivent *a priori* une loi conjuguée de Dirichlet. Premièrement, nous avons montré qu'il est possible de calculer exactement les moments de cette variable aléatoire lorsque les paramètres de la loi de Dirichlet sont tous entiers. Ensuite, nous avons comparé différentes approximations de la distribution de ϕ : l'approximation normale avec les moments exacts ou leurs approximations suggérées par Lindley (1964), Bloch et Watson (1967) ainsi que O'Hagan (1994, chapitre 10), l'expansion de Latorre et une nouvelle approximation trouvée à partir de la fonction caractéristique de ϕ et du théorème d'inversion.

En ce qui concerne l'utilisation de la loi normale comme approximation de la distribution de ϕ , elle a l'avantage d'être simple et de donner de très bon résultats lorsque les paramètres de ϕ sont grands. Par contre, cette approximation est symétrique par rapport à la moyenne, ce qui n'est pas nécessairement vrai pour la densité exacte de ϕ . De plus, cette approximation est inadéquate lorsque les paramètres sont très petits et, contrairement à l'expansion de Latorre et à

notre approximation, nous ne pouvons pas modifier le nombre de termes utilisés en fonction de l'ordre de grandeur de nos paramètres. Pour ce qui est du choix des moments, il semble qu'il est préférable d'utiliser les moments exacts si les paramètres sont entiers sinon, nous devrions alors opter pour les moments approximatifs suggérés par Bloch et Watson (1967), ces approximations étant toujours plus près de la vraie valeur que celles suggérées par Lindley (1964) ou O'Hagan (1994, chapitre 10).

Pour ce qui est des deux autres approximations, notons que le principal désavantage de l'expansion de Latorre est sa très faible vitesse de convergence vers la vraie densité de ϕ . Par contre, cette méthode fournit toujours des résultats satisfaisants lorsque l'utilisateur possède des ressources informatiques suffisantes. Pour ce qui est de la nouvelle approximation, elle fournit des résultats tout aussi satisfaisants que ceux de Latorre et ce avec beaucoup moins de termes. Nous remarquons aussi que le premier terme de cette approximation correspond toujours à l'approximation normale avec moments exacts. Cette approximation donne donc des résultats très adéquats avec peu de termes. Par contre, cette approximation a deux désavantages : comme elle utilise les coefficients $D_k(\underline{a})$, nous devons nous restreindre aux cas où tous les paramètres a_j de la loi de Dirichlet sont entiers. De plus, cette approximation a les mêmes problèmes de convergence que les expansions d'Edgeworth (voir Hall, 1992 chapitre 2), ce qui n'est pas le cas de l'expansion de Latorre. En effet, de petites valeurs de m peuvent donner des résultats plus satisfaisants que si nous avons pris un grand nombre de termes.

Ainsi, il serait intéressant d'étudier les problèmes de convergence de cette approximation. Si nous avons réussi à établir le comportement asymptotique du terme d'erreur lorsque m est fixe, des recherches devraient être entreprises sur le comportement du terme d'erreur lorsque les a_j sont fixés et que m croît.

ANNEXE A: LES PROGRAMMES

```
#Les nombres de bernoulli : ce vecteur contient les 15 premières valeurs
#des nombres bernoulli(k), lorsque k est un nombre positif pair. Ces
#nombres ont été calculés à l'aide du logiciel Mathématique.
bernoulli_c(1/6,-1/30,1/42,-1/30,5/66,-691/2730,7/6,-3617/510,43867/798,
           -174611/330, 854513/138, -236364091/2730, 8553103/6,
           -23749461029/870, 8615841276005/14322)

I_complex(real=0,imaginary=1)

fact_function(n){
#Définition des nombres factoriels.
N_1
if(n==0){N_1}
else{
for(i in 1:n){
N_N*i}}}

pair_function(k,a){
#Cette fonction nous permet de calculer exactemnt les bk(a) lorsque a est
#un entier et que k est pair, cette fonction utilise l'identité de Parseval.
  ilist_1:a
  klist_(ilist)^(-k)
  return(a^(-k)-sum(klist)+(2*pi)^(k)*(-1)^((k/2)+1)*
         bernoulli[k/2]/(2*fact(k)))
}
```

```

impair_function(k,a,b){
#Cette fonction nous permet de calculer exactement les bk(a) lorsque a est
#un entier et que k est impair.
s_0
if(a<b){for (i in a:(b-1)) {s_0+(1/i)^k}}
if(a>b){
for (i in b:(a-1)) {
s_0+(1/i)^k}
s_0-s}
return(s)}

bk_function(k,a1,a2,a3,a4){
#Cette fonction calcule les bk(a) où a est cette fois un vecteur d'entiers de
#dimension quatre.
if (k%%2==0)
{fact(k-1)*(pair(k,a1)+pair(k,a2)+pair(k,a3)+pair(k,a4))}
else
{fact(k-1)*(impair(k,a2,a1)+impair(k,a3,a4))}}

M_function(z,m){
#Cette fonction calcule les Mk(z), il retourne le vecteur (M0(z),...,Mm(z))
#de dimension (m+1). Elle utilise la relation de récurrence de la proposition
#3.3.2.
vect_NULL
vect[1]_exp(-1*(z)^2/2)
vect[2]_-z*exp(-1*(z)^2/2)
for (j in 3:(m+1)){
vect[j]_-z*vect[j-1]-(j-2)*vect[j-2]}
return(vect)}

```

```

Dk_function(a1,a2,a3,a4,m){
#Cette fonction calcule les Dk(a), retourne le vecteur (D0(a),...,Dm(a))
#de dimension (m+1). Elle utilise la relation de récurrence de la proposition
#3.3.3 .
sd_sqrt(bk(2,a1,a2,a3,a4))
bi_NULL
for(l in 1:30)
{
bi[l]_bk(1,a1,a2,a3,a4)
}
Dk_NULL
Dk[1]_(1/sqrt(2*pi))
Dk[2]_ 0
Dk[3]_0
for (j in 3:m){
v_0
for (n in 3:j){
v_v+bi[n]*Dk[j+1-n]*(j*fact(n-1))^-1*sd^-n*(-1)^n}
Dk[j+1]_v}
return(Dk)}

approximation_function(a1,a2,a3,a4,m){
#Calcule la densite de la nouvelle approximation, elle sert à faire les graphiques
#de la densité. Elle retourne une matrice de dimension 2 X 401 où chaque ligne
#correspond au vecteur (pi(x),x) où x est un point se situant à au plus
#quatre écarts-types de la moyenne du logarithme du rapport de cotes.
z_seq(-4,4,.02)
tmp_Dk(a1,a2,a3,a4,m)
mu_bk(1,a1,a2,a3,a4)
sd_sqrt(bk(2,a1,a2,a3,a4))
densite_NULL

```

```

x_NULL
for (i in 1:401) {
densite[i]_(Re(sum(tmp*M(z[i],m))))
densite[i]_densite[i]/sd
x[i]_sd*z[i]+mu
}
matrice_cbind(densite,x)
return(matrice)}

beta_function(a,b){
#Calcul de la fonction bêta, nécessaire pour calculer les expansions de
#Latorre
return(prod(1:max(min(a, b)-1,1))/prod(max(a,b):(a+b-1)))}

produit_function(i, a, b, c){
#Fonction nécessaire pour calculer les expansions hypergéométriques.
p <- 1
for(k in 0:(i - 1)) {
p_p*(a+k)*(b+k)/((c+k)*(k+1))
}
return(p)
}

hyper_function(a,b,c,m){
#Calcul de l'expansion hypergéométrique, nécessaire pour calculer les
#expansions de Latorre.
hyp_NULL
for (i in 1:m)
{hyp[i]_produit(i,a,b,c)}
return(hyp)
}

```

```

latorre_function(a1,a2,a3,a4,m){
#Calcule la densité à partir des m premiers termes de l'expansion de Latorre,
#elle sert à faire les graphiques de la densité. Elle retourne une matrice de
#dimension 2 X 401 où chaque ligne correspond au vecteur (pi(x),x) où x est
#un point se situant à au plus quatre écarts-types de la moyenne du
#logarithme du rapport de cotes.
z_seq(-4,4,.02)
mu_bk(1,a1,a2,a3,a4)
sd_sqrt(bk(2,a1,a2,a3,a4))
tmp.neg_hyper(a1+a3,a1+a2,a1+a2+a3+a4,m)
tmp.pos_hyper(a1+a2,a2+a4,a1+a2+a3+a4,m)
cte_beta(a1+a2,a3+a4)/(beta(a1,a3)*beta(a2,a4))
densite_NULL
x_NULL
for(i in 1:401){
x[i]_z[i]*sd+mu
if (x[i]>=0){
densite[i]_cte*exp(-x[i]*a2)*(1+sum(tmp.pos*(1-exp(-x[i]))^(1:m)))
}
else{
densite[i]_cte*exp(x[i]*a1)*(1+sum(tmp.neg*(1-exp(x[i]))^(1:m)))
}
}
matrice_cbind(densite,x)
return(matrice)}

Rep.app_function(x,a1,a2,a3,a4,m){
#Fonction calculant l'approximation de la fonction de répartition du
#logarithme du rapport de cotes, cette fonction est utilisée pour déterminer
#les régions de crédibilité.

```



```

z_(x-bk(1,a1,a2,a3,a4))/sqrt(bk(2,a1,a2,a3,a4))
tmp_M(z,m-1)
d_Dk(a1,a2,a3,a4,m)[2:(m+1)]
return(pnorm(z)+sum(d*tmp))}

app_function(x,a1,a2,a3,a4,m){
#Cette fonction calcule la densité de la nouvelle approximation, en un
#point x fixé.
z_(x-bk(1,a1,a2,a3,a4))/sqrt(bk(2,a1,a2,a3,a4))
tmp_M(z,m)
d_Dk(a1,a2,a3,a4,m)
return(sum(d*tmp))}

app.inf_function(a,a1,a2,a3,a4,m,alpha){
#Cette fonction retourne la borne inférieure de la région alpha-crédible, à
#partir de la nouvelle approximation. Le premier paramètre représente la
#valeur initiale (centrée et réduite) de l'algorithme de Newton que nous
#utilisons ici.
credi_NULL
credi[1]_sqrt(bk(2,a1,a2,a3,a4))*pnorm(a)+bk(1,a1,a2,a3,a4)
i_2
while(abs(Rep.app(credi[i-1],a1,a2,a3,a4,m)-alpha/2)>10^(-10)){
credi[i]_credi[i-1]-(Rep.app(credi[i-1],a1,a2,a3,a4,m)-alpha/2)/
app(credi[i-1],a1,a2,a3,a4,m)
print(Rep.app(credi[i-1],a1,a2,a3,a4,m)-alpha/2)
i_i+1
}
print(credi)
return(credi[i-1])}

app.sup_function(a,a1,a2,a3,a4,m,alpha){

```

```

#Cette fonction retourne la borne supérieure de la région alpha-crédible, à
#partir de la nouvelle approximation. Le premier paramètre représente la
#valeur initiale (centrée et réduite) de l'algorithme de Newton que nous
#utilisons ici.
credi_NULL
credi[1]_sqrt(bk(2,a1,a2,a3,a4))*pnorm(a)+bk(1,a1,a2,a3,a4)
i_2
while(abs(Rep.app(credi[i-1],a1,a2,a3,a4,m)-(1-alpha/2))>10^(-10)){
credi[i]_credi[i-1]-(Rep.app(credi[i-1],a1,a2,a3,a4,m)-(1-alpha/2))/
      app(credi[i-1],a1,a2,a3,a4,m)
print(Rep.app(credi[i-1],a1,a2,a3,a4,m)-(1-alpha/2))
i_i+1
}
print(credi)
return(credi[i-1])}

Rep.lat_function(x,a1,a2,a3,a4,m){
#Fonction calculant les m premiers termes de la fonction de répartition selon
#l'expansion de Latorre, cette fonction est utilisée pour déterminer les
#régions de crédibilité.
tmp.neg_hyper(a1+a3,a1+a2,a1+a2+a3+a4,m)
tmp.pos_hyper(a1+a2,a2+a4,a1+a2+a3+a4,m)
cte_beta(a1+a2,a3+a4)/(beta(a1,a3)*beta(a2,a4))
tmp.beta.pos_NULL
tmp.beta.neg_NULL
for(i in 1:m){
tmp.beta.neg[i]_beta(a1,i+1)
tmp.beta.pos[i]_beta(a2,i+1)
}
if (x<0){
return(cte*(exp(x*a1)/a1+sum(tmp.neg*tmp.beta.neg*

```

```

        pbeta(exp(x),a1,2:(m+1))))
    }
    else{
    return(1-cte*(exp(-x*a2)/a2+sum(tmp.pos*tmp.beta.pos*
    pbeta(1/exp(x),a2,2:(m+1))))))
    }
}

lat_function(x,a1,a2,a3,a4,m){
#Cette fonction calcule la densité selon les m premiers termes de l'expansion
#de Latorre, pour un point x fixé.
tmp.neg_hyper(a1+a3,a1+a2,a1+a2+a3+a4,m)
tmp.pos_hyper(a1+a2,a2+a4,a1+a2+a3+a4,m)
cte_beta(a1+a2,a3+a4)/(beta(a1,a3)*beta(a2,a4))
if (x>=0) {return(cte*exp(-x*a2)*(1+sum(tmp.pos*(1-exp(-x))^(1:m))))}
else {return(cte*exp(x*a1)*(1+sum(tmp.neg*(1-exp(x))^(1:m))))}
}

lat.inf_function(a,a1,a2,a3,a4,m,alpha){
#Cette fonction retourne la borne inférieure de la région alpha-crédible,
#à partir de l'expansion de Latorre. Le premier paramètre représente la
#valeur initiale (centrée et réduite) de l'algorithme de Newton que nous
#utilisons ici.
credi_NULL
credi[1]_sqrt(bk(2,a1,a2,a3,a4))*pnorm(a)+bk(1,a1,a2,a3,a4)
i_2
while(abs(Rep.lat(credi[i-1],a1,a2,a3,a4,m)-alpha/2)>10^(-10)){
credi[i]_credi[i-1]-(Rep.lat(credi[i-1],a1,a2,a3,a4,m)-alpha/2)/
lat(credi[i-1],a1,a2,a3,a4,m)
print(Rep.lat(credi[i-1],a1,a2,a3,a4,m)-alpha/2)
i_i+1
}
}

```

```

}
return(credi)}

lat.sup_function(a,a1,a2,a3,a4,m,alpha){
#Cette fonction retourne la borne supérieure de la région alpha-crédible,
#à partir de l'expansion de Latorre. Le premier paramètre représente la
#valeur initiale (centrée et réduite) de l'algorithme de Newton que nous
#utilisons ici.
credi_NULL
credi[1]_sqrt(bk(2,a1,a2,a3,a4))*pnorm(a)+bk(1,a1,a2,a3,a4)
i_2
while(abs(Rep.lat(credi[i-1],a1,a2,a3,a4,m)-(1-alpha/2))>10^(-10)){
credi[i]_credi[i-1]-(Rep.lat(credi[i-1],a1,a2,a3,a4,m)-(1-alpha/2))/
lat(credi[i-1],a1,a2,a3,a4,m)
print(Rep.lat(credi[i-1],a1,a2,a3,a4,m)-(1-alpha/2))
i_i+1
}
return(credi)}

```

BIBLIOGRAPHIE

- Abramovitz, M. et Stegun, L.A. (1965). *Handbook of mathematical functions*. Dover, New-York.
- Agresti, Alan (1990). *Categorical data analysis*. John Wiley & sons, Inc., New-York.
- Aitchison, J. et Shen, S.M. (1980). *Logistic-normal distributions: some properties and uses*. *Biometrika*, Vol. 67, 261-272.
- Albert, J.H. (1987). *Bayesian estimation of odds ratios under prior hypotheses of independence and exchangeability*. *Journal of Statistical Computation and Simulation*, Vol. 27, 251-268.
- Albert, J.H. et Gupta A.K. (1982). *Mixtures of Dirichlet distributions and estimation in contingency tables*. *The Annals of Statistics*, Vol. 10, 1261-1268.
- Albert, J.H. et Gupta A.K. (1983a). *Estimation in contingency tables using prior information*. *Journal of the Royal Statistical Society, Series B*, Vol. 45, 60-69.
- Albert, J.H. et Gupta A.K. (1983b). *Bayesian estimation methods for 2×2 contingency tables using mixtures of Dirichlet distributions*. *Journal of the American Statistical Association*, Vol. 78, 708-717.
- Altham, P.M.E. (1969). *Exact Bayesian analysis of a 2×2 contingency table and Fisher's exact significance test*. *Journal of the Royal Statistical Society, Series B*, Vol. 31, 261-269.
- Billingsley, P. (1979). *Probability and measure*. John Wiley & sons, New-York.
- Birch, M.W. (1963). *Maximum likelihood in three-way contingency tables*. *Journal of the Royal Statistical Society, Series B*, Vol. 25, 220-233.
- Bloch, D.A. et Watson, G.S. (1967). *A Bayesian study of the multinomial distribution*. *The Annals of Mathematical Statistics*, Vol. 38, 1423-1435.
- Chin, T., Marine, W., Hall, E., Gravelle, C., et Speers, J. (1961). *The influence of Salk vaccination on the epidemic pattern and the spread of the virus in the community*. *The American Journal of Hygiene*, Vol. 73, 67-94.
- Cochran, W.G. (1954). *Some methods of strengthening the common χ^2 tests*. *Biometrics*, Vol. 10, 417-451.

- Edwards, H.M. (1974). *Riemann's Zeta function*. Samuel Eilenberg and Hyman Bass, New-York.
- Fienberg, S.E. et Holland, P.W. (1970). *Methods for eliminating zero counts in contingency tables*. Random Counts on Models and structures, G.P. Patil ed., Pennsylvanie, 233-260.
- Fienberg, S.E. et Holland, P.W. (1973). *Simultaneous estimation of multinomial cell probabilities*. Journal of the American Statistical Association, Vol. 68, 683-690.
- Fisher, R.A. (1922). *On the interpretation of chi-square from contingency tables, and the calculation of P*. Journal of the Royal Statistical Society, Vol. 85, 87-94.
- Fisher, R.A. (1934). *Statistical methods for research workers*. Oliver and Boyd, Edimbourg.
- Fisher, R.A. (1935). *The design of experiments*. Oliver and Boyd, Edimbourg.
- Gart, J.J. et Zweifel, J.R. (1967). *On the bias of various estimators of the logit and its variance with application to quantal bioassay*. Biometrika, Vol. 54, 181-187.
- Good, I.J. (1976). *On the application of symmetric Dirichlet distributions and their mixtures to contingency tables*. The Annals of Statistics, Vol. 4, 1159-1189.
- Good, I.J. et Crook, J.F. (1987). *The robustness and sensitivity of the mixed-Dirichlet for "independence" in contingency tables*. The Annals of Statistics, Vol. 15, 670-693.
- Goodman, L.A. (1963). *On Plackett's test for contingency table interaction*. Journal of the Royal Statistical Society, Series B, Vol. 25, 179-188.
- Goodman, L.A. (1964). *Simultaneous confidence intervals for cross-product ratios in contingency tables*. Journal of the Royal Statistical Society, Series B, Vol. 26, 86-102.
- Gradshteyn, I.S. et Ryzhik, I.M. (1980). *Table of integrals, series and products (2^{ième} Édition)*. Academic press, Inc. LTD, Londres.
- Haldane, J.B.S. (1955). *The estimation and significance of the logarithm of a ratio of frequencies*. The Annals of Human Genetics, Vol.20, 309-311.
- Hall, P. (1992). *The bootstrap and Edgeworth expansion*. Springer-Verlag, New-York.
- Hashemi, L., Nandram, B. et Goldberg, R. (1997). *Bayesian analysis for a single 2×2 table*. Statistics in Medicine, Vol. 16, 1311-1328.
- Jeffreys, H. (1961). *Theory of Probability (3^{ième} Édition)*. Oxford University Press, Londres.
- Koehler, K. (1986). *Goodness-of-fit for log-linear models in sparse contingency tables*. Journal of the American Statistical Association, Vol. 81, 483-493.
- Koehler, K. et Larntz, K. (1980). *An empirical investigation of goodness-of-fit statistics for sparse multinomials*. Journal of the American Statistical Association, Vol. 75, 336-344.

- Lange, J. (1931). *Crime and destiny*. Allen and Unwin.
- Larntz, K. (1978). *Small-sample comparison of exact levels for chi-squared goodness-of-fit statistics*. Journal of the American Statistical Association, Vol. 73, 253-263.
- Latorre, G. (1982). *The exact posterior distribution of the cross-ratio of a 2×2 contingency table*. Journal of Statistical Computation and Simulation, Vol. 16, 19-24.
- Latorre, G. (1984). *Bayesian inference in 2×2 and $2 \times 2 \times 2$ contingency tables*. Metron, Vol. 42, 169-184.
- Latorre, G. (1985). *Bayesian comparison of two proportions*. Rivista di Statistica Applicata, Vol. 18, 3-11.
- Lindley, D.V. (1964). *The Bayesian analysis of contingency tables*. The Annals of Mathematical Statistics, Vol. 35, 1622-1643.
- Maritz, J.S. (1989). *Empirical Bayes estimation of the log odds ratio in 2×2 contingency tables*. Communications in Statistics-Theory and Methods, Vol. 18, 3215-3233.
- O'Hagan, A. (1994). *Kendall's advanced theory of statistics (Volume 2B)*. Halsted Press, New-York.
- Pearson, K. (1900). *On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*. Philosophy Magazine, Series 5, Vol. 50, 157-175.
- Pearson, K. (1904). *Mathematical contributions to the theory of evolution XIII: On the theory of contingency and its relation to association and normal correlation*. Draper's Co. Research Memoirs, Biometric Series, no. 1.
- Pearson, K. (1922). *On the χ^2 test of goodness of fit*. Biometrika, Vol. 14, 186-191.
- Plackett, R.L. (1962) *A note on interaction in contingency tables*. Communications in Statistics, series A, Vol. 7, 935-952.
- Raiffa, H. et Schlaifer, R. (1961). *Applied statistical decision theory*. Division of Research, Graduate School of Business Administration, Harvard University, Boston.
- Robert, C. (1992). *L'analyse statistique bayésienne*. Economica, Paris.
- Scheid, F. (1988). *Numerical analysis (2^{ième} Édition)*. Schaum's, New-York.
- Spivak, M. (1980). *Calculus (2^{ième} Édition)*. Publish or Perish Inc., Wilmington.
- Tocher, K.D. ((1950). *Extension of the Neyman-Pearson theory of tests to discontinuous variates*. Biometrika, Vol. 37, 130-144.

- Walker, J.S. (1991). *Fast Fourier Transforms*. CRC PRESS, Boca Raton.
- Walley, P. (1997). *Inferences from multinomial data: learning about a bag of marbles*. Journal of the Royal Statistical Society, Series B, Vol. 58, 3-57.
- Wilks, S.S. (1935). *The likelihood test of independance in contingency tables*. The Annals of Mathematical Statistics, Vol. 6, 190-196.
- Wilks, S.S. (1938). *The large-sample distribution of the likelihood ratio for testing composite hypotheses*. The Annals of Mathematical Statistics, Vol. 9, 60-62.
- Wilks, S.S. (1962). *Mathematical Statistics*. Wiley, New-York.
- Yates, F. (1934). *Contingency tables involving small numbers and the χ^2 test*. Journal of the Royal Statistical Society, Supplement, Vol. 1, 217-235.