

Sur l'estimation du rapport de deux paramètres d'intensité
poissonniens et l'estimation par fonctions de masse prédictives

par

Myriam CHABOT

Mémoire présenté au Département de mathématiques
en vue de l'obtention du grade de maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, 31 octobre 2016

Le 31 octobre 2016,

*le jury a accepté le mémoire de Madame Myriam Chabot
dans sa version finale.*

Membres du jury

Professeur Éric Marchand
Directeur de recherche
Département de mathématiques

Professeur Theodoros Nicolieris
Membre externe
Département d'économie
Université d'Athènes

Professeur Taoufik Bouezmarni
Président-rapporteur
Département de mathématiques

SOMMAIRE

Ce mémoire est consacré à l'étude du modèle statistique bivarié duquel sont issues deux variables aléatoires conditionnellement indépendantes de loi de Poisson, dont les taux ne sont pas nécessairement égaux. Tout au long de ce mémoire, l'emphase est mise sur le développement d'un cadre bayésien au problème d'estimation paramétrique pour un tel modèle. Deux thèmes principaux y sont abordés : l'inférence statistique du rapport des deux paramètres d'intensité poissonniens et les densités prédictives. Ces problèmes surviennent notamment dans le contexte d'estimation de l'efficacité d'un vaccin développé par Laurent [Lau12] de même que Laurent et Legrand [LL12], ou encore, par celui d'estimation de l'efficacité d'un traitement contre le cancer par Lindley [Lin02]. Alors que, dans ces articles, aucune contrainte paramétrique n'est imposée sur le rapport des deux taux poissonniens, une partie du mémoire abordera également ces thèmes lorsqu'il y a une contrainte restreignant le domaine du rapport sur l'intervalle $[0, 1]$. Il sera alors possible d'établir des liens avec un article sur les files d'attente d'Armero et Bayarri [AB94].

Mots-clés : Loi de Poisson bivariée, inférence statistique, statistique bayésienne, lois bêta, espace paramétrique contraint, risque, inférence prédictive.

REMERCIEMENTS

Je tiens d'abord à remercier mon directeur de maîtrise, M. Éric Marchand, pour les discussions mathématiques fort enrichissantes de même que pour son soutien et sa grande disponibilité tout au long de ma maîtrise pour m'aider à cheminer à travers mon projet de recherche. Je tiens également à souligner sa grande ouverture et sa compréhension à l'égard des projets que j'ai entrepris ces dernières années, qui seront certes un atout important pour le futur. Je remercie aussi les membres de mon jury.

Je veux également remercier mes collègues du Département de mathématiques pour les échanges mathématiques et statistiques ainsi que pour leur soutien.

Pour l'appréciable soutien financier qui me fut octroyé, merci au Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), au Fonds de recherche du Québec - Nature et technologies (FRQNT) et au Département de mathématiques.

Finalement, un merci tout particulier à mon copain, ma famille et mes amis, qui ont su m'épauler dans mes projets tout au long de ma maîtrise à travers leur écoute, leurs encouragements, leur disponibilité et leur patience.

Myriam Chabot
Sherbrooke, octobre 2016

TABLE DES MATIÈRES

SOMMAIRE	iii
REMERCIEMENTS	iv
TABLE DES MATIÈRES	v
LISTE DES FIGURES	viii
NOTATION	x
INTRODUCTION	1
CHAPITRE 1 — Concepts théoriques	7
1.1 Modélisation du problème	7
1.1.1 Modèle statistique	7
1.1.2 Hypothèses	11
1.1.3 Propriétés	12

1.2	Exemples de méthodes d'inférence statistique	14
1.2.1	Approche fréquentiste versus approche bayésienne	14
1.2.2	Estimation avec l'approche fréquentiste	25
1.2.3	Estimation avec l'approche bayésienne	31
1.3	Inférence prédictive	40
1.3.1	Estimation de densité prédictive	40
1.3.2	Évaluation des estimateurs de densité prédictive	42
CHAPITRE 2 — Répertoire de lois bêta		50
2.1	Fonctions hypergéométriques	51
2.2	Distributions bêta généralisées	57
2.3	Distributions bêta de type II généralisées	63
CHAPITRE 3 — Étude du modèle bivarié sans contrainte paramétrique		69
3.1	Estimateur à vraisemblance maximale modifié	70
3.2	Estimateurs bayésiens	71
3.3	Évaluation et comparaisons du risque des estimateurs	79
3.3.1	Risque de l'estimateur à vraisemblance maximale modifié	81
3.3.2	Risque des estimateurs bayésiens	85
3.3.3	Comparaison des risques des estimateurs	90
3.4	Inférence prédictive	95

3.4.1	Estimateur de Bayes	98
3.4.2	Espérances, variances et covariance	111
CHAPITRE 4 — Étude du modèle bivarié avec contrainte paramétrique		115
4.1	Estimateur à vraisemblance maximale modifié	116
4.2	Estimateurs bayésiens	117
4.3	Évaluation et comparaisons du risque des estimateurs	124
4.3.1	Risque de l'estimateur à vraisemblance maximale modifié avec contrainte paramétrique	124
4.3.2	Risque de l'estimateur bayésien	127
4.3.3	Comparaison des risques des estimateurs	129
4.4	Inférence prédictive	132
4.4.1	Loi de Bailey généralisée	134
4.4.2	Estimateur de Bayes	139
4.4.3	Espérances, variances et covariance	147
CONCLUSION		150
Annexe A		153
Annexe B		155
BIBLIOGRAPHIE		157

LISTE DES FIGURES

1.1	Intervalle de crédibilité de niveau 95 % pour μ si $X \mu \sim \mathcal{P}(\mu)$ à partir de l'observation $x = 17$	33
3.1	Illustration d'un intervalle de crédibilité pour ϕ	78
3.2	Fonction de risque de $\hat{\phi}_{EVMM}(\mathbf{x}) = \frac{x_1+0.5}{x_2+0.5}$	82
3.3	Fonction de risque de l'estimateur $\hat{\phi}_{B1}(\mathbf{x}) = \frac{x_1}{x_2+1}$	88
3.4	Fonction de risque de l'estimateur $\hat{\phi}_{B2}(\mathbf{x}) = \frac{x_1+0.5}{x_2+1.5}$	89
3.5	Quotient des risques de $\hat{\phi}_{EVMM}(\mathbf{x}) = \frac{x_1+0.5}{x_2+0.5}$ et $\hat{\phi}_{B1}(\mathbf{x}) = \frac{x_1}{x_2+1}$	91
3.6	Quotient des risques de $\hat{\phi}_{EVMM}(\mathbf{x}) = \frac{x_1+0.5}{x_2+0.5}$ et $\hat{\phi}_{B2}(\mathbf{x}) = \frac{x_1+0.5}{x_2+1.5}$	92
3.7	Quotient des risques de $\hat{\phi}_{B1}(\mathbf{x}) = \frac{x_1}{x_2+1}$ et $\hat{\phi}_{B2}(\mathbf{x}) = \frac{x_1+0.5}{x_2+1.5}$	94
3.8	Fonction de masse prédictive de Y_1 pour $x_1 = 17$ et $x_2 = 31$	111
3.9	Fonction de masse prédictive de Y_2 pour $x_1 = 17$ et $x_2 = 31$	111
4.1	Illustration d'un intervalle de crédibilité pour ϕ contraint à $[0, 1]$	120
4.2	Graphique de $E[\phi x_1, x_2 = 31]$ pour $a = c = d = 1$ et $b = 0$	123
4.3	Graphique de $E[\phi x_1 = 17, x_2]$ pour $a = c = d = 1$ et $b = 0$	123

4.4	Risque de l'estimateur $\hat{\phi}_{EVMMc}(\mathbf{x}) = \min\left(\frac{x_1+0,5}{x_2+0,5}, 1\right)$	125
4.5	Comparaison des estimateurs $\hat{\phi}_{EVMMc}$ et $\hat{\phi}_{EVMM}$	126
4.6	Risque de l'estimateur $\hat{\phi}_{Bc}(\mathbf{x})$ avec $(a, b, c, d) = (1, 0, 1/2, 1)$	128
4.7	Risque de l'estimateur $\hat{\phi}_{Bc}$ pour différents paramètres (a, b, c, d)	129
4.8	Comparaison des risques de l'estimateur $\hat{\phi}_{Bc}$ et d'estimateurs qui ne tiennent pas compte de la contrainte sur ϕ	130
4.9	Quotient des risques des estimateurs $\hat{\phi}_{Bc}(\mathbf{x})$ et $\hat{\phi}_{EVMMc}(\mathbf{x})$	131
4.10	Fonction de masse prédictive de Y_1 pour $x_1 = 17$ et $x_2 = 31$	146
4.11	Fonction de masse prédictive de Y_2 pour $x_1 = 17$ et $x_2 = 31$	146

NOTATION

$b(\theta, \hat{\theta})$: Biais de l'estimateur $\hat{\theta}$
$Cov(X, Y)$: Covariance de X et Y
$Cov_{\hat{q}_\pi}(X, Y)$: Covariance de X et Y sous la densité prédictive \hat{q}_π
$Skew(X)$: Coefficient d'asymétrie (Skewness)
$\xrightarrow{\mathcal{L}}$: Convergence en loi
$\hat{q}_\pi(y; x)$: Densité prédictive associée à la loi <i>a priori</i> π
$\det(A), A $: Déterminant de la matrice A
$\stackrel{d}{=}$: Égale en distribution
$EQM(\theta, \hat{\theta})$: Erreur quadratique moyenne
Θ	: Espace paramétrique
$E_\theta[g(X)]$: Espérance de la fonction $g(X)$ par rapport à une loi $X \theta$
$E_{\theta x}[p_\theta(x)]$: Espérance de la fonction $p_\theta(x)$ par rapport à la loi de θX
$E_{\hat{q}_\pi}[Y]$: Espérance de Y sous la densité prédictive \hat{q}_π
EVM	: Estimateur du maximum de vraisemblance
$\hat{\theta}$: Estimateur du paramètre θ
$B(\cdot, \cdot)$: Fonction bêta
$\Gamma(\cdot)$: Fonction gamma
$M_X(t)$: Fonction génératrice des moments de X
${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; x)$: Fonction hypergéométrique
${}_1F_1(a; b; x)$: Fonction hypergéométrique confluyente
${}_2F_1(a, b; c; x)$: Fonction hypergéométrique de Gauss
$F_1(a; b, b'; c; x, y)$: Première fonction hypergéométrique d'Appell
$L(\theta, \hat{\theta})$: Fonction de perte
$L_{KL}(q_\theta, \hat{q})$: Fonction de perte de Kullback-Leibler
$R(\theta, \hat{\theta})$: Fonction de risque
$\mathcal{L}(\theta \mathbf{x})$: Fonction de vraisemblance

\perp	: Indépendance
$I_\pi[\mathbf{x}]$: Intervalle de crédibilité
$\log(\cdot)$: Logarithme naturel
$\pi(\cdot x)$: Loi <i>a posteriori</i>
$\pi(\cdot)$: Loi <i>a priori</i>
$Bailey(a, c, d, \rho)$: Loi de Bailey
$GBailey(a, c, d, \rho_1, \rho_2)$: Loi de Bailey généralisée
$\mathcal{B}(a, b)$: Loi bêta
$\mathcal{BG3}(a, b, \lambda)$: Loi bêta généralisée à trois paramètres
$\mathcal{BG4}(a, b, \lambda, \kappa)$: Loi bêta généralisée à quatre paramètres
$\mathcal{B}_2(c, d, \tau)$: Loi bêta de type II
$\mathcal{B}_3(c, d, \kappa, \tau)$: Loi bêta de type III
$Bin(n, p)$: Loi binomiale
$\mathcal{BN}(r, p)$: Loi binomiale négative
$\mathcal{BNP}_\omega(r, p)$: Loi binomiale négative pondérée avec poids $\omega(\cdot)$
$Ber(p)$: Loi de Bernoulli
$Exp(\lambda)$: Loi exponentielle
$\mathcal{G}(a, b)$: Loi gamma d'espérance $\frac{a}{b}$ et de variance $\frac{a}{b^2}$
$\mathcal{HG}(a_1, \dots, a_p, b_1, \dots, b_q, \theta)$: Loi hypergéométrique généralisée
$\mathcal{P}(\lambda)$: Loi de Poisson univariée
$\mathcal{P2}(\lambda_1, \lambda_2)$: Loi de Poisson bivariée issue de deux variables aléatoires indépendantes $\mathcal{P}(\lambda_1)$ et $\mathcal{P}(\lambda_2)$
$I(\boldsymbol{\theta})$: Matrice d'information de Fisher
Md	: Médiane
\propto	: Proportionnel à
$RVMC$: Rapport de vraisemblance monotone croissant
$RVMD$: Rapport de vraisemblance monotone décroissant
$r_\pi(\hat{\theta})$: Risque intégré de Bayes
$(x)_n$: Symbole de Pochhammer (factorielle ascendante)
$Var_\theta(g(X))$: Variance de la fonction $g(X)$ par rapport à une loi $X \theta$
$Var_{\hat{q}_\pi}(Y)$: Variance de Y sous la densité prédictive \hat{q}_π

INTRODUCTION

Étant souvent considérée comme l’analogie discret de la loi normale, la distribution de Poisson est un modèle probabiliste couramment utilisé entre autres à travers les processus de Poisson et pour décrire des situations dans lesquelles on compte un nombre d’événements indépendants rares qui se produisent dans un intervalle de temps fixé, ou encore, dans un certain espace donné. Caractérisée par un paramètre positif λ représentant à la fois l’espérance (le taux) et la variance de la variable aléatoire, cette distribution, que l’on notera $\mathcal{P}(\lambda)$, a pour fonction de masse

$$p(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}\mathbf{1}_{\mathbb{N}}(x), \quad \lambda > 0. \quad (1)$$

Pour plus de détails sur la distribution de Poisson, voir, par exemple, [JKK05] et [Hai67].

En statistique appliquée, cette distribution se manifeste dans de multiples domaines comme la physique, l’agriculture, l’écologie, la biologie, les sciences médicales, la démographie et la finance [Hai67]. En outre, en épidémiologie, la modélisation à partir de la loi de Poisson est courante à travers l’étude de l’incidence de certaines maladies, entre autres influencée par plusieurs facteurs génétiques, biologiques, environnementaux et sociaux [FK95]. En effet, plusieurs études sur des maladies telles que le cancer et les maladies cardiovasculaires font intervenir la loi de Poisson en dénombrant le nombre de patients qui ont développé la maladie étudiée dans une population quelconque sur une période de temps donnée [FK95]. L’estimation ponctuelle ou par intervalles de confiance du taux

d'incidence (λ) dans un modèle poissonnien univarié ne date pas d'hier. Il s'agit d'un problème sur lequel plusieurs se sont penchés et pour lequel il existe plusieurs méthodes et références (pour plus de détails, voir par exemple [Hai67] et [JKK05]).

Dans de tels contextes épidémiologiques, on pourrait également vouloir étudier l'efficacité d'un traitement ou d'un vaccin sur un groupe d'individus. Pour ce faire, on peut alors considérer un modèle statistique bivarié (X_1, X_2) , duquel on obtient des observations x_1 et x_2 , où chacune d'elles est issue respectivement d'une distribution de Poisson $\mathcal{P}(\lambda S)$ et $\mathcal{P}(\mu T)$. Alors que les quantités S et T sont des constantes fixées pouvant représenter, par exemple, des tailles de populations, les taux λ et μ sont, quant à eux, des paramètres inconnus. En reprenant l'exemple décrit par Laurent et par Laurent et Legendre, respectivement dans [Lau12] et [LL12], si l'on veut tester l'efficacité d'un vaccin, les variables aléatoires X_1 et X_2 représentent respectivement le nombre de patients des groupes vacciné et contrôle (non vacciné) qui développent une maladie. Dans [Lin02], il s'agit plutôt de tester l'efficacité d'un traitement. Pour ces deux contextes, il s'avère judicieux de reparamétriser le taux λ en posant $\phi = \lambda/\mu$, ce qui permet entre autres d'obtenir l'efficacité relative du vaccin ou du traitement, donnée par $1 - \phi$. Il devient alors évident que, pour $S = T$, si $\phi < 1$, cela laisse présager un vaccin ou un traitement efficace diminuant le taux d'incidence μ , contrairement au cas où $\phi > 1$, qui témoigne plutôt d'une détérioration par rapport au groupe contrôle. Enfin, si $\phi = 1$, alors il s'agit d'une invariabilité de la situation. Ainsi, l'intérêt porté à ce modèle ainsi présenté réside davantage en l'estimation du paramètre ϕ qu'en l'estimation des paramètres initiaux λ et μ .

L'inférence statistique réalisée sur le rapport de deux paramètres d'intensité poissonniens est étudiée depuis plusieurs décennies, tant pour l'estimation ponctuelle que pour l'estimation par intervalles de confiance. Mentionnons notamment, parmi plusieurs autres, les travaux de Chapman [Cha52], Birnbaum [Bir53], Bol'shev [Bol62], Agresti et Coull [AC98],

Cousins [Cou98] et Price et Bonnett [PB00].

D'ailleurs, dans plusieurs contextes, on retrouve une restriction sur l'espace paramétrique. Depuis le début des années 1950, les problèmes d'estimation en présence d'une contrainte paramétrique retiennent l'attention de plusieurs statisticiens, dont Brunk et van Eeden, qui furent parmi les premiers. Parmi de bonnes références au sujet de l'inférence dans les espaces paramétriques contraints, on trouve notamment les livres de Robertson, Wright et Dykstra [RWD88], et, plus récemment, de van Eeden [vE06], de même que l'article de revue de Marchand et Strawderman [MS04].

Ainsi, en continuité avec la modélisation poissonnienne bivariée présentée précédemment, si on suppose qu'un vaccin ou qu'un traitement ne peut pas nuire à une certaine population à l'égard d'une maladie, alors il semble tout à fait raisonnable de supposer, *a priori*, que $\mathbb{P}(\phi > 1) = 0$. De cette manière, le support de ϕ qui en tient compte devrait se restreindre à l'intervalle $[0, 1]$. Cette contrainte, équivalente à supposer que $\lambda \leq \mu$, impose alors un ordre entre ces paramètres d'intensité poissonniens. Qu'advient-il alors des estimateurs de ϕ qui ne considèrent pas cette restriction sur le support ? Comment peut-on s'y adapter ?

L'estimation des paramètres d'intensité poissonniens sous contrainte d'ordre a notamment été étudié par Kushary et Cohen [KC91]. Dans cet article, ils considèrent le modèle tel que $X_1|\lambda_1 \sim \mathcal{P}(\lambda_1)$ et $X_2|\lambda_2 \sim \mathcal{P}(\lambda_2)$, où $X_1|\lambda_1$ et $X_2|\lambda_2$ sont indépendantes et les paramètres λ_1 et λ_2 satisfont à la contrainte d'ordre $0 \leq \lambda_1 \leq \lambda_2 < \infty$. Le principal sujet abordé concerne l'admissibilité de certains estimateurs de λ_1 et λ_2 sous la fonction de perte quadratique de même que, pour le problème d'estimation simultanée de λ_1 et λ_2 , l'inadmissibilité de l'estimateur Bayes généralisé, pour une loi *a priori* uniforme, sous les fonctions de perte $L_0(\hat{\boldsymbol{\lambda}}, \boldsymbol{\lambda}) = (\hat{\lambda}_1 - \lambda_1)^2 + (\hat{\lambda}_2 - \lambda_2)^2$ et $L_1(\hat{\boldsymbol{\lambda}}, \boldsymbol{\lambda}) = \frac{(\hat{\lambda}_1 - \lambda_1)^2}{\lambda_1} + \frac{(\hat{\lambda}_2 - \lambda_2)^2}{\lambda_2}$.

L'étude du rapport de deux paramètres d'intensité poissonniens survient également dans

la théorie des files d'attente. En effet, pour le système classique $M|M|1$ avec queue infinie, où le nombre d'arrivées est un processus de Poisson avec paramètre d'intensité λ et le temps de service de l'unique centre de service se distribue selon une loi exponentielle de paramètre μ (définie telle que l'espérance est $1/\mu$), le nombre d'individus dans le système a pour distribution limite une loi géométrique de paramètre λ/μ lorsque $\lambda < \mu$. Ainsi, cette restriction revient à imposer, dans un cadre bayésien, une loi *a priori* définie sur l'intervalle $(0, 1)$ pour le paramètre $\phi = \lambda/\mu$. Dans [AB94], Armero et Bayarri ont analysé ce problème d'inférence statistique selon une approche bayésienne, notamment en y étudiant le choix d'une famille de distributions *a priori* appropriée selon la contrainte paramétrique $0 < \phi < 1$.

D'autre part, en plus du problème d'inférence pour le ratio de deux paramètres d'intensité poissonniens, une autre question d'intérêt survient : « Comment peut-on prédire le comportement de futures variables Y_1 et Y_2 à partir d'observations disponibles pour X_1 et X_2 ? » N'étant pas née d'hier, l'estimation par densités prédictives se retrouve dans des travaux datant d'aussi loin que 1931, élaborés par Jeffreys [Jef31]. Elle fut également étudiée par Aitchison [Ait75] ainsi que Aitchison et Dunsmore [AD75] quelques décennies plus tard. Toutefois, ce n'est qu'au début du 21^e siècle que l'estimation par densités prédictives suscita une curiosité renaissante des statisticiens, qui fut notamment ranimée par des liens établis avec l'estimation de Stein. Mentionnons, par exemple, les articles de Komaki [Kom01], George, Liang et Xu [GLX06], Brown, George et Xu [BGX08] et Kato [Kat09].

Or, l'inférence prédictive pour le modèle bivarié poissonnien mérite également une attention particulière. Des résultats en lien avec cet objectif sont notamment présentés par Laurent dans [Lau12] lorsqu'il n'y a aucune contrainte sur le paramètre ϕ . D'un autre côté, lorsque ϕ est restreint à l'intervalle $(0, 1)$, on peut notamment établir des liens avec l'article d'Armero et Bayarri [AB94]. Pour les files d'attente, l'intérêt d'estimation pré-

dictive réside entre autres en la prédiction de quantités telles que le nombre de clients au total dans le système à tout instant ainsi que le temps total qu'un client y demeure, et ce, à partir d'observations passées. Bien que ce contexte d'inférence prédictive soit différent de celui qui est étudié dans ce mémoire, leur démarche s'avère ici intéressante dans la mesure où ils ont dû discuter d'un choix de lois *a priori* appropriées en lien avec leur objectif d'estimation prédictive dans un contexte de contrainte paramétrique.

Ce mémoire se divise en quatre chapitres qui visent essentiellement à présenter et à développer un cadre bayésien au problème d'estimation de paramètres d'intensité poissonniens et d'estimation de densités prédictives. Dans le Chapitre 1, on y décrit tout d'abord le modèle de Poisson bivarié employé tout au long du mémoire de même que les hypothèses qui le gouvernent. Certains résultats en lien avec ce modèle seront exposés. Des propriétés intéressantes permettront notamment d'aborder le problème d'inférence statistique sous différents angles. Toujours dans cette optique d'estimation paramétrique, quelques rappels seront faits sur des méthodes d'estimation, à la fois selon les approches fréquentiste et bayésienne, et sur l'évaluation des estimateurs qui en découlent. Enfin, la dernière partie de ce chapitre sera consacrée à la théorie entourant l'inférence prédictive.

Dans le Chapitre 2, un répertoire des lois bêta de type I et de type II et leur généralisation seront présentés. Alors que ces lois apparaîtront dans les chapitres suivants, l'importance de ce chapitre réside surtout dans la nécessité de bien définir et comprendre la structure de ces lois et les notations qui seront utilisés ultérieurement. Dans la littérature, le manque d'uniformité et l'appellation changeante de ces lois peuvent parfois porter à confusion d'un auteur à l'autre. Ainsi, l'objectif de ce chapitre consiste à bien comprendre leur genèse, leurs propriétés, les liens entre elles de même qu'à clarifier la notation et l'appellation qui seront utilisées pour la suite du mémoire. Alors que l'approche bayésienne est privilégiée dans ce mémoire, de nouvelles représentations pour les lois bêta seront proposées.

Le Chapitre 3 sera dédié à l'inférence statistique pour deux populations poissonniennes

sans restriction paramétrique. Il aura pour objectif de présenter quelques estimateurs, issus des approches fréquentiste et bayésienne, qui se baseront entre autres sur les développements effectués par Lindley [Lin02], Laurent [Lau12] et Laurent et Legrand [LL12]. De plus, ces estimateurs seront évalués et comparés selon le risque fréquentiste, afin de voir si un estimateur bayésien peut tout même être performant selon un critère fréquentiste. Dans ce cas, on verra qu'un estimateur bayésien s'avère, dans certains cas, un meilleur choix qu'un estimateur issu d'une approche fréquentiste. Puis, la dernière partie de ce chapitre portera sur les densités prédictives. Cette analyse fera intervenir la loi discrète bivariée de Bailey (en reprenant l'appellation utilisée par Laurent [Lau12]).

Finalement, le Chapitre 4 sera à l'image du Chapitre 3, mais dans le cas où l'espace paramétrique est soumis à la contrainte $\lambda \leq \mu$, ou, de manière équivalente, $\phi \in [0, 1]$. On pourra notamment faire des liens entre nos résultats et ceux présentés dans l'article d'Armero et Bayarri [AB94] dans un contexte de files d'attente. En plus, on montrera que la performance des estimateurs, selon le risque fréquentiste, peut être améliorée lorsqu'on tient compte de la contrainte lors du processus d'inférence. Enfin, pour la partie liée aux densités prédictives, il sera notamment question du développement de nouvelles distributions discrètes bivariées.

CHAPITRE 1

Concepts théoriques

1.1 Modélisation du problème

1.1.1 Modèle statistique

Considérons le modèle statistique bivarié (X_1, X_2) décrivant deux observations de comptage, où chacune d'elles est issue respectivement d'une distribution de Poisson $\mathcal{P}(\lambda S)$ et $\mathcal{P}(\mu T)$. Les taux λ et μ sont inconnus, contrairement aux quantités S et T , qui sont des constantes pouvant représenter, par exemple, des tailles de population.

De nombreux contextes font intervenir la loi de Poisson (voir, par exemple, [Hai67]). De plus, il existe un grand nombre de situations où l'utilisation d'une approximation poissonnienne est justifiée. En effet, une des plus connues est certainement l'approximation de la loi binomiale par la loi de Poisson, qui sera illustrée dans l'Exemple 1.4. Pour ce faire, nous aurons besoin des définitions et des résultats suivants.

Définition 1.1. *Soit $\{X_n\}_{n \geq 1}$ une suite de variables aléatoires. On dit que X_n converge*

en loi vers X , noté $X_n \xrightarrow{\mathcal{L}} X$, si

$$\lim_{n \rightarrow \infty} F_{X_n}(t) = F_X(t),$$

pour tout point de continuité t de F_X .

Lemme 1.2. Soit $\{a_n\}_{n \geq 1}$ une suite de nombres tels que $\lim_{n \rightarrow \infty} a_n = a$. Alors

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a_n}{n}\right)^n = e^a. \quad (1.1)$$

Démonstration. On commencera tout d'abord par montrer que pour tout $a \in \mathbb{R}$, on a

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a}{n}\right)^n = e^a.$$

En prenant le logarithme de l'expression et en utilisant la règle de l'Hospital, on a

$$\begin{aligned} \lim_{n \rightarrow \infty} n \ln \left(1 + \frac{a}{n}\right) &= \lim_{n \rightarrow \infty} \frac{\ln(1 + a/n)}{n^{-1}} \\ &= \lim_{n \rightarrow \infty} \left(\frac{-a/n^2}{(1 + a/n)} \div -n^{-2} \right) \\ &= \lim_{n \rightarrow \infty} \frac{na}{n+a} \\ &= a, \end{aligned}$$

d'où $\lim_{n \rightarrow \infty} \left(1 + \frac{a}{n}\right)^n = e^a$. Il reste à montrer que ce résultat est vrai pour toute suite de nombres $a_n \rightarrow a$.

Soit $\epsilon > 0$. Comme $\lim_{n \rightarrow \infty} a_n = a$, il existe N_ϵ tel que pour tout $n > N_\epsilon$, $|a_n - a| < \epsilon$, c'est-à-dire $a - \epsilon < a_n < a + \epsilon$. Ainsi, pour tout $n > N_\epsilon$,

$$\begin{aligned} a - \epsilon < a_n < a + \epsilon &\implies \left(1 + \frac{a - \epsilon}{n}\right)^n < \left(1 + \frac{a_n}{n}\right)^n < \left(1 + \frac{a + \epsilon}{n}\right)^n \\ &\implies e^{a-\epsilon} < \lim_{n \rightarrow \infty} \left(1 + \frac{a_n}{n}\right)^n < e^{a+\epsilon}, \end{aligned}$$

et donc, on en conclut que $\lim_{n \rightarrow \infty} \left(1 + \frac{a_n}{n}\right)^n = e^a$. □

Théorème 1.3. [CB02, Theorem 2.3.12] Soit $\{X_n\}_{n \geq 1}$ une suite de variables aléatoires telle que X_n a pour fonction génératrice des moments $M_{X_n}(t)$ pour tout $n \geq 1$. Supposons que

$$\lim_{n \rightarrow \infty} M_{X_n}(t) = M_X(t),$$

dans un voisinage de 0, et que $M_X(t)$ est une fonction génératrice des moments. Alors $X_n \xrightarrow{L} X$.

À partir des résultats précédents, on peut alors montrer l'approximation d'une loi binomiale $\mathcal{B}in(n, p)$ par une loi de Poisson, lorsque n est suffisamment grand et np suffisamment petit.

Exemple 1.4. [CB02, Généralisation de l'Exemple 2.3.13] Soit $\{X_n\}_{n \geq 1}$ une suite de variables aléatoires telles que pour tout $n \geq 1$, $X_n \sim \mathcal{B}in(n, p_n)$. On veut montrer que cette suite de variables aléatoires converge en loi vers une distribution de Poisson. Pour tout $n \geq 1$, on a

$$M_{X_n}(t) = E[e^{tX_n}] = \sum_{x=0}^n \binom{n}{x} (e^t p_n)^x (1 - p_n)^{n-x} = (p_n e^t + (1 - p_n))^n,$$

où la dernière égalité découle de la formule du binôme de Newton, soit $(a + b)^n = \sum_{i=0}^n \binom{n}{i} a^i b^{n-i}$, où $a, b \in \mathbb{R}$. En posant $p_n = \lambda_n/n$, si $\lim_{n \rightarrow \infty} \lambda_n = \lambda > 0$, on obtient

$$\lim_{n \rightarrow \infty} M_{X_n}(t) = \lim_{n \rightarrow \infty} \left(\frac{\lambda_n e^t}{n} + \left(1 - \frac{\lambda_n}{n} \right) \right)^n = \lim_{n \rightarrow \infty} \left(1 + \frac{\lambda_n (e^t - 1)}{n} \right)^n = e^{\lambda(e^t - 1)},$$

où cette dernière égalité suit du Lemme 1.2. Comme il s'agit de la fonction génératrice des moments d'une loi de Poisson avec espérance λ , il en découle du Théorème 1.3 que $X_n \xrightarrow{L} X$, où $X \sim \mathcal{P}(\lambda)$.

Plusieurs se sont penchés sur ce problème d'approximation par la loi de Poisson, par exemple, Barbour, Holst et Janson [BHJ92] de même que Arratia, Goldstein et Gordon [AGG90]. Une généralisation simple du résultat illustré à l'Exemple 1.4 est la suivante.

Théorème 1.5. [AGG90] Soient $\{X_{k,n}\}_{1 \leq k \leq n}$ un ensemble de variables aléatoires indépendantes telles que $X_{k,n} \sim \mathcal{Ber}(p_{k,n})$ pour $k \in \{1, \dots, n\}$, $\lambda_n = \sum_{k=1}^n p_{k,n}$ et $S_n = \sum_{k=1}^n X_{k,n}$. Si $n \rightarrow \infty$, $\lim_{n \rightarrow \infty} \lambda_n = \lambda$ et $\lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} p_{k,n} = 0$, alors S_n converge en loi vers une loi de Poisson de moyenne λ .

En somme, si on considère le modèle binomial tel que $X_1|\lambda \sim \mathcal{Bin}(S, \lambda)$ et $X_2|\mu \sim \mathcal{Bin}(T, \mu)$, avec S et T suffisamment grands et λ et μ petits, alors il est justifié de travailler avec les approximations poissonniennes $X_1|\lambda \stackrel{approx.}{\sim} \mathcal{P}(\lambda S)$ et $X_2|\mu \stackrel{approx.}{\sim} \mathcal{P}(\mu T)$, d'où le modèle initial proposé. En plus de la loi binomiale, d'autres distributions peuvent être approximées par la loi de Poisson, comme la loi hypergéométrique. Ainsi, l'étude du modèle poissonnien se présente comme un choix intéressant.

Rappelons que, comme mentionné plus tôt en introduction, le modèle $X_1|\lambda \sim \mathcal{P}(\lambda S)$ et $X_2|\mu \sim \mathcal{P}(\mu T)$ peut s'avérer un choix judicieux lorsque l'on veut comparer un phénomène se manifestant dans deux populations indépendantes. Cette modélisation a été étudiée par Laurent [Lau12], Laurent et Legrand [LL12] et par Lindley [Lin02] dans un contexte où l'on désire tester l'efficacité d'un vaccin ou d'un traitement. Ainsi, X_1 et X_2 représentent respectivement le nombre de patients des groupes vacciné (ou traité) et contrôle (non vacciné ou non traité). Dans ces études, afin de tester l'efficacité du vaccin (ou du traitement), il est plus intéressant de considérer le rapport $\phi = \lambda/\mu$ que les paramètres λ et μ en soi. Lindley discute d'ailleurs du choix de la reparamétrisation multiplicative plutôt qu'additive dans [Lin02].

C'est ainsi que nous procéderons pour étudier ce modèle bivarié, qui s'écrira maintenant de la manière suivante :

$$\begin{cases} X_1|\mu, \phi \sim \mathcal{P}(\mu\phi S) \\ X_2|\mu \sim \mathcal{P}(\mu T). \end{cases}$$

1.1.2 Hypothèses

En plus de ce choix de reparamétrisation, quelques hypothèses seront imposées à ce modèle, tout comme l'ont fait Laurent [Lau12], Laurent et Legrand [LL12] ainsi que Lindley [Lin02].

Tout d'abord, nous supposons que les variables aléatoires X_1 et X_2 sont indépendantes conditionnellement aux paramètres μ et ϕ . La fonction de masse conjointe de ces variables, toujours conditionnellement à μ et à ϕ , est alors donnée par :

$$p(x_1, x_2 | \mu, \phi) = \frac{S^{x_1} T^{x_2}}{x_1! x_2!} e^{-\mu(T+\phi S)} \mu^{x_1+x_2} \phi^{x_1} \mathbb{1}_{\mathbb{N} \times \mathbb{N}}(x_1, x_2). \quad (1.2)$$

De ce fait, le vecteur aléatoire $(X_1, X_2) | \mu, \phi$ se distribue selon une loi de Poisson bivariée, qui résulte ici du produit de la fonction de masse de deux lois de Poisson univariées indépendantes. Ainsi, dans le cas où deux variables aléatoires X_1 et X_2 sont indépendantes et de distribution marginale respective $\mathcal{P}(\lambda_1)$ et $\mathcal{P}(\lambda_2)$, on notera la fonction de masse conjointe $\mathcal{P}2(\lambda_1, \lambda_2)$ qui s'exprimera comme suit :

$$p(x_1, x_2) = \frac{1}{x_1! x_2!} e^{-(\lambda_1 + \lambda_2)} \lambda_1^{x_1} \lambda_2^{x_2} \mathbb{1}_{\mathbb{N} \times \mathbb{N}}(x_1, x_2). \quad (1.3)$$

Donc dans le cadre de notre modèle bivarié, on a $(X_1, X_2) | \mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$.

Puis, on supposera que les paramètres μ et ϕ sont indépendants, de telle sorte que leur densité conjointe puisse s'écrire de cette façon : $\pi(\mu, \phi) = \pi(\mu)\pi(\phi)$. Ceci a pour effet notamment de simplifier les calculs ultérieurs.

Enfin, d'autres hypothèses seront ajoutées dans les prochains chapitres, notamment sur le choix de lois *a priori*.

1.1.3 Propriétés

Dans un papier paru en 1964, Holgate a construit une classe de distributions de Poisson bivariées plus générale que celle utilisée ici, basée sur une décomposition en somme de variables aléatoires indépendantes de loi de Poisson univariée [Hol64]. Une description détaillée des distributions de Poisson multivariées se retrouve notamment dans [JKB97]. Nous énoncerons ici quelques propriétés liées à notre modèle bivarié.

La première propriété concerne la fonction génératrice des moments. La preuve de cette propriété suit directement de la fonction génératrice des moments d'une loi de Poisson univariée et de l'indépendance de X_1 et X_2 du modèle ici choisi.

Propriété 1.6. *Soit $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$. Alors la fonction génératrice des moments de \mathbf{X} est*

$$M_{\mathbf{X}}(t_1, t_2) = e^{\mu(S(e^{t_1}-1)+T(e^{t_2}-1))}. \quad (1.4)$$

Afin d'estimer le paramètre d'intérêt ϕ , une technique qui a souvent été utilisée dans la littérature résulte d'une caractéristique intéressante du vecteur aléatoire (X_1, X_2) en lien avec la distribution qui découle du conditionnement d'une des variables par rapport à la somme des deux. Pour montrer cette propriété, nous aurons besoin du lemme suivant.

Lemme 1.7. *Soient W_1, \dots, W_n des variables aléatoires indépendantes de loi de Poisson de paramètres respectifs $\lambda_1, \dots, \lambda_n$. Si $W = \sum_{i=1}^n W_i$, alors*

$$W \sim \mathcal{P}\left(\sum_{i=1}^n \lambda_i\right).$$

Démonstration. On a

$$M_W(t) = E[e^{tW}] = E\left[e^{t \sum_{i=1}^n W_i}\right] = \prod_{i=1}^n E[e^{tW_i}] = \prod_{i=1}^n e^{\lambda_i(e^t-1)} = e^{\sum_{i=1}^n \lambda_i(e^t-1)},$$

où le passage au produit des espérances s'explique par l'indépendance des W_i . □

Proposition 1.8. Soit $(X_1, X_2) \sim \mathcal{P}2(\lambda_1, \lambda_2)$. Alors $X_1|X_1 + X_2 = n \sim \mathcal{B}in\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$.

Démonstration. Du Lemme 1.7, on sait que $X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2)$. Ainsi, pour $n \in \mathbb{N}$ et $x_1 \in \{1, \dots, n\}$, on a

$$\begin{aligned} p(X_1 = x_1|X_1 + X_2 = n) &= \frac{p(X_1 = x_1)p(X_2 = n - X_1|X_1 = x_1)}{p(X_1 + X_2 = n)} \\ &= \frac{p(X_1 = x_1)p(X_2 = n - x_1)}{p(X_1 + X_2 = n)} \\ &= \frac{e^{-\lambda_1} \lambda_1^{x_1}}{x_1!} \frac{e^{-\lambda_2} \lambda_2^{n-x_1}}{(n-x_1)!} \frac{n!}{e^{-(\lambda_1+\lambda_2)}(\lambda_1 + \lambda_2)^n} \\ &= \binom{n}{x_1} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^{x_1} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-x_1}. \quad \square \end{aligned}$$

En appliquant la Proposition 1.8 à notre modèle, on peut alors conclure que la distribution de $X_1|X_1 + X_2 = n$, conditionnellement à μ et à ϕ , est une loi binomiale $\mathcal{B}in\left(n, \frac{\phi S}{\phi S + T}\right)$ et que celle de $X_2|X_1 + X_2 = n$, conditionnellement à μ et ϕ , est une loi binomiale $\mathcal{B}in\left(n, \frac{T}{\phi S + T}\right)$. En posant $p = \frac{\phi S}{\phi S + T}$, qui ne dépend pas du paramètre μ , on a une fonction injective qui relie ϕ au paramètre de proportion p de la loi conditionnelle binomiale. Ainsi, un estimateur plausible \hat{p} de p pour le modèle binomial $\mathcal{B}in(n, p)$ nous fournit l'estimation $\hat{\phi}$ de ϕ suivant :

$$\hat{\phi} = \frac{T}{S} \left(\frac{\hat{p}}{1 - \hat{p}} \right). \quad (1.5)$$

Tel que présenté par Price et Bonnet [PB00], plusieurs méthodes ont été développées et explorées afin d'estimer le paramètre p du modèle binomial $\mathcal{B}in(n, p)$. En effet, mentionnons notamment les travaux de Clopper et Pearson [CP34], Blyth et Still [BS83], Blyth [Bly86], Agresti et Coull [AC98] de même que Cousin, Hymes et Tucker [CHT10].

1.2 Exemples de méthodes d'inférence statistique

L'intérêt porté au modèle statistique décrit précédemment réside dans l'estimation des paramètres μ et ϕ . Lorsqu'il est question d'estimation paramétrique, deux grands axes de pensée fondamentalement différents rivalisent : l'approche fréquentiste et l'approche bayésienne. D'une part, dans cette section, il importe d'en décrire leurs principales spécificités. D'autre part, quelques méthodes d'estimation ponctuelle et par intervalle de confiance, qui sont explicitées entre autres dans [CB02], [Ber85], [KK92] et [Dod08], seront décrites afin de nous guider pour les prochains chapitres vers des candidats raisonnables $\hat{\mu}$ et $\hat{\phi}$. De plus, on verra comment comparer la performance des estimateurs qui en découleront selon des critères prédéfinis.

1.2.1 Approche fréquentiste versus approche bayésienne

Approche fréquentiste

L'école de pensée fréquentiste suggère que l'estimation d'un paramètre θ , qui est inconnu mais caractérisé par une valeur fixe, s'effectue à l'aide d'un échantillon $\mathbf{X} = (X_1, \dots, X_n)$ associé au paramètre d'intérêt. L'estimé pour θ résulte des n valeurs observées $\mathbf{x} = (x_1, \dots, x_n)$ et les méthodes basées sur la vraisemblance sont omniprésentes et jouissent de bonnes propriétés asymptotiques. Pour plus de détails, voir, par exemple, [CB02].

Approche bayésienne

Contrairement à l'approche fréquentiste, l'approche bayésienne, abordée par exemple dans [Ber85], étudie les paramètres à estimer vus comme variables aléatoires. Les valeurs possibles des paramètres sont décrites par une fonction de masse ou une fonction de densité, dite *loi a priori* et notée $\pi(\cdot)$. Non basé sur les données recueillies pour l'étude, le

choix des lois *a priori* peut être motivé par des considérations calculatoires, intuitives ou par des expériences réalisées dans le passé. Ainsi, même si aucune information préalable n'est disponible en lien avec le paramètre à estimer, cette approche peut être employée. Dans ce cas, on peut avoir recours à une loi *a priori* dite *non informative*, et il arrivera fréquemment qu'une telle loi ne sera pas une fonction de densité ou une fonction de masse. De plus, l'appellation *loi a priori impropre* sera utilisée pour une loi *a priori* qui, lorsqu'intégrée sur son domaine, divergera à l'infini. Difficilement interprétable, le choix d'une telle loi peut tout de même, dans plusieurs cas, être utilisée pour mener vers des estimateurs raisonnables.

Exemple 1.9. *Soit θ un paramètre à estimer, et $\Theta = \mathbb{R}_+$, l'espace paramétrique. En 1812, Laplace a utilisée une méthode basée sur la loi uniforme [Lap12]. Cette méthode avait également été introduite lors de travaux de Thomas Bayes parus en 1763 pour estimer le paramètre p du modèle binomial $\text{Bin}(n, p)$. Pour obtenir une loi non informative, l'approche probablement la plus intuitive est de ne favoriser aucune valeur sur Θ et ainsi leur attribuer un poids égal. Cela conduit alors vers la loi uniforme impropre $\pi(\theta) = \mathbb{1}_{(0, \infty)}(\theta)$. Cette loi peut être vue comme la loi limite d'une loi gamma $\mathcal{G}(a, b)$ lorsque $a = 1$ et $b \rightarrow 0$. Par contre, un inconvénient considérable d'une telle approche est le manque d'invariance sous reparamétrisation, tel qu'illustré dans [Ber85, Chapitre 3, Example 4] avec $\eta = e^\theta$.*

À la suite des critiques reçues sur la méthode naïve de Laplace, certains statisticiens se penchèrent sur une façon plus formelle pour élaborer une méthode de construction de lois *a priori* non informatives en considérant la reparamétrisation possible des paramètres. Ce thème fut entre autres étudié par Jeffreys publia une méthode encore très utilisée aujourd'hui [Jef61]. Celle-ci est basée sur la matrice d'information de Fisher $I(\theta)$, qui

s'écrit, dans le cas entre autres de familles exponentielles, comme suit :

$$I(\theta) = -E_{\theta} \left[\frac{\partial^2}{\partial \theta^2} \log f(X|\theta) \right], \quad (1.6)$$

où l'observé est X de densité $f(x|\theta)$. Ainsi, dans le cas unidimensionnel où il n'y a qu'un seul paramètre θ à estimer, la règle de Jeffreys propose de recourir aux lois *a priori* de la forme

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

Exemple 1.10. Soit $Y|\mu \sim \mathcal{P}(\mu T)$, avec $T > 0$. Alors la matrice d'information de Fisher est

$$I(\mu) = -E_{\mu} \left[\frac{\partial^2}{\partial \mu^2} \log \frac{e^{-\mu T} (\mu T)^Y}{Y!} \right] = \frac{T}{\mu^2} E_{\mu} [Y] = \frac{T^2}{\mu}.$$

Selon la règle de Jeffreys, on prend

$$\pi(\mu) \propto \frac{1}{\sqrt{\mu}} \mathbb{1}_{(0,\infty)}(\mu). \quad (1.7)$$

Il s'agit d'une loi impropre qui peut être vue comme la loi limite de la distribution $\mathcal{G}(a, b)$ avec $a = 1/2$ et $b \rightarrow 0$.

À partir de la Proposition 1.8 appliquée à notre modèle, on a vu qu'inférer sur ϕ revenait à inférer sur la proportion $p = \frac{\phi S}{\phi S + T}$ du modèle binomial $\mathcal{Bin}(n, p)$. L'exemple suivant illustre alors la règle de Jeffreys employée à ce modèle.

Exemple 1.11. Soit $X|p \sim \mathcal{Bin}(n, p)$ avec, pour fonction de masse,

$$p(x|p) = \binom{n}{x} p^x (1-p)^{n-x} \mathbb{1}_{\{0,1,\dots,n\}}(x), \quad 0 \leq p \leq 1.$$

Alors la matrice d'information de Fisher est

$$I(p) = -E_p \left[\frac{\partial^2}{\partial p^2} \log \left(\binom{n}{X} p^X (1-p)^{n-X} \right) \right]$$

$$\begin{aligned}
&= E_p \left[\frac{X}{p^2} + \frac{n - X}{(1 - p)^2} \right] \\
&= \frac{np}{p^2} + \frac{n - np}{(1 - p)^2} \\
&= \frac{n}{p(1 - p)}
\end{aligned}$$

En vertu de la règle de Jeffreys, on prend $\pi(p) \propto p^{-1/2}(1 - p)^{-1/2}\mathbb{1}_{[0,1]}(p)$, correspondant à une loi bêta $\mathcal{B}\left(\frac{1}{2}, \frac{1}{2}\right)$.

La règle de Jeffreys s'applique aussi dans le cas multidimensionnel. Si $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ est le vecteur paramétrique d'intérêt, alors la matrice d'information de Fisher $I(\boldsymbol{\theta})_{k \times k}$ est la matrice dont l'élément $I(\boldsymbol{\theta})_{ij}$ est défini comme

$$I(\boldsymbol{\theta})_{ij} = -E_{\boldsymbol{\theta}} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(X|\boldsymbol{\theta}) \right], \quad (1.8)$$

où l'observé est X de densité $f(x|\boldsymbol{\theta})$. Dans ce cas, la règle de Jeffreys stipule de choisir la loi *a priori*

$$\pi(\boldsymbol{\theta}) \propto \sqrt{\det I(\boldsymbol{\theta})}. \quad (1.9)$$

Exemple 1.12. Dans notre modèle d'intérêt, la matrice d'information de Fisher est

$$I(\boldsymbol{\theta}) = \begin{bmatrix} (T + \phi S)/\mu & -S \\ -S & \mu S/\phi \end{bmatrix}.$$

Comme $\det(I(\boldsymbol{\theta})) = ST/\phi$, en vertu de la règle de Jeffreys, on a

$$\pi(\mu, \phi) \propto 1/\sqrt{\phi}.$$

Puisque μ et ϕ sont des paramètres supposés indépendants, on peut alors poser $\pi(\phi) = \frac{1}{\sqrt{\phi}}\mathbb{1}_{\Phi}(\phi)$ et $\pi(\mu) = \mathbb{1}_{[0,\infty)}(\mu)$, deux lois impropres dans le cas où $\Theta = \mathbb{R}_+ \times \mathbb{R}_+$. Par contre, si on restreint le domaine de ϕ sur $\Phi = [0, 1]$, c'est-à-dire $\Theta = \mathbb{R}_+ \times [0, 1]$, alors $\phi \sim \mathcal{B}\left(\frac{1}{2}, 1\right)$, une loi non informative propre.

Comme en témoigne l'exemple précédent, le fait que l'espace paramétrique n'exerce pas d'influence sur la loi obtenue constitue un autre avantage de la méthode proposée par Jeffreys. D'autres méthodes ont été développées pour choisir une loi *a priori* pour le ratio de deux taux de Poisson (voir, par exemple, [YB96]). On réfère le lecteur, par exemple, à [Ber85, Chapitre 3] et à [YB96] pour plus d'information sur le choix de lois *a priori* en général.

L'idée de l'approche bayésienne est de mettre à jour la loi *a priori* en la combinant à l'information fournie par l'échantillon. L'actualisation de cette loi, dite *loi a posteriori* et notée $\pi(\cdot|x)$, est basée sur la règle de Bayes.

Définition 1.13. *Soit X une variable aléatoire (ou un vecteur aléatoire) d'une famille de densités $\mathcal{F} = \{f(\cdot|\theta), \theta \in \Theta\}$. Si $\pi(\theta)$ est la loi *a priori* de θ de fonction de répartition F_π , alors la loi *a posteriori* de θ est*

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)}, \quad (1.10)$$

où $m(x)$ est la loi marginale de X , donnée par

$$m(x) = \int_{\Theta} f(x|\theta)dF_\pi(\theta). \quad (1.11)$$

La loi *a posteriori* $\pi(\theta|x)$ est la loi utilisée pour inférer sur le(s) paramètre(s) d'intérêt.

Pour le prochain exemple, on aura besoin de définir une fonction couramment utilisée. Elle sera notamment utile dans les chapitres subséquents en lien avec la théorie des fonctions hypergéométriques.

Définition 1.14. *Le symbole de Pochhammer, également appelée factorielle ascendante, est une fonction notée $(x)_n$ et définie, pour $x \in \mathbb{R}$ et $n \in \mathbb{N}$, comme*

$$(x)_n = \begin{cases} 1 & \text{si } n = 0; \\ \prod_{k=0}^{n-1} (x+k) & \text{si } n \geq 1. \end{cases} \quad (1.12)$$

Remarque 1.15. Si x et $x+n$ sont des nombres non négatifs, on peut également l'écrire sous forme d'un quotient de fonctions gamma :

$$(x)_n = \frac{\Gamma(x+n)}{\Gamma(x)}. \quad (1.13)$$

Dans le cas de nombres non négatifs, deux propriétés utiles du symbole de Pochhammer sont présentées ci-dessous. Les démonstrations s'effectuent par calcul direct.

Propriétés 1.16. Pour $x \in \mathbb{N}$, on a

- a) $\binom{x+n-1}{n} = \frac{(x)_n}{n!}$, pour tout $n \geq 1$;
- b) $(x)_{m+n} = (x+m)_n(x)_m$, pour tout $n, m \in \mathbb{N}$.

Exemple 1.17. Soient $\mathbf{X}|\boldsymbol{\lambda} \sim \mathcal{P}2(\lambda_1 S, \lambda_2 T)$, $\lambda_1|\lambda_2 \sim \mathcal{G}(a, b)$ et $\lambda_2 \sim \mathcal{G}(c, d)$, d'où

$$\pi(\lambda_1, \lambda_2) = \frac{b^a \lambda_1^{a-1} e^{-\lambda_1 b}}{\Gamma(a)} \frac{d^c \lambda_2^{c-1} e^{-\lambda_2 d}}{\Gamma(c)}.$$

Les constantes a, b, c et d sont dites hyperparamètres. Alors, pour $(x_1, x_2) \in \mathbb{N} \times \mathbb{N}$ et $(\lambda_1, \lambda_2) \in \mathbb{R}_+ \times \mathbb{R}_+$, la loi a posteriori de $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ est

$$\begin{aligned} \pi(\lambda_1, \lambda_2 | x_1, x_2) &\propto p(x_1, x_2 | \lambda_1, \lambda_2) \pi(\lambda_1, \lambda_2) \\ &= \frac{1}{x_1! x_2!} e^{-(\lambda_1 S + \lambda_2 T)} (\lambda_1 S)^{x_1} (\lambda_2 T)^{x_2} \times \frac{b^a \lambda_1^{a-1} e^{-\lambda_1 b}}{\Gamma(a)} \times \frac{d^c \lambda_2^{c-1} e^{-\lambda_2 d}}{\Gamma(c)} \\ &\propto e^{-\lambda_1(S+b)} \lambda_1^{a+x_1-1} \times e^{-\lambda_2(T+d)} \lambda_2^{c+x_2-1} \end{aligned}$$

qui correspond produit de deux densités de lois gamma $\mathcal{G}(a+x_1, S+b)$ et $\mathcal{G}(c+x_2, T+d)$.

La constante de proportionnalité du calcul précédent est donnée par la marginale de \mathbf{X} issue de l'équation (1.11). Son calcul est illustré ci-dessous :

$$m(x_1, x_2) = \int_0^\infty \int_0^\infty \frac{1}{x_1! x_2!} e^{-(\lambda_1 S + \lambda_2 T)} (\lambda_1 S)^{x_1} (\lambda_2 T)^{x_2} \frac{b^a \lambda_1^{a-1} e^{-\lambda_1 b}}{\Gamma(a)} \frac{d^c \lambda_2^{c-1} e^{-\lambda_2 d}}{\Gamma(c)} d\lambda_1 d\lambda_2$$

$$\begin{aligned}
&= \left(\frac{S^{x_1} b^a}{x_1! \Gamma(a)} \int_0^\infty e^{-\lambda_1(S+b)} \lambda_1^{a+x_1-1} d\lambda_1 \right) \left(\frac{T^{x_2} d^c}{x_2! \Gamma(c)} \int_0^\infty e^{-\lambda_2(T+d)} \lambda_2^{c+x_2-1} d\lambda_2 \right) \\
&= \frac{S^{x_1} b^a}{x_1! \Gamma(a)} \frac{\Gamma(a+x_1)}{(S+b)^{a+x_1}} \times \frac{T^{x_2} d^c}{x_2! \Gamma(c)} \frac{\Gamma(c+x_2)}{(T+d)^{c+x_2}} \\
&= \frac{(a)_{x_1}}{x_1!} \left(\frac{b}{S+b} \right)^a \left(\frac{S}{S+b} \right)^{x_1} \times \frac{(c)_{x_2}}{x_2!} \left(\frac{d}{T+d} \right)^c \left(\frac{T}{T+d} \right)^{x_2},
\end{aligned}$$

qui est la fonction de masse conjointe de deux lois binomiales négatives indépendantes $\mathcal{BN}\left(a, \frac{b}{S+b}\right)$ et $\mathcal{BN}\left(c, \frac{d}{T+d}\right)$.

Dans l'exemple précédent, on remarque, d'une part, que même si l'on doit imposer $a, b, c, d > 0$ sur la loi *a priori* de $\boldsymbol{\lambda}$ pour qu'il s'agisse d'une loi propre, la loi *a posteriori* demeure tout de même une fonction de densité pour tout $(x_1, x_2) \in \mathbb{N} \times \mathbb{N}$ si $a, c > 0$, $b > -S$ et $d > -T$, ce qui est notamment le cas pour $\pi(\lambda_1, \lambda_2) = \lambda_1^{a-1} \lambda_2^{c-1} \mathbb{1}_{\mathbb{R}_+ \times \mathbb{R}_+}(\lambda_1, \lambda_2)$ lorsque $b = d = 0$. Cet exemple reflète alors que le fait d'avoir une loi *a priori* impropre n'est pas nécessairement un facteur limitant pour effectuer de l'inférence statistique dans le contexte bayésien, en autant que la loi *a posteriori* soit une fonction de densité. D'autre part, bien que le choix de la loi *a priori* de μ fut arbitraire ici, une propriété intéressante en découle : la loi *a posteriori* appartient à la même famille de lois que la loi *a priori*.

Définition 1.18. Soit \mathcal{F} une famille de fonctions de densité ou de fonctions de masse $f(\cdot|\theta)$, où $\theta \in \Theta$. Une classe de lois *a priori* Π est une famille conjuguée pour \mathcal{F} si la loi *a posteriori* appartient à Π pour tout $f \in \mathcal{F}$, pour tout $\pi \in \Pi$ et pour tout $x \in \mathcal{X}$.

Une technique aidant à identifier une famille conjuguée naturelle de lois *a priori* consiste à utiliser la vraisemblance de la famille paramétrique, que nous noterons $\mathcal{L}(\theta_1, \dots, \theta_k|\mathbf{x})$, où \mathbf{x} est une donnée observée de l'échantillon $\mathbf{X} = (X_1, \dots, X_n)$ de fonction de densité conjointe $f(\mathbf{x}|\theta_1, \dots, \theta_k)$. Les deux exemples qui suivent illustrent cette méthode pour

des familles paramétriques usuelles. Ces résultats pourraient notamment être généralisés pour les familles exponentielles.

Exemple 1.19. Soient $X|\lambda \sim \mathcal{P}(\lambda)$ et $\pi(\lambda)$ une loi a priori pour λ . Comme

$$\pi(\lambda|x) \propto e^{-\lambda} \lambda^x \pi(\lambda),$$

il est raisonnable de penser que la loi a priori $\mathcal{G}(a, b)$ est une famille conjuguée pour la famille poissonnienne. En effet, à l'Exemple 1.17, comme il y a indépendance entre les paramètres et les variables aléatoires conditionnellement aux paramètres, on peut alors faire un lien avec le cas univarié ici et conclure à la véracité de cette hypothèse.

Exemple 1.20. Soient $X|p \sim \mathcal{Bin}(n, p)$ et

$$\pi(p) \propto \frac{p^{a-1}(1-p)^{b-1}}{[1-(1-\lambda)p]^\kappa} \mathbb{1}_{[0,1]}(p), \quad a, b, \lambda > 0, \quad \kappa \geq 0. \quad (1.14)$$

La loi bêta $\mathcal{B}(a, b)$ est un cas particulier de cette loi, lorsque $\lambda = 1$ ou $\kappa = 0$. Cette généralisation de la loi bêta sera entre autres très utile pour l'estimation bayésienne de ϕ du modèle poissonnien à l'étude, sous la contrainte paramétrique $\phi \in [0, 1]$. Cette loi, dite loi bêta généralisée à quatre paramètres et notée $\mathcal{BG4}(a, b, \lambda, \kappa)$, sera davantage détaillée au chapitre 2.

De l'équation (1.10), on obtient

$$\pi(p|x) \propto \frac{p^{a+x-1}(1-p)^{b+n-x-1}}{[1-(1-\lambda)p]^\kappa} \mathbb{1}_{[0,1]}(p),$$

ce qui implique que $p|x$ se distribue selon une loi bêta généralisée à 4 paramètres $\mathcal{BG4}(a+x, b+n-x, \lambda, \kappa)$. Donc les lois bêta généralisées forment une famille conjuguée pour la famille de lois binomiales.

Les calculs de lois marginales et de lois *a posteriori* peuvent s'avérer difficile et guider parfois vers des lois qui ne seront calculables qu'à partir d'approximations numériques.

Ainsi, comme l'inférence est effectuée à partir de la loi *a posteriori*, il peut être alors avantageux d'utiliser la famille conjuguée pour choisir la loi *a priori* afin d'obtenir, en général, des estimateurs de forme close liés à des lois dont on connaît les caractéristiques intrinsèques. En plus des motifs calculatoires, l'intérêt porté aux familles conjuguées réside également en la capacité de voir, à travers la mise à jour des paramètres de la loi *a priori*, l'effet des données échantillonales sur cette dernière, ce qui peut en faciliter l'interprétation.

Trouver des familles de lois *a priori* conjuguées pour des modèles bivariés, comme pour notre modèle d'intérêt, peut s'avérer une tâche plus fastidieuse que dans le cas univarié. L'Exemple 1.17 permet de conclure que si notre modèle d'intérêt est paramétrisé sous la forme $(X_1, X_2)|\lambda, \mu \sim \mathcal{P}2(\lambda S, \mu T)$, alors la famille de lois *a priori* conjuguée est telle que $\lambda|\mu \sim \mathcal{G}(a, b)$ et $\mu \sim \mathcal{G}(c, d)$. Par contre, ici, notre intérêt est plutôt porté à la modélisation $(X_1, X_2)|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$. Ainsi, le prochain exemple, précédé d'un lemme auquel on fera référence, illustre le fait que la reparamétrisation sur notre modèle n'entraîne pas nécessairement la préservation d'une famille conjuguée.

Lemme 1.21. *Soient X_1 et X_2 deux variables aléatoires indépendantes distribuées respectivement selon une loi $\mathcal{G}(\alpha_1, \beta_1)$ et une loi $\mathcal{G}(\alpha_2, \beta_2)$. Posons $\lambda = \beta_1/\beta_2$. Alors la distribution du rapport $R = X_1/X_2$ est caractérisée par la fonction de densité suivante :*

$$f(r) = \frac{\lambda^{\alpha_1}}{B(\alpha_1, \alpha_2)} \frac{r^{\alpha_1-1}}{(1 + \lambda r)^{\alpha_1+\alpha_2}} \mathbb{1}_{[0, \infty)}(r). \quad (1.15)$$

Démonstration. Comme le jacobien entre x_1/x_2 et x_2 est x_2 , et que $X_1 \perp X_2$, alors, pour $r \geq 0$, on a

$$\begin{aligned} f(r) &= \int_0^\infty x_2 f_{X_1}(rx_2) f_{X_2}(x_2) dx_2 \\ &= \frac{\beta_1^{\alpha_1} \beta_2^{\alpha_2}}{\Gamma(\alpha_1) \Gamma(\alpha_2)} r^{\alpha_1-1} \int_0^\infty x_2^{\alpha_1+\alpha_2-1} e^{-x_2(\beta_2+\beta_1 r)} dx_2 \end{aligned}$$

$$\begin{aligned}
&= \frac{\beta_1^{\alpha_1} \beta_2^{\alpha_2}}{\Gamma(\alpha_1) \Gamma(\alpha_2)} r^{\alpha_1-1} \frac{\Gamma(\alpha_1 + \alpha_2)}{(\beta_2 + \beta_1 r)^{\alpha_1 + \alpha_2}} \\
&= \frac{(\beta_1/\beta_2)^{\alpha_1}}{B(\alpha_1, \alpha_2)} \frac{r^{\alpha_1-1}}{(1 + (\beta_1/\beta_2)r)^{\alpha_1 + \alpha_2}}. \quad \square
\end{aligned}$$

Remarque 1.22. La distribution introduite au Lemme 1.21 sera dite *bêta de type II* et notée $\mathcal{B}_2(\alpha_1, \alpha_2, \frac{1}{\lambda})$. La structure de celle-ci sera étudiée à la section 2.3 du chapitre 2.

Exemple 1.23. Soit $(X_1, X_2)|\lambda, \mu \sim \mathcal{P}2(\lambda S, \mu T)$ avec $S, T > 0$. À l'Exemple 1.17, on a vu que, pour cette paramétrisation, la famille de lois a priori conjuguée est telle que $\lambda|\mu \sim \mathcal{G}(a, b)$ et $\mu \sim \mathcal{G}(c, d)$. On veut illustrer ici le fait que si on reparamétrise le paramètre λ , en posant $\lambda = \mu\phi$, alors la propriété de famille conjuguée n'est pas préservée.

Du Lemme 1.21, avec les hypothèses de lois a priori énoncées ci-dessus pour λ et μ , on sait que $\phi \sim \mathcal{B}_2(a, c, \frac{d}{b})$. Comme on a supposé que μ et ϕ sont a priori indépendantes, on a

$$\pi(\mu, \phi) \propto \mu^{a-1} e^{-b\mu} \frac{\phi^{a-1}}{(1 + (b/d)\phi)^{a+c}} \mathbb{1}_{\mathbb{R}_+ \times \mathbb{R}_+}(\mu, \phi).$$

Ainsi, de l'équation (1.10), pour $(x_1, x_2) \in \mathbb{N} \times \mathbb{N}$, on obtient

$$\pi(\mu, \phi|x_1, x_2) \propto \mu^{a+x_1+x_2-1} e^{-\mu(T+b+\phi S)} \frac{\phi^{a+x_1-1}}{(1 + (b/d)\phi)^{a+c}} \mathbb{1}_{\mathbb{R}_+ \times \mathbb{R}_+}(\mu, \phi),$$

qui n'est pas de la même forme que la loi a priori.

Dans [LL12], Laurent et Legrand ont utilisé un concept similaire à celui des familles conjuguées, soit les familles semi-conjuguées. Dans cet article, ils discutent d'ailleurs des avantages que confère une telle structure par rapport à la famille conjuguée naturelle du modèle $(X_1, X_2)|\lambda, \mu \sim \mathcal{P}2(\lambda S, \mu T)$, qui seront abordés au chapitre 3. L'exemple suivant illustre ce concept.

Exemple 1.24. *Considérons notre modèle d'intérêt $(X_1, X_2)|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$, avec $S, T > 0$. En supposant que $\mu \sim \mathcal{G}(a, b)$ et que $\phi|\mu \sim \mathcal{G}(c, \mu)$, la loi conjointe a priori de μ et ϕ s'exprime comme suit :*

$$\begin{aligned}\pi(\mu, \phi) &\propto \mu^{a-1} e^{-\mu b} \mathbf{1}_{\mathbb{R}_+}(\mu) \times \mu^c \phi^{c-1} e^{-\phi \mu} \mathbf{1}_{\mathbb{R}_+}(\phi) \\ &= \mu^{a+c-1} \phi^{c-1} e^{-\mu(b+\phi)} \mathbf{1}_{\mathbb{R}_+ \times \mathbb{R}_+}(\mu, \phi).\end{aligned}$$

Pour $(x_1, x_2) \in \mathbb{N}^2$, la loi a posteriori est alors

$$\begin{aligned}\pi(\mu, \phi|x_1, x_2) &\propto p(x_1, x_2|\mu, \phi)\pi(\mu, \phi) \\ &\propto e^{-\mu(T+\phi S)} \mu^{x_1+x_2} \phi^{x_1} \times \mu^{a+c-1} \phi^{c-1} e^{-\mu(b+\phi)} \mathbf{1}_{\mathbb{R}_+ \times \mathbb{R}_+}(\mu, \phi) \\ &= \mu^{a+x_2+c+x_1-1} \phi^{c+x_1-1} e^{-\mu(T+b+(1+S)\phi)} \mathbf{1}_{\mathbb{R}_+ \times \mathbb{R}_+}(\mu, \phi) \\ &= \mu^{a+x_2-1} e^{-\mu(T+b)} \mathbf{1}_{\mathbb{R}_+}(\mu) \times \mu^{c+x_1} \phi^{c+x_1-1} e^{-\phi \mu(1+S)} \mathbf{1}_{\mathbb{R}_+}(\phi).\end{aligned}$$

Dans l'exemple précédent, on remarque que la loi a posteriori $\pi(\mu, \phi|x_1, x_2)$ n'appartient pas à la même famille de densités que la loi a priori pour tout $S > 0$, et donc, il ne s'agit pas d'une famille de lois conjuguées pour notre modèle paramétrique d'intérêt. Par contre, on remarque que cette loi a posteriori se décompose de telle sorte que $\mu|\phi, x_1, x_2 \sim \mathcal{G}(a + x_2, T + b)$ et que $\phi|\mu, x_1, x_2 \sim \mathcal{G}(c + x_1, (1 + S)\mu)$. Ainsi, chacune de ces lois conditionnelles est respectivement de la même famille que sa loi conditionnelle a priori correspondante, puisque $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi|\mu \sim \mathcal{G}(c, \mu)$. De ce fait, comme, pour chaque paramètre, sa loi a posteriori conditionnelle à l'autre paramètre est dans la même famille de lois que sa loi a priori conditionnelle à l'autre paramètre, on parlera alors d'une famille semi-conjuguée.

Définition 1.25. *Soit \mathcal{F} une famille de fonctions de densité ou de fonctions de masse $f(\cdot|\theta)$, où $\theta \in \Theta$. Une classe de lois a priori Π est une famille semi-conjuguée pour \mathcal{F} si, pour chaque paramètre θ_i du vecteur θ , sa loi a posteriori conditionnelle aux autres paramètres est dans la même famille de lois que sa loi a priori conditionnelle aux autres paramètres, pour tout $f \in \mathcal{F}$, pour tout $\pi \in \Pi$ et pour tout $\mathbf{x} \in \mathcal{X}$.*

On avait vu à l'Exemple 1.17 que la famille conjuguée naturelle, pour le modèle à l'étude $(X_1, X_2)|\lambda, \mu \sim \mathcal{P}2(\lambda S, \mu T)$, était $\lambda \sim \mathcal{G}(a, b)$ et $\mu \sim \mathcal{G}(c, d)$, où λ et μ sont indépendants. On peut alors se demander en quoi réside l'apport d'une famille semi-conjuguée par rapport à cette famille conjuguée, outre les motifs calculatoires. Dans [LL12], Laurent et Legrand discutent d'un avantage important que confère une famille semi-conjuguée, via la reparamétrisation $\lambda = \mu\phi$, par rapport à la famille naturelle conjuguée. En effet, à partir d'observations (x_1, x_2) , supposons que l'on veuille effectuer des prédictions pour des variables Y_1 et Y_2 , si $(Y_1, Y_2)|\lambda, \mu \sim \mathcal{P}2(\lambda S', \mu T')$ (cet objectif fait référence à la notion d'inférence prédictive, qui sera détaillée à la section 1.3). En supposant que μ et λ sont indépendants et tous deux de lois *a priori* gamma, ceci mène à des distributions prédictives de Y_1 et Y_2 qui sont indépendantes. Or, d'un point de vue clinique, ceci paraît peu plausible puisque l'on s'attend à ce que le nombre de personnes atteintes par une maladie Y_1 dans le groupe traité ou vacciné soit corrélé positivement au nombre de personnes atteintes Y_2 dans le groupe contrôle. Donc l'utilisation d'une famille semi-conjuguée via la reparamétrisation $\lambda = \mu\phi$ semble plus appropriée dans un tel cadre pratique.

Dans l'Exemple 1.24, comme $\phi|\mu \sim \mathcal{G}(c, \mu)$, il y a dépendance *a priori* entre les paramètres μ et ϕ . Par contre, Laurent présente dans [Lau12] une famille semi-conjuguée pour ce modèle qui suppose, *a priori*, l'indépendance entre les paramètres μ et ϕ . Cette approche sera abordée au chapitre 3.

1.2.2 Estimation avec l'approche fréquentiste

Pour présenter la théorie d'inférence statistique des prochaines sections, on considérera le modèle statistique général $X \sim p_\theta$ (ou $\mathbf{X} \sim p_\theta$ dans le cas multivarié), où le paramètre θ (ou le vecteur paramétrique $\boldsymbol{\theta}$) appartient à l'espace paramétrique Θ . On notera $\hat{\theta}(X)$,

des estimateurs de θ qui sont fonction de X . Lorsqu'une observation x est obtenue, on notera $\hat{\theta}(x)$ l'estimé qui en résulte.

Méthodes d'estimation ponctuelle

Afin d'estimer ponctuellement un paramètre d'une population, c'est-à-dire par une valeur unique calculée à partir de données échantillonnales, plusieurs méthodes ont été élaborées. Parmi celles-ci, on retrouve notamment la méthode des moments, introduite par Karl Pearson à la fin du 19^e siècle, occasionnant ainsi des avancées considérables dans la théorie de l'estimation [Dod08]. Par contre, alors qu'il n'existe pas toujours de solutions au système d'équations à résoudre à travers cette méthode, par exemple, lorsque des contraintes sont imposées sur l'espace paramétrique, cette technique peut s'avérer limitante. Ainsi, pour notre modèle, lorsqu'une contrainte sera appliquée sur l'espace paramétrique de ϕ au chapitre 4, cette méthode ne sera pas judicieuse.

Peu de temps après que la méthode de Pearson fut proposée, Fisher, statisticien des statistiques modernes à qui on attribua des éléments clés de la théorie de l'inférence statistique, proposa une méthode basée sur la fonction de vraisemblance, qui fut introduite dans ses premières publications en 1912 et explicitée dans un papier écrit en 1922 [Dod08]. Cette méthode consiste à trouver l'estimateur $\hat{\theta}(\mathbf{x}) = (\theta_1, \dots, \theta_k)$ qui maximise la vraisemblance $\mathcal{L}(\theta_1, \dots, \theta_k | \mathbf{x})$, ou encore, $\ln \mathcal{L}(\theta_1, \dots, \theta_k | \mathbf{x})$.

Plusieurs autres méthodes existent et sont détaillées par exemple dans [CB02, Chapitre 7] et dans [KK92, Chapitre 2] (pour des distributions bivariées).

Estimation par intervalles de confiance

Au lieu d'estimer un paramètre θ par une valeur unique, on peut plutôt construire un intervalle qui couvre le paramètre avec une certaine probabilité $1 - \alpha$. L'avantage que

confère cette approche par rapport à l'estimation ponctuelle est la certaine garantie, donnée par le niveau de confiance, de couvrir le paramètre d'intérêt. En construisant des intervalles, on pourrait souhaiter obtenir des fonctions L_1 , L_2 , U_1 et U_2 telles que

$$\mathbb{P}([L_1(X_1, X_2), U_1(X_1, X_2)] \ni \mu, [L_2(X_1, X_2), U_2(X_1, X_2)] \ni \phi | \mu, \phi) = 1 - \alpha,$$

avec $1 - \alpha > 0$, d'où la « certaine garantie » décrite plus tôt.

Définition 1.26. *Pour un intervalle de confiance $[L(\mathbf{X}), U(\mathbf{X})]$ d'une fonction paramétrique $g(\theta)$, la probabilité de recouvrement se définit comme*

$$\mathbb{P}([L(\mathbf{X}), U(\mathbf{X})] \ni \theta | \theta). \quad (1.16)$$

Définition 1.27. *Pour un intervalle de confiance $[L(\mathbf{X}), U(\mathbf{X})]$ d'une fonction paramétrique $g(\theta)$, le niveau de confiance est l'infimum des probabilités de recouvrement, c'est-à-dire*

$$\inf_{\theta} \mathbb{P}([L(\mathbf{X}), U(\mathbf{X})] \ni \theta | \theta). \quad (1.17)$$

Comme ce mémoire est davantage dédié à l'approche bayésienne (voir section 1.2.3), on réfère le lecteur à [CB02, Section 9.2] pour davantage de détails pour la construction d'intervalles de confiance pour une fonction paramétrique $g(\theta)$. Pour notre modèle poissonnien d'intérêt, la construction d'intervalles de confiance pour le paramètre ϕ représentant le ratio de deux paramètres poissonniens a été étudiée et décrite par exemple dans [Cou98] et dans [PB00].

Évaluation fréquentiste des estimateurs

Plusieurs critères peuvent être analysés afin d'étudier la qualité d'un estimateur. Une des façons de procéder est de se fixer une mesure de distance entre l'estimateur $\hat{\theta}$ et le paramètre θ , qui sera notée ici $L(\theta, \hat{\theta})$, et de calculer la distance moyenne, fonction de

θ . La fonction quadratique $L(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2$ et la fonction valeur absolue $L(\theta, \hat{\theta}) = |\hat{\theta} - \theta|$ sont des mesures de distance couramment utilisées pour ce calcul de perte espérée $E_\theta[L(\theta, \hat{\theta}(X))]$.

Définition 1.28. Soient $\hat{\theta}(X)$ un estimateur de θ , et $L(\theta, \hat{\theta})$ une fonction de perte. La fonction de risque se définit comme étant la perte espérée, c'est-à-dire

$$R(\theta, \hat{\theta}) = E_\theta[L(\theta, \hat{\theta}(X))], \quad \theta \in \Theta. \quad (1.18)$$

Remarque 1.29. Dans le cas particulier où la fonction de perte est la fonction quadratique $L(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2$, le risque est souvent dit « erreur quadratique moyenne », noté ici $EQM(\theta, \hat{\theta})$:

$$EQM(\theta, \hat{\theta}) = E_\theta[(\hat{\theta}(X) - \theta)^2], \quad \theta \in \Theta. \quad (1.19)$$

Le choix de la perte quadratique confère un avantage intéressant au niveau de l'interprétation du calcul de risque [CB02]. En effet, l'erreur quadratique moyenne peut se décomposer de la façon suivante :

$$E_\theta[(\hat{\theta}(X) - \theta)^2] = \text{Var}_\theta(\hat{\theta}(X)) + (E_\theta[\hat{\theta}(X)] - \theta)^2. \quad (1.20)$$

Définition 1.30. Pour un paramètre θ , le biais d'un estimateur $\hat{\theta}$ est la différence entre l'espérance de $\hat{\theta}(X)$ et θ , c'est-à-dire

$$b(\theta, \hat{\theta}) = E_\theta[\hat{\theta}(X)] - \theta, \quad \theta \in \Theta. \quad (1.21)$$

Si $b(\theta, \hat{\theta}) = 0$ pour tout $\theta \in \Theta$, on dit que l'estimateur $\hat{\theta}$ est sans biais.

Remarque 1.31. On a $E_\theta[(\hat{\theta}(X) - \theta)^2] = \text{Var}_\theta(\hat{\theta}(X)) + b(\theta, \hat{\theta})^2$. Ainsi, si un estimateur $\hat{\theta}$ est sans biais, alors $EQM(\theta) = \text{Var}(\hat{\theta}(X))$.

Selon le critère de performance basé sur le risque, on cherche alors un estimateur $\hat{\theta}$ qui offre un petit risque peu importe la valeur de θ dans l'espace paramétrique Θ .

Définition 1.32. Soient $\hat{\theta}_1(X)$ et $\hat{\theta}_2(X)$ deux estimateurs de θ . On dit que $\hat{\theta}_1(X)$ domine $\hat{\theta}_2(X)$ si, pour tout $\theta \in \Theta$, $R(\theta, \hat{\theta}_1) \leq R(\theta, \hat{\theta}_2)$ avec une inégalité stricte pour au moins une valeur de θ .

Définition 1.33. Un estimateur $\hat{\theta}(X)$ est dit admissible s'il n'existe pas d'estimateurs le dominant. Dans le cas contraire, il est dit inadmissible.

Pour un estimateur $\hat{\theta}(X)$, la quantité $\sup_{\theta \in \Theta} R(\theta, \hat{\theta})$ représente le plus grand risque que peut occasionner cet estimateur. Ainsi, on peut vouloir trouver un estimateur qui minimise cette quantité.

Définition 1.34. Un estimateur $\hat{\theta}_m$ est minimax s'il minimise le supremum du risque en θ , c'est-à-dire si

$$\sup_{\theta \in \Theta} R(\theta, \hat{\theta}_m) = \inf_{\hat{\theta}} \sup_{\theta \in \Theta} R(\theta, \hat{\theta}). \quad (1.22)$$

Pour clore cette sous-section de l'estimation selon une approche fréquentiste, on présente un exemple qui montre que pour le modèle poissonnien univarié $X|\mu \sim \mathcal{P}(\mu)$, l'estimateur $\hat{\mu}(X) = X$ est admissible pour μ sous la perte quadratique.

Exemple 1.35. Soit $X|\mu \sim \mathcal{P}(\mu)$. Montrons que $\hat{\mu}(X) = X$ est un estimateur admissible pour μ sous la perte quadratique. Cette démonstration se basera sur la démonstration du Théorème de Karlin ([LC98], p.331-332).

Comme $\hat{\mu}(X) = X$ est un estimateur sans biais de μ , le risque est

$$R(\mu, X) = E_{\mu} [(\mu - X)^2] = \text{Var}_{\mu}(X) = \mu.$$

Pour tout estimateur $\hat{\mu}(X)$, on peut écrire $E_{\mu}[\hat{\mu}(X)] = \mu + b(\mu, \hat{\mu})$, et on a $R(\mu, \hat{\mu}) = \text{Var}_{\mu}(\hat{\mu}(X)) + b(\mu, \hat{\mu})^2$. En vertu de l'inégalité de Cramér-Rao (Chapitre 5, Théorème 5.10, [LC98], p.120),

$$\text{Var}_{\mu}(\hat{\mu}(X)) \geq \frac{\left[\frac{\partial}{\partial \mu} E_{\mu}[\hat{\mu}(X)] \right]^2}{I(\mu)}, \quad (1.23)$$

où $I(\mu)$ est la matrice d'information de Fisher définie comme

$$I(\mu) = E_{\mu} \left[\left(\frac{\partial}{\partial \mu} \log p(X|\mu) \right)^2 \right]. \quad (1.24)$$

Comme $X|\mu$ appartient à une famille exponentielle (car $p(x|\mu) = \frac{1}{x!} e^{-\mu} e^{x \ln \mu}$), alors du Théorème 5.4 de [LC98], on a $I(\mu) = \frac{1}{\text{var}_{\mu}(X)} = \frac{1}{\mu}$. De ce fait, on a

$$R(\mu, \hat{\mu}) \geq \mu(1 + b'(\mu, \hat{\mu}))^2 + b(\mu, \hat{\mu})^2. \quad (1.25)$$

Supposons qu'il existe un estimateur $\hat{\mu}^*(X)$ qui domine l'estimateur X , c'est-à-dire pour tout $\mu \geq 0$,

$$R(\mu, \hat{\mu}^*) \leq R(\mu, X) = \mu,$$

avec une inégalité stricte pour au moins une valeur de μ . De l'inégalité (1.25), on aurait, pour tout $\mu \geq 0$,

$$\begin{aligned} \mu(1 + b'(\mu, \hat{\mu}^*))^2 + b(\mu, \hat{\mu}^*)^2 \leq \mu &\implies \mu b'(\mu, \hat{\mu}^*)^2 + 2\mu b'(\mu, \hat{\mu}^*) + b(\mu, \hat{\mu}^*)^2 \leq 0 \\ &\implies 2\mu b'(\mu, \hat{\mu}^*) + b(\mu, \hat{\mu}^*)^2 \leq 0. \end{aligned} \quad (1.26)$$

Pour tout $\mu \geq 0$, on déduit donc, de l'inégalité (1.26), que $b'(\mu, \hat{\mu}^*) \leq 0$ et que $b(0, \hat{\mu}^*) = 0$. Comme le biais est non croissant en μ et que $b(0, \hat{\mu}^*) = 0$, cela implique alors que $b(\mu, \hat{\mu}^*) \leq 0$ pour tout $\mu \geq 0$.

Supposons qu'il existe une valeur μ_0 telle que $b(\mu_0, \hat{\mu}^*) < 0$. Comme $b'(\mu, \hat{\mu}^*) \leq 0$, $b(\mu, \hat{\mu}^*) < 0$ et $b(\mu, \hat{\mu}^*) - b(\mu_0, \hat{\mu}^*) \leq 0$ pour tout $\mu \geq \mu_0$, on aurait

$$0 \leq \frac{1}{b(\mu, \hat{\mu}^*)} - \frac{1}{b(\mu_0, \hat{\mu}^*)} \leq -\frac{1}{b(\mu_0, \hat{\mu}^*)}. \quad (1.27)$$

De l'inégalité (1.26), pour tout $\mu \geq \mu_0$, on a

$$-\frac{b'(\mu, \hat{\mu}^*)}{b^2(\mu, \hat{\mu}^*)} \geq \frac{1}{2\mu} \implies \frac{d}{d\mu} \left(\frac{1}{b(\mu, \hat{\mu}^*)} \right) \geq \frac{1}{2\mu}$$

$$\begin{aligned}
&\implies \int_{\mu_0}^{\mu} \frac{d}{dt} \left(\frac{1}{b(t)} \right) dt \geq \int_{\mu_0}^{\mu} \frac{1}{2t} dt \\
&\implies \frac{1}{b(\mu, \hat{\mu}^*)} - \frac{1}{b(\mu_0, \hat{\mu}^*)} \geq \frac{1}{2} (\ln(\mu) - \ln(\mu_0)) \\
&\implies \lim_{\mu \rightarrow \infty} \left(\frac{1}{b(\mu, \hat{\mu}^*)} - \frac{1}{b(\mu_0, \hat{\mu}^*)} \right) \geq \lim_{\mu \rightarrow \infty} \left(\frac{1}{2} (\ln(\mu) - \ln(\mu_0)) \right) \\
&\implies \lim_{\mu \rightarrow \infty} \left(\frac{1}{b(\mu, \hat{\mu}^*)} - \frac{1}{b(\mu_0, \hat{\mu}^*)} \right) = \infty,
\end{aligned}$$

ce qui contredit l'inégalité (1.27). Donc $b(\mu, \hat{\mu}^*) = 0$ pour tout $\mu \geq 0$, d'où $R(\mu, \hat{\mu}^*(X)) \equiv \mu$. Dans ce cas, $\hat{\mu}^*(X)$ ne domine pas X , ce qui contredit l'hypothèse que $\hat{\mu}^*(X)$ domine l'estimateur X .

Donc X est un estimateur admissible pour μ . □

1.2.3 Estimation avec l'approche bayésienne

Méthodes d'estimation ponctuelle

Selon une approche bayésienne, l'inférence est effectuée à partir de la loi *a posteriori*. Une des méthodes vues précédemment, en adoptant une approche fréquentiste, consistait à choisir une valeur $\hat{\theta}$ qui maximisait la fonction de vraisemblance. Ce concept est généralisable et applicable à l'approche bayésienne en maximisant la fonction $\mathcal{L}(\theta|x) = \pi(\theta|x)$ par rapport à θ . De plus, l'espérance *a posteriori* $E[\theta|x]$ et la médiane *a posteriori* sont des estimateurs bayésiens communs pour des raisons qui seront justifiées ultérieurement. Dans certains cas, il sera facile de les obtenir contrairement à d'autres, où il sera nécessaire d'avoir recours à des méthodes numériques. On voit alors ici l'avantage que confère l'utilisation de familles conjuguées et semi-conjuguées, qui peut grandement faciliter l'ob-

tention de ces estimateurs.

Estimation par intervalles de crédibilité

L'estimation par intervalles de confiance, selon une approche fréquentiste, a été abordée précédemment. Les intervalles (ou régions) de confiance ont un analogue bayésien : les intervalles (ou régions) de crédibilité.

Définition 1.36. *Un intervalle de crédibilité de niveau $1 - \alpha$ pour un paramètre θ est un intervalle $I_\pi(\mathbf{x}) = [L_\pi(\mathbf{x}), U_\pi(\mathbf{x})]$ tel que*

$$\mathbb{P}(\theta \in I_\pi(\mathbf{x}) | \mathbf{x}) \geq 1 - \alpha.$$

Exemple 1.37. *Soient $X | \mu \sim \mathcal{P}(\mu)$ et $\mu \sim \mathcal{G}(a, b)$. De l'Exemple 1.17, on sait que $\mu | x \sim \mathcal{G}(a+x, 1+b)$. En choisissant la loi a priori $\mathcal{G}(1, 0)$, on obtient alors $\mu | x \sim \mathcal{G}(1+x, 1)$. Par exemple, en simulant une observation pour $X | \mu = 20 \sim \mathcal{P}(\mu)$, on obtient $x = 17$. Alors $\mu | x = 17 \sim \mathcal{G}(18, 1)$. Dans ce cas, comme estimateurs ponctuels, on peut prendre, par exemple, l'espérance a posteriori ou la médiane a posteriori, c'est-à-dire $E[\mu | x = 17] = 18$ ou $Md(\mu | x = 17) = 17,6678$. À partir des quantiles de la loi $\mathcal{G}(18, 1)$, on peut également estimer μ à l'aide d'un intervalle de crédibilité. Ainsi, le paramètre μ est contenu dans l'intervalle $[10,6679; 27,2186]$ avec probabilité égale à 0,95. Cet intervalle est illustré à la Figure 1.1.*

Dans l'exemple précédent, l'intervalle de crédibilité a été construit à partir des quantiles $q_{0,025}$ et $q_{0,975}$ de la distribution $\mathcal{G}(18, 1)$. Toutefois, on remarque qu'une infinité d'intervalles de crédibilité de niveau 95 % peut être obtenue à partir de cette loi. Dans une optique d'inférence statistique, il peut alors être judicieux de choisir l'intervalle le moins large. On parle alors d'intervalle de crédibilité *HPD* (*highest posterior density*). Pour plus de détails à ce sujet, voir, par exemple, [Ber85, 4.3.2. Credibles Sets].

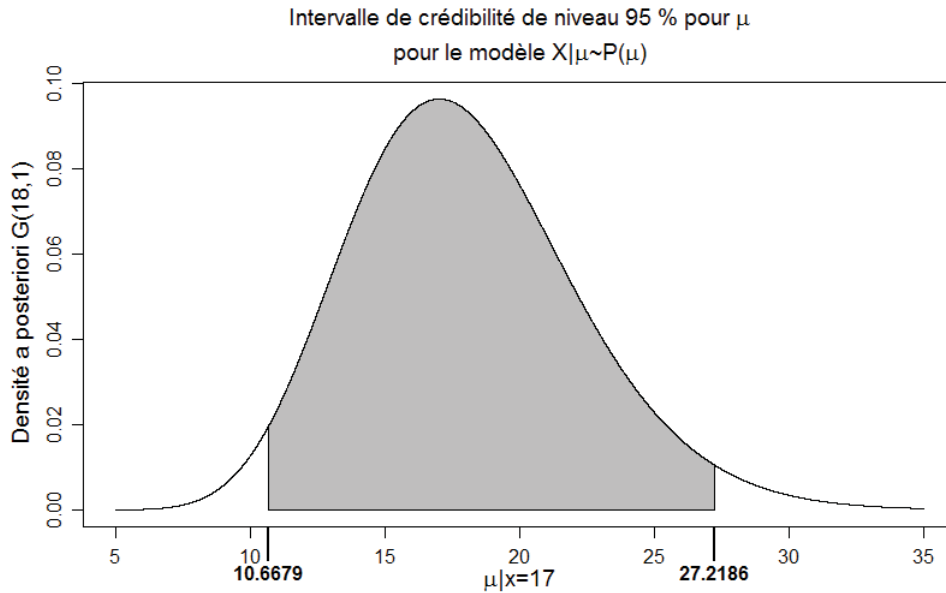


Figure 1.1 – Intervalle de crédibilité de niveau 95 % pour μ si $X|\mu \sim \mathcal{P}(\mu)$ à partir de l’observation $x = 17$

Évaluation des estimateurs issus de l’approche bayésienne

Quelques critères de performance ont été discutés dans la section faisant objet de l’évaluation fréquentiste des estimateurs. On y a notamment présenté la notion de risque, telle que définie à l’équation (1.18). On peut alors se demander si des estimateurs issus de l’approche bayésienne ont de bonnes propriétés fréquentistes. Autrement dit, est-ce que des estimateurs obtenus à partir de méthodes bayésiennes peuvent être admissibles ou encore minimax ? On discutera de cet aspect pour terminer cette section.

Puisque l’approche bayésienne considère le paramètre θ comme une variable aléatoire caractérisée par une loi *a priori* $\pi(\theta)$, on peut alors utiliser cette dernière pour calculer le risque espéré.

Définition 1.38. À partir d’une loi *a priori* $\pi(\theta)$ par rapport à une mesure σ -finie ν et

d'une fonction de perte $L(\theta, \hat{\theta})$, on définit le risque intégré de Bayes par

$$r_\pi(\hat{\theta}) = E^\theta[R(\theta, \hat{\theta})] = \int_{\Theta} R(\theta, \hat{\theta})\pi(\theta)d\nu(\theta). \quad (1.28)$$

Afin de minimiser le risque intégré de Bayes, lorsque ce dernier est fini, un simple calcul permet de constater qu'il suffit de minimiser la fonction de perte espérée a posteriori $E_{\theta|x}[L(\theta, \hat{\theta}(x))]$. En effet, à l'aide de la règle de Bayes, on obtient

$$\begin{aligned} r_\pi(\hat{\theta}) &= \int_{\Theta} R(\theta, \hat{\theta})\pi(\theta)d\nu(\theta) \\ &= \int_{\Theta} \left(\int_{\mathcal{X}} L(\theta, \hat{\theta}(x))f(x|\theta)d\mu(x) \right) \pi(\theta)d\nu(\theta) \\ &= \int_{\mathcal{X}} \left(\int_{\Theta} L(\theta, \hat{\theta}(x))\pi(\theta|x)d\nu(\theta) \right) f(x)d\mu(x), \end{aligned}$$

où μ est une mesure σ -finie sur l'espace échantillonnal de X .

Définition 1.39. *Un estimateur $\hat{\theta}_\pi(X)$ qui minimise le risque intégré de Bayes calculé pour une loi a priori π est dit estimateur de Bayes. De façon plus générale, un estimateur de Bayes $\hat{\theta}_\pi(x)$ minimise, pour tout $x \in \mathcal{X}$, l'espérance a posteriori de la perte $E_{\theta|x}[L(\theta, \hat{\theta})]$. En outre, si π est une loi a priori impropre, on dira que $\hat{\theta}_\pi(X)$ est un estimateur de Bayes généralisé.*

Exemple 1.40. *Voici quelques exemples usuels :*

a) *Pour la perte quadratique pondérée $L(\theta, \hat{\theta}) = w(\theta)(\theta - \hat{\theta})^2$, si $E[\theta^k w(\theta)|x] < \infty$ et existe pour $k = 0, 1$ et 2 et pour tout $x \in \mathcal{X}$, alors*

$$\hat{\theta}_\pi(x) = \frac{E_{\theta|x}[\theta w(\theta)]}{E_{\theta|x}[w(\theta)]}. \quad (1.29)$$

En effet, en dérivant la perte espérée a posteriori et en l'égalisant à zéro, on obtient

$$0 = \frac{d}{d\hat{\theta}} \int_{\Theta} w(\theta)(\theta - \hat{\theta})^2 \pi(\theta|x) d\nu(\theta)$$

$$\begin{aligned}
&= \frac{d}{d\hat{\theta}} \left(\int_{\Theta} w(\theta)\theta^2\pi(\theta|x)d\nu(\theta) - 2\hat{\theta} \int_{\Theta} w(\theta)\theta\pi(\theta|x)d\nu(\theta) + \hat{\theta}^2 \int_{\Theta} w(\theta)\pi(\theta|x)d\nu(\theta) \right) \\
&= -2 \int_{\Theta} w(\theta)\theta\pi(\theta|x)d\nu(\theta) + 2\hat{\theta} \int_{\Theta} w(\theta)\pi(\theta|x)d\nu(\theta),
\end{aligned}$$

d'où le résultat. Dans le cas particulier où $w(\theta) = 1$, c'est-à-dire pour la perte quadratique, l'estimateur de Bayes est l'espérance a posteriori de θ .

b) Pour la perte absolue $L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$, l'estimateur de Bayes $\hat{\theta}_{\pi}(x)$ correspond à la ou une médiane de $\pi(\theta|x)$ ([Ber85, Section 4.4.2]).

Les résultats suivants mettent en lien l'admissibilité, les estimateurs de Bayes et la minimaxité.

Théorème 1.41. *Si $\hat{\theta}_{\pi}(X)$ est un estimateur de Bayes unique pour la loi a priori $\pi(\theta)$ et que $r_{\pi}(\hat{\theta}_{\pi}) < \infty$, alors $\hat{\theta}_{\pi}(X)$ est admissible.*

Démonstration. Supposons au contraire que $\hat{\theta}_{\pi}$ soit inadmissible. Alors il existe un autre estimateur $\hat{\theta}_0$ tel que $R(\theta, \hat{\theta}_0) \leq R(\theta, \hat{\theta}_{\pi})$ pour tout $\theta \in \Theta$ avec une inégalité stricte pour au moins une valeur de θ , ce qui implique que $r_{\pi}(\hat{\theta}_0) \leq r_{\pi}(\hat{\theta}_{\pi})$. Donc $\hat{\theta}_{\pi}$ n'est pas un estimateur de Bayes unique. \square

Ainsi, dans notre modèle étudié, le théorème précédent sera utile afin d'établir l'admissibilité ou non des estimateurs obtenus. Le théorème suivant, quant à lui, permet de déduire la minimaxité à partir de certaines conditions.

Théorème 1.42. [Ber85] *Soit $\hat{\theta}$ un estimateur à risque constant, c'est-à-dire $R(\theta, \hat{\theta}) = c$ pour tout $\theta \in \Theta$. Alors*

- a) *si $\hat{\theta}$ est un estimateur de Bayes ou est admissible, alors $\hat{\theta}$ est minimax ;*
- b) *si $\hat{\theta}$ est un estimateur de Bayes (pas généralisé) unique, alors $\hat{\theta}$ est unique minimax.*

La définition suivante présente une généralisation de la notion d'estimateur de Bayes, qui permet également, dans certains cas, de déduire la minimaxité d'un estimateur.

Définition 1.43. *Un estimateur $\hat{\theta}_0$ est dit Bayes par extension lorsqu'il existe une suite de lois a priori $\{\pi_m\}_{m \geq 1}$ telles que*

$$\lim_{m \rightarrow \infty} \inf_{\hat{\theta}} r_{\pi_m}(\hat{\theta}) = \lim_{m \rightarrow \infty} r_{\pi_m}(\hat{\theta}_0).$$

Théorème 1.44 (Théorème 18, Chapitre 5, [Ber85], p.350). *Soit $\{\pi_m\}_{m \geq 1}$ une suite de lois a priori propres. S'il existe un estimateur $\hat{\theta}_0$ tel que, pour tout $\theta \in \Theta$,*

$$R(\theta, \hat{\theta}_0) \leq \lim_{m \rightarrow \infty} r_{\pi_m}(\hat{\theta}_0) < \infty,$$

alors $\hat{\theta}_0$ est minimax.

Remarque 1.45. *Du théorème précédent, on en déduit qu'un estimateur de Bayes par extension à risque constant est minimax.*

L'exemple suivant reprend le contexte de l'Exemple 1.35, mais en utilisant plutôt la perte quadratique pondérée $L(\mu, \hat{\mu}) = \frac{(\hat{\mu} - \mu)^2}{\mu}$. On montrera la minimaxité de l'estimateur $\hat{\mu}(X) = X$ selon deux approches, en utilisant quelques propriétés décrites ci-dessous.

Exemple 1.46. *Soit $X|\mu \sim \mathcal{P}(\mu)$. À l'Exemple 1.35, on a vu que, sous la perte quadratique $L(\mu, \hat{\mu}) = (\hat{\mu} - \mu)^2$, l'estimateur $\hat{\mu}(X) = X$ est admissible. Ainsi, on peut déduire que, pour la perte quadratique pondérée $L(\mu, \hat{\mu}) = \frac{(\hat{\mu} - \mu)^2}{\mu}$, qui est couramment utilisée pour les modèles impliquant des lois de Poisson¹, l'estimateur $\hat{\mu}(X) = X$ est également admissible. En effet, pour tout estimateur $\hat{\mu}(X)$, comme $R(\mu, X) \leq R(\mu, \hat{\mu})$, c'est-à-dire $E_\mu[(X - \mu)^2] \leq E_\mu[(\hat{\mu}(X) - \mu)^2]$, alors il est évident que*

$$E_\mu \left[\frac{(X - \mu)^2}{\mu} \right] \leq E_\mu \left[\frac{(\hat{\mu}(X) - \mu)^2}{\mu} \right],$$

1. Son usage courant s'explique par le fait que $\frac{(\hat{\mu} - \mu)^2}{\mu} = \frac{(\hat{\mu} - \mu)^2}{\text{Var}_\mu(X)}$.

avec une inégalité stricte pour au moins une valeur de μ . Avec cette perte quadratique pondérée, le risque devient

$$R(\mu, X) = E_\mu \left[\frac{(X - \mu)^2}{\mu} \right] = \frac{\text{Var}_\mu(X)}{\mu} = 1, \quad (1.30)$$

ce qui implique que le Théorème 1.42 est applicable, d'où le fait que X est minimax pour la perte $\frac{(\hat{\mu} - \mu)^2}{\mu}$.

Une autre manière de montrer la minimaxité de l'estimateur $\hat{\mu}(X) = X$ est de démontrer qu'il s'agit d'un estimateur de Bayes par extension et d'utiliser le Théorème 1.44. Considérons la suite de lois a priori propres $\mathcal{G}\left(1, \frac{1}{m}\right)$ pour μ , c'est-à-dire

$$\pi_m(\mu) = \frac{e^{-\mu/m}}{m}.$$

De l'Exemple 1.17, on sait que si $\mu \sim \mathcal{G}\left(1, \frac{1}{m}\right)$, alors $\mu|x \sim \mathcal{G}\left(1+x, 1+\frac{1}{m}\right)$. De plus, si $x > 0$, de l'Exemple 1.40, l'estimateur $\hat{\mu}_\pi(X)$ minimisant le risque intégré de Bayes est

$$\hat{\mu}_{\pi_m}(x) = \frac{1}{E_{\mu|x}\left[\frac{1}{\mu}\right]} = \frac{m}{m+1}x.$$

Si $x = 0$, alors l'estimateur de Bayes est $\hat{\mu}_{\pi_m}(x) = 0$, car l'intégrale

$$\int_0^\infty \frac{(\hat{\mu} - \mu)^2}{\mu} \left(1 + \frac{1}{m}\right) e^{-\mu(1+1/m)} d\mu$$

diverge à l'infini pour $\hat{\mu} \neq 0$. Donc pour tout $x \geq 0$, l'estimateur de Bayes est

$$\hat{\mu}_{\pi_m}(x) = \frac{m}{m+1}x.$$

Dans ce cas, sachant que $E^\mu[\mu] = m$, on a

$$\begin{aligned} r_{\pi_m}(\hat{\mu}_{\pi_m}) &= E^\mu E_{X|\mu} \left[\frac{\left(\frac{m}{m+1}X - \mu\right)^2}{\mu} \right] \\ &= E^\mu \left[\frac{1}{\mu} E_{X|\mu} \left[\left(\frac{m}{m+1}\right)^2 X^2 - 2\frac{m}{m+1}\mu X + \mu^2 \right] \right] \end{aligned}$$

$$\begin{aligned}
&= E^\mu \left[\left(\frac{m}{m+1} \right)^2 (1 + \mu) + \left(1 - \frac{2m}{m+1} \right) \mu \right] \\
&= \frac{m}{m+1}.
\end{aligned}$$

Comme $R(\mu, X) = 1 = \lim_{m \rightarrow \infty} \frac{m}{m+1}$, alors X est un estimateur de Bayes par extension à risque constant, et donc, minimax.

Par ailleurs, on peut montrer l'unicité de la minimaxité. Supposons qu'il existe un autre estimateur $\hat{\mu}_0$ minimax, c'est-à-dire

$$\sup_{\mu} R(\mu, \hat{\mu}_0) \leq \sup_{\mu} R(\mu, X) = 1.$$

Alors pour tout μ , $R(\mu, \hat{\mu}_0) \leq 1$. S'il existe une valeur de μ pour laquelle $R(\mu, \hat{\mu}_0) < 1$, cela contredirait l'admissibilité de X . Alors pour tout μ , $R(\mu, \hat{\mu}_0) = 1$. Comme la fonction de perte est strictement convexe, si $\hat{\mu}_0(X) \neq X$, on a

$$\begin{aligned}
\frac{\left(\left(\frac{1}{2} \hat{\mu}_0(X) + \frac{1}{2} X \right) - \mu \right)^2}{\mu} &< \frac{1}{2} \frac{(\hat{\mu}_0(X) - \mu)^2}{\mu} + \frac{1}{2} \frac{(X - \mu)^2}{\mu} \\
\implies R\left(\mu, \frac{1}{2} \hat{\mu}_0 + \frac{1}{2} X\right) &< \frac{1}{2} R(\mu, \hat{\mu}_0) + \frac{1}{2} R(\mu, X) = 1 = R(\mu, X),
\end{aligned}$$

ce qui contredirait l'admissibilité de X .

Dans l'exemple précédent, on a montré que, pour le modèle univarié $X|\mu \sim \mathcal{P}(\mu)$ et la perte quadratique pondérée $L(\mu, \hat{\mu}) = \frac{(\hat{\mu} - \mu)^2}{\mu}$, l'estimateur $\hat{\mu}(X) = X$ est Bayes par extension, admissible et unique minimax. Est-ce que ce résultat est généralisable pour le modèle poissonien bivarié ?

Exemple 1.47. [Ber85, Section 5.4, Exemple 20] Soient X_1, X_2, \dots, X_k des variables aléatoires indépendantes telles que $X_i|\mu_i \sim \mathcal{P}(\mu_i)$ pour $i = 1, \dots, k$, avec $k \geq 2$. Cleveland et Zidek ont montré, en 1975, que, pour estimer $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_k)'$ sous la

perte

$$L(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) = \sum_{i=1}^k \frac{(\hat{\mu}_i - \mu_i)^2}{\mu_i}, \quad (1.31)$$

l'estimateur $\hat{\boldsymbol{\mu}}(\mathbf{X}) = (X_1, X_2, \dots, X_k)'$ est inadmissible [CZ75], même s'il est minimax (non unique) à risque constant k , et donc, d'autres estimateurs minimax existent.

En particulier, dans le cas d'intérêt bivarié, si $(X_1, X_2) | \mu_1, \mu_2 \sim \mathcal{P}2(\mu_1, \mu_2)$ et que $L(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) = \frac{(\hat{\mu}_1 - \mu_1)^2}{\mu_1} + \frac{(\hat{\mu}_2 - \mu_2)^2}{\mu_2}$, on peut trouver un autre estimateur minimax qui domine l'estimateur $\hat{\boldsymbol{\mu}}(X_1, X_2) = (X_1, X_2)$. En effet, pour ce modèle, l'estimateur

$$\delta_{cz}(x_1, x_2) = \left(1 - \frac{c(x_1 + x_2)}{x_1 + x_2 + b}\right) x_1 + \left(1 - \frac{c(x_1 + x_2)}{x_1 + x_2 + b}\right) x_2, \quad (1.32)$$

avec

- i) $c(\cdot)$ est une fonction non décroissante ;
- ii) $0 \leq c(\cdot) \leq 2$;
- iii) $b \geq 1$,

est minimax [LC98, Théorème 6.8, Section 5.6]. Dans [LC98], on y présente en fait une généralisation pour k variables aléatoires indépendantes telles que décrites à l'Exemple 1.47. La démonstration est basée sur la loi multinomiale des variables aléatoires X_i conditionnellement à la somme $\sum_{i=1}^k X_i$, telle que montrée, dans le cas bivarié, à la Proposition 1.8.

Dans [KC91], Kushary et Cohen se sont par ailleurs intéressés à cette question d'admissibilité dans le cas où une contrainte d'ordre est imposée sur les paramètres d'intensité poissonniens. Ils y montrent notamment que si on considère les variables aléatoires indépendantes $X_1 | \lambda_1 \sim \mathcal{P}(\lambda_1)$ et $X_2 | \lambda_2 \sim \mathcal{P}(\lambda_2)$, en supposant que $0 \leq \lambda_1 \leq \lambda_2 < \infty$, alors l'estimateur $\hat{\lambda}(X_1, X_2) = X_1$ est inadmissible, sous la perte quadratique, pour estimer λ_1 . De l'Exemple 1.35, on avait pourtant vu que l'estimateur $\hat{\lambda}(X_1) = X_1$ (lorsque X_2 n'est pas observé) était admissible en l'absence d'information additionnelle sur λ_1 et λ_2 , ce qui peut paraître paradoxal en raison de l'indépendance des variables.

1.3 Inférence prédictive

Une branche de la statistique bayésienne s'intéresse à l'inférence prédictive, c'est-à-dire la prédiction d'un comportement futur aléatoire à l'aide d'observations présentement disponibles. Considérons les variables (ou vecteurs) aléatoires X et Y , qui ont pour densité respective p_θ et q_θ , où p et q sont connues. À des fins de simplification, on notera tout simplement $X \sim p_\theta$ et $Y \sim q_\theta$. À partir de l'observation X , on veut estimer la densité q_θ par la densité $\hat{q}(\cdot; x)$. On supposera pour la suite que X et Y sont des variables aléatoires indépendantes conditionnellement au paramètre θ .

1.3.1 Estimation de densité prédictive

Dans un cadre bayésien, pour le paramètre θ , on se fixe une loi *a priori* $\pi(\theta)$. À partir de celle-ci, il est possible de calculer la loi *a posteriori* de θ , donnée l'équation (1.10). Ainsi, un estimateur naturel et plausible pour q_θ , qui tient compte de l'information apportée par X , peut être déduite comme suit :

$$\begin{aligned}\hat{q}(y; x) &= \int_{\Theta} p(y, \theta|x) d\nu(\theta) \\ &= \frac{1}{m(x)} \int_{\Theta} p(x, y, \theta) d\nu(\theta) \\ &= \frac{1}{m(x)} \int_{\Theta} p_\theta(x, y) \pi(\theta) d\nu(\theta) \\ &= \frac{1}{m(x)} \int_{\Theta} p_\theta(x) q_\theta(y) \pi(\theta) d\nu(\theta) \\ &= \int_{\Theta} q_\theta(y) \pi(\theta|x) d\nu(\theta)\end{aligned}$$

où $\pi(\theta|x)$ est la densité *a posteriori* de θ étant donné x par rapport à une mesure σ -finie ν . On remarque qu'il s'agit en fait de $E_{\theta|x}[q_\theta(y)]$.

Définition 1.48. Soient $X \sim p_\theta$ et $\pi(\cdot)$ une loi *a priori* pour le paramètre θ par rapport

à une mesure σ -finie ν . La densité prédictive usuelle pour $Y \sim q_\theta$ est donnée par

$$\hat{q}_\pi(y; x) = \int_{\Theta} q_\theta(y) \pi(\theta|x) d\nu(\theta). \quad (1.33)$$

Exemple 1.49. Reprenons le contexte de l'Exemple 1.17.

On a le modèle $(X_1, X_2)|\lambda_1, \lambda_2 \sim \mathcal{P}2(\lambda_1 S, \lambda_2 T)$, $\lambda_1|\lambda_2 \sim \mathcal{G}(a, b)$ et $\lambda_2 \sim \mathcal{G}(c, d)$, où $S, T > 0$. On a montré que la loi a posteriori de (λ_1, λ_2) s'écrivait comme le produit de deux densités de lois gamma $\mathcal{G}(a + x_1, S + b)$ et $\mathcal{G}(c + x_2, T + d)$, soit

$$\pi(\lambda_1, \lambda_2|x_1, x_2) = \frac{(S + b)^{a+x_1} \lambda_1^{a+x_1-1} e^{-\lambda_1(S+b)}}{\Gamma(a + x_1)} \frac{(T + d)^{c+x_2} \lambda_2^{c+x_2-1} e^{-\lambda_2(T+d)}}{\Gamma(c + x_2)}.$$

Supposons que l'on veuille estimer la densité de $(Y_1, Y_2)|\lambda_1, \lambda_2 \sim \mathcal{P}2(\lambda_1 S', \lambda_2 T')$, où $S', T' > 0$ sont des constantes possiblement différentes de S et T . Alors, de l'équation (1.33), on obtient la densité prédictive suivante :

$$\begin{aligned} \hat{q}(\mathbf{y}; \mathbf{x}) &= \int_0^\infty \int_0^\infty \frac{e^{-S'\lambda_1} (S'\lambda_1)^{y_1}}{y_1!} \frac{e^{-T'\lambda_2} (T'\lambda_2)^{y_2}}{y_2!} \\ &\quad \times \frac{(S + b)^{a+x_1} \lambda_1^{a+x_1-1} e^{-\lambda_1(S+b)}}{\Gamma(a + x_1)} \frac{(T + d)^{c+x_2} \lambda_2^{c+x_2-1} e^{-\lambda_2(T+d)}}{\Gamma(c + x_2)} d\lambda_1 d\lambda_2 \\ &= \left(\frac{S'^{y_1} (S + b)^{a+x_1}}{y_1! \Gamma(a + x_1)} \int_0^\infty e^{-\lambda_1(S'+S+b)} \lambda_1^{a+x_1+y_1-1} d\lambda_1 \right) \\ &\quad \times \left(\frac{T'^{y_2} (T + d)^{c+x_2}}{y_2! \Gamma(c + x_2)} \int_0^\infty e^{-\lambda_2(T'+T+d)} \lambda_2^{c+x_2+y_2-1} d\lambda_2 \right) \\ &= \frac{S'^{y_1} (S + b)^{a+x_1}}{y_1! \Gamma(a + x_1)} \frac{\Gamma(a + x_1 + y_1)}{(S' + S + b)^{a+x_1+y_1}} \times \frac{T'^{y_2} (T + d)^{c+x_2}}{y_2! \Gamma(c + x_2)} \frac{\Gamma(c + x_2 + y_2)}{(T' + T + d)^{c+x_2+y_2}} \\ &= \frac{(a + x_1)_{y_1}}{y_1!} \left(\frac{S + b}{S' + S + b} \right)^{a+x_1} \left(\frac{S'}{S' + S + b} \right)^{y_1} \\ &\quad \times \frac{(c + x_2)_{y_2}}{y_2!} \left(\frac{T + d}{T' + T + d} \right)^{c+x_2} \left(\frac{T'}{T' + T + d} \right)^{y_2}, \end{aligned}$$

ce qui correspond au produit de fonctions de masse de lois binomiales négatives $\mathcal{BN}\left(a + x_1, \frac{S+b}{S'+S+b}\right)$ et $\mathcal{BN}\left(c + x_2, \frac{T+d}{T'+T+d}\right)$.

Remarque 1.50. Dans l'exemple précédent, les expressions obtenues sont également valables pour $-S < b \leq 0$ et $-T < d \leq 0$.

Une autre méthode afin d'estimer la densité q_θ est d'utiliser ce qu'on appelle des estimateurs par substitution (*plug-in estimators*). À partir de l'observation X , on trouve un estimateur $\hat{\theta}(X)$ pour θ et on estime la densité q_θ par $q_{\hat{\theta}}$.

1.3.2 Évaluation des estimateurs de densité prédictive

À l'image de l'estimation paramétrique, afin de mesurer l'efficacité d'une estimation de densité, on définit une fonction de perte $L(q, \hat{q})$ qui représente une « distance » entre les deux densités q et \hat{q} . Un choix judicieux et couramment utilisé est la perte de Kullback-Leibler [KL51], qui est définie par

$$L_{KL}(q_\theta, \hat{q}) = \int_{\mathcal{Y}} q_\theta(y) \log \frac{q_\theta(y)}{\hat{q}_\pi(y; x)} d\mu(y) = E_\theta \left[\log \frac{q_\theta(Y)}{\hat{q}_\pi(Y; x)} \right], \quad (1.34)$$

où μ est une mesure σ -finie sur l'espace échantillonnal de Y , q_θ est la densité que l'on cherche à estimer et $\hat{q}_\pi(y; x)$ est une densité prédictive. Cette fonction de perte peut être réécrite sous la forme

$$L_{KL}(q_\theta, \hat{q}) = \int_{\mathcal{Y}} q_\theta(y) \left(\frac{\hat{q}_\pi(y; x)}{q_\theta(y)} - \log \frac{\hat{q}_\pi(y; x)}{q_\theta(y)} - 1 \right) d\mu(y). \quad (1.35)$$

D'avantage évident sous cette forme, on remarque que cette fonction de perte appartient à une classe plus générale de fonctions de perte de la forme

$$\int_{\mathcal{Y}} q_\theta(y) \rho \left(\frac{\hat{q}_\pi(y; x)}{q_\theta(y)} \right) d\mu(y), \quad (1.36)$$

où ρ est décroissante sur $(0, 1)$ et croissante sur $(1, \infty)$ avec $\rho(t) \geq \rho(1) = 0$ pour tout $t > 0$. Il s'agit de fonctions de perte issues des travaux de Csiszar [Csi63] et Morimoto [Mor63] : les f -divergences de Csiszar-Morimoto, telles que décrites par exemple dans [CS04]. Pour la perte de Kullback-Leibler, on a $\rho(t) = t - \log(t) - 1$ pour tout $t > 0$.

Proposition 1.51. *La fonction de perte de Kullback-Leibler est non négative et nulle si et seulement si $\hat{q} = q_\theta$ presque partout.*

Démonstration. En vertu de l'inégalité de Jensen, comme $g(y) = -\log(y)$ est une fonction convexe, on a

$$\begin{aligned} L_{KL}(q_\theta, \hat{q}) &= E_\theta \left[-\log \frac{\hat{q}_\pi(Y; x)}{q_\theta(Y)} \right] \\ &\geq -\log E_\theta \left[\frac{\hat{q}_\pi(Y; x)}{q_\theta(Y)} \right] \\ &= -\log \int_{\mathcal{Y}} \frac{\hat{q}_\pi(y; x)}{q_\theta(y)} q_\theta(y) d\mu(y) \\ &= -\log 1 \\ &= 0. \end{aligned}$$

On a égalité si et seulement si $\frac{\hat{q}_\pi(y; x)}{q_\theta(y)}$ est constant avec probabilité un, c'est-à-dire si et seulement si $\hat{q}_\pi(y; x) \equiv q_\theta(y)$ presque partout. \square

Remarque 1.52. *On peut également le constater de l'équation (1.35), car $t - \log(t) - 1 \geq 0$ pour tout $t > 0$ et vaut 0 lorsque $t = 1$.*

Une autre classe de pertes fréquemment rencontrée est celle des pertes L_p intégrées, qui se définissent comme suit :

$$L_p(q_\theta, \hat{q}) = \int_{\mathcal{Y}} (q_\theta(y) - \hat{q}_\pi(y; x))^p d\mu(y), \quad p > 0. \quad (1.37)$$

Le cas $p = 2$ est l'analogie de la perte quadratique lors de l'estimation paramétrique. Il est évident que cette fonction de perte est non négative et qu'elle est nulle si $\hat{q}_\pi(y; x) \equiv q_\theta(y)$ presque partout. Tout comme pour l'estimation de paramètres vue dans les sections précédentes, en fixant une fonction de perte, on peut calculer le risque afin d'évaluer la performance d'un estimateur de densité $\hat{q}_\pi(y; x)$, qui constitue la perte espérée.

Définition 1.53. Soient $X|\theta \sim p_\theta$ et $Y|\theta \sim q_\theta$. Le risque d'un estimateur \hat{q} de la densité q_θ est

$$R(q_\theta, \hat{q}) = E_{X|\theta} [L(q_\theta, \hat{q})]. \quad (1.38)$$

Dans la section précédente, nous avons vu que si l'on se fixait une loi *a priori* $\pi(\theta)$ pour le paramètre θ à estimer, pour minimiser le risque intégré de Bayes, il suffisait de minimiser la perte espérée *a posteriori* $E_{\theta|x} [L(q_\theta, \hat{q})]$.

Théorème 1.54. Soient $X|\theta \sim p_\theta$, $Y|\theta \sim q_\theta$ indépendante de $X|\theta$, et $\pi(\cdot)$ une loi *a priori* pour θ par rapport à une mesure σ -finie ν pour laquelle la loi *a posteriori* existe. Alors pour la perte de Kullback-Leibler et la perte L_2 intégrée, l'estimateur de Bayes $\hat{q}_\pi(\cdot; X)$ est la densité prédictive définie à l'équation (1.33) par

$$\hat{q}_\pi(y; x) = \int_{\Theta} q_\theta(y) \pi(\theta|x) d\nu(\theta).$$

Démonstration. Pour plus de clarté, posons $q(y; x) = \int_{\Theta} q_\theta(y) \pi(\theta|x) d\nu(\theta)$. On veut donc minimiser $f(\hat{q}) = E_{\theta|x} [L(q_\theta, \hat{q})]$. Montrons alors que

$$\arg \min_{\hat{q}} f(\hat{q}) = q(y; x).$$

(i) Pour la perte de Kullback-Leibler, on a

$$\begin{aligned} f(\hat{q}) &= \int_{\Theta} \int_{\mathcal{Y}} q_\theta(y) \log \left(\frac{q_\theta(y)}{\hat{q}} \right) \pi(\theta|x) d\mu(y) d\nu(\theta) \\ &= \int_{\mathcal{Y}} \int_{\Theta} q_\theta(y) \log \left(\frac{q_\theta(y)}{q(y; x)} \frac{q(y; x)}{\hat{q}} \right) \pi(\theta|x) d\nu(\theta) d\mu(y) \\ &= \int_{\mathcal{Y}} \int_{\Theta} q_\theta(y) \log(q_\theta(y)) \pi(\theta|x) d\nu(\theta) d\mu(y) \\ &\quad - \int_{\mathcal{Y}} \int_{\Theta} q_\theta(y) \log(q(y; x)) \pi(\theta|x) d\nu(\theta) d\mu(y) \\ &\quad - \int_{\mathcal{Y}} \int_{\Theta} q_\theta(y) \log \left(\frac{\hat{q}}{q(y; x)} \right) \pi(\theta|x) d\nu(\theta) d\mu(y) \end{aligned}$$

$$\begin{aligned}
&= \int_{\mathcal{Y}} \int_{\Theta} q_{\theta}(y) \log(q_{\theta}(y)) \pi(\theta|x) d\nu(\theta) d\mu(y) \\
&\quad - \int_{\mathcal{Y}} \log(q(y;x)) q(y;x) d\mu(y) - \int_{\mathcal{Y}} q(y;x) \log\left(\frac{\hat{q}}{q(y;x)}\right) d\mu(y) \\
&\geq \int_{\mathcal{Y}} \int_{\Theta} q_{\theta}(y) \log(q_{\theta}(y)) \pi(\theta|x) d\nu(\theta) d\mu(y) \\
&\quad - \int_{\mathcal{Y}} \log(q(y;x)) q(y;x) d\mu(y) - \log\left(\int_{\mathcal{Y}} q(y;x) \frac{\hat{q}}{q(y;x)} d\mu(y)\right) \\
&= \int_{\mathcal{Y}} \int_{\Theta} q_{\theta}(y) \log(q_{\theta}(y)) \pi(\theta|x) d\nu(\theta) d\mu(y) - \int_{\mathcal{Y}} \log(q(y;x)) q(y;x) d\mu(y),
\end{aligned}$$

où l'inégalité s'explique par l'inégalité de Jensen appliquée à la fonction convexe $-\log$. Il y a égalité si et seulement si le quotient $\frac{\hat{q}}{q(y;x)}$ est constant, c'est-à-dire si et seulement si $\hat{q} = q(y;x)$ presque partout (puisque l'on veut que \hat{q} soit une densité).

(ii) Pour la perte L_2 intégrée, on a

$$\begin{aligned}
f(\hat{q}) &= \int_{\Theta} \int_{\mathcal{Y}} (\hat{q} - q_{\theta}(y))^2 \pi(\theta|x) d\mu(y) d\nu(\theta) \\
&= \int_{\mathcal{Y}} \left(\hat{q}^2 - 2\hat{q} \int_{\Theta} q_{\theta}(y) \pi(\theta|x) d\nu(\theta) + \int_{\Theta} q_{\theta}(y)^2 \pi(\theta|x) d\nu(\theta) \right) d\mu(y)
\end{aligned}$$

Pour tout x et pour tout y , on veut donc minimiser la quantité

$$\hat{q}^2 - 2\hat{q} \int_{\Theta} q_{\theta}(y) \pi(\theta|x) d\nu(\theta) + \int_{\Theta} q_{\theta}(y)^2 \pi(\theta|x) d\nu(\theta).$$

En dérivant par rapport à \hat{q} et en mettant l'expression résultante nulle, on a

$$2\hat{q} - 2 \int_{\Theta} q_{\theta}(y) \pi(\theta|x) d\theta = 0,$$

d'où $\hat{q} = \int_{\Theta} q_{\theta}(y) \pi(\theta|x) d\theta$ est la quantité qui minimise $E^{\theta|x}[L_2(q_{\theta}, \hat{q})]$. □

Il est d'ailleurs intéressant d'étudier la « calibration » sous \hat{q}_π de certaines quantités telles que l'espérance et la variance d'une variable aléatoire $Y|\theta \sim q_\theta$ à partir d'une observation indépendante $X|\theta \sim p_\theta$.

Théorème 1.55. *Soient $X|\theta \sim p_\theta$ et $Y|\theta \sim q_\theta$ indépendantes de même que $\theta \sim \pi$ telle que la loi a posteriori existe. Supposons que $E_\theta[Y] = \mu(\theta)$ et que $Var_\theta(Y) = \sigma^2(\theta)$. Alors*

$$E[Y] = E^\theta[\mu(\theta)] \quad (1.39)$$

et

$$Var(Y) = E^\theta[\sigma^2(\theta)] + Var^\theta(\mu(\theta)). \quad (1.40)$$

Démonstration. Comme $E[Y] = E^\theta[E_\theta[Y]]$ et $Var(Y) = E^\theta[Var_\theta(Y)] + Var^\theta(E_\theta[Y])$ (voir, par exemple, [CB02]), le résultat s'en suit. \square

A posteriori, les mêmes idées s'appliquent, mais en utilisant la loi a posteriori $\pi(\cdot|x)$.

Théorème 1.56. *Soient $X|\theta \sim p_\theta$ et $Y|\theta \sim q_\theta$ indépendantes de même que $\theta \sim \pi$ par rapport à une mesure σ -finie ν telle que la loi a posteriori existe. Supposons que $E_\theta[Y] = \mu(\theta)$ et que $Var_\theta(Y) = \sigma^2(\theta)$. À partir d'une observation $X = x$, si on estime q_θ par la densité prédictive usuelle $\hat{q}_\pi(y; x)$, alors*

$$E_{\hat{q}_\pi}[Y] = E_{\theta|x}[\mu(\theta)] \quad (1.41)$$

et

$$Var_{\hat{q}_\pi}(Y) = E_{\theta|x}[\sigma^2(\theta)] + Var_{\theta|x}(\mu(\theta)). \quad (1.42)$$

Démonstration. Si on utilise la densité prédictive usuelle $\hat{q}_\pi(y; x)$ pour estimer q_θ , alors

$$\begin{aligned} E_{\hat{q}_\pi}[Y] &= \int_{\mathcal{Y}} y \hat{q}_\pi(y; x) d\mu(y) \\ &= \int_{\mathcal{Y}} \int_{\Theta} y q_\theta(y) \pi(\theta|x) d\nu(\theta) d\mu(y) \end{aligned}$$

$$\begin{aligned}
&= \int_{\Theta} E_{\theta}[Y] \pi(\theta|x) d\nu(\theta) \\
&= E_{\theta|x}[E_{\theta}[Y]].
\end{aligned}$$

De manière analogue, on a $E_{\hat{q}_{\pi}}[Y^2] = E_{\theta|x}[E_{\theta}[Y^2]]$. Ainsi, pour la variance, on a

$$\begin{aligned}
\text{Var}_{\hat{q}_{\pi}}(Y) &= E_{\hat{q}_{\pi}}[Y^2] - (E_{\hat{q}_{\pi}}[Y])^2 \\
&= E_{\theta|x}[E_{\theta}[Y^2]] - (E_{\theta|x}[E_{\theta}[Y]])^2 \\
&= E_{\theta|x}[\sigma^2(\theta) + \mu(\theta)^2] - (E_{\theta|x}[\mu(\theta)])^2 \\
&= E_{\theta|x}[\sigma^2(\theta)] + E_{\theta|x}[\mu(\theta)^2] - (E_{\theta|x}[\mu(\theta)])^2 \\
&= E_{\theta|x}[\sigma^2(\theta)] + \text{Var}_{\theta|x}(\mu(\theta)). \quad \square
\end{aligned}$$

Exemple 1.57. Soient $(X_1, X_2)|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$, $(Y_1, Y_2)|\mu, \phi \sim \mathcal{P}2(\mu\phi S', \mu T')$ et $(\mu, \phi) \sim \pi$ pour laquelle la densité a posteriori existe, où $S, S', T, T' > 0$. D'une part, du Théorème 1.55, on sait que :

$$\begin{aligned}
E[Y_1] &= S' E^{\mu, \phi}[\mu\phi]; \\
E[Y_2] &= T' E^{\mu, \phi}[\mu]; \\
\text{Var}(Y_1) &= S' E^{\mu, \phi}[\mu\phi] + S'^2 \text{Var}^{\mu, \phi}(\mu\phi); \\
\text{Var}(Y_2) &= T' E^{\mu, \phi}[\mu] + T'^2 \text{Var}^{\mu, \phi}(\mu).
\end{aligned}$$

D'autre part, à partir du Théorème 1.56, on obtient des résultats analogues en calculant les espérances par rapport à la loi a posteriori $\pi(\mu, \phi|x_1, x_2)$.

Dans l'exemple précédent, il pourrait être également intéressant de calculer la covariance entre Y_1 et Y_2 .

Lemme 1.58. Soient X, Y et Z trois variables aléatoires telles que $E[X^2] < \infty$, $E[Y^2] < \infty$ et $E[Z^2] < \infty$. Alors

$$\text{Cov}(X, Y) = E[\text{Cov}(X, Y|Z)] + \text{Cov}(E[X|Z], E[Y|Z]). \quad (1.43)$$

Démonstration.

$$\begin{aligned}
Cov(X, Y) &= E[XY] - E[X]E[Y] \\
&= E[E[XY|Z]] - E[E[X|Z]]E[E[Y|Z]] \\
&= E[E[XY|Z]] - E[E[X|Z]E[Y|Z]] \\
&\quad + E[E[X|Z]E[Y|Z]] - E[E[X|Z]]E[E[Y|Z]] \\
&= E[Cov(X, Y|Z)] + Cov(E[X|Z], E[Y|Z]). \quad \square
\end{aligned}$$

Théorème 1.59. Soient $(X_1, X_2)|\theta \sim p_\theta$ et $(Y_1, Y_2)|\theta \sim q_\theta$ indépendantes, $Y_1|\theta$ et $Y_2|\theta$ également indépendantes, et $\theta \sim \pi$ telle que la loi a posteriori existe. Supposons que $E_\theta[Y_i] = \mu_i(\theta)$ et que $Var_\theta(Y_i) = \sigma_i^2(\theta)$ pour $i = 1, 2$. Alors

$$Cov(Y_1, Y_2) = Cov(\mu_1(\theta), \mu_2(\theta)). \quad (1.44)$$

De plus, à partir d'une observation (x_1, x_2) , si on estime q_θ par la densité prédictive usuelle $\hat{q}_\pi(y_1, y_2; x_1, x_2)$, alors

$$Cov_{\hat{q}_\pi}(Y_1, Y_2) = Cov_{\theta|x_1, x_2}(\mu_1(\theta), \mu_2(\theta)). \quad (1.45)$$

Démonstration. L'équation (1.44) s'obtient du Lemme 1.58 et du fait que $Y_1|\theta$ et $Y_2|\theta$ sont indépendantes.

Pour l'équation (1.45), par un raisonnement similaire à la démonstration du Théorème 1.56, on a

$$\begin{aligned}
Cov_{\hat{q}_\pi}(Y_1, Y_2) &= E_{\hat{q}_\pi}[Y_1 Y_2] - E_{\hat{q}_\pi}[Y_1]E_{\hat{q}_\pi}[Y_2] \\
&= E_{\theta|x_1, x_2}[E_\theta[Y_1 Y_2]] - E_{\theta|x_1, x_2}[E_\theta[Y_1]]E_{\theta|x_1, x_2}[E_\theta[Y_2]] \\
&= E_{\theta|x_1, x_2}[\mu_1(\theta)\mu_2(\theta)] - E_{\theta|x_1, x_2}[\mu_1(\theta)]E_{\theta|x_1, x_2}[\mu_2(\theta)] \\
&= Cov_{\theta|x_1, x_2}(\mu_1(\theta), \mu_2(\theta)). \quad \square
\end{aligned}$$

Exemple 1.60. *Considérons le modèle bivarié d'intérêt $(X_1, X_2)|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$, $(Y_1, Y_2)|\mu, \phi \sim \mathcal{P}2(\mu\phi S', \mu T')$ et $(\mu, \phi) \sim \pi$ pour laquelle la densité a posteriori existe, où $S, S', T, T' > 0$. Du Théorème 1.59, on sait a que*

$$\begin{aligned} \text{Cov}(Y_1, Y_2) &= S'T' \text{Cov}(\mu\phi, \mu) \\ &= S'T'(E[\mu^2\phi] - E[\mu\phi]E[\mu]). \end{aligned}$$

De plus, sous la densité prédictive usuelle $\hat{q}_\pi(y_1, y_2; x_1, x_2)$, on obtient un résultat analogue, mais basé sur des moments mixtes a posteriori, en calculant la covariance par rapport à la loi a posteriori $\pi(\mu, \phi|x_1, x_2)$:

$$\begin{aligned} \text{Cov}_{\hat{q}_\pi}(Y_1, Y_2) &= S'T' \text{Cov}_{\mu, \phi|x_1, x_2}(\mu\phi, \mu) \\ &= S'T'(E_{\mu, \phi|x_1, x_2}[\mu^2\phi] - E_{\mu, \phi|x_1, x_2}[\mu\phi]E_{\mu, \phi|x_1, x_2}[\mu]). \end{aligned}$$

CHAPITRE 2

Répertoire de lois bêta

Une distribution fréquemment rencontrée en statistique bayésienne comme loi *a priori* est la loi bêta à deux paramètres, dite parfois bêta de type I puisqu'elle est basée sur l'intégrale d'Euler du premier type. Cette loi, notée ici $\mathcal{B}(a, b)$, a pour fonction de densité

$$f(x) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1} \mathbf{1}_{[0,1]}(x), \quad a, b > 0, \quad (2.1)$$

où $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ est la fonction bêta. L'intérêt porté cette loi est entre autres attribuable à l'éventail de formes de densité possibles, et, comme il a été montré à l'Exemple 1.20, au fait que ce soit la famille conjuguée de la famille de lois binomiales (voir [JKB94, Chapitre 25] pour plus de détails sur cette distribution). De plus, les lois bêta constituent un choix attrayant pour modéliser des variables aléatoires prenant des valeurs sur un intervalle quelconque $[c, d]$. En effet, à partir d'une variable aléatoire $X \sim \mathcal{B}(a, b)$, dont le support est restreint à l'intervalle $[0, 1]$, on peut construire une variable aléatoire de la forme $X' = (d - c)X + c$, dont le support est l'intervalle $[c, d]$. En outre, le changement de variable $Y = \frac{X}{1-X}$ permet de travailler avec une variable aléatoire dont le support constitue les réels positifs, et ainsi d'introduire la loi bêta de

type II (sans paramètre d'échelle), notée $\mathcal{B}_2(c, d, 1)$, de fonction de densité

$$f(y) = \frac{1}{B(c, d)} \frac{y^{c-1}}{(1+y)^{c+d}} \mathbb{1}_{[0, \infty)}(y), \quad c, d > 0. \quad (2.2)$$

On remarque que cette loi, également bien connue, se présente parfois dans la littérature comme étant la loi dite bêta prime et notée $\mathcal{B}'(c, d)$. Il s'agit d'une appellation qui ne sera toutefois pas utilisée dans ce mémoire.

Des généralisations des lois bêta et des lois bêta de type II ont été développées par le passé et seront rencontrées dans les prochains chapitres de ce mémoire pour le modèle bivarié poissonnien à l'étude. Ainsi, l'objectif de ce chapitre est de mettre en lumière la construction de ces lois et de proposer de nouvelles représentations en termes de mélanges pour les généralisations de ces lois bêta (Proposition 2.15, Corollaire 2.16 et Corollaire 2.21). Nous présenterons également leurs principales caractéristiques. Alors que ces lois ont été travaillées à de nombreuses reprises, parfois présentées sous différentes formes selon les auteurs, ce chapitre a également pour but d'unifier la théorie qui s'y rattache. Pour ce faire, il importe avant tout d'introduire une classe de fonction particulières qui jouera un rôle important dans la généralisation des lois bêta : les fonctions hypergéométriques.

2.1 Fonctions hypergéométriques

Les fonctions hypergéométriques sont très présentes dans plusieurs domaines des mathématiques appliquées [Sea13]. En leur forme la plus générale, elles s'écrivent comme une série de la forme suivante :

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; x) = \sum_{n=0}^{\infty} \frac{(a_1)_n (a_2)_n \cdots (a_p)_n}{(b_1)_n (b_2)_n \cdots (b_q)_n} \frac{x^n}{n!}, \quad x, a_i, b_j \in \mathbb{R}, \quad (2.3)$$

où p correspond au nombre de paramètres a_i au numérateur, q , au nombre de paramètres b_j au dénominateur, et où les termes $(a_i)_n$ et $(b_j)_n$ sont les symboles de Pochhammer

appliqués respectivement aux paramètres a_i et b_j , pour $i = 1, \dots, p$ et $j = 1, \dots, q$. Le cas $p = 2$ et $q = 1$ est d'un intérêt particulier.

Définition 2.1. Soient $a, b, c, x \in \mathbb{R}$ avec $c \notin -\mathbb{N} \cup \{0\}$. La fonction hypergéométrique de Gauss, ou simplement fonction hypergéométrique, est une série qui s'écrit sous la forme

$${}_2F_1(a, b; c; x) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{x^n}{n!}. \quad (2.4)$$

La série définie ci-dessus converge absolument pour $|x| < 1$. C'est également le cas si $x = 1$, à condition que $c - (a + b) > 0$, et si $x = -1$, à condition que $c - (a + b) + 1 > 0$. Pour plus de détails, voir [Sla66, Section 1.1.1].

Parmi plusieurs, Euler a beaucoup travaillé sur la fonction hypergéométrique [AAR99], ce dernier ayant entre autres trouvé une représentation de cette série sous la forme d'une intégrale [Eul69]. En effet, pour $|x| < 1$, $a > 0$ et $c - a > 0$, on a

$${}_2F_1(a, b; c; x) = \frac{1}{B(a, c - a)} \int_0^1 \frac{t^{a-1} (1-t)^{c-a-1}}{(1-xt)^b} dt. \quad (2.5.a)$$

L'idée de ce passage d'une série à une intégrale réside dans le développement en série de Maclaurin de la fonction $(1-xt)^{-b}$. De plus, en appliquant à l'intégrale (2.5.a) le même changement de variables qui permet le passage de la loi $\mathcal{B}(a, b)$ à la loi $\mathcal{B}_2(c, d, 1)$ (avec $a = c$ et $b = d$), c'est-à-dire $s = \frac{t}{1-t}$, on obtient une deuxième représentation de la fonction hypergéométrique sous forme d'intégrale :

$${}_2F_1(a, b; c; x) = \frac{1}{B(a, c - a)} \int_0^{\infty} \frac{s^{a-1} (1+s)^{b-c}}{(1+(1-x)s)^b} ds. \quad (2.5.b)$$

Via d'autres transformations de la fonction hypergéométrique, qui nous seront très utiles dans les prochains chapitres, il est possible de dériver les trois équivalences présentées dans la proposition ci-dessous. Celles-ci sont attribuables à Pfaff et à Euler.

Proposition 2.2. On a :

a) *Transformation de Pfaff*

$${}_2F_1(a, b; c; x) = (1-x)^{-a} {}_2F_1\left(a, c-b; c; \frac{x}{x-1}\right) \quad (2.6.a)$$

$$= (1-x)^{-b} {}_2F_1\left(c-a, b; c; \frac{x}{x-1}\right); \quad (2.6.b)$$

b) *Transformation d'Euler*

$${}_2F_1(a, b; c; x) = (1-x)^{c-a-b} {}_2F_1(c-a, c-b; c, x). \quad (2.7)$$

Démonstration. La transformation de Pfaff découle directement de l'équation (2.5.a) à laquelle on applique le changement de variables $s = 1 - t$. Comme a et b sont interchangeables pour ${}_2F_1(a, b; c; x)$, on obtient les deux formes présentées aux équations (2.6.a) et (2.6.b). En appliquant successivement ces deux transformations, on obtient la transformation d'Euler donnée à l'équation (2.7). \square

On rappelle que la fonction hypergéométrique ${}_2F_1(a, b; c; x)$ converge absolument pour $|x| < 1$. Ainsi, le côté droit des équations associées à la transformation de Pfaff converge pour $\left|\frac{x}{x-1}\right| < 1$, c'est-à-dire pour $x < 1/2$. Cette transformation fournit donc une extension de la définition de la fonction hypergéométrique en la prolongeant sur l'ensemble des réels négatifs. Pour plus de détails, voir [AAR99, Chapitre 2].

Appell a étudié une généralisation en double série des fonctions hypergéométriques de Gauss basée sur le produit de deux d'entre elles. La construction de cette généralisation à partir des fonctions hypergéométriques de Gauss est expliquée, par exemple, dans [Sla66, Chapitre 8]. Bien que cette étude ait mené à cinq fonctions particulières, c'est-à-dire les quatre fonctions hypergéométriques dites d'Appell, de même qu'un cas particulier d'une fonction hypergéométrique, nous n'en aurons besoin que de deux pour ce mémoire. La première est présentée au lemme suivant.

Lemme 2.3. *Pour $a, b, c > 0$, on a*

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{(a)_{m+n}(b)_{m+n}}{(c)_{m+n}} \frac{x^m y^n}{m!n!} = {}_2F_1(a, b; c; x + y). \quad (2.8)$$

Démonstration.

$$\begin{aligned} {}_2F_1(a, b; c; x + y) &= \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{(x + y)^k}{k!} \\ &= \sum_{k=0}^{\infty} \sum_{m=0}^k \frac{(a)_k (b)_k}{(c)_k} \frac{x^m y^{k-m}}{m!(k-m)!} \\ &= \sum_{m=0}^{\infty} \sum_{k=m}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{x^m y^{k-m}}{m!(k-m)!} \\ &= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{(a)_{m+n} (b)_{m+n}}{(c)_{m+n}} \frac{x^m y^n}{m!n!}. \quad \square \end{aligned}$$

La deuxième fonction en double séries que nous aurons besoin dans ce mémoire est la première fonction hypergéométrique d'Appell, définie ci-dessous.

Définition 2.4. *Soient $a, b, b', x, y \in \mathbb{R}$ et $c \in \mathbb{R}^* \setminus -\mathbb{N}$. La première fonction hypergéométrique d'Appell est une double série qui s'écrit sous la forme*

$$F_1(a; b, b'; c; x, y) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{(a)_{m+n} (b)_m (b')_n}{(c)_{m+n}} \frac{x^m y^n}{m!n!}. \quad (2.9)$$

La première fonction hypergéométrique d'Appell converge absolument pour $|x| < 1$ et $|y| < 1$ (voir [Bai35, Chapitre 9]) et elle sera utilisée pour la construction d'une loi bivariée utile au modèle poissonien à l'étude, et ceci, grâce aux propriétés présentées ci-dessous. Celles-ci proviennent entre autres de la Propriété 1.16.b) du symbole de Pochhammer et de la représentation intégrale suivante de la fonction F_1 pour $a > 0$ et $c - a > 0$ [Bai35, Section 9.3] :

$$F_1(a; b, b'; c; x, y) = \frac{1}{B(a, c - a)} \int_0^1 \frac{z^{a-1} (1 - z)^{c-a-1}}{(1 - zx)^b (1 - zy)^{b'}} dz. \quad (2.10)$$

En appliquant le changement de variable $s = \frac{z}{1-z}$, on obtient également cette deuxième représentation intégrale :

$$F_1(a; b, b'; c; x, y) = \frac{1}{B(a, c-a)} \int_0^\infty \frac{s^{a-1}(1+s)^{b+b'-c}}{(1+(1-x)s)^b(1+(1-y)s)^{b'}} ds. \quad (2.11)$$

Dans [Bai35, Section 9.4], on expose quelques changements de variables, qui, appliqués à l'intégrale (2.10), mènent aux transformations suivantes :

$$F_1(a; b, b'; c; x, y) = (1-x)^{-b}(1-y)^{-b'} F_1\left(c-a; b, b'; c; -\frac{x}{1-x}, -\frac{y}{1-y}\right) \quad (2.12.a)$$

$$= (1-x)^{-a} F_1\left(a; c-b-b', b'; c; -\frac{x}{1-x}, \frac{y-x}{1-x}\right) \quad (2.12.b)$$

$$= (1-y)^{-a} F_1\left(a; b, c-b-b'; c; \frac{x-y}{1-y}, -\frac{y}{1-y}\right) \quad (2.12.c)$$

$$= (1-x)^{c-a-b}(1-y)^{-b'} F_1\left(c-a; c-b-b', b'; c; x, \frac{x-y}{1-y}\right) \quad (2.12.d)$$

$$= (1-x)^{-b}(1-y)^{c-a-b'} F_1\left(c-a; b, c-b-b'; c; \frac{y-x}{1-x}, y\right). \quad (2.12.e)$$

Ces transformations peuvent être vues comme les analogues de la transformation de Pfaff présentée à la Proposition 2.2.

Les propriétés suivantes montrent comment on peut réécrire la première fonction hypergéométrique d'Appell à l'aide de la fonction hypergéométrique de Gauss et présentent quelques cas particuliers de réduction de cette fonction.

Propriétés 2.5. [Bai35, Section 9.5] On a :

$$a) \quad F_1(a; b, b'; c; x, y) = \sum_{m=0}^{\infty} \frac{(a)_m (b)_m}{(c)_m} \frac{x^m}{m!} {}_2F_1(a+m, b'; c+m; y);$$

$$b) \quad F_1(a; b, b'; c; x, x) = {}_2F_1(a, b+b'; c; x);$$

$$c) \quad F_1(a; b, b'; b + b'; x, y) = (1 - y)^{-a} {}_2F_1\left(a, b; b + b'; \frac{x - y}{1 - y}\right).$$

Un autre choix intéressant de valeurs p et q pour la fonction hypergéométrique de forme générale ${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; x)$ est $p = q = 1$.

Définition 2.6. Soient $a, b, x \in \mathbb{R}$ avec $b \notin -\mathbb{N} \cup \{0\}$. La fonction hypergéométrique confluente est une série qui s'écrit sous la forme

$${}_1F_1(a; b; x) = \sum_{n=0}^{\infty} \frac{(a)_n x^n}{(b)_n n!}. \quad (2.13)$$

Cette série converge pour tout $x \in \mathbb{R}$, et tout comme les autres fonctions hypergéométriques présentées précédemment, la fonction hypergéométrique confluente admet une représentation sous forme d'intégrale [DLM] :

$${}_1F_1(a; b; x) = \frac{1}{B(a, b - a)} \int_0^1 e^{xt} t^{a-1} (1 - t)^{b-a-1} dt. \quad (2.14)$$

Pour obtenir ce résultat, il suffit de développer la fonction e^{xt} en série :

$$\begin{aligned} \frac{1}{B(a, b - a)} \int_0^1 e^{xt} t^{a-1} (1 - t)^{b-a-1} dt &= \frac{\Gamma(b)}{\Gamma(a)\Gamma(b - a)} \int_0^1 \sum_{n=0}^{\infty} \frac{(xt)^n}{n!} t^{a-1} (1 - t)^{b-a-1} dt \\ &= \sum_{n=0}^{\infty} \frac{\Gamma(b)}{\Gamma(a)\Gamma(b - a)} \frac{x^n}{n!} \int_0^1 t^{a+n-1} (1 - t)^{b-a-1} dt \\ &= \sum_{n=0}^{\infty} \frac{\Gamma(b)\Gamma(a + n)\Gamma(b - a)}{\Gamma(a)\Gamma(b - a)\Gamma(b + n)} \frac{x^n}{n!} \\ &= \sum_{n=0}^{\infty} \frac{(a)_n x^n}{(b)_n n!} \end{aligned}$$

Enfin, la propriété ci-dessous se démontre directement à l'aide de cette représentation sous forme d'intégrale.

Propriété 2.7. [DLM] Première transformation de Kummer :

$${}_1F_1(a; b; x) = e^x {}_1F_1(b - a; b; -x). \quad (2.15)$$

2.2 Distributions bêta généralisées

Bien que la loi bêta soit grandement utilisée comme distribution *a priori* pour le modèle binomial, Chen et Novick y voient quelques désavantages, discutés dans [CN84], qui motivent l'utilisation d'une généralisation de cette loi. En effet, ils évoquent entre autres le fait que, pour plusieurs de ces distributions, leurs queues sont généralement très minces en raison d'une croissance ou décroissance rapide de la densité. De plus, un autre point soulevé relève du fait que l'espérance et le coefficient d'asymétrie (*Skewness*) sont intimement liés. En effet, pour $X \sim \mathcal{B}(a, b)$, on rappelle (voir par exemple [JKB94, Chapitre 25]) que

$$E[X] = \frac{a}{a+b},$$

et

$$Skew(X) = \frac{2(b-a)\sqrt{a+b+1}}{(a+b+2)\sqrt{ab}}.$$

Ainsi, si $E[X] = 1/2$, alors $a = b$ et la distribution est symétrique. Si $E[X] < 1/2$, alors $a < b$ et le *Skewness* est positif. Si $E[X] > 1/2$, alors $a > b$ et le *Skewness* est négatif. Cette étroite relation entre l'espérance et le coefficient d'asymétrie peut parfois affecter la qualité de l'ajustement du modèle par rapport à la réalité. Donc l'ajout de paramètres viendra corriger cette situation.

Dans [LN82], Libby et Novick décrivent une loi bêta généralisée à trois paramètres qui résulte en fait de l'addition d'un paramètre à la loi bêta standard. Elle offre entre autres un plus grand éventail de formes de densité possibles [GN04] et permet davantage de possibilités pour les valeurs de l'espérance, de la variance, du *Skewness* et du *Kurtosis* [CN84].

Définition 2.8. *On dit que X se distribue selon une loi bêta généralisée à trois para-*

mètres, que l'on notera $\mathcal{BG3}(a, b, \lambda)$, si sa fonction de densité est de la forme

$$f(x) = \frac{\lambda^a}{B(a, b)} \frac{x^{a-1}(1-x)^{b-1}}{(1-(1-\lambda)x)^{a+b}} \mathbf{1}_{[0,1]}(x), \quad a, b, \lambda > 0. \quad (2.16)$$

Il est évident que pour $\lambda = 1$, la distribution $\mathcal{BG3}(a, b, \lambda)$ se simplifie à la distribution $\mathcal{B}(a, b)$. En outre, le lemme suivant permet d'expliquer un rôle du paramètre λ à travers la transformation du passage des lois bêta de type I aux lois bêta de type II.

Lemme 2.9. *Si $X \sim \mathcal{B}(a, b)$ et $Y \sim \mathcal{BG3}(a, b, \lambda)$, alors*

$$\frac{Y}{1-Y} \stackrel{d}{=} \frac{1}{\lambda} \frac{X}{1-X}. \quad (2.17)$$

Remarque 2.10. *On rappelle que $\frac{X}{1-X} \sim \mathcal{B}_2(a, b, 1)$, et donc le paramètre $1/\lambda$ est un paramètre d'échelle pour les lois bêta de type II associées, que l'on notera alors $\mathcal{B}_2(c, d, \tau)$, avec $a = c, b = d$ et $\tau = 1/\lambda$. Il s'agit en fait d'un cas particulier du Lemme 2.19 qui sera détaillé dans la prochaine section.*

Démonstration. Pour $s \geq 0$, on a $\mathbb{P}\left(\frac{Y}{1-Y} \leq s\right) = \mathbb{P}\left(Y \leq \frac{s}{1+s}\right)$. Alors

$$\begin{aligned} f_{\frac{Y}{1-Y}}(s) &= f_Y\left(\frac{s}{1+s}\right) \frac{1}{(1+s)^2} \\ &= \frac{\lambda^a}{B(a, b)} \frac{\left(\frac{s}{1+s}\right)^{a-1} \left(1 - \frac{s}{1+s}\right)^{b-1}}{\left(1 - (1-\lambda)\left(\frac{s}{1+s}\right)\right)^{a+b}} \frac{1}{(1+s)^2} \\ &= \frac{1}{B(a, b)} \frac{\left(\frac{s}{(1/\lambda)}\right)^{a-1}}{\left(1 + \frac{s}{(1/\lambda)}\right)^{a+b}} \frac{1}{(1/\lambda)} \\ &= f_{\frac{X}{1-X}}\left(\frac{s}{(1/\lambda)}\right) \frac{1}{(1/\lambda)}. \quad \square \end{aligned}$$

Dans [PGD89], Pham-Gia et Duong donne une interprétation géométrique à la valeur de λ qui est basée sur le lemme suivant.

Corollaire 2.11. *Soit $\lambda > 0$. Si $X \sim \mathcal{B}(a, b)$, alors $Y = \frac{X}{\lambda + (1-\lambda)X} \sim \mathcal{BG3}(a, b, \lambda)$.*

Démonstration. Le résultat suit du Lemme 2.9, car de l'équation (2.17), on a

$$Y = \frac{\frac{1}{\lambda} \frac{X}{1-X}}{\frac{1}{\lambda} \frac{X}{1-X} + 1} = \frac{X}{\lambda + (1-\lambda)X}. \quad \square$$

La loi bêta généralisée à trois paramètres peut également être obtenue à partir de deux variables aléatoires indépendantes de distributions gamma.

Proposition 2.12. [JKB94, Chapitre 25] *Soient $X_1 \sim \mathcal{G}(\alpha_1, \beta_1)$ et $X_2 \sim \mathcal{G}(\alpha_2, \beta_2)$ deux variables aléatoires indépendantes. Alors $\frac{X_1}{X_1 + X_2} \sim \mathcal{BG3}\left(\alpha_1, \alpha_2, \frac{\beta_1}{\beta_2}\right)$.*

Démonstration. Posons $\lambda = \frac{\beta_1}{\beta_2}$. On a $\lambda X_1 \sim \mathcal{G}(\alpha_1, \beta_2)$. Puis, du Lemme 1.21, on a $\frac{X_2}{\lambda X_1} \sim \mathcal{B}_2(\alpha_2, \alpha_1, 1)$, et donc, $U = \frac{\lambda X_1}{\lambda X_1 + X_2} \sim \mathcal{B}(\alpha_1, \alpha_2)$. En effet, si on pose $R = \frac{X_2}{\lambda X_1}$, alors la fonction de densité de $Y = \frac{1}{1+R}$ se calcule comme suit :

$$f_Y(y) = f_R\left(\frac{1-y}{y}\right) \cdot \left|\frac{-1}{y^2}\right| \propto y^{\alpha_1-1}(1-y)^{\alpha_2-1}, \quad 0 \leq y \leq 1.$$

Ainsi, du Corollaire 2.11, $\frac{U}{\lambda + (1-\lambda)U} \sim \mathcal{BG3}(\alpha_1, \alpha_2, \lambda)$, c'est-à-dire

$$\frac{\frac{\lambda X_1}{\lambda X_1 + X_2}}{\lambda + (1-\lambda)\left(\frac{\lambda X_1}{\lambda X_1 + X_2}\right)} = \frac{X_1}{X_1 + X_2} \sim \mathcal{BG3}(\alpha_1, \alpha_2, \lambda). \quad \square$$

D'un point de vue bayésien, pour Chen et Novick [CN84], la loi bêta généralisée à trois paramètres prend entre autres son sens comme loi *a priori* pour le modèle binomial. L'étude de ce modèle bayésien les a alors menés vers une autre généralisation de la loi bêta, mais cette fois-ci, à quatre paramètres. En effet, si $X|p \sim \mathcal{Bin}(n, p)$ et $p \sim \mathcal{BG3}(a, b, \lambda)$, alors

$$\pi(p|x) \propto \frac{p^{a+x-1}(1-p)^{b+n-x-1}}{(1-(1-\lambda)p)^{a+b}} \mathbb{1}_{[0,1]}(p),$$

et en reparamétrisant cette loi *a posteriori*, on obtient

$$\pi(p|x) \propto \frac{p^{a'-1}(1-p)^{b'-1}}{(1-(1-\lambda)p)^\kappa} \mathbb{1}_{[0,1]}(p), \quad a', b', \lambda, \kappa > 0.$$

On remarque qu'il s'agit en fait de la même loi introduite à l'Exemple 1.20 et qui forme une famille conjuguée pour la famille des lois binomiales.

Définition 2.13. *On dit que X se distribue selon une loi bêta généralisée à quatre paramètres, que l'on notera $\mathcal{BG4}(a, b, \lambda, \kappa)$, si sa fonction de densité est de la forme*

$$f(x) = \frac{1}{B(a, b)_2F_1(a, \kappa; a + b; 1 - \lambda)} \frac{x^{a-1}(1-x)^{b-1}}{(1-(1-\lambda)x)^\kappa} \mathbb{1}_{[0,1]}(x), \quad a, b, \lambda, \kappa > 0. \quad (2.18)$$

Cette fonction de densité est en fait un cas particulier des lois bêta généralisées présentées en 1976 par Exton dans [Ext76, Chapitre 7]. Elle est également utilisée dans les travaux d'Armero et Bayarri [AB94], sous l'appellation de la distribution hypergéométrique de Gauss. En outre, on remarque qu'en posant $\kappa = a + b$, cette distribution se simplifie à la distribution $\mathcal{BG3}(a, b, \lambda)$ dont la fonction de densité est donnée par l'équation (2.16). La correspondance au niveau de la constante de normalisation est obtenue puisque

$${}_2F_1(a, a + b; a + b; 1 - \lambda) = \sum_{n=0}^{\infty} \frac{(a)_n}{n!} (1 - \lambda)^n = \lambda^{-a}.$$

Par ailleurs, on peut facilement montrer que $X \sim \mathcal{B}(a, b)$ si et seulement si $1 - X \sim \mathcal{B}(b, a)$. Ce principe de réflexion se généralise de la façon suivante pour la loi bêta généralisée à quatre paramètres.

Lemme 2.14. *Soit $X \sim \mathcal{BG4}(a, b, \lambda, \kappa)$. Alors $1 - X \sim \mathcal{BG4}(b, a, 1/\lambda, \kappa)$.*

Démonstration. Soit $X \sim \mathcal{BG4}(a, b, \lambda, \kappa)$. À l'aide de la transformation de Pfaff donnée à l'équation (2.6.b), on peut réécrire la densité, pour $x \in [0, 1]$, comme suit :

$$f(x) = \frac{1}{B(a, b)_2F_1(a, \kappa; a + b; 1 - \lambda)} \frac{x^{a-1}(1-x)^{b-1}}{(1-(1-\lambda)x)^\kappa}$$

$$\begin{aligned}
&= \frac{1}{B(a, b) {}_2F_1(b, \kappa; a + b; 1 - 1/\lambda) \lambda^{-\kappa}} \frac{(1 - (1 - x))^{a-1} (1 - x)^{b-1}}{(1 - (1 - \lambda)x)^\kappa} \\
&= \frac{1}{B(a, b) {}_2F_1(b, \kappa; a + b; 1 - 1/\lambda)} \frac{(1 - x)^{b-1} (1 - (1 - x))^{a-1}}{(1 - (1 - 1/\lambda)(1 - x))^\kappa}. \quad \square
\end{aligned}$$

Les distributions bêta généralisées à quatre paramètres peuvent aussi être représentées comme des mélanges de lois bêta.

Proposition 2.15. *Soit $X \sim \mathcal{BG4}(a, b, \lambda, \kappa)$. Alors, selon la valeur de λ , X est caractérisé par les mélanges suivants :*

a) *Pour $0 < \lambda \leq 1$, $X|J = j \sim \mathcal{B}(a + j, b)$, où J est une variable aléatoire discrète sur*

$$\mathbb{N} \text{ avec fonction de masse } p(j) = \frac{1}{{}_2F_1(a, \kappa; a + b; 1 - \lambda)} \frac{(\kappa)_j (a)_j}{(a + b)_j} \frac{(1 - \lambda)^j}{j!} \mathbb{1}_{\mathbb{N}}(j);$$

b) *Pour $\lambda > 1$, $X|L = l \sim \mathcal{B}(b + l, a)$, où L est une variable aléatoire discrète sur \mathbb{N} avec*

$$\text{fonction de masse } p(l) = \frac{1}{{}_2F_1(b, \kappa; a + b; 1 - 1/\lambda)} \frac{(\kappa)_l (b)_l}{(a + b)_l} \frac{(1 - 1/\lambda)^l}{l!} \mathbb{1}_{\mathbb{N}}(l).$$

Démonstration. **a)** Si $\lambda = 1$, on remarque que J est dégénérée en 0 et on retrouve l'équivalence entre les lois $\mathcal{B}(a, b)$ et $\mathcal{BG4}(a, b, 1, \kappa)$. Pour $0 < \lambda < 1$, on a

$$\begin{aligned}
f(x) &= \frac{1}{B(a, b) {}_2F_1(a, \kappa; a + b; 1 - \lambda)} x^{a-1} (1 - x)^{b-1} (1 - (1 - \lambda)x)^{-\kappa} \\
&= \frac{1}{B(a, b) {}_2F_1(a, \kappa; a + b; 1 - \lambda)} x^{a-1} (1 - x)^{b-1} \sum_{j=0}^{\infty} \frac{(\kappa)_j}{j!} ((1 - \lambda)x)^j \\
&= \sum_{j=0}^{\infty} \frac{x^{a+j-1} (1 - x)^{b-1}}{B(a + j, b)} \frac{1}{{}_2F_1(a, \kappa; a + b; 1 - \lambda)} \frac{B(a + j, b) (\kappa)_j}{B(a, b)} \frac{(1 - \lambda)^j}{j!} \\
&= \sum_{j=0}^{\infty} \frac{x^{a+j-1} (1 - x)^{b-1}}{B(a + j, b)} \frac{1}{{}_2F_1(a, \kappa; a + b; 1 - \lambda)} \frac{(\kappa)_j (a)_j}{(a + b)_j} \frac{(1 - \lambda)^j}{j!}.
\end{aligned}$$

b) Ce résultat découle directement du Lemme 2.14 et de la partie **a)**. □

En fait, en raison de cette représentation en termes de mélanges, on remarque que les lois $\mathcal{BG4}(a, b, \lambda, \kappa)$ font partie de la famille des lois hyperbêta qui a été introduite par Marchand en 2001 [Mar01]. Cette famille comprend aussi les lois bêta décentrées et la distribution du coefficient de détermination multiple R^2 d'un échantillon de loi multinormale. Les lois hyperbêta résultent d'un mélange de loi bêta avec une loi hypergéométrique généralisée. Cette dernière, qui a été introduite par Kemp [Kem68] et qui sera notée ici $\mathcal{HG}(a_1, \dots, a_p, b_1, \dots, b_q, \theta)$, a pour fonction de densité

$$p(j) = \frac{1}{{}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; \lambda)} \frac{\prod_{i=1}^p (a_i)_j \lambda^j}{\prod_{i=1}^q (b_i)_j j!} \mathbb{1}_{\mathbb{N}}(j), \quad 0 \leq \lambda \leq 1.$$

De plus, on rappelle que la distribution $\mathcal{BG3}(a, b, \lambda)$ est en fait un cas particulier de la distribution $\mathcal{BG4}(a, b, \lambda, \kappa)$ avec $\kappa = a + b$. Dans ce cas, dans la Proposition 2.15, les fonctions de masse des variables J et L se simplifient et on obtient alors le résultat suivant.

Corollaire 2.16. *Soit $X \sim \mathcal{BG3}(a, b, \lambda)$. Alors, selon la valeur de λ , X est caractérisé par les mélanges suivants :*

- a) *Pour $0 < \lambda \leq 1$, $X|J = j \sim \mathcal{B}(a + j, b)$ et $J \sim \mathcal{BN}(a, \lambda)$;*
- b) *Pour $\lambda > 1$, $X|L = l \sim \mathcal{B}(b + l, a)$ et $L \sim \mathcal{BN}(b, 1/\lambda)$.*

Dans un autre ordre d'idées, comme ces lois généralisées surviendront lors de l'étude de notre modèle bivarié dans le but d'inférer sur ϕ , il serait utile d'avoir la fonction de répartition et les moments d'une loi $\mathcal{BG4}(a, b, \lambda, \kappa)$. De l'article de Chen et Novick [CN84], si $X \sim \mathcal{BG4}(a, b, \lambda, \kappa)$, alors pour $x \in [0, 1]$, sa fonction de répartition est donnée par

$$\mathbb{P}(X \leq x) = \frac{x^a}{aB(a, b)} \frac{F_1(a; \kappa, 1 - b; a + 1; (1 - \lambda)x, x)}{{}_2F_1(a, \kappa; a + b; 1 - \lambda)}. \quad (2.19)$$

De ce même article, en ce qui concerne les moments de X , on a

$$E[X^m] = \frac{\Gamma(a + b)\Gamma(a + m)}{\Gamma(a)\Gamma(a + b + m)} \frac{{}_2F_1(a + m, \kappa; a + b + m; 1 - \lambda)}{{}_2F_1(a, \kappa; a + b; 1 - \lambda)}, \quad m > -a. \quad (2.20)$$

Ainsi, pour une telle distribution, l'espérance de X est alors

$$E[X] = \frac{a}{a+b} \frac{{}_2F_1(a+1, \kappa; a+b+1; 1-\lambda)}{{}_2F_1(a, \kappa; a+b; 1-\lambda)}. \quad (2.21)$$

Si $\kappa = a + b$, on a vu que la loi bêta généralisée à quatre paramètres $\mathcal{BG4}(a, b, \lambda, \kappa)$ se simplifiait à la loi bêta généralisée à trois paramètres $\mathcal{BG3}(a, b, \lambda)$. Comme ${}_2F_1(a, a + b; a + b, 1 - \lambda) = \lambda^{-a}$, on peut alors déduire de l'équation (2.20) que si $X \sim \mathcal{BG3}(a, b, \lambda)$, alors les moments de X sont

$$E[X^m] = \frac{\Gamma(a+b)\Gamma(a+m)}{\Gamma(a)\Gamma(a+b+m)} \lambda^a {}_2F_1(a+m, a+b; a+b+m; 1-\lambda), \quad m > -a, \quad (2.22)$$

d'où

$$E[X] = \frac{a}{a+b} \lambda^a {}_2F_1(a+1, a+b; a+b+1; 1-\lambda). \quad (2.23)$$

En outre, si $\lambda = 1$, on a vu que ces lois généralisées se simplifiaient à une loi bêta. Or, on voit bien que, dans ce cas, l'équation (2.20) se réduit à

$$E[X^m] = \frac{\Gamma(a+b)\Gamma(a+m)}{\Gamma(a)\Gamma(a+b+m)}, \quad m > -a, \quad (2.24)$$

ce qui correspond bel et bien aux moments d'une loi bêta, qui, en particulier, nous donnent

$$E[X] = \frac{a}{a+b} \quad (2.25)$$

et

$$Var(X) = \frac{ab}{(a+b)^2(a+b+1)}. \quad (2.26)$$

2.3 Distributions bêta de type II généralisées

Au début du chapitre, on a vu que pour passer de la loi $\mathcal{B}(a, b)$ à la loi $\mathcal{B}_2(c, d, 1)$, on n'avait qu'à appliquer la transformation $T(X) = \frac{X}{1-X}$. En se basant sur celle-ci, on peut

alors également trouver des généralisations de la loi $\mathcal{B}_2(c, d, 1)$. Dans la littérature, il semble y avoir confusion quant à la dénomination de cette loi. En effet, par exemple, Pham-Gia et Duong [PGD89] font référence à la loi bêta de type II comme étant la loi F de Snedecor, et la généralisation à trois paramètres de cette loi est identifiée $\mathcal{G}\mathcal{3}\mathcal{F}$. Par contre, dans l'article de Laurent [Lau12], la loi $\mathcal{B}_2(c, d, 1)$ est plutôt identifiée comme la loi bêta prime et notée $\mathcal{B}'(c, d)$, la généralisation à trois paramètres de cette loi, comme la loi bêta de type II, et la généralisation à quatre paramètres, comme la loi bêta de type III.

Nous privilégierons ici la terminologie de Laurent, à l'exception du fait que la loi bêta prime sera également nommée la loi bêta de type II. Ce choix provient du fait que pour obtenir la généralisation à trois paramètres de la loi $\mathcal{B}_2(c, d, 1)$, il ne suffit que d'appliquer une transformation d'échelle. En effet, nous avons vu plus tôt à l'équation (2.17) que si $Y \sim \mathcal{B}\mathcal{G}3(a, b, \lambda)$ et $X \sim \mathcal{B}(a, b)$, alors

$$\frac{Y}{1-Y} = \tau \frac{X}{1-X},$$

avec $\tau = \frac{1}{\lambda}$. Comme $\frac{X}{1-X}$ est de loi $\mathcal{B}_2(a, b, 1)$, alors la généralisation à trois paramètres de cette distribution est donnée par la loi de la variable $\frac{Y}{1-Y}$. Donc pour $\tau > 0$, si $W \sim \mathcal{B}_2(c, d, 1)$, alors $\tau W \sim \mathcal{B}_2(c, d, \tau)$, et ainsi, τ est un paramètre d'échelle.

Définition 2.17. *On dit que X se distribue selon une loi bêta de type II, que l'on notera $\mathcal{B}_2(c, d, \tau)$, si sa fonction de densité est de la forme*

$$f(x) = \frac{\tau^{-c}}{B(c, d)} \frac{x^{c-1}}{\left(1 + \frac{x}{\tau}\right)^{c+d}} \mathbf{1}_{[0, \infty)}(x), \quad c, d, \tau > 0. \quad (2.27)$$

On remarque qu'il s'agit de la loi qui a été introduite au Lemme 1.21. Ainsi, cette loi peut être formée à partir du quotient de deux variables aléatoires indépendantes de lois $\mathcal{G}(\alpha_1, \beta_1)$ et $\mathcal{G}(\alpha_2, \beta_2)$. De plus, pour $\beta_1 = \beta_2$, cette loi se simplifie à la loi $\mathcal{B}_2(\alpha_1, \alpha_2, 1)$.

Dans la section précédente, on avait présenté une généralisation de la loi bêta à quatre paramètres, c'est-à-dire la loi $\mathcal{BG4}(a, b, \lambda, \kappa)$. Également basé sur la transformation $T(X) = \frac{X}{1-X}$, on peut également construire une généralisation de la loi bêta de type II à quatre paramètres, à partir de la loi $\mathcal{BG4}(a, b, \lambda, \kappa)$.

Définition 2.18. *On dit que X se distribue selon une loi bêta de type III, que l'on notera $\mathcal{B}_3(c, d, \kappa, \tau)$, avec $c, d, \tau > 0$ et $\kappa \in \mathbb{R}$, si sa fonction de densité est de la forme*

$$f(x) = \frac{1}{B(c, d)_2F_1\left(c, c + d - \kappa; c + d; 1 - \frac{1}{\tau}\right)} \frac{x^{c-1}(1+x)^{-\kappa}}{\left(1 + \frac{x}{\tau}\right)^{c+d-\kappa}} \mathbf{1}_{[0, \infty)}(x). \quad (2.28)$$

Cette famille de lois contient les lois bêta de type II. En effet, pour $\kappa = 0$, la loi $\mathcal{B}_3(c, d, \kappa, \tau)$ se réduit à la loi $\mathcal{B}_2(c, d, \tau)$, et pour $\kappa = c + d$, elle se réduit à la loi $\mathcal{B}_2(c, d, 1)$. Pour la généralisation à quatre paramètres, le paramètre τ n'en est plus un d'échelle, sauf dans le cas où $\kappa = 0$, puisque l'on retrouve la loi $\mathcal{B}_2(c, d, \tau)$.

Lemme 2.19. *Si $X \sim \mathcal{BG4}(a, b, \lambda, \kappa)$, alors $\frac{X}{1-X} \sim \mathcal{B}_3(a, b, a + b - \kappa, 1/\lambda)$. Inversement, si $Y \sim \mathcal{B}_3(c, d, \kappa, \tau)$, alors $\frac{Y}{1+Y} \sim \mathcal{BG4}(c, d, 1/\tau, c + d - \kappa)$.*

Démonstration. On a

$$\begin{aligned} f_Y(y) &= f_X\left(\frac{y}{1+y}\right) \frac{dx}{dy} \\ &= \frac{1}{B(a, b)_2F_1(a, \kappa; a + b; 1 - \lambda)} \frac{\left(\frac{y}{1+y}\right)^{a-1} \left(1 - \frac{y}{1+y}\right)^{b-1}}{\left(1 - (1 - \lambda) \frac{y}{1+y}\right)^\kappa} \mathbf{1}_{[0, 1]} \left(\frac{y}{1+y}\right) \frac{1}{(1+y)^2} \\ &= \frac{1}{B(a, b)_2F_1\left(a, \kappa; a + b; 1 - \frac{1}{\lambda}\right)} \frac{y^{a-1}(1+y)^{-(a+b-\kappa)}}{\left(1 + \frac{y}{(1/\lambda)}\right)^\kappa} \mathbf{1}_{[0, \infty)}(y). \end{aligned}$$

On procède de la même manière pour l'implication inverse. □

Corollaire 2.20. *Si $X \sim \mathcal{BG4}(a, b, \lambda, \kappa)$, alors $\frac{1-X}{X} \sim \mathcal{B}_3(b, a, a + b - \kappa, \lambda)$.*

Démonstration. Il s'agit d'une conséquence directe des lemmes 2.19 et 2.14. □

En raison de la représentation des lois bêta généralisées à quatre paramètres en termes de mélanges de lois bêta et de lois hypergéométriques généralisées, on déduit le résultat suivant.

Corollaire 2.21. *Soit $X \sim \mathcal{B}_3(c, d, \kappa, \tau)$. Alors selon la valeur de τ , X peut être représenté par les mélanges suivants :*

- a) *Pour $\tau \leq 1$, $X|J = j \sim \mathcal{B}_2(c, d + j, 1)$, où J est une variable aléatoire sur \mathbb{N} avec $p(j) \propto \frac{(1 - \tau)^j (c + d - \kappa)_j (d)_j}{j! (c + d)_j}$;*
- b) *Pour $\tau > 1$, $X|L = l \sim \mathcal{B}_2(c + l, d, 1)$, où L est une variable aléatoire sur \mathbb{N} avec $p(l) \propto \frac{(1 - 1/\tau)^l (c + d - \kappa)_l (c)_l}{l! (c + d)_l}$.*

Démonstration. Ces résultats découlent du Lemme 2.19, de la Proposition 2.15 et du Corollaire 2.16. □

Dans une optique inférentielle, la fonction de répartition d'une loi peut être utile afin de trouver des intervalles de confiance à partir des quantiles. Ainsi, si $U \sim \mathcal{B}_3(c, d, \kappa, \tau)$, avec $c, d, \tau > 0$ et $\kappa \in \mathbb{R}$, alors on a

$$\mathbb{P}(U \leq u) = C \int_0^u \frac{x^{c-1} (1+x)^{-\kappa}}{\left(1 + \frac{x}{\tau}\right)^{c+d-\kappa}} dx, \quad \text{où } C = \frac{1}{B(c, d)_2 F_1\left(c, c + d - \kappa; c + d; 1 - \frac{1}{\tau}\right)}.$$

En posant $y = \frac{x}{u-x}$, on obtient

$$\begin{aligned} \mathbb{P}(U \leq u) &= C \int_0^\infty \frac{\left(\frac{uy}{1+y}\right)^{c-1} \left(1 + \frac{uy}{1+y}\right)^{-\kappa}}{\left(1 + \frac{1}{\tau} \left(\frac{uy}{1+y}\right)\right)^{c+d-\kappa}} \frac{u}{(1+y)^2} dy \\ &= Cu^c \int_0^\infty \frac{y^{c-1} (1+y)^{d-1}}{(1 + (1+u)y)^\kappa \left(1 + \left(1 + \frac{u}{\tau}\right)y\right)^{c+d-\kappa}} dy \\ &= Cu^c B(c, 1) F_1\left(c; \kappa, c + d - \kappa; c + 1; -u, -\frac{u}{\tau}\right) \end{aligned}$$

$$= \frac{u^c}{cB(c, d)} \frac{F_1\left(c; \kappa, c + d - \kappa; c + 1; -u, -\frac{u}{\tau}\right)}{{}_2F_1\left(c, c + d - \kappa; c + d; 1 - \frac{1}{\tau}\right)}.$$

En appliquant la transformation (2.12.a) à la fonction hypergéométrique d'Appell, on a

$$\mathbb{P}(U \leq u) = \frac{u^c(1+u)^{-\kappa}(1+u/\tau)^{-(c+d-\kappa)}}{cB(c, d)} \frac{F_1\left(1; \kappa; c + d - \kappa; c + d; \frac{u}{1+u}, \frac{u/\tau}{1+u/\tau}\right)}{{}_2F_1\left(c, c + d - \kappa; c + d; 1 - \frac{1}{\tau}\right)}. \quad (2.29)$$

Dans le cas où $\kappa = 0$, c'est-à-dire en supposant plutôt que $U \sim \mathcal{B}_2(c, d, \tau)$, on a :

$$\mathbb{P}(U \leq u) = \frac{1}{cB(c, d)} \left(\frac{u/\tau}{1+u/\tau}\right)^c \left(\frac{1}{1+u/\tau}\right)^d {}_2F_1\left(1, c + d; c + 1; \frac{u/\tau}{1+u/\tau}\right). \quad (2.30)$$

Comme on a vu, dans un cadre bayésien, on peut aussi estimer ponctuellement un paramètre à partir de l'espérance *a posteriori*. En ce sens, on pourra se servir des résultats suivants sur les moments.

Théorème 2.22. *Si $X \sim \mathcal{B}_3(c, d, \kappa, \tau)$, alors ses moments sont donnés par la formule suivante :*

$$E[X^m] = \frac{\Gamma(c+m)\Gamma(d-m)}{\Gamma(c)\Gamma(d)} \frac{{}_2F_1(c+m, c+d-\kappa; c+d; 1-1/\tau)}{{}_2F_1(c, c+d-\kappa; c+d; 1-1/\tau)}, \quad c > -m, \quad d > m. \quad (2.31)$$

Démonstration.

$$\begin{aligned} E[X^m] &= \frac{1}{B(c, d) {}_2F_1(c, c+d-\kappa; c+d; 1-1/\tau)} \int_0^\infty \frac{x^{c+m-1}(1+x)^{-\kappa}}{\left(1+\frac{x}{\tau}\right)^{c+d-\kappa}} dx \\ &= \frac{B(c+m, d-m) {}_2F_1(c+m, c+d-\kappa; c+d; 1-1/\tau)}{B(c, d) {}_2F_1(c, c+d-\kappa; c+d; 1-1/\tau)}, \quad \text{si } c > -m, d > m, \\ &= \frac{\Gamma(c+m)\Gamma(d-m)}{\Gamma(c)\Gamma(d)} \frac{{}_2F_1(c+m, c+d-\kappa; c+d; 1-1/\tau)}{{}_2F_1(c, c+d-\kappa; c+d; 1-1/\tau)}. \quad \square \end{aligned}$$

Ainsi, l'espérance d'une variable aléatoire de distribution bêta de type III est

$$E[X] = \frac{c}{d-1} \frac{{}_2F_1(c+1, c+d-\kappa; c+d; 1-1/\tau)}{{}_2F_1(c, c+d-\kappa; c+d; 1-1/\tau)}, \quad d > 1. \quad (2.32)$$

De plus, comme la loi bêta de type II avec paramètre d'échelle est obtenue à partir de la loi bêta de type III en posant $\kappa = 0$, on peut alors déduire directement du Théorème 2.22 les moments d'une variable aléatoire $X \sim \mathcal{B}_2(c, d, \tau)$.

Corollaire 2.23. *Si $X \sim \mathcal{B}_2(c, d, \tau)$, alors ses moments sont donnés par la formule suivante :*

$$E[X^m] = \frac{\Gamma(c+m)\Gamma(d-m)}{\Gamma(c)\Gamma(d)} \tau^m, \quad c > -m, \quad d > m. \quad (2.33)$$

Ainsi, l'espérance et la variance d'une variable aléatoire de distribution $\mathcal{B}_2(c, d, \tau)$ sont respectivement

$$E[X] = \frac{c}{d-1} \tau, \quad d > 1, \quad (2.34)$$

et

$$Var(X) = \frac{c}{(d-1)^2} \frac{c+d-1}{d-2} \tau^2, \quad d > 2. \quad (2.35)$$

CHAPITRE 3

Étude du modèle bivarié sans contrainte paramétrique

On rappelle que le modèle bivarié à l'étude est le suivant : $(X_1, X_2)|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$, où S et T sont des constantes positives. Comme les variables aléatoires X_1 et X_2 sont conditionnellement indépendantes et de distributions marginales respectives $\mathcal{P}(\mu\phi S)$ et $\mathcal{P}(\mu T)$, la fonction de masse conjointe, conditionnellement aux paramètres μ et ϕ , s'écrit

$$p(x_1, x_2|\mu, \phi) = \frac{S^{x_1} T^{x_2}}{x_1! x_2!} e^{-\mu(T+\phi S)} \mu^{x_1+x_2} \phi^{x_1}, \quad (x_1, x_2) \in \mathbb{N}^2.$$

D'une part, ce chapitre a pour but de présenter un cadre bayésien pour ce problème et d'y extraire des lois *a posteriori* correspondantes de même que des estimateurs ponctuels pour ϕ . À l'aide du risque fréquentiste, nous les comparerons également avec l'estimateur issu de la méthode du maximum de vraisemblance. D'autre part, nous étudierons le problème d'inférence prédictive dans ce contexte. Nous présenterons des estimateurs de densité prédictive pour de futures observations $(Y_1, Y_2)|\mu, \phi \sim \mathcal{P}2(\mu\phi S', \mu T')$, avec S' et T' des constantes positives non nécessairement égales respectivement à S et à T .

3.1 Estimateur à vraisemblance maximale modifié

Le premier estimateur présenté pour ϕ sera l'estimateur du maximum de vraisemblance. Pour la suite, on notera \mathbf{X} le couple aléatoire bivarié (X_1, X_2) .

Théorème 3.1. *Soit $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$. Alors pour $x_2 > 0$, l'estimateur du maximum de vraisemblance de (μ, ϕ) est*

$$(\hat{\mu}_{EVM}(\mathbf{x}), \hat{\phi}_{EVM}(\mathbf{x})) = \left(\frac{x_2}{T}, \frac{T}{S} \left(\frac{x_1}{x_2} \right) \right). \quad (3.1)$$

Démonstration. La vraisemblance est $\mathcal{L}(\mu, \phi|\mathbf{x}) = \frac{S^{x_1} T^{x_2}}{x_1! x_2!} e^{-\mu(T+\phi S)} \mu^{x_1+x_2} \phi^{x_1}$. Pour trouver l'estimateur du maximum de vraisemblance, on doit résoudre ce système d'équations :

$$\begin{cases} \frac{\partial}{\partial \mu} \log \mathcal{L}(\mu, \phi|\mathbf{x}) = -(T + \phi S) + \frac{x_1 + x_2}{\mu} = 0 \\ \frac{\partial}{\partial \phi} \log \mathcal{L}(\mu, \phi|\mathbf{x}) = -\mu S + \frac{x_1}{\phi} = 0. \end{cases}$$

On obtient le point critique $(\hat{\mu}(\mathbf{x}), \hat{\phi}(\mathbf{x})) = \left(\frac{x_2}{T}, \frac{T x_1}{S x_2} \right)$, pour $x_2 > 0$. Il s'agit bel et bien d'un maximum, car, pour $x_2 > 0$, on a

$$\begin{vmatrix} -\frac{(x_1 + x_2)}{\hat{\mu}^2} & -S \\ -S & -\frac{x_1}{\hat{\phi}^2} \end{vmatrix} = S^2 \left(\frac{x_1 + x_2}{x_1} - 1 \right) > 0,$$

et

$$-\frac{(x_1 + x_2)}{\hat{\mu}^2} < 0. \quad \square$$

Il est à noter qu'on aurait obtenu les mêmes estimateurs à partir de la méthode des moments. De plus, l'approche conditionnelle découlant de la Proposition 1.8 du chapitre 1, dans laquelle on a $X_1|X_1 + X_2 = n \sim \mathcal{B}in\left(n, \frac{\phi S}{\phi S + T}\right)$, nous aurait aussi guider vers le même résultat. On a vu qu'inférer sur ϕ revenait à inférer sur la proportion $p = \frac{\phi S}{\phi S + T}$

via l'équation (1.5). En effet, si $W|p \sim \mathcal{Bin}(n, p)$, alors l'estimateur du maximum de vraisemblance est $\hat{p} = w/n$. Pour le modèle à l'étude, il en découle alors que, pour $x_2 > 0$,

$$\hat{\phi}(\mathbf{x}) = \frac{Tx_1}{Sx_2},$$

ce qui concorde avec ce qui a été trouvé précédemment.

Par ailleurs, on peut observer que $\phi = \frac{T}{S} \left(\frac{\mu\phi S}{\mu T} \right) = \frac{T}{S} \frac{E[X_1|\mu, \phi]}{E[X_2|\mu, \phi]}$. On remarque alors que l'estimateur $\hat{\phi}_{EVM}(\mathbf{x})$ est obtenu en remplaçant $E[X_1|\mu, \phi]$ et $E[X_2|\mu, \phi]$ par leurs EVM respectifs x_1 et x_2 , ce qui découle de la propriété d'invariance de l'EVM.

Toutefois, cet estimateur n'est pas bien défini pour toutes valeurs possibles de X_2 . En effet, si $x_1 = 0$ mais $x_2 \neq 0$, on a $\hat{\phi}(\mathbf{x}) = 0$. Par contre, lorsque $x_2 = 0$ (c'est-à-dire lorsque $\hat{p} = 1$), cet estimateur n'est pas défini. Nous appliquerons alors une correction pour qu'il soit défini sur tout le support de $\mathbf{x} \in \mathbb{N}^2$. La correction proposée ici sera d'ajouter la valeur 0,5 à x_1 et à x_2 . C'est une stratégie qui a été notamment utilisée par Price et Bonett dans [PB00] pour améliorer la performance générale de certaines méthodes d'estimation basées sur le maximum de vraisemblance. Nous définissons alors l'estimateur du maximum de vraisemblance modifié comme suit :

$$\hat{\phi}_{EVMM}(\mathbf{x}) = \frac{T}{S} \left(\frac{x_1 + 0,5}{x_2 + 0,5} \right). \quad (3.2)$$

Cet estimateur, issu d'une méthode d'approche fréquentiste, sera comparé plus tard avec certains estimateurs bayésiens, qui seront présentés dans la prochaine section.

3.2 Estimateurs bayésiens

Dans cette section, les paramètres μ et ϕ sont vus comme des variables aléatoires. Lindley [Lin02], Laurent [Lau12] de même que Laurent et Legrand [LL12] ont travaillé sur ce

problème d'estimation selon une approche bayésienne. À l'image de ce qu'ils ont fait, on supposera que μ et ϕ sont indépendants, de telle sorte que $\pi(\mu, \phi) = \pi(\mu)\pi(\phi)$. Comme Lindley l'explique dans [Lin02], cette hypothèse revient à considérer, par exemple, dans un contexte où l'on veut estimer l'efficacité d'un traitement, que le traitement a un effet multiplicatif d'un facteur ϕ et non un effet additif.

Comme seconde hypothèse, on supposera que $\mu|\phi \sim \mathcal{G}(a, b)$, où $a, b > 0$. On admettra aussi des lois *a priori* impropres, soit celles pour lesquelles $a > 0$ et $T < b \leq 0$, afin que la loi *a posteriori* soit bien définie. Ainsi, on a

$$\pi(\mu, \phi) = \frac{b^a \mu^{a-1} e^{-b\mu}}{\Gamma(a)} \mathbb{1}_{[0, \infty)}(\mu) \pi(\phi). \quad (3.3)$$

Ce choix de densité *a priori* $\pi(\mu|\phi)$ découle principalement de motifs calculatoires en lien avec les éventuels calculs de densités *a posteriori*.

Un premier estimateur pour ϕ , présenté par Lindley [Lin02], est obtenu en considérant μ comme un paramètre de nuisance. L'idée de cette méthode est de calculer la valeur espérée de $p(\mathbf{x}|\mu, \phi)$ par rapport à la loi *a priori* de $\mu|\phi$, et ainsi, d'obtenir une vraisemblance qui ne dépend que de ϕ . La proposition suivante donne quelques résultats en lien avec notre modèle étudié lorsque μ est considéré comme un paramètre de nuisance.

Proposition 3.2. *Soient $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$ et $\mu|\phi \sim \mathcal{G}(a, b)$, avec $a, b > 0$. Alors la fonction de masse de $\mathbf{X}|\phi$ est*

$$p(\mathbf{x}|\phi) = \frac{(a)_{x_1+x_2}}{x_1!x_2!} \frac{b^a (\phi S)^{x_1} T^{x_2}}{(T + b + \phi S)^{a+x_1+x_2}}. \quad (3.4)$$

De plus, on a :

- i) $X_1|\phi \sim \mathcal{BN}\left(a, \frac{b}{b+\phi S}\right)$;
- ii) $X_2|\phi \sim \mathcal{BN}\left(a, \frac{b}{T+b}\right)$;
- iii) $X_1|X_2 = x_2, \phi \sim \mathcal{BN}\left(a + x_2, \frac{T+b}{T+b+\phi S}\right)$.

Démonstration. Pour $\mathbf{x} \in \mathbb{N}^2$, on a

$$\begin{aligned}
p(\mathbf{x}|\phi) &= \int_0^\infty p(\mathbf{x}|\phi, \mu)\pi(\mu|\phi)d\mu \\
&= \int_0^\infty p(\mathbf{x}|\phi, \mu)\pi(\mu)d\mu \quad (\text{puisque } \mu \text{ et } \phi \text{ sont indépendantes}) \\
&= \frac{S^{x_1}T^{x_2}}{x_1!x_2!} \frac{b^a}{\Gamma(a)} \phi^{x_1} \int_0^\infty e^{-\mu(T+b+\phi S)} \mu^{a+x_1+x_2-1} d\mu \\
&= \frac{(a)_{x_1+x_2}}{x_1!x_2!} \frac{b^a(\phi S)^{x_1}T^{x_2}}{(T+b+\phi S)^{a+x_1+x_2}}.
\end{aligned}$$

On peut réécrire cette densité de la manière suivante :

$$p(\mathbf{x}|\phi) = \frac{(a+x_2)_{x_1}}{x_1!} \left(\frac{T+b}{T+b+\phi S} \right)^{a+x_2} \left(\frac{\phi S}{T+b+\phi S} \right)^{x_1} \times \frac{(a)_{x_2}}{x_2!} \left(\frac{b}{T+b} \right)^a \left(\frac{T}{T+b} \right)^{x_2},$$

d'où $X_1|X_2 = x_2, \phi \sim \mathcal{BN}\left(a+x_2, \frac{T+b}{T+b+\phi S}\right)$ et $X_2|\phi \sim \mathcal{BN}\left(a, \frac{b}{T+b}\right)$. Pareillement, on obtient de $p(\mathbf{x}|\phi)$ que $X_1|\phi \sim \mathcal{BN}\left(a, \frac{b}{b+\phi S}\right)$. \square

Remarque 3.3. De la Proposition 3.2, on a que $X_2|\phi \sim \mathcal{BN}\left(a, \frac{b}{T+b}\right)$. Toutefois, cette loi ne fait pas intervenir le paramètre ϕ , et donc, on a aussi $X_2 \sim \mathcal{BN}\left(a, \frac{b}{T+b}\right)$. Ainsi,

$$E[X_2] = \frac{Ta}{b} \quad \text{et} \quad \text{Var}(X_2) = \frac{Ta(T+b)}{b^2} = \left(\frac{T+b}{b} \right) E[X_2]. \quad (3.5)$$

Comme Lindley le fait remarquer [Lin02], bien que $X_2|\mu$ se distribue selon une loi de Poisson, où $E[X_2|\mu] = \text{Var}(X_2|\mu)$, l'espérance et la variance de la loi marginale X_2 ne sont pas égales. En effet, telle que présentée à l'équation (3.5), la variance de X_2 lui est supérieure d'un facteur $\frac{T+b}{b}$, ce qui implique une plus grande variabilité relative que lorsque le paramètre μ est connu. De plus, comme mentionné par Lindley [Lin02], la loi marginale de X_2 peut servir à choisir les hyperparamètres a et b dans un cadre pratique. Cette méthode sera illustrée à l'Exemple 3.15.

De l'équation (3.4), on trouve

$$\mathcal{L}(\phi|\mathbf{x}) \propto \frac{\phi^{x_1}}{(T + b + \phi S)^{a+x_1+x_2}}, \quad (3.6)$$

et donc,

$$\log \mathcal{L}(\phi|\mathbf{x}) \propto x_1 \log(\phi) - (a + x_1 + x_2) \log(T + b + \phi S). \quad (3.7)$$

Cette log-vraisemblance est maximisée en

$$\hat{\phi}_{B1}(\mathbf{x}) = \left(\frac{T + b}{S} \right) \left(\frac{x_1}{x_2 + a} \right), \quad a > 0, \quad b > -T. \quad (3.8)$$

Une analyse très intéressante de cet estimateur, que nous avons noté $\hat{\phi}_{B1}(\mathbf{x})$, a été présentée par Lindley [Lin02] et est basée sur la Remarque 3.3, c'est-à-dire sur le fait que $X_2 \sim \mathcal{BN}\left(a, \frac{b}{T+b}\right)$. En effet, pour $x_2 > 0$ et $\rho = \frac{Ta}{bx_2} = \frac{E[X_2]}{x_2}$, on a

$$\hat{\phi}_{B1}(\mathbf{x}) = \frac{T x_1}{S x_2} \left(\frac{T + b}{T + \rho b} \right) = \hat{\phi}_{EVM}(\mathbf{x}) \left(\frac{T + b}{T + \rho b} \right).$$

On rappelle, comme expliqué plus tôt, que $\phi = \frac{T}{S} \frac{E[X_1|\mu, \phi]}{E[X_2|\mu, \phi]}$ et que $\hat{\phi}_{EVM}(\mathbf{x}) = \frac{T x_1}{S x_2}$. Ainsi, il semble que le facteur $\frac{T+b}{T+\rho b}$ vienne apporter une correction à cet estimateur du maximum de vraisemblance. Trois cas se présentent à nous selon la valeur de ρ :

- 1) Supposons que $\rho = 1$. Dans ce cas, on a $x_2 = E[X_2]$. Donc $\hat{\phi}_{B1}(\mathbf{x}) = \hat{\phi}_{EVM}(\mathbf{x})$.
- 2) Supposons que $\rho < 1$. Dans ce cas, on a $x_2 > E[X_2]$. Avec la valeur observée x_2 de X_2 supérieure à la moyenne, l'estimé bayésien $\hat{\phi}_{B1}(\mathbf{x})$ ajuste à la hausse $\hat{\phi}_{EVM}(\mathbf{x}) = \frac{T x_1}{S x_2}$, car il est plausible que ce dernier sous-estime $\phi = \frac{T}{S} \frac{E[X_1|\mu, \phi]}{E[X_2|\mu, \phi]}$. Donc $\hat{\phi}_{B1}(\mathbf{x}) > \hat{\phi}_{EVM}(\mathbf{x})$.
- 3) Supposons que $\rho > 1$. Dans ce cas, on a $x_2 < E[X_2]$. Par un raisonnement analogue au cas précédent, il est plausible que $\hat{\phi}_{EVM}(\mathbf{x})$ surestime ϕ et donc, $\hat{\phi}_{B1}(\mathbf{x}) < \hat{\phi}_{EVM}(\mathbf{x})$.

Le second estimateur bayésien pour ϕ qui sera analysé sera l'espérance *a posteriori*. On rappelle que l'intérêt porté à l'espérance *a posteriori* provient du fait qu'il s'agit de l'estimateur de Bayes pour la fonction de perte quadratique, tel que montré à l'Exemple 1.40. En ce sens, le prochain théorème présente les lois *a posteriori* dans leur forme générale, c'est-à-dire lorsque ϕ a pour densité $\pi(\phi)$. On suppose toujours que $\mu|\phi \sim \mathcal{G}(a, b)$.

Théorème 3.4. *Soit $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$. Si $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi \sim \pi$ telle que la densité *a posteriori* $\pi(\phi|\mathbf{x})$ existe, alors*

$$\mu|\phi, \mathbf{x} \sim \mathcal{G}(a + x_1 + x_2, T + b + \phi S), \quad (3.9)$$

et les densités *a posteriori* de μ et de ϕ sont respectivement

$$\pi(\mu|\mathbf{x}) \propto \int_0^\infty e^{-\mu(T+b+\phi S)} \mu^{a+x_1+x_2-1} \phi^{x_1} \pi(\phi) \mathbf{1}_{[0,\infty)}(\mu) d\phi, \quad (3.10)$$

et

$$\pi(\phi|\mathbf{x}) \propto \frac{\phi^{x_1} \pi(\phi)}{(T + b + \phi S)^{a+x_1+x_2}}. \quad (3.11)$$

Démonstration. Comme

$$\begin{aligned} \pi(\mu, \phi|\mathbf{x}) &\propto p(\mathbf{x}|\mu, \phi) \pi(\mu, \phi) \\ &\propto e^{-\mu(T+\phi S)} \mu^{x_1+x_2} \phi^{x_1} \mu^{a-1} e^{-b\mu} \mathbf{1}_{[0,\infty)}(\mu) \pi(\phi) \\ &= (T + b + \phi S)^{a+x_1+x_2} e^{-\mu(T+b+\phi S)} \mu^{a+x_1+x_2-1} \mathbf{1}_{[0,\infty)}(\mu) \times \frac{\phi^{x_1} \pi(\phi)}{(T + b + \phi S)^{a+x_1+x_2}}, \end{aligned}$$

et que $\pi(\mu, \phi|\mathbf{x}) = \pi(\mu|\phi, \mathbf{x}) \times \pi(\phi|\mathbf{x})$, alors $\mu|\phi, \mathbf{x} \sim \mathcal{G}(a + x_1 + x_2, T + b + \phi S)$ et $\pi(\phi|\mathbf{x}) \propto \frac{\phi^{x_1} \pi(\phi)}{(T + b + \phi S)^{a+x_1+x_2}}$.

Pour la densité *a posteriori* de μ , on a

$$\pi(\mu|\mathbf{x}) = \int_0^\infty \pi(\mu, \phi|\mathbf{x}) d\phi \propto \int_0^\infty e^{-\mu(T+b+\phi S)} \mu^{a+x_1+x_2-1} \mathbf{1}_{[0,\infty)}(\mu) \phi^{x_1} \pi(\phi) d\phi. \quad \square$$

Remarque 3.5. *Le Théorème 3.4 demeure valable pour les lois impropres avec $\pi(\mu) \propto \mu^{a-1} e^{-b\mu}$ pour $-T < b \leq 0$.*

Quant à l'interprétation de ce théorème, tout d'abord, on remarque que la densité de $\mu|\mathbf{x}$ est toujours représentée comme un mélange de loi $\mathcal{G}(a+x_1+x_2, T+b+\phi S)$, où $\phi|\mathbf{x}$ a une densité telle qu'en (3.11). Ensuite, on remarque que les densités *a priori* et *a posteriori* de μ sont toutes deux de la famille des lois gamma, et ce, peu importe la distribution de ϕ . Ce résultat est en accord avec l'Exemple 1.17, où l'on a vu que la famille de lois *a priori* $\mathcal{G}(a, b)$ forme une famille conjuguée pour la famille des lois de Poisson.

Dans les articles de Lindley [Lin02], Laurent [Lau12] et Laurent et Legrand [LL12], la même densité *a priori* pour ϕ est proposée, soit

$$\phi \sim \mathcal{B}_2 \left(c, d, \frac{T+b}{S} \right).$$

Théorème 3.6. *Soient $\mathbf{X}|\mu, \phi \sim \mathcal{P}_2(\mu\phi S, \mu T)$, $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi \sim \mathcal{B}_2 \left(c, d, \frac{T+b}{S} \right)$, où $a, c > 0$, $b > -T$ et $d > -a$. Alors*

$$\phi|\mathbf{x} \sim \mathcal{B}_2 \left(c+x_1, d+a+x_2, \frac{T+b}{S} \right).$$

Démonstration. Du Théorème 3.4, on a

$$\pi(\phi|\mathbf{x}) \propto \frac{\phi^{c+x_1-1}}{(T+b+\phi S)^{a+x_1+x_2} \left(1 + \frac{S}{T+b}\phi\right)^{c+d}} \propto \frac{\phi^{c+x_1-1}}{\left(1 + \frac{S}{T+b}\phi\right)^{c+x_1+d+a+x_2}}. \quad \square$$

Le Lemme 3.1 présenté dans l'article de Laurent [Lau12] décrit la densité *a posteriori* obtenue pour ϕ , mais dans un cadre plus général. En effet, on y considère un paramètre d'échelle quelconque plutôt que de le restreindre à $\frac{T+b}{S}$. De l'équation (3.11), on a alors

$$\pi(\phi|\mathbf{x}) \propto \frac{\phi^{c+x_1-1} \left(1 + \frac{S}{T+b}\phi\right)^{-(a+x_1+x_2)}}{\left(1 + \frac{\phi}{\rho}\right)^{c+d}} \mathbb{1}_{[0, \infty)}(\phi),$$

d'où $\phi|\mathbf{x} \sim \frac{T+b}{S} \times \mathcal{B}_3 \left(c+x_1, d+a+x_2, a+x_1+x_2, \frac{S}{T+b}\rho \right)$. En posant $\rho = \frac{T+b}{S}$, on voit bien que la densité se simplifie à celle d'une loi $\mathcal{B}_2 \left(c+x_1, d+a+x_2, \frac{T+b}{S} \right)$. Ainsi, le fait d'imposer judicieusement $\rho = \frac{T+b}{S}$ est un choix basé entre autres sur des motifs

calculatoires et on remarque que ce choix mène à une famille semi-conjuguée pour la famille de lois à l'étude. C'est d'ailleurs ce même choix qui a été fait dans [Lin02], [Lau12] et [LL12].

Du Théorème 3.6, on peut alors trouver un deuxième estimateur issu de méthodes bayésiennes basé sur l'espérance *a posteriori* $E[\phi|x_1, x_2]$, qui, pour une loi bêta de type II, est donnée par l'équation (2.34). On a alors l'estimateur suivant :

$$\hat{\phi}_{B2}(\mathbf{x}) = \left(\frac{T+b}{S} \right) \left(\frac{x_1+c}{x_2+a+d-1} \right), \quad a, c > 0, \quad b > -T, \quad d > 1-a. \quad (3.12)$$

Cet estimateur sera étudié à la section 3.3.

Pour clore cette section, on propose un exemple qui illustre la construction d'un intervalle de crédibilité à partir des lois ci-haut obtenues.

Exemple 3.7. Soient $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$, $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi \sim \mathcal{B}_2(c, d, \rho)$ avec $a, c, \rho > 0$, $b > -T$ et $d > 1-a$. On sait que

$$\phi|\mathbf{x} \sim \frac{T+b}{S} \times \mathcal{B}_3\left(c+x_1, d+a+x_2, a+x_1+x_2, \frac{S}{T+b}\rho\right).$$

On peut notamment construire un intervalle de crédibilité $1-\alpha$, en choisissant par exemple $[q_{\alpha/2}, q_{1-\alpha/2}]$, où $q_{\alpha/2}$ et $q_{1-\alpha/2}$ sont des quantiles de cette loi de telle façon que

$$\mathbb{P}(q_{\alpha/2} \leq \phi \leq q_{1-\alpha/2}|\mathbf{x}) = 1-\alpha.$$

Pour ce faire, on peut s'aider de l'équation (2.29), donnant la fonction de répartition d'une loi bêta de type III, ou, dans un cas plus simple, de l'équation (2.30), lorsque $\rho = \frac{T+b}{S}$, puisque dans ce cas, $\phi|x_1, x_2 \sim \mathcal{B}_2\left(c+x_1, d+a+x_2, \frac{T+b}{S}\right)$.

Par exemple, supposons que $a=1$ et $b=0$, c'est-à-dire que μ est de loi impropre uniforme sur les réels positifs, que $c=d=\frac{1}{2}$, et finalement, que $S=T=1$. Si $\phi \sim \mathcal{B}_2\left(c, d, \frac{T+b}{S}\right)$, alors on obtient la loi *a posteriori* suivante :

$$\phi|x_1, x_2 \sim \mathcal{B}_2\left(x_1 + \frac{1}{2}, x_2 + \frac{3}{2}, 1\right).$$

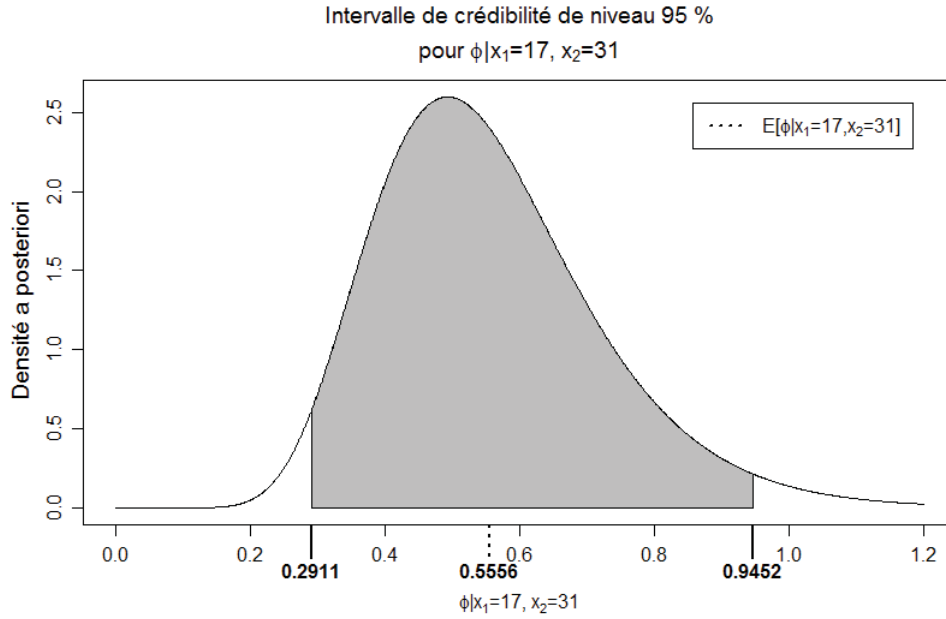


Figure 3.1 – Illustration d’un intervalle de crédibilité pour ϕ

À partir d’une simulation avec $X_1|\mu, \phi \sim \mathcal{P}(21)$ et $X_2|\mu, \phi \sim \mathcal{P}(30)$, c’est-à-dire $\mu = 30$ et $\phi = 0,7$, on obtient $x_1 = 17$ et $x_2 = 31$.

L’estimateur ponctuel de ϕ basé sur l’espérance a posteriori est $\hat{\phi}_{B2}(17, 31) = \frac{5}{9}$. Cet estimateur concorde avec l’estimateur $\hat{\phi}_{EVM}(x)$, tel que défini à l’équation (3.2). On remarque en fait que ce sera le cas pour tous les estimateurs $\hat{\phi}_{B2}(x)$ tels que $a + d = \frac{3}{2}$, $c = \frac{1}{2}$ et $b = 0$. En outre, de l’équation (3.8), toujours pour $a = 1$ et $b = 0$, on a $\hat{\phi}_{B1}(17, 31) = 0,53125$, ce qui est légèrement inférieur à la valeur des autres estimateurs et, cette valeur s’éloigne un peu plus de la réelle valeur $\phi = 0,7$.

Par ailleurs, à partir de la loi a posteriori de ϕ , un intervalle de crédibilité de niveau 95 % pour ϕ est donné par $I_\pi(\phi) = [0,2911; 0,9452]$. Les résultats obtenus semblent plausibles puisque la simulation a été faite avec $\phi = 0,7$. La Figure 3.1 illustre la loi a posteriori de ϕ à partir des observations $x_1 = 17$ et $x_2 = 31$ de même que l’intervalle de crédibilité décrit ci-dessus.

3.3 Évaluation et comparaisons du risque des estimateurs

Depuis le début du chapitre 3, nous avons obtenu trois estimateurs ponctuels, qui, on le rappelle, sont les suivants :

$$\hat{\phi}_{EVMM}(\mathbf{x}) = \frac{T}{S} \left(\frac{x_1 + 0,5}{x_2 + 0,5} \right);$$

$$\hat{\phi}_{B1}(\mathbf{x}) = \left(\frac{T+b}{S} \right) \left(\frac{x_1}{x_2+a} \right), \quad a > 0, \quad b > -T;$$

$$\hat{\phi}_{B2}(\mathbf{x}) = \left(\frac{T+b}{S} \right) \left(\frac{x_1+c}{x_2+a+d-1} \right), \quad a, c > 0, \quad b > -T, \quad d > 1-a.$$

On remarque que l'estimateur du maximum de vraisemblance modifié $\hat{\phi}_{EVMM}(\mathbf{x})$ est un cas particulier de l'estimateur bayésien $\hat{\phi}_{B2}(\mathbf{x})$. En effet, tel qu'illustré à l'Exemple 3.7, pour $b = 0$, $c = \frac{1}{2}$ et $a + d = \frac{3}{2}$, on a que $\hat{\phi}_{EVMM}(\mathbf{x}) = \hat{\phi}_{B2}(\mathbf{x})$ pour tout $\mathbf{x} \in \mathbb{N}^2$. De plus, on constate que ces estimateurs font tous partie de la famille d'estimateurs $\mathcal{F}(\mathbf{x}) = \left\{ \frac{\alpha_1 x_1 + \beta_1}{\alpha_2 x_2 + \beta_2} \mid \alpha_1, \alpha_2, \beta_2 > 0, \beta_1 \geq 0 \right\}$.

Au début des années 50, Chapman a prouvé qu'il n'existait pas d'estimateurs sans biais pour ϕ avec variance finie [Cha52]. Il y a toutefois montré que l'estimateur $\frac{x_1}{x_2+1}$ (pour notre modèle avec $S = T = 1$), qui appartient à la famille d'estimateurs $\mathcal{F}(\mathbf{x})$ ci-haut décrite, est presque sans biais. Ce résultat est présenté dans la proposition qui suit.

Proposition 3.8. [Cha52] Soit $\mathbf{X} \mid \mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$. Alors on a

$$E_{\mu, \phi} \left[\frac{T}{S} \left(\frac{X_1}{X_2+1} \right) \right] = \phi (1 - e^{-\mu T}), \quad (3.13)$$

et donc, l'estimateur $\frac{T}{S} \left(\frac{X_1}{X_2+1} \right)$ est presque sans biais pour ϕ lorsque μ n'est pas trop petit.

Démonstration. On a

$$\begin{aligned}
E_{\mu,\phi} \left[\frac{1}{X_2 + 1} \right] &= \sum_{x_2=0}^{\infty} \frac{e^{-(\mu T)} (\mu T)^{x_2}}{(x_2 + 1)!} \\
&= \frac{e^{-\mu T}}{\mu T} \sum_{x_2=0}^{\infty} \frac{(\mu T)^{x_2+1}}{(x_2 + 1)!} \\
&= \frac{e^{-\mu T}}{\mu T} (e^{\mu T} - 1) \\
&= \frac{1 - e^{-\mu T}}{\mu T},
\end{aligned}$$

et le résultat suit puisque $X_1 | \mu, \phi \perp X_2 | \mu, \phi$. \square

Pour cette section, on procédera à l'étude des trois estimateurs ponctuels trouvés dans la section précédente en évaluant et en comparant leur risque fréquentiste tel que défini à l'équation (1.18). Pour tous les exemples de calcul, on supposera que $S = T = 1$. Le risque sera évalué en fonction du paramètre d'intérêt ϕ pour différentes valeurs fixées de μ . De plus, on utilisera la fonction de perte quadratique, d'où

$$R((\mu, \phi), \hat{\phi}) = \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} \left(\phi - \hat{\phi}(x_1, x_2) \right)^2 \frac{e^{-\mu(1+\phi)}}{x_1! x_2!} \mu^{x_1+x_2} \phi^{x_1}. \quad (3.14)$$

On remarque qu'avec $S = T = 1$ et $b = 0$, les trois estimateurs ponctuels présentés précédemment sont de la forme $\hat{\phi}(x_1, x_2) = \frac{x_1 + \beta_1}{x_2 + \beta_2}$. Dans ce cas particulier, pour μ fixé, le risque est un polynôme de degré 2 en ϕ , tel que montré dans le théorème suivant. Celui-ci pourra nous être utile dans l'interprétation ultérieure des courbes de risque obtenues.

Théorème 3.9. *Soit $\mathbf{X} | \mu, \phi \sim \mathcal{P}2(\mu\phi, \mu)$. Pour toute valeur fixée de $\mu \geq 0$, le risque fréquentiste d'un estimateur ponctuel $\hat{\phi}(\mathbf{x}) = \frac{x_1 + \beta_1}{x_2 + \beta_2}$, avec $\beta_1 \geq 0$ et $\beta_2 > 0$, sous la fonction de perte quadratique, est de la forme*

$$R((\mu, \phi), \hat{\phi}) = A(\mu)\phi^2 + B(\mu)\phi + C(\mu), \quad (3.15)$$

avec

$$A(\mu) = E_{\mu} \left[\left(\frac{\mu}{X_2 + \beta_2} - 1 \right)^2 \right], \quad (3.16)$$

$$B(\mu) = \mu(1 + 2\beta_1)E_\mu \left[\frac{1}{(X_2 + \beta_2)^2} \right] - 2\beta_1 E_\mu \left[\frac{1}{X_2 + \beta_2} \right] \quad (3.17)$$

et

$$C(\mu) = \beta_1^2 E_\mu \left[\frac{1}{(X_2 + \beta_2)^2} \right]. \quad (3.18)$$

Démonstration.

$$\begin{aligned} R((\mu, \phi), \hat{\phi}) &= E_{X_1, X_2 | \mu, \phi} \left[\left(\frac{X_1 + \beta_1}{X_2 + \beta_2} - \phi \right)^2 \right] \\ &= E_{X_2 | \mu, \phi} \left[\frac{1}{(X_2 + \beta_2)^2} E_{X_1 | X_2, \mu, \phi} \left[(X_1 + \beta_1 - \phi(X_2 + \beta_2))^2 \right] \right] \\ &= E_{X_2 | \mu, \phi} \left[\frac{1}{(X_2 + \beta_2)^2} E_{X_1 | X_2, \mu, \phi} \left[(X_1 - \mu\phi + \beta_1 + \phi(\mu - (X_2 + \beta_2)))^2 \right] \right] \\ &= E_{\mu, \phi} \left[\frac{\mu\phi + (\beta_1 + \phi(\mu - (X_2 + \beta_2)))^2}{(X_2 + \beta_2)^2} \right], \text{ car } E_{X_1 | X_2, \mu, \phi} \left[(X_1 - \mu\phi)^2 \right] = \mu\phi \\ &\hspace{25em} \text{et } E_{X_1 | X_2, \mu, \phi} \left[X_1 - \mu\phi \right] = 0 \\ &= \phi^2 E_{\mu, \phi} \left[\left(\frac{\mu}{X_2 + \beta_2} - 1 \right)^2 \right] \\ &\quad + \phi \left(\mu(1 + 2\beta_1) E_{\mu, \phi} \left[\frac{1}{(X_2 + \beta_2)^2} \right] - 2\beta_1 E_{\mu, \phi} \left[\frac{1}{X_2 + \beta_2} \right] \right) \\ &\hspace{25em} + \beta_1^2 E_{\mu, \phi} \left[\frac{1}{(X_2 + \beta_2)^2} \right]. \end{aligned}$$

□

3.3.1 Risque de l'estimateur à vraisemblance maximale modifié

La Figure 3.2 représente graphiquement le risque de l'estimateur $\hat{\phi}_{EVMM}(\mathbf{x}) = \frac{x_1 + 0,5}{x_2 + 0,5}$ pour quelques valeurs fixées de μ .

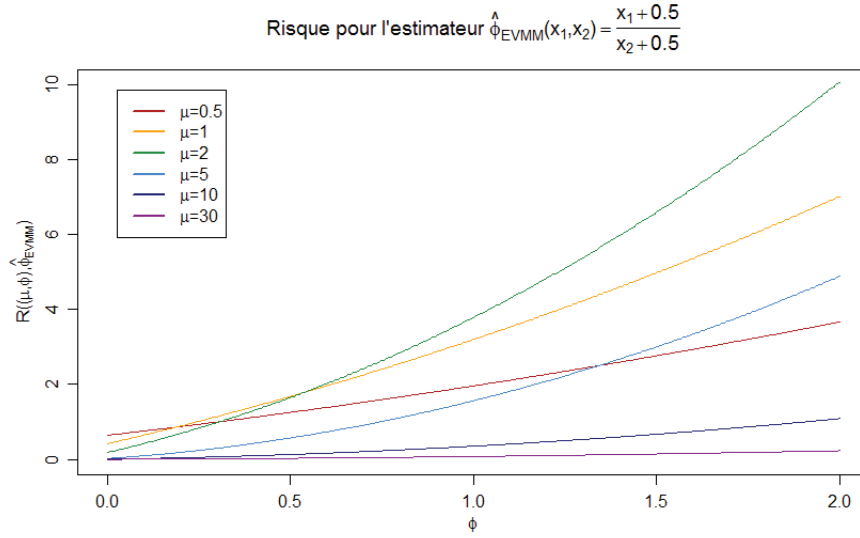


Figure 3.2 – Fonction de risque de $\hat{\phi}_{EVMM}(\mathbf{x}) = \frac{x_1 + 0.5}{x_2 + 0.5}$

Exemple 3.10. Pour $\mu = 2$, on peut trouver numériquement la fonction de risque selon ϕ en appliquant le Théorème 3.9. En effet, on a

$$E \left[\left(\frac{2}{X_2 + 0,5} - 1 \right)^2 \right] \approx 1,3428; \quad E \left[\frac{1}{(X_2 + 0,5)^2} \right] \approx 0,7257; \quad E \left[\frac{1}{X_2 + 0,5} \right] \approx 0,6400,$$

d'où $R((2, \phi), \hat{\phi}_{EVMM}) = 1,3428\phi^2 + 2,2628\phi + 0,1814$. Cette fonction coïncide avec la courbe de risque obtenue pour $\mu = 2$ du graphique de la Figure 3.2.

Dans la Figure 3.2, pour toutes les valeurs de μ étudiées, le risque est moindre pour de petites valeurs de ϕ , et ainsi, il est croissant par rapport ϕ . Pour ϕ très près de 0, le risque paraît décroissant en μ , soit une caractéristique qui n'est pas conservée lorsque ϕ augmente. On peut expliquer cette caractéristique à l'aide du Théorème 3.9. En effet, pour $\phi \approx 0$, on a

$$R((\mu, 0), \hat{\phi}_{EVMM}) = 0,5^2 E_\mu \left[\frac{1}{(X_2 + 0,5)^2} \right].$$

Afin d'étudier le comportement de la fonction de risque ci-dessus selon μ , on définira la

notion de famille de rapport de vraisemblance monotone croissant et une propriété s'y rattachant.

Définition 3.11. Une famille de densités $\mathcal{F} = \{p_\theta | \theta \in \Theta \subseteq \mathbb{R}\}$ d'une variable aléatoire X a un rapport de vraisemblance monotone croissant (respectivement décroissant), noté ici *RVMC* (respectivement *RVMD*), si pour tout $\theta_2 > \theta_1$, le rapport $\frac{p_{\theta_2}(x)}{p_{\theta_1}(x)}$ est une fonction monotone croissante (respectivement décroissante) en x .

Lemme 3.12 (Inégalité de covariance généralisée). Soient X une variable aléatoire et g et h des fonctions telles que $\text{Var}(g(X)) < \infty$ et $\text{Var}(h(X)) < \infty$. Si g et h sont croissantes, alors $\text{Cov}(h(X), g(X)) \geq 0$, c'est-à-dire $E[h(X)g(X)] \geq E[h(X)]E[g(X)]$.

Démonstration. Soient X_1, X_2 deux variables aléatoires de même loi que X , et g et h deux fonctions croissantes. On a

$$\begin{aligned} E[(g(X_1) - g(X_2))(h(X_1) - h(X_2))] \\ &= E[g(X_1)h(X_1)] + E[g(X_2)h(X_2)] \\ &\quad - E[g(X_2)]E[h(X_1)] - E[g(X_1)]E[h(X_2)], \quad \text{car } X_1 \perp X_2 \\ &= 2E[g(X)h(X)] - 2E[g(X)]E[h(X)] \\ &= 2\text{Cov}(g(X), h(X)). \end{aligned}$$

Comme g et h sont des fonctions croissantes, alors $g(X_1) - g(X_2)$ et $h(X_1) - h(X_2)$ seront toujours du même signe, c'est-à-dire positifs si $X_1 > X_2$ et négatifs si $X_1 < X_2$, d'où

$$2\text{Cov}(g(X), h(X)) = E[(g(X_1) - g(X_2))(h(X_1) - h(X_2))] \geq E[0] = 0. \quad \square$$

Remarques 3.13. 1) Si g et h sont décroissantes, alors on a le même résultat, car dans ce cas, $-g$ et $-h$ sont croissantes.

2) Si g est croissante et h décroissante (ou vice-versa), alors $\text{Cov}(h(X), g(X)) \leq 0$.

Proposition 3.14. Soit $\mathcal{F} = \{p_\theta : \theta \in \Theta \subseteq \mathbb{R}\}$ une famille de densités RVMC d'une variable aléatoire X . Si h est une fonction croissante (respectivement décroissante) en x , alors $\gamma(\theta) = E_\theta[h(X)]$ est une fonction croissante (respectivement décroissante) en θ .

Démonstration. Soit $\theta_2 > \theta_1$. Pour h croissante, on veut montrer que $\gamma(\theta_2) \geq \gamma(\theta_1)$.

$$\begin{aligned}
\gamma(\theta_2) &= E_{\theta_2}[h(X)] \\
&= \int_{\mathcal{X}} h(x)p_{\theta_2}(x)d\mu(x) \\
&= \int_{\mathcal{X}} h(x)\frac{p_{\theta_2}(x)}{p_{\theta_1}(x)}p_{\theta_1}(x)d\mu(x) \\
&= E_{\theta_1}\left[h(X)\frac{p_{\theta_2}(X)}{p_{\theta_1}(X)}\right] \\
&\geq E_{\theta_1}[h(X)]E_{\theta_1}\left[\frac{p_{\theta_2}(X)}{p_{\theta_1}(X)}\right] \\
&= \gamma(\theta_1)\int_{\mathcal{X}}\frac{p_{\theta_2}(x)}{p_{\theta_1}(x)}p_{\theta_1}(x)d\mu(x) \\
&= \gamma(\theta_1),
\end{aligned}$$

où l'inégalité découle du Lemme 3.12. □

Pour étudier le comportement de $R((\mu, 0), \hat{\phi}_{EVMM})$, on utilise le résultat de la Proposition 3.14. Comme $X_2|\mu \sim \mathcal{P}(\mu)$, alors pour tout $\mu_2 > \mu_1$, on a

$$\frac{p_{\mu_2}(x_2)}{p_{\mu_1}(x_2)} = \frac{\left(\frac{e^{-\mu_2}\mu_2^{x_2}}{x_2!}\right)}{\left(\frac{e^{-\mu_1}\mu_1^{x_2}}{x_2!}\right)} = e^{-(\mu_2-\mu_1)}\left(\frac{\mu_2}{\mu_1}\right)^{x_2}.$$

Il s'agit d'une fonction monotone croissante en x_2 , et donc, cette famille de lois de Poisson est RVMC. Or, comme $\frac{1}{(x_2+0,5)^2}$ est une fonction décroissante en x_2 , alors, de la Proposition 3.14, $R((\mu, 0), \hat{\phi}_{EVMM})$ est une fonction décroissante en μ . Ceci explique ainsi pourquoi le risque paraît décroissant en μ pour $\phi \approx 0$ à la Figure 3.2.

De plus, toujours pour $\phi \approx 0$, on peut borner supérieurement le risque de l'estimateur $\hat{\phi}_{EVMM}(x_1, x_2)$. En effet, dans ce cas particulier, on a

$$\begin{aligned} R\left((\mu, 0), \frac{x_1 + 0,5}{x_2 + 0,5}\right) &= 0,5^2 E\left[\frac{1}{(X_2 + 0,5)^2}\right] \\ &= \sum_{x_2=0}^{\infty} \frac{1}{(2x_2 + 1)^2} \frac{e^{-\mu} \mu^{x_2}}{x_2!} \\ &< \sum_{x_2=0}^{\infty} \frac{e^{-\mu} \mu^{x_2}}{(x_2 + 1)!} \\ &= \frac{1 - e^{-\mu}}{\mu}. \end{aligned}$$

On a $\lim_{\mu \rightarrow \infty} \frac{1 - e^{-\mu}}{\mu} = 0$ et cette borne supérieure est décroissante pour tout $\mu \geq 0$, car $\frac{d}{d\mu} \left(\frac{1 - e^{-\mu}}{\mu}\right) < 0$.

Par ailleurs, on remarque que pour de très grandes valeurs de μ , par exemple $\mu = 30$, le taux de croissance du risque semble plus faible, et ainsi, on peut penser que le risque demeure assez faible pour de très grandes valeurs de ϕ .

3.3.2 Risque des estimateurs bayésiens

Une des difficultés avec l'approche bayésienne est de choisir les valeurs des hyperparamètres a , b , c et d que l'on retrouve dans les lois *a priori*.

D'une part, dans les analyses présentées dans [Lin02], on adopte une approche bayésienne subjective. En effet, comme $X_2 \sim \mathcal{BN}\left(a, \frac{b}{T+b}\right)$, le choix des hyperparamètres a et b peut être basé sur les formules d'espérance et de variance de X_2 données à l'équation (3.5). Par exemple, dans un contexte où l'on veut tester l'efficacité d'un traitement (ou vaccin), il semble raisonnable de penser que les cliniciens ont une idée du comportement sans traitement (ou vaccin) du nombre de patients atteints. Ensuite, à nouveau en se basant sur certaines hypothèses du cliniciens, les hyperparamètres c et d peuvent être posés à

partir de la densité *a priori* de ϕ . Par exemple, dans la section 3.2, on a supposé que $\phi \sim \mathcal{B}_2\left(c, d, \frac{T+b}{S}\right)$, et donc, $E[\phi] = \frac{c(1+b)}{d-1}$, pour $d > 1$, et $Var(\phi) = \frac{c(c+d-1)(1+b)^2}{(d-1)^2(d-2)} = E[\phi] \frac{(c+d-1)(1+b)}{(d-1)(d-2)}$, pour $d > 2$. On peut alors se demander si les hyperparamètres c et d peuvent être choisis de telle sorte que $E[\phi]$ et $Var(\phi)$ correspondent à des valeurs fixées *a priori*. En effet, c'est possible. En utilisant les formules décrites ci-dessus pour $E[\phi]$ et $Var(\phi)$, on a, pour $b > -1$,

$$c = \frac{E[\phi]}{Var(\phi)(1+b)} (E[\phi](E[\phi] + b + 1) + Var(\phi)); \quad (3.19)$$

$$d = \frac{E[\phi](E[\phi] + b + 1)}{Var(\phi)} + 2. \quad (3.20)$$

D'autre part, dans [LL12], il est question de deux approches possibles : l'une dite semi-informative et l'autre dite non informative. Dans le premier cas, comme Lindley le fait, on utilise l'information disponible sur le taux de base μ pour trouver des valeurs raisonnables de a et b , alors qu'on utilise une loi *a priori* non informative pour ϕ . Par exemple, pour ce modèle, Laurent et Legrand suggèrent, selon la règle de Jeffreys, $c = \frac{1}{2}$ et $d = 0$, de telle sorte que la loi *a priori* impropre $\mathcal{B}_2\left(\frac{1}{2}, 0, \frac{T+b}{S}\right)$ soit utilisée pour ϕ [LL12]. Dans le deuxième cas, une approche complètement non informative est proposée. Comme vu à l'Exemple 1.10, selon la règle de Jeffreys, une loi *a priori* non informative pour le paramètre d'une loi de Poisson est la loi impropre $\mathcal{G}\left(\frac{1}{2}, 0\right)$. Et comme dans le cas précédent, on suppose $\phi \sim \mathcal{B}_2\left(\frac{1}{2}, 0, \frac{T+b}{S}\right)$. Par contre, dans notre cas, comme nous utilisons l'espérance *a posteriori* comme estimateur de ϕ , ces choix de lois *a priori* ne pourront pas être pris, car on doit avoir $d > 1 - a$, tel qu'établi à l'équation (3.12).

Exemple 3.15. Soient $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi, \mu)$, $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi \sim \mathcal{B}_2(c, d, 1+b)$ avec $a, c > 0$, $b > -1$ et $a + d > 1$. On illustre ici les approches subjective et semi-informative afin de choisir les hyperparamètres.

Selon une approche subjective ou semi-informative, on a recours tout d'abord au fait que

$E[X_2] = \frac{a}{b}$ et $Var(X_2) = \left(\frac{1+b}{b}\right) E[X_2]$. Par exemple, si on suppose que $E[X_2] = 30$ et $Var(X_2) = 45$, alors un choix raisonnable pour la loi a priori pour μ est $a = 60$ et $b = 2$.

Ensuite, pour le choix de c et de d , si on adopte une approche semi-informative, on peut utiliser la règle de Jeffreys, et ainsi, poser, par exemple, $c = \frac{1}{2}$ et $d = 0$. Si on adopte plutôt l'approche entièrement subjective, on peut utiliser les équations (3.19) et (3.20). Par exemple, si on suppose a priori qu'il n'y a aucune différence entre les taux des deux groupes, on peut poser $E[\phi] = 1$ avec $Var(\phi) = 0,1$. On trouve alors $c = \frac{41}{3}$ et $d = 42$.

Si on reprend les données utilisées à l'Exemple 3.7, c'est-à-dire $x_1 = 17$ et $x_2 = 31$, l'approche semi-informative mène à l'estimateur bayésien $\hat{\phi}_{B2}(17, 31) \approx 0,5833$ alors que l'approche subjective mène à $\hat{\phi}_{B2}(17, 31) \approx 0,6970$.

On procède maintenant à l'étude du risque de l'estimateur $\hat{\phi}_{B1}(\mathbf{x}) = (1+b) \left(\frac{x_1}{x_2+a}\right)$. Comme mentionné précédemment, le choix des hyperparamètres a et b peut être fait selon une approche subjective ou informative. Quelques fonctions de risque ont été illustrées en Annexe A pour différents choix de vecteurs (a, b) , et, pour chacun des graphiques qui y sont présentés, pour différentes valeurs de μ . Parmi les six graphiques, celui provenant du choix $(a, b) = (1, 0)$, c'est-à-dire celui pour lequel μ est issu d'une loi impropre uniforme sur les réels positifs, semble assez performant par rapport aux autres. Il est représenté à la Figure 3.3. Il s'agit également de l'estimateur abordé à la Proposition 3.8.

Du Théorème 3.9, avec $\beta_1 = 0$ et $\beta_2 = 1$, on a

$$R\left((\mu, \phi), \frac{x_1}{x_2+1}\right) = \phi^2 E\left[\left(\frac{\mu}{X_2+1} - 1\right)^2\right] + \phi\mu E\left[\frac{1}{(X_2+1)^2}\right]. \quad (3.21)$$

Pour tout $\mu \geq 0$, l'équation (3.21) permet de voir directement que le risque est nul en $\phi = 0$, comme on peut le constater également à la Figure 3.3, et qu'il est croissant en ϕ pour tout $\phi \geq 0$. De plus, en observant le graphique, on remarque que, pour de très grandes valeurs de μ , le risque est assez faible et croît plus lentement que pour de petites

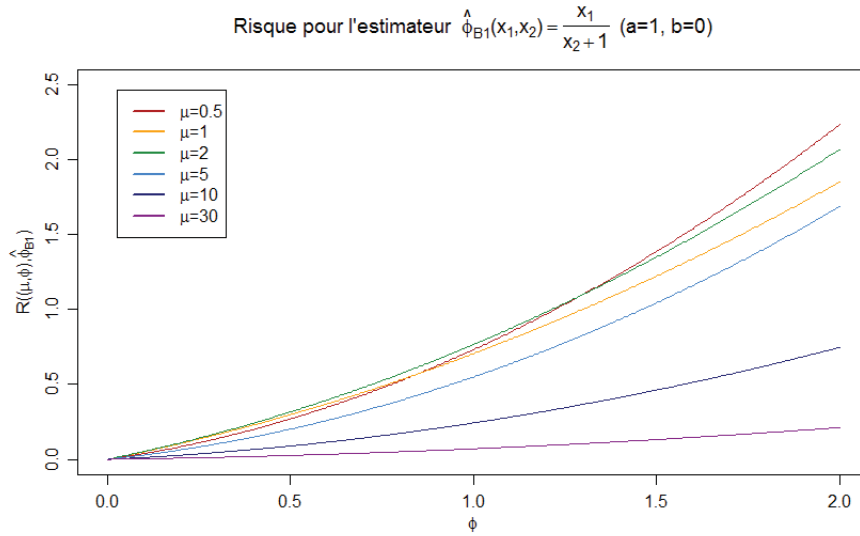


Figure 3.3 – Fonction de risque de l'estimateur $\hat{\phi}_{B1}(\mathbf{x}) = \frac{x_1}{x_2+1}$

valeurs de μ .

Exemple 3.16. À l'image de l'Exemple 3.10, pour $\mu = 2$, on peut trouver numériquement la fonction de risque selon ϕ en appliquant le Théorème 3.9. En effet, on a

$$E \left[\left(\frac{2}{X_2 + 1} - 1 \right)^2 \right] \approx 0,2678; \quad E \left[\frac{1}{(X_2 + 1)^2} \right] \approx 0,2493;$$

d'où $R((2, \phi), \hat{\phi}_{B1}) = 0,2678\phi^2 + 0,4986\phi$. Cette fonction coïncide avec la courbe de risque obtenue pour $\mu = 2$ du graphique de la Figure 3.3.

Enfin, on procède à l'étude du risque de l'estimateur $\hat{\phi}_{B2}(\mathbf{x}) = (1 + b) \left(\frac{x_1 + c}{x_2 + a + d - 1} \right)$. On retrouve en Annexe B des graphiques qui illustrent ce risque, qui a été évalué pour différents triplets $(a + d, b, c)$, et, pour chacun des graphiques présentés, pour différentes valeurs de μ . On voit clairement la forme parabolique des courbes des graphiques 4 et 5. Contrairement aux estimateurs $\hat{\phi}_{EVMM}(\mathbf{x})$ et $\hat{\phi}_{B1}(\mathbf{x})$ précédemment étudiés, les courbes de risque ne sont pas toujours croissantes, le sommet de la parabole se situant parfois à

une valeur de ϕ positive. Tout comme il l'a été fait pour l'estimateur $\hat{\phi}_{B1}(\mathbf{x})$, on a choisi un estimateur parmi ceux tracés en Annexe B, soit $\hat{\phi}_{B2}(\mathbf{x}) = \frac{x_1+0,5}{x_2+1,5}$, dont le risque est présenté à la Figure 3.4. On peut à nouveau se servir du Théorème 3.9 pour avoir la forme du risque en fonction de ϕ pour μ fixé, en posant $\beta_1 = 0,5$ et $\beta_2 = 1,5$.

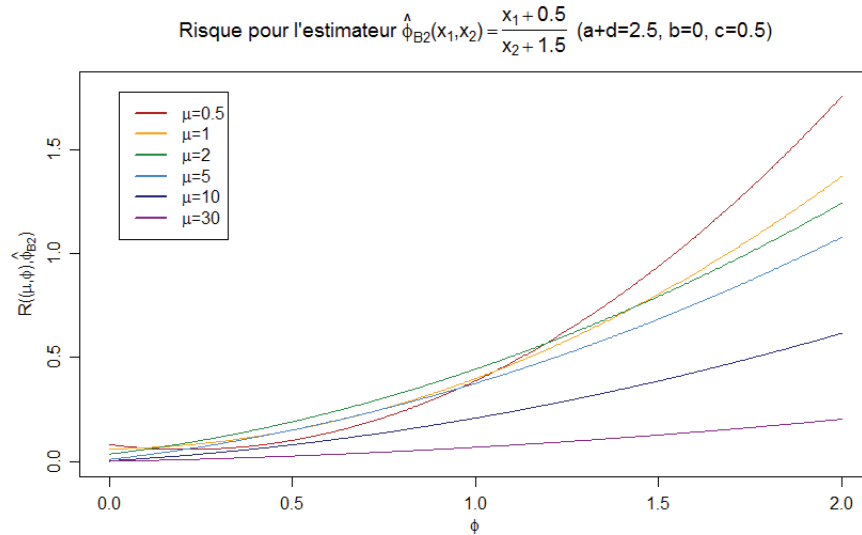


Figure 3.4 – Fonction de risque de l'estimateur $\hat{\phi}_{B2}(\mathbf{x}) = \frac{x_1+0.5}{x_2+1.5}$

Par un raisonnement analogue à celui effectué pour l'estimateur $\hat{\phi}_{EVMM}(\mathbf{x})$, pour $\phi = 0$, on peut montrer que le risque est décroissant en μ . Pour μ assez petit (par exemple, dans le graphique ci-dessus, $\mu = 0,5$), on remarque que le sommet des paraboles représentant le risque se situe à une valeur de $\phi > 0$ de telle sorte que la fonction de risque est décroissante sur un petit intervalle à partir de $\phi = 0$, puis croissante sur le reste de \mathbb{R}^+ . Finalement, comme c'était le cas pour les deux estimateurs étudiés précédemment, le risque demeure faible pour de très grandes valeurs de μ et croît plus lentement que pour de faibles valeurs de μ , pour ϕ suffisamment grand.

Exemple 3.17. *Tout comme les exemples 3.10 et 3.16, pour $\mu = 2$, on peut trouver numériquement la fonction de risque selon ϕ en appliquant le Théorème 3.9. En effet, on*

a

$$E \left[\left(\frac{2}{X_2 + 1,5} - 1 \right)^2 \right] \approx 0,1943; \quad E \left[\frac{1}{(X_2 + 1,5)^2} \right] \approx 0,1386; \quad E \left[\frac{1}{X_2 + 1,5} \right] \approx 0,3400,$$

d'où $R((2, \phi), \hat{\phi}_{B2}) = 0,1943\phi^2 + 0,2144\phi + 0,03465$. Cette fonction coïncide avec la courbe de risque obtenue pour $\mu = 2$ du graphique de la Figure 3.3.

3.3.3 Comparaison des risques des estimateurs

Nous procéderons ici à une comparaison des trois estimateurs choisis précédemment, c'est-à-dire $\hat{\phi}_{EVMM}(\mathbf{x}) = \frac{x_1+0,5}{x_2+0,5}$, $\hat{\phi}_{B1}(\mathbf{x}) = \frac{x_1}{x_2+1}$ et $\hat{\phi}_{B2}(\mathbf{x}) = \frac{x_1+0,5}{x_2+1,5}$ à partir de leur risque fréquentiste. Pour ce faire, on peut analyser la différence des risques ou leur quotient. Ici, une analyse préliminaire a permis de conclure que graphiquement, le quotient des risques semblait donner davantage d'information que leur différence.

Tout d'abord, comparons les estimateurs $\hat{\phi}_{EVMM}(\mathbf{x}) = \frac{x_1+0,5}{x_2+0,5}$ et $\hat{\phi}_{B1}(\mathbf{x}) = \frac{x_1}{x_2+1}$ à l'aide du quotient $\frac{R((\mu, \phi), \hat{\phi}_{B1})}{R((\mu, \phi), \hat{\phi}_{EVMM})}$. Pour tout $\mathbf{x} \in \mathbb{N}^2$, on a $\hat{\phi}_{B1}(\mathbf{x}) < \hat{\phi}_{EVMM}(\mathbf{x})$. Toutefois, à la Figure 3.5, le graphique permet de voir qu'aucun de ces estimateurs ne domine l'autre pour tout μ .

En effet, on voit que le quotient des risques pour $\mu = 0,5$ est croissant et traverse la droite $y = 1$ de telle sorte que pour cette valeur de μ , l'estimateur $\hat{\phi}_{B1}(\mathbf{x})$ est plus performant que l'estimateur $\hat{\phi}_{EVMM}(\mathbf{x})$ pour de petites valeurs de ϕ , et vice-versa. On voit d'ailleurs dans les Figures 3.2 et 3.3 que, pour $\mu = 0,5$, le risque associé à $\hat{\phi}_{B1}$ est nul pour $\phi = 0$, contrairement à celui pour $\hat{\phi}_{EVMM}$, mais qu'il a un taux de croissance plus élevé que ce dernier. On peut expliquer ce comportement en prenant le cas limite dégénéré, c'est-à-dire $\mu = 0$. En utilisant le résultat du Théorème 3.9, on trouve $R((0, \phi), \hat{\phi}_{EVMM}) = (\phi - 1)^2$ et $R((0, \phi), \hat{\phi}_{B1}) = \phi^2$. Ainsi,

$$\frac{R((0, \phi), \hat{\phi}_{B1})}{R((0, \phi), \hat{\phi}_{EVMM})} < 1 \iff \phi < \frac{1}{2}.$$

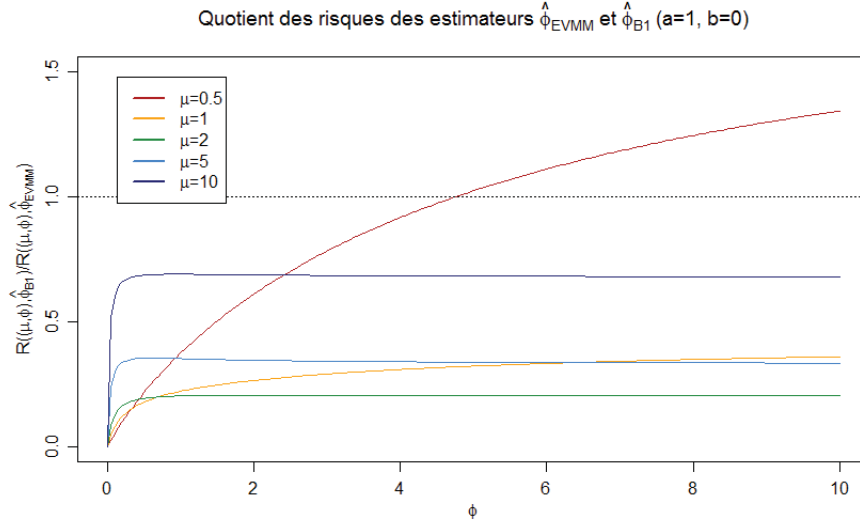


Figure 3.5 – Quotient des risques de $\hat{\phi}_{EVMM}(\mathbf{x}) = \frac{x_1+0,5}{x_2+0,5}$ et $\hat{\phi}_{B1}(\mathbf{x}) = \frac{x_1}{x_2+1}$

Par ailleurs, on remarque que certaines courbes de ce graphique semblent demeurer à peu près constantes pour des valeurs suffisamment grandes de ϕ , c'est-à-dire $\frac{R((\mu, \phi), \hat{\phi}_{B1})}{R((\mu, \phi), \hat{\phi}_{EVMM})} \approx c_\mu$, pour $c_\mu \in \mathbb{R}_+$. Ce comportement de stabilisation s'explique à partir de la représentation sous forme quadratique des fonctions de risque décrite au Théorème 3.9. En effet, pour deux estimateurs $\hat{\phi}_1$ et $\hat{\phi}_2$ de la forme $\frac{x_1+\beta_1}{x_2+\beta_2}$ avec $\beta_1 \geq 0$ et $\beta_2 > 0$, on a

$$\lim_{\phi \rightarrow \infty} \frac{R((\mu, \phi), \hat{\phi}_1)}{R((\mu, \phi), \hat{\phi}_2)} = \frac{A_1(\mu)}{A_2(\mu)},$$

où $A_1(\mu)$ et $A_2(\mu)$ représentent respectivement la fonction $A(\mu)$ (voir l'équation (3.16)) pour $\hat{\phi}_1$ et $\hat{\phi}_2$. Ainsi, toutes les fonctions de quotient de risques peuvent être représentées par des quotients de polynômes. Par exemple, si on reprend les courbes de risque trouvées aux exemples 3.10 et 3.16, pour $\mu = 2$, on a

$$\frac{R((2, \phi), \hat{\phi}_{B1})}{R((2, \phi), \hat{\phi}_{EVMM})} \approx \frac{0,2678\phi^2 + 0,4986\phi}{1,3428\phi^2 + 2,2628\phi + 0,1814},$$

soit une fonction qui tend vers $\frac{0,2678}{1,3428} \approx 0,1994$ lorsque $\phi \rightarrow \infty$. Ceci coïncide avec la courbe représentée par $\mu = 2$ à la Figure 3.5. De plus, il semble que plus μ est grand,

plus la fonction se rapproche rapidement de la valeur de convergence et plus cette valeur semble se rapprocher de 1.

En somme, pour ϕ suffisamment grand, un estimateur en domine un autre si son terme $A(\mu)$ est inférieur à celui de l'autre estimateur. Bien qu'à partir du graphique, cela ne semble pas le cas pour la courbe associée à $\mu = 0,5$, on peut confirmer qu'il y a stabilisation de la fonction pour de plus grandes valeurs de ϕ . En effet, à partir du Théorème 3.9, on trouve, en évaluant numériquement les espérances,

$$\frac{R((1/2, \phi), \hat{\phi}_{B1})}{R((1/2, \phi), \hat{\phi}_{EVMM})} \approx \frac{0,3860\phi^2 + 0,3458\phi}{0,1940\phi^2 + 1,1246\phi + 0,64355}.$$

Ce quotient tend vers 1,9897 lorsque $\phi \rightarrow \infty$, et ainsi, dans ce cas, $\hat{\phi}_{EVMM}$ domine $\hat{\phi}_{B1}$ pour tout ϕ suffisamment grand.

Ensuite, la Figure 3.6 illustre la comparaison des estimateurs $\hat{\phi}_{EVMM}(\mathbf{x}) = \frac{x_1+0,5}{x_2+0,5}$ et $\hat{\phi}_{B2}(\mathbf{x}) = \frac{x_1+0,5}{x_2+1,5}$ à partir du quotient des risques $\frac{R((\mu, \phi), \hat{\phi}_{EVMM})}{R((\mu, \phi), \hat{\phi}_{B2})}$.

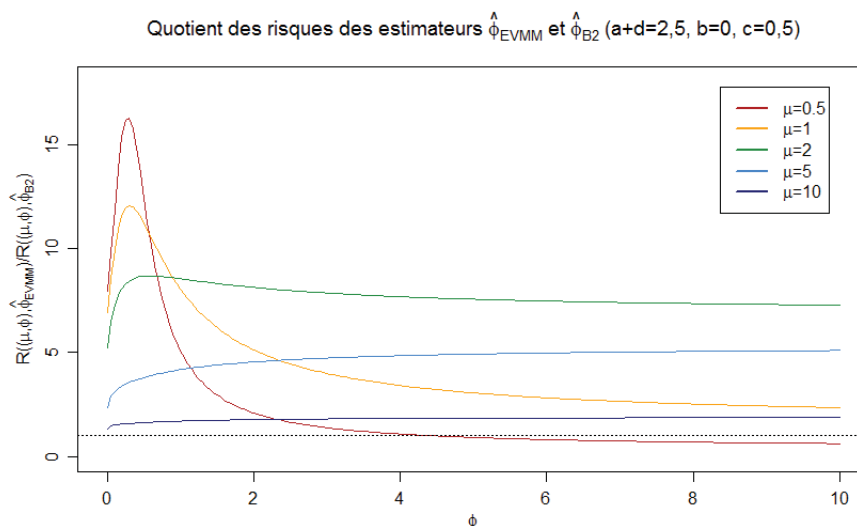


Figure 3.6 – Quotient des risques de $\hat{\phi}_{EVMM}(\mathbf{x}) = \frac{x_1+0,5}{x_2+0,5}$ et $\hat{\phi}_{B2}(\mathbf{x}) = \frac{x_1+0,5}{x_2+1,5}$

Tout comme pour la comparaison d'estimateurs précédente, le graphique permet de voir

qu'aucun de ces estimateurs ne domine l'autre pour toutes valeurs de μ , bien qu'on ait $\hat{\phi}_{B2}(\mathbf{x}) < \hat{\phi}_{EVMM}(\mathbf{x})$ pour tout $\mathbf{x} \in \mathbb{N}^2$. Encore une fois, la courbe représentée par $\mu = 0,5$ traverse la droite $y = 1$ de telle sorte que le risque de l'estimateur $\hat{\phi}_{EVMM}(\mathbf{x})$ soit supérieur à celui de l'estimateur $\hat{\phi}_{B2}(\mathbf{x})$ pour des valeurs de ϕ inférieures environ à 5. Lors de la comparaison d'estimateurs précédente, on avait vu que pour le cas dégénéré $\mu = 0$, on avait $R((\mu, \phi), \hat{\phi}_{EVMM}) = (\phi - 1)^2$. Par un raisonnement analogue, on trouve que $R((0, \phi), \hat{\phi}_{B2}) = (\phi - \frac{1}{3})^2$. Ainsi,

$$\frac{R((0, \phi), \hat{\phi}_{EVMM})}{R((0, \phi), \hat{\phi}_{B2})} < 1 \iff \phi > \frac{2}{3}.$$

En outre, le Théorème 3.9 nous permet également de trouver numériquement les fonctions de risque, ce qui nous permet d'obtenir que

$$\frac{R((1/2, \phi), \hat{\phi}_{EVMM})}{R((1/2, \phi), \hat{\phi}_{B2})} \approx \frac{0,1940\phi^2 + 1,1246\phi + 0,64355}{0,5308\phi^2 - 0,2254\phi + 0,08125},$$

et donc, que le quotient tend vers 0,3655 lorsque $\phi \rightarrow \infty$. Pour μ près de 0, l'estimateur $\hat{\phi}_{EVMM}$ domine donc l'estimateur $\hat{\phi}_{B2}$ pour ϕ suffisamment grand. Toutefois, on voit que sur les autres courbes de la Figure 3.4, le quotient semble vouloir se stabiliser au-dessus de la droite $y = 1$ de telle sorte que l'estimateur $\hat{\phi}_{B2}$ domine l'estimateur $\hat{\phi}_{EVMM}$ pour de telles valeurs de μ et de ϕ .

Par ailleurs, peu importe la valeur de μ , l'estimateur $\hat{\phi}_{B2}$ domine l'estimateur $\hat{\phi}_{EVMM}$ lorsque ϕ est près de 0. Il suffit de constater que pour $\phi = 0$ le quotient des risques est équivalent au quotient des fonctions $C(\mu)$ telles que définies à l'équation (3.18). Ainsi, dans ce cas-ci, pour tout $\mu \geq 0$, on a

$$C_{EVMM}(\mu) = 0,5^2 E \left[\frac{1}{(X_2 + 0,5)^2} \right] > 0,5^2 E \left[\frac{1}{(X_2 + 1,5)^2} \right] = C_{B2}(\mu),$$

et donc, $\frac{C_{EVMM}(\mu)}{C_{B2}(\mu)} > 1$, d'où le résultat.

Enfin, la Figure 3.7 illustre la comparaison des estimateurs $\hat{\phi}_{B1}(\mathbf{x}) = \frac{x_1}{x_2+1}$ et $\hat{\phi}_{B2}(\mathbf{x}) = \frac{x_1+0.5}{x_2+1.5}$.

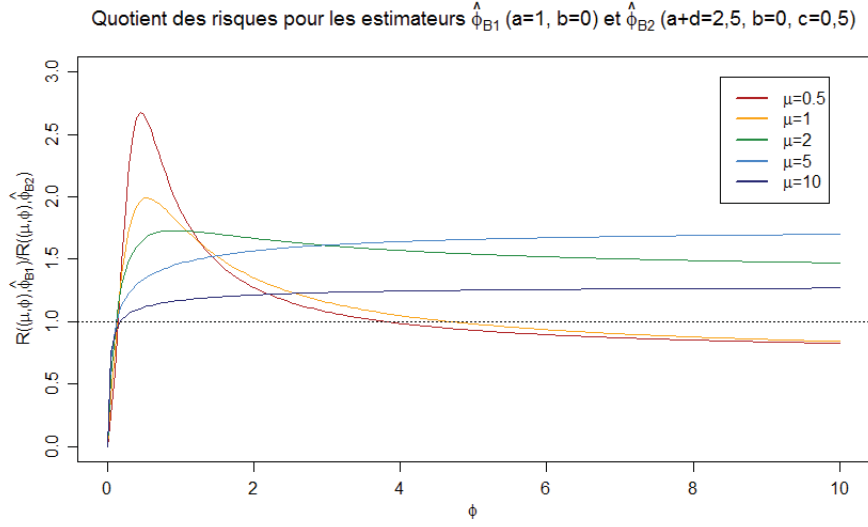


Figure 3.7 – Quotient des risques de $\hat{\phi}_{B1}(\mathbf{x}) = \frac{x_1}{x_2+1}$ et $\hat{\phi}_{B2}(\mathbf{x}) = \frac{x_1+0,5}{x_2+1,5}$

En ce qui concerne la comparaison des risques, pour tout μ , il n'y a pas un estimateur qui domine l'autre selon les courbes obtenues dans le graphique. En effet, pour $\phi \approx 0$, l'estimateur $\hat{\phi}_{B1}(\mathbf{x})$ semble plus performant que l'estimateur $\hat{\phi}_{B2}(\mathbf{x})$ pour tout $\mu > 0$. Ceci peut s'expliquer par le fait que pour $\phi = 0$, on a $\hat{\phi}_{B1}(\mathbf{x}) = 0$, d'où $R((\mu, 0), \hat{\phi}_{B1}) = 0$, et $\hat{\phi}_{B2}(0, x_2) = \frac{0,5}{x_2+1,5} > 0$, d'où $R((\mu, 0), \hat{\phi}_{B2}) > 0$.

Au fur et à mesure que ϕ s'éloigne de 0, pour les courbes représentant $\mu = 2$, $\mu = 5$ et $\mu = 10$, le quotient des risques semble également se stabiliser au-dessus de la droite $y = 1$, suggérant ainsi que pour ϕ et μ suffisamment grands, l'estimateur $\hat{\phi}_{B2}(\mathbf{x})$ est plus performant que l'autre. Toutefois, pour de faibles valeurs de μ , par exemple, pour $\mu = 0,5$ et pour $\mu = 1$, bien que le quotient des risques soit supérieur à 1 pour ϕ légèrement supérieur à 0, ce quotient repasse sous cette valeur pour ϕ suffisamment grand. Donc l'estimateur $\hat{\phi}_{B1}$ domine l'estimateur $\hat{\phi}_{B2}$ lorsque μ est suffisamment petit et ϕ suffisamment grand.

3.4 Inférence prédictive

La dernière section de ce chapitre concerne l'inférence prédictive. Supposons qu'après avoir observé le vecteur aléatoire $\mathbf{X} = (X_1, X_2)$, on veuille estimer la densité de futures observations $\mathbf{Y} = (Y_1, Y_2)$. Tout comme pour $\mathbf{X}|\mu, \phi$, on suppose que $\mathbf{Y}|\mu, \phi \sim \mathcal{P}2(\mu\phi S', \mu T')$, où S' et T' sont des constantes positives pouvant être différentes de S et T . Ainsi, la densité que l'on cherche à estimer est

$$q(\mathbf{y}|\mu, \phi) = \frac{S'^{y_1} T'^{y_2}}{y_1! y_2!} e^{-\mu(T'+\phi S')} \mu^{y_1+y_2} \phi^{y_1}, \quad (3.22)$$

que nous noterons plutôt $q_{\mu, \phi}(\mathbf{y})$ pour la suite. En continuité avec l'estimation ponctuelle, on suppose que $\mu|\phi \sim \mathcal{G}(a, b)$ et que $\phi \sim \mathcal{B}_2\left(c, d, \frac{T+b}{S}\right)$.

D la Proposition 3.2, on constate entre autres que X_2 est indépendant du paramètre ϕ . De plus, comme les distributions de $X_1|\phi$ et $X_1|X_2 = x_2, \phi$ sont des binomiales négatives, les densités de X_1 et de $X_1|X_2 = x_2$ s'écrivent toujours comme un mélange de telles lois. En effet, d'une part, pour X_1 , on a

$$X_1|\phi \sim \mathcal{BN}\left(a, \frac{b}{b + \phi S}\right) \text{ avec } \phi \sim \mathcal{B}_2\left(c, d, \frac{T+b}{S}\right),$$

ou, de manière équivalente,

$$X_1|U = u \sim \mathcal{BN}\left(a, \frac{1}{1+u}\right) \text{ avec } U \sim \mathcal{B}_2\left(c, d, \frac{T+b}{b}\right).$$

D'autre part, pour $X_1|X_2 = x_2$, on a

$$X_1|X_2 = x_2, \phi \sim \mathcal{BN}\left(a + x_2, \frac{T+b}{T+b + \phi S}\right) \text{ avec } \phi \sim \mathcal{B}_2\left(c, d, \frac{T+b}{S}\right),$$

ou, de manière équivalente,

$$X_1|X_2 = x_2, V = v \sim \mathcal{BN}\left(a + x_2, \frac{1}{1+v}\right) \text{ avec } V \sim \mathcal{B}_2(c, d, 1).$$

Le prochain théorème, qui peut être vu comme un corollaire de la Proposition 3.2, présente les densités qui résultent des mélanges précédents. Il s'agit en fait d'un cas particulier du Résultat 3.1 de l'article de Laurent [Lau12], auquel on présentera une preuve par calcul direct contrairement à Laurent, qui utilise plutôt des liens entre des lois préalablement développées dans son article.

Théorème 3.18. *Soient $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$, $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi \sim \mathcal{B}_2\left(c, d, \frac{T+b}{S}\right)$ avec $a, b, c, d, S, T > 0$. Alors on a les fonctions de masse sur \mathbb{N} (ou \mathbb{N}^2) suivantes :*

1) *Fonction de masse de \mathbf{X}*

$$p_{\mathbf{X}}(\mathbf{x}) = \left(\frac{b}{T+b}\right)^a \frac{(a)_{x_1+x_2} (c)_{x_1} (d)_{a+x_2}}{x_1! x_2! (c+d)_{a+x_1+x_2}} \left(\frac{T}{T+b}\right)^{x_2}; \quad (3.23)$$

2) *Fonction de masse marginale de X_1*

$$p_{X_1}(x_1) = \frac{\left(\frac{b}{T+b}\right)^c (a)_{x_1}}{B(c, d) x_1!} B(c+x_1, a+d)_2 F_1\left(c+x_1, c+d; a+c+d+x_1; \frac{T}{T+b}\right); \quad (3.24)$$

3) *Fonction de masse conditionnelle de X_1 sachant X_2*

$$p_{X_1|X_2}(x_1|x_2) = \frac{(a+x_2)_{x_1} B(c+x_1, d+a+x_2)}{x_1! B(c, d)}. \quad (3.25)$$

Démonstration. On se base sur les résultats de la Proposition 3.2.

1) On a

$$\begin{aligned} p_{\mathbf{X}}(\mathbf{x}) &= \frac{(a)_{x_1+x_2} b^a S^{x_1} T^{x_2}}{x_1! x_2!} \\ &\quad \times \int_0^\infty \frac{\phi^{x_1}}{(T+b+\phi S)^{a+x_1+x_2}} \times \frac{\left(\frac{T+b}{S}\right)^{-c}}{B(c, d)} \frac{\phi^{c-1}}{\left(1+\frac{S}{T+b}\phi\right)^{c+d}} d\phi \\ &= \frac{(a)_{x_1+x_2} b^a S^{x_1} T^{x_2} \left(\frac{T+b}{S}\right)^{-c}}{x_1! x_2! B(c, d) (T+b)^{a+x_1+x_2}} \int_0^\infty \frac{\phi^{c+x_1-1}}{\left(1+\frac{S}{T+b}\phi\right)^{a+c+d+x_1+x_2}} d\phi \end{aligned}$$

$$\begin{aligned}
&= \frac{(a)_{x_1+x_2}}{x_1!x_2!} \frac{b^a S^{x_1} T^{x_2} \left(\frac{T+b}{S}\right)^{-c}}{B(c, d)(T+b)^{a+x_1+x_2}} \frac{B(c+x_1, a+d+x_2)}{\left(\frac{T+b}{S}\right)^{-(c+x_1)}} \\
&= \left(\frac{b}{T+b}\right)^a \frac{(a)_{x_1+x_2}(c)_{x_1}(d)_{a+x_2}}{x_1!x_2!(c+d)_{a+x_1+x_2}} \left(\frac{T}{T+b}\right)^{x_2}.
\end{aligned}$$

2) On a

$$\begin{aligned}
p_{X_1}(x_1) &= \frac{(a)_{x_1}}{x_1!} b^a S^{x_1} \int_0^\infty \frac{\phi^{x_1}}{(b+\phi S)^{a+x_1}} \times \frac{\left(\frac{T+b}{S}\right)^{-c}}{B(c, d)} \frac{\phi^{c-1}}{\left(1+\frac{S}{T+b}\phi\right)^{c+d}} d\phi \\
&= \frac{(a)_{x_1}}{x_1!} \frac{b^a S^{x_1}}{b^{a+x_1}} \frac{1}{B(c, d)} \left(\frac{S}{T+b}\right)^c \int_0^\infty \frac{\phi^{c+x_1-1} \left(1+\frac{S}{b}\phi\right)^{-(a+x_1)}}{\left(1+\frac{S}{T+b}\phi\right)^{c+d}} d\phi.
\end{aligned}$$

En posant $U = \frac{S}{b}\phi$, on obtient

$$\begin{aligned}
p_{X_1}(x_1) &= \frac{(a)_{x_1}}{x_1!} \frac{1}{B(c, d)} \left(\frac{b}{T+b}\right)^c \int_0^\infty \frac{u^{c+x_1-1} (1+u)^{-(a+x_1)}}{\left(1+\frac{b}{T+b}u\right)^{c+d}} du \\
&= \frac{\left(\frac{b}{T+b}\right)^c}{B(c, d)} \frac{(a)_{x_1}}{x_1!} B(c+x_1, a+d) {}_2F_1\left(c+x_1, c+d; a+c+d+x_1; \frac{T}{T+b}\right),
\end{aligned}$$

où la dernière égalité découle d'une représentation sous forme d'intégrale de la fonction hypergéométrique telle que donnée à l'équation (2.5.b).

3) On a

$$\begin{aligned}
p_{X_1|X_2}(x_1|x_2) &= \frac{(a+x_2)_{x_1}}{x_1!} (T+b)^{a+x_2} S^{x_1} \\
&\quad \times \int_0^\infty \frac{\phi^{x_1}}{(T+b+\phi S)^{a+x_1+x_2}} \times \frac{\left(\frac{T+b}{S}\right)^{-c}}{B(c, d)} \frac{\phi^{c-1}}{\left(1+\frac{S}{T+b}\phi\right)^{c+d}} d\phi
\end{aligned}$$

$$\begin{aligned}
&= \frac{(a+x_2)_{x_1}}{x_1!} \left(\frac{S}{T+b}\right)^{c+x_1} \frac{1}{B(c,d)} \int_0^\infty \frac{\phi^{c+x_1-1}}{\left(1 + \frac{S}{T+b}\phi\right)^{a+x_1+x_2+c+d}} d\phi \\
&= \frac{(a+x_2)_{x_1}}{x_1!} \left(\frac{S}{T+b}\right)^{c+x_1} \frac{B(c+x_1, d+a+x_2)}{B(c,d)} \left(\frac{T+b}{S}\right)^{c+x_1} \\
&= \frac{(a+x_2)_{x_1}}{x_1!} \frac{B(c+x_1, d+a+x_2)}{B(c,d)}. \quad \square
\end{aligned}$$

Remarque 3.19. *On a vu précédemment que quelques fonctions de masses de ce théorème pouvaient être réécrites sous forme de mélange. Celles-ci sont d'ailleurs présentées à travers l'article de Laurent [Lau12], de même que leur construction à partir de mélange. En effet, on remarque tout d'abord que la loi marginale de X_1 est en fait la loi dite, par Laurent [Lau12], Poisson-Gamma-Bêta, tel que montré dans le Résultat 3.1 de ce même article. De plus, la loi conditionnelle de $X_1|X_2 = x_2$ est présentée dans l'article de Laurent [Lau12] comme étant la loi Poisson-Gamma-Bêta prime.*

Ces lois permettent entre autres d'analyser l'impact de l'information fournie par des données recueillies sur la prédiction de données futures. Par exemple, la fonction de masse de $X_1|X_2 = x_2$ est une fonction de masse prédictive bayésienne pour X_1 sous la loi *a priori* donnée sur (μ, ϕ) . Dans la prochaine section, on verra comment utiliser des données actuelles pour prédire des données futures.

3.4.1 Estimateur de Bayes

On passe maintenant au problème d'analyse prédictive qui vise à prédire des tendances futures (Y_1, Y_2) en extrayant l'information fournie par les données actuelles (X_1, X_2) . Tel que vu à l'équation (1.33), une densité prédictive usuelle pour $\mathbf{Y} \sim q_{\mu, \phi}(\mathbf{y})$ est donnée

par

$$\hat{q}_\pi(\mathbf{y}; \mathbf{x}) = \int_0^\infty \int_0^\infty q_{\mu, \phi}(\mathbf{y}) \pi(\mu, \phi | \mathbf{x}) d\mu d\phi,$$

où $\pi(\mu, \phi | \mathbf{x})$ est la densité *a posteriori* de (μ, ϕ) . Tout comme ce fut le cas pour l'estimation ponctuelle dans un cadre bayésien, on supposera que $\mu | \phi \sim \mathcal{G}(a, b)$ et que $\phi \sim \mathcal{B}_2\left(c, d, \frac{T+b}{S}\right)$, avec $a, c > 0$, $b > -T$ et $d > -a$, et ainsi, du Théorème 3.4 et du Théorème 3.6, $\mu | \phi, \mathbf{x} \sim \mathcal{G}(a + x_1 + x_2, T + b + \phi S)$ et $\phi | \mathbf{x} \sim \mathcal{B}_2\left(c + x_1, d + a + x_2, \frac{T+b}{S}\right)$.

Avant de poursuivre, on définira une loi qui apparaîtra dans l'analyse prédictive du modèle bivarié à l'étude. Cette loi a été présentée dans [Lau12] comme étant la loi dite de Bailey, soit une appellation que l'on conservera ici.

Définition 3.20. *On dit que $\mathbf{X} = (X_1, X_2)$ se distribue selon la loi de Bailey, notée $\text{Bailey}(a, c, d, \rho)$, si sa fonction de masse s'écrit*

$$p(\mathbf{x}) = \left(\frac{\rho}{1+\rho}\right)^a \frac{(a)_{x_1+x_2} (c)_{x_1} (d)_{x_2}}{x_1! x_2! (c+d)_{x_1+x_2}} \left(\frac{1}{1+\rho}\right)^{x_1+x_2} \mathbb{1}_{\mathbb{N}^2}(\mathbf{x}), \quad a, c, d, \rho > 0. \quad (3.26)$$

Remarque 3.21. *On peut facilement montrer qu'il s'agit d'une fonction de masse à l'aide des propriétés de la première fonction hypergéométrique d'Appell, comme l'explique Laurent dans [Lau12]. En effet, à partir des Propriétés 2.5 b) et c), on a l'égalité suivante*

$$F_1(a; c, d; c+d; z, z) = (1-z)^{-a} {}_2F_1(a, c; c+d; 0),$$

$$d'où \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} (1-z)^a \frac{(a)_{x_1+x_2} (c)_{x_1} (d)_{x_2}}{x_1! x_2! (c+d)_{x_1+x_2}} z^{x_1+x_2} = 1.$$

En reparamétrisant par $z = \frac{1}{1+\rho}$, on obtient bien

$$\sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} \left(\frac{\rho}{1+\rho}\right)^a \frac{(a)_{x_1+x_2} (c)_{x_1} (d)_{x_2}}{x_1! x_2! (c+d)_{x_1+x_2}} \left(\frac{1}{1+\rho}\right)^{x_1+x_2} = 1.$$

Pour $a = c + d$, la loi de Bailey se réduit au produit de deux lois binomiales négatives indépendantes, c'est-à-dire si $\mathbf{X} \sim \text{Bailey}(c+d, c, d, \rho)$, alors $X_1 \sim \mathcal{BN}\left(c, \frac{\rho}{1+\rho}\right)$ et $X_2 \sim \mathcal{BN}\left(d, \frac{\rho}{1+\rho}\right)$. De plus, la loi de Bailey admet une représentation sous forme de mélange de lois de Poisson indépendantes.

Proposition 3.22. *Si $(X_1, X_2) \sim \text{Bailey}(a, c, d, \rho)$, alors $\mathbf{X} = (X_1, X_2)$ admet la représentation sous forme de mélange suivante :*

$$\mathbf{X}|\mu, \phi \sim \mathcal{P}2\left(\frac{\mu\phi}{\rho}, \frac{\mu}{\rho}\right), \quad \mu|\phi \sim \mathcal{G}(a, 1 + \phi) \quad \text{et} \quad \phi \sim \mathcal{B}_2(c, d, 1).$$

Démonstration. On a

$$\begin{aligned} p(x_1, x_2) &= \left(\frac{\rho}{1 + \rho}\right)^a \frac{(a)_{x_1+x_2}(c)_{x_1}(d)_{x_2}}{x_1!x_2!(c+d)_{x_1+x_2}} \left(\frac{1}{1 + \rho}\right)^{x_1+x_2} \\ &= \left(\frac{\rho}{1 + \rho}\right)^a \left(\frac{1}{1 + \rho}\right)^{x_1+x_2} \frac{(a)_{x_1+x_2}}{x_1!x_2!} \frac{B(c + x_1, d + x_2)}{B(c, d)} \\ &= \frac{\rho^{-(x_1+x_2)} \left(1 + \frac{1}{\rho}\right)^{-(a+x_1+x_2)} \Gamma(a + x_1 + x_2)}{x_1!x_2!\Gamma(a)B(c, d)} \int_0^\infty \frac{\phi^{c+x_1-1}}{(1 + \phi)^{c+d+x_1+x_2}} d\phi \\ &= \frac{\rho^{-(x_1+x_2)}}{x_1!x_2!\Gamma(a)B(c, d)} \int_0^\infty \frac{\Gamma(a + x_1 + x_2) \left((1 + \phi) \left(1 + \frac{1}{\rho}\right)\right)^{-(a+x_1+x_2)} \phi^{c+x_1-1}}{(1 + \phi)^{c+d-a}} d\phi \\ &= \frac{\rho^{-(x_1+x_2)}}{x_1!x_2!\Gamma(a)B(c, d)} \int_0^\infty \int_0^\infty \frac{e^{-\mu(1+\phi)(1+1/\rho)} \mu^{a+x_1+x_2-1} \phi^{c+x_1-1}}{(1 + \phi)^{c+d-a}} d\mu d\phi \\ &= \int_0^\infty \int_0^\infty \frac{e^{-\mu\phi/\rho} (\mu\phi/\rho)^{x_1}}{x_1!} \frac{e^{-\mu/\rho} (\mu/\rho)^{x_2}}{x_2!} \frac{(1 + \phi)^a e^{-\mu(1+\phi)} \mu^{a-1}}{\Gamma(a)} \\ &\quad \times \frac{1}{B(c, d)} \frac{\phi^{c-1}}{(1 + \phi)^{c+d}} d\mu d\phi. \end{aligned}$$

□

Corollaire 3.23. *Si $\mathbf{X} \sim \text{Bailey}(a, c, d, \rho)$, alors*

$$X_1 + X_2|\psi \sim \mathcal{P}(\psi),$$

avec $\psi \sim \mathcal{G}(a, \rho)$.

Démonstration. Du Lemme 1.7 et de la Proposition 3.22, $X_1 + X_2 | \mu, \phi \sim \mathcal{P}2\left(\frac{\mu(1+\phi)}{\rho}\right)$, où $\mu | \phi \sim \mathcal{G}(a, 1+\phi)$ et $\phi \sim \mathcal{B}_2(c, d, 1)$. Posons $\psi = \frac{\mu(1+\phi)}{\rho}$. Or, comme la fonction génératrice des moments de $\mu | \phi$ est $M_{\mu | \phi}(t) = \left(1 - \frac{t}{1+\phi}\right)^{-a}$ pour $t < 1 + \phi$, alors

$$\begin{aligned} M_\psi(t) &= E^\phi \left[E^{\mu | \phi} \left[e^{t \frac{\mu(1+\phi)}{\rho}} \right] \right] \\ &= E^\phi \left[M_{\mu | \phi} \left(\frac{t(1+\phi)}{\rho} \right) \right] \\ &= E^\phi \left[\left(1 - \frac{t(1+\phi)}{(1+\phi)\rho} \right)^{-a} \right] \\ &= \left(1 - \frac{t}{\rho} \right)^{-a}, \end{aligned}$$

ce qui correspond à la fonction génératrice des moments d'une loi $\mathcal{G}(a, \rho)$. □

Corollaire 3.24. *Si $\mathbf{X} \sim \text{Bailey}(a, c, d, \rho)$, alors*

$$E[X_1] = \frac{ac}{\rho(c+d)} \text{ et } E[X_2] = \frac{ad}{\rho(c+d)}.$$

Démonstration. Pour $\mu | \phi \sim \mathcal{G}(a, 1+\phi)$ et $\phi \sim \mathcal{B}_2(c, d, 1)$, de la Proposition 3.22, comme

$$E[X_1] = E^{\mu, \phi} [E_{\mu, \phi} [X_1]] = \frac{1}{\rho} E^\phi [\phi E^{\mu | \phi} [\mu]] = \frac{a}{\rho} E^\phi \left[\frac{\phi}{1+\phi} \right],$$

on a

$$E[X_1] = \frac{a}{\rho B(c, d)} \int_0^\infty \frac{\phi^{(c+1)-1}}{(1+\phi)^{c+d+1}} d\phi = \frac{aB(c+1, d)}{\rho B(c, d)} = \frac{ac}{\rho(c+d)}.$$

Puis, de la Proposition 3.22 et du Corollaire 3.23, on a

$$\begin{aligned} E[X_2] &= E^{\mu, \phi} [E_{\mu, \phi} [X_1 + X_2]] - E[X_1] \\ &= E^{\mu, \phi} \left[\frac{\mu(1+\phi)}{\rho} \right] - E[X_1] \end{aligned}$$

$$\begin{aligned}
&= \frac{a}{\rho} - \frac{ac}{\rho(c+d)} \\
&= \frac{ad}{\rho(c+d)}. \quad \square
\end{aligned}$$

La Proposition 3.22 implique que $\mathbf{X}|Z, W \sim \mathcal{P}2(Z, W)$ avec $Z \stackrel{d}{=} \frac{\mu\phi}{\rho}$ et $W \stackrel{d}{=} \frac{\mu}{\rho}$, où $\mu|\phi \sim \mathcal{G}(a, 1 + \phi)$ et $\phi \sim \mathcal{B}_2(c, d, 1)$. Ainsi, les marginales d'une loi de Bailey et la somme des variables s'expriment comme des mélanges de lois de Poisson. Cette représentation permet d'ailleurs de constater que lorsque $a = c + d$, les lois marginales de Z et de W sont des lois de gamma, comme le montre dans le lemme suivant.

Lemme 3.25. Soient $Z = \frac{\mu\phi}{\rho}$ et $W = \frac{\mu}{\rho}$ pour $\rho > 0$, $\mu|\phi \sim \mathcal{G}(a, 1 + \phi)$ et $\phi \sim \mathcal{B}_2(c, d, 1)$. Si $a = c + d$, alors on a

$$Z \sim \mathcal{G}(c, \rho) \text{ et } W \sim \mathcal{G}(d, \rho).$$

Démonstration. Soit $t > 0$. Pour $Z = \frac{\mu\phi}{\rho}$, il suffit de montrer que $M_Z(t) = \left(1 + \frac{t}{\rho}\right)^{-c}$.

On a

$$\begin{aligned}
M_Z(t) &= E^\phi \left[E^{\mu|\phi} \left[e^{t\frac{\mu\phi}{\rho}} \right] \right] \\
&= E^\phi \left[M_{\mu|\phi} \left(\frac{t\phi}{\rho} \right) \right] \\
&= E^\phi \left[\left(1 - \frac{t\phi}{\rho(1+\phi)} \right)^{-(c+d)} \right] \\
&= \frac{1}{B(c, d)} \int_0^\infty \frac{\phi^c}{\left(\left(1 - \frac{t\phi}{\rho(1+\phi)} \right) (1 + \phi) \right)^{c+d}} d\phi \\
&= \frac{1}{B(c, d)} \int_0^\infty \frac{\phi^c}{\left(1 + \left(1 - \frac{t}{\rho} \right) \phi \right)^{c+d}} d\phi
\end{aligned}$$

$$= \left(1 + \frac{t}{\rho}\right)^{-c},$$

car de la loi $\mathcal{B}_2\left(c, d, \frac{1}{1-\frac{t}{\rho}}\right)$, l'intégrale vaut $B(c, d) \left(1 + \frac{t}{\rho}\right)^{-c}$.

Pour $W = \frac{\mu}{\rho}$, si $w \geq 0$, on a

$$\begin{aligned} f_W(w) &= \rho \int_0^\infty \frac{(1+\phi)^{c+d} (\rho w)^{c+d-1} e^{-(1+\phi)\rho w}}{\Gamma(c+d)} \cdot \frac{1}{B(c,d)} \frac{\phi^{c-1}}{(1-\phi)^{c+d}} d\phi \\ &= \frac{\rho^{c+d} w^{c+d-1} e^{-\rho w} \Gamma(c)}{B(c,d) \Gamma(c+d) (\rho w)^c} \\ &= \frac{\rho^d w^{d-1} e^{-\rho w}}{\Gamma(d)}. \end{aligned} \quad \square$$

Comme mentionné précédemment, les lois marginales de X_1 et X_2 d'une loi de Bailey dans le cas où $a = c + d$ sont des lois binomiales négatives. On peut alors se servir de l'Exemple 1.17 et du lemme précédent pour le montrer.

Corollaire 3.26. *Soit $\mathbf{X} \sim \text{Bailey}(c + d, c, d, \rho)$. Alors*

$$X_1 \sim \mathcal{BN}\left(c, \frac{\rho}{1+\rho}\right) \text{ et } X_2 \sim \mathcal{BN}\left(d, \frac{\rho}{1+\rho}\right).$$

Bien que les lois marginales d'une loi de Bailey puissent s'exprimer sous forme de mélange de lois de Poisson, on peut tout de même calculer la densité résultante, dans un cas plus général que celui du Corollaire 3.26, tel que présentée dans la prochaine proposition.

Proposition 3.27. *Si $\mathbf{X} \sim \text{Bailey}(a, c, d, \rho)$, alors les fonctions de masse marginales de X_1 et X_2 sont respectivement*

$$p_{X_1}(x_1) = \left(\frac{\rho}{1+\rho}\right)^a \frac{(a)_{x_1} (c)_{x_1}}{x_1! (c+d)_{x_1}} \left(\frac{1}{1+\rho}\right)^{x_1} {}_2F_1\left(a+x_1, d; c+d+x_1; \frac{1}{1+\rho}\right) \mathbf{1}_{\mathbb{N}}(x_1), \quad (3.27)$$

et

$$p_{X_2}(x_2) = \left(\frac{\rho}{1+\rho}\right)^a \frac{(a)_{x_2}(d)_{x_2}}{x_2!(c+d)_{x_2}} \left(\frac{1}{1+\rho}\right)^{x_2} {}_2F_1\left(a+x_2, c; c+d+x_2; \frac{1}{1+\rho}\right) \mathbb{1}_{\mathbb{N}}(x_2). \quad (3.28)$$

Démonstration. Pour la fonction de masse de X_1 , on

$$\begin{aligned} p_{X_1}(x_1) &= \sum_{x_2=0}^{\infty} \left(\frac{\rho}{1+\rho}\right)^a \frac{(a)_{x_1+x_2}(c)_{x_1}(d)_{x_2}}{x_1!x_2!(c+d)_{x_1+x_2}} \left(\frac{1}{1+\rho}\right)^{x_1+x_2} \mathbb{1}_{\mathbb{N}}(x_1) \\ &= \left(\frac{\rho}{1+\rho}\right)^a \left(\frac{1}{1+\rho}\right)^{x_1} \frac{(a)_{x_1}(c)_{x_1}}{x_1!(c+d)_{x_1}} \mathbb{1}_{\mathbb{N}}(x_1) \sum_{x_2=0}^{\infty} \frac{(a+x_1)_{x_2}(d)_{x_2}}{x_2!(c+d+x_1)_{x_2}} \left(\frac{1}{1+\rho}\right)^{x_2} \\ &= \left(\frac{\rho}{1+\rho}\right)^a \frac{(a)_{x_1}(c)_{x_1}}{x_1!(c+d)_{x_1}} \left(\frac{1}{1+\rho}\right)^{x_1} {}_2F_1\left(a+x_1, d; c+d+x_1; \frac{1}{1+\rho}\right) \mathbb{1}_{\mathbb{N}}(x_1). \end{aligned}$$

On procède de manière analogue pour $p_{X_2}(x_2)$. \square

En utilisant la transformation de Pfaff des fonctions hypergéométriques donnée à la Proposition 2.2, on remarque que les fonctions de masse marginales d'une distribution de Bailey sont en fait des lois que Laurent a définies dans [Lau12], soient les lois Poisson-gamma-inverse bêta.

Le prochain théorème présente un estimateur pour $q_{\mu, \phi}(\mathbf{y})$ basé sur la densité prédictive donnée à l'équation (1.33). Il s'agit en fait d'un cas particulier du Résultat 3.2 de l'article de Laurent [Lau12]. La fonction de masse introduite à la Proposition 3.27 servira entre autres dans les calculs de ces densités prédictives. Afin d'en faciliter la démonstration, nous utiliserons le lemme suivant.

Lemme 3.28. *Pour $c, d, \alpha, \tau > 0$ et $\kappa \in \mathbb{R}$, on a*

$$\int_0^{\infty} \frac{x^{c-1}(1+\alpha x)^{-\kappa}}{\left(1+\frac{x}{\tau}\right)^{c+d-\kappa}} dx = \alpha^{-c} B(c, d) {}_2F_1\left(c, c+d-\kappa; c+d; 1-\frac{1}{\alpha\tau}\right).$$

Démonstration. Posons $w = \alpha x$. Alors on a

$$\begin{aligned} \int_0^\infty \frac{x^{c-1}(1+\alpha x)^{-\kappa}}{\left(1+\frac{x}{\tau}\right)^{c+d-\kappa}} dx &= \int_0^\infty \frac{\left(\frac{w}{\alpha}\right)^{c-1} (1+w)^{-\kappa} \frac{1}{\alpha} dw}{\left(1+\frac{w}{\alpha\tau}\right)^{c+d-\kappa}} \\ &= \alpha^{-c} \int_0^\infty \frac{w^{c-1}(1+w)^{-\kappa}}{\left(1+\frac{w}{\alpha\tau}\right)^{c+d-\kappa}} dw \\ &= \alpha^{-c} B(c, d) {}_2F_1\left(c, c+d-\kappa; c+d; 1-\frac{1}{\alpha\tau}\right), \end{aligned}$$

où la dernière égalité découle d'une représentation sous forme d'intégrale de la fonction hypergéométrique telle que donnée à l'équation (2.5.b). \square

Théorème 3.29. [*Lau12, Result 3.2*] Soient $\mathbf{Y}|\mu, \phi \sim \mathcal{P}2(\mu\phi S', \mu T')$, $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$, $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi \sim \mathcal{B}_2\left(c, d, \frac{T+b}{S}\right)$, où $a, c > 0$, $b > -T$, $d > -a$ et où S, T, S' et T' sont des constantes positives. Alors pour $\tau = \frac{S'}{S}$ et $t = \frac{T'}{T+b}$, la fonction de masse prédictive telle que définie à l'équation (1.33) est

$$\begin{aligned} (i) \quad \hat{q}_\pi(\mathbf{y}; \mathbf{x}) &= \left(\frac{1}{t+1}\right)^{a+x_1+x_2} \frac{(a+x_1+x_2)_{y_1+y_2} (c+x_1)_{y_1} (a+d+x_2)_{y_2}}{y_1! y_2! (a+c+d+x_1+x_2)_{y_1+y_2}} \\ &\quad \times \left(\frac{\tau}{t+1}\right)^{y_1} \left(\frac{t}{t+1}\right)^{y_2} {}_2F_1\left(c+x_1+y_1, a+x_1+x_2+y_1+y_2; \right. \\ &\quad \left. a+c+d+x_1+x_2+y_1+y_2; 1-\frac{1+\tau}{1+t}\right), \quad (3.29) \end{aligned}$$

et les fonctions de masse prédictives marginales correspondantes sont

$$\begin{aligned} (ii) \quad \hat{q}_\pi(y_1; \mathbf{x}) &= \left(\frac{1}{1+\tau}\right)^{a+x_1+x_2} \frac{(a+x_1+x_2)_{y_1} (c+x_1)_{y_1}}{y_1! (a+c+d+x_1+x_2)_{y_1}} \left(\frac{\tau}{1+\tau}\right)^{y_1} \\ &\quad \times {}_2F_1\left(a+x_1+x_2+y_1, a+d+x_2; a+c+d+x_1+x_2+y_1; \frac{\tau}{1+\tau}\right), \quad (3.30) \end{aligned}$$

et

$$(iii) \quad \hat{q}_\pi(y_2; \mathbf{x}) = \left(\frac{1}{1+t} \right)^{a+x_1+x_2} \frac{(a+x_1+x_2)_{y_2} (a+d+x_2)_{y_2}}{y_2! (a+c+d+x_1+x_2)_{y_2}} \left(\frac{t}{1+t} \right)^{y_2} \\ \times {}_2F_1 \left(a+x_1+x_2+y_2, c+x_1; a+c+d+x_1+x_2+y_2; \frac{t}{1+t} \right). \quad (3.31)$$

Démonstration. Nous démontrerons (i) par calcul direct. Une fois la partie (i) démontrée, nous utiliserons cette équation pour démontrer (ii) et (iii).

(i) Posons $\rho = \frac{T+b}{S}$. On a

$$\hat{q}_\pi(\mathbf{y}; \mathbf{x}) = \int_0^\infty \int_0^\infty p(\mathbf{y}|\mu, \phi) \pi(\mu|\phi, \mathbf{x}) \pi(\phi|\mathbf{x}) d\mu d\phi \\ = \int_0^\infty \int_0^\infty \frac{e^{-\mu\phi S'} (\mu\phi S')^{y_1}}{y_1!} \frac{e^{-\mu T'} (\mu T')^{y_2}}{y_2!} \\ \times \frac{(T+b+\phi S)^{a+x_1+x_2} \mu^{a+x_1+x_2-1} e^{-\mu(T+b+\phi S)}}{\Gamma(a+x_1+x_2)} \\ \times \frac{\rho^{-(c+x_1)}}{B(c+x_1, d+a+x_2)} \frac{\phi^{c+x_1-1}}{\left(1+\frac{\phi}{\rho}\right)^{x_1+x_2+a+c+d}} d\mu d\phi.$$

En posant $K_1(\mathbf{y}) = \frac{S^{y_1} T^{y_2}}{y_1! y_2!} \frac{\rho^{-(c+x_1)}}{\Gamma(a+x_1+x_2) B(c+x_1, d+a+x_2)}$, on obtient

$$\hat{q}_\pi(\mathbf{y}; \mathbf{x}) = K_1(\mathbf{y}) \int_0^\infty (T+b+\phi S)^{a+x_1+x_2} \frac{\phi^{c+x_1+y_1-1}}{\left(1+\frac{\phi}{\rho}\right)^{x_1+x_2+a+c+d}} \\ \times \int_0^\infty e^{-\mu(T+T'+b+(S+S')\phi)} \mu^{a+x_1+x_2+y_1+y_2-1} d\mu d\phi \\ = K_1(\mathbf{y}) \int_0^\infty \frac{\phi^{c+x_1+y_1-1}}{(T+b+\phi S)^{-(a+x_1+x_2)} \left(1+\frac{\phi}{\rho}\right)^{x_1+x_2+a+c+d}} \\ \times \frac{\Gamma(a+x_1+x_2+y_1+y_2)}{(T+T'+b+(S+S')\phi)^{a+x_1+x_2+y_1+y_2}} d\phi.$$

En posant $K_2(\mathbf{y}) = K_1(\mathbf{y}) \frac{\Gamma(a+x_1+x_2+y_1+y_2)(T+b)^{a+x_1+x_2}}{(T+T'+b)^{a+x_1+x_2+y_1+y_2}}$ et $\alpha = \frac{S+S'}{T+T'+b}$, on obtient

$$\hat{q}_\pi(\mathbf{y}; \mathbf{x}) = K_2(\mathbf{y}) \int_0^\infty \frac{\phi^{c+x_1+y_1-1} (1 + \alpha\phi)^{-(a+x_1+x_2+y_1+y_2)}}{\left(1 + \frac{\phi}{\rho}\right)^{c+d}} d\phi,$$

et, en vertu du Lemme 3.28, on a

$$\begin{aligned} \hat{q}_\pi(\mathbf{y}; \mathbf{x}) &= K_2(\mathbf{y}) \alpha^{-(c+x_1+y_1)} B(c+x_1+y_1, a+d+x_2+y_2) \\ &\quad \times {}_2F_1\left(c+x_1+y_1, c+d; a+c+d+x_1+x_2+y_1+y_2; 1 - \frac{1}{\alpha\rho}\right) \\ &= K_2(\mathbf{y}) \alpha^{-(c+x_1+y_1)} B(c+x_1+y_1, a+d+x_2+y_2) (\alpha\rho)^{c+x_1+y_1} \\ &\quad \times {}_2F_1(c+x_1+y_1, a+x_1+x_2+y_1+y_2; \\ &\quad \quad \quad a+c+d+x_1+x_2+y_1+y_2; 1 - \alpha\rho) \\ &= K_2(\mathbf{y}) B(c+x_1+y_1, a+d+x_2+y_2) \rho^{c+x_1+y_1} {}_2F_1(c+x_1+y_1, \\ &\quad \quad \quad a+x_1+x_2+y_1+y_2; a+c+d+x_1+x_2+y_1+y_2; 1 - \alpha\rho). \end{aligned}$$

En remplaçant $K_2(\mathbf{y})$ par sa valeur, on obtient l'équation (3.29).

(ii) Pour ce qui est de la fonction de masse prédictive marginale de Y_1 , on a

$$\begin{aligned} \hat{q}_\pi(y_1; \mathbf{x}) &= \sum_{y_2=0}^{\infty} \hat{q}_\pi(y_1, y_2; \mathbf{x}) \\ &= \sum_{y_2=0}^{\infty} \frac{(a+x_1+x_2)_{y_1+y_2} (c+x_1)_{y_1} (a+d+x_2)_{y_2}}{y_1! y_2! (a+c+d+x_1+x_2)_{y_1+y_2}} \frac{\tau^{y_1} t^{y_2}}{(t+1)^{a+x_1+x_2+y_1+y_2}} \\ &\quad \times {}_2F_1(c+x_1+y_1, a+x_1+x_2+y_1+y_2; \\ &\quad \quad \quad a+c+d+x_1+x_2+y_1+y_2; 1 - \frac{1+\tau}{1+t}) \end{aligned}$$

$$\begin{aligned}
&= \frac{(a+x_1+x_2)_{y_1}(c+x_1)_{y_1}}{y_1!(a+c+d+x_1+x_2)_{y_1}} \frac{\tau^{y_1}}{(t+1)^{a+x_1+x_2+y_1}} \\
&\quad \times \sum_{y_2=0}^{\infty} \frac{(a+x_1+x_2+y_1)_{y_2}(a+d+x_2)_{y_2}}{y_2!(a+c+d+x_1+x_2+y_1)_{y_2}} \left(\frac{t}{t+1}\right)^{y_2} \\
&\quad \times \left(\frac{1+\tau}{1+t}\right)^{-(a+x_1+x_2+y_1+y_2)} {}_2F_1\left(a+d+x_2+y_2, a+x_1+x_2+y_1+y_2; \right. \\
&\quad \left. a+c+d+x_1+x_2+y_1+y_2; 1-\frac{1+t}{1+\tau}\right) \\
&= \frac{(a+x_1+x_2)_{y_1}(c+x_1)_{y_1}}{y_1!(a+c+d+x_1+x_2)_{y_1}} \frac{\tau^{y_1}}{(1+\tau)^{a+x_1+x_2+y_1}} \\
&\quad \times \sum_{y_2=0}^{\infty} \frac{(a+x_1+x_2+y_1)_{y_2}(a+d+x_2)_{y_2}}{y_2!(a+c+d+x_1+x_2+y_1)_{y_2}} \left(\frac{t}{1+\tau}\right)^{y_2} \\
&\quad \times \sum_{n=0}^{\infty} \frac{(a+d+x_2+y_2)_n(a+x_1+x_2+y_1+y_2)_n}{n!(a+c+d+x_1+x_2+y_1+y_2)_n} \left(\frac{\tau-t}{1+\tau}\right)^n \\
&= \frac{(a+x_1+x_2)_{y_1}(c+x_1)_{y_1}}{y_1!(a+c+d+x_1+x_2)_{y_1}} \frac{\tau^{y_1}}{(1+\tau)^{a+x_1+x_2+y_1}} \\
&\quad \times \sum_{y_2=0}^{\infty} \sum_{n=0}^{\infty} \frac{(a+x_1+x_2+y_1)_{y_2+n}(a+d+x_2)_{y_2+n}}{(a+c+d+x_1+x_2+y_1)_{y_2+n}} \frac{\left(\frac{t}{1+\tau}\right)^{y_2} \left(\frac{\tau-t}{1+\tau}\right)^n}{y_2!n!},
\end{aligned}$$

et, en vertu du Lemme 2.3, on obtient

$$\begin{aligned}
\hat{q}_\pi(y_1; \mathbf{x}) &= \left(\frac{1}{1+\tau}\right)^{a+x_1+x_2} \frac{(a+x_1+x_2)_{y_1}(c+x_1)_{y_1}}{y_1!(a+c+d+x_1+x_2)_{y_1}} \left(\frac{\tau}{1+\tau}\right)^{y_1} \\
&\quad \times {}_2F_1\left(a+x_1+x_2+y_1, a+d+x_2; a+c+d+x_1+x_2+y_1; \frac{\tau}{1+\tau}\right).
\end{aligned}$$

(iii) Pour ce qui est de la fonction de masse prédictive marginale de Y_2 , on a

$$\hat{q}_\pi(y_2; \mathbf{x}) = \sum_{y_1=0}^{\infty} \hat{q}_\pi(y_1, y_2; \mathbf{x})$$

$$\begin{aligned}
&= \sum_{y_1=0}^{\infty} \frac{(a+x_1+x_2)_{y_1+y_2} (c+x_1)_{y_1} (a+d+x_2)_{y_2}}{y_1! y_2! (a+c+d+x_1+x_2)_{y_1+y_2}} \frac{\tau^{y_1} t^{y_2}}{(t+1)^{a+x_1+x_2+y_1+y_2}} \\
&\quad \times {}_2F_1(c+x_1+y_1, a+x_1+x_2+y_1+y_2; \\
&\quad \quad \quad a+c+d+x_1+x_2+y_1+y_2; 1 - \frac{1+\tau}{1+t}) \\
&= \frac{(a+x_1+x_2)_{y_2} (a+d+x_2)_{y_2}}{y_2! (a+c+d+x_1+x_2)_{y_2}} \frac{t^{y_2}}{(1+t)^{a+x_1+x_2+y_2}} \\
&\quad \times \sum_{y_1=0}^{\infty} \frac{(a+x_1+x_2+y_2)_{y_1} (c+x_1)_{y_1}}{y_1! (a+c+d+x_1+x_2+y_2)_{y_1}} \left(\frac{\tau}{1+t}\right)^{y_1} \\
&\quad \quad \times \sum_{n=0}^{\infty} \frac{(c+x_1+y_1)_n (a+x_1+x_2+y_1+y_2)_n}{n! (a+c+d+x_1+x_2+y_1+y_2)_n} \left(\frac{t-\tau}{1+t}\right)^n \\
&= \frac{(a+x_1+x_2)_{y_2} (a+d+x_2)_{y_2}}{y_2! (a+c+d+x_1+x_2)_{y_2}} \frac{t^{y_2}}{(1+t)^{a+x_1+x_2+y_2}} \\
&\quad \times \sum_{y_1=0}^{\infty} \sum_{n=0}^{\infty} \frac{(a+x_1+x_2+y_2)_{y_1+n} (c+x_1)_{y_1+n}}{(a+c+d+x_1+x_2+y_2)_{y_1+n}} \frac{\left(\frac{\tau}{1+t}\right)^{y_1} \left(\frac{t-\tau}{1+t}\right)^n}{y_1! n!},
\end{aligned}$$

et, en vertu du Lemme 2.3, on obtient

$$\begin{aligned}
\hat{q}_\pi(y_2; \mathbf{x}) &= \left(\frac{1}{1+t}\right)^{a+x_1+x_2} \frac{(a+x_1+x_2)_{y_2} (a+d+x_2)_{y_2}}{y_2! (a+c+d+x_1+x_2)_{y_2}} \left(\frac{t}{1+t}\right)^{y_2} \\
&\quad \times {}_2F_1\left(a+x_1+x_2+y_2, c+x_1; a+c+d+x_1+x_2+y_2; \frac{t}{1+t}\right).
\end{aligned}$$

□

De cette démonstration, on voit que si $\rho = \frac{1}{\alpha}$, c'est-à-dire $\frac{S'}{S} = \frac{T'}{T+b}$, alors la fonction de masse $\hat{q}_\pi(\mathbf{y}; \mathbf{x})$ se réduit à une loi bivariée définie plus tôt telle que présentée ici.

Corollaire 3.30. [Lau12] Soient $\mathbf{Y}|\mu, \phi \sim \mathcal{P}2(\mu\phi S', \mu T')$, $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$, $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi \sim \mathcal{B}_2\left(c, d, \frac{T+b}{S}\right)$, où $a, c > 0$, $b > -T$, $d > -a$ et où S, T, S' et T' sont des constantes positives telles que $\frac{S'}{S} = \frac{T'}{T+b}$. Alors pour $\delta = \frac{S'}{S}$, la fonction de masse prédictive donnée en (3.29) est

$$\hat{q}_\pi(\mathbf{y}; \mathbf{x}) = \left(\frac{1}{1+\delta}\right)^{a+x_1+x_2} \frac{(a+x_1+x_2)_{y_1+y_2} (c+x_1)_{y_1} (a+d+x_2)_{y_2}}{y_1! y_2! (a+c+d+x_1+x_2)_{y_1+y_2}} \left(\frac{\delta}{\delta+1}\right)^{y_1+y_2},$$

ce qui correspond à une loi *Bailey* $(a + x_1 + x_2, c + x_1, d + a + x_2, \frac{1}{\delta})$.

Remarques 3.31. 1) Dans [Lau12], Laurent a exprimé la fonction de masse prédictive conjointe du Théorème 3.29 à l'aide de l'opération \star sur la loi de *Bailey*, notée

$$\mathbf{y}|\mathbf{x} \sim \left(\frac{S'}{S}, \frac{T'}{T+b} \right) \star \text{Bailey}(a + x_1 + x_2, c + x_1, d + a + x_2, 1).$$

2) Les fonctions de masse prédictives marginales de Y_1 et Y_2 du Théorème 3.29 sont de mêmes lois que les marginales d'une loi de *Bailey*, et donc, elles peuvent s'écrire sous la forme d'un mélange de loi de *Poisson*, comme vu précédemment.

Exemple 3.32. Supposons que l'on cherche à estimer la fonction de masse de $\mathbf{Y}|\mu, \phi \sim \mathcal{P}2(\mu\phi, \mu)$ en sachant que l'on a observé \mathbf{x} , où $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi, \mu)$. Du Théorème 3.29, en posant, par exemple, $a = 1, b = 0, c = d = 1/2$ et $S = T = S' = T' = 1$, on trouve

$$\hat{q}_\pi(\mathbf{y}; \mathbf{x}) = \left(\frac{1}{2} \right)^{x_1+x_2+1} \frac{(x_1 + x_2 + 1)_{y_1+y_2} (x_1 + 1/2)_{y_1} (x_2 + 3/2)_{y_2}}{y_1! y_2! (x_1 + x_2 + 2)_{y_1+y_2}} \left(\frac{1}{2} \right)^{y_1+y_2} \mathbf{1}_{\mathbb{N}^2}(y_1, y_2),$$

c'est-à-dire une loi *Bailey* $(x_1+x_2+1, x_1+1/2, x_2+3/2, 1)$, et pour les densités marginales prédictives, on trouve

$$\begin{aligned} \hat{q}_\pi(y_1; \mathbf{x}) &= \left(\frac{1}{2} \right)^{x_1+x_2+1} \frac{(x_1 + x_2 + 1)_{y_1} (x_1 + 1/2)_{y_1}}{y_1! (x_1 + x_2 + 2)_{y_1}} \left(\frac{1}{2} \right)^{y_1} \\ &\quad \times {}_2F_1 \left(x_1 + x_2 + 1 + y_1, x_2 + \frac{3}{2}; x_1 + x_2 + 2 + y_1; \frac{1}{2} \right), \end{aligned}$$

et

$$\begin{aligned} \hat{q}_\pi(y_2; \mathbf{x}) &= \left(\frac{1}{2} \right)^{x_1+x_2+1} \frac{(x_1 + x_2 + 1)_{y_2} (x_1 + 3/2)_{y_2}}{y_2! (x_1 + x_2 + 2)_{y_2}} \left(\frac{1}{2} \right)^{y_2} \\ &\quad \times {}_2F_1 \left(x_1 + x_2 + 1 + y_2, x_2 + \frac{1}{2}; x_1 + x_2 + 2 + y_2; \frac{1}{2} \right). \end{aligned}$$

Par exemple, pour $x_1 = 17$ et $x_2 = 31$, les fonctions de masse marginales prédictives de Y_1 et Y_2 sont représentées respectivement à la Figure 3.8 et à la Figure 3.9. Elles pourraient être entre autres comparées aux fonctions de masse trouvées au Théorème 3.18 pour voir l'impact qu'a le fait d'avoir des données observées en lien avec les paramètres inconnus.

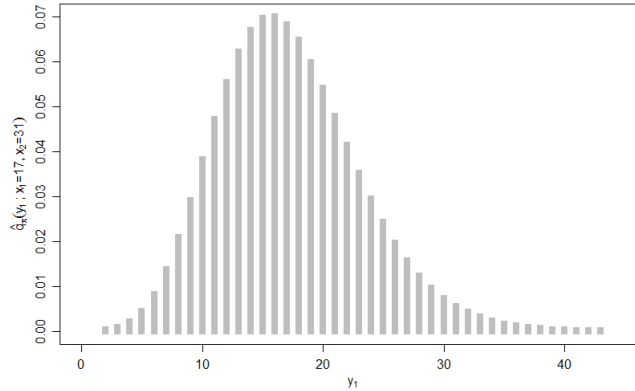


Figure 3.8 – Fonction de masse prédictive de Y_1 pour $x_1 = 17$ et $x_2 = 31$

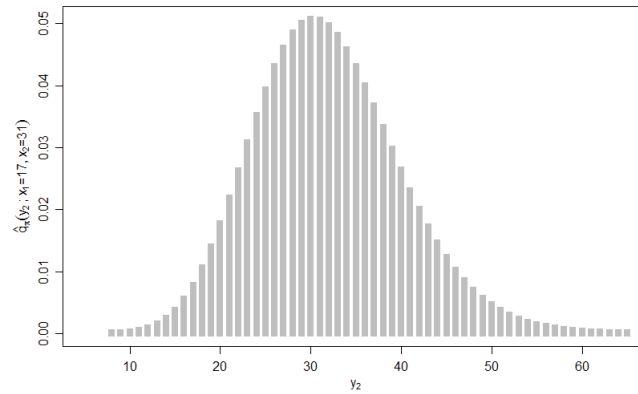


Figure 3.9 – Fonction de masse prédictive de Y_2 pour $x_1 = 17$ et $x_2 = 31$

3.4.2 Espérances, variances et covariance

Comme abordé au chapitre 1, il peut être intéressant d'étudier la « calibration » sous \hat{q}_π de certaines quantités telles que l'espérance et la variance d'une variable aléatoire $Y|\theta \sim q_\theta$ à partir d'une observation indépendante $X|\theta \sim p_\theta$. Voyons voir ce qu'il en est dans le modèle bivarié ci-présent étudié. On supposera à nouveau que $\mu|\phi \sim \mathcal{G}(a, b)$ et que $\phi \sim \mathcal{B}_2\left(c, d, \frac{T+b}{S}\right)$. Pour faciliter l'obtention de ces éventuelles expressions, les deux lemmes qui suivent seront utiles.

Lemme 3.33. Soient $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi \sim \mathcal{B}_2\left(c, d, \frac{T+b}{S}\right)$, où $a, c > 0$, $b > -T$, $d > -a$

et $S, T > 0$. Alors

$$E[\mu^m \phi^n] = \frac{(a)_m}{b^m} \cdot \frac{\Gamma(c+n)\Gamma(d-n)}{\Gamma(c)\Gamma(d)} \left(\frac{T+b}{S}\right)^n, \quad m > -a, \quad n < d. \quad (3.32)$$

Démonstration. Cela suit des moments d'une loi gamma et du Corollaire 2.23. \square

Lemme 3.34. Soient $\mu|\phi, \mathbf{x} \sim \mathcal{G}(a + x_1 + x_2, T + b + \phi S)$ et $\phi|\mathbf{x} \sim \mathcal{B}_2\left(c + x_1, d + a + x_2, \frac{T+b}{S}\right)$, où $\mathbf{x} \in \mathbb{N}^2$, $a, c > 0$, $b > -T$, $d > -a$, et $S, T > 0$. Alors

$$E[\mu^m \phi^n | \mathbf{x}] = \frac{1}{S^n (T+b)^{m-n}} \frac{(a+x_1+x_2)_m (c+x_1)_n (d+a+x_2)_{m-n}}{(a+c+d+x_1+x_2)_m}, \quad m, n, m-n \geq 0. \quad (3.33)$$

Démonstration. On a $E[\mu^m \phi^n | \mathbf{x}] = E_{\phi|\mathbf{x}}[\phi^n E_{\mu|\phi, \mathbf{x}}[\mu^m]]$.

Comme $E_{\mu|\phi, \mathbf{x}}[\mu^m] = \frac{(a+x_1+x_2)_m}{(T+b)^m (1+\frac{S}{T+b}\phi)^m}$, alors

$$\begin{aligned} E[\mu^m \phi^n | \mathbf{x}] &= \frac{(a+x_1+x_2)_m}{(T+b)^m} E_{\phi|\mathbf{x}} \left[\frac{\phi^n}{\left(1 + \frac{S}{T+b}\phi\right)^m} \right] \\ &= \frac{(a+x_1+x_2)_m}{(T+b)^m} \frac{\left(\frac{T+b}{S}\right)^{-(c+x_1)}}{B(c+x_1, d+a+x_2)} \int_0^\infty \frac{\phi^{c+x_1+n-1}}{\left(1 + \frac{S}{T+b}\phi\right)^{c+x_1+d+a+x_2+m}} d\phi \\ &= \frac{(a+x_1+x_2)_m}{(T+b)^m} \frac{\left(\frac{T+b}{S}\right)^{-(c+x_1)}}{B(c+x_1, d+a+x_2)} \frac{B(c+x_1+n, d+a+x_2+m-n)}{\left(\frac{T+b}{S}\right)^{-(c+x_1+n)}} \\ &= \frac{1}{S^n (T+b)^{m-n}} \frac{(a+x_1+x_2)_m (c+x_1)_n (d+a+x_2)_{m-n}}{(a+c+d+x_1+x_2)_m}. \quad \square \end{aligned}$$

Ces moments mixtes *a posteriori* permettent entre autres de vérifier un résultat de Laurent [Lau12] exprimant la covariance entre Y_1 et Y_2 en tenant compte des observations x_1 et x_2 , dont la démonstration était laissée au lecteur. En effet, à partir du

Lemme 3.34, du Théorème 1.59 et des expressions qui en découlent telles que présentées à l'Exemple 1.60, on trouve

$$Cov_{\hat{q}_\pi}(\mathbf{Y}) = \left(\frac{S'}{S}\right) \left(\frac{T'}{T+b}\right) \frac{(a+x_1+x_2)(c+x_1)(d+a+x_2)(c+d-a)}{(a+c+d+x_1+x_2)^2(a+c+d+x_1+x_2+1)}. \quad (3.34)$$

Comme le fait remarquer Laurent dans [Lau12], cette covariance est toujours positive.

Les moments mixtes trouvées aux lemmes précédents permettent également de trouver des expressions pour les espérances et les variances des variables de $Y_1, Y_2 | \mu, \phi \sim \mathcal{P}2(\mu\phi S', \mu T')$ qui se basent sur le Théorème 1.55 et le Théorème 1.56. Quelques expressions appliquées à notre modèle à l'étude avaient été trouvées à l'Exemple 1.57. Le théorème suivant illustre l'application pour les espérances.

Théorème 3.35. *Soient $\mathbf{Y} | \mu, \phi \sim \mathcal{P}2(\mu\phi S', \mu T')$, $\mathbf{X} | \mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$, $\mu | \phi \sim \mathcal{G}(a, b)$, $\phi \sim \mathcal{B}_2\left(c, d, \frac{T+b}{S}\right)$, où $a, c > 0$, $b > -T$, $d > -a$ et $S, S', T, T' > 0$. Alors on a*

$$E[Y_1] = \frac{S'(T+b)}{S} \frac{ac}{b(d-1)}, \quad d > 1, \quad (3.35)$$

et

$$E[Y_2] = \frac{T'a}{b}, \quad b > 0. \quad (3.36)$$

À partir d'observations $(x_1, x_2) \in \mathbb{N}^2$, sous les fonctions de masse prédictives obtenues au Théorème 3.29, on a

$$E_{\hat{q}_\pi}[Y_1; \mathbf{x}] = \left(\frac{S'}{S}\right) \frac{(a+x_1+x_2)(c+x_1)}{a+c+d+x_1+x_2}, \quad (3.37)$$

et

$$E_{\hat{q}_\pi}[Y_2; \mathbf{x}] = \left(\frac{T'}{T+b}\right) \frac{(a+x_1+x_2)(d+a+x_2)}{a+c+d+x_1+x_2}. \quad (3.38)$$

Démonstration. Cela suit des théorèmes 1.55 et 1.56 et des lemmes 3.33 et 3.34. \square

Remarque 3.36. *Au Corollaire 3.30, on a vu que dans le cas où $\frac{S'}{S} = \frac{T'}{T+b}$, alors $\hat{q}_\pi(\mathbf{y}; \mathbf{x})$ correspond à la fonction de masse d'une loi Bailey $\left(a+x_1+x_2, c+x_2, d+a+x_2, \frac{1}{\delta}\right)$, où*

$\delta = \frac{S'}{S} = \frac{T'}{T+b}$. Du Corollaire 3.24, des expressions avaient été trouvées pour l'espérance des variables d'une loi de Bailey. Ce résultat concorde avec les équations (3.37) et (3.38).

Exemple 3.37. Reprenons le contexte et les données de l'Exemple 3.32. On a

$$E_{\hat{q}_\pi}[Y_1; \mathbf{x}] = \frac{(x_1 + x_2 + 1)(x_1 + 1/2)}{x_1 + x_2 + 2},$$

et

$$E_{\hat{q}_\pi}[Y_2; \mathbf{x}] = \frac{(x_1 + x_2 + 1)(x_2 + 3/2)}{x_1 + x_2 + 2}.$$

À partir de l'observation $\mathbf{x} = (17, 31)$, on obtient $E_{\hat{q}_\pi}[Y_1; \mathbf{x}] = 17,15$ et $E_{\hat{q}_\pi}[Y_2; \mathbf{x}] = 31,85$.

CHAPITRE 4

Étude du modèle bivarié avec contrainte paramétrique

On rappelle que le modèle étudié peut s'écrire ainsi :

$$\begin{cases} X_1 | \mu, \phi \sim \mathcal{P}(\mu\phi S) \\ X_2 | \mu, \phi \sim \mathcal{P}(\mu T). \end{cases}$$

Comme on a vu, dans un contexte de vaccination, cette paramétrisation confère l'avantage de pouvoir mesurer directement l'efficacité du traitement prodigué à l'une des deux populations à l'étude à travers le paramètre d'intérêt ϕ . Dans le chapitre 3, aucune contrainte n'a été mise sur ce paramètre. À partir de méthodes d'inférence bayésiennes, une loi *a priori* raisonnable pour ϕ pouvait alors être définie sur l'intervalle $[0, \infty)$. Par contre, dans plusieurs contextes, on peut vouloir imposer une contrainte d'ordre sur les paramètres $\lambda = \mu\phi$ et μ , par exemple, $\lambda \leq \mu$. Pour qu'une telle condition soit respectée, il faut alors restreindre le support de ϕ à l'intervalle $[0, 1]$. Ainsi, si l'on pense que le vaccin ne peut pas nuire aux patients, on serait alors avisé de stipuler *a priori* que $\mathbb{P}(\phi > 1) = 0$. Aussi, si l'on sait que le vaccin est efficace, la question demeure de mesurer le degré d'amélioration et une loi *a priori* pour ϕ de support $[0, 1]$ en est une conséquence

logique. Pareillement, pour la file $M|M|1$ avec queue infinie, avec un temps entre les arrivées de loi $\mathcal{Exp}(\lambda)$ et des temps de service de loi $\mathcal{Exp}(\mu)$, si l'on suppose que la distribution limite existe, il est alors nécessaire d'imposer $\lambda < \mu$, c'est-à-dire ϕ de support $[0, 1]$.

D'une part, ce chapitre a pour but de présenter un cadre bayésien pour ce problème avec contrainte d'ordre paramétrique et d'y extraire des lois *a posteriori* correspondantes de même que des estimateurs ponctuels pour ϕ . À l'aide du risque fréquentiste, nous évaluerons ces estimateurs, les comparerons avec celui issu de la méthode du maximum de vraisemblance et illustrerons l'importance d'utiliser l'information disponible sur l'ordre paramétrique lors du développement d'estimateurs. D'autre part, tout comme ce fut le cas dans le chapitre 3, nous présenterons des estimateurs de fonction de masse prédictive pour de futures observations $(Y_1, Y_2)|\mu, \phi \sim \mathcal{P}2(\mu\phi S', \mu T')$, avec S' et T' des constantes positives non nécessairement égales respectivement à S et à T .

4.1 Estimateur à vraisemblance maximale modifié

De la Proposition 1.8, on sait que $X_1|X_1 + X_2 = n \sim \mathcal{Bin}\left(n, \frac{\phi S}{\phi S + T}\right)$. Ainsi, inférer sur ϕ revient à inférer sur la proportion $p = \frac{\phi S}{\phi S + T}$ de la distribution conditionnelle binomiale par l'équation (1.5).

Au chapitre 3, on a vu que, lorsqu'aucune condition n'est imposée sur le support de ϕ , l'estimateur du maximum de vraisemblance pour p est $\hat{p} = \frac{x_1}{x_1 + x_2}$, ce qui mène à l'estimateur $\hat{\phi}(\mathbf{x}) = \frac{T x_1}{S x_2}$. Par contre, comme cet estimateur n'est pas défini pour $x_2 = 0$, on applique une correction pour ainsi obtenir l'estimateur $\hat{\phi}_{EVMM}(\mathbf{x}) = \frac{T}{S} \left(\frac{x_1 + 0,5}{x_2 + 0,5} \right)$. Ici, la contrainte imposée sur le support de ϕ revient à restreindre p sur l'intervalle $\left[0, \frac{S}{S+T}\right]$. L'estimateur du maximum de vraisemblance avec contrainte pour \hat{p} , que nous

distinguerons de celui trouvé au chapitre 3 par la notation c en indice, devient alors

$$\hat{p}_{EVMc}(\mathbf{x}) = \min\left(\frac{x_1}{x_1 + x_2}, \frac{S}{S + T}\right) = \begin{cases} \frac{x_1}{x_1 + x_2} & \text{si } x_1 \leq \frac{Sx_2}{T} \\ \frac{S}{S + T} & \text{si } x_1 > \frac{Sx_2}{T}, \end{cases} \quad (4.1)$$

ce qui conduit à l'estimateur pour ϕ

$$\hat{\phi}_{EVMc}(\mathbf{x}) = \min\left(\frac{Tx_1}{Sx_2}, 1\right).$$

De la même manière qu'au chapitre précédent, comme cet estimateur n'est pas défini pour $x_2 = 0$, on applique une correction en ajoutant 0,5 aux valeurs de x_1 et x_2 , ce qui mène à l'estimateur suivant :

$$\hat{\phi}_{EVMMc}(\mathbf{x}) = \min\left(\frac{T}{S} \left(\frac{x_1 + 0,5}{x_2 + 0,5}\right), 1\right). \quad (4.2)$$

4.2 Estimateurs bayésiens

Tout comme dans le cas sans contrainte sur le support de ϕ , on suppose que μ et ϕ sont indépendants, d'où $\pi(\mu, \phi) = \pi(\mu)\pi(\phi)$ et $\mu|\phi \sim \mathcal{G}(a, b)$.

On peut établir des ressemblances entre ce problème d'estimation paramétrique et celui abordé par Armero et Bayarri en lien avec les files d'attente dans [AB94]. En effet, on y retrouve notamment la question entourant le choix d'une loi *a priori* appropriée pour ϕ (noté ρ dans cet article), qui doit tenir compte de la contrainte $\phi < 1$ afin que le système soit en équilibre et que la distribution limite de la taille du système existe. Une discussion intéressante à ce sujet y est présentée et est motivée par leur objectif d'inférence prédictive de certaines quantités d'intérêt des files d'attente $M|M|1$ en équilibre. En effet, Armero et Bayarri y expliquent notamment pourquoi la famille conjuguée *a priori* naturelle n'est pas un choix approprié dans ce contexte, et cette analyse les mène vers une famille de lois

qu'ils définissent comme étant la famille des distributions hypergéométriques de Gauss. On remarque que cette distribution correspond en fait à la distribution bêta généralisée à quatre paramètres décrite à la section 2.2.

Dans notre cas, avec $\phi \in [0, 1]$, il est raisonnable de supposer que $\phi \sim \mathcal{B}(c, d)$, c'est-à-dire

$$\pi(\phi) = \frac{\phi^{c-1}(1-\phi)^{d-1}}{B(c, d)} \mathbf{1}_{[0,1]}(\phi), \quad c, d > 0. \quad (4.3)$$

En ce qui concerne les lois *a posteriori*, du Théorème 3.4, on a déjà montré que $\mu|\mathbf{x}, \phi \sim \mathcal{G}(a + x_1 + x_2, T + b + \phi S)$. Le théorème suivant présente les lois *a posteriori* de μ et de ϕ .

Théorème 4.1. *Soient $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \phi T)$, $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi \sim \mathcal{B}(c, d)$, où $a, c, d > 0$ et $b > -T$. Alors les lois *a posteriori* de μ et ϕ ont pour densités respectivement*

$$\pi(\mu|\mathbf{x}) \propto e^{-\mu(S+T+b)} \mu^{a+x_1+x_2-1} {}_1F_1(d; c+d+x_1; \mu S) \mathbf{1}_{[0,\infty)}(\mu) \quad (4.4)$$

et

$$\phi|\mathbf{x} \sim \mathcal{BG4} \left(c + x_1, d, \frac{S + T + b}{T + b}, a + x_1 + x_2 \right). \quad (4.5)$$

Démonstration. D'une part, de l'équation (3.10), on a

$$\pi(\mu|\mathbf{x}) \propto e^{-\mu(T+b)} \mu^{a+x_1+x_2-1} \mathbf{1}_{[0,\infty)}(\mu) \times \int_0^1 e^{-\mu\phi S} \phi^{c+x_1-1} (1-\phi)^{d-1} d\phi,$$

qui, de l'équation 2.14, se réécrit

$$\begin{aligned} \pi(\mu|\mathbf{x}) &\propto e^{-\mu(T+b)} \mu^{a+x_1+x_2-1} {}_1F_1(c+x_1; c+d+x_1; -\mu S) \mathbf{1}_{[0,\infty)}(\mu) \\ &= e^{-\mu(S+T+b)} \mu^{a+x_1+x_2-1} {}_1F_1(d; c+d+x_1; \mu S) \mathbf{1}_{[0,\infty)}(\mu). \end{aligned}$$

D'autre part, de l'équation (3.11), on a

$$\pi(\phi|x_1, x_2) \propto \frac{\phi^{c+x_1-1}(1-\phi)^{d-1}}{(T+b+\phi S)^{a+x_1+x_2}} \mathbf{1}_{[0,1]}(\phi) \propto \frac{\phi^{c+x_1-1}(1-\phi)^{d-1}}{\left(1 + \frac{S}{T+b}\phi\right)^{a+x_1+x_2}} \mathbf{1}_{[0,1]}(\phi),$$

d'où $\phi|\mathbf{x} \sim \mathcal{BG4} \left(c + x_1, d, \frac{S+T+b}{T+b}, a + x_1 + x_2 \right)$. □

Exemple 4.2. Soient $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$, $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi \sim \mathcal{B}(c, d)$ avec $a, c, d > 0$ et $b > -T$. On sait que $\phi|\mathbf{x} \sim \mathcal{BG}4\left(c + x_1, d, \frac{S+T+b}{T+b}, a + x_1 + x_2\right)$ et on peut construire un intervalle de crédibilité bayésien $[q_{\alpha/2}, q_{1-\alpha/2}]$ de niveau $1 - \alpha$, où $q_{\alpha/2}$ et $q_{1-\alpha/2}$ sont des quantiles d'ordre $\alpha/2$ et $1 - \alpha/2$ de cette loi.

À l'Exemple 3.7, avec $x_1 = 17$, $x_2 = 31$, $S = T = 1$, $a = 1$, $b = 0$ et $c = d = 1/2$, à partir de la fonction de répartition d'une loi bêta de type II, nous avons calculé l'intervalle bayésien de crédibilité 95 % pour ϕ suivant :

$$I_\pi(\phi) = [0, 2911; 0, 9452].$$

Or, si nous avons comme information supplémentaire que $\phi < 1$, alors, du modèle bayésien construit dans cette section, on sait que $\phi|\mathbf{x} \sim \mathcal{BG}4\left(c + x_1, d, \frac{S+T+b}{T+b}, a + x_1 + x_2\right)$. On peut alors calculer un intervalle bayésien de crédibilité 95 % à partir de quantiles de cette loi. Avec $a = c = d = 1$ et $b = 0$, on trouve alors

$$I_\pi(\phi) = [0, 3466; 0, 8733].$$

Les intervalles sont un peu différents, mais cette différence serait accentuée si $\frac{x_1}{x_2}$ était plus près de 1. La Figure 4.1 illustre la densité a posteriori de ϕ pour cet exemple, c'est-à-dire pour $a = c = d = S = T = 1$, $b = 0$, $x_1 = 17$ et $x_2 = 31$.

Voici des expressions pour les espérances a posteriori dans le contexte du Théorème 4.1.

Corollaire 4.3. Soient $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \phi T)$, $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi \sim \mathcal{B}(c, d)$, où $a, c, d, S, T > 0$ et $b > -T$. Alors

$$E[\mu|\mathbf{x}] = \frac{a + x_1 + x_2}{S + T + b} \frac{{}_2F_1\left(d, a + x_1 + x_2 + 1; c + d + x_1; \frac{S}{S+T+b}\right)}{{}_2F_1\left(d, a + x_1 + x_2; c + d + x_1; \frac{S}{S+T+b}\right)}, \quad (4.6)$$

et

$$E[\phi|\mathbf{x}] = \frac{c + x_1}{c + d + x_1} \frac{{}_2F_1\left(d, a + x_1 + x_2; c + d + x_1 + 1; \frac{S}{S+T+b}\right)}{{}_2F_1\left(d, a + x_1 + x_2; c + d + x_1; \frac{S}{S+T+b}\right)}. \quad (4.7)$$

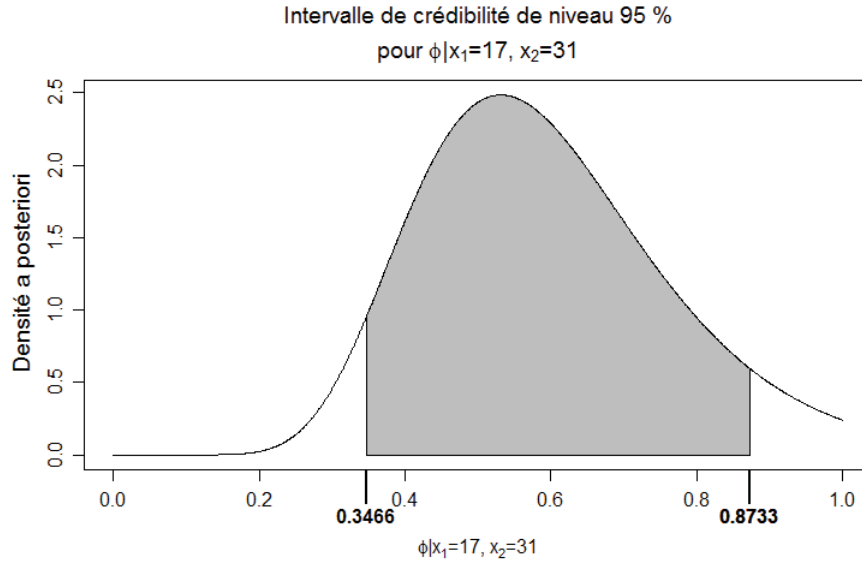


Figure 4.1 – Illustration d’un intervalle de crédibilité pour ϕ contraint à $[0, 1]$

Démonstration. Pour l’espérance *a posteriori* de μ , on a

$$\begin{aligned}
 E[\mu|x_1, x_2] &= (a + x_1 + x_2)E\left[\frac{1}{T + b + \phi S}\right] \\
 &= \frac{a + x_1 + x_2}{B(c + x_1, d)_2F_1\left(c + x_1, a + x_1 + x_2; c + d + x_1; \frac{-S}{T+b}\right)} \\
 &\quad \times \int_0^1 \frac{1}{(T + b + \phi S)} \frac{\phi^{c+x_1-1}(1 - \phi)^{d-1}}{\left(1 + \left(\frac{S}{T+b}\right)\phi\right)^{a+x_1+x_2}} d\phi.
 \end{aligned}$$

De l’identité (2.5.a), on a

$$\begin{aligned}
 &\int_0^1 \frac{1}{(T + b + \phi S)} \frac{\phi^{c+x_1-1}(1 - \phi)^{d-1}}{\left(1 + \left(\frac{S}{T+b}\right)\phi\right)^{a+x_1+x_2}} d\phi \\
 &= \frac{B(c + x_1, d)_2F_1\left(c + x_1, a + x_1 + x_2 + 1; c + d + x_1; \frac{-S}{T+b}\right)}{T + b}.
 \end{aligned}$$

Puis, en utilisant la transformation de Pfaff donnée à l’équation (2.6.b), on a

$$E[\mu|x_1, x_2] = \frac{a + x_1 + x_2}{T + b} \frac{\left(\frac{T+b}{S+T+b}\right)^{a+x_1+x_2+1} {}_2F_1\left(d, a + x_1 + x_2 + 1; c + d + x_1; \frac{S}{S+T+b}\right)}{\left(\frac{T+b}{S+T+b}\right)^{a+x_1+x_2} {}_2F_1\left(d, a + x_1 + x_2; c + d + x_1; \frac{S}{S+T+b}\right)}$$

$$= \frac{a + x_1 + x_2}{S + T + b} \frac{{}_2F_1\left(d, a + x_1 + x_2 + 1; c + d + x_1; \frac{S}{S+T+b}\right)}{{}_2F_1\left(d, a + x_1 + x_2; c + d + x_1; \frac{S}{S+T+b}\right)}.$$

Pour l'espérance *a posteriori* de ϕ , de l'équation (2.21), on a

$$\begin{aligned} E[\phi|x_1, x_2] &= \frac{c + x_1}{c + d + x_1} \frac{{}_2F_1\left(c + x_1 + 1, a + x_1 + x_2; c + d + x_1 + 1; \frac{-S}{T+b}\right)}{{}_2F_1\left(c + x_1, a + x_1 + x_2; c + d + x_1; \frac{-S}{T+b}\right)} \\ &= \frac{c + x_1}{c + d + x_1} \frac{{}_2F_1\left(d, a + x_1 + x_2; c + d + x_1 + 1; \frac{S}{S+T+b}\right)}{{}_2F_1\left(d, a + x_1 + x_2; c + d + x_1; \frac{S}{S+T+b}\right)}, \end{aligned}$$

où la dernière égalité s'obtient également de la transformation de Pfaff donnée en (2.6.b). \square

Exemple 4.4. À partir des mêmes hypothèses que celles de l'Exemple 4.2, on trouve $E[\mu|x_1 = 17, x_2 = 31] \approx 31,2409$ et $E[\phi|x_1 = 17, x_2 = 31] \approx 0,5839$.

Le théorème suivant se rapport à un comportement plausible des lois *a posteriori* de $\phi|\mathbf{x}$.

Théorème 4.5. Soit $\phi|\mathbf{x} \sim \mathcal{BG4}\left(c + x_1, d, \frac{S+T+b}{T+b}, a + x_1 + x_2\right)$ avec $a, c, d > 0$, $b > -T$ et $\mathbf{x} \in \mathbb{N}^2$. Alors la famille de distributions a un rapport de vraisemblance monotone

i) croissant en ϕ , pour x_2 fixe et x_1 le paramètre;

ii) décroissant en ϕ , pour x_1 fixe et x_2 le paramètre.

Ainsi, $E[\phi|\mathbf{x}]$ est une fonction croissante en x_1 et décroissante en x_2 .

Démonstration. Soient $x_1, x'_1 \in \mathbb{N}$ avec $x'_1 > x_1$. Alors pour tout $x_2 \in \mathbb{N}$, on a

$$f_1(\phi) = \frac{\pi(\phi|x'_1, x_2)}{\pi(\phi|x_1, x_2)} = K_1(x_1, x'_1, x_2) \left(\frac{\phi}{T + b + S\phi} \right)^{x'_1 - x_1},$$

où $K_1(x_1, x'_1, x_2)$ est une fonction positive. Comme $f_1(\phi)$ est croissante en ϕ , la famille de distributions de $\phi|x_1, x_2$, pour x_2 fixé, appartient à une famille de densités RVMC en

x_1 (voir Définition 3.11). En vertu de la Proposition 3.14, en posant $h(\phi) = \phi$, qui est une fonction croissante, on a que $E[\phi|\mathbf{x}]$ est une fonction croissante en x_1 .

De manière similaire, pour $x_2, x'_2 \in \mathbb{N}$ avec $x'_2 > x_2$, et $x_1 \in \mathbb{N}$ fixe, on a

$$f_2(\phi) = \frac{\pi(\phi|x_1, x'_2)}{\pi(\phi|x_1, x_2)} = K_2(x_1, x_2, x'_2)(T + b + S\phi)^{x_2 - x'_2},$$

où $K_2(x_1, x_2, x'_2)$ est une fonction positive, qui décroît en ϕ . Le résultat s'en suit. \square

L'exemple ci-dessous illustrera le Théorème 4.5. De plus, on y comparera les estimateurs obtenus avec l'estimateur $\hat{\phi}_{EVMMc}$.

Exemple 4.6. *Sous les mêmes hypothèses qu'à l'Exemple 4.2, avec $a = c = d = S = T = 1$ et $b = 0$, alors*

$$E[\phi|\mathbf{x}] = \left(\frac{x_1 + 1}{x_1 + 2} \right) \frac{{}_2F_1\left(1, x_1 + x_2 + 1; x_1 + 3; \frac{1}{2}\right)}{{}_2F_1\left(1, x_1 + x_2 + 1; x_1 + 2; \frac{1}{2}\right)}.$$

Si on fixe $x_2 = 31$, alors $E[\phi|x_1, x_2 = 31]$ est croissante en x_1 . Pour cette valeur et d'autres valeurs de x_2 , cette fonction est illustrée à la Figure 4.2, de même qu'une comparaison avec l'estimateur du maximum de vraisemblance modifié. On remarque tout d'abord que ces deux estimateurs semblent plutôt similaires, sauf lorsque x_1 est près de la valeur de x_2 fixée. En effet, lorsque x_1 est suffisamment plus petit que x_2 , les deux estimateurs sont à peu près équivalents ($\hat{\phi}_{EVMMc}(x_1, x_2)$ est légèrement inférieur à l'autre), mais à mesure que x_1 se rapproche de x_2 , l'estimateur $\hat{\phi}_{EVMMc}(x_1, x_2)$ devient supérieur à l'estimateur $E[\phi|x_1, x_2]$ et le demeure. Enfin, ce dernier tend vers 1 quand x_1 tend vers l'infini.

Si on fixe plutôt $x_1 = 17$, alors $E[\phi|x_1 = 17, x_2]$ est décroissante en x_2 , comme illustré à la Figure 4.3. Cet estimateur a également été tracé pour d'autres valeurs de x_2 fixées et comparé à l'estimateur du maximum de vraisemblance modifié. Tout comme à la Figure 4.2, on remarque que ces deux estimateurs sont similaires sauf lorsque $x_2 \approx x_1$. En

effet, pour x_2 suffisamment plus petit que x_1 , l'estimateur $\hat{\phi}_{EVMMc}(x_1, x_2)$ est supérieur à l'estimateur $E[\phi|x_1, x_2]$, avec un écart plus prononcé lorsque x_2 est près de x_1 . Puis, lorsque x_2 devient suffisamment plus grand que x_1 , les deux estimateurs sont plutôt semblables ($\hat{\phi}_{EVMMc}(x_1, x_2)$ devient légèrement inférieur à $E[\phi|x_1, x_2]$), et ils tendent tous deux vers 0 lorsque x_2 tend vers l'infini.

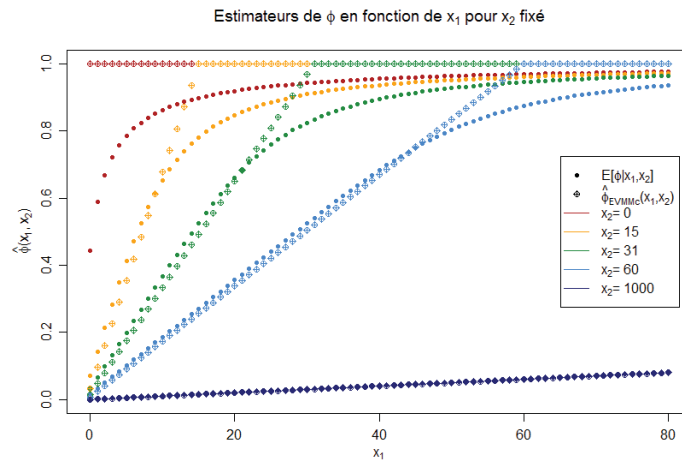


Figure 4.2 – Graphique de $E[\phi|x_1, x_2 = 31]$ pour $a = c = d = 1$ et $b = 0$

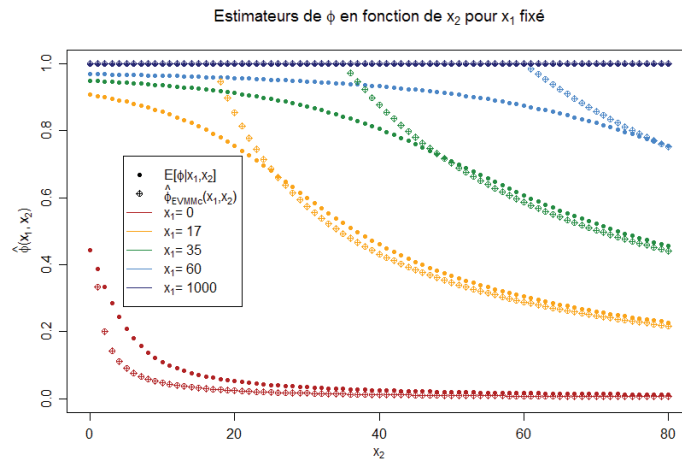


Figure 4.3 – Graphique de $E[\phi|x_1 = 17, x_2]$ pour $a = c = d = 1$ et $b = 0$

4.3 Évaluation et comparaisons du risque des estimateurs

Dans ce chapitre, nous avons obtenu l'estimateur à vraisemblance maximale modifié avec contrainte et un estimateur bayésien issu de l'espérance *a posteriori*, c'est-à-dire :

$$\hat{\phi}_{EVMMc}(\mathbf{x}) = \min \left(\frac{T}{S} \left(\frac{x_1 + 0, 5}{x_2 + 0, 5} \right), 1 \right);$$

$$\hat{\phi}_{Bc}(\mathbf{x}) = \frac{c + x_1}{c + d + x_1} \frac{{}_2F_1 \left(d, a + x_1 + x_2; c + d + x_1 + 1; \frac{S}{S+T+b} \right)}{{}_2F_1 \left(d, a + x_1 + x_2; c + d + x_1; \frac{S}{S+T+b} \right)}, \quad a, c, d > 0, b > -T.$$

Alors que les estimateurs obtenus au chapitre 3 faisaient tous partie de la famille d'estimateurs $\mathcal{F}(\mathbf{x}) = \left\{ \frac{\alpha_1 x_1 + \beta_1}{\alpha_2 x_2 + \beta_2} \mid \alpha_1, \alpha_2, \beta_2 > 0, \beta_1 \geq 0 \right\}$, on remarque que $\hat{\phi}_{Bc}$ est d'une forme plus complexe, ce qui en rend l'analyse plus difficile.

Pour l'étude des deux estimateurs $\hat{\phi}_{EVMMc}$ et $\hat{\phi}_{Bc}$, on comparera leurs risques. On analysera également les gains réalisés en tenant compte de la contrainte sur le support de ϕ . Tout comme ce fut le cas au chapitre 3, on supposera que $S = T = 1$ et on utilisera la fonction de perte quadratique pour évaluer le risque en fonction du paramètre d'intérêt ϕ pour différentes valeurs de μ .

4.3.1 Risque de l'estimateur à vraisemblance maximale modifié avec contrainte paramétrique

On s'intéresse ici au risque de l'estimateur $\hat{\phi}_{EVMMc}$. La Figure 4.4 représente graphiquement le risque de cet estimateur pour quelques valeurs de μ .

On remarque tout d'abord que pour $\phi \approx 0$, le risque est décroissant en μ , tout comme c'était le cas lorsqu'il n'y avait pas de contrainte sur le support de ϕ . En effet, pour

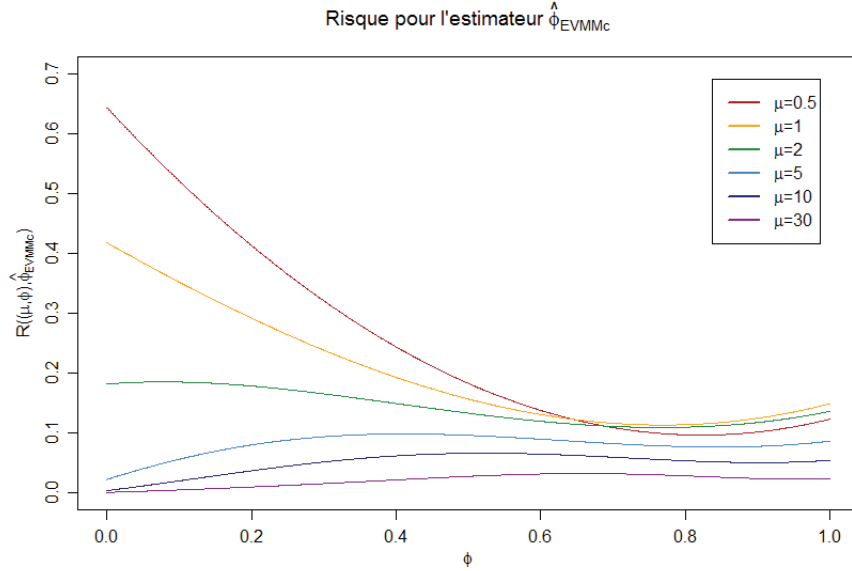


Figure 4.4 – Risque de l'estimateur $\hat{\phi}_{EVMMc}(\mathbf{x}) = \min\left(\frac{x_1+0,5}{x_2+0,5}, 1\right)$

une telle valeur de ϕ , on a $x_1 \approx 0$, et donc $\hat{\phi}_{EVMMc}(0, x_2) = \min\left(\frac{0,5}{x_2+0,5}, 1\right) = \frac{0,5}{x_2+0,5} = \hat{\phi}_{EVMM}(0, x_2)$. Par le même raisonnement que celui présenté à la section 3.3.1, on peut expliquer pourquoi le risque décroît en μ pour $\phi \approx 0$.

Dans un autre ordre d'idées, on peut également se demander quelle est l'importance de considérer l'estimateur $\hat{\phi}_{EVMMc}$ plutôt que l'estimateur $\hat{\phi}_{EVMM}$ lorsque l'on a l'information que ϕ est restreint à l'intervalle $[0, 1]$. Graphiquement, on peut voir à la Figure 4.5 que le quotient $\frac{R((\mu, \phi), \hat{\phi}_{EVMMc})}{R((\mu, \phi), \hat{\phi}_{EVMM})}$ est toujours inférieur à 1 pour toutes les valeurs de μ étudiées. De plus, on voit via la courbe représentant $\mu = 30$ que pour une valeur de μ suffisamment grande, le risque de ces estimateurs semblent se comporter de manière similaire pour plusieurs valeurs de ϕ , puisque le quotient est près de la droite $y = 1$. Or, il semble que les gains réalisés sur le risque, en tenant compte de la contrainte imposée sur ϕ , soient d'autant plus considérables lorsque μ est de petite valeur. Ainsi, dans le cas où $0 \leq \phi \leq 1$, l'estimateur $\hat{\phi}_{EVMM}$ est inadmissible et il est entre autres dominé par l'estimateur $\hat{\phi}_{EVMMc}$.

En effet, on peut facilement montrer que $R((\mu, \phi), \hat{\phi}_{EVMMc}) < R((\mu, \phi), \hat{\phi}_{EVMM})$, car si $\hat{\phi}_{EVMM}(\mathbf{x}) > 1$, alors $(\phi - 1)^2 < (\phi - \hat{\phi}_{EVMM}(\mathbf{x}))^2$.

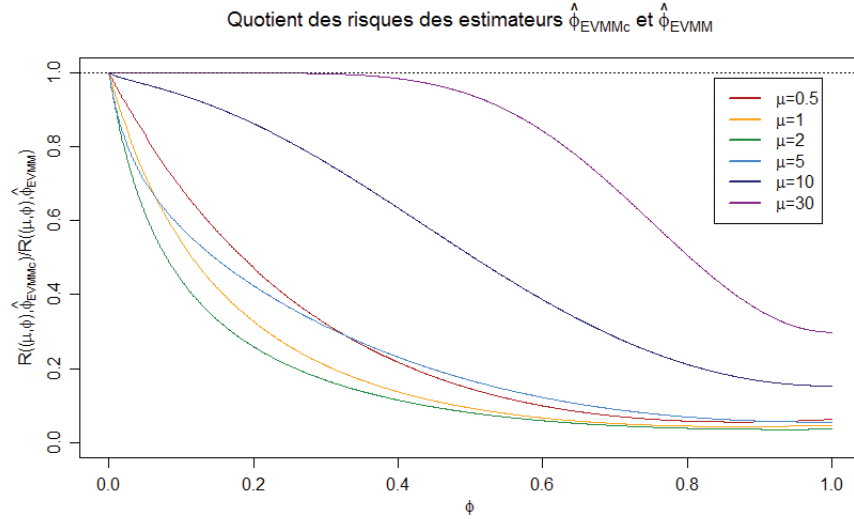


Figure 4.5 – Comparaison des estimateurs $\hat{\phi}_{EVMMc}$ et $\hat{\phi}_{EVMM}$

La question se pose également en ce qui concerne l'estimateur $\hat{\phi}_{B2}$ du chapitre 3. Contrairement à l'estimateur comparé précédemment, il n'y a pas un estimateur qui domine l'autre dans ce cas pour toutes valeurs de ϕ . En effet, il semble que pour ϕ suffisamment petit, l'estimateur $\hat{\phi}_{B2}$ soit meilleur que l'estimateur $\hat{\phi}_{EVMMc}$. Par exemple, supposons que $\mu = 5$ et $\phi = 0, 1$. Alors $R((5, 0.1), \hat{\phi}_{EVMMc}) \approx 0, 0558$ et $R((5, 0.1), \hat{\phi}_{B2}) \approx 0, 0308$, et donc $\frac{R((5, 0.1), \hat{\phi}_{EVMMc})}{R((5, 0.1), \hat{\phi}_{B2})} \approx 1, 8117 > 1$, d'où le fait que l'estimateur $\hat{\phi}_{B2}$ est plus performant à l'égard du risque étudié. Toutefois, on observe le comportement inverse lorsque ϕ est suffisamment grand. Dans ce cas, l'estimateur $\hat{\phi}_{EVMMc}$, qui tient compte de la contrainte pour ϕ , présente un risque plus faible. Par exemple, si $\mu = 5$ et $\phi = 0, 9$. Alors $R((5, 0.9), \hat{\phi}_{EVMMc}) \approx 0, 0773$ et $R((5, 0.9), \hat{\phi}_{B2}) \approx 0, 3234$, et donc $\frac{R((5, 0.9), \hat{\phi}_{EVMMc})}{R((5, 0.9), \hat{\phi}_{B2})} \approx 0, 2390 < 1$.

4.3.2 Risque de l'estimateur bayésien

On veut ici analyser le risque de l'estimateur bayésien $\hat{\phi}_{Bc}$. Tout comme ce fut expliqué à la section 3.3, le choix des hyperparamètres avec l'approche bayésienne peut se faire selon une approche subjective, semi-informative ou non informative.

Dans le cas d'une approche subjective, on avait vu au chapitre 3 que, comme $X_2 \sim \mathcal{BN}\left(a, \frac{b}{T+b}\right)$, les choix des hyperparamètres a et b pouvaient être basés sur les formules d'espérance et de variance de X_2 , soit une méthode pouvant également être utilisée ici. Puis, comme $\phi \sim \mathcal{B}(c, d)$, on a $E[\phi] = \frac{c}{c+d}$ et $Var(\phi) = \frac{cd}{(c+d)^2(c+d+1)} = \frac{E[\phi](1-E[\phi])}{c+d+1}$. Ainsi, les hyperparamètres c et d peuvent être choisis de telle sorte que $E[\phi]$ et $Var(\phi)$ correspondent à des valeurs fixées *a priori*. On obtient alors

$$c = E[\phi] \left(\frac{E[\phi](1-E[\phi])}{Var(\phi)} - 1 \right); \quad (4.8)$$

$$d = \frac{E[\phi](1-E[\phi])^2}{Var(\phi)} + E[\phi] - 1. \quad (4.9)$$

Dans le cas d'une approche non informative, telle qu'abordée au chapitre 1, la règle de Jeffreys stipule, pour un vecteur paramétrique $\boldsymbol{\theta}$, de prendre $\pi(\boldsymbol{\theta}) \propto \sqrt{\det I(\boldsymbol{\theta})}$, où $I(\boldsymbol{\theta})$ est la matrice d'information de Fisher définie à l'équation (1.8). À l'Exemple 1.12, on a vu que pour le modèle ici étudié, selon la règle de Jeffreys, il fallait choisir $\pi(\mu, \phi) \propto 1/\sqrt{\phi}$. Ainsi, il suffit de prendre la loi impropre uniforme pour μ , qui, rappelons-le, peut être vue comme la loi limite d'une loi $\mathcal{G}(a, b)$ lorsque $a = 1$ et $b \rightarrow 0$, de même que $\phi \sim \mathcal{B}(1/2, 1)$. Du Théorème 4.1, on trouve alors, en posant $S = T = 1$, $\phi|\mathbf{x} \sim \mathcal{BG4}(x_1 + 1/2, 1, 2, x_1 + x_2 + 1)$, et du Corollaire 4.3,

$$\hat{\phi}_{Bc}(\mathbf{x}) = E[\phi|\mathbf{x}] = \left(\frac{x_1 + 1/2}{x_1 + 3/2} \right) \frac{{}_2F_1(1, x_1 + x_2 + 1; x_1 + 5/2; 1/2)}{{}_2F_1(1, x_1 + x_2 + 1; x_1 + 3/2; 1/2)}. \quad (4.10)$$

Le risque de cet estimateur est illustré à la Figure 4.6 pour différentes valeurs de μ .

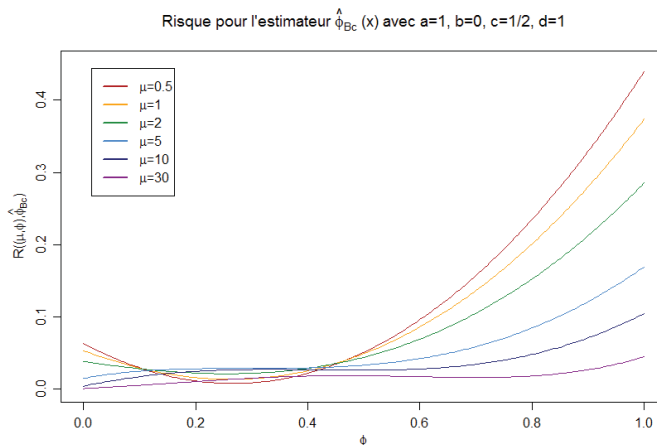
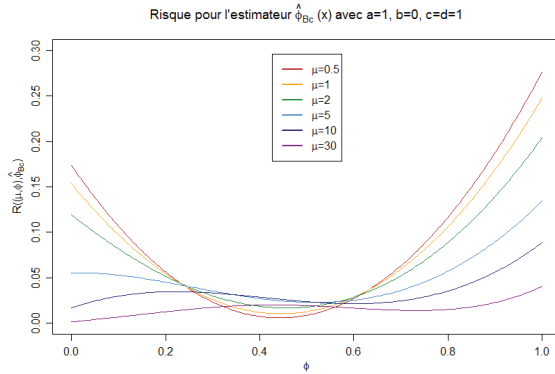


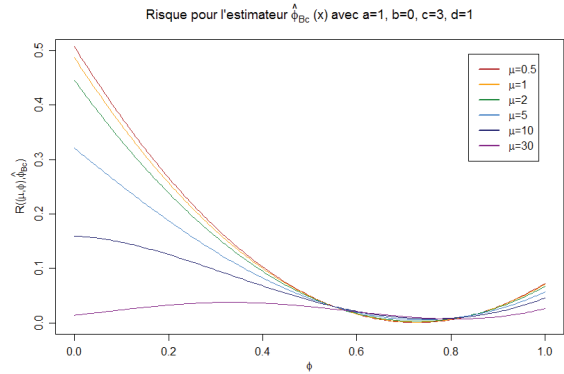
Figure 4.6 – Risque de l’estimateur $\hat{\phi}_{Bc}(\mathbf{x})$ avec $(a, b, c, d) = (1, 0, 1/2, 1)$

Tout d’abord, on remarque à nouveau la forme de parabole pour de petites valeurs de μ , par exemple, pour les courbes telles que $\mu = 0,5$ et $\mu = 1$. Le sommet pour ces courbes, soit l’endroit où le risque est à son minimum, se trouve environ en $\phi = 1/3$. Il s’agit en fait de l’espérance *a priori* de ϕ . C’est d’ailleurs un comportement que l’on observe pour différents vecteurs paramétriques (a, b, c, d) , tel qu’illustré à la Figure 4.7. En effet, le sommet des paraboles pour de petites valeurs de μ se situe toujours environ en $\phi = \frac{c}{c+d}$. Puis, pour $\phi \approx 0$ et pour $\phi \approx 1$, le risque est décroissant en μ .

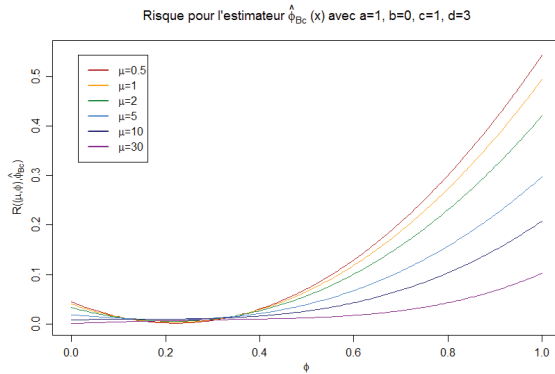
Par ailleurs, afin de voir s’il y a des gains sur la performance de l’estimateur en tenant compte de la contrainte, on a comparé le risque de l’estimateur $\hat{\phi}_{Bc}$ issu de l’approche non informative de Jeffreys à celui des estimateurs $\hat{\phi}_{EVMM}$ et $\hat{\phi}_{B2}$ du chapitre 3. Les risques sont présentés à la Figure 4.8. Contrairement à la dominance observée de $\hat{\phi}_{EVMMc}$ sur $\hat{\phi}_{EVMM}$ sous la contrainte $\phi \in [0, 1]$, l’estimateur $\hat{\phi}_{Bc}$ ne domine aucun des estimateurs qui ont été décrits au chapitre 3. On peut observer que $\hat{\phi}_{Bc}(\mathbf{x})$ est plus performant que $\hat{\phi}_{EVMM}(\mathbf{x})$ pour tout $\phi \in [0, 1]$ lorsque μ est assez faible. Par contre, le comportement inverse est observé si ϕ est suffisamment faible et μ suffisamment élevé. Pour ce qui est de la comparaison entre $\hat{\phi}_{Bc}(\mathbf{x})$ et $\hat{\phi}_{B2}(\mathbf{x})$, l’estimateur bayésien tenant compte de la



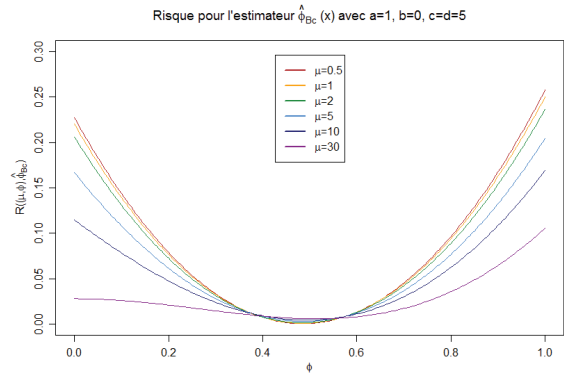
(a) Graphique 1 : $(a, b, c, d) = (1, 0, 1, 1)$



(b) Graphique 2 : $(a, b, c, d) = (1, 0, 3, 1)$



(c) Graphique 3 : $(a, b, c, d) = (1, 0, 1, 3)$



(d) Graphique 4 : $(a, b, c, d) = (1, 0, 5, 5)$

Figure 4.7 – Risque de l'estimateur $\hat{\phi}_{Bc}$ pour différents paramètres (a, b, c, d)

contrainte semble général plus performant que l'autre, sauf pour μ suffisamment grand et ϕ suffisamment petit, ou encore, pour μ près de 0 et ϕ près de 1.

4.3.3 Comparaison des risques des estimateurs

Parmi la famille d'estimateurs $\hat{\phi}_{Bc}$, nous analyserons celui issu de l'approche non informative, c'est-à-dire celui tel que $(a, b, c, d) = (1, 0, 1/2, 1)$, dont le risque est représenté à la Figure 4.6. Ainsi, pour la suite, la notation $\hat{\phi}_{Bc}(\mathbf{x})$ fera référence à cet estimateur

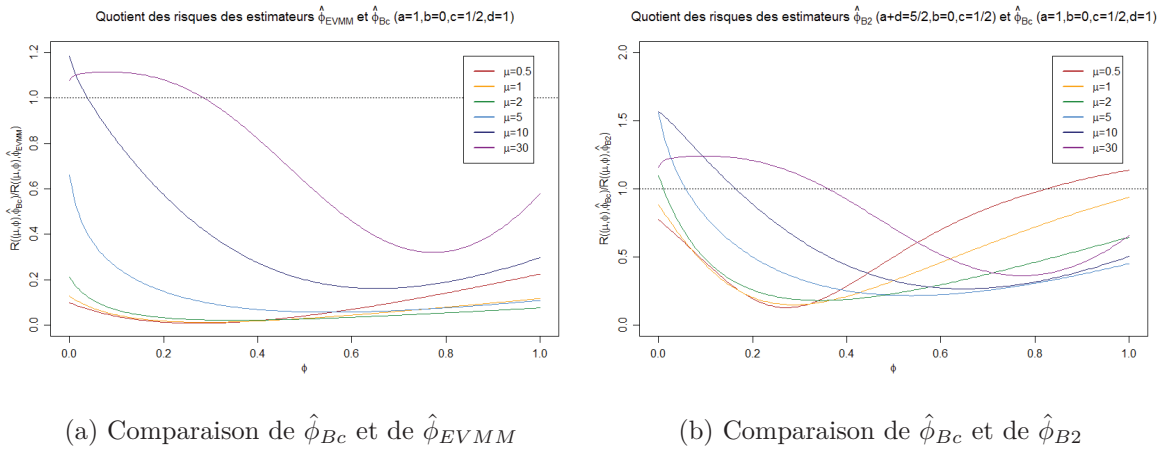


Figure 4.8 – Comparaison des risques de l’estimateur $\hat{\phi}_{Bc}$ et d’estimateurs qui ne tiennent pas compte de la contrainte sur ϕ

bayésien. Nous procéderons ici à la comparaison des risques des deux estimateurs abordés précédemment qui tiennent compte de la contrainte paramétrique sur ϕ , c’est-à-dire $\hat{\phi}_{EVMMc}(\mathbf{x})$ et $\hat{\phi}_{Bc}(\mathbf{x})$, et ce, en étudiant le quotient $\frac{R((\mu, \phi), \hat{\phi}_{Bc})}{R((\mu, \phi), \hat{\phi}_{EVMMc})}$, tel qu’illustré à la Figure 4.9.

L’analyse du quotient de ces risques est plutôt difficile en raison de la forme complexe de l’estimateur $\hat{\phi}_{Bc}$. De plus, le paramètre de nuisance μ complique l’étude puisque le comportement des courbes est différent selon sa valeur. Graphiquement, on peut tout de même observer certaines caractéristiques.

Tout d’abord, pour de plus faibles valeurs de ϕ , l’estimateur bayésien proposé est plus performant, en général, que celui issu d’une approche fréquentiste. Par contre, plus μ augmente, plus l’estimateur bayésien semble risqué pour de telles valeurs de ϕ . En effet, on voit que les courbes représentant $\mu = 10$ et $\mu = 30$ sont au-dessus de l’axe $y = 1$ lorsque ϕ est très faible, contrairement aux valeurs inférieures de μ . Ceci concorde d’ailleurs avec la Figure 4.4, où l’on peut voir que le risque est plutôt élevé pour de petites valeurs de ϕ .

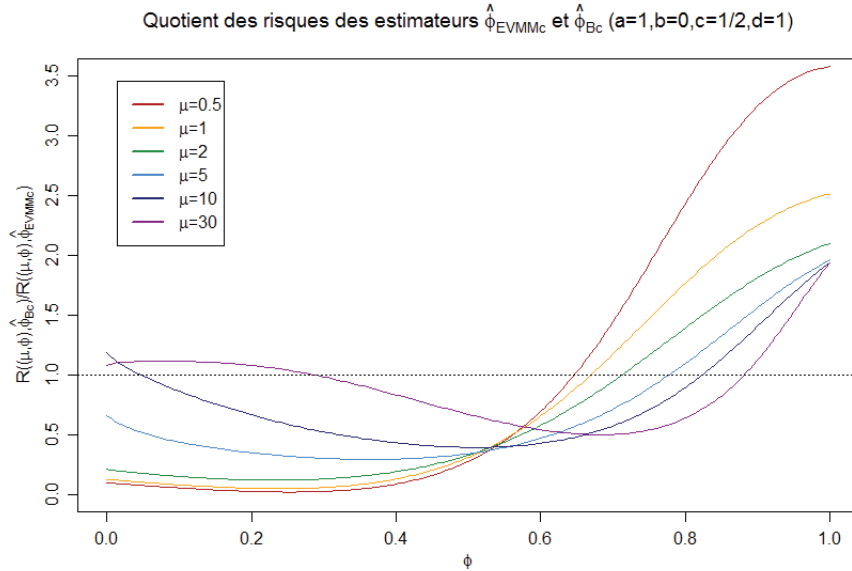


Figure 4.9 – Quotient des risques des estimateurs $\hat{\phi}_{Bc}(\mathbf{x})$ et $\hat{\phi}_{EVMMc}(\mathbf{x})$

Ensuite, lorsque ϕ est de valeur modérée, l'estimateur de Bayes est un meilleur choix d'estimateur que celui du maximum de vraisemblance modifié.

Puis, lorsque ϕ est près de 1, l'estimateur $\hat{\phi}_{EVMMc}$ a un risque plus faible que celui de l'estimateur $\hat{\phi}_{Bc}$, qui présentait un risque plutôt élevé pour de telles valeurs de ϕ à la Figure 4.6. En s'éloignant de $\phi = 1$, l'estimateur bayésien devient moins risqué que celui issu du maximum de vraisemblance, et on remarque que les courbes ne croisent pas l'axe $y = 1$ à la même valeur de ϕ . En effet, plus μ est petit, plus le point de croisement est éloigné de $\phi = 1$.

En somme, l'estimateur $\hat{\phi}_{Bc}$ est meilleur pour de petites valeurs de ϕ (sauf si μ est suffisamment grand) de même que pour des valeurs modérées de ϕ , alors que l'estimateur $\hat{\phi}_{EVMMc}$ est meilleur pour des valeurs de ϕ près de 1. Par contre, on avait soulevé précédemment que pour certaines valeurs de μ , le risque semblait influencé par le choix des paramètres c et d , le sommet des paraboles se déplaçant vers $\frac{c}{c+d}$. Il pourrait alors être

intéressant de voir comment se comparent les estimateurs des approches et fréquentiste et bayésienne selon différents choix de paramètres c et d . Par exemple, pour des valeurs de ϕ près de 1, est-ce que l'estimateur $\hat{\phi}_{EVMMc}$ est toujours le meilleur choix lorsque $\frac{c}{c+d}$ est près de 1 ?

4.4 Inférence prédictive

Dans cette dernière section, nous cherchons à estimer la fonction de masse de futures observations $\mathbf{Y} = (Y_1, Y_2)$ à partir des observations x_1 et x_2 . Il s'agit d'une section analogue à la section 3.4, mais ici, nous tiendrons compte de la contrainte $0 \leq \phi \leq 1$. On suppose à nouveau que $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$ et que $\mathbf{Y}|\mu, \phi \sim \mathcal{P}2(\mu\phi S', \mu T')$, où $S, S', T, T' > 0$. On rappelle que la fonction de masse de $\mathbf{Y}|\mu, \phi$ est

$$q_{\mu, \phi}(\mathbf{y}) = \frac{S'^{y_1} T'^{y_2}}{y_1! y_2!} e^{-\mu(T' + \phi S')} \mu^{y_1 + y_2} \phi^{y_1}.$$

Pour parvenir à l'estimer, on utilisera l'équation (1.33), et ce, en considérant les lois *a priori* $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi \sim \mathcal{B}(c, d)$.

Il peut être tout d'abord intéressant d'obtenir la distribution lorsqu'aucune information antérieure n'est disponible sur les variables. De la Proposition 3.2, c'est-à-dire en considérant μ comme un paramètre de nuisance de loi $\mathcal{G}(a, b)$, on a entre autres trouvé que les variables $X_1|\phi$, $X_2|\phi$ et $X_1|X_2, \phi$ sont toutes de distribution binomiale négative. En particulier, la loi de $X_2|\phi$ est indépendante de ϕ . Au Théorème 3.18, soit un résultat basé sur le Résultat 3.1 de [Lau12], on a trouvé les fonctions de masse résultantes des variables X_1 , X_2 et $X_1|X_2$ en supposant en plus que $\phi \sim \mathcal{B}_2\left(c, d, \frac{T+b}{S}\right)$.

Ainsi, le prochain théorème sera l'analogie de ce résultat, mais en considérant plutôt $\phi \sim \mathcal{B}(c, d)$, soit une hypothèse en continuité avec la section précédente et en accord avec la contrainte paramétrique imposée sur ϕ dans ce chapitre.

Théorème 4.7. Soient $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$, $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi \sim \mathcal{B}(c, d)$ avec $a, b, c, d > 0$. Alors, pour $\psi = \frac{S}{b}$ et $\xi = \frac{S}{T+b}$, on a les fonctions de masse suivantes :

1) Fonction de masse de \mathbf{X}

$$p_{\mathbf{X}}(\mathbf{x}) = \left(\frac{b}{S+T+b} \right)^a \frac{(a)_{x_1+x_2} (c)_{x_1}}{x_1! x_2! (c+d)_{x_1}} \frac{S^{x_1} T^{x_2}}{(S+T+b)^{x_1+x_2}} \\ \times {}_2F_1 \left(d, a+x_1+x_2; c+d+x_1; \frac{S}{S+T+b} \right); \quad (4.11)$$

2) Fonction de masse marginale de X_1

$$p_{X_1}(x_1) = \left(\frac{1}{1+\psi} \right)^a \frac{(a)_{x_1} (c)_{x_1}}{x_1! (c+d)_{x_1}} \left(\frac{\psi}{\psi+1} \right)^{x_1} \\ \times {}_2F_1 \left(d, a+x_1; c+d+x_1; \frac{\psi}{1+\psi} \right); \quad (4.12)$$

3) Fonction de masse conditionnelle de X_1 sachant X_2

$$p_{X_1|X_2}(x_1|x_2) = \left(\frac{1}{1+\xi} \right)^{a+x_2} \frac{(a+x_2)_{x_1} (c)_{x_1}}{x_1! (c+d)_{x_1}} \left(\frac{\xi}{1+\xi} \right)^{x_1} \\ \times {}_2F_1 \left(d, a+x_1+x_2; c+d+x_1; \frac{\xi}{1+\xi} \right). \quad (4.13)$$

Démonstration. Pour arriver à ces fonctions de masse, on utilise les résultats de la Proposition 3.2. Pour celle de \mathbf{X} , on a

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{(a)_{x_1+x_2}}{x_1! x_2!} b^a S^{x_1} T^{x_2} \int_0^1 \frac{\phi^{x_1}}{(T+b+\phi S)^{a+x_1+x_2}} \times \frac{\phi^{c-1} (1-\phi)^{d-1}}{B(c, d)} d\phi \\ = \frac{(a)_{x_1+x_2}}{x_1! x_2!} \frac{b^a S^{x_1} T^{x_2}}{B(c, d) (T+b)^{a+x_1+x_2}} \int_0^1 \frac{\phi^{c+x_1-1} (1-\phi)^{d-1}}{\left(1 + \frac{S}{T+b} \phi\right)^{a+x_1+x_2}} d\phi \\ = \frac{(a)_{x_1+x_2}}{x_1! x_2!} \frac{b^a S^{x_1} T^{x_2}}{B(c, d) (T+b)^{a+x_1+x_2}} \\ \times B(c+x_1, d) {}_2F_1 \left(c+x_1, a+x_1+x_2; c+d+x_1; \frac{-S}{T+b} \right),$$

et le résultat découle de la transformation de Pfaff donnée à la Proposition 2.2. On procède de manière analogue pour les fonctions de masse de X_1 et $X_1|X_2$. \square

De ce théorème, on reconnaît que la loi marginale de X_1 et la loi conditionnelle de X_1 étant donné X_2 sont des distributions qui ont été abordées au chapitre 3. En effet, elles font partie de la même famille de lois que celle des marginales d'une loi de Bailey décrite à la Proposition 3.27. On rappelle que cette famille de lois a été décrite dans l'article de Laurent [Lau12] au Lemme 2.2 et appelée la distribution Poisson-gamma-inverse bêta.

Les lois qui viennent d'être abordées se présentent dans un contexte bayésien où l'on n'a aucune information antérieure sur les variables. Toutefois, tel que présenté au début de la section, on s'intéressera plutôt au cas d'estimation de la fonction de masse $q_{\mu,\phi}$ lorsque l'on dispose d'observations X_1 et X_2 , avec $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi \sim \mathcal{B}(c, d)$. Pour bien comprendre les fonctions de masse prédictives qui seront présentées, il importe avant tout de définir une nouvelle famille de lois.

4.4.1 Loi de Bailey généralisée

On rappelle que la loi de Bailey définie en (3.26) peut être construite à partir des Propriétés 2.5 b) et c) de la première fonction hypergéométrique d'Appell. On peut construire une généralisation de cette loi en n'utilisant qu'une seule de ces propriétés. En effet, en partant de la propriété c) et en utilisant un développement similaire à celui de la loi de Bailey, on a, pour $|u| < 1$, $|v| < 1$, et $\frac{u-v}{1-v} < 1$,

$$F_1(a; b, b'; b + b'; u, v) = (1 - v)^{-a} {}_2F_1\left(a, b; b + b'; \frac{u - v}{1 - v}\right).$$

La réécriture de la fonction hypergéométrique d'Appell sous la forme d'une double série, telle que définie à l'équation (2.9), nous amène donc à la définition suivante, en posant $u = \frac{1}{1+\rho_1}$ et $v = \frac{1}{1+\rho_2}$.

Définition 4.8. *On dit que \mathbf{X} se distribue selon la loi de Bailey généralisée, notée $\mathcal{GBailey}(a, c, d, \rho_1, \rho_2)$ si sa fonction de masse s'écrit*

$$p(\mathbf{x}) = \frac{1}{{}_2F_1\left(a, c; c + d; \frac{\rho_2 - \rho_1}{\rho_2(1 + \rho_1)}\right)} \left(\frac{\rho_2}{1 + \rho_2}\right)^a \frac{(a)_{x_1+x_2} (c)_{x_1} (d)_{x_2}}{x_1! x_2! (c + d)_{x_1+x_2}} \times \left(\frac{1}{1 + \rho_1}\right)^{x_1} \left(\frac{1}{1 + \rho_2}\right)^{x_2} \mathbf{1}_{\mathbb{N}^2}(\mathbf{x}), \quad (4.14)$$

où $a, c, d, \rho_1, \rho_2 > 0$ et $\frac{\rho_2 - \rho_1}{\rho_2(1 + \rho_1)} < 1$.

Remarque 4.9. *En utilisant la transformation de Pfaff décrite à la Proposition 2.2, on peut réécrire la fonction de masse bivariée donnée en (4.14) de la manière suivante :*

$$p(\mathbf{x}) = \frac{1}{{}_2F_1\left(a, d; c + d; \frac{\rho_1 - \rho_2}{\rho_1(1 + \rho_2)}\right)} \left(\frac{\rho_1}{1 + \rho_1}\right)^a \frac{(a)_{x_1+x_2} (c)_{x_1} (d)_{x_2}}{x_1! x_2! (c + d)_{x_1+x_2}} \times \left(\frac{1}{1 + \rho_1}\right)^{x_1} \left(\frac{1}{1 + \rho_2}\right)^{x_2} \mathbf{1}_{\mathbb{N}^2}(\mathbf{x}).$$

Pour $a = c + d$, la fonction de masse en (4.14) s'écrit

$$p(\mathbf{x}) = \frac{(c)_{x_1}}{x_1!} \left(\frac{\rho_1}{1 + \rho_1}\right)^c \left(\frac{1}{1 + \rho_1}\right)^{x_1} \mathbf{1}_{\mathbb{N}}(x_1) \times \frac{(d)_{x_2}}{x_2!} \left(\frac{\rho_2}{1 + \rho_2}\right)^d \left(\frac{1}{1 + \rho_2}\right)^{x_2} \mathbf{1}_{\mathbb{N}}(x_2),$$

et ainsi, X_1 et X_2 sont des variables aléatoires indépendantes respectivement de loi $\mathcal{BN}\left(c, \frac{\rho_1}{1 + \rho_1}\right)$ et $\mathcal{BN}\left(d, \frac{\rho_2}{1 + \rho_2}\right)$.

La motivation initiale de la construction de cette loi était de trouver une généralisation de la loi de Bailey qui permettrait de faire un lien avec la fonction de masse prédictive $\hat{q}_\pi(\mathbf{y}; \mathbf{x})$ qui apparaît lorsque l'on suppose que $\phi \sim \mathcal{B}(c, d)$. L'idée a alors été de relâcher les propriétés qui avaient servi à la construction de la loi de Bailey (Propriétés 2.5 b) et c)), soit en tenant maintenant compte uniquement de la Propriété 2.5 c). Cette tentative n'a pas permis d'arriver au résultat escompté, mais un lien a pu être établi entre la famille de lois univariées obtenues à partir des marginales de cette loi et les marginales prédictives $\hat{q}_\pi(y_1; \mathbf{x})$ et $\hat{q}_\pi(y_2; \mathbf{x})$ du modèle étudié. De ce fait, l'intérêt ici porté à la loi de Bailey généralisée réside en ses lois marginales.

Théorème 4.10. *Si $\mathbf{X} \sim \mathcal{GBailey}(a, c, d, \rho_1, \rho_2)$, alors les fonctions de masse marginales sont données par*

$$p_{X_1}(x_1) = \frac{1}{{}_2F_1\left(d, a; c + d; \frac{\rho_1 - \rho_2}{\rho_1(1 + \rho_2)}\right)} \left(\frac{\rho_1}{1 + \rho_1}\right)^a \frac{(a)_{x_1}(c)_{x_1}}{x_1!(c + d)_{x_1}} \\ \times \left(\frac{1}{1 + \rho_1}\right)^{x_1} {}_2F_1\left(d, a + x_1; c + d + x_1; \frac{1}{1 + \rho_2}\right) \mathbf{1}_{\mathbb{N}}(x_1), \quad (4.15)$$

et

$$p_{X_2}(x_2) = \frac{1}{{}_2F_1\left(c, a; c + d; \frac{\rho_2 - \rho_1}{\rho_2(1 + \rho_1)}\right)} \left(\frac{\rho_2}{1 + \rho_2}\right)^a \frac{(a)_{x_2}(d)_{x_2}}{x_2!(c + d)_{x_2}} \\ \times \left(\frac{1}{1 + \rho_2}\right)^{x_2} {}_2F_1\left(c, a + x_2; c + d + x_2; \frac{1}{1 + \rho_1}\right) \mathbf{1}_{\mathbb{N}}(x_2). \quad (4.16)$$

Démonstration. Pour ce qui est de la fonction de masse de X_1 , on a

$$p_{X_1}(x_1) = \sum_{x_2=0}^{\infty} p(x_1, x_2) \\ = \frac{\left(\frac{\rho_1}{1 + \rho_1}\right)^a}{{}_2F_1\left(a, d; c + d; \frac{\rho_1 - \rho_2}{\rho_1(1 + \rho_2)}\right)} \frac{(a)_{x_1}(c)_{x_1}}{x_1!(c + d)_{x_1}} \left(\frac{1}{1 + \rho_1}\right)^{x_1} \\ \times \sum_{x_2=0}^{\infty} \frac{(a + x_1)_{x_2}(d)_{x_2}}{x_2!(c + d + x_1)_{x_2}} \left(\frac{1}{1 + \rho_2}\right)^{x_2} \mathbf{1}_{\mathbb{N}}(\mathbf{x}) \\ = \frac{1}{{}_2F_1\left(d, a; c + d; \frac{\rho_1 - \rho_2}{\rho_1(1 + \rho_2)}\right)} \left(\frac{\rho_1}{1 + \rho_1}\right)^a \frac{(a)_{x_1}(c)_{x_1}}{x_1!(c + d)_{x_1}} \left(\frac{1}{1 + \rho_1}\right)^{x_1} \\ \times {}_2F_1\left(d, a + x_1; c + d + x_1; \frac{1}{1 + \rho_2}\right) \mathbf{1}_{\mathbb{N}}(x_1),$$

d'où le résultat. On procède de manière analogue pour la fonction de masse de X_2 à partir de l'équation (4.14). \square

On remarque que les lois marginales d'une loi de Bailey généralisée sont de la même famille de lois. Ce qui suit nous permettra de mieux comprendre la structure de cette famille de lois.

Définition 4.11. On dit que X se distribue selon une loi binomiale négative pondérée avec poids $\omega(\cdot)$, qu'on notera ici $\mathcal{BNP}_\omega(r, p)$, si elle a pour fonction de masse

$$p(x) = \frac{1}{E[\omega(Z)]} \omega(x) \frac{(r)_x}{x!} p^r (1-p)^x \mathbb{1}_{\mathbb{N}}(x), \quad 0 \leq p \leq 1, \quad r > 0, \quad (4.17)$$

où $Z \sim \mathcal{BN}(r, p)$, et ω est une fonction $\mathbb{N} \rightarrow \mathbb{R}_+$ telle que $E[\omega(Z)] < \infty$.

On peut alors représenter les lois marginales d'une loi de Bailey généralisée sous forme de mélange à l'aide de la proposition suivante.

Proposition 4.12. Si Y est une variable aléatoire discrète avec fonction de masse

$$p(y) = \frac{\theta_1^a}{{}_2F_1\left(d, a; c+d; 1 - \frac{\theta_2}{\theta_1}\right)} \frac{(a)_y (c)_y}{y! (c+d)_y} (1 - \theta_1)^y {}_2F_1(d, a+y; c+d+y; 1 - \theta_2) \mathbb{1}_{\mathbb{N}}(y),$$

où $0 < \theta_1 < 1$ et $0 < \theta_2 < 1$, alors elle admet la représentation sous forme de mélange suivante :

$$Y|T = t \sim \mathcal{BNP}_\omega(c, \theta_1) \quad \text{et} \quad T \sim \mathcal{BN}(d, \theta_2),$$

où $\omega(y; t) = \frac{(a)_{y+t}}{(c+d)_{y+t}}$. En particulier, pour $a = c + d$, Y et T sont indépendantes et $Y \sim \mathcal{BN}(c, \theta_1)$.

Démonstration. Pour $y \in \mathbb{N}$, on a

$$\begin{aligned} p(y) &\propto \frac{(a)_y (c)_y}{y! (c+d)_y} (1 - \theta_1)^y {}_2F_1(d, a+y; c+d+y; 1 - \theta_2) \\ &= \sum_{t=0}^{\infty} \frac{(a)_y (c)_y}{y! (c+d)_y} (1 - \theta_1)^y \times \frac{(d)_t (a+y)_t (1 - \theta_2)^t}{(c+d+y)_t t!} \\ &= \sum_{t=0}^{\infty} \frac{(a)_{y+t}}{(c+d)_{y+t}} \frac{(c)_y}{y!} (1 - \theta_1)^y \times \frac{(d)_t}{t!} (1 - \theta_2)^t \\ &\propto \sum_{t=0}^{\infty} \frac{(a)_{y+t}}{(c+d)_{y+t}} \frac{(c)_y}{y!} \theta_1^c (1 - \theta_1)^y \times \frac{(d)_t}{t!} \theta_2^d (1 - \theta_2)^t. \quad \square \end{aligned}$$

Remarque 4.13. On peut trouver directement la constante de normalisation $K = \frac{1}{E[\omega(Z)]}$ de la proposition précédente. En effet, on a

$$\begin{aligned} K^{-1} &= \sum_{y=0}^{\infty} \frac{(a)_{y+t} (c)_y}{(c+d)_{y+t} y!} \theta_1^c (1-\theta_1)^y \\ &= \frac{(a)_t}{(c+d)_t} \theta_1^c \sum_{y=0}^{\infty} \frac{(a+t)_y (c)_y}{(c+d+t)_y} \frac{(1-\theta_1)^y}{y!} \\ &= \frac{(a)_t}{(c+d)_t} \theta_1^c {}_2F_1(a+t, c; c+d+t; 1-\theta_1). \end{aligned}$$

Le dernier résultat de cette section permet de calculer l'espérance d'une variable aléatoire telle que définie à la Proposition 4.12.

Proposition 4.14. Si Y est une variable aléatoire discrète avec fonction de masse

$$p(y) = \frac{\theta_1^a}{{}_2F_1\left(d, a; c+d; 1-\frac{\theta_2}{\theta_1}\right)} \frac{(a)_y (c)_y}{y! (c+d)_y} (1-\theta_1)^y {}_2F_1(d, a+y; c+d+y; 1-\theta_2) \mathbf{1}_{\mathbb{N}}(y), \quad (4.18)$$

où $0 < \theta_1 < 1$ et $0 < \theta_2 < 1$, alors

$$E[Y] = \frac{ac}{c+d} \left(\frac{1}{\theta_1} - 1 \right) \frac{{}_2F_1\left(d, a+1; c+d+1; 1-\frac{\theta_2}{\theta_1}\right)}{{}_2F_1\left(d, a; c+d; 1-\frac{\theta_2}{\theta_1}\right)}. \quad (4.19)$$

Démonstration. Posons $K = \frac{\theta_1^a}{{}_2F_1\left(d, a; c+d; 1-\frac{\theta_2}{\theta_1}\right)}$. Alors on a

$$\begin{aligned} E[Y] &= K \sum_{y=0}^{\infty} y \frac{(a)_y (c)_y}{y! (c+d)_y} (1-\theta_1)^y {}_2F_1(d, a+y; c+d+y; 1-\theta_2) \\ &= K \sum_{y=1}^{\infty} \frac{(a)_y (c)_y}{(y-1)! (c+d)_y} (1-\theta_1)^y {}_2F_1(d, a+y; c+d+y; 1-\theta_2) \\ &= K \sum_{w=0}^{\infty} \frac{(a)_{w+1} (c)_{w+1}}{w! (c+d)_{w+1}} (1-\theta_1)^{w+1} {}_2F_1(d, a+1+w; c+1+d+w; 1-\theta_2) \end{aligned}$$

$$= K \frac{ac}{c+d} (1-\theta_1) \sum_{w=0}^{\infty} \frac{(a+1)_w (c+1)_w}{w! (c+1+d)_w} (1-\theta_1)^w \\ \times {}_2F_1(d, a+1+w; c+1+d+w; 1-\theta_2).$$

Comme l'équation (4.18) représente une fonction de masse sur \mathbb{N} , on obtient

$$E[Y] = \frac{\theta_1^a}{{}_2F_1(d, a; c+d; 1-\frac{\theta_2}{\theta_1})} \frac{ac}{c+d} (1-\theta_1) \times \frac{{}_2F_1(d, a+1; c+d+1; 1-\frac{\theta_2}{\theta_1})}{\theta_1^{a+1}},$$

et le résultat de l'équation (4.19) s'en suit. \square

Remarque 4.15. Dans le cas où $a = c + d$, l'espérance $E[Y]$ se réduit à l'expression $\frac{c(1-\theta_1)}{\theta_1}$, ce qui correspond bel et bien à l'espérance d'une loi $\mathcal{BN}(c, \theta_1)$, en concordance avec la Proposition 4.12.

4.4.2 Estimateur de Bayes

Nous en sommes à calculer la fonction de masse prédictive pour $q_{\mu, \phi}(\mathbf{y})$ telle que décrite à l'équation (1.33). Il s'agit en fait de l'analogie du Résultat 3.2 de l'article de Laurent [Lau12] dans le cas où $\phi \sim \mathcal{B}(c, d)$ (motivé par la contrainte $\phi \in [0, 1]$).

Théorème 4.16. Soient $\mathbf{Y}|\mu, \phi \sim \mathcal{P}2(\mu\phi S', \mu T')$, $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$, $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi \sim \mathcal{B}(c, d)$, où $a, c, d > 0$, $b > -T$ et $S, S', T, T' > 0$. Alors, pour $\tau_1 = \frac{S+T+b}{S'}$, $\tau_2 = \frac{T+b}{S+S'}$, $\psi_1 = \frac{S+T+b}{T'}$ et $\psi_2 = \frac{T+T'+b}{S}$, la fonction de masse prédictive donnée en (1.33) est, pour $\mathbf{x} \in \mathbb{N}^2$ et $\mathbf{y} \in \mathbb{N}^2$,

$$(i) \quad \hat{q}_\pi(\mathbf{y}; \mathbf{x}) = \frac{\left(\frac{1}{1+1/\tau_1+1/\psi_1}\right)^{a+x_1+x_2}}{{}_2F_1\left(d, a+x_1+x_2; c+d+x_1; \frac{1+\psi_1}{\psi_1(1+\psi_2)}\right)} \frac{(a+x_1+x_2)_{y_1+y_2} (c+x_1)_{y_1}}{y_1! y_2! (c+d+x_1)_{y_1}} \\ \times \left(\frac{1}{1+\tau_1+\tau_1/\psi_1}\right)^{y_1} \left(\frac{1}{1+\psi_1+\psi_1/\tau_1}\right)^{y_2} \\ \times {}_2F_1\left(d, a+x_1+x_2+y_1+y_2; c+d+x_1+y_1; \frac{\tau_2(1+\psi_2)}{\psi_1\psi_2-1}\right), \quad (4.20)$$

et les fonctions de masse prédictives marginales correspondantes s'expriment sous la forme des mélanges suivants :

$$(ii) \quad Y_1|U = u \sim \mathcal{BNP}_{\omega_1} \left(c + x_1, \frac{\tau_1}{1 + \tau_1} \right) \quad \text{et} \quad U \sim \mathcal{BN} \left(d, \frac{\tau_2}{1 + \tau_2} \right), \quad (4.21)$$

$$\text{où } \omega_1(z) = \frac{(a+x_1+x_2)_{z+u}}{(c+d+x_1)_{z+u}};$$

$$(iii) \quad Y_2|V = v \sim \mathcal{BN} \left(a + x_1 + x_2 + v; \frac{\psi_1}{1 + \psi_1} \right) \quad \text{et} \quad V \sim \mathcal{BNP}_{\omega_2} \left(d, \frac{\psi_1\psi_2 - 1}{\psi_1(1 + \psi_2)} \right), \quad (4.22)$$

$$\text{où } \omega_2(z) = \frac{(a+x_1+x_2)_z}{(c+d+x_1)_z}.$$

Remarque 4.17. La fonction de masse prédictive marginale $\hat{q}_\pi(y_1; \mathbf{x})$ fait partie de la même famille de loi univariée que celle des marginales de la loi de Bailey généralisée. Elle peut donc être réécrite sous la forme suivante :

$$\begin{aligned} \hat{q}_\pi(y_1; \mathbf{x}) &= \frac{\left(\frac{\tau_1}{1+\tau_1} \right)^{a+x_1+x_2}}{{}_2F_1 \left(d, a + x_1 + x_2; c + d + x_1; \frac{\tau_1 - \tau_2}{\tau_1(1+\tau_2)} \right)} \frac{(a + x_1 + x_2)_{y_1} (c + x_1)_{y_1}}{y_1! (c + d + x_1)_{y_1}} \\ &\quad \times \left(\frac{1}{1 + \tau_1} \right)^{y_1} {}_2F_1 \left(d, a + x_1 + x_2 + y_1; c + d + x_1 + y_1; \frac{1}{1 + \tau_2} \right). \end{aligned} \quad (4.23)$$

Toutefois, la densité prédictive marginale $\hat{q}_\pi(y_2; \mathbf{x})$ admet une structure légèrement différente de celle de Y_1 , bien qu'elle fasse tout de même intervenir la classe de lois binomiales négatives pondérées. De l'équation (4.22), par calcul direct, on trouve

$$\begin{aligned} \hat{q}_\pi(y_2; \mathbf{x}) &= \frac{\left(\frac{\psi_1}{1+\psi_1} \right)^{a+x_1+x_2}}{{}_2F_1 \left(d, a + x_1 + x_2; c + d + x_1; \frac{1+\psi_1}{\psi_1(1+\psi_2)} \right)} \frac{(a + x_1 + x_2)_{y_2}}{y_2!} \\ &\quad \times \left(\frac{1}{1 + \psi_1} \right)^{y_2} {}_2F_1 \left(d, a + x_1 + x_2 + y_2; c + d + x_1; \frac{1}{1 + \psi_2} \right). \end{aligned} \quad (4.24)$$

Démonstration du Théorème 4.16. Nous démontrerons (i) par calcul direct. Nous utiliserons ensuite cette équation pour démontrer (ii) et (iii) par l'obtention des fonctions de masse données aux équations (4.23) et (4.24).

- (i) Des Théorèmes 3.4 et 4.1, on sait que $\mu|\phi, \mathbf{x} \sim \mathcal{G}(a + x_1 + x_2, T + b + \phi S)$ et que $\phi|\mathbf{x} \sim \mathcal{BG4}(c + x_1, d, \frac{S+T+b}{T+b}, a + x_1 + x_2)$.

De l'équation (1.33), on a

$$\begin{aligned} \hat{q}_\pi(\mathbf{y}; \mathbf{x}) &= \int_0^1 \int_0^\infty p(y_1, y_2 | \mu, \phi) \pi(\mu | \phi, x_1, x_2) \pi(\phi | x_1, x_2) d\mu d\phi \\ &= \int_0^1 \int_0^\infty \frac{e^{-\mu\phi S'} (\mu\phi S')^{y_1}}{y_1!} \frac{e^{-\mu T'} (\mu T')^{y_2}}{y_2!} \\ &\quad \times \frac{(T + b + \phi S)^{a+x_1+x_2} \mu^{a+x_1+x_2-1} e^{-\mu(T+b+S\phi)}}{\Gamma(a + x_1 + x_2)} \times \frac{\phi^{c+x_1-1} (1-\phi)^{d-1}}{(T + b + S\phi)^{a+x_1+x_2}} \\ &\quad \times \frac{(S + T + b)^{a+x_1+x_2}}{B(c + x_1, d) {}_2F_1\left(d, a + x_1 + x_2; c + d + x_1; \frac{S}{S+T+b}\right)} d\mu d\phi. \end{aligned}$$

En posant $K_1(\mathbf{y}) = \frac{S^{y_1} T^{y_2}}{y_1! y_2! \Gamma(a+x_1+x_2)} \frac{(S+T+b)^{a+x_1+x_2}}{B(c+x_1, d) {}_2F_1\left(d, a+x_1+x_2; c+d+x_1; \frac{S}{S+T+b}\right)}$, on obtient

$$\begin{aligned} \hat{q}_\pi(\mathbf{y}; \mathbf{x}) &= K_1(\mathbf{y}) \int_0^1 \phi^{c+x_1+y_1-1} (1-\phi)^{d-1} \\ &\quad \times \int_0^\infty e^{-\mu(T+T'+b+\phi(S+S'))} \mu^{a+x_1+x_2+y_1+y_2-1} d\mu d\phi \\ &= K_1(\mathbf{y}) \int_0^1 \phi^{c+x_1+y_1-1} (1-\phi)^{d-1} \\ &\quad \times \frac{\Gamma(a + x_1 + x_2 + y_1 + y_2)}{(T + T' + b + \phi(S + S'))^{a+x_1+x_2+y_1+y_2}} d\phi. \end{aligned}$$

En posant $K_2(\mathbf{y}) = K_1(\mathbf{y}) \frac{\Gamma(a+x_1+x_2+y_1+y_2)}{(T+T'+b)^{a+x_1+x_2+y_1+y_2}}$, on obtient

$$\begin{aligned} \hat{q}_\pi(\mathbf{y}; \mathbf{x}) &= K_2(\mathbf{y}) \int_0^1 \frac{\phi^{c+x_1+y_1-1} (1-\phi)^{d-1}}{\left(1 + \frac{S+S'}{T+T'+b}\phi\right)^{a+x_1+x_2+y_1+y_2}} d\phi \\ &= K_2(\mathbf{y}) \times B(c + x_1 + y_1, d) \\ &\quad \times {}_2F_1\left(c + x_1 + y_1, a + x_1 + x_2 + y_1 + y_2; c + d + x_1 + y_1; -\frac{S + S'}{T + T' + b}\right) \end{aligned}$$

$$\begin{aligned}
&= K_2(\mathbf{y}) \times B(c + x_1 + y_1, d) \left(\frac{T + T' + b}{S + S' + T + T' + b} \right)^{a+x_1+x_2+y_1+y_2} \\
&\quad \times {}_2F_1 \left(d, a + x_1 + x_2 + y_1 + y_2; c + d + x_1 + y_1; \frac{S + S'}{S + S' + T + T' + b} \right),
\end{aligned}$$

où la dernière égalité découle de la transformation de Pfaff donnée en (2.6.b). Le résultat suit directement en substituant $K_2(\mathbf{y})$ puis $K_1(\mathbf{y})$.

(ii) Passons maintenant à la fonction de masse prédictive marginale de Y_1 , que nous calculerons de la manière suivante :

$$\hat{q}_\pi(y_1; \mathbf{x}) = \sum_{y_2=0}^{\infty} \hat{q}_\pi(y_1, y_2; \mathbf{x}).$$

Posons

$$\begin{aligned}
K_3(y_1) &= \frac{\left(\frac{S+T+b}{S+T+b+S'+T'} \right)^{a+x_1+x_2}}{{}_2F_1 \left(d, a + x_1 + x_2; c + d + x_1; \frac{S}{S+T+b} \right)} \\
&\quad \times \frac{(a + x_1 + x_2)_{y_1} (c + x_1)_{y_1}}{y_1! (c + d + x_1)_{y_1}} \left(\frac{S'}{S + T + b + S' + T'} \right)^{y_1}.
\end{aligned}$$

Comme $(a+x_1+x_2)_{y_1+y_2} = (a+x_1+x_2+y_1)_{y_2} (a+x_1+x_2)_{y_1}$, on a, de l'équation (4.20),

$$\begin{aligned}
\hat{q}_\pi(y_1; \mathbf{x}) &= K_3(y_1) \sum_{y_2=0}^{\infty} \frac{(a + x_1 + x_2 + y_1)_{y_2}}{y_2!} \left(\frac{T'}{S + T + b + S' + T'} \right)^{y_2} \\
&\quad \times {}_2F_1 \left(d, a + x_1 + x_2 + y_1 + y_2; c + d + x_1 + y_1; \frac{S + S'}{S + S' + T + T' + b} \right) \\
&= K_3(y_1) \sum_{y_2=0}^{\infty} \frac{(a + x_1 + x_2 + y_1)_{y_2}}{y_2!} \left(\frac{T'}{S + T + b + S' + T'} \right)^{y_2} \\
&\quad \times \sum_{n=0}^{\infty} \frac{(d)_n (a + x_1 + x_2 + y_1 + y_2)_n}{n! (c + d + x_1 + y_1)_n} \left(\frac{S + S'}{S + S' + T + T' + b} \right)^n \\
&= K_3(y_1) \sum_{y_2=0}^{\infty} \sum_{n=0}^{\infty} \frac{(a + x_1 + x_2 + y_1)_{n+y_2} (d)_n}{n! y_2! (c + d + x_1 + y_1)_n} \frac{T'^{y_2} (S + S')^n}{(S + S' + T + T' + b)^{n+y_2}}
\end{aligned}$$

$$\begin{aligned}
&= K_3(y_1) \sum_{y_2=0}^{\infty} \sum_{n=0}^{\infty} \frac{(a+x_1+x_2+y_1)_n (a+x_1+x_2+y_1+n)_{y_2} (d)_n}{n! y_2! (c+d+x_1+y_1)_n} \\
&\quad \times \frac{T^{y_2} (S+S')^n}{(S+S'+T+T'+b)^{n+y_2}} \\
&= K_3(y_1) \sum_{n=0}^{\infty} \frac{(a+x_1+x_2+y_1)_n (d)_n}{n! (c+d+x_1+y_1)_n} \left(\frac{S+S'}{S+S'+T+T'+b} \right)^n \\
&\quad \times \sum_{y_2=0}^{\infty} \frac{(a+x_1+x_2+y_1+n)_{y_2}}{y_2!} \left(\frac{T'}{S+S'+T+T'+b} \right)^{y_2} \\
&= K_3(y_1) \sum_{n=0}^{\infty} \frac{(a+x_1+x_2+y_1)_n (d)_n}{n! (c+d+x_1+y_1)_n} \left(\frac{S+S'}{S+S'+T+T'+b} \right)^n \\
&\quad \times \left(\frac{S+S'+T+T'+b}{S+S'+T+b} \right)^{a+x_1+x_2+y_1+n}
\end{aligned}$$

En posant $K_4(y_1) = K_3(y_1) \left(\frac{S+S'+T+T'+b}{S+S'+T+b} \right)^{a+x_1+x_2+y_1}$, on obtient

$$\begin{aligned}
\hat{q}_\pi(y_1; \mathbf{x}) &= K_4(y_1) \sum_{n=0}^{\infty} \frac{(a+x_1+x_2+y_1)_n (d)_n}{n! (c+d+x_1+y_1)_n} \left(\frac{S+S'}{S+S'+T+b} \right)^n \\
&= K_4(y_1) \times {}_2F_1 \left(d, a+x_1+x_2+y_1; c+d+x_1+y_1; \frac{S+S'}{S+S'+T+b} \right),
\end{aligned}$$

et le résultat suit en substituant $K_4(y_1)$ puis $K_3(y_1)$.

(iii) Passons maintenant à $\hat{q}_\pi(y_2; \mathbf{x})$, que nous calculerons de la manière suivante :

$$\hat{q}_\pi(y_2; \mathbf{x}) = \sum_{y_1=0}^{\infty} \hat{q}_\pi(y_1, y_2; \mathbf{x}).$$

Posons

$$\begin{aligned}
K_5(y_2) &= \frac{\left(\frac{S+T+b}{S+T+b+S'+T'} \right)^{a+x_1+x_2}}{{}_2F_1 \left(d, a+x_1+x_2; c+d+x_1; \frac{S}{S+T+b} \right)} \frac{(a+x_1+x_2)_{y_2}}{y_2!} \\
&\quad \times \left(\frac{T'}{S+T+b+S'+T'} \right)^{y_2}.
\end{aligned}$$

Comme $(a + x_1 + x_2)_{y_1+y_2} = (a + x_1 + x_2 + y_2)_{y_1}(a + x_1 + x_2)_{y_2}$, on a, de (4.20),

$$\begin{aligned}\hat{q}_\pi(y_2; \mathbf{x}) &= K_5(y_2) \sum_{y_1=0}^{\infty} \frac{(a + x_1 + x_2 + y_2)_{y_1}(c + x_1)_{y_1}}{y_1!(c + d + x_1)_{y_1}} \left(\frac{S'}{S + T + b + S' + T'} \right)^{y_1} \\ &\quad \times {}_2F_1 \left(d, a + x_1 + x_2 + y_1 + y_2; c + d + x_1 + y_1; \frac{S + S'}{S + S' + T + T' + b} \right) \\ &= K_5(y_2) \sum_{y_1=0}^{\infty} \frac{(a + x_1 + x_2 + y_2)_{y_1}(c + x_1)_{y_1}}{y_1!(c + d + x_1)_{y_1}} \left(\frac{S'}{S + T + b + S' + T'} \right)^{y_1} \\ &\quad \times {}_2F_1 \left(a + x_1 + x_2 + y_2 + y_1, d; c + d + x_1 + y_1; \frac{S + S'}{S + S' + T + T' + b} \right).\end{aligned}$$

De cette dernière expression, on constate, à partir de la Propriétés 2.5 a) que cette somme est en fait la première fonction hypergéométrique d'Appell suivante :

$$F_1 \left(a + x_1 + x_2 + y_2; c + x_1, d; c + d + x_1; \frac{S'}{S + S' + T + T' + b}, \frac{S + S'}{S + S' + T + T' + b} \right). \quad (4.25)$$

On remarque également que le quatrième paramètre de cette fonction est en fait la somme des deux précédents, ce qui implique que l'on peut utiliser la Propriété 2.5 c). Ainsi, on obtient :

$$\begin{aligned}\hat{q}_\pi(y_2; \mathbf{x}) &= K_5(y_2) \left(\frac{S + S' + T + T' + b}{T + T' + b} \right)^{a+x_1+x_2+y_2} \\ &\quad \times {}_2F_1 \left(a + x_1 + x_2 + y_2, c + x_1; c + d + x_1; \frac{-S}{T + T' + b} \right). \\ &= K_5(y_2) \left(\frac{S + S' + T + T' + b}{T + T' + b} \right)^{a+x_1+x_2+y_2} \\ &\quad \times {}_2F_1 \left(c + x_1, a + x_1 + x_2 + y_2; c + d + x_1; \frac{-S}{T + T' + b} \right).\end{aligned}$$

L'équation (4.24) est obtenue en appliquant la transformation de Pfaff donnée en (2.6.b) à cette dernière équation et en substituant $K_5(y_2)$. \square

Exemple 4.18. *Supposons que l'on cherche à estimer la fonction de masse de $\mathbf{Y}|\mu, \phi \sim \mathcal{P}2(\mu\phi, \mu)$ en sachant que l'on a observé \mathbf{x} , où $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi, \mu)$ et que $\mathbf{X}|\mu, \phi$ et $\mathbf{Y}|\mu, \phi$ sont indépendantes. Supposons en plus que ϕ est contraint à l'intervalle $[0, 1]$. Du Théorème 4.16, en considérant par exemple la loi a priori $\pi(\mu, \pi) \propto 1/\sqrt{\phi}$ non informative pour (μ, ϕ) (c'est-à-dire $a = 1, b = 0, c = 1/2$ et $d = 1$), on trouve*

$$\hat{q}_\pi(\mathbf{y}; \mathbf{x}) = \frac{\left(\frac{1}{2}\right)^{x_1+x_2+1}}{{}_2F_1\left(1, x_1 + x_2 + 1; x_1 + \frac{3}{2}; \frac{1}{2}\right)} \frac{(x_1 + x_2 + 1)_{y_1+y_2} \left(x_1 + \frac{1}{2}\right)_{y_1}}{y_1! y_2! \left(x_1 + \frac{3}{2}\right)_{y_1}} \left(\frac{1}{4}\right)^{y_1+y_2} \\ \times {}_2F_1\left(1, x_1 + x_2 + 1 + y_1 + y_2; x_1 + \frac{3}{2} + y_1; \frac{1}{2}\right), \quad (4.26)$$

et pour les fonctions de masse marginales prédictives, on trouve

$$\hat{q}_\pi(y_1; \mathbf{x}) = \frac{\left(\frac{2}{3}\right)^{x_1+x_2+1}}{{}_2F_1\left(1, x_1 + x_2 + 1; x_1 + \frac{3}{2}; \frac{1}{2}\right)} \frac{(x_1 + x_2 + 1)_{y_1} \left(x_1 + \frac{1}{2}\right)_{y_1}}{y_1! \left(x_1 + \frac{3}{2}\right)_{y_1}} \left(\frac{1}{3}\right)^{y_1} \\ {}_2F_1\left(1, x_1 + x_2 + 1 + y_1; x_1 + \frac{3}{2} + y_1; \frac{2}{3}\right), \quad (4.27)$$

et

$$\hat{q}_\pi(y_2; \mathbf{x}) = \frac{\left(\frac{2}{3}\right)^{x_1+x_2+1}}{{}_2F_1\left(1, x_1 + x_2 + 1; x_1 + \frac{3}{2}; \frac{1}{2}\right)} \frac{(x_1 + x_2 + 1)_{y_2}}{y_2!} \left(\frac{1}{3}\right)^{y_2} \\ {}_2F_1\left(1, x_1 + x_2 + 1 + y_2; x_1 + \frac{3}{2}; \frac{1}{3}\right). \quad (4.28)$$

Ces fonctions de masse marginales ont été représentées aux Figures 4.10 et 4.11. Tout comme au chapitre 3, ces fonctions de masse pourraient être notamment comparées à celles trouvées au Théorème 4.1 pour voir l'impact qu'a le fait d'avoir des données observées en lien avec les paramètres inconnus.

En outre, on remarque que, pour cet exemple particulier avec $(x_1, x_2) = (17, 31)$, $a = 1, b = 0, c = 1/2$ et $d = 1$, ces fonctions de masse prédictives sont graphiquement très semblables à celles obtenues au chapitre 3 à l'Exemple 3.32 pour le choix de lois

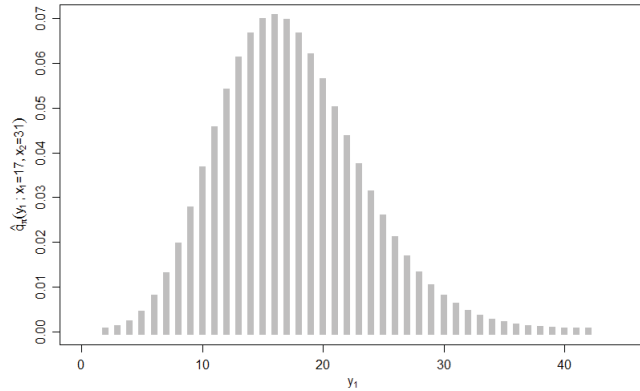


Figure 4.10 – Fonction de masse prédictive de Y_1 pour $x_1 = 17$ et $x_2 = 31$

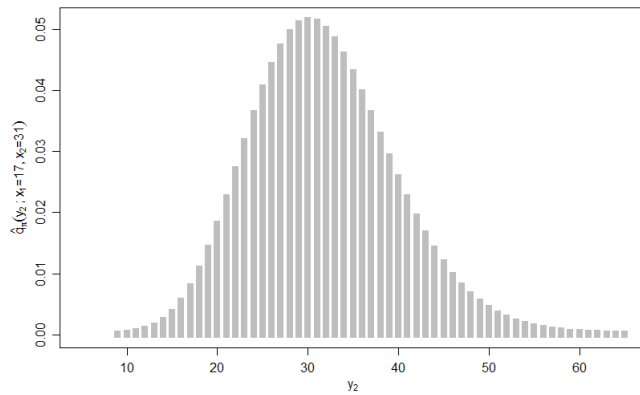


Figure 4.11 – Fonction de masse prédictive de Y_2 pour $x_1 = 17$ et $x_2 = 31$

a priori spécifique qui avait été fait. Ainsi, bien la forme analytique des fonctions de masse prédictives obtenues aux chapitres 3 et 4 soient différentes, on se peut demander si certaines approximations de ces fonctions peuvent être faites afin d'expliquer la similitude considérable que l'on observe graphiquement pour cet exemple en particulier.

4.4.3 Espérances, variances et covariance

Cette dernière partie traite de la « calibration » sous \hat{q}_π de l'espérance de $\mathbf{Y}|\mu, \phi$ à partir d'une observation indépendante \mathbf{x} , à l'image de la section 3.4.2 du chapitre 3, mais en tenant compte de la contrainte $\phi \in [0, 1]$. On se basera notamment sur les lemmes suivants pour obtenir certaines expressions.

Lemme 4.19. *Soient $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi \sim \mathcal{B}(c, d)$, où $a, b, c, d > 0$. Alors*

$$E[\mu^m \phi^n] = \frac{(a)_m (c)_n}{b^m (c+d)_n}, \quad m > -a, \quad n > -c. \quad (4.29)$$

Lemme 4.20. *Soient $\mu|\phi, \mathbf{x} \sim \mathcal{G}(a + x_1 + x_2, T + b + \phi S)$ et $\phi|\mathbf{x} \sim \mathcal{BG4}(c + x_1, d, \frac{S+T+b}{T+b}, a + x_1 + x_2)$, où $\mathbf{x} \in \mathbb{N}^2$, $a, c, d > 0$, $b > -T$ et $S, T > 0$. Alors*

$$E[\mu^m \phi^n | \mathbf{x}] = \frac{1}{(S+T+b)^m} \frac{(a+x_1+x_2)_m (c+x_1)_n}{(c+d+x_1)_n} \times \frac{{}_2F_1\left(d, a+x_1+x_2+m; c+d+x_1+n; \frac{S}{S+T+b}\right)}{{}_2F_1\left(d, a+x_1+x_2; c+d+x_1; \frac{S}{S+T+b}\right)}, \quad m, n \geq 0. \quad (4.30)$$

Démonstration. On a

$$\begin{aligned} E[\mu^m \phi^n | \mathbf{x}] &= \frac{(a+x_1+x_2)_m}{(T+b)^m} E^{\phi|\mathbf{x}} \left[\frac{\phi^n}{\left(1 + \frac{S}{T+b}\phi\right)^m} \right] \\ &= \frac{(a+x_1+x_2)_m}{(T+b)^m} \frac{1}{B(c+x_1, d) {}_2F_1\left(c+x_1, a+x_1+x_2; c+d+x_1; -\frac{S}{T+b}\right)} \\ &\quad \times \int_0^\infty \frac{\phi^{c+x_1+n-1} (1-\phi)^d}{\left(1 + \frac{S}{T+b}\phi\right)^{a+x_1+x_2+m}} d\phi \\ &= \frac{(a+x_1+x_2)_m}{(T+b)^m} \\ &\quad \times \frac{B(c+x_1+n, d) {}_2F_1\left(c+x_1+n, a+x_1+x_2+m; c+d+n+x_1; -\frac{S}{T+b}\right)}{B(c+x_1, d) {}_2F_1\left(c+x_1, a+x_1+x_2; c+d+x_1; -\frac{S}{T+b}\right)} \end{aligned}$$

$$= \frac{(a + x_1 + x_2)_m (c + x_1)_n}{(S + T + b)^m (c + d + x_1)_n} \frac{{}_2F_1\left(d, a + x_1 + x_2 + m; c + d + n + x_1; \frac{S}{S+T+b}\right)}{{}_2F_1\left(d, a + x_1 + x_2; c + d + x_1; \frac{S}{S+T+b}\right)}.$$

□

Au chapitre 3, les moments mixtes *a posteriori* avaient notamment permis de vérifier la formule de covariance de Y_1 et Y_2 sous \hat{q}_π qui était donnée dans [Lau12]. Or, à partir de la densité prédictive trouvée au Théorème 4.16, le Lemme 4.20 et le Théorème 1.59 nous permettent de trouver que la covariance s'exprime comme

$$\begin{aligned} \text{Cov}_{\hat{q}_\pi}(\mathbf{Y}) = & \frac{S'T'\alpha\gamma}{(S + T + b)^2(\gamma + \delta)} \left((\alpha + 1) \frac{{}_2F_1\left(\delta, \alpha + 2; \gamma + \delta + 1; \frac{S}{S+T+b}\right)}{{}_2F_1\left(\delta, \alpha; \gamma + \delta; \frac{S}{S+T+b}\right)} \right. \\ & \left. - \alpha \frac{{}_2F_1\left(\delta, \alpha + 1; \gamma + \delta + 1; \frac{S}{S+T+b}\right) {}_2F_1\left(\delta, \alpha + 1; \gamma + \delta; \frac{S}{S+T+b}\right)}{{}_2F_1\left(\delta, \alpha; \gamma + \delta; \frac{S}{S+T+b}\right)^2} \right), \end{aligned} \quad (4.31)$$

où $\alpha = a + x_1 + x_2$, $\gamma = c + x_1$ et $\delta = d$.

La forme est beaucoup plus complexe que celle trouvée au chapitre 3. De ce fait, la question suivante (ouverte) se pose : « Est-ce qu'il existerait une bonne approximation de la covariance donnée à l'équation (4.31) ? »

Enfin, les moments mixtes trouvées aux lemmes précédents permettent en plus de trouver des expressions pour les espérances de Y_1 et Y_2 .

Proposition 4.21. *Soient $\mathbf{Y}|\mu, \phi \sim \mathcal{P}2(\mu\phi S', \mu T')$, $\mathbf{X}|\mu, \phi \sim \mathcal{P}2(\mu\phi S, \mu T)$, $\mu|\phi \sim \mathcal{G}(a, b)$ et $\phi \sim \mathcal{B}(c, d)$, où $a, c, d > 0$, $b > -T$ et $S, S', T, T' > 0$. Alors*

$$E[Y_1] = \frac{S'ac}{b(c+d)}, \quad (4.32)$$

et

$$E[Y_2] = \frac{T'a}{b}, \quad b > 0. \quad (4.33)$$

À partir d'observations $(x_1, x_2) \in \mathbb{N}^2$, sous les fonctions de masse prédictives définies au Théorème 4.16, pour $\alpha = a + x_1 + x_2$, $\gamma = c + x_1$ et $\delta = d$, on a

$$E_{\hat{q}_\pi}[Y_1; \mathbf{x}] = \frac{S'\alpha\gamma}{(S+T+b)(\gamma+\delta)} \frac{{}_2F_1\left(\delta, \alpha+1; \gamma+\delta+1; \frac{S}{S+T+b}\right)}{{}_2F_1\left(\delta, \alpha; \gamma+\delta; \frac{S}{S+T+b}\right)} \quad (4.34)$$

et

$$E_{\hat{q}_\pi}[Y_2; \mathbf{x}] = \left(\frac{T'\alpha}{S+T+b}\right) \frac{{}_2F_1\left(\delta, \alpha+1; \gamma+\delta; \frac{S}{S+T+b}\right)}{{}_2F_1\left(\delta, \alpha; \gamma+\delta; \frac{S}{S+T+b}\right)}. \quad (4.35)$$

Démonstration. Cela suit des théorèmes 1.55 et 1.56 et des lemmes 4.19 et 4.20. \square

Exemple 4.22. Reprenons le contexte de l'Exemple 4.18. De la Proposition 4.21, on a

$$E_{\hat{q}_\pi}[Y_1; \mathbf{x}] = \frac{(x_1+x_2+1)(x_1+1/2)}{(2x_1+3)} \frac{{}_2F_1(1, x_1+x_2+2; x_1+5/2; 1/2)}{{}_2F_1(1, x_1+x_2+1; x_1+3/2; 1/2)}$$

et

$$E_{\hat{q}_\pi}[Y_2; \mathbf{x}] = \left(\frac{x_1+x_2+1}{2}\right) \frac{{}_2F_1(1, x_1+x_2+2; x_1+3/2; 1/2)}{{}_2F_1(1, x_1+x_2+1; x_1+3/2; 1/2)}.$$

À partir de l'observation $\mathbf{x} = (17, 31)$, on obtient $E_{\hat{q}_\pi}[Y_1; \mathbf{x}] = 17,3212$ et $E_{\hat{q}_\pi}[Y_2; \mathbf{x}] = 31,6788$, ce qui est relativement près des valeurs trouvées au chapitre 3, soit respectivement 17,15 et 31,85.

CONCLUSION

En somme, l'inférence pour le rapport de deux paramètres d'intensité poissonniens $\phi = \lambda/\mu$, qui survient dans de nombreuses applications telles que les files d'attente (abordées notamment par Armero et Bayarri [AB94]) et l'étude de l'efficacité d'une vaccin ou d'un traitement (abordée notamment par Laurent [Lau12], Laurent et Legrand [LL12] de même que Lindley [Lin02]) est certes un sujet qui soulève de multiples questions intéressantes. Dans certains contextes, ce problème d'estimation paramétrique est soumis à des restrictions qui contraignent l'espace paramétrique et qui doivent être considérées lors de la recherche d'estimateurs aux propriétés intéressantes.

Tout d'abord, en adoptant une approche inférentielle bayésienne, l'espace paramétrique contraint vient notamment motiver la sélection de certaines lois *a priori* pour les paramètres. Le choix de lois *a priori* a permis d'explorer plusieurs distributions en lien avec les distributions bêta de type II (lorsqu'il n'y a aucune contrainte paramétrique sur ϕ) et bêta (lorsque ϕ est contraint à l'intervalle $[0, 1]$). De nouvelles représentations sous forme de mélanges ont d'ailleurs pu être construites pour des généralisations de ces lois. Les lois bêta et leurs généralisations se sont avérées ici des choix judicieux, entre autres pour des motifs calculatoires et contextuels, mais ce problème inférentiel avec contrainte paramétrique mériterait certainement d'être étudié à partir de lois *a priori* différentes.

Ensuite, des estimateurs bayésiens issus des lois *a posteriori* ont été analysés selon le cri-

tère du risque fréquentiste. Tant pour les contextes avec et sans contrainte paramétrique, la comparaison d'estimateurs bayésiens avec un estimateur du maximum de vraisemblance modifié a révélé que les estimateurs bayésiens pouvaient avoir de bonnes propriétés fréquentistes. En effet, dans certaines situations variant notamment selon la valeur des paramètres μ et ϕ et des hyperparamètres des lois *a priori*, les estimateurs bayésiens présentaient un risque plus faible que l'estimateur du maximum de vraisemblance modifié. Afin de mieux comprendre ces comportements observés à partir d'exemples numériques, il pourrait être intéressant d'approfondir la structure des fonctions de risque obtenues, peut-être à partir de certaines approximations qui simplifieraient leur forme plutôt complexe, pour voir si certains résultats analytiques peuvent être obtenus. Ou encore, au lieu de considérer le risque en fonction du paramètre d'intérêt ϕ pour μ fixé, étudier (numériquement et analytiquement) son comportement lors de l'estimation simultanée de (μ, ϕ) avec une fonction de perte $(\mu - \hat{\mu})^2 + (\phi - \hat{\phi})^2$.

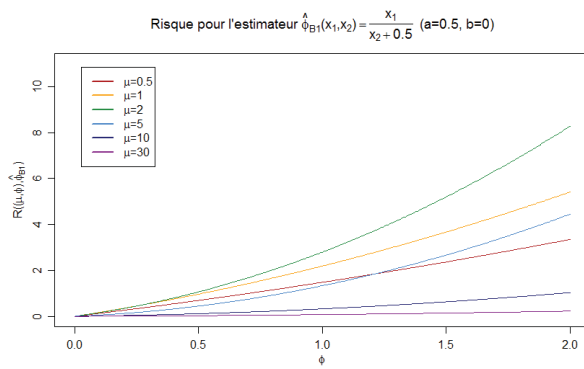
Par ailleurs, lorsque ϕ est contraint à l'intervalle $[0, 1]$, on a pu constater que l'estimateur du maximum de vraisemblance modifié trouvé au chapitre 3 était inadmissible, ce qui motive l'importance de tenir compte d'une contrainte suggérée par un quelconque contexte lors de la recherche d'estimateurs. L'inadmissibilité et l'admissibilité d'estimateurs pour deux paramètres d'intensité poissonniens ordonnés (ce qui revient à restreindre leur ratio sur l'intervalle $[0, 1]$) ont d'ailleurs été étudiées Kushary et Cohen [KC91]. Il pourrait être intéressant de voir si certains résultats obtenus peuvent être adaptés en cas de reparamétrisation telle celle étudiée dans ce travail.

Puis, lors de l'étude de l'inférence prédictive, plusieurs fonctions de masse univariées et bivariées se sont présentées. D'une part, dans le cas général pour ϕ non contraint, Laurent avait déjà introduit la loi de Bailey [Lau12], qui survient comme fonction de masse prédictive bivariée pour de futures observations poissonniennes et qui peut être représenté sous forme de mélange de lois de Poisson. D'autre part, dans le cas où l'espace

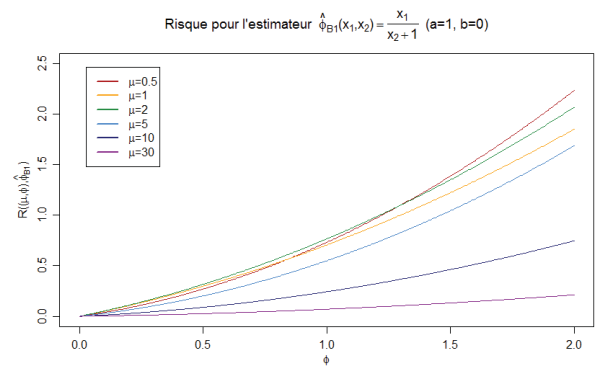
paramétrique de ϕ est contraint à l'intervalle $[0, 1]$, l'idée est venue de généraliser la loi de Bailey dans la perspective de mieux comprendre la fonction de masse bivariée prédictive obtenue. Bien qu'une correspondance entre la loi de Bailey généralisée et la fonction de masse prédictive bivariée n'a pu être établie, cette généralisation a pris son importance lors de l'approfondissement de la structure des lois univariées qui en découlent. En effet, une des lois marginales prédictives appartient à la même famille de lois que les lois marginales de la loi de Bailey généralisée, qui peuvent être représentées sous forme de mélange de lois binomiales négatives pondérées.

Enfin, outre celles mentionnées précédemment, plusieurs avenues pour le problème inférentiel du rapport de deux paramètres d'intensité poissonniens sont encore à explorer. L'approfondissement des liens avec les files d'attente pourrait mener à de nouveaux résultats. En outre, dans le contexte suggéré par Laurent [Lau12] et Laurent et Legrand [LL12] portant sur l'estimation de l'efficacité relative d'un vaccin, il est raisonnable de penser que ce paramètre peut être influencé par certaines caractéristiques intrinsèques aux individus participant à l'étude. Ainsi, on peut se demander comment il serait possible d'introduire des covariables au modèle. La modélisation proposée par Karlis et Meligkotsidou dans [KM05] serait une piste à envisager, eux qui proposent d'ailleurs des méthodes d'estimation basées sur le maximum de vraisemblance et d'autres selon une approche bayésienne. Dans un autre ordre d'idées, les lois de Poisson ne s'avèrent pas nécessairement la meilleure famille de lois à utiliser pour décrire le comportement des données. Le manque de contrôle de leur variabilité est d'ailleurs un problème bien connu des lois de Poisson puisque la variance correspond également au paramètre d'intensité poissonnien (espérance). Or, d'autres modèles discrets offrant une plus grande variabilité, comme les lois binomiales négatives, pourraient être une alternative intéressante. Cette famille de lois, construites à partir d'un mélange de lois de Poisson avec un paramètre d'intensité distribué selon une loi Gamma, ouvriraient fort probablement de nouvelles perspectives.

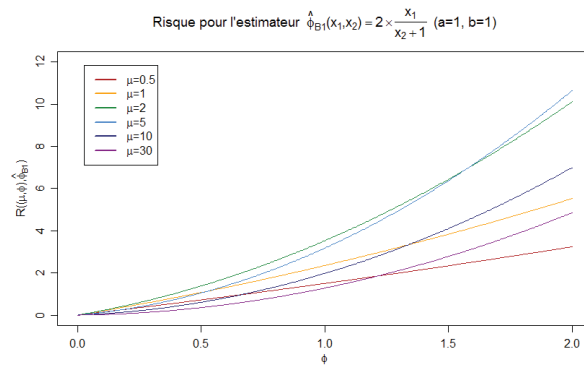
ANNEXE A



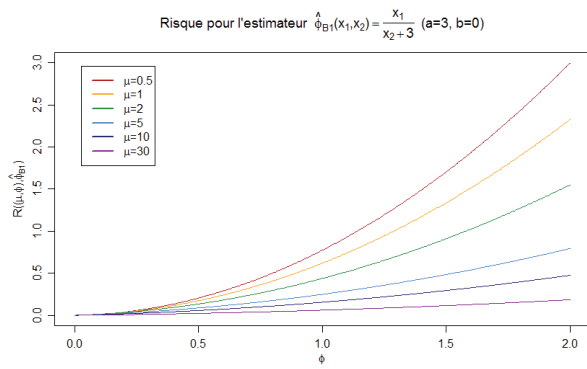
(a) Graphique 1 : $(a, b) = (0.5, 0)$



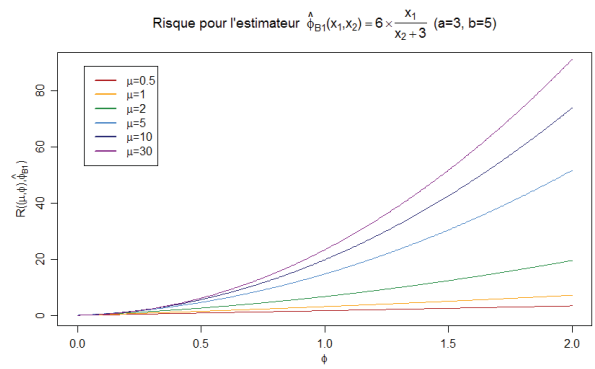
(b) Graphique 2 : $(a, b) = (1, 0)$



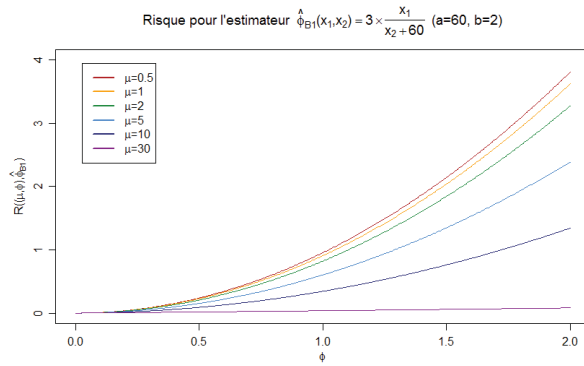
(c) Graphique 3 : $(a, b) = (1, 1)$



(d) Graphique 4 : $(a, b) = (3, 0)$

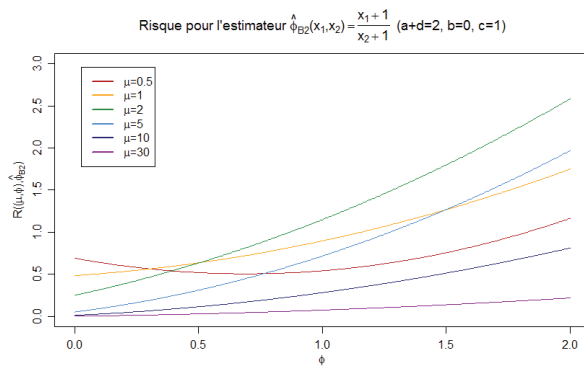


(e) Graphique 5 : $(a, b) = (3, 5)$

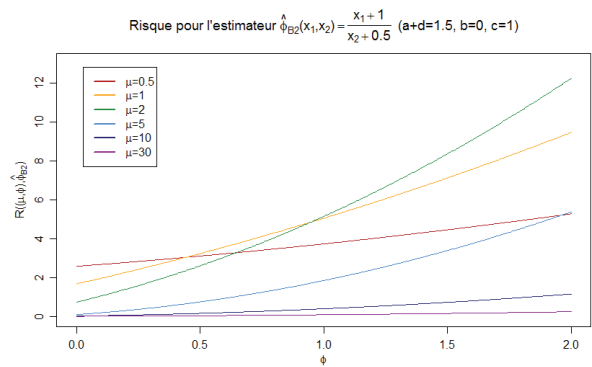


(f) Graphique 6 : $(a, b) = (60, 2)$

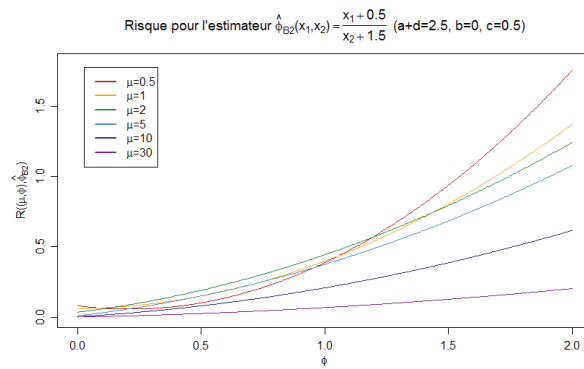
ANNEXE B



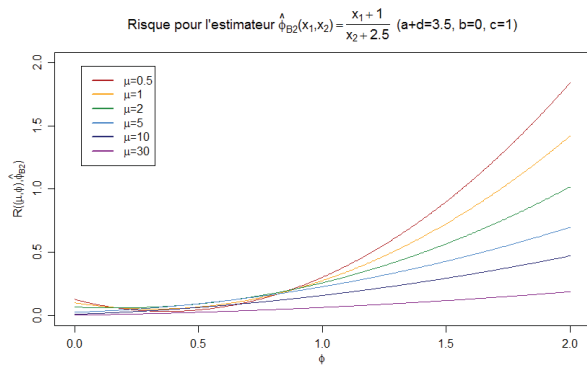
(a) Graphique 1 : $(a + d, b, c) = (2, 0, 1)$



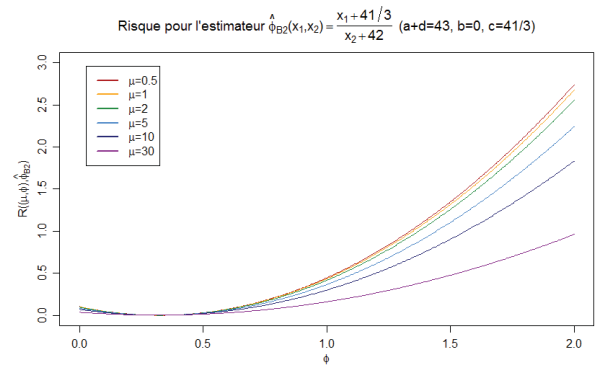
(b) Graphique 2 : $(a + d, b, c) = (1.5, 0, 1)$



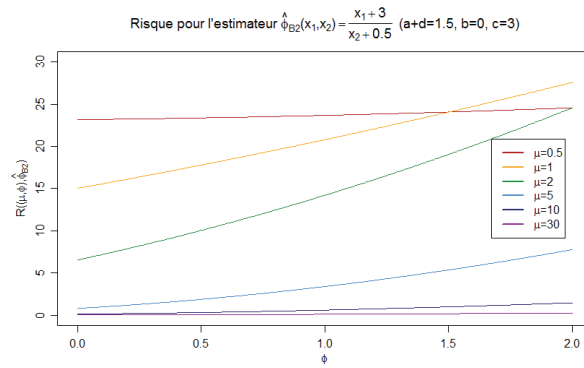
(c) Graphique 3 : $(a + d, b, c) = (2.5, 0, 0.5)$



(d) Graphique 4 : $(a + d, b, c) = (3.5, 0, 1)$



(e) Graphique 5 : $(a + d, b, c) = (43, 0, 41/3)$



(f) Graphique 6 : $(a + d, b, c) = (1.5, 0, 3)$

Bibliographie

- [AAR99] George E. ANDREWS, Richard ASKEY et Ranjan ROY : *Special Functions*. Cambridge University Press, 1999. Cambridge Books Online.
- [AB94] C. ARMERO et M. J. BAYARRI : Prior Assessments for Prediction in Queues. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 43(1): 139–153, 1994.
- [AC98] Alan AGRESTI et Brent A. COULL : Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions. *The American Statistician*, 52(2):119–126, 1998.
- [AD75] J. AITCHISON et I. R. DUNSMORE : *Statistical Prediction Analysis*. Cambridge University Press, 1975. Cambridge Books Online.
- [AGG90] Richard ARRATIA, Larry GOLDSTEIN et Louis GORDON : Poisson Approximation and the Chen-Stein Method. *Statistical Science*, 5(4):403–424, 1990.
- [Ait75] J. AITCHISON : Goodness of prediction fit. *Biometrika*, 62(3):547–554, 1975.
- [Bai35] Wilfrid Norman BAILEY : Generalized Hypergeometric Series. *Cambridge University Press*, 1935.
- [Ber85] James O. BERGER : *Statistical Decision Theory and Bayesian Analysis*. Springer Science & Business Media New York, 1985.

- [BGX08] Lawrence D. BROWN, Edward I. GEORGE et Xinyi XU : Admissible Predictive Density Estimation. *The Annals of Statistics*, 36(3):1156–1170, 06 2008.
- [BHJ92] Andrew D. BARBOUR, Lars HOLST et Svante JANSON : *Poisson Approximation*. Clarendon Press Oxford, 1992.
- [Bir53] Allan BIRNBAUM : Some Procedures for Comparing Poisson Processes or Populations. *Biometrika*, 40(3/4):447–449, 1953.
- [Bly86] Colin R. BLYTH : Approximate Binomial Confidence Limits. *Journal of the American Statistical Association*, 81(395):843–855, 1986.
- [Bol62] L. N. BOL'SHEV : Comparison of Parameters of Poisson Distributions. *Theory of Probability & Its Applications*, 7(1):113–114, 1962.
- [BS83] Colin R. BLYTH et Harold A. STILL : Binomial Confidence Intervals. *Journal of the American Statistical Association*, 78(381):108–116, 1983.
- [CB02] George CASELLA et Roger L. BERGER : *Statistical Inference*, volume 2. Duxbury, 2002.
- [Cha52] Douglas G. CHAPMAN : On Tests and Estimates for the Ratio of Poisson Means. *Annals of the Institute of Statistical Mathematics*, 4(1):45–49, 1952.
- [CHT10] Robert D. COUSINS, Kathryn E. HYMES et Jordan TUCKER : Frequentist Evaluation of Intervals Estimated for a Binomial Parameter and for the Ratio of Poisson Means. *Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment*, 612(2):388–398, 2010.
- [CN84] James J. CHEN et Melvin R. NOVICK : Bayesian Analysis for Binomial Models with Generalized Beta Prior Distributions. *Journal of Educational and Behavioral Statistics*, 9(2):163–175, 1984.
- [Cou98] R.D. COUSINS : Improved Central Confidence Intervals for the Ratio of Poisson Means. *Nuclear Instruments and Methods in Physics Research Section*

- A : Accelerators, Spectrometers, Detectors and Associated Equipment*, 417(2-3):391–399, 1998.
- [CP34] Charles J. CLOPPER et Egon S. PEARSON : The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika*, 26(4):404–413, 1934.
- [CS04] I. CSISZÁR et P.C. SHIELDS : *Information Theory and Statistics : A Tutorial*. Foundations and trends in communications and information theory. Now Publishers, 2004.
- [Csi63] I. CSISZÁR : Eine informationstheoretische Ungleichung und ihre anwendung auf den Beweis der ergodizität von Markoffschen Ketten. *Publ. Math. Inst. Hungar. Acad.*, 8:95–108, 1963.
- [CZ75] M. Lawrence CLEVENSON et James V. ZIDEK : Simultaneous Estimation of the Means of Independent Poisson Laws. *Journal of the American Statistical Association*, 70(351a):698–705, 1975.
- [DLM] NIST Digital Library of Mathematical Functions. <http://dlmf.nist.gov/13>, Release 1.0.10 of 2015-08-07. Online companion to [OLBC10].
- [Dod08] Yadolah DODGE : *The Concise Encyclopedia of Statistics*. Springer Science & Business Media, 2008.
- [Eul69] Leonhard EULER : Institutiones Calculi Integralis, II. *Opera Omnia*, 12:221–230, 1769.
- [Ext76] Harold EXTON : *Multiple Hypergeometric Functions and Applications*. E. Horwood, 1976.
- [FK95] W. Dana FLANDERS et David G. KLEINBAUM : Basic Models for Disease Occurrence in Epidemiology. *International Journal of Epidemiology*, 24(1):1–7, 1995.

- [GLX06] Edward I. GEORGE, Feng LIANG et Xinyi XU : Improved Minimax Predictive Densities under Kullback-Leibler Loss. *The Annals of Statistics*, 34(1):78–91, 2006.
- [GN04] Arjun K. GUPTA et Saralees NADARAJAH : *Handbook of Beta Distribution and its Applications*. CRC Press, 2004.
- [Hai67] Frank A. HAIGHT : *Handbook of the Poisson distribution*. New York : Wiley, 1967. Collection : Operations Research Society of America. Publications in operations research. no. 11.
- [Hol64] Philip HOLGATE : Estimation for the Bivariate Poisson Distribution. *Biometrika*, 51(1-2):241–245, 1964.
- [Jef31] Harold JEFFREYS : *Scientific Inference*. Cambridge University Press, 1931.
- [Jef61] Harold JEFFREYS : *The Theory of Probability*. Oxford University Press, 1961.
- [JKB94] Norman L. JOHNSON, Samuel KOTZ et Narayanaswamy BALAKRISHNAN : *Continuous Univariate Distributions*, volume 1-2. New York : John Wiley & Sons, 2 édition, 1994.
- [JKB97] Norman L. JOHNSON, Samuel KOTZ et N. BALAKRISHNAN : *Discrete Multivariate Distributions*. Wiley, New York, 1997.
- [JKK05] Norman L JOHNSON, Adrienne W KEMP et Samuel KOTZ : *Univariate Discrete Distributions*. John Wiley & Sons, 3 édition, 2005.
- [Kat09] Kengo KATO : Improved Prediction for a Multivariate Normal Distribution with Unknown Mean and Variance. *Annals of the Institute of Statistical Mathematics*, 61(3):531–542, 2009.
- [KC91] Debashis KUSHARY et Arthur COHEN : Estimation of Ordered Poisson Parameters. *Sankhyā : The Indian Journal of Statistics, Series A (1961-2002)*, 53(3):334–356, 1991.

- [Kem68] Adrienne W. KEMP : A Wide Class of Discrete Distributions and the Associated Differential Equations. *Sankhyā : The Indian Journal of Statistics, Series A (1961-2002)*, 30(4):401–410, 1968.
- [KK92] Subrahmaniam KOCHERLAKOTA et Kathleen KOCHERLAKOTA : *Bivariate discrete distributions*, volume 132 de *Statistics : A Series of Textbooks and Monographs*. Marcel Dekker Inc., 1992.
- [KL51] Solomon KULLBACK et Richard A. LEIBLER : On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.
- [KM05] Dimitris KARLIS et Loukia MELIGKOTSIDOU : Multivariate Poisson regression with covariance structure. *Statistics and Computing*, 15(4):255–265, 2005.
- [Kom01] Fumiyasu KOMAKI : A Shrinkage Predictive Distribution for Multivariate Normal Observables. *Biometrika*, 88(3):859–864, 2001.
- [Lap12] Pierre Simon marquis de LAPLACE : *Théorie analytique des probabilités*. Ve. Courcier, Paris, 1812.
- [Lau12] Stéphane LAURENT : Some Poisson mixtures distributions with a hyperscale parameter. *Brazilian Journal of Probability and Statistics*, 26(3):265–278, 08 2012.
- [LC98] Erich L. LEHMANN et George CASELLA : *Theory of Point Estimation*, volume 31. Springer Science & Business Media, 1998.
- [Lin02] Dennis V. LINDLEY : A Holistic Analysis of Poisson Data with Application to a Trial of Selenium and Cancer Deaths. *Sankhyā : The Indian Journal of Statistics, Series A (1961-2002)*, 64(3):722–739, 2002.
- [LL12] Stéphane LAURENT et Catherine LEGRAND : A Bayesian framework for the ratio of two Poisson rates in the context of vaccine efficacy trials. *ESAIM : Probability and Statistics*, 16:375–398, 2012.

- [LN82] David L. LIBBY et Melvin R. NOVICK : Multivariate Generalized Beta Distributions with Applications to Utility Assessment. *Journal of Educational Statistics*, 7(4):271–294, 1982.
- [Mar01] Éric MARCHAND : Point Estimation of the Coefficient of Determination. *Statistics & Risk Modeling*, 19(2):137–154, 2001.
- [Mor63] Tetsuzo MORIMOTO : Markov Processes and the H-Theorem. *Journal of the Physical Society of Japan*, 18(3):328–331, 1963.
- [MS04] Eric MARCHAND et William E. STRAWDERMAN : *Estimation in restricted parameter spaces : A review*, volume 45 de *Lecture Notes–Monograph Series*, pages 21–44. Institute of Mathematical Statistics, Beachwood, Ohio, USA, 2004.
- [OLBC10] Frank W.J. OLVER, Daniel W. LOZIER, Ronald F. BOISVERT et Charles W. CLARK : *NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, NY, 2010. Print companion to [DLM].
- [PB00] Robert M. PRICE et Douglas G. BONETT : Estimating the ratio of two Poisson rates. *Computational Statistics & Data Analysis*, 34(3):345–356, 2000.
- [PGD89] T. PHAM-GIA et Q.P. DUONG : The Generalized Beta-and F-distributions in Statistical Modelling. *Mathematical and Computer Modelling*, 12(12):1613–1625, 1989.
- [RWD88] T. ROBERTSON, F.T. WRIGHT et R.L. DYKSTRA : *Order Restricted Statistical Inference*. Chichester (UK) etc. : John Wiley & Sons, 1988.
- [Sea13] James B. SEABORN : *Hypergeometric Functions and Their Applications*, volume 8. Springer Science & Business Media, 2013.
- [Sla66] L. J. SLATER : Generalized Hypergeometric Functions. *Cambridge University Press*, 1966.

- [vE06] Constance van EEDEN : Restricted Parameter Space Estimation Problems.
Springer New York, 2006.
- [YB96] Ruoyong YANG et James O. BERGER : *A Catalog of Noninformative Priors*.
Institute of Statistics and Decision Sciences, Duke University, 1996.